XXI. ULUSAL VE IV. ULUSLARARASI

BİYOİSTATİSTİK KONGRESİ

XXIst NATIONAL & IVth INTERNATIONAL BIOSTATISTICS CONGRESS

26-29 EKİM/OCTOBER 2019

Belconti Resort Otel, Belek

ANTALYA / TURKEY

Kongre Başkanları / Congress Chairs

Doç. Dr. S. Kenan KÖSE
Ankara Üniversitesi

Prof. Dr. Erdem KARABULUT
Biyoistatistik Derneği Başkanı

Düzenleme Kurulu / Organization Committee

Prof. Dr. İsmet DOĞAN
Afyonkarahisar Sağlık Bilimleri Üniversitesi

Prof. Dr. Atilla Halil ELHAN
Ankara Üniversitesi

Doç. Dr. Gökmen ZARARSIZ
Erciyes Üniversitesi

Bilimsel Sekretarya / Scientific Secretariat

Dr. Osman DAĞ
osman.dag@outlook.com

Dr. Dinçer GÖKSÜLÜK
dincer.goksuluk@gmail.com

Uzm. İrem KAR
iremkar.biostat@gmail.com

Uzm. Batuhan BAKIRARAR
batuhan_bakirarar@hotmail.com

Organizasyon Sekretarya / Organization Secretariat

Coraltravel

Sercan BAŞ
T: +90 212 247 46 46 F: +90 212 225 72 64
E: sercan.bas@coraltatil.com
www.coraltatil.com

E: info@biyoistatistikkongresi.org
www.biyoistatistikkongresi.org

www.biyoistatistikkongresi.org

XXI. ULUSAL
ve
IV. ULUSLARARASI
BİYOİSTATİSTİK KONGRESİ
BİLDİRİ ÖZETLERİ KİTABI

Editör
Prof. Dr. Erdem Karabulut

Yardımcı Editörler
Dr. Osman Dağ
Dr. Dinçer Göksülük
Uzm. İrem Kar
Uzm. Batuhan Bakırarar

26-29 Ekim 2019
Belek / ANTALYA

XXI. NATIONAL and IV. INTERNATIONAL BIOSTATISTICS CONGRESS BOOK of ABSTRACTS

Editor in Chief
Prof. Erdem Karabulut

Associate Editors
Osman Dağ, *PhD*
Dinçer GÖKSÜLÜK, *PhD*
İrem KAR, *MSc*
Batuhan Bakırarar, *MSc*

October 26-29, 2019
Belek / ANTALYA

Kongre Onursal Başkanı

Prof. Dr. Kadir SÜMBÜLOĞLU	H.Ü. Tıp Fak. Biyoistatistik AD E. Öğretim Üyesi/Biyoistatistik Derneği Onursal Başkanı

Kongre Eşbaşkanları

Prof. Dr. Erdem KARABULUT	Hacettepe Üniversitesi
Doç. Dr. S. Kenan KÖSE	Ankara Üniversitesi

Düzenleme Kurulu

Prof. Dr. Erdem KARABULUT	Hacettepe Üniversitesi	Başkan
Doç. Dr. S. Kenan KÖSE	Ankara Üniversitesi	Başkan
Prof. Dr. İsmet DOĞAN	Afyonkarahisar Sağlık Bilimleri Üniversitesi	Üye
Prof. Dr. Atilla Halil ELHAN	Ankara Üniversitesi	Üye
Doç. Dr. Gökmen ZARARSIZ	Roche Diagnostics GmbH, Germany	Üye
Dr. Osman DAĞ	Hacettepe Üniversitesi	Bilimsel Sekreterya
Dr. Dinçer GÖKSÜLÜK	Hacettepe Üniversitesi	Bilimsel Sekreterya
Uzm. İrem KAR	Ankara Üniversitesi	Bilimsel Sekreterya
Uzm. Batuhan BAKIRARAR	Ankara Üniversitesi	Bilimsel Sekreterya

Honorary President of the Congress

Prof. Kadir SÜMBÜLOĞLU	Hacettepe University, Emeritus Faculty Member of Biostatistics / Honorary President of Biostatistics Association

Congress Co-chairs

Prof. Erdem KARABULUT	Hacettepe University
Assoc. Prof. S. Kenan KÖSE	Ankara University

Organizing Committee

Prof. Erdem KARABULUT	Hacettepe University	Co-chairs
Assoc. Prof. S. Kenan KÖSE	Ankara University	Co-chairs
Prof. İsmet DOĞAN	Afyonkarahisar Health Sciences University	Members
Prof. Atilla Halil ELHAN	Ankara University	Members
Assoc. Prof. Gökmen ZARARSIZ	Roche Diagnostics GmbH, Germany	Members
Osman DAĞ, PhD	Hacettepe University	Scientific secretariat
Dinçer GÖKSÜLÜK, PhD	Hacettepe University	Scientific secretariat
İrem KAR, MSc	Ankara University	Scientific secretariat
Batuhan BAKIRARAR, MSc	Ankara University	Scientific secretariat

Bilimsel Danışma Kurulu

Prof. Dr. Ahmet ÖZTÜRK	Erciyes Üniversitesi
Prof. Dr. Alan TENNANT	University of Lucerne, Switzerland
Prof. Dr. Atilla Halil ELHAN	Ankara Üniversitesi
Prof. Dr. Ayşe Canan YAZICI GÜVERCİN	Aydın Üniversitesi
Prof. Dr. Bahar TAŞDELEN	Mersin Üniversitesi
Prof. Dr. Beyza AKDAĞ ŞAHİN	Pamukkale Üniversitesi
Prof. Dr. Bülent ÇELİK	Gazi Üniversitesi
Prof. Dr. Cemil ÇOLAK	İnönü Üniversitesi
Prof. Dr. Erdem KARABULUT	Hacettepe Üniversitesi
Prof. Dr. Ergun KARAAĞAOĞLU	Hacettepe Üniversitesi
Prof. Dr. Ersin ÖĞÜŞ	Başkent Üniversitesi
Prof. Dr. Ertuğrul ÇOLAK	Eskişehir Osmangazi Üniversitesi
Prof. Dr. Fezan ŞAHİN MUTLU	Eskişehir Osmangazi Üniversitesi
Prof. Dr. Gülşah SEYDAOĞLU	Çukurova Üniversitesi
Prof. Dr. Hafize SEZER	Aydın Üniversitesi
Prof. Dr. İlker ERCAN	Uludağ Üniversitesi
Prof. Dr. İlker ETİKAN	Yakındoğu Üniversitesi, K.K.T.C.
Prof. Dr. İmran KURT ÖMÜRLÜ	Adnan Menderes Üniversitesi
Prof. Dr. İsmet DOĞAN	Afyonkarahisar Sağlık Bilimleri Üniversitesi
Prof. Dr. Mehtap AKÇİL OK	Başkent Üniversitesi
Prof. Dr. Melody GHAHRAMANİ	Winnipeg Üniversitesi, Canada
Prof. Dr. Meriç YAVUZ ÇOLAK	Başkent Üniversitesi
Prof. Dr. Mevlut TÜRE	Adnan Menderes Üniversitesi
Prof. Dr. Ömer SATICI	Dicle Üniversitesi
Prof. Dr. Peter W. ROSE	University of California, USA
Prof. Dr. Pınar ÖZDEMİR	Hacettepe Üniversitesi
Prof. Dr. Reha ALPAR	Hacettepe Üniversitesi
Prof. Dr. Saim YOLOĞLU	İnönü Üniversitesi
Prof. Dr. Sanjib BASU	University of Illinois, USA
Prof. Dr. Setenay Kevser DİNÇER ÖNER	Osmangazi Üniversitesi
Prof. Dr. Seval KUL	Gaziantep Üniversitesi
Prof. Dr. Sıddık KESKİN	Van Yüzüncü Yıl Üniversitesi
Prof. Dr. Vildan SÜMBÜLOĞLU	Sanko Üniversitesi
Prof. Dr. Yasemin YAVUZ	Ankara Üniversitesi
Prof. Dr. Yusuf ÇELİK	Biruni Üniversitesi
Prof. Dr. Z. Nazan ALPARSLAN	Çukurova Üniversitesi
Prof. Dr. Zeki AKKUŞ	Dicle Üniversitesi
Doç. Dr. Alex R. de LEON	University of Calgary, Canada

Doç. Dr. Beyza DOĞANAY ERDOĞAN	Ankara Üniversitesi
Doç. Dr. Canan BAYDEMİR	Kocaeli Üniversitesi
Doç. Dr. Cengiz BAL	Eskişehir Osmangazi Üniversitesi
Doç. Dr. Deniz SIĞIRLI	Uludağ Üniversitesi
Doç. Dr. Derya GÖKMEN	Ankara Üniversitesi
Doç. Dr. Ferhan ELMALI	İzmir Katip Çelebi Üniversitesi
Doç. Dr. Gökhan OCAKOĞLU	Uludağ Üniversitesi
Doç. Dr. Gökmen ZARARSIZ	Roche Diagnostics GmbH, Germany
Doç. Dr. Gülhan OREKİCİ TEMEL	Mersin Üniversitesi
Doç. Dr. Güven ÖZKAYA	Uludağ Üniversitesi
Doç. Dr. Harika GÖZÜKARA BAĞ	İnönü Üniversitesi
Doç. Dr. İlker ÜNAL	Çukurova Üniversitesi
Doç. Dr. Jale KARAKAYA	Hacettepe Üniversitesi
Doç. Dr. Prof. Lynn SLEEPER	Harvard Medical School, USA
Doç. Dr. M. Agah TEKİNDAL	Selçuk Üniversitesi
Doç. Dr. Nurhan DOĞAN	Afyonkarahisar Sağlık Bilimleri Üniversitesi
Doç. Dr. Recai YÜCEL	University at Albany-SUNY, USA
Doç. Dr. S. Kenan KÖSE	Ankara Üniversitesi
Doç. Dr. Selcen YÜKSEL	Yıldırım Beyazıt Üniversitesi
Doç. Dr. Semra ERDOĞAN	Mersin Üniversitesi
Doç. Dr. Şengül CANGÜR	Düzce Üniversitesi
Doç. Dr. Ünal ERKORKMAZ	Sakarya Üniversitesi
Doç. Dr. Yaşar SERTDEMİR	Çukurova Üniversitesi
Yrd. Doç. Dr. Duchwan RYU	Northen Illinois University, USA
Yrd. Doç. Dr. Umut ÖZBEK	Icahn School of Medicine, Mount Sinai, USA
Dr. Bernd KLAUS	European Mol. Bio. Lab.,Germany

Scientific Advisory Board

Prof. Ahmet ÖZTÜRK	Erciyes University
Prof. Alan TENNANT	University of Lucerne, Switzerland
Prof. Atilla Halil ELHAN	Ankara University
Prof. Ayşe Canan YAZICI GÜVERCİN	Aydın University
Prof. Bahar TAŞDELEN	Mersin University
Prof. Beyza AKDAĞ ŞAHİN	Pamukkale University
Prof. Bülent ÇELİK	Gazi University
Prof. Cemil ÇOLAK	İnönü University
Prof. Erdem KARABULUT	Hacettepe University
Prof. Ergun KARAAĞAOĞLU	Hacettepe University
Prof. Ersin ÖĞÜŞ	Başkent University
Prof. Ertuğrul ÇOLAK	Eskişehir Osmangazi University
Prof. Fezan ŞAHİN MUTLU	Eskişehir Osmangazi University
Prof. Gülşah SEYDAOĞLU	Çukurova University
Prof. Hafize SEZER	Aydın University
Prof. İlker ERCAN	Uludağ University
Prof. İlker ETİKAN	Yakındoğu University, K.K.T.C.
Prof. İmran KURT ÖMÜRLÜ	Adnan Menderes University
Prof. İsmet DOĞAN	Afyonkarahisar Health Sciences University
Prof. Mehtap AKÇİL OK	Başkent University
Prof. Melody GHAHRAMANİ	University of Winnipeg, Canada
Prof. Meriç YAVUZ ÇOLAK	Başkent University
Prof. Mevlut TÜRE	Adnan Menderes University
Prof. Ömer SATICI	Dicle University
Prof. Peter W. ROSE	University of California, USA
Prof. Pınar ÖZDEMİR	Hacettepe University
Prof. Reha ALPAR	Hacettepe University
Prof. Saim YOLOĞLU	İnönü University
Prof. Sanjib BASU	University of Illinois, USA
Prof. Setenay Kevser DİNÇER ÖNER	Osmangazi University
Prof. Seval KUL	Gaziantep University
Prof. Sıddık KESKİN	Van Yüzüncü Yıl University
Prof. Vildan SÜMBÜLOĞLU	Sanko University
Prof. Yasemin YAVUZ	Ankara University
Prof. Yusuf ÇELİK	Biruni University
Prof. Z. Nazan ALPARSLAN	Çukurova University
Prof. Zeki AKKUŞ	Dicle University
Assoc. Prof. Alex R. de LEON	University of Calgary, Canada

Assoc. Prof. Beyza DOĞANAY ERDOĞAN	Ankara University
Assoc. Prof. Canan BAYDEMİR	Kocaeli University
Assoc. Prof. Cengiz BAL	Eskişehir Osmangazi University
Assoc. Prof. Deniz SIĞIRLI	Uludağ University
Assoc. Prof. Derya GÖKMEN	Ankara University
Assoc. Prof. Ferhan ELMALI	İzmir Katip Çelebi University
Assoc. Prof. Gökhan OCAKOĞLU	Uludağ University
Assoc. Prof. Gökmen ZARARSIZ	Roche Diagnostics GmbH, Germany
Assoc. Prof. Gülhan OREKİCİ TEMEL	Mersin University
Assoc. Prof. Güven ÖZKAYA	Uludağ University
Assoc. Prof. Harika GÖZÜKARA BAĞ	İnönü University
Assoc. Prof. İlker ÜNAL	Çukurova University
Assoc. Prof. Jale KARAKAYA	Hacettepe University
Assoc. Prof. Lynn SLEEPER	Harvard Medical School, USA
Assoc. Prof. M. Agah TEKİNDAL	Selçuk University
Assoc. Prof. Nurhan DOĞAN	Afyonkarahisar Health Sciences University
Assoc. Prof. Recai YÜCEL	University at Albany-SUNY, USA
Assoc. Prof. S. Kenan KÖSE	Ankara University
Assoc. Prof. Selcen YÜKSEL	Yıldırım Beyazıt University
Assoc. Prof. Semra ERDOĞAN	Mersin University
Assoc. Prof. Şengül CANGÜR	Düzce University
Assoc. Prof. Ünal ERKORKMAZ	Sakarya University
Assoc. Prof. Yaşar SERTDEMİR	Çukurova University
Assist. Prof. Duchwan RYU	Northen Illinois University, USA
Assist. Prof. Umut ÖZBEK	Icahn School of Medicine, Mount Sinai, USA
Ph.D. Bernd KLAUS	European Mol. Bio. Lab.,Germany

ÖNSÖZ

Sayın Hocalarım ve Sevgili Meslektaşlarım,

Biyoistatistik Derneği tarafından düzenlenen XXI. Ulusal ve IV. Uluslararası Biyoistatistik kongresini 26-29 Ekim 2019 tarihleri arasında Antalya, Belconti Resort Hotel Belek'te gerçekleştirecek olmanın sevincini ve heyecanını yaşıyoruz.

Biyoistatistik; tıp ve sağlık bilimleri alanında yapılacak çalışmaların planlanması ve çalışma bulgularının değerlendirilip bir karara varılması süreçleri için istatistiksel yöntemleri kullanan ve bu süreçlere yönelik yeni yöntemler geliştiren bir bilim alanıdır. Son yıllarda sağlık alanında veri kalitesi, miktarı ve çeşitliliğindeki artış, klasik yöntemlerin yerine daha yeni, daha karmaşık ve sorunların çözümüne odaklı biyoistatistiksel yaklaşımlara olan gereksinimi artırmıştır. Söz konusu gereksinimler; tıp bilimleri ile biyoistatistik anabilim dallarının daha da iç içe geçmesini sağlamıştır.

Kişiselleştirilmiş tıp, tanı, prognoz ve tedavi etkinliklerinin planlanması, koruyucu hekimlik uygulamalarının geliştirilmesi, risk faktörlerinin saptanması, ilişkilerinin belirlenmesi, uyumlulukların ortaya konulması, sağlık hizmetlerinin planlanması, sunumu ve sağlık politikalarının geliştirilmesi gibi birçok konuda hekimler ve diğer sağlık çalışanları, birlikte çalışacağı Biyoistatistik uzmanı arayışındadır. Bu nedenle, yetkin biyoistatistik uzmanlarına duyulan gereksinim giderek artmaktadır.

Kongremizde tüm konuları içerecek şekilde bilimsel açıdan zengin bir program oluşturmayı hedefledik. Kongrenin ilk günü yine katılımcıların istekleri doğrultusunda içeriği belirlenen temel ve ileri düzey kurslara ayrılmıştır. Kongremizin, araştırmacıların deneyimlerini paylaşmalarına, biyoistatistik alanındaki güncel konuları takip etmelerine, katılacakları kurslar, konferanslar ve sözlü/poster sunumları ile eksikliğini hissettikleri konularda daha donanımlı hale gelmelerine olanak sağlayacağını umuyoruz.

Biyoistatistik topluluğunun değerli üyelerini ve araştırmacıları kongremizde aramızda görmekten büyük onur ve mutluluk duyacağız. Tüm bilim insanlarını kongremize davet ederken, düzenleme komitesi adına hepinize en içten sevgi ve saygılarımı sunarız.

Kongre Düzenleme Kurulu Adına,
Doç. Dr. S. Kenan KÖSE
Prof. Dr. Erdem KARABULUT
Kongre Başkanları

PREFACE

Dear Colleagues,

We are experiencing the joy and excitement of being the XXI. National and IV. International Biostatistics Congress organized by the Biostatistics Association to be held in Belconti Resort Hotel Belek, Antalya on October 26-29, 2019.

In our congress, we aimed to create a program rich in all aspects of science. On the first day of the congress, basic and advanced courses which are determined according to the requests of the participants are arranged. We hope that our congress will allow researchers to share their experiences, to follow current issues in biostatistics, and to become better equipped with the courses, conferences and oral / poster presentations they will attend.

We will be honored and happy to see the esteemed members of the biostatistics community and health researchers in our congress. While we invite all scientists to our congress, we offer our sincere respect to all of you on behalf of the organizing committee.

On behalf of the Organizing Committee,

Assoc. Prof. S. Kenan KÖSE
Prof. Erdem KARABULUT
Congress Chairs

BİLİMSEL PROGRAM / SCIENTIFIC PROGRAM

XXI. Ulusal ve IV. Uluslararası Biyoistatistik Kongresi
XXI. National and IV. International Biostatistics Congress
Ekim / October 26 – 29, 2019 - Belconti Hotel, Belek/Antalya

Kongre Programı / Congress Program

1. Gün / Day 1 (27 Ekim/October): Ana Salon / Main Hall

	Saat/Hour		Sunum Başlığı/Presentation Title
	09:30	**10:00**	**Açılış Konuşmaları / Opening Speeches**
***Oturum Başkanları / Chairs**: Prof. A. Ergun KARAAĞAOĞLU, Assoc. Prof. Nurhan DOĞAN*			
Oturum1 / Session 1 (Ödül Oturumu / Award Session)	**10:00**	**10:15**	An Interactive R Shiny Application to Facilitate Comparing Classification Methods: ClasComp *Uğur TOPRAK*
	10:15	**10:30**	Dairesel(Circular) Verinin Tanıtılması ve Gerçek Veri Seti Üzerine Uygulama *Hülya BİNOKAY*
	10:30	**10:45**	Aynı Veri Setleri Üzerinde Bazı Etki Büyüklüğü Hesaplama Yöntemlerinin Tahmin Başarılarının Karşılaştırılması *Duygu KORKMAZ*
	10:45	**11:00**	ULSAM Veri Setinde Tek Nükleotid Polimorfizmi Veri ile Makine Öğrenmesi Yöntemlerinin Kullanımı *Ragıp Onur ÖZTORNACI*
	11:00	**11:15**	**Kahve Arası / Coffee Break**
***Oturum Başkanları / Chairs**: Prof. Meriç YAVUZ ÇOLAK, Assoc. Prof. İlker ÜNAL, Assist. Prof. Osman DEMİR*			
Oturum 2 / Session 2 (Bioinformatics, Omics Data Analysis)	**11:15**	**11:30**	RNA-Dizileme Verilerinin Kesikli Dağılıma Dayalı Yöntemler ile Sınıflandırılması *Dinçer GÖKSÜLÜK*
	11:30	**11:45**	Genetik ve Klinik Verilerin Birleştirilmesinde Regresyon Modeli Artık Değerlerinin Kullanımı *Erdal COŞGUN*
	11:45	**12:00**	Mikrodizi Kanser Verilerinin Öznitelik Seçimi ve Sınıflandırılması *Özlem ARIK*
	12:00	**12:15**	Makine Öğrenmesi ile Genetik Varyasyon Çeşitlerinin Tahminlenmesi *Su ÖZGÜR*
	12:15	**12:30**	Genom-Boyu İlişki Çalışmalarında Makine Öğrenimi Yöntemleri ve Derin Öğrenme Yönteminin Farklı Örnek Genişliklerinde Performanslarının Değerlendirilmesi *Ragıp Onur ÖZTORNACI*
	12:30	**14:00**	**Öğle Yemeği Arası / Lunch Break**
***Oturum Başkanları/Chairs**:*			
I.T.	**14:00**	**14:45**	Dünyada ve Türkiye'de Biyoistatistiğin Gelişimi *Kadir SÜMBÜLOĞLU*

	14:45	**15:00**	**Kahve Arası / Coffee Break**
***Oturum Başkanları/Chairs**: Prof. Seval KUL, Assoc. Prof. S. Kenan KÖSE, Assist. Prof. Can ATEŞ*			
Oturum 3 / Session 3 (Statistical Modeling - II)	**15:00**	**15:15**	Faktör Sayısının Belirlenmesinde Velicer'in Map Testi, Horn'un Paralel Analizi ve Kaiser Yöntemlerinin Karşılaştırılması: Besin Gücü Ölçeği Uygulaması *Mehtap AKÇİL OK*
	15:15	**15:30**	(R x C) Çapraz Tablolarda Kullanılan Bağımsızlık Testlerinin Karşılaştırılması *Merve BAŞOL*
	15:30	**15:45**	A Simulation Study on a Novel Marginalized Model with Probit Link for Bivariate Longitudinal Binary Data *Özlem İLK DAĞ*
	15:45	**16:00**	*Risk Kestiriminde Gizli Sınıf Kümeleme Analizi ve Kestirim Eğrisi'nin Birlikte Kullanımı* *Didem DERİCİ YILDIRIM*
	16:00	**16:15**	Kesintiye Uğramış Zaman Serilerinin Analizi: Segmented Regresyon ya da Joinpoint Regresyon *Tanyeli GÜNEYLİGİL KAZAZ*
	16:15	**16:30**	(2 x 2) Boyutlu Simetrik Dengesiz Tasarımlarda Değerlendiriciler Arası Uyumun Değerlendirilmesi *Tuğçe ŞENÇELİKEL*
I.T.: Invited Talk			

2. Gün/Day 2 (28 Ekim/October): Ana Salon/Main Hall			
Saat/Hour			**Sunum Başlığı/Presentation Title**
***Oturum Başkanları/Chairs:** Prof. Yasemin YAVUZ, Assoc. Prof. Beyza DOĞANAY, Assist. Prof. Sevilay KARAHAN*			
Oturum 4 / Session 4 (Survival Analysis, Diagnostic Tests)	**09:00**	**09:15**	Soldan Sansürlü Veriler İçin Parametrik Ters Hazard Modelleri: Simülasyon Çalışması *Mustafa Agah TEKİNDAL*
	09:15	**09:30**	Çok-Değerlendiricili Çok-Vakalı ROC Çalışmalarında Farklı Dizayn Türlerinin Radyolojik Görüntüleme Verileri Üzerinde Uygulanması *Fatma KAYMAKAMTORUNLARI DENİZ*
	09:30	**09:45**	Çok-Değerlendiricili Çok-Testli Tasarıma Sahip Tanı Doğruluğu Çalışmalarının Karşılaştırılmasında Güç Analizi *Gülçin AYDOĞDU*
	09:45	**10:00**	Kesişen İki Yaşam Fonksiyonu İçin Kullanılan İstatistiksel Testlerin Karşılaştırılması *Hülya ÖZEN*
	10:00	**10:15**	Pankreas Kanseri verilerinin Paylaşılmış Zayıflık modeli ile Analizi *Şirin ÇETİN*
	10:15	**10:30**	Beyin EEG Ölçümlerini Bayesçi Sınıflandırma *Selim DÖNMEZ*
	10:30	**10:45**	**Kahve Arası / Coffee Break**
***Oturum Başkanları/Chairs**: Prof. Yavuz SANİSOĞLU, Assoc. Prof. M. Agah TEKİNDAL, Assist. Prof. Duygu AYDIN HAKLI*			
Oturum 5 / Session 5 (Statistical Methods, Meta Analysis)	**10:45**	**11:00**	A Novel Test of Normality Based on the Concordance Correlation Coefficient *Osman DAĞ*
	11:00	**11:15**	*Glasgow Koma Skoru Güvenirliği: Cronbach Alpha için Meta Analizi* *Pınar GÜNEL KARADENİZ*
	11:15	**11:30**	Hypoxia inducible factor 2 alpha as a marker of arteriovenous fistula maturation in patients with end stage renal disease *Eray EROĞLU*
	11:30	**11:45**	Polikistik Over Sendromu Olan Kadınlarda Nesfatin Düzeyleri: Bir Meta Analiz Çalışması *Pınar GÜNEL KARADENİZ*

	12:00	**14:00**	**Öğle Yemeği Arası / Lunch Break**
***Oturum Başkanları/Chairs**: PhD. Erdal COŞGUN*			
Oturum 6 / Session 6 (Invited Talks)	**14:00**	**14:45**	Lifescience, Genomics and Bioinformatics on Cloud Based Systems *Mutlu DOĞRUEL* Azure UK Genomics- Life Science Lead at Microsoft, Cambridge, United Kingdom
	14:45	**15:10**	UÇEP 2020 Güncellemesinde Biyoistatistiğin Yeri *Bahar TAŞDELEN*
	15:10	**15:35**	Tükiye'deki Tıp Fakültelerinde Mezuniyet Öncesi Verilen Biyoistatistik Eğitiminin İçeriğinin Standartlaştırılması - Bir Delphi Çalışması *Selcen YÜKSEL*
	15:35	**15:45**	**Kahve Arası / Coffee Break**
	15:45	**17:00**	**Conference Meeting**
3. Gün/Day 3 (29 Ekim/October): Ana Salon/Main Hall			
Saat/Hour			**Sunum Başlığı/Presentation Title**
***Oturum Başkanları/Chairs**: Assoc. Prof. Gülhan TEMEL, Assist. Prof. Didem DERİCİ YILDIRIM*			
Oturum 7 / Session 7 (Kısa Sözlü Sunumlar/Short Oral Presentations)	**09:00**	**09:05**	Comparison of Diagnostic Performance of Binary - Multliclass ROC Analysis and Multinomial Logistic Regression by Using Glasgow Coma Scale Levels *Arzu BAYGÜL*
	09:05	**09:10**	Değişim Noktası Analizi *İsmet DOĞAN*
	09:10	**09:15**	Alzheimer Hastalarında Bayesci Hızlandırılmış Başarısızlık Süresi Modelinin Cox Oransal Hazard Modeliyle Karşılaştırılması: Bir Gerçek Veri Uygulaması *Ilgın Asena CEBECİ*
	09:15	**09:20**	Engelli Kadınlarda Psikiyatrik Belirti Dağılımı *İsmail YILDIZ*
	09:20	**09:25**	Türkiye'de 2009-2017 Yılları Arasında Enfeksiyon'a Bağlı Hastalık Mortalitesinde Trendlere Genel Bir Bakış *Nurhan DOĞAN*
	09:25	**09:30**	Yüksek Kayıp Gözlem Oranına Sahip Çalışmalarda Derin Öğrenme Uygulaması(VUR ve IYE) Su ÖZGÜR
	09:30	**09:35**	Baş Boyun Diseksiyonu Yapilan Kanserli Olgularda Skapular Diskineziyi Tahminlemek İçin Görüntüleme Yöntemlerinin Performansinin Değerlendirilmesi Su ÖZGÜR
	09:35	**09:40**	Akciğer Kanserinde Yarışan Riskler Şirin ÇETİN
	09:40	**09:45**	Biyoistatistik eğitim algımın yolculuğu ve diğer öğrencilerde bunu geliştirmek için bir fikir *Bayazıt CAN*
	09:45	**10:00**	**Kahve Arası / Coffee Break**
	10:00	**10:30**	**Kapanış Konuşmaları / Closing Speeches**

XXI. Ulusal ve IV. Uluslararası Biyoistatistik Kongresi
XXI. National and IV. International Biostatistics Congress
Ekim / October 26 – 29, 2019 - Belconti Hotel, Belek/Antalya

Kongre Programı / Congress Program

1. Gün / Day 1 (27 Ekim/October): Salon II / Hall II

Saat/Hour		Sunum Başlığı/Presentation Title
Oturum Başkanları / Chairs: *Prof. Mehtap AKÇİL OK, Assoc. Prof. Ünal ERKORKMAZ, Assist. Prof. Mesut AKYOL*		
Oturum 2 / Session 2 (Statistical Modeling - I)	**11:15 – 11:30**	Potential Risk Factors for Breast Cancer Subtypes from a North Cyprus Cohort Analysis *Ayse ÜLGEN*
	11:30 – 11:45	Bilgilendirici Küme Boyutlu Verilerin Modellemesine Bir Yaklaşım Önerisi *Betül DAĞOĞLU HARK*
	11:45 – 12:00	Gerçek Veri ve Benzetim Çalışmalarıyla Yarı En Küçük Kareler Regresyonu Yöntemi *Erdoğan ASAR*
	12:00 – 12:15	Model Uyum Ölçütlerini Çok Değişkenli Kontamine Normal Dağılım ile Karşılaştıran Bir Simülasyon *İlkay DOĞAN*
	12:15 – 12:30	Faktöriyel Tasarımlarda Veri Dönüştürme Yöntemlerinin Varyans Analizinde Gerçekleşen I. Tip Hata ve Testin Gücü Üzerine Etkisi *Yeliz KAŞKO ARICI*
	12:30 – 14:00	**Öğle Yemeği Arası / Lunch Break**
Oturum Başkanları/Chairs*: Prof. Nazan ALPARSLAN, Assoc. Prof. Jale KARAKAYA, Assoc. Prof. Yaşar SERTDEMİR*		
Oturum 3 / Session 3 (Other Statistical Methods)	**15:00 – 15:15**	Randomize Kontrollü Çalışmalarda Kayıp Veri Sorunu Üzerine *Aslıhan ALHAN*
	15:15 – 15:30	In silico Investigation of Potential Biomarkers of Lung Adenocarcinoma *Ceren SUCULARLI*
	15:30 – 15:45	Acil Servislerdeki Çocuk Alt Solunum Yolu Enfeksiyonları: Bibliyometrik Analiz *Hadiye DEMİRBAKAN*
	15:45 – 16:00	Randomize Olmayan Klinik Çalışmalarda En Uygun Eşleştirme Analizi için Makine Öğrenme Algoritmaları ile Yeni Propensity Skor Tahmin Modellerinin Geliştirilmesi ve R Shiny ile İnteraktif Bir Web Uygulaması *Emre DEMİR*
	16:00 – 16:15	Türkiye'de Bebek Ölüm Riskinin Mekan-zamansal Bayesci Modeller ile İncelenmesi *Sade KILIÇ YILDIRIM*
	16:15 – 16:30	Ensemble-Based Machine Learning Prediction of Heart Disease Presence with The Cleveland Database *Setia PRAMANA*

2. Gün/Day 2 (28 Ekim/October): Salon II / Hall II

Saat/Hour		Sunum Başlığı/Presentation Title
Oturum Başkanları/Chairs*: Prof. İsmet DOĞAN, Assoc. Prof. Semra ERDOĞAN, Assoc. Prof. Cengiz BAL*		
Oturum 4 / Session 4 (Machine	**09:00 – 09:15**	Naive Bayes Sınıflayıcı ve Markov Modellerinin Sınıflandırma Performansının Karşılaştırılması *Adem DOĞANER*
	09:15 – 09:30	Doğrusal Olmayan Temel Bileşenler Analizi İçin Yapay Sinir Ağları Yaklaşımı *Canan DEMİR*

	09:30	**09:45**	Deep Learning Based Imbalanced Data Classification for Drug Discovery *Selçuk KORKMAZ*
	09:45	**10:00**	Tıbbi Görüntü İşleme İle Tanı Koymada Veri Madenciliği Yöntemlerinin Sınıflama Performanslarının Karşılaştırılması *Hanife AVCI*
	10:00	**10:15**	Gizli Ağaç Modelleri ve Gizli Sınıf Ağaç Analizi Yaklaşımı *Meriç YAVUZ ÇOLAK*
	10:15	**10:30**	Yoğun Bakım Ünitelerinde Yatan Hastalara İlişkin Mortalite ve Yatış Süresine Etki Eden Faktörlerin Veri Madenciliği Yöntemleriyle İncelenmesi *Erkin SÜLEKLİ*
	10:30	**10:45**	**Kahve Arası / Coffee Break**
***Oturum Başkanları/Chairs**: Prof. Fezan MUTLU, Prof. Bahar TAŞDELEN, Assist. Prof. İsmail YILDIZ*			
Oturum 5 / Session 5 (Machine Learning - II)	**10:45**	**11:00**	Transkriptom Veri Seti Üzerinde Derin Öğrenme Yöntemi İle Klasik Veri Madenciliği Yöntemlerinin Sınıflama Performanslarının Karşılaştırılması *Merve KAŞIKCI*
	11:00	**11:15**	Plantar Fasiit Tanısında Performans Değerlendirme için Çapraz Doğrulama Yöntemlerinin Karşılaştırılması *Senem KOÇ*
	11:15	**11:30**	Çoklu Sınıflamada Sınıflama Performansı Testlerinin Bir Uygulama Üzerinde Karşılaştırılması *Tanyeli GÜNEYLİGİL KAZAZ*
	11:30	**11:45**	Apriori Algoritması İle Yoğun Bakımlarda Yapılan Tetkiklerin Modellenmesi *Meryem GÖRAL YILDIZLI*
	11:45	**12:00**	Comparative Analysis Of MNA Scores With Linear Mixed Effect Models And Linear Regression Models in Patients With Peg Feed *Neslihan GÖKMEN*

SÖZLÜ BİLDİRİLER / ORAL PRESENTATIONS

Sayfa

S1. Adem DOĞANER. Naive Bayes Sınıflayıcı ve Markov Modellerinin Sınıflandırma Performansının Karşılaştırılması 2

S2. Mustafa Agah TEKİNDAL, Can ATEŞ, Çağatay BÜYÜKUYSAL. Soldan Sansürlü Veriler İçin Parametrik Ters Hazard Modelleri: Simülasyon Çalışması 4

S3. Aslıhan ALHAN, Derya GÖKMEN. On The Issue Of Missing Data In Randomized Controlled Trials 6

S4. Ayse ULGEN, Wentian Li, Ozlem GURKUT. Potential Risk Factors for Breast Cancer Subtypes from a North Cyprus Cohort Analysis 8

S5. Dinçer Göksülük, A. Ergun Karaağaoğlu. RNA-Dizileme Verilerinin Kesikli Dağılıma Dayalı Yöntemler ile Sınıflandırılması 9

S6. Betül DAĞOĞLU HARK, Zeliha Nazan ALPARSLAN. Bilgilendirici Küme Boyutlu Verilerin Modellemesine Bir Yaklaşım Önerisi 11

S7. Canan DEMİR, Sıddık KESKİN. Doğrusal Olmayan Temel Bileşenler Analizi İçin Yapay Sinir Ağları Yaklaşımı 13

S8. Didem DERİCİ YILDIRIM, Merve TÜRKEGÜN, Bahar TAŞDELEN, Tuğba ARIKOĞLU, Semanur KUYUCU. Risk Kestiriminde Kestirim Eğrisi ve Gizli Sınıf Kümeleme Analizi'nin Birlikte Kullanımı 14

S9. Duygu KORKMAZ, İlker ÜNAL. Aynı Veri Setleri Üzerinde Bazı Etki Büyüklüğü Hesaplama Yöntemlerinin Tahmin Başarılarının Karşılaştırılması 16

S10. Emre DEMİR, Serdal Kenan KÖSE. Randomize Olmayan Klinik Çalışmalarda En Uygun Eşleştirme Analizi için Makine Öğrenme Algoritmaları ile Yeni Propensity Skor Tahmin Modellerinin Geliştirilmesi ve R Shiny ile İnteraktif Bir Web Uygulaması 18

S11. Erdal COŞGUN. Genetik ve Klinik Verilerin Birleştirilmesinde Regresyon Modeli Artık Değerlerinin Kullanımı 20

S12. Erdoğan ASAR, Erdem KARABULUT. Quasi-Least Squares Regression Method with Real Data and Simulation Studies 24

S13. Fatma KAYMAKAMTORUNLARI DENİZ, Gülçin AYDOĞDU, Yunus ÇAKMAK, Yasemin YAVUZ, Elif PEKER. Çok-Değerlendiricili Çok-Vakalı ROC Çalışmalarında Farklı Tasarım Türlerinin Radyolojik Görüntüleme Verileri Üzerinde Uygulanması 25

S14. Gülçin AYDOĞDU, Yunus ÇAKMAK, Fatma KAYMAKAMTORUNLARI DENİZ, Yasemin YAVUZ. Çok-Değerlendiricili Çok-Testli (MRMC) Tasarıma Sahip Tanı Doğruluğu Çalışmalarının Karşılaştırılmasında Güç Analizi 26

S15. Hanife AVCI, Jale KARAKAYA. Tıbbi Görüntü İşleme İle Tanı Koymada Veri Madenciliği Yöntemlerinin Sınıflama Performanslarının Karşılaştırılması 27

S16. Hülya ÖZEN, Cengiz BAL, Ertuğrul ÇOLAK. Kesişen İki Yaşam Fonksiyonu İçin Kullanılan İstatistiksel Testlerin Karşılaştırılması 28

S17. İlkay DOĞAN. Model Uyum Ölçütlerini Çok Değişkenli Kontamine Normal Dağılım ile Karşılaştıran Bir Simülasyon Çalışması 30

S18. Merve KAŞIKCI, Erdal COŞGUN, Erdem KARABULUT. Transkriptom Veri Seti Üzerinde Derin Öğrenme Yöntemi İle Klasik Veri Madenciliği Yöntemlerinin Sınıflama Performanslarının Karşılaştırılması 32

S19. Mehtap AKÇİL OK. Faktör Sayısının Belirlenmesinde Velicer'in Map Testi, Horn'un Paralel Analizi ve Kaiser Yöntemlerinin Karşılaştırılması: Besin Gücü Ölçeği Uygulaması 36

S20. Meriç YAVUZ ÇOLAK, Ersin ÖĞÜŞ. Gizli Ağaç Modelleri ve Gizli Sınıf Ağaç Analizi Yaklaşımı 38
S21. Merve BAŞOL, Ebru ÖZTÜRK, Sevilay KARAHAN. R x C Çapraz Tablolarda Kullanılan Bağımsızlık Testlerinin Karşılaştırılması 39
S22. Meryem GÖRAL YILDIZLI, Zeliha Nazan ALPARSLAN. Apriori Algoritması İle Yoğun Bakımlarda Yapılan Tetkiklerin Modellenmesi 41
S23. Neslihan GÖKMEN, Arzu BAYGÜL, Aydın ERAR. Comparative Analysis of Mna Scores With Linear Mixed Effect Models And Linear Regression Models in Patients with Peg Feed 43
S24. Hülya BİNOKAY, Yaşar SERTDEMİR. Dairesel(Circular) Verinin Tanıtılması ve Gerçek Veri Seti Üzerine Uygulama 45
S25. Gül INAN, Ozlem ILK. İki Elemanlı İki Değişkenli Uzunlamasına Veri için Probit Bağlantılı Marjinalleştirilmiş Özgün bir Model Üzerine bir Simülasyon Çalışması 49
S26. Özlem ARIK, Erdem KARABULUT. Mikrodizi Kanser Verilerinin Öznitelik Seçimi ve Sınıflandırılması 51
S27. Demet ARI YILMAZ, Pınar GÜNEL KARADENİZ, Vildan SÜMBÜLOĞLU. Glasgow Koma Skoru Güvenirliği: Cronbach Alpha için Meta Analizi 55
S28. R. Onur ÖZTORNACI, Bahar TAŞDELEN, Andrew P. MORRIS, Hamzah SYED. ULSAM Veri Setinde Tek Nükleotid Polimorfizmi Veri ile Makine Öğrenimi Yöntemlerinin Kullanımı 57
S29. Sade KILIÇ YILDIRIM, Celal Reha ALPAR, Erdem KARABULUT, Atilla Halil ELHAN. Türkiye'de Bebek Ölüm Riskinin Mekan-zamansal Bayesci Modeller ile İncelenmesi 61
S30. Ceren SUCULARLI. Akciğer Adenokarsinomu için Potansiyel Biyobelirteçlerin In Silico Araştırılması 63
S31. Setia PRAMANA, Mieke NURMALASARI. Ensemble-Based Machine Learning Prediction of Heart Disease Presence with The Cleveland Database 65
S32. Eray EROĞLU, İsmail KOÇYİĞİT, Ciğdem KARAKÜKÇÜ, Aydın TUNÇAY, Gökmen ZARARSIZ, Davut EREN, Murat Hayri SİPAHİOGLU, Bülent TOKGÖZ, Kutay TAŞDEMİR, Oktay OYMAK. Hypoxia inducible factor 2 alpha as a marker of arteriovenous fistula maturation in patients with end stage renal disease 66
S33. Senem KOÇ, Leman TOMAK. Plantar Fasiit Tanısında Performans Değerlendirme için Çapraz Doğrulama Yöntemlerinin Karşılaştırılması 67
S34. Yeliz KAŞKO ARICI, Mustafa MUHİP ÖZKAN. Faktöriyel Tasarımlarda Veri Dönüştürme Yöntemlerinin Varyans Analizinde Gerçekleşen I. Tip Hata ve Testin Gücü Üzerine Etkisi 69
S35. Su ÖZGÜR, Erdal COŞGUN, Kemal Uğur TÜFEKCİ, Pembe KESKİNOĞLU, Görsev YENER, Şermin GENÇ, Mehmet N. ORMAN. Makine Öğrenmesi İle Genetik Varyasyon Çeşitlerinin Tahminlenmesi 71
S36. Şirin ÇETİN, Osman DEMİR, İsa DEDE, Özge PASİN, Emre KUYUCU. Pankreas Kanseri verilerinin Paylaşılmış Zayıflık modeli ile analizi 75
S37. Tanyeli GÜNEYLİGİL KAZAZ, İlkay DOĞAN, Seval KUL, Neslihan BAYRAMOĞLU TEPE. Çoklu Sınıflama Durumunda Sınıflama Performansı Testlerinin Bir Uygulama ile Karşılaştırılması 76
S38. Uğur TOPRAK, Beyza DOĞANAY ERDOĞAN, Esin ÖĞüŞ. An Interactive R Shiny Application to Facilitate Comparing Classification Methods: ClasComp 78
S39. H. Yağmur ZENGİN, Tuğçe ŞENÇELİKEL, Ersin ÖĞÜŞ. 2×2 Boyutlu Simetrik Dengesiz Tasarımlarda Değerlendiriciler Arası Uyumun Değerlendirilmesi 79
S40. Osman Dag, Anil Dolgun. Konkordans Korelasyon Katsayısı Temelli Yeni Bir Normallik Testi 80
S41. Burçin ALTINBAŞ, Pınar GÜNEL KARADENİZ. Polikistik Over Sendromu Olan Kadınlarda Nesfatin Düzeyleri: Bir Meta Analiz Çalışması 82
S42. Pınar GÜNEL KARADENİZ, Tanyeli GÜNEYLİGİL KAZAZ. Kesintiye Uğramış Zaman Serilerinin Analizi: Segmented Regresyon ya da Joinpoint Regresyon 84
S43. Hadiye DEMİRBAKAN, Demet ARI YILMAZ, Pınar GÜNEL KARADENİZ, Vildan SÜMBÜLOĞLU. Acil Servislerdeki Çocuk Alt Solunum Yolu Enfeksiyonları: Bibliyometrik Analiz 86

S44. Ragıp Onur ÖZTORNACI, Erdal COŞGUN, Bahar TAŞDELEN. Genom-Boyu İlişki Çalışmalarında Makine Öğrenimi Yöntemleri ve Derin Öğrenme Yönteminin Farklı Örnek Genişliklerinde Performanslarının Değerlendirilmesi 88

KISA SÖZLÜ BİLDİRİLERİ / SHORT ORAL PRESENTATIONS

Sayfa

K1. Arzu BAYGÜL, Neslihan GÖKMEN. Comparison of Diagnostic Performance of Binary - Multliclass ROC Analysis and Multinomial Logistic Regression by Using Glasgow Coma Scale Levels 93

K2. İsmet DOĞAN, Nurhan DOĞAN. Değişim Noktası Analizi 94

K3. Ilgın Asena CEBECİ, Damla ÖZTÜRK, Başak DOĞAN, Nural BEKİROĞLU. Alzheimer Hastalarında Bayesci Hızlandırılmış Başarısızlık Süresi Modelinin Cox Oransal Hazard Modeliyle Karşılaştırılması: Bir Gerçek Veri Uygulaması 98

K4. İsmail YILDIZ, Ömer SATICI, Remzi OTO. Engelli Kadınlarda Psikiyatrik Belirti Dağılımı 100

K5. Su ÖZGÜR, Pembe KESKİNOĞLU, Hande Melike HALAÇ, Aybüke Cansu KALKAN, Seher ÖZYÜREK, Ali BALCI, Ersoy DOĞAN, Ahmet Ömer İKİZ, Arzu GENÇ. Baş Boyun Diseksiyonu Yapilan Kanserli Olgularda Skapular Diskineziyi Tahminlemek için Görüntüleme Yöntemlerinin Performansinin Değerlendirilmesi 101

K6. Su ÖZGÜR, Pembe KESKİNOĞLU, Erdal COŞGUN, Timur KÖSE, Ahmet KESKİNOĞLU Yüksek Kayıp Gözlem Oranına Sahip Çalışmalarda Derin Öğrenme Uygulaması (VUR ve İYE) 105

K7. Bayazıt CAN. Biyoistatistik Eğitim Algımın Yolculuğu ve Diğer Öğrencilerde Bunu Geliştirmek için Bir Fikir 109

K8. Şirin ÇETİN, İsa DEDE. Akciğer Kanserinde Yarışan Riskler 110

K9. Nurhan DOĞAN, Neşe Demirtürk, İsmet DOĞAN. Türkiye'de 2009-2017 Yılları Arasında Enfeksiyon'a Bağlı Hastalık Mortalitesinde Trendlere Genel Bir Bakış 111

TAM METİN BİLDİRİLER / FULL TEXT ORAL PRESENTATIONS

Sayfa

T1. Erdoğan ASAR, Erdem KARABULUT. Gerçek Veri ve Benzetim Çalışmalarıyla Yarı En Küçük Kareler Regresyonu Yöntemi 114

T2. Didem DERİCİ YILDIRIM, Merve TÜRKEGÜN, Bahar TAŞDELEN, Tuğba ARIKOĞLU, Semanur KUYUCU. Risk Kestiriminde Kestirim Eğrisi ve Gizli Sınıf Kümeleme Analizi'nin Birlikte Kullanımı 128

T3. Duygu KORKMAZ, İlker ÜNAL. Aynı Veri Setleri Üzerinde Bazı Etki Büyüklüğü Hesaplama Yöntemlerinin Tahmin Başarılarının Karşılaştırılması 140

T4. Emre DEMİR, Serdal Kenan KÖSE. Randomize Olmayan Klinik Çalışmalarda En Uygun Eşleştirme Analizi için Makine Öğrenme Algoritmaları ile Yeni Propensity Skor Tahmin Modellerinin Geliştirilmesi ve R Shiny ile İnteraktif Bir Web Uygulaması 149

T5. Hanife AVCI, Jale KARAKAYA. Tıbbi Görüntü İşleme İle Tanı Koymada Veri Madenciliği Yöntemlerinin Sınıflama Performanslarının Karşılaştırılması 163

T6. Hülya BİNOKAY, Yaşar SERTDEMİR. Dairesel(Circular) Verinin Tanıtılması Gerçek Veri Seti Üzerine Uygulama 171

T7. Neslihan GÖKMEN, Arzu BAYGÜL, Aydın ERAR. Comparatıve Analysis of Mna Scores With Linear Mıxed Effect Models And Linear Regressıon Models in Patients With Peg Feed 180

T8. R. Onur ÖZTORNACI, Bahar TAŞDELEN, Andrew P. MORRIS, Hamzah SYED. ULSAM Veri Setinde Tek Nükleotid Polimorfizmi Verileri ile Makine Öğrenimi Yöntemlerinin Kullanımı 192

T9. Sade KILIÇ YILDIRIM, Celal Reha ALPAR, Erdem KARABULUT, Atilla Halil ELHAN. Türkiye'de Bebek Ölüm Riskinin Mekan-zamansal Bayesci Modeller ile İncelenmesi 199

T10. Selim DÖNMEZ, Özer ÖZAYDIN. Beyin EEG Ölçümlerini Bayesçi Sınıflandırma 208

T11. Uğur TOPRAK, Beyza DOĞANAY ERDOĞAN, Ersin ÖĞÜŞ. Sınıflama Metotlarının Karşılaştırılmasını Kolaylaştırmak için İnteraktif Bir R Shiny Uygulaması: ClasComp 213

T12. YÜKSEL S., ALKAN A., DEMİR P., SANİSOĞLU S.Y., ERCAN İ., TÜRE M., KARAHAN S., ATEŞ C., ELMALI F., ERDOĞAN S., KAYA M.O., ŞAHİN B., YAVUZ Z., YILDIRIM D. Türkiye'deki Tıp Fakültelerinde Mezuniyet Öncesi Verilen Biyoistatistik Eğitiminin İçeriğinin Standartlaştırılması-Bir Delphi Çalışması 220

T13. Selçuk KORKMAZ. Deep Learning Based Imbalanced Data Classification for Drug Discovery 271

SÖZLÜ BİLDİRİLER
ORAL PRESENTATIONS

S1

Naive Bayes Sınıflayıcı ve Markov Modellerinin Sınıflandırma Performansının Karşılaştırılması

Adem DOĞANER

Kahramanmaraş Sütçü İmam Üniversitesi, Tıp Fakültesi, Biyoistatistik ve Tıbbi Bilişim Anabilim Dalı, Kahramanmaraş

Amaç: Makine öğrenme yöntemleri içerisinde sıklıkla kullanılan yöntemlerden biride Stokastik modellerdir. Stokastik modeller başarılı sınıflandırıcılar olmasına karşın her veri setinde yüksek sınıflandırma performansı sağlayamamaktadır. Naive Bayes ve Markov Modelleri gibi güçlü sınıflandırıcılar aynı veri seti için farklı performans sergileyebilmektedir. Bu nedenle sınıflayıcının özelliklerine göre veri setine uygun yöntemi belirlemek önem arz etmektedir. Bu çalışmada Naive Bayes sınıflayıcı ve Markov Modellerinin sınıflandırma performansı karşılaştırılmış ve Markov modellerinin sınıflandırma performansının artırılması için farklı önerilerde bulunulmuştur.

Yöntem: Çalışmada Naive Bayes Sınıflayıcı ve Markov Modellerinden biri olan Saklı Markov Modeli kullanılmıştır. Saklı Markov Modeli bir durum dizisi(state sequence) ve durumlara bağlı olarak ortaya çıkan gözlemlerden oluşmaktadır. Modelde durumlar(states) doğrudan görünmemektedir ve gözlemler(observation) aracılığı ile tahmin edilmektedir. Modelde başlangıç olasılıkları vektörü, geçiş olasılıkları matrisi ve emisyon olasılıkları matrisi elde edilmektedir. Saklı Markov modeli herhangi bir t anında, bir durum için bir gözlem ortaya çıktığını esas alarak tahmin gerçekleştirmektedir. Bu çalışmada her bir durum için bir gözlemin ortaya çıktığı saklı markov modelinin yanı sıra her bir durum için birden fazla gözlem ortaya çıktığı saklı markov modeli de oluşturulmuştur. Oluşturulan bu modelde her gözlem dizisi için emisyon olasılıkları matrisi üretilmiştir. Emisyon olasılıkları matrisinde en yüksek olasılığa sahip durumlar belirlenmiş ve çoğunluk oyu (majority voting) yöntemiyle durumlar tahmin edilmiştir. Veri seti olarak Kahramanmaraş ilinin 6 aylık alerjen hava raporları dikkate alınmıştır. Nefes alma konforu durum olarak kabul edilirken gözlem dizilerini oluşturan değişkenler ise, polen durumu, hava sıcaklığı, nem oranı, rüzgar hızı ve yağış oranı dikkate alınmıştır. Sürekli değişkenler kategorize edilerek kesikli değişkene dönüştürülmüştür. Veri seti Naive Bayes sınıflandırıcı ve Markov Modelleri ile sınıflandırılmıştır. Sınıflandırma performanslarının değerlendirmesinde doğruluk, sensivite, spesivite, metrikleri kullanılmıştır.

Bulgular: Naive Bayes ve Markov modellerinin sınıflandırma performansları karşılaştırılmıştır. Markov Modelleri Naive Bayes sınıflandırıcıya göre daha yüksek sınıflandırma performansı göstermiştir.

Sonuç: Durum dizisi içeren verilerin sınıflandırılmasında durumlar birden fazla gözlem içermesi durumunda Markov modellerinin Naive Bayes sınıflandırıcıya göre daha yüksek performans sağladığı gözlemlenmiştir. Ayrıca önerilen Markov Modelinin mevcut Saklı Markov modelinin sınıflandırma performansını artırdığı gözlenmiştir.

Anahtar Kelimeler: Naive Bayes Sınıflayıcı, Markov Modelleri, Makine Öğrenme

S1

Comparison of Classification Performances of Naïve Bayes Classifier and Markov Models

Adem DOĞANER

Kahramanmaras Sutcu Imam University, Faculty of Medicine, Department of Biostatistics and Medical Informatics, Kahramanmaras

Aim: Stochastic models are one of the methods commonly used in machine learning. Although stochastic models are successful classifiers, they do not provide high classification performance in each data set. Powerful classifiers such as Naive Bayes and Markov Models can perform differently for the same data set. Therefore, it is important to determine the appropriate method for the data set according to the characteristics of the classifier. In this study, the classification performance of Naive Bayes classifier and Markov models are compared and different suggestions are made to improve the classification performance of Markov models.

Method: In the study, Hidden Markov Model is used as one of the Markov Models and Naive Bayes Classifier. Hidden Markov Model consists of a state sequence and observations which occur depending on the situations. The states do not appear directly in the model and are estimated through observations. In the model, initial probability vector, transition probability matrix and emission probability matrix are obtained. Hidden Markov model realizes the prediction based on the occurrence of an observation for a situation at any t moment. In this study, a hidden Markov model is created, where more than one observation occurs for each state as well as a hidden Markov model where one observation occurs for each state. In this model, emission probability matrix is produced for each observation series. The highest probability states are determined in the emission probability matrix and the states are estimated by majority voting method. 6 months-long weather allergy reports of Kahramanmaraş province were taken into consideration as data set. While the breathing comfort is accepted as the state, the variables forming the observation series and taken into consideration are the pollen condition, air temperature, humidity rate, wind speed and rainfall rate. Continuous variables are categorized and transformed into discrete variables. The data set is classified with Naive Bayes classifier and Markov Models. Accuracy, sensitivity, specificity metrics were used to evaluate the classification performances.

Results: The classification performances of Naive Bayes and Markov models are compared. Markov Models showed higher classification performance than Naive Bayes classifier.

Conclusion: It was observed that Markov models provide higher performance than Naive Bayes classifier if the states contain more than one observation in the classification of the data containing the state series. It was also observed that the proposed Markov Model improves the classification performance of the existing Hidden Markov model.

Keywords: Naive Bayes Classifier, Markov Models, Machine Learning

S2

Soldan Sansürlü Veriler İçin Parametrik Ters Hazard Modelleri: Simülasyon Çalışması

Mustafa Agah Tekindal[1], Can Ateş[2], Çağatay Büyükuysal[3]

[1] *Selçuk Üniversitesi, Veteriner Fakültesi, Biyoistatistik ABD, Konya*
[2] *Van Yüzüncü Yıl Üniversitesi, Tıp Fakültesi, Biyoistatistik ABD, Van*
[3] *Zonguldak Bülent Ecevit Üniversitesi, Tıp Fakültesi, Biyoistatistik ABD, Zonguldak*

Amaç: Hekimlik alanlarında soldan sansürlü verisi bulunan birçok çalışmaya rastlanmaktadır. Bu çalışmalarda sansürlü veriler çeşitli yöntemlerle en az yanlı sonuç verecek şekilde değerlendirilmektedir. Çalışmada, soldan sansürlü veriler için parametrik ters hazard modelleri için hangi örneklem genişliğinde ve hangi dağılımda ne kadar yanlılık olduğu belirlenmeye çalışılmıştır. En iyi dağılım AIC, AICC, HQIC ve CAIC bilgi kriterleri kullanılarak belirlenmiştir.

Yöntem: Üstel, Log-Normal, Inverse(Ters) Normal, Gamma, Genelleştirilmiş Gamma, Ters Gamma, Log-Logistik, Weibull, Ters Weibull, Genelleştirilmiş Ters Weibull, Esnek Weibull, Marshal-Olkin, Güçlü Genelleştirilmiş Weibull, Modifiye Weibull ve Gompertz dağılımları kullanılarak %20-%30 ve %40 soldan sansür oranlarında simülasyon çalışması yapılmıştır. Sansür oranı, 0'ın sansürlü bir gözlemi temsil ettiği, aksi takdirde 1 olan bir sansür göstergesi oluşturularak sağlanmıştır. Gösterge daha sonra rastgele gözlem oranına atanarak sansürleme belirlenmiştir. Simülasyon çalışmaları her biri güvenilirliği sağlamak için 5000 kez tekrar edilmiştir. Hangi dağılım modelinin en uygun olduğunu değerlendirmek için AIC, AICC, HQIC ve CAIC bilgi kriterleri kullanılmıştır.

Bulgular: Yapılan benzetim çalışmaları sonucunda, soldan sansürlü veriler için parametrik ters hazard modelleri için kullanılacak en iyi dağılımın, sırasıyla Log-Logistic, Log-Normal, Inverse Gaussian ve Gamma dağılımı ile birlikte Genelleştirilmiş Ters Weibull dağılımı olduğunu göstermektedir. Ayrıca Marshal-Olkin dağılımının, Modifiye Weibull, Genelleştirilmiş Gamma, Gamma ve Esnek Weibull dağılımlarına göre üstün performans gösterdiğini söyleyebiliriz.

Sonuç: Tüm olası senaryolar hakkında simülasyon çalışması yapmak mümkün olamayacağı için birbirine yakın senaryolar genellikle diğerlerinin yerine tercih edilmektedir. Belirtilen senaryolardaki simülasyon çalışmalarında tam bir tutarlılık elde edilememiştir. Sansür oranının %20-%30 ve %40 olduğu durumlarda soldan sansürlü veriler için parametrik ters hazard modelleri için kullanılacak en uygun dağılımın Log logistic dağılımı olduğu kanısına varılmıştır.

Anahtar Sözcükler: Soldan Sansür, Parametrik Ters Hazard Modelleri, Simülasyon, Genelleştirilmiş Ters Weibull dağılımı

S2

Parametric Reversed Hazards Model For Left-Censored Data: A Simulation Study

Mustafa Agah Tekindal[1], Can Ateş[2], Çağatay Büyükuysal[3]

[1]*Selçuk University, Faculty Of Veterinary Medicine, Department of Biostatistics, Konya*
[2]*Van Yüzüncü Yıl University, Faculty of Medicine, Department of Biostatistics, Van*
[3]*Zonguldak Bülent Ecevit University, Faculty of Medicine, Department of Biostatistics, Zonguldak*

Aim: There are many studies with left-censored data in the fields of medicine. In these studies, censored datas are evaluated in a way that gives the least biased results by various methods. In this study, it is tried to determine how much bias ocur in which sample size and distribution for parametric inverse hazard models for left-censored datas. The best distribution identified by AIC, AICC, HQIC and CAIC information criterias.

Method: A simulation study was conducted with a left-censored ratios of 20%, 30% and 40% for Exponential, Log-Normal, Inverse Normal, Gamma, Generalized Gamma, Inverse Gamma, Log-Logistic, Weibull, Inverse Weibull, Generalized Inverse Weibull, Flexible Weibull, Marshal-Olkin, Power-Generalized Weibull, Modified Weibull and Gompertz distributions. Censoring rate is provided by creating a censor indicator where 0 represents a censored , otherwise 1. The indicator was then assigned to a random observation rate to determine censoring. Simulation stuied were repeated 5.000 times to provide reliability. AIC, AICC, HQIC and CAIC information criterias were used to evaluate which distribution model was most appropriate.

Results: As a result of the simulation study, best distributions for the parametric reversed hazards models Log-Logistic, Log-Normal, Inverse Gaussian and Gamma and Generalized Inverse Weibull respectively. Also, we can specify that Marshal-Olkin has much better performance to Modified Weibull, Generalized Gamma, Gamma ve Flexible Weibull distributions.

Conclusion: Since it is not possible to simulate all possible scenarios, close-up scenarios are often preferred over others. A complete consistency has not been achieved in the simulation studies in the mentioned scenarios. It has been conclueded as Log-Logistic distribution is the most appropriate distribution for parametric reversed hazard models for lef-censored datas with a censor rate of 20%, 30% and 40%.

Keywords: Left-Censored, Parametric Reversed Hazards Models, Simulation, Generalized Inverse Weibull Distribution

S3

On The Issue Of Missing Data In Randomized Controlled Trials

Aslıhan ALHAN[1], Derya GÖKMEN[2]

[1]Ufuk University, Faculty of Medicine, Department of Biostatistics, Ankara, Turkey
[2]Ankara Univeristy, *Faculty of Medicine, Department of Biostatistics, Ankara, Turkey*

Aim: The purpose of this study is to introduce various approaches to the solutions of problems faced in case of missing data in Randomized Controlled Trials (RCT) studies, which has an influential position in interventional researches.

Method: In interventional studies, patients might be excluded from the study due to various reasons such as an inability to reach the participant, inappropriateness of the application, side effects of the drug. The researchers may encounter problems such as the omission of randomization, decreasing statistical power of the study, bias or failure to reflect the clinically expected. In such cases, the Intention to Treated Analysis (ITT), Per Protocol Analysis (PP), and As-Treated Analysis (AT) approaches are used to minimize the errors caused by missing data. In these approaches, the last observation carried forward (LOCF), multiple imputations, full information maximum likelihood, and repeated measures mixed effect is used to estimate the missing data.

Results: The ITT method is used in randomized controlled trials when they have missing data. Because ITT analysis is conservative, results can be biased and Type II error can be increased. Therefore, most researchers have recommended sensitivity analysis by evaluating the results considering different conditions. Furthermore, different methods should be applied for the experimental purposes. If the trial's purpose is purely explanatory, a per-protocol or an as-treated analysis might be a great alternative. Among these missing data estimation methods, the multiple imputations approach is the most effective method when the missing data rate is less than 20%. If there is more than 20% missing data, type I error increases and statistical power decreases.

Conclusion: According to the type of missing data, missing data rate, and purpose of the experiment, missing data analysis can be performed with sensitivity analysis; however, the accuracy and impartiality of these results are controversial, and a definite result cannot be achieved. Therefore, even these missing data estimate methods are used, it is recommended to report the amount of missing data by following the CONSORT or STROBE guidelines.

Keywords: Randomized Controlled Trials (RCT), Missing Data, Intention to Treated Analysis (ITT), Per Protocol Analysis (PP) and As-Treated Analysis (AT)

References

1. Armijo-Olivo, S., Warren, S., & Magee, D. (2009). Intention to treat analysis, compliance,drop-outs and how to deal with missing datain clinical research: a review. Physical Therapy Reviews, 36-49.
2. Gökmen, D. (2018). Bilimsel Araştırma Yöntemleri. Ankara: Ankara Üniversitesi Yayınevi.

3. Gupta, S. (2011). Intention-to-treat concept: A review. Perspect Clin Res., 109-12.
4. Heritier, S., Gebski, V., & Keech, A. (2003). Inclusion of patients in clinical trial analysis:the intention-to-treat principle. The Medical Journal of Australia:MJA, 438-440.
5. Hernán, M., & Robins, J. (2017). Per-Protocol Analyses of Pragmatic Trials. The New England Journal of Medicine, 1391-1398.
6. Lydersen, S. (2019). Missing Data: Is It All The Same? June 2019 Annals of the Rheumatic Diseases 78(Suppl 2):48, (s. 1-48). Madrid.
7. Molnar, F., Hutton, B., & Fergusson, D. (2008). Does analysis using "last observation carried forward" introduce bias in dementia research? CMAJ, 751–753.
8. Naimi, A., Cole, S., & Kennedy, E. (2017). An introduction to g methods. International Journal of Epidemiology, 756–762.
9. Ranganathan, P., Pramesh, C. S., & Aggarwal, R. (2016;7). Common pitfalls in statistical analysis: Intention-to-treat versus perprotocol. 144-6.
10. Tilbrook, H., Hewitt, C., Aplin, J., Semlyen, A., Trewhela, A., Watt, I., & Torgerson, D. (2014). Compliance effects in a randomised controlled trial of yoga for chronic low back pain: a methodological study. Physiotherapy, 256–262.
11. Weitkunat, R., Baker, G., & Lüdicke, F. (2016). Intention-to-Treat Analysis but for Treatment Intention: How should Consumer Product Randomized Controlled Trials be Analyzed? International Journal of Statistics in Medical Research, 90-98.

S4

Potential Risk Factors for Breast Cancer Subtypes from a North Cyprus Cohort Analysis

Ayse Ulgen[1,2],Wentian Li[3], Ozlem Gurkut[4]

[1]*Department of Epidemiology, Mailman School of Public Health, Columbia University, New York, NY, USA*

[2]*Dr Fazil Kucuk Faculty of Medicine, Eastern Mediterranean University, Famagusta, TRNC*

[3]*The Robert S. Boas Center for Genomics and Human Genetics, The Feinstein Institute for Medical Research, Northwell Health, Manhasset, NY, USA*

[4]*Dr. Burhan Nalbantoglu State Hospital (BNSH), Faculty of Medicine, Yakin Dogu University, Nicosia, TRNC*

Aim: Determine risk factors for breast cancer subtypes for North Cyprus population.

Method: More than three hundred North Cyprus breast cancer patients with subtype information from are surveyed from the Dr. Burhan Nalbantoglu State Hospital (BNSH) in Nicosia, North Cyprus between 2006-2015 for their demographic, reproductive, genetic, epidemiological factors. This represented around 40% of total breast cancer cases that exist in the archives during this period. The breast cancer subtypes: Estrogen receptor (ER) +/-, Progesterone receptor (PR) +/-, and human epidermal growth factor 2 (HER2) +/- status, are determined. We carried out single-variable, multiple variable, regularized (automatic variation selection) regressions such as LASSO, with risk factors as independent variables, and breast cancer subtypes as dependent variables.

Results: Despite the fact that our cohort differs significantly from some larger cohorts (e.g., the Breast Cancer Family Registry (BCFR) with samples from USA/Canada/Australia) in age, menopause status, age of menarche, parity, education level, oral contraceptive use, breast feeding, the distribution of breast subtypes is not significantly different. The subtype distribution in our cohort is also not different from another study on a Turkish cohort. Using regularized regressions, we show that the estrogen-receptor-positive (ER+) subtype is positively related to post-menopause and negatively associated with hormone therapy; the estrogen-receptor-positive and progesterone-receptor-positive (ER+/PR+) subtype is positively associated with breast feeding, and negatively associated with hormone therapy status. On the other hand, the human epidermal growth factor 2 positive (HER2+) subtype, which itself is negatively correlated with ER+ and ER+/PR+, is positively related to having first-degree-relative with cancer, and negatively associated with post-menopause. Single and multiple regression also identify older age to be positively correlated to ER+ and ER+/PR+ subtypes, and negatively correlated to HER2+ subtype.

Conclusion: Assuming ER+ and ER+/PR+ subtypes to have better prognostic, then post-menopause and breast-feeding are beneficial, and hormone therapy treatment is detrimental.

Keywords: breast cancer, epidemiology, regularized regression, reproductive risk factors, estrogen receptor, progesterone receptor, human epidermal receptor 2.

S5

RNA-Dizileme Verilerinin Kesikli Dağılıma Dayalı Yöntemler ile Sınıflandırılması

Dinçer Göksülük, A. Ergun Karaağaoğlu

Biyoistatistik Anabilim Dalı, Hacettepe Üniversitesi, Ankara, Türkiye

Amaç: Bu çalışmadaki amaç çeşitli makine öğrenme yöntemlerinin RNA-dizileme teknolojisi ile elde edilen gen ifade verilerinin sınıflandırılmasında nasıl performans gösterdiğini incelemek ve kesikli dağılıma dayalı yöntemlerin beklenildiği gibi daha iyi bir performans gösterip göstermediğini değerlendirmektir.

Yöntem: RNA-dizileme gen ifade verileri mikro-dizilim verilerinden farklı olarak pozitif tamsayılar olarak ölçülebilen kesikli değerlerden oluşmaktadır. Literatürde kullanılan yöntemlerin çok büyük bir bölümü mikro-dizilim gen ifade verileri gibi sürekli dağılım özelliği gösteren ve aşırı yaygınlık (overdispersion) göstermeyen veri yapıları için uygundur. RNA-dizileme verileri ise varyansı ortalamasından daha büyük değerler alan ve aşırı yaygın dağılım gösteren kesikli verilerden oluşmaktadır. Bu nedenle RNA-dizileme verilerinin sınıflandırılmasında iki farklı yol izlenebilir: (i) kesikli dağılımlara dayalı yöntemleri kullanmak ve (ii) RNA-dizileme verilerine sürekli dağılım sergileyecek şekilde dönüşüm uygulamak [1]. Kesikli dağılımlar için önerilen sınıflama algoritmaları Poisson Doğrusal Ayırma Analizi (PDAA) [2] ve Negatif Binom Doğrusal Ayırma Analizi (NBDAA)'dir [3]. Ayrıca, verilen değişken kümesi içerisinden belirli bir alt grubu seçerek daha az sayıda değişken kümesi ile sınıflama yapan (sparse) seyrek PDAA (sPDAA) ve seyrek NBDAA (sNBDAA) yöntemleri de değerlendirilmiştir. Dönüşüm uygulanan RNA-dizileme verilerinin sınıflandırılmasında ise En Yakın Küme Merkezleri (EYKM) [4] ve Voom-tabanlı En Yakın Küme Merkezleri (voomEYKM) [5] algoritmaları kullanılmıştır. Bu çalışmada, kapsamlı bir benzetim çalışması yürütülmüş ve farklı senaryolar altında (örneklem genişliği, aşırı yaygınlık miktarı, anlamlı gen sayısı vb.) incelenen sınıflama algoritmalarının performansları değerlendirilmiştir. Yöntemlerin performansları ayrıca gerçek veri setleri kullanılarak karşılaştırılmıştır. Çalışma kapsamında kullanılan ve literatürde sıklıkla tercih edilen çeşitli sınıflama algoritmaları bir R/Bioconductor paketi olan MLSeq paketi ile araştırmacıların kullanımına sunulmuştur [1].

Bulgular: Sınıflama performansları açısından değerlendirildiğinde NBDAA yöntemi aşırı yaygınlık miktarı arttıkça en yüksek başarı yüzdesine sahip yöntem olmuştur. Bizim önerdiğimiz yöntem olan sNBDAA yöntemi ise çoğu durumda daha az sayıda gen kullanarak NBDAA'dan daha iyi performans sergilemiştir. Modelde kullanılan gen sayısı dikkate alındığında seyrek yöntemler içerisinde en az sayıda gen kullanan yöntem voomEYKM yöntemi olmuştur. Bazı durumlarda voomEYKM yöntemi en az sayıda gen ile en yüksek başarı oranını gösterebilmiştir. Benzetim çalışması sonuçlarına göre yöntemlerin performansları üzerinde etkili olan en temel faktörlerin aşırı yaygınlık düzeyi ve veri setinde yer alan anlamlı genlerin miktarı olduğu görülmüştür.

Sonuç: RNA-dizileme verilerinin yapısının bozulmadan kesikli dağılımlar ile sınıflandırılması metodolojik açıdan daha uygundur. Sınıflama performansları açısından değerlendirildiğinde ise veri dönüşümlerine bağlı olarak yöntemleri aşırı uyum sorunu gösterebileceği dikkate alınmalıdır. Kesikli dağılımlar ile sınıflama yapılmak istendiğinde ise, özellikle aşırı yaygınlık söz

konusu olduğunda, NBDAA ve sNBDAA yöntemleri tercih edilmelidir. voomNSC yöntemi en az sayıda gen seçerek sınıflama yapabildiği için bu yöntem ayrıca değişken seçim algoritması amacı ile de kullanılabilir.

Anhtar Kelimeler: RNA-Dizileme, gen ifade verileri, sınıflama, NBDAA, PLDA

Kaynaklar

1. Goksuluk, D., Zararsiz, G., Korkmaz, S., Eldem, V., Klaus, B., Ozturk, A., & Karaagaoglu, A. E. (2019). MLSeq: Machine Learning Interface to RNA-Seq Data.
2. Witten, D. M. (2011). Classification and clustering of sequencing data using a Poisson model. *The Annals of Applied Statistics*, *5*(4), 2493-2518.
3. Dong, K., Zhao, H., Tong, T., & Wan, X. (2016). NBLDA: negative binomial linear discriminant analysis for RNA-Seq data. *BMC Bioinformatics*, *17*(1), 369.
4. Tibshirani, R., Hastie, T., Narasimhan, B., & Chu, G. (2003). Class prediction by nearest shrunken centroids, with applications to DNA microarrays. *Statistical Science*, *18*(1), 104-117.
5. Zararsiz, G., Goksuluk, D., Klaus, B., Korkmaz, S., Eldem, V., Karabulut, E., & Ozturk, A. (2017). voomDDA: discovery of diagnostic biomarkers and classification of RNA-seq data. *PeerJ*, *5*, e3890.

S6

Bilgilendirici Küme Boyutlu Verilerin Modellemesine Bir Yaklaşım Önerisi

Betül DAĞOĞLU HARK[1], Zeliha Nazan ALPARSLAN[2]

[1]Fırat Üniversitesi Tıp Fakültesi, Biyoistatistik ve Tıbbi Bilişim Anabilim Dalı, Elazığ
[2]Çukurova Üniversitesi Tıp *Fakültesi, Biyoistatistik Anabilim Dalı, Adana*

Amaç: Kümelenmiş verilerdeki en önemli özellik, gözlemlerin bilinen kümeler içinde birbirine bağlı olmasıdır. Kümelenmiş veri analizi, hem bireysel düzeydeki faktörlerin hem de küme düzeyindeki faktörlerin sonuç değişkeni üzerindeki etkilerinin incelenmesini sağlar. Kümelenmiş verilerdeki bu etkilerin marjinal modellenmesinde genelleştirilmiş kestirim eşitliği (GEE) kullanılır. Elde edilen sonuç değişkeni ile kümenin büyüklüğü arasındaki bağımlılığı ifade eden bir kavram olan bilgilendirici küme boyutu (ihmal edilemez küme boyutu), kümelenmiş veri yapılarında özel bir durumdur. Bu durumda, marjinal model esas olarak iki farklı yaklaşımla uydurulur ve tahminler yapılır. Bu yaklaşımlar, küme-içi yeniden örneklendirme (within-cluster resampling, WCR) ve küme ağırlıklı genelleştirilmiş kestirim eşitliği (cluster weighted generalized estimating equations, CWGEE)'dir.

Kayıp veri mekanizmasına bağlı olarak kümelenmiş verilerin marjinal modellemesinde kullanılan yöntemlerden biri, iki katlı robust GEE (DR-GEE)'dir. Bağımsız çalışan korelasyon matrisi temeline dayanan DR-GEE yaklaşımının küme boyutuna göre ağırlıklandırılması önerimizin(DR-GEE(1)), mevcut yaklaşımlarla karşılaştırılması amaçlanmaktadır.

Yöntem: Periodontal hastalık varlığına neden olan faktörleri belirlemek amacıyla 206 bireyin dahil edildiği bir çalışmanın verileri kullanılmıştır. Çalışmada incelenen kümeyi oluşturan bireylerin alt çene dişleri ile ilgilenildiğinden maksimum küme büyüklüğü 16, minimum küme büyüklüğü 4'tür. Periodontal hastalık varlığı 6 açıklayıcı değişken ile modellenmektedir. Bu değişkenler; yaş, sigarayı bırakma durumu, sigara içme durumu, plak varlığı, kırsal bölgede yaşama ve beyaz ırk mensubu olmak şeklindedir. Parametre tahminleri öncelikle GEE, WCR ve CWGEE yaklaşımı ile elde edilmiş, daha sonra çalışmanın amacına yönelik olarak yeni önerilen çıkarsama yöntemine(DR-GEE(1)) ilişkin değerlendirmeler yapılmış ve dört yöntemle elde edilen bulgular değerlendirilmiştir.

Bulgular: GEE yaklaşımı ile elde edilen parametre tahminlerinin standart hataları diğer üç yaklaşımdan farklı değildir. Ancak literatürde yapılan simülasyon çalışmaları, bilgilendirici küme boyutu varlığında GEE yaklaşımının parametre tahminlerinde yanlılığa neden olduğunu göstermiştir. Bununla beraber, küme boyutunun etkisini düzeltmeye yönelik olarak geliştirilen WCR ve CWGEE yaklaşımları, benzer standart hatalara sahip tahminler vermiştir. Bu iki yönteme ek olarak, çalışma kapsamında geliştirilen DR-GEE(1) yaklaşımı ile elde edilen parametre tahminlerinin standart hataları, literatürde kullanılan diğer yöntemlere göre daha düşüktür.

Sonuç: Yapılan gerçek veri çalışması ile bilgilendirici küme boyutuna sahip kümelenmiş verilerin modellenmesi için yeni yaklaşım önerisi hâlihazırda kullanılan yöntemlerle kıyaslanmıştır. Elde edilen bulgular ışığında, önerilen yöntemin daha düşük standart hataya sahip tahminler verdiği görülmüştür.

Anahtar Kelimeler: Kümelenmiş veri, bilgilendirici küme boyutu, GEE, DR-GEE, WCR, CWGEE

S6

A Proposal for the Modeling of Informative Cluster Size Data

Betül DAĞOĞLU HARK[1], Zeliha Nazan ALPARSLAN[2]

[1]Fırat University Faculty of Medicine, Department of Biostatistics and Medical Informatics, Elazığ
[2] Çukurova University Faculty of *Medicine, Department of Biostatistics, Adana*

Aim: The most important feature of clustered data is that the observations are interconnected within known clusters. Clustered data analysis allows the analysis of the effects of both individual-level and cluster-level factors on the outcome variable and generalized estimation equation (GEE) is used in the marginal modeling of these effects. The informative cluster size (the non-ignorable cluster size), a concept that expresses the dependence between the outcome variable and the size of the cluster, is a special case in the clustered data structures. In this case, the marginal model is essentially fitted and predictions are performed with two different approaches: within-cluster resampling (WCR) and cluster weighted generalized estimating equations (CWGEE).

Doubly robust GEE (DR-GEE) is a method used in the marginal modeling of clustered data with missing data mechanism. We propose that the DR-GEE approach based on independent working correlation matrix be weighted according to the cluster size (DR-GEE(1)) and aim to compare the results with the current approaches.

Method: Data from a study including 206 individuals were used to determine the factors that cause periodontal disease. The maximum cluster size was 16 and the minimum cluster size was 4, since only the lower jaw teeth of the individuals involved in the study were considered. The presence of periodontal disease is modeled by 6 explanatory variables. These variables are; age, smoking cessation, smoking status, plaque presence, living in rural areas and being a member of white race. Parameter estimations were firstly obtained by GEE, WCR and CWGEE approaches, and then, calculations were made using the proposed inference method (DR-GEE (1)) and the results obtained by four methods were compared.

Results: The standard errors of the parameter estimates obtained by the GEE approach are not different from the other three approaches. However, simulation studies conducted in the literature show that GEE approach causes bias in parameter estimation in the presence of informative cluster size. The WCR and CWGEE approaches developed to improve the effect of cluster size yielded estimates with similar standard errors. In addition to these two methods, the standard errors of the parameter estimates obtained with the proposed method, i.e. DR-GEE(1) were lower than the other methods currently used.

Conclusion: The newly proposed approach for modeling clustered data with informative cluster size is applied on real data and comparison was made with the methods currently used. According to the results, it was seen that the proposed method gives estimates with lower standard errors.

Keywords: Clustered data, informative cluster size, GEE, DR-GEE, WCR, CWGEE

S7

Doğrusal Olmayan Temel Bileşenler Analizi İçin Yapay Sinir Ağları Yaklaşımı

Canan DEMİR[1], Sıddık KESKİN[2]

[1]*Van Yüzüncü Yıl Üniversitesi, Sağlık Hizmetleri Meslek Yüksek Okulu, Van*
[2]*Van Yüzüncü Yıl Üniversitesi, Tıp Fakültesi, Biyoistatistik Anabilim Dalı, Van*

Amaç: Doğrusal Olmayan Temel Bileşenler Analizi (DOTBA), aralarında doğrusal veya doğrusal olmayan ilişki bulunan veri kümeleri için rakamsal ve görsel sonuçlar veren açıklayıcı bir boyut indirgeme yöntemidir. DOTBA'da, sürekli değişkenlerin yanı sıra, kategorik ve sıralı değişkenler de aynı anda analize dâhil edilebilir. Analizde gözlenen değişkenler arasındaki ilişkilerin doğrusal olduğu varsayımına gerek yoktur. Bu çalışmada amaç yapay sinir ağları temelinde, doğrusal olmayan boyut indirgeme için DOTBA yönteminin etkinliğini göstermektir.

Yöntem: DOTBA'nın uygulaması için giriş katmanı, kodlama katmanı, darboğaz katmanı, kod çözme katmanı ve çıkış katmanı olmak üzere toplam 5 katman içeren bir ağ mimarisi kullanıldı. Nöron sayısı 18-6-2-6-18 olarak seçilen ağ mimarisinde iterasyon sayısı 3000 olarak belirlendi. Kodlama ve kod çözme katmanında doğrusal olmayan hiperbolik tanjant fonksiyonu, diğer katmanlarda ise doğrusal fonksiyon kullanıldı. Eğitim algoritması olarak CGD (Conjugate Gradient Descent) algoritması tercih edildi.

Bulgular: Uygulamada veri seti olarak 422 hastaya ait 19 değişkenli hipotiroidi verisi kullanılmıştır. Yapay sinir ağlarında doğrusal olmayan temel bileşenler analizi kullanılarak elde edilen sonuçlar, tablo ve grafikler halinde sunularak yorumlanmıştır. İlk iki temel bileşen, DOTBA' da toplam varyansın %95.65'ini açıklamıştır.

Sonuç: DOTBA'nın yüksek bir varyans açıklama oranı ile başarılı sonuçlar verdiği ve böylece ileriye dönük yapılacak tahminlerde kullanılabileceği vurgulanmıştır.

Anahtar Kelimeler: Doğrusal Olmayan Temel Bileşenler Analizi, Hipotiroidi, Temel Bileşenler Analizi

S8

Risk Kestiriminde Kestirim Eğrisi ve Gizli Sınıf Kümeleme Analizi'nin Birlikte Kullanımı

Didem DERİCİ YILDIRIM[1], Merve TÜRKEGÜN[1], Bahar TAŞDELEN[1], Tuğba ARIKOĞLU[2], Semanur KUYUCU[2]

[1] Mersin Üniversitesi Tıp Fakültesi, Biyoistatistik Anabilim Dalı, MERSİN
[2] Mersin Üniversitesi Tıp Fakültesi, Çocuk Sağlığı ve Hastalıkları Anabilim Dalı, MERSİN

Amaç: Hastalık riski hakkında bilgi veren biyomarkerların seçimi kronik hastalıkların tedavisinde önemli bir yere sahiptir. Bu çalışmada; "Referans Test" olmadığı durumda birey odaklı sınıflama yaklaşımlarından biri olan Gizli Sınıf Kümeleme Analizi (GSKA) ile yine her bir birey için riskin hesaplanabildiği Kestirim Eğrisi (KE) yönteminin birlikte kullanımı araştırılacak olup, kronik bir hastalık olan astım tanısı koymada klinik öneme sahip parametrelerden yararlanılacaktır.

Yöntem: Çalışmaya 6-24 yaş arasındaki astım tanısı alan ve bir yıldır takibi yapılan 60 hasta ve sağlıklı 51 birey dahil edilmiştir. Astımı predikte etmede önemli değişkenler olan Exhaled nitric oxide (FeNO), Total IgE ve Asymmetric Dimethylarginine (ADMA) biyomarkerları kullanılmıştır. Hastaların düşük ve yüksek hastalık risk eşikleri, yaş ve cinsiyet düzeltmesi yapılan lojistik regresyon analizi ile belirlenmiştir. KE çizilerek, düşük ve yüksek riskli hastaları tahmin etmede kullanılacak markerlar belirlenmiştir. Marker seçimleri, sınıflama performans istatistiklerinden True Positive Fraction (TPF) ≥0.60 ve False Positive Fraction (FPF) ≤0.30 kriterleri dikkate alınarak değerlendirilmiştir. Gizli sınıf kümeleme analizi ile elde edilen iki sınıf için KE çizilmiştir. Her iki yöntemin kestirim sonuçları karşılaştırılmıştır.

Bulgular ve Sonuç: TPF ve FPF değerleri dikkate alındığında (TPF=0.88 [95%CI=0.765-0.954]), FPF=0.19.[95%CI=0.094-0.359]) FeNO markerı düşük riskli hastaları sınıflandırmada kullanılacak en iyi biyomarkerdır. Yüksek riskli hastaları sınıflandırmada kullanılacak en iyi marker ise Total IgE'dir (TPF=0.917 [95%CI=0.741-0.950], FPF=0.118 [95%CI=0.010-0.110]). GSKA sonucunda bireylerin %40.5'i birinci örtük sınıfa, 59.5'i ise ikinci örtük sınıfa atanmıştır. GSKA ile elde edilen yeni risk grupları için yapılan KE sonuçlarına göre de yine FeNO ile yüksek ve düşük riskli hastalar sınıflandırılabilmektedir. Ancak aynı başarı ADMA için gözlenememiştir.

Anahtar Kelimeler: Kestirim Eğrisi, Gizli Sınıf Kümeleme Analizi, Biyomarker

S8

Usage of Predictiveness Curve and Latent Cluster Analysis with Together for Risk Estimation

Didem DERICI YILDIRIM[1], Merve TURKEGUN[1], Bahar TASDELEN[1], Tugba ARIKOGLU[2], Semanur KUYUCU[2]

[1] *Mersin University Faculty of Medicine, Biostatistics and Medical Informatics, Mersin*
[2] *Mersin University Faculty of Medicine, Pediatrics, Mersin*

Aim: The selection of biomarkers which provide information about disease risk has an important role in the treatment of chronic diseases. In this study, it is aimed to use the Latent Class Cluster Analysis (LC Cluster) which is a person-centered method and PC which is calculated the risk for each individual with together for diagnosis of asthma utilizing the clinical parameters for this disease.

Method: The children and young adults attending to the department with asthma and aged 6–24 years were enrolled. They had been followed up in our clinic for at least one year. Sixty patients with asthma and fifty one healthy people were included to the study. Significant parameters of asthma prediction such as Exhaled nitric oxide (FeNO), Total IgE and Asymmetric Dimethylarginine (ADMA) were evaluated. The low and thresholds of disease were calculated with multiple logistic regression adjusted by age and gender. True Positive Fraction (TPF) ≥0.60 and False Positive Fraction (FPF) ≤0.30 were considered for selection of significant biomarkers. Predictiveness Curve was drawn for reference groups and LC Clusters. The results of PC were compared for biomarkers.

Results and Conclusion: According to TPF and FPF, Feno was the best biomarker to classify low-risk patients. TPF (0.88.[95%CI=0.765-0.954]) and FPF (0.19.[95%CI=0.094-0.359]) were calculated for FeNO. Total IgE is the best biomarker to classify high-risk patients. TPF=0.917 [0.741-0.950] ve FPF=0.118[0.010-0.110] were calculated for Total IgE. The 40.5 percentage of cohort was in the first latent class and 59.5 percentage of cohort was in the second latent class according to LC Cluster analysis. FeNO was the best biomarker again based on PC results with latent clusters for both risk groups. However, ADMA was not a significant parameter to predict asthma risk.

Keywords: Predictiveness Curve, Latent Class Cluster Analysis, Biomarker

S9

Aynı Veri Setleri Üzerinde Bazı Etki Büyüklüğü Hesaplama Yöntemlerinin Tahmin Başarılarının Karşılaştırılması

Duygu KORKMAZ[1,2], İlker ÜNAL[1]

[1]Çukurova Üniversitesi Tıp Fakültesi Biyoistatistik Anabilim Dalı, Adana
[2]Yüzüncü Yıl Üniversitesi Tıp Fakültesi Tıp Eğitimi Anabilim Dalı, Van

Giriş: İstatistiksel analiz sonuçlarını rapor ederken, gruplar arasında gözlenen farklılıkların istatistiksel olarak anlamlı olmasının yanı sıra, bu gözlenen farklılıkların pratik öneme sahip olduğunu gösterecek bazı etki büyüklüklerini sunmak da çok önemlidir. Etki büyüklüğü hesaplama yöntemleri arasında en sık kullanılan ve en iyi bilinenleri Eta kare ve Cohen'dir. Kullanılan diğer bazı yöntemler ise Epsilon kare ve Omega karedir.

Amaç: Bu çalışmada amaç, sık kullanılan dört etki büyüklüğü hesaplama yöntemlerinin popülasyon değerini tahmin etme başarılarının farklı senaryolarda üretilen verilerle karşılaştırılmasıdır.

Yöntem: Farklı grup sayısı (3, 4, 5, 10) ve farklı etki büyüklüğü derecesi (küçük, orta, büyük) senaryolarında 1.000.000'luk veri setleri, popülasyon veri setleri olarak üretilmiştir. Bu veri setlerinden istenilen örnek büyüklüğünde (5, 10, 20, 30, 50, 100) örneklemler çekilmiştir. Çekilen bu örneklemlerden Eta kare, Cohen, Epsilon kare ve Omega kare yöntemleri ile hesaplanan etki büyüklüklerinin popülasyon etki büyüklüklerini tahmin etme başarısı karşılaştırılmıştır.

Bulgular: Genel olarak Omega kare ve Epsilon kare hesaplama yöntemleri popülasyon parametre tahmininde Eta kare ve Cohen hesaplama yöntemlerinden daha başarılı sonuçlar vermiştir. Küçük orta ve büyük etki büyüklüklerini tahmin etme başarısı en iyi olan yöntem Omega kare olarak belirlenmiştir. Grup sayısı arttıkça (5,10) Epsilon kare tahminleri de Omega kareye benzer şekilde tüm örneklerde (5,10,20,30,50,100) en iyi tahminlere ulaşmıştır. Eta kare ve Cohen f tahminleri ise gözlem sayıları küçük olduğunda başarısız sonuçlar vermiştir.

Sonuç: Sık kullanılmasına rağmen Eta kare ve Cohen f'in, popülasyona ait etki büyüklüğünün gerçek değerini tahmin etme başarısının Omega kare ve Epsilon kare kadar iyi olmadığı görülmüştür. Grup sayısı 10 olduğunda ise çalışmaya alınan etki büyüklüğü hesaplama yöntemlerinin hepsinde tahmin başarısı oldukça düşmektedir (gerçek değerden uzaklaşmaktadırlar).Etki büyüklüğünün derecesi tahminlerdeki başarılar üzerinde etkili değildir. Tüm senaryolar dikkate alındığında Omega ve Epsilon karenin diğer etki büyüklüğü hesap yöntemlerinden daha yansız ve birbirlerine yakın sonuçlar verdiği görülmüştür.

Anahtar Kelimeler: Etki büyüklüğü, Eta kare, Cohen, Epsilon kare, Omega kare

S9

Comparison of Some Effect Size Calculation Methods in Estimation of Effect Size on Same Data Sets

Duygu KORKMAZ[1,2], İlker ÜNAL[1]

[1]*Çukurova Üniversitesi Tıp Fakültesi Biyoistatistik Anabilim Dalı, Adana*
[2]*Yüzüncü Yıl Üniversitesi Tıp Fakültesi Tıp Eğitimi Anabilim Dalı, Van*

Aim: In reporting the results of statistical analysis, in addition to differences observed between groups it is important to present some effect sizes measures to show that these observed differences are of practical importance. The most commonly used and best-known methods of effect size measures are Eta square and Cohen. Other methods used are Epsilon square and Omega square. The aim of this study is to compare the performance of four commonly used effect size estimation methods using simulated data with a variety of experimental conditions.

Method: Data sets containing 1,000,000 observations are generated in different groups (3, 4, 5, 10) and different effect size (small, medium, large) scenarios as population data sets. Samples with the desired sample size (5, 10, 20, 30, 50) are taken from these data sets. From these samples Eta square, Cohen, Epsilon square and Omega square methods are compared to estimate the population effect size successfully.

Results: In general, Omega square and Epsilon square calculation methods yielded more successful results in population parameter estimation than Eta square and Cohen calculation methods. Omega square is the best method to estimate small, medium and large effect sizes. As the number of groups increased (5,10), Epsilon square estimates reached the best estimates in all samples (5,10,20,30,50,100) similar to Omega square. Estimates of Eta square and Cohen f failed when the number of observations was small.

Conclusion: Although frequently used, the success of Eta square and Cohen f to estimate the true value of the effect size of the population was not as good as Omega square and Epsilon square. When the number of groups is 10, the predicted success of the effect size calculation methods decreases considerably (they deviate from the actual value). When all scenarios are considered, it is seen that Omega and Epsilon squares are more unbiased and close to each other than other effect size calculation methods.

Keywords: Effect size, Eta square, Cohen, Epsilon square, Omega square

S10

Randomize Olmayan Klinik Çalışmalarda En Uygun Eşleştirme Analizi için Makine Öğrenme Algoritmaları ile Yeni Propensity Skor Tahmin Modellerinin Geliştirilmesi ve R Shiny ile İnteraktif Bir Web Uygulaması

Emre DEMİR[1], Serdal Kenan KÖSE[2]

[1]Hitit Üniversitesi, Tıp Fakültesi, Biyoistatistik Anabilim Dalı, Çorum
[2]Ankara Üniversitesi, Tıp Fakültesi, Biyoistatistik Anabilim Dalı, Ankara

Amaç: Deneysel araştırmalarda deney sonucunu etkileyebilecek karıştırıcı etkenler randomizasyon ile kontrol edilirken gözlemsel çalışmalarda deneklerin araştırma gruplarına atanma olasılığı eşit olmadığı için karıştırıcı değişkenler göz ardı edilerek ulaşılan sonuçlarda yanlılık bulunmaktadır. Bu nedenle gözlemsel çalışmalarda bir tedavinin nedensel etkisi araştırılırken yaygın olarak Propensity skor (PS)'a dayalı eşleştirme analizi kullanılmaktadır. Bu çalışmada PS tahmini için klasik yöntem olan Lojistik Regresyon (LR)'a alternatif olarak Makine öğrenme (ML) ve bazı yaygın kullanılan yöntemlerin başarısı benzetim çalışması ile değerlendirilerek yeni bir kombine PS tahmin modeli geliştirmek amaçlanmıştır.

Yöntem: Birinci benzetim çalışmasında (10000 benzetim) etkileşim etkilerinin ve doğrusal olmayan ilişkilerin bulunduğu modellerde LR ile rekabet edebilecek algoritmaların belirlenmesi için öncelikle LR için avantajlı olan modelde (Senaryo A: sadece temel etki ve toplamsal model) n=500 ve n=1000 örneklem büyüklüklerinde üretilen veri seti kullanılarak 21 farklı tahminci ve LR ile PS'ler hesaplanmıştır. PS'ler arasındaki uyumu değerlendirmek için sınıf içi korelasyon katsayısı kullanılmıştır. İkinci benzetim çalışmasında (1000 benzetim) A senaryosuna ek olarak etkileşim etkilerinin ve kuadratik terimlerinde bulunduğu B ve C senaryolarında ilk benzetim sonucunda başarılı bulunan 12 algoritma ve LR ile elde edilen PS'ler kullanılarak eşleştirme analizleri yapılmıştır. Elde edilen denge (Tedavi grubunda ortalama tedavi etkisi, yanlılık, standardize edilmiş ortalamalar arasındaki farklar, varyans oranları ve kolmogorov smirnov değerleri) sonuçlarına göre başarılı tahminciler ile yeni bir kombine tahminci model oluşturularak üçüncü benzetim çalışması yapılmıştır. Üçüncü benzetim çalışmasında (1000 benzetim) üç farklı senaryoda kombine tahminci ve LR ile elde edilen PS'ler kullanılarak yapılan eşleştirme analizi denge sonuçları karşılaştırılmıştır. PS tahmini ve eşleştirme analizi için interaktif web uygulaması R Shiny paketi ile geliştirilmiştir.

Bulgular: Birinci benzetim sonucunda LR ile en uyumlu algoritmalar BAYESGLM, GLMNET, GAM, STEPAIC, DBARTS, GBM, EARTH, QDA, KSVM, CFOREST, NNET ve IPREDBAGG olarak belirlenmiştir. İkinci benzetim sonucunda 1000 örneklem büyüklüğünde A senaryosunda LR ile rekabet edebilen, B ve C senaryolarında LR'den daha başarılı olan KSVM, NNET, DBARTS, QDA, EARTH ve STEPAIC algoritmaları ile kombine tahmin modeli oluşturulmuştur. Üçüncü benzetim sonucunda sırasıyla A, B ve C senaryolarında LR %0,496; %10,160; %17,762 ve kombine tahminci %0,527; %7,405; %7,672 yanlılık ile kestirimde bulunmuştur.

Sonuç: Denge sonuçlarına göre n=500 ve n=1000 örneklem büyüklüğünde kombine tahmin modelinin senaryo zorlaştıkça LR'ye göre daha az yanlılıkla tedavi etkisini tahmin ettiği belirlenmiştir. Özellikle n=1000 örneklem büyüklüğünde en zor senaryo olan C senaryosunda kombine tahmincinin LR'den %10,09 daha az yanlı kestirimde bulunması ve LR'nin yüksek bir yanlılık kestiriminde bulunması ML veya diğer algoritmaların PS tahmininde kullanılmasının yanlılığı azaltarak gözlemsel çalışmaların güvenirliğini arttıracağını göstermektedir.

Anahtar Kelimeler: Propensity skor, eşleştirme, lojistik regresyon, makine öğrenme, gözlemsel çalışmalar

S11

Genetik ve Klinik Verilerin Birleştirilmesinde Regresyon Modeli Artık Değerlerinin Kullanımı

Erdal COŞGUN

Data Scientist, Genomics Team, Microsoft Research & AI, Redmond, WA, USA

Amaç: Büyük ölçekli tıbbi çalışmalarda en önemli veri kaynağı elektronik kayıt sistemleridir. Bu kayıt sistemlerinde saklanan bilgiler günlük takip bilgileri, verilen ilaçların dozları, görüntüleme sistemlerinin verileri, uygulanacak tedavinin detayları ve son zamanlarda genetik testlerden elde edilen büyük boyutlu verilerdir. Bütün bu veri kaynaklarının hızlı şekilde ayrı ayrı analizi ve birlikte modellenmesi gereklilik haline gelmiştir. Biyoistatistik uzmanlarının da klinik veriler ve diğer veri kaynaklarını birleşleştirerek sağlık uzmanlarına daha yüksek doğruluk oranına sahip karar destek modellerini oluşturması beklenmektedir. Bu çalışmada elektronik kayıt sistemlerinden elde edilen bu kaynakların aynı anda analiz edilmesi için yeni yeni kullanılan yöntemlerden "artık değer yaklaşımı" çalışılmıştır. Sonuçlar genetik varyasyonlar ve klinik değişkenlerin ayrı ayrı kullanıldığı model ile karşılaştırılmıştır.

Yöntem: Çalışma ilaç dozu tahminleme çalışması olarak dizayn edilmiştir. VarSim yöntemi ile elde edilen 50 yapay tüm genom sekans verisi ve yine yapay yaş, boy, kilo, ALT, GGT test sonuçları doz tahminlemesi amacıyla kullanılmıştır. Tüm genom sekans verisinin boyutu hasta başına 100 GB ve analiz edilen veri matrisi boyutu 56 milyon sütun ve 50 satırdan oluşmaktadır. Analizler Microsoft Azure bulut platformundaki SQL Veritabanı, Microsoft Genomics service ve Veri Bilimi sanal bilgisayarlarında Python programlama dili ile gerçekleştirilmiştir. Kullanılan en yüksek hafıza değeri 1.6 terabyte ve en yüksek disk bölümü 2 terabayt'tır. Tüm analizler 1000 defa tekrarlanmış ve ortalama değerler raporlanmıştır. Çalışmanın ilk aşamasında genetik veriler bağımsız bileşenler analizi ile boyutları azaltılarak (500 bileşene) klinik verilerle birlikte regresyon modellerine dahil edilmiştir. Bir sonraki aşamada elde edilen birinci seviye regresyon modellerinin artık dağılımları bağımlı, sadece genetik varyasyon verileri ise bağımsız değişkenler olarak kullanılmıştır. Kullanılan tahmin modelleri Support Vector Machine (SVR), Random Forest (RFR), Generalized Boosted regresyon (GBR) algoritmalarıdır.

Bulgular: Elde edilen sonuçlara göre artık değer yaklaşımı için regresyon modelleri içerisinde ortalama en yüksek açıklayıcılık katsayısı (R^2) değeri %72.3 ile GBR'de, en düşük değer %65.3 ile SVR'de gözlenmiştir. Klasik yaklaşım (ayrı ayrı değişken kullanımı) ile elde edilen modelin ortalama açıklayıcılık katsayısı ise en yüksek %61.4 de kalmıştır (Tablo 1).

Tablo 1. ML algoritmalarına göre "Artık yaklaşımı" ve "klasik yöntemin" ilaç doz tahminleme performansları

Performans Ölçütü (Ortalama)	RVR	SVR	GBR	Klasik Yaklaşım
R^2	68.9	65.3	72.3	61.4

Sonuç: Bu çalışmanın en önemli özelliği büyük boyutlu, çok sayıdaki değişkenin olduğu regresyon modellerinde az sayıdaki klinik verinin, genetik veri içerisinde etki kaybına uğradığını göstermesidir. Uygulanan yaklaşım çoklu gen analizi (Polygenic), gen-çevre etkileşim, görüntü, hasta takip ve biyokimyasal test sonuçları gibi farklı veri kaynaklarını birleştirmede araştırıcılara kolaylık sağlayacaktır. Tüm bunların yanında hesaplama platformları bakımından lokal bilgisayarlar üzerinde gerçekleştirilmesi çok zor olan bu analizlerin zaman içerisinde bulut sistemlerine aktarılması önerilmektedir.

Anahtar Kelimeler: Yeni Nesil Sekanslama, Makine Öğrenmesi, Elektronik Sağlık Kaydı

Kaynaklar

1. Regier, A. A., Farjoun, Y., Larson, D. E., Krasheninina, O., Kang, H. M., Howrigan, D. P., … Hall, I. M. (2018). Functional equivalence of genome sequencing analysis pipelines enables harmonized variant calling across human genetics projects. Nature Communications, 9(1), 1–8. https://doi.org/10.1038/s41467-018-06159-4
2. Van der Auwera, G. A., Carneiro, M. O., Hartl, C., Poplin, R., del Angel, G., Levy-Moonshine, A., … DePristo, M. A. (2013). From fastQ data to high-confidence variant calls: The genome analysis toolkit best practices pipeline. Current Protocols in Bioinformatics, (SUPL.43)
3. Voss, K., Gentry, J., & Van der Auwera, G. (2017). Full-stack genomics pipelining with GATK4 + WDL + Cromwell. https://doi.org/https://doi.org/10.7490/f1000research.1114631.1
4. Talwalkar, Ameet, et al. "SM a SH" Bioinformatics 30.19 (2014): 2787-2795
5. Bartenhagen, Christoph, and Martin Dugas. "RSVSim" Bioinformatics 29.13 (2013): 1679-1681
6. Mu, John C., et al. "VarSim" Bioinformatics 31.9 (2014): 1469-1471.

S11

Use of Regression Model's Residual Values in Combining Genetic and Clinical Data

Erdal COSGUN

Data Scientist, Genomics Team, Microsoft Research & AI, Redmond, WA, USA

Aim: The most important data source in large-scale medical studies is electronic recording systems. The information stored in these recording systems are: daily follow-up information, applied drug dosages, data from imaging systems, details of treatment to be applied, and large-scale data obtained recently from genetic testing. Rapid analysis and co-modeling of all these data sources has become critical. Biostatistics experts are also expected to combine clinical data and other data sources to form decision support models with higher accuracy for healthcare professionals. In this study, the `residual value approach`, which is one of the methods newly used to analyze these sources obtained from electronic recording systems at the same time, is studied. The results were compared with the model using genetic variations and clinical variables separately.

Method: The study was designed as a drug dosage prediction study. 50 artificial genome sequence data obtained by VarSim method and artificial age, height, weight, ALT, GGT test results were used for drug dosage prediction in the study. The entire genome sequence data is 100 GB per patient and the analyzed data matrix size is 56 million columns and 50 rows. The analysis were performed with the SQL Database on Microsoft Azure cloud platform, Microsoft Genomics service and Data Science Virtual Machines with Python programming language. The highest memory value used is 1.6 terabytes and the highest disk partition is 2 terabytes. All analysis were repeated 1000 times and average values were reported. In the first stage of the study, the genetic data dimension were reduced in size by independent components analysis (500 components) and the clinical data added in the final regression models. The residual distributions of the risk regression models to the first level obtained in the next step were dependent, and only genetic variation data were used as independent variables. Prediction models that used are Support Vector Machine (SVR), Random Forest (RFR), Generalized Boosted regression (GBR) algorithms.

Results: According to the results, mean maximum coefficient of determination value was found in GBR with 72.3% and SVR with 65.3% in regression models for residuals value approach. The mean coefficient of determination of the model obtained with the use of classical approach – seperately use of variables - remained the highest at 61.4% (Table 1).

Table 1. Performances of 'Residual approach' and 'classical method' in drug dosage prediction by ML algorithms

Performance Criteria (Average)	RVR	SVR	GBR	Classic Approach
R^2	68.9	65.3	72.3	61.4

Conclusion: The most important feature of this study is that a few clinical data have lost their effect within the large genetic data in clinical regression models. The approach will provide researchers with the ability to combine different data sources such as multiple gene analysis (polygenic), gene-environment interaction, image, patient tracking, and biochemical test results. In addition, it is recommended that these analysis, which are very difficult to perform on local computers in terms of computing platforms, be transferred to cloud systems over time.

Keywords: Next Generation Sequencing, Machine Learning, Electronic Health Record

References

1. Regier, A. A., Farjoun, Y., Larson, D. E., Krasheninina, O., Kang, H. M., Howrigan, D. P., … Hall, I. M. (2018). Functional equivalence of genome sequencing analysis pipelines enables harmonized variant calling across human genetics projects. Nature Communications, 9(1), 1–8. https://doi.org/10.1038/s41467-018-06159-4
2. Van der Auwera, G. A., Carneiro, M. O., Hartl, C., Poplin, R., del Angel, G., Levy-Moonshine, A., … DePristo, M. A. (2013). From fastQ data to high-confidence variant calls: The genome analysis toolkit best practices pipeline. Current Protocols in Bioinformatics, (SUPL.43)
3. Voss, K., Gentry, J., & Van der Auwera, G. (2017). Full-stack genomics pipelining with GATK4 + WDL + Cromwell. https://doi.org/https://doi.org/10.7490/f1000research.1114631.1
4. Talwalkar, Ameet, et al. "SM a SH" Bioinformatics 30.19 (2014): 2787-2795
5. Bartenhagen, Christoph, and Martin Dugas. "RSVSim" Bioinformatics 29.13 (2013): 1679-1681
6. Mu, John C., et al. "VarSim" Bioinformatics 31.9 (2014): 1469-1471.

S12

Quasi-Least Squares Regression Method with Real Data and Simulation Studies

Erdoğan ASAR[1], Erdem KARABULUT[2]

[1] *Turkish Statistical Institute*
[2] *Department of Biostatistics, Hacettepe University, Ankara, Turkey*

Aim: In this study, it is aimed to introduce Quasi-Least Squares Regression (QLS) obtained by expanding Generalized Estimation Equations (GEE) and to compare QLS and GEE results obtained from the applications.

Method: QLS is not widely used in our country. This method is an associated data analysis method based on a two-stage calculation. For the purposes of this study, the real data related health and the data obtained from a simulation study were analyzed and the results were evaluated. In the study conducted with real data; the effects of time and group variables on four dependent variables were investigated using 114 observations of 38 patients who had orthodontic treatment. In the simulation study, 9 datasets with 1000 replicates were produced for the 3 correlation structures and 3 different values determined. 36 cases results obtained from applying four working correlation structures on per data set were evaluated.

Findings and Results: As a result of the studies, according to the simulation study; In general, QLS has superiority over GEE in terms of the efficiency of estimations. In order to make better comparisons of GEE and QLS results, "Markov" working correlation structure should be applicable to GEE as well as QLS. Within the scope of the application made with real data; Among the working correlation structures used in QLS, the highest correlation value was obtained by "Markov " working correlation structure. Single maxillofacial surgery and double maxillofacial surgery as a group variable on all outcome variables were not significant. In both real data and simulated data sets; while the convergence problem occurred for GEE in the implementation of the "tri-diagonal" working correlation structure, it was obtained that this problem did not occur in the QLS.

Keywords: Regression, Generalized Estimating Equation, Quasi-Least Squares Regression, Markov correlation structure

S13

Çok-Değerlendiricili Çok-Vakalı ROC Çalışmalarında Farklı Tasarım Türlerinin Radyolojik Görüntüleme Verileri Üzerinde Uygulanması

Fatma KAYMAKAMTORUNLARI DENİZ[1], Gülçin AYDOĞDU[2], Yunus ÇAKMAK[3], Yasemin YAVUZ[4], Elif PEKER[5]

[1]Ankara Üniversitesi, Tıp Fakültesi, Biyoistatistik Anabilim Dalı, Ankara.
[2]Artuklu Üniversitesi,İktisadi İdari Bilimler Fakültesi,İşletme Bölümü, Mardin.
[3]Ankara Üniversitesi, Tıp Fakültesi, Radyoloji Anabilim Dalı, Ankara.

Tanı testleri klinik çalışmalarda hasta ve sağlıklı bireylerin çeşitli tanı yöntemleri ve laboratuvar testlerinin sonuçlarına dayanarak hasta olup olmadıklarını belirlemede kullanılan araçlardır. Bu noktada kullanılan tanı testinin ne derece doğrulukla hasta bireyleri sağlıklı bireylerden ayırt edebildiği büyük önem taşımaktadır. Tanı testi performanslarının değerlendirilmesi için en yaygın kullanılan yöntem ROC analizidir. ROC çalışmaları radyoloji alanında, tanısal görüntüleme sistemlerinin doğruluğunu değerlendirmek için yaygın olarak kullanılmaktadır. Bu alanda yapılan çalışmalarda genellikle değerlendirmeler sübjektif kriterlere dayanır. Bu durum tanı testi sonuçlarının değerlendiricinin görsel, bilişsel ve algısal yeteneklerinden etkilenmesine neden olur. Bunun yanında bir diğer değişkenlik kaynağı da çalışmaya seçilen hastaların ve sağlıklıların bireysel özellikleridir (hastalık süresi, komorbidite farklılıkları, demografik özellikleri v.b.). Bu nedenle sübjektif kriterlere dayanan tanı testi çalışmalarında tanı testi doğruluğunu incelerken ve tanı testlerini karşılaştırırken değerlendiriciden kaynaklı yanlılığı azaltmak amacıyla Çok-Değerlendiricili Çok-Vakalı (Multi-Reader Multi-Case, MRMC) yöntemler geliştirilmiştir.

MRMC ROC çalışmalarında, farklı veri yapılarına göre düzenlenmiş çeşitli faktöriyel deneme tasarımları mevcuttur. Bu çalışmalarda iki ya da daha fazla testi karşılaştırmak için her bir tanı testi altında her değerlendiricinin her bir vakayı değerlendirdiği (a paired-case, a paired-reader) geleneksel tasarım, her bir test altında eşit sayıda farklı değerlendiriciler tarafından tüm vakaların değerlendirildiği (a paired-case, unpaired-reader) tasarım, yine her bir test için aynı değerlendiricilerin eşit sayıdaki farklı vakaları değerlendirdiği (unpaired-case, a paired-reader) tasarım, her bir değerlendiricinin tüm testleri kullanarak eşit sayıdaki farklı vaka gruplarını değerlendirdiği hibrid (hybrid) tasarım ve her bir değerlendiricinin tüm testler altında ait olduğu grup içerisindeki her bir vakayı değerlendirdiği karma (mixed) tasarım olmak üzere farklı tasarımlar mevcuttur. Bu çalışmada, MRMC veri yapısına uygun veriler üzerinde yapılan çalışmalarda, hangi durumlarda hangi tasarımların tercih edilmesi gerektiğini belirlemek için, belirtilen tasarım türleri incelenmiş olup OR (Obuchowski and Rockette) yöntemi kullanılarak, radyolojik görüntüleme verileri üzerinde bu tasarımlar uygulanmıştır.

Anahtar kelimeler: Çok-Değerlendiricili Çok-Vakalı ROC, OR

S14

Çok-Değerlendiricili Çok-Testli (MRMC) Tasarıma Sahip Tanı Doğruluğu Çalışmalarının Karşılaştırılmasında Güç Analizi

Gülçin AYDOĞDU[1], Yunus ÇAKMAK[2], Fatma KAYMAKAMTORUNLARI DENİZ[1], Yasemin YAVUZ[1]

[1]Ankara Üniversitesi, Tıp Fakültesi, Biyoistatistik Anabilim Dalı, Ankara
[2]Artuklu Üniversitesi, İktisadi İdari Bilimler Fakültesi, İşletme Bölümü, Mardin

İşlem karakteristik eğrisi (ROC) analizi, radyolojik görüntüleme çalışmaları için sürekli veya sıralı cevaplarla tanı testlerinin performansını değerlendirmek ve karşılaştırmak için iyi bir yöntemdir. Bir testin ne kadar iyi performans gösterdiğini değerlendirmek veya testlerin performansını karşılaştırmak için her değerlendirici ve her test için ROC eğrileri ve ilgili doğruluk tahminleri hesaplanır. Bu tür çalışmalarda vakalar arasında ve değerlendiriciler arasında değişkenlik vardır. Bu nedenle, sonuçların hem ilgili vakaya hem de değerlendirici kitlelerine genellenmesi önemlidir. Bu hedefi gerçekleştiren yöntemler genellikle çok değerlendiricili çok testli (MRMC) yöntemleri olarak adlandırılır.

MRMC çalışmalarında farklı tasarım türleri bulunur. Bu tasarımlardan en yaygın olarak kullanılan geleneksel tasarımdır. Bu tasarımda her okuyucunun yorumlaması gereken çok sayıda vaka bulunur. Örneklem büyüklüğü ve güç hesaplamaları, MRMC ROC çalışması için çok önemlidir. MRMC çalışmalarında örneklem büyüklüğü sadece vaka sayısına bağlı olmayıp aynı zamanda değerlendirici sayısına da bağlıdır. Toplam yorum sayısının elde edilmesi hem değerlendirici sayısının azalıp artmasına hem de vaka sayısının azalıp artmasına bağlıdır. MRMC çalışmalarını analiz etmek için genellikle, karma etkili varyans (ANOVA) analizi modelleri kullanılır. Bu çalışmalarda Obuchowski-Rockette (OR) yöntemi yaygın olarak kullanılır. OR metodu kullanılarak MRMC ROC çalışmalarının gücünü tahmin etmek için çeşitli eşitlikler kullanılır. Bu tür çalışmaların gücünü hesaplamak için parametrik olmayan yöntemler de geliştirilmiştir. Eunhee Kim vd 2014'de Delong vd önerdiği U-istatistik teorisini genişleterek AUC leri tahmin etmek ve karşılaştırmak için parametrik olmayan yeni bir yöntem önermiştir. Bu yöntem MRMC çalışmalarında gücü hesaplamada yaygın olarak kullanılır. Bu çalışmada MRMC tasarımına uygun verilerin örneklem büyüklüğünü ve gücünü hesaplamak için kullanılan yöntemler incelenmiş olup, bu yöntemler kullanılarak geleneksel tasarıma uygun veriler üzerinde tanı testlerinin güç analizleri yapılarak elde edilen sonuçlar karşılaştırılır. Yapılan uygulama Karotis darlıkları üzerine yapılan bir çalışmada, hastaların görüntüleri damar içerisindeki dansite değişimleri üzerinden değerlendirilmiştir. Hastaların değerlendirilmesinde, 2 ayrı yöntem kullanılarak ölçümler yapılmıştır. İki farklı değerlendirici tarafından yapılan değerlendirmeler sonucunda 2 tanı testi karşılaştırılmak istenir. Çalışmada 38 sağlıklı 16 hasta toplam 54 vaka bulunur. Yapılan bu çalışmada tanı testlerinin güç analizleri yapıldığında değerlendirici sayısı, vaka sayısı ve toplam yorum sayısı arttığında güç artar. Ayrıca değerlendirici sayısı ve vaka sayısı sabit iken etki büyüklüğü arttığında güç artar. Bu analizler R programı ve lowa üniversitesi radyoloji anabilim dalı tarafından yazılmış olan güç analizi ve örneklem büyüklüğü hesaplama programı kullanılarak yapıldı. Sonuç olarak toplam örneklem büyüklüğü, etki büyüklüğü veya değerlendirici sayısı arttığında güç artar.

Anahtar Kelimeler: ROC, Çok değerlendiricili çok testli tasarım, OR, U-istatistiği, güç analizi

S15

Tıbbi Görüntü İşleme İle Tanı Koymada Veri Madenciliği Yöntemlerinin Sınıflama Performanslarının Karşılaştırılması

Hanife AVCI[1], Jale KARAKAYA[1]

[1]*Hacettepe Üniversitesi Tıp Fakültesi, Biyoistatistik Anabilim Dalı, Ankara*

Amaç: Çalışmanın amacı, görüntü işleme sonucu elde edilen verilerin veri madenciliği algoritmalarındaki performanslarını incelemek ve görüntü işleme sonuçlarının, klinik bulguları (mamografi, ultrason, tanı testleri vb.)sınıflama başarısına olan katkısını araştırmaktır.

Yöntem: Dünyada ve ülkemizde en çok görülen ölüm nedenlerinden birisi kanserdir. Kanserde erken tanı koymak, ölüm oranını azaltması ve tedavi başarısını artırması açısından oldukça önemlidir. Tıbbi görüntüleme teknikleri, kanserin saptanmasında ve sınıflandırılmasında yaygın olarak kullanılmaktadır.Tanı koyma aşamasında hekimlere yardımcı olmak için bilgisayar destekli tanı modelleri (Computer-aided diagnostic models, CADx) önerilmiştir. Bu bilgisayar destekli tanı modelleri, iyi huylu ve kötü huylu kitleleri otomatik olarak ayırarak hem doğruluk hem de zaman gereksinimi açısından tanı sürecine önemli ölçüde katkı sağlar.Kanserlerin pek çok türü bulunmaktadır. Kadınlarda en sık görülen kanser türü ise meme kanseridir. Meme kanserinin araştırılmasında, radyologlara kolaylık sağlamak ve birden fazla radyoloğun incelemesi gereken görüntülerde iş yükünü azaltmak amacıyla görüntü işleme teknikleri tercih edilebilir. Bu çalışma kapsamında, meme kanseri sınıflandırması ve tanı sürecinde doktorlara yardımcı olmak amacıyla tıbbi görüntü işleme tekniklerinden yararlanılacaktır.Çalışmada açık erişimli bir veri tabanından elde edilecek meme kanseri veri setleri kullanılacaktır.Hastalardan elde edilen görüntüler üzerinde uygulanacak birinci aşama görüntü ön işlemedir.Daha sonra özellik çıkarımı işlemi aşamasında, görüntü ön işleme aşamasından geçirilerek lezyonların bulunması ve her bir görüntüye ait özelliklerin çıkarımı işlemleri gerçekleştirilecektir. Bu çalışmada, gerçek hastalık sınıflamasına ait sonuçlarla görüntü işlemeden elde edilen sayısal verilerin uyumu ve sonuçların tanı performansını ne kadar arttırdığı incelenecektir. Daha sonra klasik veri madenciliği yöntemlerinden Rastgele Orman (Random Forest), Yapay Sinir Ağları ve Destek Vektör Makineleri algoritmalarının sınıflama performansları karşılaştırılacaktır. Çalışma sonunda sınıflama yöntemlerinin performansları karşılaştırılırken duyarlılık, seçicilik, pozitif kestirim değeri, negatif kestirim değeri, doğru sınıflama oranı, ROC eğrisinin altında kalan alan, F ölçüsü gibi ölçütler kullanılacaktır.

Bulgular: Tıbbi görüntü işleme tekniklerinin hastalık durumunu belirlemedeki başarısı belirlenecek, klasik veri madenciliği algoritmaları ile sınıflama başarısındaki farklılıklar ortaya konulacaktır.

Sonuç: Sonuç olarak hastalara ilişkin görüntüler üzerinde uygulanacak görüntü işleme teknikleri ile klasik veri madenciliği algoritmalarının sınıflama performanslarını tıp alanında bir uygulama yaparak karşılaştırmak ve böylece hekime daha doğru, güvenilir ve hızlı tanı koymasında yardımcı olmaktır. Böylece her bir algoritmanın sınıflandırmadaki performansı değerlendirilmiş olacaktır.

S16

Kesişen İki Yaşam Fonksiyonu İçin Kullanılan İstatistiksel Testlerin Karşılaştırılması

Hülya ÖZEN[1], Cengiz BAL[1], Ertuğrul ÇOLAK[1]

[1]Eskişehir Osmangazi Üniversitesi Tıp Fakültesi Biyoistatistik AD, Eskişehir

Amaç: İki yaşam fonksiyonunun benzerliğinin araştırılması yaşam analizi çalışmalarında sıklıkla kullanılmaktadır. Literatürde en çok tercih edilen yöntem log-rank testidir, ancak bu yöntem orantılı hazardlar varsayımının ihlal edildiği kesişen yaşam fonksiyonlarını değerlendirmede yanıltıcı sonuçlar sunmaktadır. Kesişen yaşam fonksiyonları birçok farklı yapıda oluşabilmektedir. Tek başına tüm koşulları uygun bir şekilde değerlendirilecek bir test ise bulunmamaktadır. Bu nedenle kesişen yaşam fonksiyonlarını değerlendirmek için literatürde farklı karşılaştırma testleri önerilmiştir. Bu çalışmada literatürde önerilen güncel yöntemlerden parçalanmış log-rank testleri ve ağırlıklandırılmış Lin ve Wang testlerinin, log-rank ve ağırlıklandırılmış log-rank testleri ile karşılaştırılması amaçlanmaktadır.

Yöntem: Monte Carlo benzetim tekniği kullanılarak orantılı hazardlar varsayımının sağlandığı ve kesişen yaşam fonksiyonları ile bu varsayımın ihlal edildiği senaryolarda karşılaştırma testlerinin zayıf ve güçlü yönleri incelendi. Farklı örneklem büyüklüğü ve sansür oranlarında karşılaştırma testlerinin tip 1 hata oranları ve güçleri kıyaslandı. Karşılaştırmalar parçalanmış log-rank, Lin ve Wang testleri, log-rank ve ağırlıklandırılmış log-rank (Wilcoxon, Tarone-Ware, Peto, Modified Peto, Fleming-Harrington) testleri kullanılarak gerçekleştirildi. Yaşam sürelerinin türetimi farklı parametreli üstel ve Weibull dağılımları ile yapıldı. Sağdan sansürlü mekanizma oluşturmak için standart tekdüze dağılımdan faydalanıldı. Benzetim çalışmalarında tekrar sayısı 2000 kabul edildi. Grupların örneklem büyüklükleri birbirine eşit (n1=n2=50 ve 100) ve farklı (n1=40, n2=80; n1=50, n2=100) olarak belirlendi. Sansür oranları her iki grupta birbirine eşit ve %0, %20, %40 ve %60 olarak belirlendi. Anlamlılık düzeyi ise α=0.05 kabul edildi. Her bir karşılaştırma testinin gücü; anlamlı fark bulunan benzetim sonuçlarının tekrar sayısına oranı ile elde edildi.

Bulgular: Simülasyon çalışmaları sonucunda tüm testler %5'e yakın tip 1 hata oranları sunmaktadır. Orantılı hazardlar varsayımı altında, log-rank en başarılı karşılaştırma testidir. Kesişen yaşam fonksiyonlarına ait senaryolarda ise en başarılı sonuçlar parçalanmış log-rank testlerine aittir. Lin ve Wang testleri ve ağırlıklandırılmış log-rank testleri, kesişen yaşam fonksiyonlarına ait senaryolarda parçalanmış log-rank testlerinden oldukça düşük güç oranları sunmaktadır. Örneklem büyüklüğündeki artış testlerin performansını arttırmakta, ancak sansür oranındaki artış olumsuz bir etki yaratmaktadır. Yaşam fonksiyonlarının kesişim noktasının çalışmanın başlangıcından sonuna doğru ilerlemesi, testlerin performansını olumsuz yönde etkilemektedir. En büyük etki log-rank, en küçük etki ise parçalanmış log-rank testlerinde görülmektedir.

Sonuç: İki yaşam fonksiyonunun karşılaştırılmasında orantılı hazardlar varsayımı sağlandığında log-rank testinin, kesişen yaşam fonksiyonlarının karşılaştırılmasında ise parçalanmış log-rank testlerinin kullanılması önerilmektedir.

Anahtar Kelimeler: Ağırlıklandırılmış Lin ve Wang testleri, Kesişen yaşam fonksiyonları, Log-rank testi, Parçalanmış log-rank testleri, Yaşam analizi

S16

Comparison of the Statistical Tests for Two Crossing Survival Functions

Hulya OZEN[1], Cengiz BAL[1], Ertugrul COLAK[1]

[1]*Eskisehir Osmangazi University, Faculty of Medicine, Department of Biostatistics, Eskisehir*

Aim: Investigating the homogeneity of two survival functions is frequently used in survival analysis studies. The most preferred method in the literature is log-rank test, but this method provides misleading results in the assessment of crossing survival functions in which the assumption of proportional hazards is violated. Crossing survival functions can occur in many different structures. There is no single test that will evaluate all conditions alone. For this reason, different comparison tests have been proposed in the literature in order to evaluate crossing survival functions. The aim of this study is to compare partitioned log-rank tests and weighted Lin and Wang tests with log-rank and weighted log-rank tests.

Method: We investigated the weak and strong aspects of the comparison tests in the scenarios where the assumption of proportional hazards was obtained and violated with crossing survival functions with using the Monte Carlo simulation technique. Type 1 error rates and powers of the comparison tests were compared at different sample sizes and censorship rates. Comparisons were performed using partitioned log-rank tests, Lin and Wang tests, log-rank and weighted log-rank (Wilcoxon, Tarone-Ware, Peto, Modified Peto, Fleming-Harrington) tests. Survival times were derived with exponential and Weibull distributions with different parameters. Standard uniform distribution was used to create a right censored mechanism. In the simulation studies, the number of repetitions was accepted as 2000. The sample sizes of the groups were determined as equal (n1 = n2 = 50 and 100) and unequal (n1 = 40, n2 = 80; n1 = 50, n2 = 100). Censorship rates were determined to be equal in both groups as 0%, 20%, 40% and 60%. Significance level was accepted as $\alpha = 0.05$. The power of each comparison test was obtained by the ratio of the number of significant results to the number of repetitions.

Results: As a result of the simulation studies, all tests offer type 1 error rates close to 5%. Under the assumption of proportional hazards, log-rank is the most powerful comparison test. In the scenarios of crossing survival functions, the most successful results belong to partitioned log-rank tests. Lin and Wang tests and weighted log-rank tests offer considerably lower power ratios than partitioned log-rank tests in scenarios of crossing survival functions. The increase in the sample size increases the performance of the tests, but the increase in the censorship rate has an adverse effect. The progress of the crossing point of survival functions affected the performance of the tests adversely. The largest effect is seen in log-rank test while the smallest effect is seen in partitioned log-rank tests.

Conclusion: It is recommended to use log-rank test when the proportional hazards assumption is valid and the use of partitioned log-rank tests to compare crossing survival functions.

Keywords: Weighted Lin and Wang tests, Crossing survival functions, Log-rank test, Partitioned log-rank tests, Survival analysis

S17

Model Uyum Ölçütlerini Çok Değişkenli Kontamine Normal Dağılım ile Karşılaştıran Bir Simülasyon Çalışması

İlkay DOĞAN

Gaziantep Üniversitesi, Tıp Fakültesi, Biyoistatistik Anabilim Dalı, Gaziantep

Amaç: Bu çalışmada çok değişkenli kontamine normal dağılıma sahip, farklı örneklem büyüklerinde ve korelasyonlarda veri türetilerek yapısal eşitlik modellemesinde kullanılan model uyum ölçütlerinden en uygun olanını belirlemek amaçlanmıştır.

Yöntem: Yapısal eşitlik modellemesinde parametre tahmin yöntemleri olarak kullanılan ADF (Asymptotically Distribution Free), ML (Maximum Likelihood) ve GLS (Generalized Least Squares) yöntemleri kullanılmıştır. Ayrıca simülasyon çalışmasında 100, 150, 250, 500, 1000 ve 5000 örneklem büyüklerinde ve farklı korelasyonlarda çok değişkenli normal dağılmayan veri setleri türetilerek model uyum ölçütleri olan AGFI (Adjusted Goodness of Fit Index), GFI (Goodness of Fit Index), CFI (Comparative Fit Index), NFI (Normed Fit Index), NNFI (Non-Normed Fit Index), IFI (Incremental Fit Index), RMSEA (Root Mean Square Error of Approximatin) ve RMR (Root Mean Square Residuals) değerleri incelenmiştir. Simülasyon çalışması SAS programı yardımıyla PROC IML ve PROC CALIS prosedürleri kullanılarak 1000 tekrarlama ile gerçekleştirilmiştir.

Bulgular: Düşük korelasyona sahip veri setlerinde 100 ve 150 örneklem büyüklüklerinde tüm parametre tahmin yöntemleri için optimizasyon problemleri gözlenmiştir. Çok değişkenli normal dağılmayan veri setlerinde kullanılan ADF parametre tahmin yönteminde AGFI, GFI ve RMSEA değerlerinin örneklem büyüklüklerinden ve korelasyon değerlerinden etkilenmediği belirlenmiştir. Ancak CFI, NFI, NNFI ve RMR değerlerinin ise hem örneklem büyüklüğünden hem korelasyon değerlerinden etkilendiği tespit edilmiştir.

Sonuç: Çok değişkenli normal dağılmayan veri setlerinde ADF yöntemi kullanılırken örneklem büyüklüğü 250'den fazla olmalıdır. Bununla birlikte ADF yöntemi tercih edildiğinde AGFI, GFI ve RMSEA model uyum ölçütlerinin kullanılabileceği sonucuna varılmıştır.

Anahtar kelimeler: Doğrulayıcı faktör analizi, Yapısal eşitlik modellemesi, Model uyum ölçütleri

S17

A Simulation Study Comparing Model Fit Measures with Multivariate Contaminated Normal Distribution

İlkay DOĞAN

Gaziantep University, Faculty of Medicine, Department of Biostatistics, Gaziantep

Aim: In this study, it was aimed to determine the most appropriate model fit measure used in structural equation modeling by deriving data from a multivariate contaminated normal distribution with different sample sizes and correlations.

Method: ADF (Asymptotically Distribution Free), ML (Maximum Likelihood) and GLS (Generalized Least Squares) which are used as parameter estimation methods in structural equation modeling were used. Besides, data sets were obtained from a multivariate non-normal distribution with both 100, 150, 250, 500, 1000, and 5000 sample sizes and different correlations. AGFI (Goodness of Fit Index), GFI (Goodness of Fit Index), CFI Fit Index), NFI (Non-Normed Fit Index), IFI (Incremental Fit Index), RMSEA (Root Mean Square Error of Approximation) and RMR (Root Mean Square Residuals) values were examined as model fit measures in the simulation study. The simulation study was performed with the help of SAS program with 1000 replications using PROC IML and PROC CALIS procedures.

Results: Optimization problems were observed for all parameter estimation methods in 100 and 150 sample sizes in low correlation data sets. In ADF parameter estimation method used in multivariate non-normal distributed data sets; AGFI, GFI and RMSEA values were not affected by sample sizes and correlation values. However, it was found that CFI, NFI, NNFI, and RMR values were affected by both sample sizes and correlation values.

Conclusion: Sample size should be more than 250 when using ADF method in multivariate non-normal distributed data sets. Moreover, it is concluded that AGFI, GFI and RMSEA model fit measures can be used when the ADF method is preferred.

Keywords: Confirmatory factor analysis, Structural equation modeling, Model fit measures

S18

Transkriptom Veri Seti Üzerinde Derin Öğrenme Yöntemi İle Klasik Veri Madenciliği Yöntemlerinin Sınıflama Performanslarının Karşılaştırılması

Merve KAŞIKCI[1], Erdal COŞGUN[2], Erdem KARABULUT[1]

[1]Hacettepe Üniversitesi Tıp Fakültesi Biyoistatistik Anabilim Dalı, Ankara
[2]Ph.D-Data Scientist, Genomics Team, Microsoft AI & Research, Redmond, WA, USA

Amaç: Bu çalışmada, son yıllarda hastalık ve gen arasındaki ilişkileri incelemek için tercih edilen yöntemlerden biri olan RNA-dizileme verilerinin sınıflandırılmasında Derin Öğrenme ile klasik sınıflandırma yöntemlerinin performanslarının karşılaştırılması amaçlanmıştır. Çalışmada farklı özelliklere sahip iki veri seti kullanılmış, bu veri setleri üzeride çeşitli gen filtreleme yöntemleri uygulanmıştır. Bu sayede, sınıflama yöntemlerinin performansları farklı durumlar için de incelenmiştir.

Yöntem: Çalışmada akciğer ve böbrek kanseri olmak üzere iki farklı kanser türüne ait, RNA-dizileme yöntemi ile edilmiş gen ifade verileri kullanılmıştır. Akciğer kanseri veri seti iki sınıflı olup sınıf dağılımları dengelidir. Böbrek kanseri veri setinde ise üç sınıf bulunmaktadır ve gözlemlerin sınıflara dağılımı dengesizdir. Farklı olarak ifade edilmiş genler, Partek Flow yazılımı kullanılarak, DESeq2 ve GSA modelleri kullanılarak bulunmuştur. Farklı ifade edilmiş olarak bulunan gen setleri, üç farklı yanlış bulgu oranı (%1, %2 ve %5) kullanılarak filtrelenmiştir. Böylece sınıflama yöntemlerinin her veri setinde kullanılması ile sonuçların geçerliliği hakkında yorum yapabilmek amaçlanmıştır. Elde edilen veri setlerine normalizasyon ve dönüşüm uygulanmış, ardından %80 eğitim, %20 test seti olarak ayrılarak Rastgele Orman, Destek Vektör Makineleri, Yapay Sinir Ağı ve Derin Öğrenme yöntemleri ile sınıflandırılmıştır. Belleği 64 GB olan, 8 çekirdekli, 2 NVDIA GEFORCE GTX 1050 GPU'ya sahip özel bir bilgisayar Microsoft Azure bulut bilişim ortamında kullanılmış, bu sayede analizler oldukça hızlı bir şekilde sonuçlandırılmıştır. Derin Öğrenme modelleri için, çeşitli yapay sinir ağları algoritmalarının uygulanabildiği, Microsoft'un derin öğrenme için geliştirdiği açık kaynak kodlu bir araç olan Microsoft Cognitive Toolkit (CNTK) kullanılmıştır.

Bulgular: Akciğer kanseri veri seti DESeq2 yöntemi ile analiz edildiğinde, farklı filtreler için elde edilen gen sayıları yaklaşık olarak on beş bin iken GSA yönteminde yaklaşık olarak on bindir. DESeq2 yöntemi ile böbrek kanseri veri setinde ortalama dokuz bin gen seçilmiş iken GSA yöntemi ile altı bin gen seçilmiştir. Akciğer kanseri bilgilerini içeren veri setlerinde, genel olarak en yüksek performans değerleri Derin Öğrenme yöntemi ile elde edilmiştir. Böbrek kanseri veri setlerinde ise Derin Öğrenme ve Destek Vektör Makineleri'nin yüksek değerler aldığı görülmüştür. Genel olarak Derin Öğrenme ve Destek Vektör Makineleri, doğruluk, F ölçütü, Kappa gibi performans ölçüleri bakımından çalışmada en yüksek ya da ikinci en yüksek değerleri almıştır.

Sonuç: Daha fazla sayıda gen içeren ve dengeli sınıf dağılımı gösteren akciğer kanseri veri setlerinde genel olarak Derin Öğrenme yöntemi, klasik sınıflama yöntemlerine göre daha başarılı sonuçlar vermiştir ve kullanılması önerilmektedir.

Anahtar Kelimeler: RNA dizileme, Veri Madenciliği, Sınıflama Yöntemleri, Derin Öğrenme

Kaynaklar

1. A. Jabeen, N. Ahmad, K. Raza. Machine Learning-Based State-of-the-Art Methods for the Classification of RNA-Seq Data. Classification in BioApps, 2017, pp 133-172.
2. N.M. Elaraby, M. Elmogy, S.Barakat. Deep Learning: Effective Tool for Big Data Analytics. International Journal of Computer Science Engineering, 2016, pp 254-262.
3. Urda D, Montes-Torres J, Moreno F, Franco L, Jerez JM. Deep Learning to Analyze RNA-Seq Gene Expression Data. IWANN. 2017:50-59.

S18

Comparison of Classification Performance for Deep Learning Method and Classical Data Mining Methods on Transcriptome Data Set

Merve KAŞIKCI[1], Erdal COŞGUN[2], Erdem KARABULUT[1]

[1]*Hacettepe University Faculty of Medicine Department of Biostatistics, Ankara*
[2]*Ph.D-Data Scientist, Genomics Team, Microsoft AI & Research, Redmond, WA, USA*

Aim: In this study, it is aimed to compare the performance of Deep Learning and classical classification methods on the classification of RNA-sequencing data which is one of the preferred methods for examining the relationship between disease and gene. Two data sets with different characteristics were used and various gene filtering methods were applied to these data sets in this study. By this means, the performances of the classification methods have been examined for different situations.

Method: Gene expression data of two different cancer types were used. The lung cancer data set has two classes and class distributions are balanced. There are three classes in the renal cell cancer data set and the number of observations in the classes is uneven. To find differentially expressed genes, DESeq2 and GSA models were applied by using Partek Flow Genomic Analysis Software. Various false discovery rates (1%, 2% and %5) were utilized to filter gene sets. In this way, it was aimed to comment on the validity of the results by using classification methods in each data set. The obtained data sets were normalized, transformed, divided to train set (80%) and test set (20%). After these stages, they were classified with different classification methods. These methods are Random Forest, Support Vector Machines, Artificial Neural Networks, and Deep Learning. A special computer with 64 GB of memory, 8 cores, 2 NVIDIA GEFORCE GTX 1050 GPUs was used in the Microsoft Azure cloud computing environment. Thus, the analyses were completed very quickly. For the Deep Learning model, Microsoft Cognitive Toolkit (CNTK), an open-source tool developed by Microsoft for deep learning, is used, where various artificial neural network algorithms can be applied.

Results: The lung cancer data set was analyzed by the DESeq2 model and the number of genes obtained for the different filters was approximately fifteen thousand. When the same data set was analyzed by the GSA method, the number of genes obtained for the different filters was found approximately ten thousand. Similarly, approximately nine thousand genes were selected with DESeq2 model whereas nearly six thousand genes were found differentially expressed with the GSA model. In the lung cancer data sets, the highest performance values were obtained mostly by the Deep Learning method. In the kidney cancer data sets, Deep Learning and Support Vector Machines were found to have high values. In general, Deep Learning and Support Vector Machines obtained the highest or second highest values in terms of performance measures such as accuracy, F measure and Kappa.

Conclusion: In the lung cancer datasets that contain more genes and show a balanced class distribution, Deep Learning method generally has more successful results than classical classification methods and it is recommended to use.

Keywords: RNA sequencing, data mining, classification methods, Deep Learning

References

1. A. Jabeen, N. Ahmad, K. Raza. Machine Learning-Based State-of-the-Art Methods for the Classification of RNA-Seq Data. Classification in BioApps, 2017, pp 133-172.
2. N.M. Elaraby, M. Elmogy, S.Barakat. Deep Learning: Effective Tool for Big Data Analytics. International Journal of Computer Science Engineering, 2016, pp 254-262.
3. Urda D, Montes-Torres J, Moreno F, Franco L, Jerez JM. Deep Learning to Analyze RNA-Seq Gene Expression Data. IWANN. 2017:50-59.

S19

Faktör Sayısının Belirlenmesinde Velicer'in Map Testi, Horn'un Paralel Analizi ve Kaiser Yöntemlerinin Karşılaştırılması: Besin Gücü Ölçeği Uygulaması

Mehtap AKÇİL OK

Başkent Üniversitesi, Sağlık Bilimleri Fakültesi, Beslenme ve Diyetetik Bölümü, Ankara

Amaç: Yapı geçerliği çalışmalarında geliştirilen ya da başka bir kültüre uyarlanan ölçme aracının boyut sayısını belirlemek için alan yazında kullanılan çok sayıda istatistiksel ölçüt bulunmaktadır. Bunlardan çoğunun yapılan simülasyon çalışmaları sonucunda araştırmacılara yanlış yön gösterdiğine ilişkin bulgulara ulaşılmıştır. Bu araştırmada faktör (boyut) sayısına karar vermede kullanılan; Doğrulayıcı Faktör Analizi, Kaiser, Horn'un Paralel Analiz ve Velicer'in MAP testi yöntemleri ile elde edilen faktör sayılarının karşılaştırılması amaçlanmıştır.

Yöntem: Ölçek çalışmalarında boyut sayısına karar vermede en yaygın kullanılan istatistiksel yöntemler "Açıklayıcı ve Doğrulayıcı Faktör Analizleridir (AFA; DFA)". Özellikle temel bileşenler tahmin yöntemiyle uygulanan AFA'da, uygun faktör sayısına Kaiser'in özdeğerin 1'den büyük olması kuralına göre karar verilmektedir. Eğer kuramsal bilgi söz konusu ise DFA gibi yapısallığa dayanan eşitlik modelleri ile boyut sayısına karar verilir. Bu yöntemlerin avantaj ve dezavantajları son yıllarda oldukça tartışılmakta ve diğer yöntemlerle karşılaştırılarak karar verilmesi önerilmektedir. Bu yöntemlerden biri Velicer'in MAP testi (Velicer'in ortalama kısmi korelasyon testine dayanan özdeğer) diğeri ise Horn'un (1965) önerdiği ve daha sonra Brain O'Connor (2000) tarafından geliştirilen, gerçek veriyle aynı sayıda katılımcı ve madde sayısı içeren çeşitli simülasyonlarla üretilen korelasyon matrisi sonuçları ile gerçek veri seti korelasyon matrisi öz değer sonuçlarını karşılaştıran "Paralel analizi" yöntemidir. Bu yöntemlerin karşılaştırılmasında, lezzetli besinlerin yaygın olarak bulunduğu ortamlarda, metabolik ihtiyaç olmadan bireylerin, besin ve beslenme ile ilgili duygu ve düşüncelerini değerlendirme amacıyla kullanılan ve Lowe, MR ve arkadaşları (2009) tarafından geliştirilmiş ve Türkçe'ye uyarlanması tarafımızdan yapılan "Besin Gücü Ölçeği (BGÖ)" verisi kullanılmıştır. Orijinal BGÖ, başlangıçta 21 maddeden oluşmakta olup analizler sonucu 15 maddeden ve 3 faktörden oluşan bir ölçektir. Analizlerde IBM AMOS 21.0 ve SPSS 25.0 için geliştirilmiş Syntaxlardan yararlanılmıştır.

Bulgular: AFA sonucunda 21 maddeden 6'sının faktör yükleri 0.30'un altında olduğu için ölçekten çıkarılmıştır. Aynı maddeler orijinal ölçekte de çıkarılan maddelerdir. Ölçek Keiser'in öz değeri 1'den büyük olma kısıtı altında 3 faktöre ayrılmıştır. Elde edilen sonuçlara ve orijinal ölçeğe göre "iki seviyeli DFA" uygulanmıştır. DFA'den hesaplanan uyum istatistiklerinin orijinal ölçek ile aynı sonuçlar verdiği ve boyut sayısı üç bulunarak "konfirme" edilmiştir. MAP testi sonuçlarına göre ölçeğe ait öz değerler ile ortalama kısmi korelasyonlarının dördüncü dereceden kuvvetine ilişkin öz değerler incelendiğinde 3.basamakta en küçük öz değere (0.0244) ulaşılmıştır. Bu sonuç, ölçek maddelerinin 3 boyut altında toplandığını göstermektedir. Paralel analizinde gerçek veriden hesaplanan öz değerlerin, Monte Carlo Simülasyonu ile türetilen rasgele veriden hesaplanan öz değerlerden büyük olduğu basamak faktör sayısını vermektedir. Bu analize göre de 3.basamakta gerçek veriden hesaplanan öz değer 1.546 iken simülasyondan hesaplanan öz değer 1.280 olduğu için faktör sayısı 3 olarak elde edilmiştir.

Sonuç: BGÖ verisine tüm faktör sayısı belirleme yöntemleri uygulandığında orjinalinde olduğu gibi üç faktöre ayrılabileceği şeklinde sonuçlar bulunmuştur. Özellikle ölçek geliştirme çalışmalarında faktör sayısına doğru karar vermek için yaygın kullanılan yöntemlerin dışında MAP testi ve Paralel analizi yöntemlerinin kullanılması da yapısal geçerliğin gücünü artıracağı sonucuna varılmıştır.

Anahtar Kelimeler: Velicer'in MAP testi, Paralel Analiz, Faktör Analizi, Besin Gücü Ölçeği

S20

Gizli Ağaç Modelleri ve Gizli Sınıf Ağaç Analizi Yaklaşımı

Meriç YAVUZ ÇOLAK, Ersin ÖĞÜŞ

Başkent Üniversitesi, Tıp Fakültesi, Biyoistatistik Ana Bilim Dalı, Ankara

Çalışmanın amacı, Gizli Ağaç grafik modellerini kuramsal olarak tanıtmak ve temel özelliklerini vermektir. Gizli Ağaç Modelleri – Latent Tree Models"- (LTM) hakkında değerlendirme yapılacak ve Gizli Sınıf Ağaç Analizi – Latent Class Tree Analysis (LCTA) yönteminin sınıflandırma problemlerinde kullanımı ortaya konulacaktır. Sağlık alanında kullanımları, avantaj ve dezavantajları, mevcut diğer yöntemlerle karşılaştırmaları, çözümlemede kullanılacak olan yazılımlar, paket programlar incelenecektir.

Gizli Ağaç Analizi, gözlemlenen değişkenler arasındaki korelasyonları, yaprak düğümlerinin gözlenen değişkenleri, iç düğümlerin gizli değişkenleri temsil ettiği LTM olarak adlandırılan ağaç modelini kullanan grafiksel modelleme yöntemidir. Kesikli değişkenleri olan ağaç yapılı bir Bayes ağıdır. Sürekli değişkenler için olan farklı modeller de vardır. Rastgele değişkenler kümesi arasındaki korelasyonu gizli değişkenler ağacı kullanılarak modellemeye çalışır. LTM ile manifest değişkenler arasındaki karmaşık ilişki modellenebilir. Denetimsiz öğrenme için bir model sınıfı olarak önerilmiş olup; kümeleme, çok bölümlü kümeleme, ve yoğunluk tahmini gibi çeşitli problemlere uygulanmıştır. Bir popülasyondaki homojen alt grupları belirlemeye yönelik olarak sosyal bilimlerde ve tıpta özellikle son yıllarda yaygın olarak kullanılan bir yöntemdir. Konu belirleme, deep leraning olasılıksal modelleme gibi pek çok makine öğrenme yöntemlerine yeni ve faydalı bakış açısı sağlar. Filogenetik analiz, ağ tomografisi, bilgisayarlı görüntüleme, nedensel modelleme ve veri kümeleme alanlarında yaygın olarak kullanılmaktadır: LTM'ler, Gizli Markov modelleri, Brownian Motion Ağacı modeli, Ising modeli ve filogenetikte kullanılan birçok popüler model gibi bilinen diğer modelleri de kapsamaktadır. LTM'ler co-occurence modellemede, çok boyutlu kümeleme analizinde, belgelerdeki kelimelerin birlikte oluşma modellerini incelemede, kullanıcıların eş-tüketimini modellemede, hastalardaki semptomların birlikteliklerini modellemede de oldukça sık kullanılmaktadır. LCT analizi, veri yapısındaki alt grupları en iyi şekilde tanımlamak için gizli sınıf ağaçlara dayanan model tabanlı bir hiyerarşik kümeleme yöntemidir. Gizli Sınıf analizi, bir dizi kategorik değişkene dayanan anlamlı kümeleri bulmak için yaygın olarak kullanılmaktadır. Gizli Sınıf analizindeki en temel problem, çok fazla sayıda sınıf olması durumunda bu sınıfları yorumlamanın oldukça zor olmasıdır. Bu probleme LCT analizi alternatif yaklaşım olarak önerilmiştir. Az sayıda temel sınıf içeren bir çözümleme ile başlayıp daha sonraki aşamalarda alt sınıflara ayrılabilen bir yaklaşım izlenmektedir. Belirtilen sayıda sınıfa verilen tüm sınıfların aynı anda oluşturulduğu bir tahmin yöntemi kullanmak yerine LCT analizinde, sınıflar sırayla iki alt sınıfa ayrılarak, karşılıklı olarak bağlı sınıfların hiyerarşik bir yapısı elde edilir. Ortaya çıkan ağaç yapısı, sınıfların nasıl oluştuğu ve farklı sınıflardaki çözümlerin birbirleriyle nasıl bağlantılı olduğu konusunda net bir fikir vermektedir.Çalışmada kuramsal olarak tanıtılan ve literatürde uygulamaları ve popülaritesi son yıllarda giderek artan bu yöntemlerin, araştırmacılara ilham kaynağı olacağı ve yol göstereceği düşünülmektedir.

Anahtar Kelimeler: Gizli Ağaç Modelleri, Gizli Sınıf Ağaç Analizi, Sınıflandırma

S21

R x C Çapraz Tablolarda Kullanılan Bağımsızlık Testlerinin Karşılaştırılması

Merve BAŞOL[1], Ebru ÖZTÜRK[2], Sevilay KARAHAN[2]

[1] Bolu Abant İzzet Baysal Üniversitesi Tıp Fakültesi Biyoistatistik ve Tıbbi Bilişim Anabilim Dalı, Bolu
[2] Hacettepe Üniversitesi Tıp Fakültesi Biyoistatistik Anabilim Dalı, Ankara

Amaç: Klinik çalışmalarda sıklıkla elde edilen ve çapraz tablolar yardımı ile kolayca özetlenebilen kategorik değişkenlerin birbirinden bağımsız ya da ilişkili olup olmadığını test etmede kullanılan birçok yöntem vardır. Pearson ki kare testi bu yöntemlerden en sık tercih edilen ve en bilinen yöntem olsa da bazı durumlarda bu testin kullanımı önerilmemektedir [1]. Kategorik değişkenlerin bağımsızlığını/ ilişkisini farklı koşullar altında test etmek amacı ile çapraz tablolar için geliştirilen farklı istatistiksel yöntemler vardır. Bu çalışma kapsamında (i) Pearson ki kare testi, (ii) Olabilirlik Oran Testi, (iii) Freeman-Tukey testi, (iv) Cressie- Read testi ve (v) Fisher- Freeman – Halton Kesin testi kullanılacaktır. Bu çalışmanın amacı, çapraz tablolar için geliştirilen bu yöntemlerin olumlu ya da olumsuz yanlarını göstermek ve yöntemleri karşılaştırmaktır.

Yöntem: Yöntemlerin performansını değerlendirmek ve birbirleri ile karşılaştırmak için bir benzetim çalışması yapılmıştır. Benzetim çalışmasında farklı boyutlardaki çapraz tablolar için sonuçlar elde edilmiştir. Yöntemler, farklı etki genişliğinde (küçük, orta, büyük), farklı satır yüzdelerinde (dengeli, yarı dengeli, dengesiz) ve farklı örneklem büyüklüklerinde (küçük, orta, büyük) incelenmiştir. Her bir kombinasyon için 10000 veri seti elde edilmiştir. Sonuçlar, elde edilen güç düzeyi, birinci tip hata düzeyi ve güven aralığı genişlikleri üzerinden değerlendirilmiştir. Analizler, R 3.5.1 [2] programı ile "rTableICC" [3] ve "ggplot2"[4] paketleri kullanılarak gerçekleştirilmiştir.

Bulgular: Benzetim çalışmasından elde edilen sonuçlara göre gözlem sayısının artması ile her test benzer sonuçlar vermektedir. Gözlem sayısının az ve satır yüzdelerinin dengeli olduğu durumda en yüksek güç düzeyi olabilirlik oran ve Freeman Tukey test istatistiğinde elde edilmiştir. Ayrıca, benzetim çalışmasından elde edilen test istatistiklerinin %2.5 ve %97.5 değerleri incelendiğinde olabilirlik oran ve Freeman Tukey testlerinin daha geniş test istatistik aralığına sahip olduğu görülmüştür. Satır yüzdesi dengeli olduğu durumda tüm etki büyüklükleri için Freeman- Tukey testinin gücü Pearson ki kare, Cressie-Read ve Fisher- Freeman – Halton Kesin testlerinden daha yüksektir. Ancak, yeterli örneklem genişliğinde satır yüzdesi dengesiz olduğu durumda Cressie-Read ya da Fisher- Freeman – Halton Kesin testi önerilebilir. Olabilirlik oran testinin birinci tip hata düzeyi tüm senaryolarda diğer yöntemlere göre daha yüksektir. Bu testin yokluk hipotezini red etme eğiliminde olduğu söylenebilir [5].

Sonuç: R x C çapraz tablolarda değişkenler arasındaki farklılığı / ilişkiyi test etmede etki genişliği, örneklem tasarımı ve/ veya örneklem genişliğine dikkat edilmelidir.

Kaynaklar:

1. Agresti, A. (2002). Categorical Data Analysis. Hoboken: John Wiley & Sons.

2. R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

3. Demirhan, H. (2016). RTableICC: An R Package for Random Generation of 22K and RC Contingency Tables. The R Journal, 8 (1), 48. doi:10.32614/rj-2016-004.

4. Wickham, H., & Sievert, C. (2016). ggplot2: Elegant graphics for data analysis. S.l.: Springer

5. Rudas, T. (1986). A Monte Carlo comparison of the small sample behaviour of the Pearson, the likelihood ratio and the Cressie-Read statistics. Journal of Statistical Computation and Simulation. 24:107-120.

S22

Apriori Algoritması İle Yoğun Bakımlarda Yapılan Tetkiklerin Modellenmesi

Meryem GÖRAL YILDIZLI[1], Zeliha Nazan ALPARSLAN[1]

[1]Çukurova Üniversitesi Tıp Fakültesi, Biyoistatistik Anabilim Dalı, Adana

Amaç: Sağlık alanındaki teknolojilerin gelişimi, hizmet sunumunda kalitenin iyileştirilmesi ve sağlık kurumları arasındaki rekabetin çoğalmasıyla sağlık alanındaki harcamalar hızla artış göstermektedir. Sağlık kurumlarının mevcut finans kaynakları ile artan harcamaları karşılayıp varlıklarını sürdürebilmeleri için ciddi önlemler alması zorunlu hale gelmiştir. Çalışmanın amacı, hekimlerin tanı/tedavi hizmeti sunumlarına yardımcı olabilecek hastanın profiline uygun istem şablonları önermektir. Böylelikle hem hizmet kalitesi ve hızını arttırılıp, hem de sağlık harcamalarının önemli bileşeni olan laboratuvar tetkik harcamalarını azaltabilir.

Yöntem: Sağlık kurumları hastalık tanılarını belirlemede referans aldıkları laboratuvar test sonuçları hastane bilgi yönetim sistemlerinin yoğun ve önemli veri kaynağıdır. Veri madenciliği yöntemleri ile bu veriler analiz edilip, tıbbi kararlarında sağlık personeline ve kurumların efektif yönetilmesi için kurum yöneticilerine destek sağlayacak bilgiler sunulabilir. Bu amaçla 2018 yılında yoğun bakımda yatan hastalara yapılan tetkik, demografik ve tedavi sürecindeki verileri içeren bir veri seti oluşturularak, bu veri seti üzerine açık kaynak kodu olan WEKA programı ile veri madenciliği birliktelik kuralı algoritmalarından Apriori algoritması uygulanmıştır. Daha sonra Apriori algoritmasının öne çıkardığı birliktelik kurallarını içeren tetkikler belirlenip, bu tetkiklerin istendiği hastaların demografik ve tedavi sürecindeki tıbbi bilgilerinden oluşturulan verilerin ortak özellikleri ile gruplara göre istem modelleri şekillendirilmiştir.

Bulgular: Öncelikle 2018 yılında yoğun bakımda yatan hastalara yapılan 63686 kayıt içeren laboratuvar istemindeki tetkiklerin verileri düzenlenmiş. 'Apriori Algoritması' ile analiz edilip algoritma gereği güven (confidence) değerinin 0.95'den büyük/eşit, kuralın önemini gösteren kaldıraç (lift) değerlerinin büyük olduğu birlikte istenen tetkikler ayıklanmıştır. Daha sonra bu tetkiklerden birliktelik koşullarına uyan 17 kural belirlenip, kuralın sağlandığı 15786 kayıt içeren veri seti oluşturulmuştur. Son olarak da hastaların tıbbi ve demografik verileri kullanılarak bölüm bazlı istem şablonları oluşturulmuştur.

Sonuç: Hekimlere sağlık hizmetleri sunumlarında daha etkili ve doğru karar vermelerine yardımcı olmak için tasarlanan karar destek sistemleri, hastane bilgi yönetim sistemlerinin bir parçası olarak kullanılmaktadır. Çalışmada bu sistemlerin bir modülü olabilecek istem şablonları tanımlanmıştır. Yoğun bakım hastalarına yapılan laboratuvar testlerinin hastanın bölümü ve demografik özelliklerine göre belirlenen örüntülerin şablon olarak hekimlere sağlanması efektif sağlık hizmeti sunumuna destek olacaktır. Aynı zamanda belirlenen tetkik istem şablonlarının yoğun bakımların özellikleri açısından ilişkisi klinisyenler tarafından yorumlanarak akılcı istemlerin yapılması laboratuvar tetkiklerin maliyetini düşürerek tedavi masraflarını azaltabilir.

Anahtar Kelimeler : Apriori, yoğun bakım, laboratuvar tetkik

S22

Modeling of Laboratory Tests in Intensive Care Units with Apriori Algorithm

Meryem GÖRAL YILDIZLI[1], Zeliha Nazan ALPARSLAN[1]

[1] *Çukurova University Faculty of Medicine Department of Biostatistics, Adana*

Aim: Expenditures in the field of health are increasing rapidly with the development of technologies, improving the quality of health care provision and increasing competition among health institutions. Health institutions have become obliged to take serious measures in order to meet the increasing expenditures with the available financial resources and to survive. The aim of the study was to propose pattern of ordering lab test for patiens in intensive care units appropriate to the patient's profile that could help physicians to provide diagnostic / treatment services. In this way, both the quality and speed of service can be increased, as well as the important component of health expenditures can reduce laboratory test expenditures.

Method: The laboratory test results that health institutions refer to in the diagnosis of disease are intensive and important data sources of hospital information management systems. Data mining methods can be used to analyze these data and provide support to physicians in their medical decisions and the effective management of managers' institutions. For this purpose, in 2018, a data set containing the data in the lab tests, demographics and treatment process was created for the patients hospitalized in the intensive care unit and the Apriori algorithm which is an open source code WEKA program and data mining association rule algorithms were applied. The clinicial ordering models were then determined according to the common characteristics of the demographic and medical information of the patients to whom the tests involving important association rules were performed according to the result of the Apriori algorithm.

Results: First of all, the data of the laboratory tests including 63686 records made to the inpatients in intensive care unit in 2018 were prepared. As a result of the Apriori algorithm, rules with confidence values greater than or equal to 0.95 and big leverage value indicating the importance of the rule are selected. Then, A set of data containing 15786 records was created from the lab tests that provided 17 rules. Finally, department-based ordering patterns were formed using the medical and demographic data of the patients.

Conclusion : Decision support systems, which are designed to help physicians to make more effective and correct decisions in their health care services, are used as part of hospital information management systems. In the study, the ordering pattern which can be a module of these systems are defined. Ordering pattern created from laboratory tests performed in intensive care patients according to the patient's department and demographic characteristics can provide support to physicians to provide effective health care services. At the same time, rational laboratory test ordering prepared by clinicians using designated ordering pattern can reduce treatment costs by reducing laboratory costs.

Keywords : Apriori, intensive care units, laboratory tests

S23

Comparative Analysis of Mna Scores With Linear Mixed Effect Models And Linear Regression Models in Patients with Peg Feed

Neslihan GÖKMEN[1], Arzu BAYGÜL[2], Aydın ERAR[3]

[1] Istanbul Technical University Basic Sciences, Istanbul
[2] Koç University Biostatistics, Istanbul
[3] Mimar Sinan Fine Arts University, Statistics, Istanbul

Mixed effects models are powerful tools for longitudinal data analysis. It incorporates the time points of the measurements as a continuous variable and characterized by the ability to model the mean response level as a combination of population characteristics and individual-specific effects shared by all individuals. Therefore, it is frequently used in health studies.

Aim: The aim of this study is to present the performance evaluation of linear mixed effect models and classical linear regression models used in longitudinal data analysis based on the health data. For this purpose, it was aimed to determine the parameters affecting the MNA (Mini Nutritional Assessment) score in PEG feed patients.

Method: The MNA scores,which were measured at baseline, 3rd months, 6th months and 12th months after PEG implantation from 53 patients included in the study .After PEG implantation, survival time was recorded on a daily basis and mortality was included in the analysis as a group variable. Linear mixed effect models were estimated by using the restricted maximum likelihood estimation method, which is the best linear unbiased estimator. Least squares estimation was used in linear regression modeling. Linear mixed effect models with fixed and random effects were modeled as univariate and multivariate. In the selection of variables, the significance of the model was evaluated by ANOVA test. Information criteria (AIC,BIC-Akaike, Bayesian information criterion) were used to select the appropriate model.

Results: Univariate model with random effects performance of four different models was found to be higher than the others. While the mean square errors and confidence interval width were lower, the model significance was high. The random effects constant varies by measurement and group, and the inclusion of individual effects in the model also strengthens the model performance. MNA score is the most important factor affecting mortality of patients are said to have shown a decrease of 12.2 units compared to those living for MNA levels.

Conclusion: According to the findings and application results in the literature; the performance of the linear mixed effect model was found to be higher than the multiple regression model. Since it is not possible to provide multiple linear regression assumptions in longitudinal and unstable data, it is preferable to use linear mixed effect models.

Keywords: linear mixed effects models, fixed effect, random effect, multiple linear regression models

References

1. Wu, H., Zhang, J. (2006). Nonparametric Regression Methods for Longitudinal Data Analysis, John Wiley & Sons, Inc.

2. Armitage, P., Colton, T. (1998). Encyclopedia of Biostatistics, Wiley, New York , 4: 2657-2662.

3. Patterson, H.D., Thompson, R. (1971). Recovery of inter-block information when block sizes are unequal, Biometrica, 58(3): 545-554.

4. Doğanay, B. (2007). "Uzunlamasına Çalışmaların Analizinde Karma Etki Modelleri", Ankara Üniversitesi Sağlik Bilimleri Enstitüsü Biyoistatistik Anabilim Dali Yüksek Lisans Tezi, Ankara.

5. Koç, T.,Cengiz, M.A. (2012). Genelleştirilmiş Lineer Karma Modellerde Tahmin Yöntemlerinin Uygulamalı Karşılaştırılması. Karaelmas Fen ve Mühendislik Dergisi, 2 (2): 47-52.

6. Vittinghoff, E., Shiboski, S.C., Glidden, D.V., Mcculloch, C.E. (2005), Regression Methods in Biostatistics: Linear, Logistics, Survival and Repeated Measrures Models, Springer Science&Business Media Inc.

7. Hedeker D., Gibbons R.D. (2006). Longitudinal Data Analysis, John Wiley & Sons, Inc., Hoboken, New Jersey.

8. Fitzmaurice, G., Laird, N., Ware, J. (2004). Applied Longitudinal Analysis, Wiley Series in Probabilty and Statistics.

9. Kuznetsova, A., Brockhoff, P.B., Christensen, R.H.B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models, Journal of Statistical Software, 82 (13): 1-26.

10. Morrel, C. H. (1998). Likelihood ratio testing of variance components in the linear mixedeffects model using restricted maximum likelihood. Biometrics 54: 1560–1568.

11. Langkamp-Henken, B. (2006). Usefulness of the MNA® in the Long-Term And Acute-Care Settings Within the United States, The Journal of Nutrition, Health & Aging. 10 (6):502-509.

12. Chan, M., Lim, Y.P., Ernest, A., Tan, T.L. (2010). Nutritional assessment in an Asian nursing home and its association with mortality. The Journal of Nutrition, Health & Aging. 14 (1): 23-28.

S24

Dairesel(Circular) Verinin Tanıtılması ve Gerçek Veri Seti Üzerine Uygulama

Hülya BİNOKAY[1], Yaşar SERTDEMİR[1]

[1] Çukurova Üniversitesi Tıp Fakültesi Biyoistatistik Anabilim Dalı, Adana

Amaç: Dairesel veri, belirli bir başlangıç noktası ve dönüş yönü (saat yönü veya tersi) olan birim çemberin çevresi üzerinde nokta olarak gösterilebilen gözlemler topluluğuna denir. Olaylar periyodik olarak gerçekleştiğinde (haftanın günlerinde ya da günün saatlerinde) gözlenen olayların oluşturduğu verilerdir. Periyodik yapısından dolayı doğrusal verilerden farklıdır ayrıca dairesel veriler aralıklı ölçek içerisinde değerlendirilmektedir. Bu veriler yönlü ya da yönsel veriler, açısal veriler olarak da bilinmektedir.

Sağlık alanında hastalıkların ortaya çıkma zamanları ve ölümlerin zamana göre dağılımı dairesel veriyi oluşturmaktadır. Örneğin, gün içinde acil servisine başvuru saatleri, hastaların acil servise ulaştırılma süreleri, epilepsi hastasının nöbet geçirme zamanları ve yıl içerisinde acile alerjik astım atağı ile başvuran hastaların başvuru zamanları dairesel veri olarak değerlendirilebilir. Bu veriler doğrusal yapıda olan verilerden farklı tanımlayıcı istatistik formüllere ve dağılımlara sahiptir.

Bu çalışmada, gerçek bir veri seti üzerinden, dairesel verilerin tanımlayıcı istatistiklerinin hesaplanması, önemlilik testlerinin incelenmesi ve görselleştirilmesi amaçlanmıştır.

Yöntem: Dairesel verilerin analizinde kullanılan hesaplamaların gösterilmesi amacıyla bir ilimizdeki özel ve devlet hastanelerine 2016 yılında koah şikayeti ile başvuran hastaların başvuru tarihleri, cinsiyetleri ve kükürt değerlerinin bulunduğu veri seti kullanılmıştır. Cinsiyet gruplarına göre Başvuruların ve havadaki kükürt değerinin yüksek olduğu günlere karşılık gelen açıların önemlilik testleri ve tanımlayıcı istatistikleri hesaplanmıştır. Düzgün dağılıma uygunluğu belirlemek için Rayleigh, testi ve gruplar arası farklar için Watson Wheeler veya Williams testleri kullanılmıştır. Analizler için R 3.3.3 paket programı ve görselleştirmeler için NCSS ve SPSS paket programları kullanılmıştır.

Bulgular ve Sonuç:

Başvuruların yüksek olduğu günler göz önüne alındığında koah hastası olan erkek ve kadınların ortalama başvuru zamanı kış aylarında olmaktadır. Cinsiyete göre bakıldığında başvuru tarihleri açısından fark gözlenmiştir.

Havadaki kükürt değerinin yüksek olduğu durum incelendiğinde koah hastası olan erkek ve kadınların ortalama başvuru zamanı kış aylarında olmaktadır. Cinsiyete göre bakıldığında başvuru tarihleri açısından fark gözlenmiştir.

Kükürdün yüksek olduğu grupta ortalama başvuru zamanı kış aylarında iken düşük olduğu grupta nisan aylarındadır. Düşük ve yüksek kükürt grupları için başvuru tarihleri bakımından farklılık gözlenmiştir. Koah atakları daha çok kış aylarında oluşmaktadır ve havadaki kükürt değerlerinin yüksek olduğu günlerde ataklar artmaktadır. Erkeklerde kadınlara göre daha çok görülmektedir. Sonuç olarak mevsimsel olarak meydana gelen hastalıkların dairesel veriler için uygulanabilirliği gözlenmiştir.

	Başvuruların yüksek olduğu günler		Kükürt gruplarına göre		Havadaki Kükürt değerinin yüksek olduğu durum	
Tanımlayıcı İstatistikler	Erkek (n=11079)	Kadın (n=4397)	So2≥20 (n=1410)	so2<20 (n=14993)	Erkek (n=1025)	Kadın (n=385)
Ortalama yön ($\bar{\theta}$)	35.65^0	55.33^0	1.41^0	93.01^0	1.30^0	1.69^0
Ortalama vektör uzunluk ($\bar{R}$)	0.10	0.11	0.87	0.06	0.87	0.86
	p= 0.039		p<0.001		p<0.001	
Dairesel varyans (V)	0.90	0.89	0.13	0.94	0.13	0.14
Standart sapma (ν)	2.12	2.10	0.54	2.36	0.54	0.55
Von Mises yoğunlaşma parametresi (κ)	0.21	0.22	4.01	0.12	4.05	3.93

Anahtar Kelimeler: Dairesel veri, Von Mises, Doğrusal veri.

S24

Introduction of Circular data and Application on Data Set

Hülya BİNOKAY[1], Yaşar SERTDEMİR[1]

[1] *Cukurova University Faculty of Medicine Department of Biostatistics, Adana*

Introduction: Circular data is a collection of observations that can be represented as dots on the perimeter of the unit circle with a given starting point and direction of rotation (clockwise or counterclockwise). These are the data generated by events observed when events occur periodically (days of the week or hours of the day). It is different from linear data due to its periodic structure and also circular data is evaluated on an interval scale. These data are also known as directional or angular data. In the field of health, the occurrence of diseases and the distribution of deaths over time constitute circular data. For example, the hours of admission to the emergency department during the day, the time of delivery to the emergency department, the time of seizure of epilepsy patients and the time of admission of patients with an allergic asthma attack to the emergency department during the year can be considered as circular data. Descriptive statistical formulas and distributions are different for Circular and Linear data.

Aim: In this study, we aimed to calculate the descriptive statistics of circular data, to examine significance tests and visualize data using a real data set.

Method: In order to show the calculations used in the analysis of circular data, a dataset containing the administration dates and gender of patients with COPD hospitalized at private or public hospitals in a province in 2016. Additional information of Air pollution (SO_2) for each day was registered. Significance tests and descriptive statistics of the angles corresponding to the days of hospitalization were calculated. To test the uniform distribution assumption we applied the Rayleigh test and to test for differences among groups the Watson Wheeler or Williams tests were applied.

R 3.3.3 program was used for analysis and NCSS program was used for visualizations.

Results and Conclusion: Considering the days with high(>25person per day) admissions, the average hospitalization time for men and women with COPD was January. There was a difference in terms of hospitalization dates according to gender.

The analysis of hospitalizations for days with high air pollution (SO_2>20) showed that the average admission time for men and women with COPD was winter. We observed a significant difference in terms of mean direction according to gender.In our analysis we also observed significant differences for mean direction between days with SO_2≥20 and SO_2<20, the mean direction was January and April respectively.COPD attacks occur mostly in the winter and attacks increase days with SO_2≥20. Consequently, the applicability of seasonal diseases for circular data was observed.

	High administration days (>25 person per day)				High SO_2 days ($SO_2 \geq 20$)	
	Male (n=11079)	Female (n=4397)	$SO_2 \geq 20$ (n=1410)	$SO_2 < 20$ (n=14993)	Male (n=1025)	Female (n=385)
Mean Direction ($\bar{\theta}$)	35.65^0	55.33^0	1.41^0	93.01^0	1.30^0	1.69^0
Mean Resultant length ($\bar{R}$)	0.10	0.11	0.87	0.06	0.87	0.86
	p= 0.039		p<0.001		p<0.001	
Circular Variance (V)	0.90	0.89	0.13	0.94	0.13	0.14
Circular Standard Deviation (ν)	2.12	2.10	0.54	2.36	0.54	0.55
Concentration Parameter (κ)	0.21	0.22	4.01	0.12	4.05	3.93

Keywords: Circular Data, Von Mises, Lineer Data

S25

İki Elemanlı İki Değişkenli Uzunlamasına Veri için Probit Bağlantılı Marjinalleştirilmiş Özgün bir Model Üzerine bir Simülasyon Çalışması

Gül INAN[1], Ozlem ILK[2]

[1] İstanbul Teknik Universitesi, Matematik Bölümü, İstanbul
[2] Orta Doğu Teknik Universitesi, İstatistik Bölümü, Ankara

Amaç: Bu çalışmada, iki elemanlı iki değişkenli uzunlamasına verinin analizi için, probit bağlantılı marjinalleştirilmiş rassal etkili bir model önerilmiş ve başarısı değerlendirilmiştir.

Yöntemler: Önerilen model iki seviyeden oluşmaktadır: İlk seviye bir lojistik regresyon modeli aracılığıyla ortalama cevap değişkenlerini açıklayıcı değişkenlerle ilişkilendirmektedir; ve ikinci seviye bir probit regresyon modeli içerisine rassal kesişim terimleri dahil etmektedir. Rassal etkiler kovaryans matrisi değiştirilmiş Cholesky ayrıştırma yöntemi ile bağımlılık ve varyans parçalarına ayrıştırılmıştır. Daha sonra, bağımlılık ve varyans parametrelerine dair daha iyi açıklamalar getirmesi ve rassal etkiler kovaryans matrisinde hesaplanması gereken parametre sayısının azalmasını sağlamasından dolayı, ortaya çıkan parametrelerin kısıtsız halleri düşük boyutta açıklayıcı değişkenlerle modellenmiştir. Önerilen modeldeki parametrelerin, en çok olabilirlik yöntemi parametre ve standart hata tahmin edicilerini hesaplamak için veri klonlaması algoritması kullanılmıştır. Önerilen modelin geçerliliği farklı senaryolar altında R'da bir Monte Carlo benzetim çalışması aracılığıyla incelenmiştir.

Bulgular: Genellikle küçük yanlılık ve orta düzeyde hata kareler ortalaması gözlenmiştir. Örneğin, iki açıklayıcı değişken varken, 150 şahısın 4 farklı zamanda tekrarlı ölçümlerinin alındığı alt senaryo için, ortalama parametrelerinin yanlılığı 0,01 ila 0,03 arasında değişmektedir; ve hata kareler ortalaması 0,007 ila 0,029 arasındadır. Aynı alt senaryo için, bağımlılık ve varyans parametrelerine dair yanlılık -0,04 ila -0,01 arasında değişmektedir; ve hata kareler ortalaması 0,014 ila 0,314 arasındadır.

Sonuç: Simülasyon çalışmamız modelimizin, özellikle orta uzunluktaki uzunlamasına veri için, iyi performans verdiğini göstermektedir.

Anahtar Kelimeler: Birden çok korelasyonlu rassal etkiler, Çok seviyeli modeller, Kovaryans matrisi ayrıştırması

S25

A Simulation Study on a Novel Marginalized Model with Probit Link for Bivariate Longitudinal Binary Data

Gül INAN[1], Ozlem ILK[2]

[1] *Istanbul Technical University, Department of Mathematics, Istanbul*
[2] *Middle East Technical University, Department of Statistics, Ankara*

Aim: In this study, we propose and evaluate a marginalized random effects model with probit link function for the analysis of bivariate longitudinal data with binary responses.

Methods: The proposed model consists of two levels: The first level links the marginal mean of responses with covariates through a logistic regression model; and the second level includes random intercepts within a probit regression model. The covariance matrix of the random effects is further decomposed into its dependence and variance components via modified Cholesky decomposition method. The unconstrained version of resulting parameters is modelled in terms of covariates with low-dimensional regression parameters. This provides better explanations and a reduction in the number of parameters to be estimated in random effects covariance matrix. Data cloning algorithm is used to compute the maximum likelihood estimates and standard errors of the parameters. The validity of the proposed model is assessed through a Monte Carlo simulation study under different scenarios in R.

Results: Generally, quite small biases and moderate MSEs are observed. For instance, for the sub-scenario with two covariates observed on 150 subjects repeatedly at 4 different time points, the bias values for the mean parameters range from 0.01 to 0.03; and the MSEs range from 0.007 to 0.029. For the same sub-scenario, the bias values for the dependence and variance parameters range between -0.04 and -0.01; and the MSEs vary between 0.014 and 0.314.

Conclusion: Our simulation study show that the model performs well in longitudinal data, especially with medium length series.

Keywords: Covariance matrix decomposition, Multilevel models, Multiple correlated random effects

S26

Mikrodizi Kanser Verilerinin Öznitelik Seçimi ve Sınıflandırılması

Özlem ARIK[1], Erdem KARABULUT[2]

[1]Kütahya Sağlık Bilimleri Üniversitesi, Tıp Fakültesi, Biyoistatistik Anabilim Dalı, KÜTAHYA
[2]Hacettepe Üniversitesi, Tıp Fakültesi, Biyoistatistik Anabilim Dalı, ANKARA

Amaç: Biyoinformatik; biyoloji, bilgisayar, matematik, istatistik ve genetik alanlarını kapsayan disiplinler arası bir bilim dalıdır ve tıpta önemli yeri vardır. Biyoinformatiğin en önemli araştırma alanlarından birisi gen analizidir ve bu alanda DNA mikrodizi teknolojisi çok yönlü bir teknolojidir. Mikrodizi verilerinin sahip olduğu yüksek p (öznitelik sayısı) ve düşük n (örnek sayısı) parametreleri, makina öğrenme algoritmalarının başarı oranını olumsuz yönde etkilemektedir. Bu sebeple mikrodizi veri setlerinin modellenmesinde öznitelik seçimi önemli bir işlem adımıdır. Genetik verilerin incelenmesinde verinin çok büyük boyutlarda olması nedeniyle çeşitli veri madenciliği yöntemleri de geliştirilmiştir. Veri madenciliğinde verilerin içerdiği ortak özellikleri kullanarak söz konusu bu verileri ayrıştırmak ve yeni eklenecek verilerin hangi sınıfa dahil olacağına karar vermek için kullanılan yöntemlerden birisi sınıflamadır. Bu çalışmadaki temel amaç; NCBI-GEO veri tabanından elde edilen meme kanseri, lösemi ve lenfoma veri setleri üzerinde öznitelik seçimi yaparak sınıflama modellerini elde etmek ve performanslarını ölçmektir.

Yöntem: Bu çalışmada; meme kanseri, lösemi ve lenfoma türlerine ait mikrodizi gen ifade verileri kullanıldı. Veri ile ilgili ön işlemeden sonra R programında *genefilter* paketi içerisinde yer alan varFilter() fonksiyonu yardımıyla öznitelik seçimi yapıldı. Veri madenciliği sınıflama yöntemlerinden karar ağaçları(C4.5), random forest ve bayes sınıflandırıcı: naive bayes kullanıldı ve doğruluk ve ROC eğrisi altında kalan alan (AUC) ölçüler aracılığıyla yöntemlerin performansları elde edildi. Sınıflama modellerinde yer alan önemli öznitelikler belirlendi.

Bulgular: Meme kanseri, lösemi ve lenfoma veri setleri üzerinde öznitelik seçimi yapıldıktan sonra elde edilen sınıflama modellerine ait bulgular;

Tablo 1: Model Performans Ölçüleri

Veri Seti	Sınıflama Modeli	Doğruluk	AUC
Meme	C4.5	0.82	0.88
	Random Forest	0.73	0.71
	Naive Bayes	0.64	0.65
Lösemi	C4.5	1	1
	Random Forest	0.94	0.90
	Naive Bayes	0.88	0.80
Lenfoma	C4.5	1	1
	Random Forest	0.75	0.75
	Naive Bayes	0.75	0.75

Sonuç: R programı aracılığıyla yapılan çalışmada, Tablo 1`de verilen değerlere göre Meme, lösemi ve lenfoma veri setleri için en iyi sınıflama yöntemi C4.5`tir. Meme ve lösemi veri setleri için random forest yöntemi naive-bayes yöntemine göre daha iyi performansa sahiptir. Veri setleri üzerinde öznitelik seçimi yapıldıktan sonra kullanılan sınıflama modellerinin oluşturulmasında önemli yer tutan özniteliklerin önem dereceleri Şekil 1`deki grafikler üzerinde gösterilmektedir. C4.5,random forest ve naive bayes sınıflama modellerinde yer alan önemli öznitelikler, veri setlerinin her biri için farklıdır.

Anahtar sözcükler: Veri Madenciliği, Biyoinformatik, Öznitelik Seçimi, Sınıflama.

Kaynaklar

Korkem, E. (2013), "Mikroarray Gen Ekspresyon Veri Setlerinde Random Forest Ve Naıve Bayes Sınıflama Yöntemleri Yaklaşımı", Yüksek Lisans Tezi, Hacettepe Üniversitesi, Ankara.

Özkan, Y. ve Erol, Ç.S.,(2019), "Kanser Biyoinformatiğinde Yapay Zeka", Papatya Yayıncılık Eğitim, İstanbul.

Zararsız, G., Öztürk A., Elmalı, F., Köprü, M., Coşgun, E., Karabulut, E. (2011), "Gen Ekspresyon Verilerinde En iyi Gen Kümelerinin Belirlenmesi Üzerine Bir Uygulama", XIII. Ulusal Biyoistatistik Kongresi. Ankara, Türkiye.

https://www.ncbi.nlm.nih.gov/gds

https://www.bioconductor.org/

S26

Attribute Selection and Classification of Microarray Cancer Data

Özlem ARIK[1], Erdem KARABULUT[2]

[1]Kutahya Health Science University, Faculty of Medicine, Department of Biostatistics, KÜTAHYA
[2]Hacettepe University, Faculty of Medicine, Department of Biostatistics, ANKARA

Aim: Bioinformatics is an interdisciplinary field of science covering biology, computers, mathematics, statistics and genetics and has an important place in medicine. One of the most important research areas of bioinformatics is gene analysis and DNA microarray technology is a versatile technology in this field. The high p (number of attributes) and low n (number of samples) parameters of the microarray data adversely affect the success rate of the machine learning algorithms. For this reason, feature selection is an important process in the modeling of microarray data sets. In the investigation of genetic data, various data mining methods have been developed due to the large size of the data. Classification is one of the methods used in data mining to parse data and determine the class of new data, using the common features contained in the data. The aim of this study is to obtain classification models and measure their performance by selecting attributes on breast cancer, leukemia and lymphoma datasets obtained from NCBI-GEO database.

Method: In this study; microarray gene expression data of breast cancer, leukemia and lymphoma species were used. After preprocessing of the data, feature selection was made with the help of varFilter () function in the *genefilter* package in R program. Decision trees (C4.5), random forest and naive bayes were used as data mining classification methods, and the performances of the methods were obtained by measuring accuracy and area under the ROC curve (AUC). Important attributes of classification models were determined.

Results: After feature selection on breast cancer, leukemia and lymphoma data sets, the obtained results for classification models are;

Table 1: Performance Criteria

Datasets	Classification Model	Accuracy	AUC
Breast	C4.5	0.82	0.88
	Random Forest	0.73	0.71
	Naive Bayes	0.64	0.65
Leukemia	C4.5	1	1
	Random Forest	0.94	0.90
	Naive Bayes	0.88	0.80
Lymphoma	C4.5	1	1
	Random Forest	0.75	0.75
	Naive Bayes	0.75	0.75

Conclusion: According to the values given in Table 1, the best classification method for Breast, Leukemia and Lymphoma data sets is C4.5. Random forest method has better performance for breast and

leukemia data sets than naive-bayes method. After the selection of attributes on the data sets, the important attributes in the classification models used are different for each of the data sets.

Keywords: Data Mining, Bioinformatic, Feature Selection, Classification.

References

Korkem, E. (2013), "Mikroarray Gen Ekspresyon Veri Setlerinde Random Forest Ve Naıve Bayes Sınıflama Yöntemleri Yaklaşımı", Yüksek Lisans Tezi, Hacettepe Üniversitesi, Ankara.

Özkan, Y. ve Erol, Ç.S.,(2019), "Kanser Biyoinformatiğinde Yapay Zeka", Papatya Yayıncılık Eğitim, İstanbul.

Zararsız, G., Öztürk A., Elmalı, F., Köprü, M., Coşgun, E., Karabulut, E. (2011), "Gen Ekspresyon Verilerinde En iyi Gen Kümelerinin Belirlenmesi Üzerine Bir Uygulama", XIII. Ulusal Biyoistatistik Kongresi. Ankara, Türkiye.

https://www.ncbi.nlm.nih.gov/gds

https://www.bioconductor.org/

S27

Glasgow Koma Skoru Güvenirliği: Cronbach Alpha için Meta Analizi

Demet ARI YILMAZ[1], Pınar GÜNEL KARADENİZ[2], Vildan SÜMBÜLOĞLU[2]

[1]SANKO Üniversitesi, Tıp Fakültesi, Acil Tıp Anabilim Dalı, Gaziantep
[2]SANKO Üniversitesi, Tıp Fakültesi, Biyoistatistik Anabilim Dalı, Gaziantep

Amaç: Bu çalışmanın amacı acil servis ve yoğun bakım ünitelerinde sık kullanılan bir ölçek olan Glasgow Koma Skoru'nun güvenirliği için bir meta analiz çalışması yürütmektir.

Yöntem: Reith ve arkadaşlarının (2016) "The reliability of the Glasgow Coma Scale: a systematic review" isimli makalesinde yer alan ve Cronbach alpha katsayıları raporlanmış olan 10 çalışma ile buna ek olarak 2016 yılı sonrası için MEDLINE, EMBASE, CINAHL ve Google Scholar veri tabanlarında taranan 6 makale meta analize dâhil edilmiştir. Birden fazla değerlendiricinin yer aldığı çalışmalar için iki defa raporlanan Cronbach alpha katsayılarının ikisi de dikkate alınmıştır. Meta analiz, R programında "metafor" paketi kullanılarak yürütülmüştür. Heterojenlik I^2 istatistiği ve Q testi ile değerlendirilmiş ve rasgele etkiler meta analizi uygulanarak genel Cronbach alpha değeri hesaplanmıştır. Yayın yanlılığı; sıra korelasyon testi "ranktest", regresyon testi "regtest" ve huni (funnel) grafiği ile değerlendirilmiştir.

Bulgular: Toplam 16 çalışmadan elde edilen toplam 19 Cronbach alpha değerleri 0.69 ile 0.99 arasında değişmekteydi. Yürütülen meta analiz sonucunda I^2 istatistiği %95 ve Q testi için $p<0.001$ olarak elde edilmiştir. Bulgular heterojenliğin varlığını gösterdiğinden rasgele etkiler meta analizi uygulanarak genel Cronbach alpha değeri 0.87 (%95 Güven Aralığı: 0.84-0.90) olarak hesaplanmıştır. Bu değerlere ait forest grafiği elde edilmiştir. Huni grafiği çizdirilerek yayın yanlılığı görsel olarak değerlendirildiğinde çalışmaların simetrik olmadığı gözlenmiştir. Sıra korelasyon ve regresyon testleriyle huni grafiğinin asimetrisi araştırırlmış ve her ikisinin sonucuna göre asimetrik bulunmamıştır (sırasıyla; p=0.298, p=0.057). Bu bulgular doğrultusunda yayın yanlılığı olmadığı sonucuna varılmıştır.

Sonuç: Çalışma sonucunda elde edilen bulgular doğrultusunda Glasgow Koma Skoru için genel Cronbach alpha değeri, bir ölçeğin güvenirliği için kabul edilebilir sınırlar içinde bulunmuştur.

Anahtar Kelimeler: Cronbach alpha, güvenirlik için meta analiz, Glasgow koma skoru

S27

Reliability of Glasgow Coma Score: Meta-Analysis for Cronbach Alpha

Demet ARI YILMAZ[1], Pınar GÜNEL KARADENİZ[2], Vildan SÜMBÜLOĞLU[2]

[1]SANKO University, Faculty of Medicine, Department of Emergency Medicine, Gaziantep
[2]SANKO University, Faculty of Medicine, Department of Biostatistics, Gaziantep

Aim: The aim of this study was to conduct a meta-analysis for the reliability of the Glasgow Coma Score, a commonly used scale in emergency and intensive care units.

Method: Ten studies which take part in Reith et al. (2016) study entitled "The reliability of the Glasgow Coma Scale: a systematic review" and in which Cronbach alpha coefficients were reported, and in addition six studies which were searched in MEDLINE, EMBASE, CINAHL and Google Scholar databases for 2016 and later were included in the meta-analysis. For studies involving more than one rater, Cronbach's alpha coefficients reported twice were both considered. Meta-analysis was carried out using the "metafor" package in the R program. Heterogeneity was evaluated by I^2 statistic and Q test and random effects meta-analysis was applied, overall Cronbach alpha value was calculated. Publication bias was assessed via rank correlation test "ranktest", via regression test "regtest" and funnel graph.

Results: A total of 19 Cronbach alpha values obtained from 16 studies ranged from 0.69 to 0.99. As a result of the meta-analysis, I^2 statistic was 95% and $p < 0.001$ for Q test. Since these findings represent heterogeneity, random effects meta-analysis were applied and the overall Cronbach alpha value was calculated as 0.87 (95% Confidence Interval: 0.84-0.90). Forest plot of these values was obtained. When the publication bias was evaluated visually by using funnel plot graph it was observed that the studies were not symmetrical. Asymmetry of the funnel plot was investigated by rank correlation and regression tests and it is found that funnel plot was not asymmetric with respect to the results of both tests ($p=0.298$, $p=0.057$, respectively). Based on these findings, it was concluded that there was no publication bias.

Conclusion: According to the findings of the study, the overall Cronbach alpha value for the Glasgow Coma Score was found to be within acceptable limits for the reliability of a scale.

Keywords: Cronbach alpha, meta-analysis for reliability, Glasgow coma score

S28

ULSAM Veri Setinde Tek Nükleotid Polimorfizmi Veri ile Makine Öğrenimi Yöntemlerinin Kullanımı

R. Onur ÖZTORNACI[1], Bahar TAŞDELEN, Andrew P. MORRIS[2], Hamzah SYED[3]

[1] Mersin University Faculty of Medicine, Department of Biostatistics and Medical Informatics, Mersin, Turkey
[2] School of Biological Sciences, Division of Evolution & Genomic Sciences, University Of Manchester, Manchester, UK
[3] Faculty of Pop Health Sciences, ICH Genetics & Genomic Medicine Prog., University College London, London, UK

Amaç: GWAS verilerindeki hastalıkla ilişkili tek nükleotid polimorfizmi (TNP) verilerini kullanarak tahmin ve sonuçları doğrulamak için makine öğrenimine (ML) başvurmak, genetik araştırmalarda sıklıkla kullanılmaktadır. ML ile ilgili sorunlardan birisi de değişken seçimi ve veri eğitimi için doğru yöntemi bulmaktır [1,2]. Bu çalışmanın amacı, 1178 yaşlı erkek katılımcının ULSAM tip 2 diyabet kohortundan TNP verilerini kullanarak vaka-kontrol durumunu tahmin etmek için yaygın olarak kullanılan iki değişken seçim yöntemini ve üç ML modelini karşılaştırmaktadır.

Yöntem: Verilerin analizi esnasında, standart genom boyu ilişki analizi (GWAS) kalite kontrol prosedüründen geçmiş olan veri ile PLINK kullanılarak yapıldı. GWAS sonuçlarına göre; birçok SNP, genom boyunda önemi olan tip 2 diyabet ile ilişkilendirildi. Tek değişkenli (ki-kare) analiz ve topaklanma (clumping) yöntemi ile p-değeri için eşik değeri 0,25 değişken seçiminde kullanıldı. Tip 2 diyabet ile ilişkili olan TNP'ler, üç farklı ML yöntemi uygulanarak tespit edilmeye çalışmıştır bu yöntemler; (i) Destek vektör makineleri (SVM), (ii) C5.0 ve (iii) rastgele ormandır (RF). Bu yöntemler vaka kontrol sınıflandırma sonuçlarının doğruluğunu karşılaştırmak için de kullanılmıştır.

Bulgular: GWAS sonuçlarından, TNP'ler rs11850812, rs12597362, rs16869618, rs12094313 ve rs11163137'nin tip 2 diyabet ile yüksek derecede ilişkili olduğu bulunmuştur. Ki-kare testi ile TNP seçiminde ML yöntemlerini kullanarak, en iyi sınıflandırma algoritması, vakaları ve kontrolleri sınıflandırmada yüksek doğru sınıflama oranına sahip olan DVM' dir.

Sonuç: ML tekniklerinin kullanımı geniş bir veri kümesi yelpazesinde yaygın olarak kullanılmaktadır. Bu teknikler klasik istatistiksel yaklaşımların aksine herhangi bir varsayımda bulunulmasını gerektirmez. Ancak, değişken seçimi ML genomik verilerle kullanırken önemli bir sorundur. Bu çalışmanın sonuçları Ki-kare testinin kümeleme (clumping) yöntemine göre daha iyi bir alternatif olarak değişken seçiminde kullanılabileceğini göstermektedir [3,4]. DVM sonuçları, bu veri setinde ML kullanmak için en doğru model olarak gösterilmiştir. Bu sonuçlara göre, ML yöntemlerinin, genomik verilerdeki TNP arasında bilinemeyen örüntüleri ortaya çıkartmak için, önsel bir istatistiksel çıkarım yöntemi ile kullanımı fayda sağlamaktadır.

Anahtar Kelimeler: Sınıflandırma, Tek nükleotid polimorfizmi, Makine Öğrenimi, GWAS

Referanslar

1. Cosgun, E., Limdi, N. A., & Duarte, C. W. (2011). High-dimensional pharmacogenetic prediction of a continuous trait using machine learning techniques with application to warfarin dose prediction in African Americans. Bioinformatics, 27(10), 1384-1389.

2. Chen, X., Zhang, M., Wang, M., Zhu, W., Cho, K., & Zhang, H. (2009, December). Memory management in genome-wide association studies. In BMC proceedings (Vol. 3, No. 7, p. S54). BioMed Central.

3. Hosmer Jr, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). Applied logistic regression (Vol. 398). John Wiley & Sons

4. Alpar, R. (2011). Çok Değişkenli İstatistiksel Yöntemler, Ankara: Detay Yayıncılık.

S28

The Use of Machine Learning Methods On SNP Data to Predict the Case-Control Status of Individuals in The ULSAM Dataset

R. Onur ÖZTORNACI[1], Bahar TAŞDELEN, Andrew P. MORRIS[2], Hamzah SYED[3]

[1] Mersin University Faculty of Medicine, Department of Biostatistics and Medical Informatics, Mersin, Turkey
[2] School of Biological Sciences, Division of Evolution & Genomic Sciences, University Of Manchester, Manchester, UK
[3] Faculty of Pop Health Sciences, ICH Genetics & Genomic Medicine Prog., University College London, London, UK

Aim: Machine learning methods for uncovering significant SNPs in GWAS data are becoming increasingly used in genetic research. An issue with machine learning is finding the correct method for feature selection and data training [1,2]. This work compares two widely used variable selection methods and three machine learning models to predict case-control status using SNP data from the ULSAM type 2 diabetes cohort of 1178 elderly male participants.

Method: Analysis of the data was performed using PLINK, following a standard GWAS quality control procedure. According to the GWAS results, many SNPs were associated with type 2 diabetes at genome-wide significance. Univariate (chi-square) analysis and clumping were applied for feature selection with a p-value threshold of 0.05. Irrelevant SNPs had been pruned by applying three different machine learning methods; (i) Support vector machine (SVM), (ii) C5.0 and (iii) random forest. These methods were used to compare the accuracy of the case-control classification results.

Results: From the GWAS results, the SNPs rs11850812, rs12597362, rs16869618, rs12094313 and rs11163137 were found to be highly associated with type 2 diabetes. Using machine learning methods with Chi-square test SNP selection, the best classification algorithm was SVM which had a 99% accuracy for classifying cases and controls.

Conclusion: The use of machine learning techniques is widely used for a wide range of data sets. These techniques do not require any assumptions to be made in contrast with classical statistical approaches. However, feature selection is one crucial problem when using machine learning with genomic data. The results of this article show that the Chi-squared test could be used for feature selection as a better alternative to the clumping method [3,4]. SVM results have been shown as the most accurate model for using machine learning on this dataset. According to these results, machine learning methods are beneficial to uncovering patterns within genomic data that would be missed when using methods with prior statistical inference.

Key Words: Classification, Machine Learning, Genome-wide Association Studies (GWAS), SNP Data

References

1. Cosgun, E., Limdi, N. A., & Duarte, C. W. (2011). High-dimensional pharmacogenetic prediction of a continuous trait using machine learning techniques with application to warfarin dose prediction in African Americans. Bioinformatics, 27(10), 1384-1389.

2. Chen, X., Zhang, M., Wang, M., Zhu, W., Cho, K., & Zhang, H. (2009, December). Memory management in genome-wide association studies. In BMC proceedings (Vol. 3, No. 7, p. S54). BioMed Central.

3. Hosmer Jr, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). Applied logistic regression (Vol. 398). John Wiley & Sons

4. Alpar, R. (2011). Çok Değişkenli İstatistiksel Yöntemler, Ankara: Detay Yayıncılık.

S29

Türkiye'de Bebek Ölüm Riskinin Mekan-zamansal Bayesci Modeller ile İncelenmesi

Sade Kılıç Yıldırım[1,2], Celal Reha Alpar[1], Erdem Karabulut[1], Atilla Halil Elhan[3]

[1]Türkiye İstatistik Kurumu, Ankara
[2]Hacettepe Üniversitesi Tıp Fakültesi Biyoistatistik Anabilim Dalı, Ankara
[3]Ankara Üniversitesi Tıp Fakültesi Biyoistatistik Anabilim Dalı, Ankara

Amaç: Bu çalışmada 2009 ve 2017 yılları arasında Türkiye'de illerdeki bebek ölüm riskini mekan, zaman ve mekan-zaman etkileşimi ile belirlemek, risk haritaları elde etmek ve etkileyen risk faktörlerini incelemek amaçlanmıştır.

Yöntem: Bebek ölümlerinin sayısının Türkiye'de 2009 ve 2017 yılları arasında her il için Poisson dağılımına sahip olduğu varsayılmıştır. Bağlantı fonksiyonu (log) ile dağılımın bilinmeyen parametresi bebek ölüm riski, mekan-zamansal Bayesci modelleriyle modellenmiştir. Modeller, R yazılımında INLA (entegre iç içe Laplace yaklaşımı) paketi ile uygulanmıştır. En iyi model sapma bilgi kriterine (DIC) göre belirlenmiştir. İlin kişi başına gayri safi yurtiçi hasılası ve anne yaşının bebek ölüm riski üzerindeki etkileri, en iyi mekan-zamansal model ve mekan-zaman kavramı olmadan genelleştirilmiş doğrusal model ile incelenmiştir.

Bulgular: 2009'dan 2017'ye kadar, yapılandırılmamış mekansal etki ve yapılandırılmış zamansal etki etkileşimine sahip en iyi model için, 1'den büyük riski olan il sayısı azalmaktadır. İlin kişi başına gayri safi yurtiçi hasılası; genelleştirilmiş doğrusal model ve mekan-zamansal model için bebek ölümü riski üzerinde azaltıcı etkiye sahiptir. Genelleştirilmiş doğrusal model için anne yaşının 20'den küçük olması ve 39'dan büyük olması risk üzerinde artırıcı etkiye sahiptir, ancak mekan-zamansal model için risk üzerinde bir etkisi yoktur. Risk faktörlerinin bulunduğu zaman-mekansal model, genelleştirilmiş doğrusal modelden daha düşük sapma bilgi kriterine (DIC) sahiptir.

Sonuç: Mekan-zamansal modelde anne yaşı risk üzerinde etkili değildir, ancak sapma bilgi kriteri (DIC) nedeniyle genelleştirilmiş doğrusal modelden daha çok tercih edilebilir. Bu nedenle, bebek ölüm riski yalnızca etkileyen faktörler açısından incelenmemeli, etkileyen risk faktörleri ile birlikte zaman, mekan ve zaman-mekan etkileşimi de göz önünde bulundurularak incelenmelidir.

Anahtar Kelimeler: entegre iç içe Laplace yaklaşımı, mekan-zamansal Bayesci model, bebek ölüm riski.

S29

Examination of Infant Mortality Risk in Turkey with Spatio-temporal Bayesian Models

Sade Kılıç YILDIRIM[1,2], Celal Reha ALPAR[2], Erdem KARABULUT[2], Atilla Halil ELHAN[3]

[1]Turkish Statistical Institute, Ankara
[2]Hacettepe University Medical Faculty Department of Biostatistics, Ankara
[3]Ankara University Medical Faculty Department of Biostatistics, Ankara

Aim: In this study it is aimed to determine the infant mortality risk for each province in Turkey between 2009 and 2017 years by including the concepts of space. time and space-time interaction. to obtain risk maps and to examine the risk factors affecting.

Method: It is assumed that the number of infant deaths has a Poisson distribution for each province in Turkey years between 2009 and 2017. Infant mortality risk which is unknown parameter of distribution with link function (log) is modeled with spatio-temporal Bayesian models. The models are applied with INLA (integrated nested Laplace approximations) package in R software. The best model is determined according to the deviance information criterion (DIC). The effects of per capita gross domestic product of the province and the maternal age on infant mortality risk are examined with the best spatio-temporal model and with linear model without the concepts of space and time.

Results: From 2009 to 2017. the number of provinces with the risk greater than 1 is decreased for the best model with unstructured spatial and structured temporal interaction random effect. Per capita gross domestic product of the province for the linear model (RR (Relative risk) = 0.881; 95% Cl (credibility interval) (0.875;0.887)) and spatio-temporal model (RR=0.907; 95% Cl (0.841;0.988) has negative effect on infant mortality risk. Maternal age smaller 20 (RR=1.021; 95% Cl (1.018; 1.024) and older than 40 (RR=1.08; 95% Cl (1.073; 1.087)) has positive effect on risk for the linear model. but no effect on risk for the spatio-temporal model. DIC of the spatio-temporal model with risk factors and linear model is 5753.27 and 8558.05 respectively. The contribution of spatial and temporal interaction effect and the temporal effect. to the variability in risk explained by the spatio-temporal model is 52.61 % and 29.41 % respectively. with adding risk factors to the model the contribution of these effects is 59.70 % and 14.89% respectively.

Conclusion: The contribution of the temporal effect to the variability in risk modeled with spatio-temporal model adding risk factors decreases markedly compared to model without risk factors. but interaction effect increases. In spatio-temporal model maternal age has no effect on risk. but because of DIC it is more preferable than linear model. Hence infant mortality risk should not be examined only in terms of affecting factors. it is very important to consider the effect of time. space and time-space interaction with affecting factors.

S30

Akciğer Adenokarsinomu için Potansiyel Biyobelirteçlerin *In Silico* Araştırılması

Ceren SUCULARLI[1]

[1]*Biyoinformatik Anabilim Dalı, Hacettepe Üniversitesi, Ankara*

Amaç: Akciğer adenokarsinomu, ileri evrede düşük sağkalım oranı gösteren ve yüksek insidansı olan akciğer kanseri türlerinden biridir. Kanser gen ekspresyonu veri setlerini kullanarak gen ekspresyonu değişikliklerini bulmak, kanserlerin moleküler mekanizmalarını tanımlamak ve potansiyel kanser biyobelirteçleri önermek için kullanılmıştır. Bu çalışmada amacım, potansiyel akciğer adenokarsinom biyobelirteçlerini saptamak için bilgisayımsal yöntemler kullanarak akciğer adenokarsinomunda gen ekspresyonu değişikliklerini bulmak ve sonrasında ekspresyonu istatistiksel olarak anlamlı değişen genleri fonksiyonel annotasyon, protein-protein etkileşim ağları ve hasta sağkalımı açısından analiz etmekti.

Yöntem: Halka açık akciğer adenokarsinomu RNA-seq veri setlerini kontrolleri ile analiz ettim. Ekspresyonu anlamlı olarak değişen genlerin fonksiyonel annotasyonları Kyoto Encyclopedia of Genes and Genomes (KEGG) yolakları ve Biyolojik İşlev Gen Ontolojisi terimleri için gerçekleştirildi. İfadesi anlamlı olarak artan genler protein-protein etkileşim ağları oluşturularak analiz edildi ve merkez genler ileri araştırma için seçildi. Seçilen genlerin gen ekspresyonlarının hasta sağkalımı üzerindeki etkisi akciğer kanseri veri setlerinde araştırıldı.

Bulgular: 957 gen, 247 ekspresyonu artan ve 705 ekspresyonu azalan, akciğer adenokarsinomunda istatistiksel olarak anlamlı değişiyor olarak saptandı ($|\log_2 FC| > 2$ ve q-value < 0.001). Ekspresyonu artan genler çoğunlukla hücre döngüsü ve hücre bölünmesi ile ilgili terimlere annotate oldu (Benjamini-Hochberg corrected p-value < 0.05). Ekspresyonu artan genlerden seçilenlerin bazıları akciğer adenokarsinomu hastalarında genel hasta sağkalımı ile anlamlı bir ilişki gösterdi (log rank p-value <0.05).

Sonuç: Çalışmam akciğer adenokarsinomu için potansiyel prognostik ve prediktif biyobelirteçler önermiştir.

Anahtar Kelimeler: Gen ekspresyonu, RNA-seq, protein-protein etkileşim ağları, hasta sağkalımı

S30

In silico Investigation of Potential Biomarkers of Lung Adenocarcinoma

Ceren SUCULARLI[1]

[1]*Department of Bioinformatics, Hacettepe University, Ankara*

Aim: Lung adenocarcinoma is one of the major types of lung cancer, which shows poor survival rate in the advanced stage and has a high incidence rate. Finding gene expression alterations using cancer gene expression datasets has been used to identify molecular mechanisms of cancers and to propose potential cancer biomarkers. In this study, my aim was to find gene expression alterations in lung adenocarcinoma and further analyze differentially expressed genes in terms of functional annotation, protein-protein interaction networks and patient survival by using computational methods to find potential lung adenocarcinoma biomarkers.

Method: I analyzed publicly available RNA-seq datasets of lung adenocarcinoma with controls. Functional annotation of differentially expressed genes to Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways and Gene Ontology terms for Biological Process were performed. Differentially expressed upregulated genes were investigated by generating protein-protein interaction networks and hub genes were selected for further analysis. The effect of gene expressions of selected genes on patient survival was investigated in lung cancer datasets.

Results: 952 genes, 247 upregulated and 705 downregulated, were identified as differentially expressed in lung adenocarcinoma ($|\log_2 FC| > 2$ and q-value < 0.001). Upregulated genes mostly annotated to cell cycle and cell division related terms (Benjamini-Hochberg corrected p-value < 0.05). Expression of several selected upregulated genes showed significant association with overall survival of lung adenocarcinoma patients (log rank p-value < 0.05).

Conclusion: My study proposed potential prognostic and predictive biomarkers for lung adenocarcinoma.

Keywords: Gene expression, RNA-seq, protein-protein interaction networks, patient survival

S31

Ensemble-Based Machine Learning Prediction of Heart Disease Presence with The Cleveland Database

Setia PRAMANA[1], Mieke NURMALASARI[2]

[1]Department of Computational Statistics, Politeknik Statistika STIS, Jakarta, Indonesia
[2]Department of Health Information Management, Faculty of Health and Sciences, Universitas Esa Unggul, Jakarta, Indonesia

Disease diagnosis in heart disease extremely important because it is the first killer in the world. Detection of the presence of heart disease in patients can be optimized by make classifications based on the factors. Using only a single prediction algorithm can lead to non-robust and overfitting prediction. It can be solved by using ensemble methods by using multiple learning algorithms to obtain better predictive performance. In this study, the heart disease Cleveland database with 14 features such as age, sex, chest pain type, resting blood pressure, serum cholesterol were used. Two ensemble-based algorithms e.g., random forest, gradient boosting and super learner are carried out and compared to predict the heart disease presence. The result shows that random forest provide better prediction. It is shown that chest pain type, number of major vessels coloured by fluoroscopy, defect type, ST depression induced by exercise and maximum heart rate achieved are the most important predictors for heart disease presence.

Keywords: heart disease, health, random forest, gradient boosting, classification

S32

Hypoxia inducible factor 2 alpha as a marker of arteriovenous fistula maturation in patients with end stage renal disease

Eray EROĞLU[1], İsmail KOÇYIĞIT[2], Ciğdem KARAKÜKÇÜ[3], Aydın TUNÇAY[4],Gökmen ZARARSIZ[5],Davut EREN[6], Murat Hayri SIPAHIOGLU[2], Bülent TOKGÖZ[2], Kutay TAŞDEMIR[4], Oktay OYMAK[2]

[1]*Karolinska Institutet, Department of Clinical Science Intervention and Technology, Division of Renal Medicine and Baxter Novum, Stockholm, Sweden*
[2]*Erciyes University School of Medicine, Department of Internal Medicine, Division of Nephrology, Kayseri, Turkey*
[3]*Kayseri City Hospital, Department of Biochemistry, Kayseri, Turkey*
[4]*Erciyes University School of Medicine, Department of Cardiovascular Surgery, Kayseri, Turkey*
[5]*Roche Diagnostics GmbH, Personalized Health Care, Munich, Germany*
[6] *Erciyes University School of Medicine, Department of Internal Medicine, Kayseri, Turkey*

Aim: Arteriovenous fistula (AVF) failure is one of the most important clinical problems in patients with end-stage renal disease (ESRD). A specific biomarker of AVF maturation has not been identified yet. Hypoxia inducible factor 2α (HIF-2α) is tended to be increase in hypoxic conditions and it might be related with hypoxic vascular complications of ESRD. Thus we aimed to investigate the association between HIF-2α and AVF maturation in ESRD patients whom had chosen hemodialysis for the first renal replacement therapy and fistula had been created for the first time.

Methods: This prospective cohort study was conducted in 21 voluntary healthy subjects and 50 patients with end-stage renal disease who were admitted for their first renal replacement therapy and were available for AVF creation. Venography imaging was performed in all patients before AVF creation. Eight patients were excluded due to inadequate veins after venography imaging, and also six patients were excluded due to postoperative thrombosis and one patient declined to participate during follow-up. Total 35 patients were included final analysis. The blood samples were analyzed a day before the fistula operation, and serum levels of HIF-2α were measured.

Results: Compared to healthy subjects, ESRD patients had significantly higher level of HIF-2α. [1.3(1.0-1.9) vs 2.2(1.6-3.0)] (p=0.002). Patients were divided into two groups after the evaluation of AVF maturation, as the mature group and the failure group. Serum HIF-2α level was 1.7(1.1-1.8) in mature group however it was 3.1(2.8-3.3 in failure group (p< 0.001) (Figure-1). Multiple logistic regression analyses showed that HIF-2α independently predicted AVF maturation. The ROC curve analysis showed that HIF-2α >2.65 had an ability of predicting AVF maturation failure with the 87% sensitivity and 94% specificity [AUC:0.947, 95% CI (0.815-0.994), p< 0.001] (Figure-2).

Conclusion: Increased level of HIF-2-α could be a marker of failure of AVF maturation in patients with ESRD.

Keywords: Arteriovenous fistula, Hypoxia, Vascular remodeling

S33

Plantar Fasiit Tanısında Performans Değerlendirme için Çapraz Doğrulama Yöntemlerinin Karşılaştırılması

Senem KOÇ[1], Leman TOMAK[1]

[1] Ondokuz Mayıs Üniversitesi Biyostatistik and Tıp Bilişimi, Samsun

Amaç: Çapraz doğrulama, istatistiksel analiz sonuçlarının bağımsız bir veri setine nasıl genelleştirildiğinin değerlendirilmesinde kullanılan bir yöntemdir. Bu çalışmanın amacı, plantar fasiit teşhisi için farklı makine öğrenme algoritmaları kullanarak farklı çapraz doğrulama yöntemlerinin uyumluluğunu belirlemektir.

Metot: Çapraz doğrulama, makine öğrenme modellerinin başarılarının değerlendirilmesi için kullanılan bir yöntemdir. Bu yöntemde veri seti iki bölüme ayrılır. Bir bölüm bir modele uyması için kullanılan eğitim seti, diğer bölüm ise tahmin hatasını tahmin ederek bir modelin performansını değerlendiren bir test setidir. Farklı çapraz doğrulama yöntemleri vardır ve bu çalışmada dört farklı yöntem uygulanmıştır. Bunlar, bootstrap yöntemi, 10-kat çapraz doğrulama, tekrarlanan çapraz doğrulama ve birini dışarıda bırak çapraz doğrulama yöntemidir. Değerlendirme amacıyla iki farklı makine öğrenme algoritması, Destek Vektör Makineleri (DVM) ve Rastgele Orman (RO) kullanılmıştır. Plantar fasiit teşhisi için Kaggle'dan bir veri seti kullanılmıştır. Plantar fasiit, topuk ağrısının en sık görülen nedenlerinden biridir. Bu veri seti, 384 gözlem ile hedef değişken olan plantar fasiitin belirlenmesi için kullanılan 12 değişkenden oluşmaktadır. Analizler R yazılımı kullanılarak yapıldı.

Bulgular: Elde edilen sonuçlara göre, doğruluk değeri karşılaştırma olarak kullanıldığında DVM algoritması için tekrarlı çapraz doğrulama 0,8704, bootstrap ile 0,9074, 10 kat çapraz doğrulama ile 0,8704 ve birini dışarıda bırak yöntemi çaraz doğrulama 0,8889 olarak elde edilmiştir. RO için tekrarlanan çapraz doğrulama değeri 0,9907, 10 kat çapraz doğrulama 0,9815, bootstrap ile 0,9907 ve birini dışarıda bırak çapraz geçerleme 0,9970'dir.

Sonuç: RF, SVM algoritmasına göre daha iyi performans göstermektedir. Bootstrap yönteminde kullanılan değer ve tekrarlı çapraz geçerleme için kullanılan tekrar sayısı doğruluk üzerine etkilidir. Sonuç olarak birini dışarıda bırak çapraz doğrulama, farklı çapraz doğrulama yöntemlerine kıyasla daha iyi performans göstermektedir.

Anahtar Sözcükler: Çapraz- Geçerleme, Destek Vektör Makineleri, Rastgele Orman, Plantar Fasciitis.

S33

Compatibility of Cross-Validation Methods for Performance Evaluation for Diagnosing Plantar Fasciitis

Senem KOÇ[1], Leman TOMAK[1]

[1] *Ondokuz Mayıs University Biyostatistik and Medical Informatics, Samsun*

Aim: Cross-validation (CV) is a method that is used for the evaluation of how the results of statistical analysis generalize to an independent data set. The aim of this study is to determine the compatibility of different CV methods by using different machine learning algorithms for diagnosing plantar fasciitis.

Method: CV is a data resampling method which involves fitting the same analyze method multiple times using different subsets of the data set. The data set is split into two parts. One part is the training set which is used to fit a model, the other part is a testing set to evaluate the performance of a model by estimating the prediction error. There were different cross validation methods and in this study, we applied four different methods. These are bootstrap, 10-fold cross validation (CV), repeated CV and leave one out CV. For evaluation purposes two different machine learning algorithms, Support Vector Machines (SVM) and Random Forest (RF) were used. We used a data set from Kaggle for the diagnosis plantar fasciitis. Plantar fasciitis is one of the most common causes of heel pain. The data set consist of 384 observations and 12 variables with a target variable for determining plantar fasciitis. The analyses were carried out using R software.

Results: According to the results obtained, when using accuracy as a metric repated CV is 0.8704, bootstrap is 0.9074, 10 fold CV is 0.8704 and leave one out CV is 0.8889 by SVM algorithm. For RF repeated CV is 0.9907, 10 fold CV is 0.9815, bootstrap is 0.9907 and leave one out CV is 0.9970.

Conclusion: RF shows better performance according to SVM algorithm. The number of bootstrap effects also the accuracy so the number of repeats by repeated CV effect it. Overall leave one out CV shows better performance compared to different CV methods.

Keywords: Cross- Validation, Support Vector Machines, Random Forest, Plantar Fasciitis.

S34

Faktöriyel Tasarımlarda Veri Dönüştürme Yöntemlerinin Varyans Analizinde Gerçekleşen I. Tip Hata ve Testin Gücü Üzerine Etkisi

Yeliz KAŞKO ARICI[1], Mustafa Muhip ÖZKAN[2]

[1]Ordu Üniversitesi, Tıp Fakültesi, Biyoistatistik ve Tıbbi Bilişim A.D., Ordu
[2]Ankara University, Ziraat Fakültesi, Biyometri ve Genetik A.D., Ankara

Amaç: Faktöriyel tasarımlı çalışmalardan elde edilen verilerin analizinde en yaygın kullanılan istatistik yöntem faktöriyel varyans analizidir. Bunun sebebi faktörlerin esas etkilerinin yanı sıra faktörler arasındaki etkileşimin de interaksiyon terimine ait hipotez kontrolü ile ortaya konulabilmesidir. Çalışılan sürekli değişkenler her zaman varyans analizinin varsayımlarını yerine getirmeyebilir. Özellikle homojen olmayan alt grup varyansları araştırıcıların en büyük problemidir. Varyans analizine alternatif olarak kullanılacak yöntemler de çoğunlukla interaksiyon bilgisinin kaybına neden olmaktadır. Bu sebeple araştırıcılar verilerini varyans analizine uygun hale getirmek amacıyla veri dönüştürme yöntemlerine başvurabilmektedir. Bu çalışmada, alt grup varyansları homojen olmayan faktöriyel tasarımlı çalışmalarda veri dönüştürme yöntemlerinin interaksiyon terimine ait hipotez kontrolünde gerçekleşecek I. tip hata ve testin gücü değerleri üzerine etkilerinin simülasyon yaklaşımı ile araştırılması amaçlandı.

Yöntem: Çalışmada N(0,1) dağılımından tesadüf sayıları üretilmiş ve Monte Carlo simülasyon yöntemiyle 50.000 simülasyon denemesi sonucunda I. tip hata ve testin gücü değerleri hesaplandı. Tüm hesaplamalar fortran programlama dilinde yazılmış programlar yardımıyla yapıldı. İnteraksiyona ilişkin hipotez kontrolü, sırasıyla 2×2, 2×3, 2×5, 3×3, 3×4, 3×5, 2×2×2 ve 2×3×4 faktöriyel düzenlerde yapıldı. Alt gruplardaki gözlem sayısı (n) her birinde eşit olmak üzere sırasıyla 3, 4, 5, 8, 10 ve 20 olarak belirlendi. Alt gruplardan birinin varyansı diğerlerinin varyansının sırasıyla 2, 3, 5, 9 ve 16 katı kadar artırılarak heterojen hale getirildi. Veri dönüştürme yöntemi olarak karekök, logaritmik ve açı transformasyon yöntemleri kullanıldı. Veri dönüştürme uygulandıktan sonra alt grup varyanslarının homojenlik kontrolü Hartley testi ile yapıldı. Testin gücü değerleri sırasıyla 0.5, 1.0, 1.5 etki büyüklükleri (d) için hesaplandı.

Bulgular: Tüm deneme koşullarında hem I. tip hata hem de testin gücü bakımından veri dönüştürme yöntemleri benzer sonuçlar gösterdi. Alt grupların varyansları arasındaki heterojenlik düzeyi arttıkça kararlaştırılan I. tip hata olasılığı korunamadı ve n kaç olursa olsun artış gösterdi. Faktör ve faktör seviye sayısındaki artışın da I. tip hata olasılıklarında artışa sebep olduğu görüldü. Testin gücü d=0.5 için alt grup varyansları arasındaki heterojenlik düzeyi arttıkça yükselirken, d=1.0 ve d=1.5 için alt grup varyansları arasındaki heterojenlik düzeyi arttıkça ciddi düşüşler gösterdi. Heterojenlik düzeyi özellikle 3 katın üzerinde ve n<10 olduğunda varyans analizinin güvenilirliği tüm deneme koşullarında ciddi olarak azaldı. Faktör sayısındaki artış durumun ciddiyetini arttırdı.

Sonuç: Alt grup varyansları homojen olmayan faktöriyel tasarımlı çalışmalarda, veri dönüştürme yöntemlerinin kullanılması faktöriyel varyans analizinde interaksiyon terimine ait hipotez kontrolünde varılacak olan kararların güvenilirliğini ciddi bir biçimde etkilemektedir. Dolayısıyla veri dönüştürme kullanılmasından kaçınılmalıdır.

Keywords: Faktöriyel tasarım, varyansların homojenliği, veri dönüştürme, I. tip hata, testin gücü

S34

The Effect of Data Transformation Methods on the Type I Error and Test Power of ANOVA in Factorial Designs

Yeliz KAŞKO ARICI[1], Mustafa Muhip ÖZKAN[2]
[1]*Ordu University, Faculty of Medicine, Department of Biostatistics and Medical Information, Ordu*
[2]*Ankara University, Faculty of Agriculture, Biometry and Genetics Unit, Ankara*

Objective: The most common statistical method used in the analysis of data obtained from factorial design studies is factorial ANOVA. The reason for this is that in addition to the main effects of the factors, their interactions can be demonstrated by hypothesis test of the interaction term. The studied continuous variables may not always fulfill the assumptions of ANOVA. Especially non-homogeneous subgroup variances form the biggest problem for researchers. As a result, with the aim of making data appropriate for ANOVA, researchers may apply data transformation methods. In this study, the data transformation method for hypothesis test belonging to the interaction term for subgroup variances in non-homogeneous factorial design studies will be completed with the aim of researching the effects on the type I error rates and power of the test with the simulation approach.

Method: In the study, random numbers were generated from the N (0,1) distribution and calculated type I error rates and power of the test for the test at the end of 50.000 simulation trials with the Monte Carlo simulation method. All calculations were performed with the aid of programs written in the Fortran programming language. Hypothesis test related to interaction was performed with 2×2, 2×3, 2×5, 3×3, 3×4, 3×5, 2×2×2 and 2×3×4 factorial designs, respectively. Number of observations in subgroups were equal for each and were determined as 3, 4, 5, 8, 10 and 20. The variance for one of the subgroups increased the variance for the others by 2, 3, 5, 9 and 16 times to make it heterogeneous. For data transformation methods, the square root, logarithmic and arcsine transformation methods were used. After applying data transformation, the homogeneity of the subgroup variances was checked with the Hartley test. The power values for the test were calculated to have 0.5, 1.0 and 1.5 effect sizes (d).

Results: For all trial conditions, the data transformation methods provided similar results for both type I error and power of the test. As the heterogeneity level between the subgroup variances increased, the agreed type I error probability could not be maintained and increased at all number of observations. The increase in factor and factor level numbers was observed to cause an increase in type I error probabilities. For test power, as the level of heterogeneity between subgroup variances increased, it increased with d=0.5, while as the heterogeneity between the subgroup variances increased it displayed severe falls for d=1.0 and d=1.5. Heterogeneity level, especially more than 3 times and with n<10, reduced the reliability of ANOVA severely under all trial conditions. The increase in factor numbers increased the severity of this situation.

Conclusion: In factorial designed studies without homogeneity for subgroup variances, the use of data transformation methods affects the reliability of decisions made for hypothesis test of the interaction term belonging to factorial ANOVA in a severe manner. As a result, the use of data transformation should be avoided.

Keywords: Factorial design, homogeneity of variances, data transformation, type I error, test power

S35

MAKİNE ÖĞRENMESİ İLE GENETİK VARYASYON ÇEŞİTLERİNİN TAHMİNLENMESİ

Su ÖZGÜR[1], Erdal COŞGUN[2], Kemal Uğur TÜFEKCİ[3], Pembe KESKİNOĞLU[4], Görsev YENER[5], Şermin GENÇ[3], Mehmet N. ORMAN[1]

[1]Ege Üniversitesi Tıp Fakültesi Biyoistatistik ve Tıbbi Bilişim AD, İzmir
[2]Genomics Team, Microsoft Research, Redmond, WA, USA
[3]İzmir Biyotıp ve Genom Merkezi, İzmir
[4]Dokuz Eylül Üniversitesi Tıp Fakültesi Biyoistatistik ve Tıbbi Bilişim AD, İzmir
[5]Dokuz Eylül Üniversitesi Tıp Fakültesi Nöroloji Anabilim Dalı, İzmir

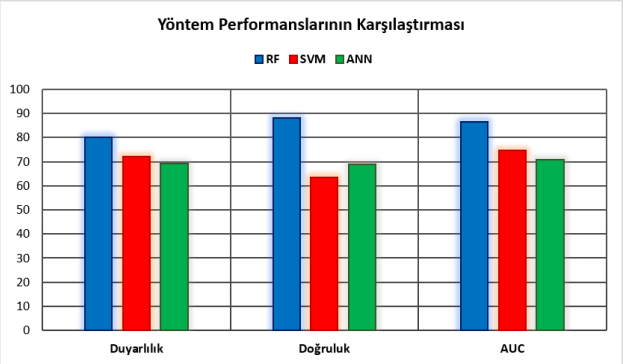
Amaç: Gelişen yeni nesil sekanslama (YNS) teknolojileriyle son yıllarda yapılan çalışmalar kanser, kalp hastalıkları, diyabet, Crohn hastalığı, Alzheimer gibi kompleks hastalıkların temel olarak, DNA moleküllerinde belirli bir yerde tek bir nükleotidin (A, T, C ve G) değişmesinden veya çoklu nükleotid varyantlarını içeren genlerden kaynaklandığını göstermiştir. Karmaşık hastalıklar, karmaşık genetik mimari ile karakterize edildiğinden çeşitli genleri ve değişken etkilerini içerir, bu da değişkenleri fenotipik farklılıklarla bağlama görevini araştırmadaki en büyük zorluklardan biri yapar. DNA moleküllerinin genetik bilgisini taşıyan tek nükleotid polimorfizmi (SNP), genetik varyantların en temel birimidir. SNP'ler, kodlama yapan, kodlama yapmayan veya genler arasındaki bölgeler dahil olmak üzere, DNA dizisini etkileyebilir. Bu çalışmada YNS verilerinin ikincil analizleri sonucu elde edilen variant calling format (VCF) verisinin makine öğrenmeye (ML) dayalı yöntemlerle farklı varyant çeşitlerini: SNP, MNP (Multiple Nucleotide Polymorphisms)/complex'ten ayırt ediciliğinin tahminlenmesi amaçlanmıştır.

Yöntem: Çalışmaya 8 Alzheimer hastası (AH) ve 8 sağlıklı birey dahil edilmiştir. AH olguları Dokuz Eylül Üniversitesi Tıp Fakültesi Nöroloji AD Demans polikliniğinde klinik inceleme, nöropsikolojik ve beyin omurilik sıvısı (BOS) AH biyobelirteç değerlendirmesiyle tanı almış, takip edilmekte olan hastalardan seçilmiştir. Kontrol grubu, hastalarla benzer demografik özelliklere sahip gönüllü sağlıklı kişilerden oluşturulmuştur. AH ve sağlıklı bireylerden elde edilen single end YNS (Human-exome) verisi Illumina NextSeq 500 ile fastq formatında elde edilmiştir. Kalite kontrolü gerçekleştirildikten sonra Bowtie yöntemiyle referans genoma (Hg 38) hizalanmıştır. Hizalanan veriye Variant Calling için en sık kullanılan ardışık düzen olan (pipeline) FreeBayes yöntemi uygulanarak SNP'leri, indelleri ve complex varyantları tanımlayan VCF verisi elde edilmiştir. Bu verideki; DP (Read depth), DPB (Total read depth per bp at the locus), AC (Allele count), AF (Allele frequency), RO (Reference allele observation count), AO (Alternate allele observation count), QR (Sum of quality of the reference observations), QA (Sum of quality of the alternate observations), SAP (Strand balance probability for the alternate allele), ODDS (The log odds ratio of the best genotype combination to the second-best) değişkenleri, random forest (RF), karar destek makineleri (SVM) ve yapay sinir ağları (ANN) yöntemleriyle SNP'lerin doğru şekilde ayırt edilmesi için kullanılmıştır.

Bulgular: SNP'leri tahminlemek için üç yöntem kullanılmıştır.

Tablo 1. SNP'leri tahminlemede kullanılan yöntemlerin karşılaştırması

	RF	SVM	ANN
Duyarlılık	80,00	88,0	86,70
Doğruluk	72,03	63,64	74,83
AUC	69,20	68,88	70,70



Sonuç: SNP olarak tanımlanan varyantların doğru ayırt edilmesi için kullanılan ML yöntemlerinin sonuçlarına göre en yüksek doğruluğun ve AUC oranının ANN'de olduğu görülmektedir. Ancak SNP olarak tanımlanan varyantların tahminlemesinde SVM daha yüksek performans göstermiştir. Özellikle son yıllarda ağaç (decision tree) tabanlı ML algoritmalarının gösterdiği yüksek performans göz önüne alındığında, bağımsız değişkenler arasında düşük-orta korelasyonun ağaç bölünmelerinde tahmine katkısı olduğu hipotezi bu çalışma için de doğrulanamamıştır. Böylelikle NN tabanlı derin öğrenme (DL) gibi yöntemlerin araştırmalarda uygulanması kaçınılmaz görünmektedir.

Teşekkür: *Bu çalışma TÜBİTAK tarafından 217S584 proje numarası ile desteklenmiştir.*

Anahtar Kelimeler: Yeni Nesil Sekanslama, Variant Calling Format, Makine Öğrenmesi

Kaynaklar

1. Li, B., Zhang, N., Wang, Y. G., George, A. W., Reverter, A., & Li, Y. (2018). Genomic Prediction of Breeding Values Using a Subset of SNPs Identified by Three Machine Learning Methods. *Frontiers in genetics*, *9*, 237. doi:10.3389/fgene.2018.00237

2. Nannya, Y., Taura, K., Kurokawa, M., Chiba, S., Ogawa, S. 2007. Evaluation of Genome-Wide Power of Genetic Association Studies Based on Empirical Data from The HapMap Project. Human Molecular Genetics, 16(20), 2494–2505

3. Garrison E, Marth G. Haplotype-based variant detection from short-read sequencing. arXiv preprint arXiv:1207.3907 [q-bio.GN] 2012

4. Piskol R, Ramaswami G, Li JB. Reliable Identification of Genomic Variants from RNA-Seq Data, Am J Hum Genet. 2013 Oct 3;93(4):641-51. doi: 10.1016/j.ajhg.2013.08.008. Epub 2013 Sep 26.

S35

ESTIMATION OF GENETIC VARIATION TYPES VIA MACHINE LEARNING

Su ÖZGÜR[1], Erdal COŞGUN[2], Kemal Uğur TÜFEKCİ[3], Pembe KESKİNOĞLU[4], Görsev YENER[5], Şermin GENÇ[3], Mehmet N. ORMAN[1]

[1]Ege University Faculty of Medicine Department of Biostatistics and Medical Informatics, İzmir
[2]Genomics Team, Microsoft Research, Redmond, WA, USA
[3]İzmir Biomedicine and Genome Center, İzmir
[4]Dokuz Eylül University Faculty of Medicine Department of Biostatistics and Medical Informatics, İzmir
[5]Dokuz Eylül University Faculty of Medicine Department of Neurology, İzmir

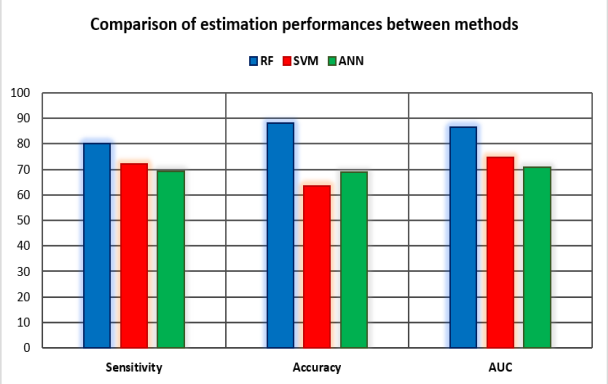
Aim: Recent studies with developing new generation sequencing technologies (NGS) have shown that complex diseases such as cancer, heart disease, diabetes, Crohn's disease, Alzheimer are mainly caused by the change of a single nucleotide (A, T, C and G) in a specific place in DNA molecules or genes containing multiple nucleotide variants. Since complex diseases are characterized by complex genetic architecture, they involve various genes and their variable effects. This makes it one of the biggest challenges in investigating the task of linking variables with phenotypic differences. The single nucleotide polymorphism (SNP) that carries the genetic information of DNA molecules is the most fundamental unit of genetic variants. SNPs may affect the DNA sequence, including the regions encoding, non-coding, or between genes. In this study, it is aimed to predict the differentiation of different variant types (SNP, MNP/Complex) of VCF (variant calling format) data which is obtained from secondary analysis of NGS data by using machine learning methods.

Methods: Eight patients with Alzheimer's disease (AD) and eight healthy subjects were included in the study. Patients with AD were selected from the patients who were diagnosed and followed up in Dokuz Eylul University Facult of Medicine Neurology Dementia outpatient clinic with clinical examination, neuropsychological and cerebrospinal fluid (CSF) evaluation of AH biomarker. The control group consisted of healthy volunteers with similar demographic characteristics. Single end NGS (Human-exome) data obtained from AD and healthy subjects were obtained via Illumina NextSeq 500 in fastq format. The data was aligned to the reference genome (Hg 38) by the Bowtie method after quality control procedures were completed. The FreeBayes method, which is the most commonly used consecutive order (pipeline) for variant calling, was applied to the aligned data to obtain VCF data that describes SNPs, indices and complex variants. In this data; DP (Read depth), DPB (Total read depth per bp at the locus), AC (Allele count), AF (Allele frequency), RO (Reference allele observation count), AO (Alternate allele observation count), QR (Sum of Quality of the reference observations, QA (Sum of quality of the alternate observations), SAP (Strand balance probability for the alternate allele), ODDS (The log odds ratio of the best genotype combination to the second-best) variables, random forest (RF), decision support machines (SVM) and artificial neural networks (ANN) methods were used to accurately distinguish SNPs.

Results: Three methods were used to estimate SNPs.

Table 1. Comparison of methods used to estimate SNPs

	RF	SVM	ANN
Sensitivity	80,00	88,0	86,70
Accuracy	72,03	63,64	74,83
AUC	69,20	68,88	70,70



Conclusions: According to the results of ML methods used to correctly distinguish variants defined as SNP, the highest accuracy and AUC ratio were obtained in ANN. However, SVM performed better in estimating variants identified as SNP. Especially in recent years, considering the high performance of the decision tree based ML algorithms, the hypothesis that the low-medium correlation between independent variables contributes to the estimation of tree splits could not be confirmed in this study. Thus, it seems inevitable to apply NN-based deep learning (DL) methods in research.

Acknowledgment: *This study was supported by TÜBİTAK with project number 217S584.*

Keywords: Next Generation Sequencing, Variant Calling Format, Machine Learning

References

1. Li, B., Zhang, N., Wang, Y. G., George, A. W., Reverter, A., & Li, Y. (2018). Genomic Prediction of Breeding Values Using a Subset of SNPs Identified by Three Machine Learning Methods. *Frontiers in genetics*, *9*, 237. doi:10.3389/fgene.2018.00237

2. Nannya, Y., Taura, K., Kurokawa, M., Chiba, S., Ogawa, S. 2007. Evaluation of Genome-Wide Power of Genetic Association Studies Based on Empirical Data from The HapMap Project. Human Molecular Genetics, 16(20), 2494–2505

3. Garrison E, Marth G. Haplotype-based variant detection from short-read sequencing. arXiv preprint arXiv:1207.3907 [q-bio.GN] 2012

4. Piskol R, Ramaswami G, Li JB. Reliable Identification of Genomic Variants from RNA-Seq Data, Am J Hum Genet. 2013 Oct 3;93(4):641-51. doi: 10.1016/j.ajhg.2013.08.008. Epub 2013 Sep 26.

S36

Pankreas Kanseri verilerinin Paylaşılmış Zayıflık modeli ile analizi

Şirin ÇETİN[1], Osman DEMİR[1], İsa DEDE[2], Özge PASİN[3], Emre KUYUCU[1]

[1]Tokat Gaziosmanpaşa Üniversitesi Tıp Fakültesi Biyoistatistik Anabilim Dalı,
[2]Mustafa Kemal Üniversitesi Tıp Fakültesi Tıbbi Onkoloji Anabilim Dalı,
[3]İstanbul Üniversitesi, İstanbul Tıp Fakültesi, Biyoistatistik Anabilim Dalı.

Amaç: Pankreas kanseri verilerinin yapıları incelendiğinde, genellikle bireyler arasında bir heterojenlik söz konusudur. Veride heterojenliğin mevcut olduğu durumlarda klasik yaşam analizi çözümlemelerinin kullanılması, modelin yeteri kadar açıklanamamasına ve yanlı sonuçların oluşmasına neden olabilmektedir. Dolayısı ise heterojen yapıya sahip verilerin yaşam analizlerinde zayıflık modellerinin kullanılması önerilmektedir. Yapılan bu çalışmada, heterojen yapıya sahip pankreas kanseri tanısı almış 100 hastanın yaşam analizi verilerinin modellenmesi ve sağkalım sürelerinin belirlenmesi amaçlanmıştır.

Yöntem: Modelleme de paylaşılmış zayıflık modeli, dağılıma uygunluk için ise Üstel, Weibull, log-normal, Gompertz regresyon modelleri kullanılmıştır. Model seçiminde, Akaike bilgi kriteri, Bayesci bilgi kriteri ve grafiksel yöntemlerine bakılarak değerlendirme yapılmıştır. Kurulan paylaşılmış zayıflık modelinde pankreas kanseri hastaları zayıflık süreleri koşullu bağımsız olmak üzere aynı zayıflığı paylaşmaktadır.

Bulgular: Modellerin anlamlılığını test etmek için olabilirlik oran test istatistiği kullanılmış ve tüm modellerin istatistiksel olarak anlamlı olduğu sonucuna varılmıştır ($p < 0.05$). Sonuçta pankreas kanseri veri seti için en uygun modelin gamma zayıflık terimi içeren üstel dağılımı ile kurulan model olduğuna karar verilmiştir. Pankreas kanseri hastalarının yaşam süresini etkileyen CA19- 9 ($p=0.03$) prognostik faktörünün yanı sıra ek hastalık (diyabet) veya genetik faktörlerininde etkili olabileceği söylenebilir. Bu ek hastalık(diyabet) veya genetik faktörlerin herhangi bir pankreas kanseri hastasında riski 1.43 kat artırdığı belirlenmiştir.

Sonuç: Sonuç olarak, veri yapısının heterojen olduğu ve hastalığı açıklayacak tüm açıklayıcı değişkenlerin modele dâhil edilmediği durumlarda, klasik yaşam analizi yöntemleri yerine zayıflık modellerinin kullanılması önerilmektedir.

S37

Çoklu Sınıflama Durumunda Sınıflama Performansı Testlerinin Bir Uygulama ile Karşılaştırılması

Tanyeli GÜNEYLİGİL KAZAZ[1], İlkay DOĞAN[1], Seval KUL[1], Neslihan BAYRAMOĞLU TEPE[2]

[1]Gaziantep Üniversitesi, Tıp Fakültesi, Biyoistatistik Anabilim Dalı, Gaziantep
[2]Gaziantep Üniversitesi, Tıp Fakültesi, Kadın Hastalıkları ve Doğum Anabilim Dalı, Gaziantep

Amaç: Sınıflama performansı testleri yardımı ile ikiden fazla tanı grubunu ayırt etmek ve ayırt etme performanslarını karşılaştırmak amaçlanmıştır.

Yöntem: Gaziantep Üniversitesi Kadın Doğum Anabilim dalı tarafından değerlendirilen 41 hastaya ait veriler kullanılmıştır. Hastalar perkreta, inkreta, akreta olmak üzere üç gruba ayrılmıştır. Ölçülen YKL-40 değerinin belirlenen grupları ayırt etme performansı üç yönlü ROC (Receiver Operating Characteristic) analizi, LASSO, kNN (k Nearest Neighborhood), Random Forest, SVM (Support Vector Machine), Naive Bayes, AdaBoost yöntemleri ile karşılaştırılmalı olarak incelenmiştir. Parametre tahminlerinde, örnek sayısının azlığı ve grupların dengesizliği nedeni ile one-leave out crosvalidation yöntemi kullanılmıştır. Verinin %50si modeli oluşturmak için diğer %50 si ise modelin validasyonu için kullanmıştır. Analizler R 3.6.1 paket programı (TrinROC paketi) ve Orange Canvas 3.22 programı ile gerçekleştirilmiştir.

Bulgular: Yapılan analizler sonucunda test verisi için Random Forest, kNN, AdaBoost, Naive Bayes, SVM, LASSO yöntemleriyle AUC (Area Under the Curve) değerleri sırasıyla 0,803; 0,782; 0,775; 0,718; 0,668; 0,613 bulunmuştur. Validasyon verisi için ise AUC değerleri sırasıyla 0,602; 0,595; 0,519; 0,490; 0,417; 0,461 olarak belirlenmiştir. Üç yönlü ROC analizi sonucunda ise VUS (Volume Under the ROC surface) değeri tüm verilerden yapılan analiz sonucunda 0,669 olarak bulunmuştur. Bu değer %50lik alt gruplara ayrıldığında 0,011 VUS değerine düşmüştür. Buna ek olarak üç yönlü ROC analizinde kesim noktaları c_1=30,98 ve c_2=37,98 olarak hesaplanmıştır.

Sonuç: Farklı yöntemlerle elde edilen sınıflama performansları dikkate alındığında tüm yöntemlerin validasyon verisinde ayırma performansının düştüğü, Random Forest yönteminin ayırma performansının en iyi olduğu ve üç yönlü ROC analizinin örnek hacminden çok etkilendiğini göstermektedir. Elde edilen bu sonuçlar bu veri setine ilişkin olmakla birlikte bu yöntemlerin sınıflandırma performansı simülasyon çalışmalarıyla desteklenmelidir.

Anahtar kelimeler: Sınıflandırma Performans testleri, AUC , VUS, TrinROC, Üç Yönlü ROC Analizi

S37

Comparison of Classification Performance Tests with an Application in Multiple Classification

Tanyeli GÜNEYLİGİL KAZAZ[1], İlkay DOĞAN[1], Seval KUL[1], Neslihan BAYRAMOĞLU TEPE[2]

[1]Gaziantep University, Faculty of Medicine, Department of Biostatistics, Gaziantep
[2] Gaziantep University, Faculty of Medicine, Department of Obstetrics and Gynecology, Gaziantep

Aim: With the help of classification performance tests, it is aimed to separate more than two diagnostic groups and compare their performance.

Method: The data of 41 patients evaluated by Gaziantep University Obstetrics Department were used. The patients were divided into three groups as percreta, increta and acreta. The performance of the measured YKL-40 value to separate the determined groups were investigated with three-way ROC (Receiver Operating Characteristic) analysis, LASSO, kNN (k Nearest Neighborhood), Random Forest, SVM (Support Vector Machine), Naive Bayes, AdaBoost methods. One-leave out cross-validation method was used in parameter estimations due to the small sample size and unbalance of the groups. Half of the data was used to create the model and the other 50% was used to validate the model. Analyzes were performed with R 3.6.1 package program (TrinROC package) and Orange Canvas 3.22 program.

Results: As a result of the analyses, for test dataset AUC (Area Under the Curve) values of Random Forest, kNN, AdaBoost, Naive Bayes, SVM, LASSO methods were found 0.803; 0.782; 0.775; 0.718; 0.668; 0,613; respectively. Moreover, for validation data AUC values were found 0,602; 0,595; 0,519; 0,490; 0,417; 0,461; respectively. VUS (Volume Under the ROC surface) value was found to be 0,669 as a result of three-way ROC analysis from all data. This value is divided into sub-groups of 50% has decreased to VUS=0,011. In addition, the cut-off points in the three-way ROC analysis were calculated as c_1 = 30,98 and c_2 = 37,98.

Conclusion: When the classification performances obtained by different methods are taken into consideration, it is seen that the separation performance was decreased in validation dataset for all methods, the separation performance of Random Forest method was the best and the three-way ROC analysis was highly affected from sample size. Although these results are related to this data set, classification performance of these methods should be supported by simulation studies.

Key words: Classification Performance Tests, AUC, VUS, TrinROC, Three Way ROC Analysis

S38

An Interactive R Shiny Application to Facilitate Comparing Classification Methods: ClasComp

Uğur TOPRAK[1], Beyza DOĞANAY ERDOĞAN[2], Esin ÖĞüŞ[1]

[1]Başkent University Faculty of Medicine, Department of Biostatistics
[2]Ankara University Faculty of Medicine, Department of Biostatistics

Aim: Classification is the issue of defining, on the grounds of observations, which a new observation belongs to a set of categories. This basic and widely used technique forms the basis of decision making processes. In our study we aim to assemble powerful classification algorithms under an R Shiny application in order to perform all in one solutions for who deals with mostly commonly such cases. Since Shiny apps can be deployed on the Web, end users with zero knowledge of R can use them.

Method: Recently, due to the fast pace progression in data science, one can find numerous classification techniques offered under many platforms. We synthesis the widely used classification algorithms under the same roof by utilizing R Shiny. Such as Random Forest, Neural Networks, glmnet, Linear Discriminant Analysis (LDA) and eXtreme Gradient Boosting (XGBoost). We divide the data analysis process into three parts; exploratory data analysis, fitting and validation and model interpretation.

Exploratory data analysis section offers uploading your own data, screening data structure, investigating missing values and correlated variables. In addition, we allow users to clean and transform their data according to ReCiPe method where the primary object is to transform the long lists into a limited number of indicator scores.

Fitting and validation section starts with splitting the data into training and validation set. In the background, each method performs their specific process simultaneously.

Model interpretation section has not yet been fully completed, however we expect to build a module for reporting model comparisons in structured format.

Results: Having defined what are the key concepts of our application, fitting and validation part holds significant importance in the analysis process. This section can best be treated under two divisions: Model training and model validation. Model training illustrates numerous information such as performance statistics of the models given both graphically and numerically, investigating model specifications individually and short feedback on training results. Model validation, also covers variable importance plots for each method in addition to model training information.

Conclusion: The present study was designed to show how our application offers resourceful solutions for time consuming problems. The most obvious argument to emerge from this study is that contributing the softwares which are used in classification by providing a compact user interface, ease of access via web and comparing results instantly. Further improvements still need to be carried out in order to report model interpretations in a well-established format.

Keywords: Classification, Shiny, Machine Learning

S39

2×2 Boyutlu Simetrik Dengesiz Tasarımlarda Değerlendiriciler Arası Uyumun Değerlendirilmesi

H. Yağmur ZENGİN[1], Tuğçe ŞENÇELİKEL[2], Ersin ÖĞÜŞ[2]

[1]Hacettepe Üniversitesi Biyoistatistik Anabilim Dalı,
[2]Başkent Üniversitesi Biyoistatistik Anabilim Dalı

Cohen'in Kappa katsayısı, iki değerlendirici arasındaki uyumun belirlenmesinde ölçümler sınıflanabilir nitel türde olduğunda yaygın kullanılan bir uyum katsayısıdır. Ancak, bazı durumlarda Cohen'in Kappa katsayısı paradoks olarak da adlandırılan hatalı sonuçlar üretmektedir. Bu paradokslardan biri, marjinal toplamların neredeyse mükemmel simetrik ve dengesiz olduğu durumlarda, değerlendiriciler arası yüksek uyum söz konusu olmasına rağmen düşük Kappa değerlerinin gözlenmesi biçiminde ortaya çıkar. Diğeri ise marjinal toplamlardaki simetrik dengesizliğin azaldığı ya da asimetrik olduğu durumlarda değerlendiriciler arası uyum düşük olmasına rağmen yüksek Kappa değerlerinin elde edilmesi şeklinde tanımlanmıştır. 1980-1990 yılları arasında birçok araştırmacı bu paradoksları teorik olarak irdelemiştir. Günümüzde ise bu paradokslara karşı bağışık uyum katsayılarının geliştirilmesi amacıyla çalışmalar hala devam etmektedir.

Amaç: Bu çalışmada, 2×2 boyutlu simetrik dengesiz tasarımlarda Cohen'in Kappa katsayısının ve değerlendirici ile değerlendirme sınıf sayısı iki iken kullanılan diğer bazı uyum katsayılarının gerçek uyumu tahmin performansları benzetim çalışması ile araştırılmıştır. Bu kapsamda, araştırmacılara farklı senaryolar altında bu tür bir tasarım ile ilişkili paradoksa karşı güçlü alternative katsayı önerileri sunulması amaçlanmıştır.

Yöntem: R'da uygulanan benzetim çalışmasında, örnek genişliklerinin n=50, 100, 200, 500, 1000; gerçek uyumun 0,10, 0,60, 0,90 ve değerlendiricilerin marjinal olasılıklarının eşit olarak 0,50 ile simetrik dengesizlik için 0,60, 0,75, 0,95 olarak alındığı farklı senaryolarda, Cohen'in Kappa katsayısı ve ona alternatif olarak kullanılan Gwet'in AC1, Scott'ın Pi, Brennan-Prediger Kappa, Krippendorf'un Alfa, Bennet ve arkadaşlarının Sigma ve Goodman-Kruskal Lamda katsayılarının gerçek uyumu tahmin etmedeki başarısı 10.000 tekrar sayısında Ortalama Hata Kareler (OHK) hesaplanarak incelenmiştir.

Bulgular: Örnek genişliğindeki artış tüm katsayılarda OHK'de azalmayı sağlarken; örnek genişliği 200'ün altında ve gerçek uyum yüksek olduğunda Sigma ve Brennan-Prediger'in Kappa katsayısı en iyi performansı gösterirken ardından Gwet AC1 gelmiştir ancak sırasıyla Krippendorf'un Alpha, Cohen'in Kappa, Scot'ın Pi katsayılarında yüksek Ortalama Hata Kareler elde edilmiştir. En kötü performansı Goodman-Kruskal Lamda göstermiştir. Ancak, marjinal olasılıkların 0,75 ve üstünde fakat gerçek uyumun orta ve düşük olduğu durumda örnek genişliğinin artması ile azalmasına rağmen S katsayısı, Brennan-Prediger'in kappa katsayısı, Gwet AC1 gibi katsayıların daha yüksek Ortalama Hata Kare değerlerine sahip olduğu görülmüştür. Tasarım dengeli hale geldiğinde tüm katsayılar birbirine yakın sonuçlar vermesine rağmen tüm senaryolarda en iyi performansı sırasıyla Gwet AC1 ile Brennan-Prediger Kappa ve Sigma katsayısı göstermiştir.

Sonuç: Benzetim çalışmasında görülmüştür ki örnek genişliğinin düşük olduğu simetrik dengesiz çalışmalarda gerçekte uyumun yüksek olduğu düşünülüyor ise S ve Brennan-Prediger'in kappa katsayıları; uyumun orta veya düşük düzeyde olması bekleniyor ise Krippendorf alfa; dengeli tasarımlarda ise genel olarak Gwet AC1'in kullanılması önerilmektedir.

Anahtar Kelimeler: Cohen'in Kappa Katsayısı, Paradoks, Uyum

S40

Konkordans Korelasyon Katsayısı Temelli Yeni Bir Normallik Testi

Osman DAĞ[1], N. Anıl DOLGUN[2]

[1] Biyoistatistik Anabilim Dalı, Hacettepe Üniversitesi, Ankara, Türkiye;
[2] Fen Fakültesi, RMIT Üniversitesi, Melbourne, Avustralya.

Amaç: Bu çalışmadaki öncelikli amaç konkordans korelasyon katsayısı (KKK) temelli yeni bir normallik testi önermektir.

Yöntem: Lin (1989), KKK'yi orjinden 45 derecelik açı ile geçen çizgiden değişimi ölçerek iki ölçüm arasındaki uyumu değerlendiren bir ölçüm olarak önermiştir. Bu çalışmada, gözlenen değerlerlerle yokluk hipotezi altında beklenen değerlerlerin ne kadar uyuşup uyuşmadığını incelenerek, KKK'nin normalliği test etmek için kullanılabileceği gösterilmiştir.

Bulgular: KKK'nin yokluk hipotezi altında dağılımı incelenmiştir ve test geliştirildikten sonra Monte Carlo benzetim çalışmasıyla diğer normallik testleri ile karşılaştırılmıştır. Tip I hata değerleri incelendiğinde, tüm örneklem genişliklerinde önerilen normallik testi, Alexander-Darling test ve Cramer-von Mises testi nominal seviyesini yakalamışlardır. Örneklem genişliği arttıkça, Shapiro-Wilk ve Kolmogorov-Smirnov testleri daha tutucu olmaktadır. Örneklem genişliği azaldıkça, Jarque-Bera test tutucu davranmaktadır. Düşük örneklem büyüklüğünde, Pearson ki-kare testi için tip I hatası nominal seviyesinden yüksek çıkmaktadır. Tüm örneklem genişliklerinde Shapiro-Francia testin tip I hatası nominal seviyesinden yüksek olarak bulunmuştur. Güç değerleri incelendiğinde, veri lojistik, student t ve laplace dağılımlarından üretildiğinde, önerilen normallik testi ile Shapiro-Francia testi daha iyi performans göstermektedir. Shapiro-Francia testin tip I hatası nominal seviyesinden yüksek olması nedeniyle, önerilen testten biraz daha yüksek güce sahiptir. Veri tekdüze dağılımdan üretildiğinde, en güçlü test Shapiro-Wilk testidir.

Sonuç: Bu çalışmada, KKK temelli yeni bir normallik testi önerilmektedir. Önerilen normallik testi ile diğer çok bilinen testler Monte-Carlo beenzetim çalışmasıyla karşılaştırılmıştır. Özetle, önerdiğimiz test çoğu senaryoda diğer normallik testlerinden iyi performans gösterdiği için, önerilen testimizin pratikte kullanımı faydalı görünmektedir.

Anahtar Kelimeler: Konkordans Korelasyon Katsayısı, Normallik Testi, Uyum İyiliği Testi.

References

1. Lin, L. I. (1989). A concordance correlation coefficient to evaluate reproducibility. Biometrics, 45, 255-268.

S40

A Novel Test of Normality Based on the Concordance Correlation Coefficient

Osman DAG[1], Anıl DOLGUN[2]

[1] Department of Biostatistics, Hacettepe University, Ankara, Turkey;
[2] School of Science, RMIT University, Melbourne, Australia.

Aim: In this study, our primary focus was to propose a novel normality test based on the concordance correlation coefficient (CCC).

Method: Lin (1989) introduced the CCC as a reproducibility index, which evaluates the agreement between two measurements by measuring the variation from the 45º line through the origin. In this paper, we showed that this statistic can be used to test the normality, by examining how well the observed values concord with the hypothesized (expected) ones under the null hypothesis.

Results: The null distribution properties of CCC is also investigated and compared with its well-known counterparts via Monte Carlo simulation. When the type I errors are examined; our proposed normality test, Anderson-Darling test and Cramer-von Mises test hold the nominal size for all sample sizes. Shapiro-Wilk and Kolmogorov-Smirnov tests get conservative as the sample size gets larger. Jarque-Bera test gets conservative as the sample size gets smaller. Pearson chi-square test has inflated type I error for small sample size. Shapiro-Francia test has inflated type I error for all sample sizes. When the powers are examined; our proposed test and Shapiro-Francia test perform better to detect non-normality when the data generated from lojistic, student t, laplace distributions. Due to the inflation of type I error of Shapiro-Francia test, it has little higher power than our proposed test. Shapiro-Wilk is the most powerful one to detect non-normality when the data generated from uniform distribution.

Conclusion: In this study, we proposed a new normality test based on the CCC. We compared our proposed test to other well-known normality tests. To sum up, the use of our proposed test seems to be beneficial, since it performs well under most scenarios and takes advantage from other normality tests.

Keywords: Concordance Correlation Coefficient, Normality Test, Goodness-of-Fit Test.

References

1. Lin, L. I. (1989). A concordance correlation coefficient to evaluate reproducibility. Biometrics, 45, 255-268.

S41

Polikistik Over Sendromu Olan Kadınlarda Nesfatin Düzeyleri: Bir Meta Analiz Çalışması

Burçin ALTINBAŞ[1], Pınar GÜNEL KARADENİZ[2]

[1]SANKO Üniversitesi, Tıp Fakültesi, Fizyoloji Anabilim Dalı, Gaziantep
[2]SANKO Üniversitesi, Tıp Fakültesi, Biyoistatistik Anabilim Dalı, Gaziantep

Amaç: Bu çalışmanın amacı, polikistik over sendromu olan kadınlarda nesfatin-1 düzeylerinin sağlıklı kadınlardan farklı olup olmadığını meta analiz ile ortaya koymaktır.

Yöntem: Pubmed ve Web of Science veri tabanları **"**polycystic over syndrom" ve "nesfatin" anahtar kelimeleri kullanılarak taranmış ve toplamda 7 makaleye ulaşılmıştır. Bu makalelerin biri sıçanlarda yapıldığından biri de ilgili istatistiği raporlamadığından analiz dışında bırakılmıştır. Meta analiz, R programında "metafor" paketi kullanılarak yürütülmüştür. Heterojenlik I^2 istatistiği ve Q testi ile değerlendirilmiş ve rasgele etkiler meta analizi uygulanarak genel etki büyüklüğü standartlaştırılmış ortalamalar farkı hesaplanmıştır. Çalışmalara ait etki büyüklükleri görsel olarak forest grafiği ile gösterilmiştir. Yayın yanlılığı huni (funnel) grafiği ile değerlendirilmiştir.

Bulgular: Toplam 5 çalışmadan elde edilen etki büyüklükleri -1.56 ile 41.75 arasında değişmektedir. Yürütülen meta analiz sonucunda I^2 istatistiği %99.9 ve Q testi için $p<0.001$ olarak elde edilmiştir. Bulgular heterojenliğin varlığını gösterdiğinden rasgele etkiler meta analizi uygulanarak genel etki büyüklüğü 7.81 (%95 Güven Aralığı: -8.45 ile 24.06) olarak hesaplanmıştır. Bu 5 çalışmadan birinin etki büyüklüğü aşırı değer olarak kabul edilmiş ve meta analiz bunun dışında kalan 4 çalışma üzerinden tekrar yürütülmüştür. İkinci analizde I^2 istatistiği %95 ve Q testi için $p<0.001$ olarak elde edilmiştir. Rasgele etkiler meta analizi sonucunda genel etki büyüklüğü -0.41 (%95 Güven Aralığı:-1.43 ile 0.60) olarak bulunmuştur. Her iki meta analiz sonucuna göre etki büyüklüğü için güven aralıkları sıfırı içerdiğinden etkinin anlamlı olmadığı sonucuna ulaşılmıştır ($p>0.05$).

Sonuç: Meta analiz sonucunda elde edilen bulgular doğrultusunda nesfatin düzeyinin polikistik over sendromuna sahip kadınlarda sağlıklı kadınlara göre farklı olduğunun bir kanıtı elde edilememiştir. Literatürde bu konuda yapılan çalışma sayısı az olduğundan meta analize dâhil edilen çalışma sayısının azlığı bir kısıt oluşturmaktadır. Bu konuda yeni çalışmalar yapıldıkça, daha fazla sayıda çalışmanın dahil edildiği bir meta analizin daha güvenilir sonuçlar vereceği düşünülmektedir.

Anahtar Kelimeler: Polikistik over sendromu, nesfatin düzeyi, meta analiz

S41

Nesfatin Levels in Women with Polycystic Ovary Syndrome: A Meta-Analysis

Burçin ALTINBAŞ[1], Pınar GÜNEL KARADENİZ[2]

[1] SANKO University, Faculty of Medicine, Department of Physiology, Gaziantep
[2] SANKO University, Faculty of Medicine, Department of Biostatistics, Gaziantep

Aim: The aim of this study was to determine whether nesfatin-1 levels were different in women with polycystic ovary syndrome from healthy women via meta-analysis.

Method: PubMed and Web of Science databases were searched by using keywords “polycystic over syndrome” and “nesfatin” and a total of seven articles were reached. Two of these articles were excluded from the analysis as one was carried out in rats and one did not report the relevant statistics. Meta-analysis was performed using the “metafor” package in the R program. Heterogeneity was evaluated by I^2 statistic and Q test and random effects meta-analysis was applied to calculate the overall effect size, standardized mean difference. The effect sizes of the studies are presented visually by using forest plot. The publication bias was evaluated via funnel plot.

Results: The effect sizes obtained from a total of five studies ranged from -1.56 to 41.75. As a result of the meta-analysis, I^2 statistic was 99.9% and $p < 0.001$ for Q test. Since the findings represent the presence of heterogeneity, random effects meta-analysis was carried out and the overall effect size was found as 7.81 (95% Confidence Interval: -8.45 and 24.06). The effect size of one of these five studies was accepted as outlier and the meta-analysis was carried out again including the remaining four studies. In the second analysis, I^2 statistic was 95% and $p < 0.001$ for Q test. As a result of random effects meta-analysis, overall effect size was -0.41 (95% Confidence Interval: -1.43 and 0.60). According to both meta-analysis results, it was concluded that the effect was not significant because the confidence intervals included zero for the effect sizes ($p > 0.05$).

Conclusion: Based on the findings obtained from the meta-analyses, there was no evidence that nesfatin levels are different in women with polycystic ovary syndrome compared to healthy women. Since there is few studies on this subject in the literature, including small number of studies into meta-analysis is a limitation of our study. As far as further studies are conducted on this subject it is considered that a meta-analysis with more studies will provide more reliable results.

Keywords: Polycystic ovary syndrome, nesfatin levels, meta-analysis

S42

Kesintiye Uğramış Zaman Serilerinin Analizi: Segmented Regresyon ya da Joinpoint Regresyon

Pınar GÜNEL KARADENİZ[1], Tanyeli GÜNEYLİGİL KAZAZ[2]

[1]SANKO Üniversitesi, Tıp Fakültesi, Biyoistatistik Anabilim Dalı, Gaziantep
[2]Gaziantep Üniversitesi, Tıp Fakültesi, Biyoistatistik Anabilim Dalı, Gaziantep

Amaç: Bu çalışmanın amacı, kesintiye uğramış zaman serilerinin analizinde kullanılan segmented (parçalara ayrılmış, bölünmüş) regresyon analizinin ve joinpoint regresyon analizinin örnek bir veri seti üzerinde uygulamasını göstermek ve bu iki modeli tanıtmaktır.

Yöntem: Mide kanserine bağlı ölüm oranlarının zamana bağlı değişimini değerlendirmek için segmented regresyon analizi ve joinpoint regresyon analizi uygulanmıştır. Veriler, Türkiye İstatistik Kurumunun 2007-2018 yılları arasındaki veri tabanından elde edilmiştir. Segmented regresyon analizi, R yazılımında "Segmented" paketi kullanılarak; joinpoint regresyon analizi ise Joinpoint Regression Program, 4.7.0.0 versiyonu kullanılarak yürütülmüştür.

Bulgular: Hem segmented regresyon analizi ile hem de joinpoint regresyon analizi ile yıllara bağlı ölüm oranları trendinde değişiklik gözlenen nokta 2014 yılı olmuştur. Segmented regresyon analizi ile trenddeki bu değişim Davies testi sonucu istatistiksel olarak anlamlı bulunmuştur ($p<0.001$). Joinpoint regresyon analizi ile de benzer şekilde kırılımın olduğu bu nokta istatistiksel olarak anlamlı bulunmuştur ($p=0.020$). Joinpoint regresyon analizi ile 2007-2014 arasındaki yıllık yüzde değişim %5.4 (%95 Güven Aralığı: 3.3 ile 7.7) ve istatistiksel olarak anlamlı bulunmuş; 2014-2018 arasındaki yıllık yüzde değişim ise %-2.2 (%95 Güven Aralığı: -6.9 ile 2.7) elde edilmiş fakat istatistiksel olarak anlamlı bulunmamıştır. Ayrıca 2007-2018 arası ortalama yıllık yüzde değişim incelendğinde %2.6 (%95 Güven Aralığı: 0.7 ile 4.5) ve istatistiksel olarak anlamlı bulunmuştur.

Sonuç: Çalışma bulgularından yola çıkarak bu değişimin sebepleri araştırılabilir ve toplum sağlığını ilgilendiren çıkarımlar yapılabilir. Bu çalışma ile kesintiye uğramış zaman serileri için uygun olan iki modelin uygulaması örnek bir veri seti üzerinde gösterilmiştir. Ancak model performanslarının karşılaştırılmasına yönelik simülasyon çalışmalarının yürütülmesi önerilmektedir.

Anahtar Kelimeler: kesintiye uğramış zaman serileri, segmented regresyon, joinpoint regresyon

S42

Interrupted Time Series Analysis: Segmented Regression or Joinpoint Regression

Pınar GÜNEL KARADENİZ[1], Tanyeli GÜNEYLİGİL KAZAZ[2]

[1]SANKO University, Faculty of Medicine, Department of Biostatistics, Gaziantep
[2]Gaziantep University, Faculty of Medicine, Department of Biostatistics, Gaziantep

Aim: The aim of this study was to demonstrate the application of segmented regression analysis and joinpoint regression analysis used for the analysis of interrupted time series on a sample data set and to introduce these two models.

Method: Segmented regression analysis and joinpoint regression analysis were performed to evaluate the time-dependent changes in gastric cancer mortality. Data were obtained from the Turkish Statistical Institute database between the years 2007 and 2018. Segmented regression analysis was carried out by using "Segmented" package in R software; the joinpoint regression analysis was carried out by using the Joinpoint Regression Program, version 4.7.0.0.

Results: Both segmented regression analysis and joinpoint regression analysis revealed that the trend for mortality has changed in 2014. In segmented regression analysis change in trend was found statistically significant by the Davies test ($p < 0.001$). Similarly, this period was found to be statistically significant by joinpoint regression analysis ($p = 0.020$). In joinpoint regression analysis, the annual percent change between 2007 and 2014 was 5.4% (95% Confidence Interval: 3.3 to 7.7) and statistically significant; the annual percent change between 2014-2018 was -2.2% (95% Confidence Interval: -6.9 to 2.7) but was not statistically significant. In addition, when the average annual percent change between 2007-2018 was examined, it is found to be 2.6% (95% Confidence Interval: 0.7 to 4.5) and statistically significant.

Conclusion: Based on the findings of the study, the causes of this change can be investigated and inferences concerning public health can be made. In this study, the application of two models suitable for interrupted time series was demonstrated on a sample data set. However, it is recommended that simulation studies should be conducted to compare the model performances.

Keywords: Interrupted time series, segmented regression, joinpoint regression.

S43

Acil Servislerdeki Çocuk Alt Solunum Yolu Enfeksiyonları: Bibliyometrik Analiz

Hadiye DEMİRBAKAN[1], Demet ARI YILMAZ[2], Pınar GÜNEL KARADENİZ[3], Vildan SÜMBÜLOĞLU[3]

[1]SANKO Üniversitesi, Tıp Fakültesi, Tıbbi Mikrobiyoloji Anabilim Dalı, Gaziantep
[2]SANKO Üniversitesi, Tıp Fakültesi, Acil Tıp Anabilim Dalı, Gaziantep
[3]SANKO Üniversitesi, Tıp Fakültesi, Biyoistatistik Anabilim Dalı, Gaziantep

Amaç: Bu çalışmada, acil servislerdeki çocuk alt solunum yolu enfeksiyonlarını konu alan araştırmaların literatür taramasını yaparak üzerinde en çok çalışılan konuların belirlenmesi ve bu konuda çalışma yapacak araştırmacılara yol gösterecek, literatürü özetleyen bilgiler sunulması amaçlanmıştır.

Yöntem: Bibliyometrik analiz; kitaplar, web siteleri, bildiri kitapları için kullanılabilen bir literatür tarama ve özetleme yöntemidir. Sağlık alanında çoğunlukla bir araştırma makalesinin etkisini ölçmek için kullanılır. Bibliyometrik yöntemler bir yayının kendinden sonraki araştırmalar üzerinde ne kadar etki yaptığını, alanına ne kadar katkıda bulunduğunu değerlendirir. Bu değerlendirme, genellikle araştırmanın yayınlandıktan sonra ne kadar atıf aldığının belirlenmesi ile yapılır. Bu çalışmada, Web of Science veri tabanında yer alan veriler kullanılmıştır. "lower respiratory", "emergency", "pediatric", "child/children" anahtar kelimeleri kullanılarak elde edilen makaleler bir Acil Tıp uzmanı ve bir Tıbbi Mikrobiyoloji uzmanı tarafından bağımsız şekilde incelenmiş ve konu ile ilgili en çok atıf alan 50 makale belirlenmiştir. Bu makalelerin; toplam atıf sayıları, yıl başına ortalama atıf sayıları, yıllara göre dağılımı, dergi dağılımı, makale türü, yazar sayısı ve ülke dağılımı bilgileri raporlanmıştır.

Bulgular: Acil servislerdeki çocuk alt solunum yolu enfeksiyonlarını konu alan ve en çok atıf alan ilk 50 makalenin atıf sayıları 23-964 arasında değişiyordu. Bu konuda en yüksek atıf alan makalelerin 6'sı (% 12) 2014 yılında yayınlanmıştır. Toplam atıf sayısı 100'ün üzerinde olan 6 (% 12) makale; yıl başına ortalama atıf sayısı 10'un üzerinde olan 10 (% 20) makale bulunuyordu. 46 (% 92) makale orijinal araştırma idi. Yarısına yakını (% 42) 6'dan az yazarla yürütülmüş ve (% 44) pediatri dergilerinde yayınlanmıştı. Makalelerin 26'sında Amerika Birleşik Devletleri'nden yazarlar yer alıyordu.

Sonuç: Bu çalışmada, acil servislerdeki çocuk alt solunum yolu enfeksiyonlarını konu alan araştırmaların bibliyometrik analizleri yapılarak bu konuda araştırma yapacak olan araştırmacılara yön gösterecek önemli bilgiler ortaya konulmuştur.

S43

Pediatric Lower Respiratory Tract Infections in Emergency Departments: Bibliometric Analysis

Hadiye DEMİRBAKAN[1], Demet ARI YILMAZ[2], Pınar GÜNEL KARADENİZ[3], Vildan SÜMBÜLOĞLU[3]

[1]SANKO University, Faculty of Medicine, Department of Medical Microbiology, Gaziantep
[2]SANKO University, Faculty of Medicine, Department of Emergency Medicine, Gaziantep
[3]SANKO University, Faculty of Medicine, Department of Biostatistics, Gaziantep

Aim: In this study, it is aimed to determine the most studied topics on children's lower respiratory tract infections in emergency departments in the literature and to provide information which summarizes the literature and guides the researchers studied in this field.

Method: Bibliometric analysis is a literature search and summarization method that can be used for books, web sites and papers. In the health field, bibliometrics are mostly used to measure the influence or impact of research articles. Bibliometric methods evaluate how much a publication has impact on subsequent researches and how much it contributes to the relevant field. This evaluation is usually done by counting the number of times the article is cited after it is published. In this study, the data in Web of Science database was used. The articles obtained using the keywords “lower respiratory”, “emergency”, “pediatric”, “child / children” were examined independently by an Emergency Medicine specialist and a Medical Microbiologist and the 50 most cited articles were identified. The total number of citations, average number of citations per year, distribution by years, journal distribution, article type, number of authors and country distribution of these articles were reported.

Results: Citation numbers of the first 50 most frequently cited articles on pediatric lower respiratory tract infections in emergency departments ranged from 23 to 964. Six (12 %) of the most cited articles were published in 2014. There were 6 (12 %) articles with total citations exceeding 100 and 10 (20 %) articles with an average of over 10 citations per year. 46 (92 %) articles were original research. Nearly half of the articles (42 %) were published by less than 6 authors and (44 %) were published in pediatric journals. 26 of the articles included authors from the United States.

Conclusion: In this study, bibliometric analysis of the studies on pediatric lower respiratory tract infections in emergency departments was carried out and important information was provided to the researchers who will conduct research on this subject.

S44

Genom-Boyu İlişki Çalışmalarında Makine Öğrenimi Yöntemleri ve Derin Öğrenme Yönteminin Farklı Örnek Genişliklerinde Performanslarının Değerlendirilmesi

Ragıp Onur ÖZTORNACI[1], Erdal COŞGUN[2], Bahar TAŞDELEN[1]

[1] Mersin Üniversitesi Tıp Fakültesi, Biyoistatistik ve Tıbbi Bilişim Anabilim Dalı, Mersin
[2] Microsoft Research, 14820 NE 36th Street, Building 99, Redmond, WA, 98052, USA

Amaç: Makine öğrenimi (ML) ve derin öğrenme (DL) yöntemlerinin genetik ilişki çalışmalarında (GWAS) kullanımı giderek yaygınlaşmaktadır. Bu yöntemler, hem klasik GWAS metodolojisine ek olarak elde edilen sonuçların doğrulanmasına yardımcı olurken hem de GWAS prosedürü takip edilirken kullanılan yüksek eşit değeri nedeniyle göz ardı edilebilen hastalıkla ilişkili tek nükleotid polimorfizmlerinin (TNP) yakalanmasına yarar sağlayabilir. Bu nedenle; bu çalışmanın amacı, en yaygın kullanılan ML yöntemleri ve DL yönteminin farklı örnek genişliklerinde kıyaslanarak en doğru hasta-kontrol ayırımını yapan sınıflama yöntemini bulmaktır.

Yöntem: ML ve DL yöntemleri için farklı örnek genişliklerinde aynı prevelansa ve minör allel frekansına sahip ve eşit sayıda hasta-kontrol içeren veriler üretilmiştir. Tüm veri setlerinde, tüm yöntemler için 10-katlı çapraz geçerlilik yöntemi uygulandı, tüm yöntemler için, %70 eğitim veri seti %30 test veri seti olarak belirlenerek modellerin geçerlilikleri ve tahmin güçleri sınandı. Simülasyonlar için PLINK yazılımı, ML için R programlama dili ve DL için de Python programlama dili Microsoft Azure GPU makinalar ile kullanmıştır.

Bulgular: Örnek genişliği; N=200 için; en iyi sınıflama performanslarını ML yöntemleri içinde, Random Forest (RF) 0.73 Doğruluk (Acc.) (%95 GA [0.60;0.83], 0.73 Sensitivite, 0.73 Spesifite) ve CART 0.73 Acc. (%95 GA [0.60;0.83], 0.70 Sensitivite, 0.76 Spesifite) ile gösterirken, DL 0.60 Acc. (%95 GA [0.52;0.64], 0.65 Sensitivite, 0.57 Spesifite) değeri elde etmiştir. N=600 için; ML yöntemlerinden DVM 0.78 Acc. (%95 GA [0.72;0.84], 0.84 Sensitivite, 0.73 Spesifite) ile en iyi sınıflama başarısını elde ederken, DL 0.64 Acc. (%95 GA [0.62;0.63], 0.60 Sensitivite, 0.57 Spesifite) değeri elde etmiştir. Örnek genişliği maksimum olan 1000'e çıkartıldığında, hem ML hem de DL için doğru sınıflama başarısının değişiminin etkilenmediği söylenebilir; ML içinde, Destek vektör makineleri (DVM) 0.75 Acc. (%95 GA [0.69;0.79], 0.74 Sensitivite, 0.75 Spesifite) değeri elde ederken, DL 0.64 (%95 GA [0.62;0.65], 0.58 Sensitivite, 0.60 Spesifite) elde ettiği görülmüştür.

Sonuç: ML tabanlı çalışmalarda en çok merak edilen konu yüksek boyutlu verilerin olduğu çalışmalarda hangi yöntemin daha iyi tahminlime yapacağıdır. Bu tip karşılaştırma çalışmalarının en zayıf noktası gerçek anlamda büyük veri uygulamalarının yapılamamasıdır. Bu çalışmada genetik araştırmalarda en sık kullanılan GWAS veri seti ile artık klasikleşen ML yöntemleri ve DL yöntemleri karşılaştırılmıştır. Hazırlanan deney düzeni genetik varyasyon verilerinin ilişkilerini, örneklem genişliklerini göz önüne alınmıştır. Son yıllarda özellikle bağımsız değişkenler arasında ilişkilerin çok yüksek olduğu veri setlerinde yüksek tahminlime oranı DVM yöntemi edilmektedir. Bizim çalışmamızda örneklem genişliğinden bağımsız olarak en yüksek sonuç DVM ile elde edilmiştir. Bu da genetik verilerde sık karşılaşılan doğrusal olmayan ML yöntemlerinin başarısını kanıtlamaktadır.

Anahtar Kelimeler: Sınıflandırma, Tek nükleotid polimorfizmi, Makine Öğrenimi, Derin Öğrenme.

Referanslar

1. [1] Szymczak, S., Biernacka, J. M., Cordell, H. J., González-Recio, O., König, I. R., Zhang, H., & Sun, Y. V. (2009). Machine learning in genome-wide association studies. *Genetic epidemiology*, *33*(S1), S51-S57.
2. [2] Teo, Y. Y. (2008). Common statistical issues in genome-wide association studies: a review on power, data quality control, genotype calling and population structure. *Current opinion in lipidology*, *19*(2), 133-143.
3. [3] Hong, E. P., & Park, J. W. (2012). Sample size and statistical power calculation in genetic association studies. *Genomics & informatics*, *10*(2), 117.

S44

Evaluation of Machine Learning Methods and Deep Learning Method Performance in Different Sample Size in Genome-Wide Association Studies

Ragıp Onur ÖZTORNACI[1], Erdal COŞGUN[2], Bahar TAŞDELEN[1]

[1]Mersin University Faculty of Medicine, Department of Biostatistics and Medical Informatics, Mersin, Turkey
[2] Microsoft Research, 14820 NE 36th Street, Building 99, Redmond, WA, 98052, USA

Aim: The use of machine learning (ML) and deep learning (DL) methods in genetic relationship studies (GWAS) are becoming increasingly common. These methods can both help confirm the results obtained in addition to the classical GWAS methodology, and can also help capture single-nucleotide polymorphisms (TNP) that can be ignored due to the high equal value used when following the GWAS procedure [1,2]. Therefore; The aim of this study was to compare the most commonly used ML methods and DL methods at different sample size and find the most accurate case-control classification method.

Method: For ML and DL methods, data with the same prevalence and minor allele frequency at different sample size and containing an equal number of patient-controls were produced [3]. In all data sets, 10-fold cross validation method was applied for all methods, 70% training data set was determined as 30% test data set for all methods and validity and predictive power of the models were tested. PLINK software for simulations, R programming language for ML and Python programming language for DL have been used with Microsoft Azure GPU machines.

Results: Sample width; N = 200; best classification performances in ML methods, Random Forest (RF) 0.73 Accuracy (Acc.) (95% GA [0.60; 0.83], 0.73 Sensitivity, 0.73 Specificity) and CART 0.73 Accuracy (Acc.) (95% GA [0.60; 0.83], 0.70 Sensitivity, 0.76 Specificity), while DL 0.60 Acc. (95% CI [0.52; 0.64], 0.65 Sensitivity, 0.57 Specificity). For N = 600; MLM methods of DVM 0.78 Acc. (95% CI [0.72; 0.84], 0.84 Sensitivity, 0.73 Specificity), DL achieved 0.64 Acc. (95% CI [0.62; 0.63], 0.60 Sensitivity, 0.57 Specificity). When the sample size is increased to a maximum of 1000, it can be said that the change in correct classification success for both ML and DL are not affected; In ML, Support vector machines (SVM) 0.75 Acc. (95% GA [0.69; 0.79], 0.74 Sensitivity, 0.75 Specificity), whereas DL obtained 0.64 (95% GA [0.62; 0.65], 0.58 Sensitivity, 0.60 Specificity).

Conclusion: In ML-based studies, the most curious issue is which method will make better estimation in studies with high dimensional data. The weakest point of this kind of comparison studies are the lack of real big data applications. In this study, the most commonly used GWAS dataset was compared with the classical ML methods and DL methods. The relationships between the genetic variation data and the sample size were taken into consideration. In recent years, especially in the data sets where the relationships between independent variables are very high, the ratio of estimation to SVM is high. In our study, the highest result was obtained with DVM regardless of sample size. This proves the success of nonlinear ML methods that are frequently encountered in genetic data.

Keywords: Classification, Single nucleotide polymorphism, Machine Learning, Deep Learning.

References

1. [1] Szymczak, S., Biernacka, J. M., Cordell, H. J., González-Recio, O., König, I. R., Zhang, H., & Sun, Y. V. (2009). Machine learning in genome-wide association studies. Genetic epidemiology, 33 (S1), S51-S57.
2. [2] Teo, Y. Y. (2008). Common statistical issues in genome-wide association studies: a review on power, data quality control, genotype calling and population structure. Current opinion in lipidology, 19 (2), 133-143.
3. [3] Hong, E. P., & Park, J. W. (2012). Sample size and statistical power calculation in genetic association studies. Genomics & informatics, 10 (2), 117.

KISA SÖZLÜ BİLDİRİLER
SHORT ORAL PRESENTATIONS

K1

Comparison of Diagnostic Performance of Binary - Multliclass ROC Analysis and Multinomial Logistic Regression by Using Glasgow Coma Scale Levels

Arzu BAYGÜL[1], Neslihan GÖKMEN[2]

[1] Department of Biostatistics, Koç University, İstanbul – TURKEY
[2] Basic Sciences, İstanbul University, İstanbul - TURKEY

Aim: The use of Receiver Operator Characteristic (ROC) analysis for model selection and threshold optimization has become a standard practice for the design of two class pattern recognition systems. Multi-class operating characteristics are a generalization of two classes of receiving operator characteristics. A challenge to this method is the complexity of the calculation with an increasing number of classes. In this study, multi-class ROC calculations are presented and a 3-category classification study is performed. For this purpose, it was aimed to determine the AUC for CRP according to Glasgow Coma Scale (GCS).

Method: The CRP levels of the 45 patients included in the study were measured and GcS scores were collected for three different risk statement which were low risk, moderate risk and high risk. To determine the classification power of GCS multi-class ROC analysis were used and multinomial logistic regression was utilized to evaluate the affect of GCS to CRP. Analyzes were carried out using R-pROC package.

Results: According to pairwise calculations of AUC in binary levels, AUC is found higher than >.70. The average of pairwise AUC values are calculated equal to multi-class AUC value which is 0.706. The results show the power of GKS levels is significant for CRP. According to the results of multinomial logistic regression analysis, when the low risk level was taken as the reference category, the effect of CRP was found to be significant at the modarate risk level. This result is in parallel with pairwise AUC (low vs. modarate) values calculated in binary.

Conclusion: According to the findings and application results; the average of pairwise AUC values are equivalent to multi-class ROC analysis. AUC values of significant levels in multinominal logistic regression were also high for GCS application.

Keywords: multiclass ROC, multinomial logistic regression, AUC

K2

Değişim Noktası Analizi

İsmet DOĞAN[1] Nurhan DOĞAN[1]

[1]Afyonkarahisar Sağlık Bilimleri Üniversitesi, Tıp Fakültesi,
Biyoistatistik ve Tıbbi Bilişim Anabilim Dalı, Afyonkarahisar-TÜRKİYE

Amaç: Bu çalışmada, aynı veride bir veya birkaç değişim noktasını tespit etmek için yaygın olarak kullanılan değişim noktası yöntemleri (Birikimli Toplam - CUSUM, Pettitt Yöntemi ve Hata Kareler Ortalaması Minimizasyonu Yöntemi) gözden geçirilmiştir.

Yöntem: Yapısal değişim problemlerini test etmek istatistikte önemli bir konudur. Değişim noktası belirlemede öncelikle dikkate alınan veri, t zamanından önceki ve t zamanından sonraki veriler olmak ikiye parçalanmaktadır. Değişim noktası tespitinde kullanılan klasik istatistiksel formüller, t zamanından önceki ve sonraki verilere ait olasılık dağılımlarını göz önünde bulundurmakta ve iki dağılım önemli ölçüde farklı ise t zaman noktasını bir değişim noktası olarak kabul etmektedir. Bu ifade matematiksel olarak aşağıdaki şekilde ifade edilmektedir; $\{Y_t\}$ ilgilenilen değişkene ait t anındaki gözlenen değer olsun. F_1 ve F_2 ise sırasıyla t zamanından önceki ve sonraki verilerin sahip olduğu dağılımları göstersin. Eğer,

$$\begin{matrix} \{Y_t \sim F_1 & t < t^* \} \\ \{Y_t \sim F_2 & t \geq t^* \} \end{matrix}$$

ise t^* zamanı bir değişim noktasıdır. Değişim Noktası Analizi'nde amaç önceden belirlenen bazı yanlış alarm oranlarını da (α) dikkate alarak değişimin başladığı andan itibaren mümkün olan en kısa sürede t^* belirlemektir. Bilimsel olarak problem,

$H_0 : X_1, X_2, \ldots\ldots, X_n \sim F_1$
$H_1 : X_i \sim F_1,\ X_j \sim F_2, \qquad i = 1, 2, \ldots\ldots, v-1,\ \ j = v, v+1, \ldots\ldots, n,$

hipotezinin test edilmesidir. Burada $X_1, X_2, \ldots\ldots, X_n$ bağımsız gözlem değerlerine ait örneklemi, F_1ve F_2 ise sırasıyla f_1 ve f_2 yoğunluk fonksiyonlarına karşılık gelen dağılım fonksiyonlarını göstermektedir. F_1 ve F_2 dağılım fonksiyonlarının bilinmesi gerekli değildir. Bilinmeyen v parametresi, $2 \leq v \leq n$ değişim noktası olarak isimlendirilir.

CUSUM

$X_1, X_2, \ldots\ldots, X_n$ n tane ardışık veri olsun. Bu verilerden yararlanarak $S_0, S_1, S_2, \ldots\ldots, S_n$ ile gösterilen birikimli toplamlar ve bunlardan yararlanarak değişim noktası aşağıdaki gibi hesaplanmaktadır:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}, \qquad S_i = S_{i-1} + (x_i - \bar{x}), \qquad S_0 = 0$$

$$Değişim\ Noktası \rightarrow |S_k| = \max_{i=0,1,\ldots\ldots,n} |S_i|$$

Pettitt Yöntemi

$$D_{ij} = sgn(x_i - x_j), \quad U_{k,n} = \sum_{i=1}^{k} \sum_{j=k+1}^{n} D_{ij}$$

$$Değişim\ Noktası \rightarrow K_n = \max_{1 \le k \le n} |U_{k,n}|$$

Hata Kareler Ortalaması Minimizasyonu Yöntemi

$$\bar{x}_1 = \frac{\sum_{i=1}^{k} x_i}{k}, \quad \bar{x}_2 = \frac{\sum_{i=k+1}^{n} x_i}{n-k}$$

$$MSE(k) = \sum_{i=1}^{k} (x_i - \bar{x}_1)^2 + \sum_{i=k+1}^{n} (x_i - \bar{x}_2)^2$$

$$Değişim\ Noktası \rightarrow \min_{1 \le k \le n} MSE(k)$$

Sonuç ve Tartışma: Birçok uygulamada, yapısal değişimin ne zaman gerçekleştiğini bilmek faydalıdır. Değişim Noktası Analizi hem uygulamalı hem de teorik istatistikte önemlidir. Literatürde, bu çalışmada dikkate alınan değişim noktası tespit yöntemlerinin karşılaştırıldığı bir çalışma bulunmamaktadır. Farklı veri setlerinden elde edilen simülasyon sonuçlarına göre, değişim noktalarını tespit etmek için bu yöntemlerin basit ve etkili yöntemler olduğu sonucuna ulaşılmıştır.

Anahtar Kelimeler: Zaman serileri analizi, ardışık veri analizi, değişim noktası belirleme, CUSUM, Pettitt yöntemi, hata kareler ortalaması minimizasyonu

K2

Change-Point Analysis (CPA)

İsmet DOĞAN[1] Nurhan DOĞAN[1]

[1]Afyonkarahisar Health Sciences University, Faculty of Medicine, Department of Biostatistics and Medical Informatics, Afyonkarahisar-TURKEY

Aim: In this study, the three most popular CPA methods to detect one or several change points, CUSUM, Pettitt Method and Mean Squared Error Minimization Method are reviewed.

Method: Testing on structural change problems has been an important issue in statistics. A typical statistical formulation of change-point detection is to consider probability distributions from which data in the past and present intervals are generated, and regard the target time point as a change point if two distributions are significantly different. Let $\{Y_t\}$ be an observed process in discrete time. The time moment t^* is a change point with F_1 and F_2 distribution functions if

$$\begin{Bmatrix} Y_t \sim F_1 & t < t^* \\ Y_t \sim F_2 & t \geq t^* \end{Bmatrix}$$

The goal is to detect t^* as soon as possible after its occurence, under some predefined false-alarm rate α. Formally, the problem is that of hypothesis testing:

$$H_0 : X_1, X_2, \ldots\ldots, X_n \sim F_1$$

$$H_1 : X_i \sim F_1,\ X_j \sim F_2, \qquad i = 1, 2, \ldots\ldots, v-1,\ \ j = v, v+1, \ldots\ldots, n,$$

where $X_1, X_2, \ldots\ldots, X_n$ is a sample of independent observations, F_1 and F_2 are distribution functions with corresponding density functions f_1 and f_2, respectively. The distribution functions F_1 and F_2 are not necessary known. The unknown parameter $v,\ 2 \leq v \leq n$ is called a change point.

CUSUM

Let $X_1, X_2, \ldots\ldots, X_n$ represent the n data points. From this, the cumulative sums $S_0, S_1, S_2, \ldots\ldots, S_n$ are calculated. The cumulative sums are calculated as follows:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}, \qquad S_i = S_{i-1} + (x_i - \bar{x}), \qquad S_0 = 0$$

$$Change\ Point \rightarrow |S_k| = \max_{i=0,1,\ldots\ldots,n} |S_i|$$

Pettitt Method

$$D_{ij} = sgn(x_i - x_j), \quad U_{k,n} = \sum_{i=1}^{k} \sum_{j=k+1}^{n} D_{ij}$$

$$Change\ Point \rightarrow K_n = \max_{1 \le k \le n} |U_{k,n}|$$

Mean Squared Error Minimization

$$\bar{x}_1 = \frac{\sum_{i=1}^{k} x_i}{k}, \quad \bar{x}_2 = \frac{\sum_{i=k+1}^{n} x_i}{n-k}$$

$$MSE(k) = \sum_{i=1}^{k} (x_i - \bar{x}_1)^2 + \sum_{i=k+1}^{n} (x_i - \bar{x}_2)^2$$

$$Change\ Point \rightarrow \min_{1 \le k \le n} MSE(k)$$

Results and Conclusion: In many applications, it is useful to know when the structural change occurred. CPA is important in both applied and theoretical statistics. In the literature, there is no comparative study involving the change point detection methods considered. According to the results obtained from our simulations for different conditions, it is concluded that these three methods are simple and effective methods to detect change points.

Keywords: Time series analysis, sequential data analysis, change point detection, CUSUM, Pettitt method, Mean Squared Error Minimization

K3

Alzheimer Hastalarında Bayesci Hızlandırılmış Başarısızlık Süresi Modelinin Cox Oransal Hazard Modeliyle Karşılaştırılması: Bir Gerçek Veri Uygulaması

Ilgın Asena CEBECİ[1], Damla ÖZTÜRK[2], Başak DOĞAN[2], Nural BEKİROĞLU[3]

[1] Marmara Üniversite Sağlık Bilimleri Enstitüsü, Biyoistatistik Anabilim Dalı, İstanbul
[2] Marmara Üniversite, Diş Hekimliği Fakültesi, Periodontoloji Anabilim Dalı, İstanbul
[3] Marmara Üniversitesi, Tıp Fakültesi, Biyoistatistik AnabilimDalı, İstanbul

Amaç: Bayesci yaklaşımda önsel bilgi ile örneklemden sağlanan bilgi birleştirilerek parametrenin sonsal dağılımı elde edilir. Bu çalışmanın amacı, ağız sağlığı göstergelerinin Alzheimer hastalarındaki demans şiddeti durumuna göre önsel dağılım kullanılarak Bayesci Hızlandırılmış Başarısızlık Süresi (HBS) modeliyle tahminlenmesidir.

Yöntem: Marmara Üniversitesi Periodontoloji Bölümü ile gerçekleştirilen bu çalışmada İstanbul'da iki farklı merkezden 36 Alzheimer hastasına ulaşılmıştır. Takip süresi Alzheimer hastalık süresi ve başarısızlık ileri demans olarak çalışmaya alınmıştır. Toplanan 30 bağımsız değişkenin tek değişkenli analizi sonucunda Geriatrik Ağız Sağlığı Değerlendirme İndeksinin fonksiyonel boyutu (GOHAI fonksiyonel), DMFT (çürük, kayıp, dolgulu dişler) indeksi ve Ağız Diş Bakım Sıklığı istatistiksel olarak anlamlı çıkmıştır. Önce Cox Oransal Hazard (COH) ile tahmin modeli oluşturulmuş daha sonra Bayesci HBS modeli ile analiz edilmiştir. Bayesci HBS modeli Weibull dağılımı ve parametre tahmini için Markov Zinciri Monte Carlo (MCMC) simülasyonu kullanılmıştır. Bayesci HBS çok değişkenli analizinde önsel dağılımlar; GOHAI fonksiyonel değişkeni için beta (8, 2) dağılımı ve DMFT değişkeni için normal (23, 4) dağılımı olarak alınmıştır. Ağız Diş Bakım Sıklığı değişkeni için günde en az bir defa ağız bakımı yapanlar referans kategori olarak kabul edilmiştir. Model uyumu Sapma Bilgi kriteri (DIC) ile değerlendirilmiştir. Analizler Stata paket programı ile gerçekleştirilmiştir.

Bulgular: COH modeli analizine göre, tek değişkenli analizde anlamlı olan üç değişken arasından sadece GOHAI fonksiyonel değişkeni istatistiksel olarak anlamlı çıkmıştır [Hazard Oran:0.77, %95 Güven Aralığı: 0.65 – 0.97]. Bayesci HBS modelinde sonucunda GOHAI fonksiyonel [Zaman Oranı(ZO):1.50, %95 Kredi Aralığı: 1.24-1.91], DMFT [ZO:0.95, %95 Kredi Aralığı: 0.89–1.02], Ağız Diş Bakımı Sıklığı haftada 2 defadan az olma durumu [ZO:4.45, %95 Kredi Aralığı: 0.79–18.84] ve haftada birkaç defa olma durumu [ZO:1.94, %95 Kredi Aralığı: 0.41–5.95] olarak hesaplanmıştır. Bayesian HBS modelinde sadece GOHAI fonksiyonel değişkeni istatistiksel olarak anlamlı bulunmuştur. Hasta ağız sağlığının fonksiyonel kalitesindeki artış hızının 1.50 olması; demansın ileri seviyelere geçme süresini yarıladığı anlamına gelmekte ve hastaların ileri demans seviyesine geçişini yavaşlatmaktadır. Modelin, DIC değeri ve MCMC simülasyonu için Kabul Hızı (Acceptance rate) sırasıyla 84.45 ve 0.34 olarak hesaplanmıştır.

Sonuç: Sonuç olarak, Alzheimer hastaları gibi çalışılması zor bir hasta grubunda söz konusu olan küçük örneklemin taşıdığı bilgiler çok önemli ve değerlidir. Böylesi küçük örneklem çalışmalarında Bayesci yaklaşımın COH yaklaşımına göre hem klinik hem de biyoistatistiksel açıdan daha iyi sonuçlar verdiği gözlenmiştir.

Anahtar kelimeler: Bayesci Hızlandırılmış Başarısızlık Süresi Modeli, Alzheimer, Cox Oransal Hazard Modeli

K3

Bayesian Accelerated Failure Time Model Application in Alzheimer's Disease Patients and Its Comparison with Cox Proportional Hazard Model: A Real Data Application

Ilgın Asena CEBECİ[1], Damla ÖZTÜRK[2], Başak DOĞAN[2], Nural BEKİROĞLU[3]

[1]Marmara University Institute of Health Sciences, Department of Biostatistics, Istanbul
[2]Marmara University Faculty of Dentistry, Department of Periodontology, Istanbul
[3]Marmara University School of Medicine, Department of Biostatistics, Istanbul

Aim: In Bayesian approach, priors and information obtained from the sample are used to estimate parameters' posterior distributions. Aim of this study is to estimate oral health indicators for Alzheimer's disease patients according to dementia severity through utilizing informative priors with Bayesian Accelerated Failure Time (AFT) model.

Method: This study is carried out with Marmara University Department of Periodontology, and 36 Alzheimer's disease patients from two centers in Istanbul were participated. Follow up time was Alzheimer's disease duration, and the failure event was having severe dementia. The univariate analysis was performed for 30 covariates, amongst them Geriatric Oral Health Assessment Index functional dimension (GOHAI functional), DMFT (number of decayed, missing and filled teeth) index and Oral Care frequency were found statistically significant. Firstly, Cox Proportional Hazard (CPH) model was performed on the survival data. Then Bayesian AFT model was performed with Weibull distribution and Markov Chain Monte Carlo (MCMC) simulation for the parameters' estimation of posterior distribution. Prior distributions for the covariates of GOHAI functional and DMFT were beta (8, 2) distribution, and normal (23, 4) distribution respectively. Daily oral care was used as a reference category for the variable Oral Care frequency. Deviation Information Criteria (DIC) of the models was calculated to evaluate the model fit. The analyzes were performed with Stata program.

Results: CPH model showed that the GOHAI functional as the only statistically significant covariate [Hazard Ratio:0.77, 95% Confidence Interval: 0.65-0.97]. Results of the multivariate model were as follows; GOHAI functional [Time Ratio(TR):1.50, 95% Credible interval 1.24-1.91]. DMFT [TR:0.95, 95% Credible interval 0.89–1.02], Oral Care frequency of less than 2 times a week [TR:4.45, 95% Credible interval: 0.79–18.84], and Oral Care frequency of several times a week [TR: 1.94, 95% Credible interval 0.41–5.95]. Bayesian AFT model indicated that the GOHAI functional as the only statistically significant covariate. The TR value of 1.50 could be interpreted as the functional dimension of oral health increases, patient's progression of severe dementia slows down. The DIC value of the model and the Acceptance rate for MCMC simulation were 84.45 and 0.34, respectively.

Conclusion: Information obtained from a small sample is very important and valuable especially for a group of patients where it was difficult to conduct a research, such as Alzheimer's disease. In our study, we found that that the Bayesian approach seems to have better clinical and statistical results than the CPH approach in small sample.

Keywords: Bayesian Accelerated Failure Time Model, Alzheimer's Disease, Cox Proportional Hazard Model

K4

Engelli Kadınlarda Psikiyatrik Belirti Dağılımı

İsmail YILDIZ, Ömer SATICI, Remzi OTO

Biyoistatistik Anabilim Dalı, Tıp Fakültesi, Dicle Üniversitesi
Ruh Sağlığı ve Hastalıkları Anabilim Dalı, Tıp Fakültesi Dicle Üniversitesi

Amaç: Engelliler toplumun önemli bir bölümünü oluşturmalarına ve bu yönde yapılmış bir çok yasal iyileştirici düzenlemeye karşın, hala bir çok psiko-sosyal sorunlar yumağında yaşamlarını sürdürmek zorunda kalmaktadırlar. Genel olarak yaşanan sorunlar, engelli, kadın olunca daha zor durumlara hazırlıklı olması kaçınılmaz olmaktadır. Yaşanan ekonomik, sosyal, kültürel sorunlar ve engellilere yönelik kalıp yargılar, önyargılar engelli bireylerde ruhsal sorunlara da yol açmaktadır. Bu çalışmada, daha önceki çalışmalarda eksik görülen engelli kadınlar ele alındı. Diyarbakır ili Engelliler derneğine kayıtlı 600 Kadın üyeden ulaşabilen 92 Engelli kadında Psikiyatrik Belirti Dağılımı incelendi.

Yöntem: Çalışmanın verileri 2016 yılında toplanmaya başlandı. Herhangi bir alanda fiziksel engeli bulunan kadınların yaşadıkları psikiyatrik sorunları saptamak, sorunların çözümü için yapılabilecekleri planlamak açısından, bireylerin sosyal ve demografik özelliklerini saptamaya yönelik, Geçerlilik ve Güvenirlik Analizi yapılmış, Engelli birey Sosyo-Demografik anket formu ile SCL90-R Psikiyatrik Belirti tarama testi uygulandı. Engelli kadınlarda, çalışma örnekleminin başlangıç özellikleri olarak; kişilerin medeni hali, eğitimi, mesleği, engellilik nedeni, bedensel engel tipi incelendi. Engelli kadınlarda, çalışma örnekleminin başlangıç özellikleri ile Kadınlarda Yas Gurupları için; somatizasyon, obsesifkompulsiz, int, depresyon, kaygı anksiyete, öfke düşmanlık, fobik anksiyete, paranoid düşünce, psikotizm ve 10 ek madde; Kolmogorov –Smirnov, Binom ve Odds Ratio testleri kullanılarak analiz edildi.

Bulgular: Engelli kadınlarda, çalışma örnekleminin başlangıç özellikleri açısından, incelediğimiz değişkenlerden sırasıyla; medeni hali, eğitimi, mesleği, engellilik nedeni, bedensel engel tipi, bedensel engellilik eğitimi alıp almadığı, sivil toplum kuruluşu üyeliği, sivil toplum kuruluşuna aktif katılıp katılmadığı ve anketteki üç sorunu açıklayıp açıklaması değişkenleri aşısından anlamlı farklar gözlendi($p<0.001$).

Sonuç: Engelli kadınların önemli bir bölümünün, psikiyatrik belirti alanlarında müdahaleyi gerektirecek düzeylerde sorunlar yaşadıkları saptanmıştır. Her üç engelli kadından biri (%35.9) sorunlarını herhangi bir bedensel belirti ile ifade etmek durumunda kalmaktadır. Kişilerarası ilişkiler alanı (%26.1) ile, öfke ve düşmanlık (%26.1) ve uyku-yeme sorunları(ek maddeler %25.0) alanlarında daha yüksek olmak üzere sorunlar yaşadıkları saptanmıştır. Çalışmamızda sırasıyla; Somatizasyon, Obsesif Kompulsif, Kişilerarası İlişkiler, Depresyon Ve Fobik Anksiyete gibi psikolojik belirtiler, kadınlarda en sık tespit edilmiştir. Bu durum kadın engelliler için yeni bir dezavantajlı durum yaratmış olup, engelli kadınların toplumsal işlevlerini sürdürmelerinde yeni bir sorun alanına yol açmaktadır. Psikiyatride engellilik halinin özellikle kadınlarda ayrıntılı olarak değerlendirilmesi, koruyucu ve önleyici ruh sağlığı programları geliştirilmesi açısından önem taşımaktadır.

K5

Baş Boyun Diseksiyonu Yapilan Kanserli Olgularda Skapular Diskineziyi Tahminlemek için Görüntüleme Yöntemlerinin Performansinin Değerlendirilmesi

Su ÖZGÜR[1], Pembe KESKİNOĞLU[2], Hande Melike HALAÇ[3], Aybüke Cansu KALKAN[4], Seher ÖZYÜREK[5], Ali BALCI[3], Ersoy DOĞAN[6], Ahmet Ömer İKİZ[6], Arzu GENÇ[5]

[1]Ege Üniversitesi Tıp Fakültesi, Biyoistatistik ve Tıbbi Bilişim AD, İzmir
[2]Dokuz Eylül Üniversitesi Tıp Fakültesi, Biyoistatistik ve Tıbbi Bilişim AD, İzmir
[3]Dokuz Eylül Üniversitesi Tıp Fakültesi, Radyoloji AD, İzmir
[4]Dokuz Eylül Üniversitesi Sağlık Bilimleri Enstitüsü, İzmir
[5]Dokuz Eylül Üniversitesi Tıp Fakültesi, Fizik Tedavi ve Rehabilitasyon Yüksekokulu, İzmir
[6]Dokuz Eylül Üniversitesi Tıp Fakültesi, Kulak Burun Boğaz AD, İzmir

Amaç: Baş boyun kanserlerinde boyun diseksiyonu gerçekleştirilirken aksesuar sinir hasarı sık görülmektedir. Aksesuar sinir hasarı trapez kasının işlevini etkiler ve skapular diskinezi görülür. Operasyon öncesi de kitlenin yerleşim yeri nedeniyle bu sinirin etkilenimi görülebilmektedir. Bu pilot çalışmanın amacı boyun kanserli hastalarda diseksiyon öncesi skapular diskinezi varlığını tahminlemede kas kitle ölçümünün manyetik rezonans (MR) ve ultrasonografi (USG) ile performanslarının değerlendirmektir.

Yöntem: Baş boyun kanserli 14 hasta (11 erkek, 3 kadın) çalışmaya dahil edilmiştir. 14 hastanın hem sağ, hem sol tarafından kas kalınlıkları ölçümü USG ve MR görüntüleme ile değerlendirilmiş olup skapular diskinezi varlığı ise McClure ve ark. tarafından tanımlanan prosedüre göre Skapular Diskinezi Testi ile belirlenmiştir. Ölçümlerin diskineziyi öngörme performansları ROC analizi ile değerlendirilmiştir. Trapez kası kalınlıkları alt, orta ve üst trapez olmak üzere 3 ayrı bölgeden değerlendirilmiştir. Görüntülemelerde kas yapısını en iyi ortaya koyan bölümler orta-üst ve alt olarak sıralanmaktadır. Bu nedenle üst ve orta bölümdeki ölçümlerin performansı daha ön planda incelenmiştir.

Bulgular: Sekiz hastada diskinezi saptanmamıştır. Diğer hastalarda hafif/orta/keskin bulgular elde edilmiştir. USG ile sağ üst trapez bölgesi diskinezi varlığı ROC eğrisi altında kalan alan istatistik olarak değerlendirildiğinde ayırt edici saptanmıştır (AUC=0,917, p=0,010). İncelenen 14 hastada lezyonların sağ taraf ağırlıklı olması nedeni ile "n" yeterliğinden kaynaklı ayırt edicilik sağ tarafta saptanmıştır.

Sonuç: Aksesuar sinir hasarı gelişen durumda, trapez kas işlev yitimi gerçekleşmeden, sorunu invaziv olmayan ucuz, hastaya rahatsızlık vermeyen görüntüleme yöntemleri ile erken tespit etmek önemlidir. Bu ön çalışmanın sonuçlarına göre üst trapez bölgesinde diskinetik hareket varlığı USG kas kalınlığı ölçümleri ile tahminlenebilmektedir. Bulgu ve sonuçlarımız operasyon öncesi belirtiler için geçerlidir. Operasyon sonrası 6. aya kadar izlem ile aynı görüntüleme tekniklerinin disknezi ile birlikte, trapez kas fonksiyonu olan omuz hareketlerinin sınırlanmasının gelişimi (donuk omuz) için ayırt ediciliği boylamsal olarak çalışılacaktır. Engelliliğin erken tanı ile önlenmesi, şiddetinin azaltılması için tanı testlerinin performansının incelenmesi önem taşımaktadır.

Teşekkür: Bu çalışma, Dokuz Eylül Üniversitesi özgelirleri ile Bilimsel Araştırma Projeleri Koordinasyon Birimi tarafından 2017.KB.SAG.053 proje numarası ile desteklenmektedir

Anahtar Kelimeler: ROC, skapular diskinezi, trapez kas kalınlığı

Tablo 1.Trapez kası disknezisini belirlemede görüntüleme yöntemlerinin performansları

Kas	Yöntem	AUC	p değeri	CI
Preop üst trapez R	**USG**	0,917	0,010	0,767 - 1,000
	MR	0,563	0,699	0,244 - 0,881
Preop üst trapez L	**USG**	0,729	0,156	0,449 - 1,000
	MR	0,646	0,366	0,339 - 0,953
Preop orta trapez R	**USG**	0,813	0,043	0,567 - 1,000
	MR	0,688	0,245	0,350 - 1,000
Preop orta trapez L	**USG**	0,792	0,071	0,527 - 1,000
	MR	0,583	0,606	0,251 - 0,916
Preop alt trapez R	**USG**	0,729	0,156	0,436 - 1,000
	MR	0,667	0,302	0,352 - 0,981
Preop alt trapez L	**USG**	0,51	0,949	0,181 - 0,840
	MR	0,656	0,333	0,350 - 0,963

p<0,05 Anlamlılık düzeyi, AUC: Eğri altında kalan alan CI: Güven Aralığı R: Sağ L: Sol

Kaynaklar

1. Cliona O'Sullivan, Jim Meaney, Gerard Boyle, John Gormley, Maria Stokes Man Ther., The validity of Rehabilitative Ultrasound Imaging for measurement of trapezius muscle thickness. 2009 Oct; 14(5): 572–578. Published online 2009 Mar 5. doi: 10.1016/j.math.2008.12.005
2. Michener L.A., Subasi Yesilyaprak S.S., Seitz A.L., Timmons M.K., Walsworth M.K, Supraspinatus tendon and subacromial space parameters measured on ultrasonographic imaging in subacromial impingement syndrome, Knee Surgery Sports Traumatology Arthroscopy, 23, 363-69, (2015).
3. Nibu K., Ebihara Y., Ebihara M., Kawabata K., Onitsuka T., Fujii T., Saikawa M., Quality of life after neck dissection: a multicenter longitudinal study by the Japanese Clinical Study Group on Standardization of Treatment for Lymph Node Metastasis of Head and Neck Cancer, International Journal of Clinical Oncology,15, 33-8, (2010).
4. O'Sullivan C., Bentman S., Bennett K., Stokes M., Rehabilitative ultrasound imaging of the lower trapezius muscle: technical description and reliability, Journal of Orthopaedic and Sports Physical Therapy, 37, 620-26, (2007).

K5

Evaluation of the Performance of Imaging Methods in Cancer Patients Who Has Head and Neck Dissection to Estimate Scapular Dyskinesia

Su ÖZGÜR[1], Pembe KESKİNOĞLU[2], Hande Melike HALAÇ[3], Aybüke Cansu KALKAN[4], Seher ÖZYÜREK[5], Ali BALCI[3], Ersoy DOĞAN[6], Ahmet Ömer İKİZ[6], Arzu GENÇ[5]

[1]Ege University Faculty of Medicine Department of Biostatistics and Medical Informatics, İzmir
[2]Dokuz Eylul University Faculty of Medicine Department of Biostatistics and Medical Informatics, İzmir
[3]Dokuz Eylul University Faculty of Medicine, Department of Radyology, İzmir
[4]Dokuz Eylul University, Institute of Health Sciences, İzmir
[5]Dokuz Eylul University Faculty of Medicine, School of Physical Therapy And Rehabilitation Department of Physiotheraphy and Rehabilitation, İzmir
[6]Dokuz Eylul University Faculty of Medicine, Department of Otolaryngology, İzmir

Aim: Accessory nerve damage is common in neck and neck dissection in head and neck cancers. Accessory nerve damage affects the function of the trapezoidal muscle and scapular dyskinesia occurs. Preoperatively, this nerve may be affected due to the location of the mass. The aim of this pilot study was to evaluate the performance of muscle mass measurement by MRI and USG in predicting the presence of scapular dyskinesia before dissection in patients with neck cancer.

Method: Fourteen patients (11 males, 3 females) with head and neck cancer were included in the study. Imaging measurements were taken from both right and left neck of 14 patients. ROC analysis was used to evaluate the predictive performance for dyskinesia. Measurements were taken from different regions of the muscle mass (lower, middle and upper). In the imaging, the sections that best reveal the muscle structure are listed as middle-upper and lower. Therefore, the performance of the measurements in the upper and middle sections has been examined in the foreground.

Results: Dyskinesia was not detected in eight patients. Mild / moderate / sharp dyskinetic findings were obtained in other patients. The presence of dyskinesia in the upper right trapezoidal region by USG was found to be discriminative when the area under the ROC curve was evaluated statistically (AUC = 0.917, p = 0.010). Since the masses were located on the right-sided in the majority of patients, the number of cases on the right-sided was sufficient, and the discrimination was detected on the right-sided.

Conclusion: In the case of accessory nerve damage, it is important to detect the problem early with non-invasive, inexpensive, uncomfortable imaging methods before the loss of trapezoidal muscle function. According to the results of this preliminary study, the presence of dyskinetic motion in the upper trapezoid region can be estimated by USG muscle thickness measurements. Our findings and results are valid for preoperative symptoms. In the follow-up up to the 6th postoperative period, the same imaging techniques will be studied longitudinally for the development of the restriction of shoulder movements with dysfunctional and trapezoidal muscle function (dull shoulder). It is important to examine the performance of diagnostic tests to prevent disability with early diagnosis and to reduce its severity.

Acknowledgment: This study was supported by Dokuz Eylul University Scientific Research Projects Coordination Unit with project number 2017.KB.SAG.053

Keywords: ROC, scapular dyskınesia, trapezius muscle thickness

Table 1. Performances of imaging methods in determining the dyskinesia of triple muscle

Muscle	Methods	AUC	p value	CI
Preop upper trapezoidal R	**USG**	0,917	0,010	0,767 - 1,000
	MR	0,563	0,699	0,244 - 0,881
Preop upper trapezoidal L	**USG**	0,729	0,156	0,449 - 1,000
	MR	0,646	0,366	0,339 - 0,953
Preop middle trapezoidal R	**USG**	0,813	0,043	0,567 - 1,000
	MR	0,688	0,245	0,350 - 1,000
Preop middle trapezoidal L	**USG**	0,792	0,071	0,527 - 1,000
	MR	0,583	0,606	0,251 - 0,916
Preop lower trapezoidal R	**USG**	0,729	0,156	0,436 - 1,000
	MR	0,667	0,302	0,352 - 0,981
Preop lower trapezoidal L	**USG**	0,51	0,949	0,181 - 0,840
	MR	0,656	0,333	0,350 - 0,963

p<0,05 Significance level, AUC: Area under the curve CI: Confidence Interval R:Right L:Left

References

1. Cliona O'Sullivan, Jim Meaney, Gerard Boyle, John Gormley, Maria Stokes Man Ther., The validity of Rehabilitative Ultrasound Imaging for measurement of trapezius muscle thickness. 2009 Oct; 14(5): 572–578. Published online 2009 Mar 5. doi: 10.1016/j.math.2008.12.005
2. Michener L.A., Subasi Yesilyaprak S.S., Seitz A.L., Timmons M.K., Walsworth M.K, Supraspinatus tendon and subacromial space parameters measured on ultrasonographic imaging in subacromial impingement syndrome, Knee Surgery Sports Traumatology Arthroscopy, 23, 363-69, (2015).
3. Nibu K., Ebihara Y., Ebihara M., Kawabata K., Onitsuka T., Fujii T., Saikawa M., Quality of life after neck dissection: a multicenter longitudinal study by the Japanese Clinical Study Group on Standardization of Treatment for Lymph Node Metastasis of Head and Neck Cancer, International Journal of Clinical Oncology,15, 33-8, (2010).
4. O'Sullivan C., Bentman S., Bennett K., Stokes M., Rehabilitative ultrasound imaging of the lower trapezius muscle: technical description and reliability, Journal of Orthopaedic and Sports Physical Therapy, 37, 620-26, (2007).

K6

Yüksek Kayıp Gözlem Oranına Sahip Çalışmalarda Derin Öğrenme Uygulaması (VUR ve İYE)

Su ÖZGÜR[1], Pembe KESKİNOĞLU[2], Erdal COŞGUN[3], Timur KÖSE[1], Ahmet KESKİNOĞLU[4]

[1]Ege Üniversitesi Tıp Fakültesi Biyoistatistik ve Tıbbi Bilişim AD, İzmir
[2]Dokuz Eylül Üniversitesi Tıp Fakültesi Biyoistatistik ve Tıbbi Bilişim AD, İzmir
[3]Genomics Team, Microsoft Research, Redmond, WA, USA
[4]Ege Üniversitesi Tıp Fakültesi, Çocuk Sağlığı ve Hastalıkları AD, Pediatrik Nefroloji BD, İzmir

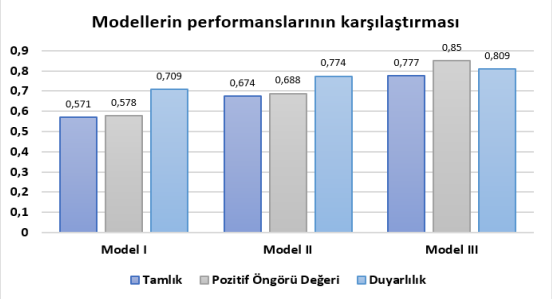
Amaç: Gerçek hayattaki veriler eksik, gürültülü ve tutarsız olma eğilimindedir. Verileri yok sayma yaklaşımı genellikle sınıf etiketi eksik olduğunda (temel hedefin sınıflandırma olduğu varsayılırsa) veya birçok nitelik rasgele olarak satırlarda eksik olduğunda yapılır. Fakat, veri setinde böyle satırların yüzdesi yüksekse, tahminlerin düşük performansına neden olur. Derin öğrenme (DL) temelli yaklaşımlar, veri setindeki eksik değerleri ihmal / yok etmeden veri setini değerlendirebilmektedir. Vezikoüreteral reflü (VUR) ve idrar yolu enfeksiyonu (İYE) ile ilgili hastane kayıtlarında, demografik ve klinik özelliklere, laboratuvar bulgularına ve görüntüleme verilerine sahip tam olguların sayısı çok düşüktür.

Yöntem: Bu çalışmada, Ege Üniversitesi Tıp Fakültesi Hastanesi ve SBÜ Tepecik Eğitim ve Araştırma Hastanesi'nden retrospektif kesitsel bir çalışma ile elde edilen 611 pediatrik hastanın (425 VUR, 186 İYE, %22,9 kayıp oranı) verisi kullanılmıştır. CNTK (The Microsoft Cognitive Toolkit); fiziksel bulgular, laboratuvar ve görüntüleme bulgularını içeren 39 özellik ile farklı modelleri değerlendirmek için kullanmıştır. CNTK, Microsoft Research tarafından geliştirilen bir DL çerçevesidir. CNTK, sinir ağlarını yönlendirilmiş bir grafik aracılığıyla bir dizi hesaplama adımı olarak tanımlar. Derin sinir ağı uygulaması 128 gizli katman ve L1 ve L2 = 0.001 seçimi ile sonlandırılmıştır. Son eğitim modeli için Epoch değeri 5 ve yineleme sayısı 800'dür. Test / eğitim seti oranı tüm analiz için 20/80'dir.

Bulgular: VUR / İYE'nin ayırıcı tanısında DL yöntemiyle üç farklı model değerlendirilmiştir. Model I'e ultrasonografik değişken içermeyen, 14 klinik ve 16 laboratuvar değişkeni dahil edildi. Model II'ye ekstra iki ultrasonografik değişken eklenmiştir. Model III'e 14 klinik, 16 laboratuvar ve 10 ultrasonografik değişken (9 hidronefroz ve 1 dilatasyon) dahil edilmiştir. Model III; 0,777 tamlık, 0,809 duyarlılık ve pozitif öngörü değeri (PPV) 0,850 ile en yüksek performansı göstermiştir (Tablo 1).

Table 1. Modellerin performanslarının karşılaştırması

	Model I	Model II	Model III
Tamlık	0,571	0,674	0,777
PPV	0,578	0,688	0,850
Duyarlılık	0,709	0,774	0,809



Sonuç: Özellikle klinik araştırmalarda, eksik değerleri olmayan veri setleri elde etmek zordur. DL modelleriyle, araştırmacılar var olan tüm verileri sentetik atama yöntemleri uygulamadan kullanabilirler. Öte yandan, derin öğrenmeyle; daha doğru tahminler yapmak ve tekrarlanabilir analizlerle hızlı, güvenilir sonuçlar almak mümkündür.

Teşekkür: Bu çalışma TÜBİTAK tarafından 114S011 proje numarası ile desteklenmiştir.

Anahtar Kelimeler: Derin Öğrenme, Veri Madenciliği, Kayıp Gözlemler, VUR, İYE

Kaynaklar

1. Dong Y, Peng CY., Principled missing data methods for researchers, Springerplus. 2013 May 14;2(1):222. doi: 10.1186/2193-1801-2-222. Print 2013 Dec.
2. Sinharay S, Stern HS, Russell D (2001) The use of multiple imputation for the analysis of missing data. Psychol Meth 6(4):317–329. doi:10.1037/1082-989X.6.4.317
3. https://www.microsoft.com/en-us/research/wp-content/uploads/2016/02/DeepLearning-NowPublishing-Vol7-SIG-039.pdf
4. LeCun, Yann, Yoshua Bengio, and Geoffrey Hinton. "Deep learning." Nature 521, no. 7553 (2015): 436-444.

K6

Applied Deep Learning on Studies with High Missing Ratio (VUR and UTI)

Su ÖZGÜR[1], Pembe KESKINOĞLU[2], Erdal COŞGUN[3], Timur KÖSE[1], Ahmet KESKINOĞLU[4]

[1]Ege University Faculty of Medicine, Department of Biostatistics and Medical Informatics
[2]Dokuz Eylul University Faculty of Medicine, Department of Biostatistics and Medical Informatics
[3]Genomics Team, Microsoft Research, Redmond, WA, USA
[4]Ege University Children's Hospital, Department of Pediatric Nephrology

Aim: Real-world data tends to be incomplete, noisy and inconsistent. Ignore the data row approach is usually done when the class label is missing (assuming goal is classification) or many attributes are randomly missing from the row(not just one). But clearly it leads to poor performance if the percentage of such rows is high. Deep learning (DL) based approaches can evaluate the dataset without omit/impute missing value from dataset. In the hospital records for vesicoureteral reflux(VUR) and urinary tract infection(UTI), the complete number of cases with demographic and clinical characteristics, laboratory findings and imaging data are very low.

Method: The data set used in this study included 611 pediatric patients(425 with VUR, 186 with UTI, 22.9% missing ratio) from Ege University and Tepecik Training and Research Hospital obtained through a retrospective cross-sectional study. CNTK have used for evaluating different models for 39 features (physical findings, laboratory and imaging findings). This is a DL framework developed by Microsoft Research. CNTK describes neural networks as a series of computational steps via a directed graph. Deep neural network implementation finalized with 128 hidden layer and L1 and L2=0.001 selection. Epoch value is 5 and numbers of iterations are 800 for final training model. Testing/training sample ratio is 20/80 for the entire analysis.

Results: Three different models were evaluated by DL for differential diagnosis of VUR/UTI. In Model I, 14 clinical and 16 laboratory variables without any ultrasonographic variables were included. Extra two ultrasonographic variables were added to Model II. In Model III, 14 clinical, 16 laboratory, and 10 ultrasonographic variables (9 hydronephrosis and 1 dilatation) were included. Model III showed the best performance with high accuracy (0.777), precision(0.850) and recall(0.809) for differential diagnosis (Table 1) via DL.

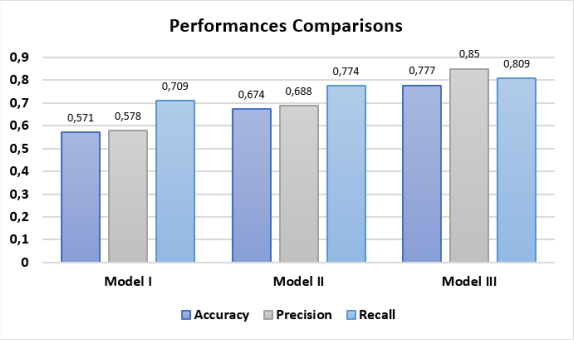
Conclusions: Especially in clinical researches, it is difficult to obtain datasets without missing values. With DL models, researchers can use the all existed data without adding synthetic imputations. On the other hand, it is possible to make more accurate estimates and obtaining results with fast, reliable and repeatable analysis.

Acknowledgment: This study was supported by TÜBİTAK with project number 114S011.

Keywords: Deep Learning, Data Mining, Missing Values, VUR, IYE

Table 1. Performance of the models

Criteria	Model I	Model II	Model III
Accuracy	0.571	0.674	0.777
Precision	0.578	0.688	0.850
Recall	0.709	0.774	0.809



References

1. Dong Y, Peng CY., Principled missing data methods for researchers, Springerplus. 2013 May 14;2(1):222. doi: 10.1186/2193-1801-2-222. Print 2013 Dec.
2. Sinharay S, Stern HS, Russell D (2001) The use of multiple imputation for the analysis of missing data. Psychol Meth 6(4):317–329. doi:10.1037/1082-989X.6.4.317
3. https://www.microsoft.com/en-us/research/wp-content/uploads/2016/02/DeepLearning-NowPublishing-Vol7-SIG-039.pdf
4. LeCun, Yann, Yoshua Bengio, and Geoffrey Hinton. "Deep learning." Nature 521, no. 7553 (2015): 436-444.

K7

Biyoistatistik Eğitim Algımın Yolculuğu ve Diğer Öğrencilerde Bunu Geliştirmek için Bir Fikir

Bayazıt CAN

Tıp Fakültesi, Trakya Üniversitesi, Edirne

Amaç: Tıp Fakültesi'nde aldığım birinci sınıf biyoistatistik dersleri sonucunda nasıl bir biyoistatistik algısı edindiğimi ve okuduğum biri akademik(THE ANALYSİS of BİOLOGİCAL DATA) ve 1 popüler(OUTLİERS Malcolm Gladwell) iki kitapla bu algımın nasıl geliştiğini anlatmak,eğer öğretim üyeleri tarafından uygun görülürse ilgili akademik kitabı dilimizekazandırmak için çalışma başlatmak.

Yöntem ve Bulgular: Tıp fakültesine gelmeden önce özellikle sinirbilim ve beslenme ile ilgili birçok popüler bilim kitabı okudum. Dikkatimi çeken noktalardan bir tanesi doktorların bazı konularda mücadele halinde olması ve anlaşamamaları idi.Fiziksel ve kimyasal kanunlara göre çalışan insan vücudu üzerine yapılan araştırmaların birbirlerinden farklı şeyler söylemeleri beni rahatsız etti.Çünkü mantıksal olarak insanların geldikleri popülasyonlar farklı olsa bile vücut yapılar aynıydı ve bilimde aynı meselede birden fazla doğru olamazdı. Ancak karmaşık insan vücudu ve karmaşık popülasyonlar üzerinde yapılan araştırmalar bizi idealist bir yaklaşımdan çok gerçekçi bir anlayışa zorluyor.Çünkü tıbbi meselelerin arkasındaki fiziksel gerçekliği her zaman tam olarak bilemiyoruzBu gerçekçi anlayışı sağlıklı şekilde edinmek de ancak biyoistatistiğin doğru kullanımı ile mümkün gözüküyor. Biyoistatistiğin doğru kullanımı ile gerçeğe yakın sonuçlar elde edebiliriz.Okulda harcadığım biyoistatistik saatlerinden çıkarttığım dersler bunlar.

Daha önemlisinden bahsetmek, yani biyoistatistik algımın nasıl genişlediğine gelmek istiyorum.Geçen yıl üniversitesi kütüphanesinde Malcolm Gladwell isimli gazetecinin Outliers/Çizginin Dışındakiler isimli kitabını gördüm. Kitap etkileyici bir biçimde başarı kavramında biyoistatistiksel gerçeklerin edindiği rolü gösteriyordu. Bu kitap beni üniversite kütüphanesinde biyoistatistik ile ilgili diğer kitaplara göz atmaya itti ve harika bir kitap buldum. THE ANALYSİS of BİOLOGİCAL DATA.Bu kitap yüksek düzeyde bir matematiksel altyapı gerektirmeden ilginç biyolojik ve tıbbi örneklerle biyoistatistiğin önemini gösteriyor ve çeşitli makalelerle biyoloji istatistik ilişkisinin nasıl geliştiğini anlatıyor.Aynı zamanda biyoistatistiğin yalnızca klinik araştırmalarda değil; günlük hayatta daha geniş bir hayat algısı yakalamakta anahtar rolde olduğunu, insana daha aydın bir düşünce yapısı kazandırabileceğini gösteriyor.Bunun da öğrencilere kazandırılması çok önemli.

Sonuç: The Analysis of Biological Data kitabının Türkçeye çevrilmesi Türkiye'deki biyoistatistik algısının değişmesine katkı sunar kanaatindeyim.

K8

Akciğer Kanserinde Yarışan Riskler

Şirin ÇETİN[1], İsa DEDE[2]

[1] Tokat Gaziosmanpaşa Üniversitesi Tıp Fakültesi Biyoistatistik Anabilim Dalı
[2] Mustafa Kemal Üniversitesi Tıp Fakültesi Tıbbi Onkoloji Anabilim Dalı

Amaç: Kaplan-Meier (K-M) tekniği yaşam çözümlemesi verileri için yaşam olasılığını kestirmede en popüler yaklaşımdır, fakat diğer başarısızlık olaylarını değerlendirme imkânı yoktur. K-M tekniğinde yaşam olasılıkları hesaplanırken tek bir başarısızlık riski dikkate alınmakta olup, diğer başarısızlık riskleri sansürlü olarak kabul edilmektedir. Bu durum ise veri kaybına sebebiyet vermektedir. Dolayısıyla veri ve sansür özelliklerine bağlı olarak doğru modelin seçilmesi önem kazanmaktadır. Bu çalışmada amacımız akciğer kanseri hastalarında tüm başarısızlık nedenlerini dikkate alan yarışan riskler modelini değerlendirmektir.

Yöntem: Bu amaç ile, Mustafa Kemal Üniversitesi Tıp Fakültesi Tıbbi Onkoloji Polikliniğine Ocak 2017 ile Ağustos 2019 tarihleri arasında başvuran akciğer kanseri tanısı almış 150 hastanın verileri incelenmiştir. Histolojik olarak tanısı doğrulanan hastalar çalışmaya alınmıştır. Hastaların tanı anındaki klinik özellikleri ve hematalojik parametrelerin prognoz üzerine etkileri değerlendirilmiştir. Akciğer kanseri hastaları için yaşam çözümlemesi, K-M yöntemi ve yarışan riskler dikkate alınarak birikimli sıklık yöntemi ile yapılmıştır. Hematolojik parametrelerin kesim değerleri ise ölüm olayı temel alınarak ROC eğrisi yardımıyla belirlenmiştir. Analizler R programında "finalfit" paketi kullanılarak yapılmıştır.

Bulgular: K-M tekniğinin akciğer kanseri dışı ölümleri sansürlü veri olarak kabul etmesinden dolayı yaşam olasılıklarının daha düşük sonuçlar verdiği bulgulanmıştır.

Sonuç: Bu sebeple yaşam çözümlemelerinde tüm başarısızlık risklerinin ilgilenildiği durumlarda yarışan risk analizini kullanmanın daha isabetli sonuçlar vereceği ifade edilebilir.

K9

Türkiye'de 2009-2017 Yılları Arasında Enfeksiyon'a Bağlı Hastalık Mortalitesinde Trendlere Genel Bir Bakış

Nurhan Doğan[1], Neşe Demirtürk[2], İsmet Doğan[1]

[1]Afyonkarahisar Sağlık Bilimleri Üniversitesi Tıp Fakültesi, Biyoistatistik ve Tıbbi Bilişim Anabilim Dalı, Afyonkarahisar
[2]Afyonkarahisar Sağlık Bilimleri Üniversitesi Tıp Fakültesi, Enfeksiyon Hastalıkları Anabilim Dalı, Afyonkarahisar

Amaç: Bu çalışmanın temel amacı, 2009-2017 döneminde enfeksiyon hastalıkları mortalite oranındaki değişimi yaş ve cinsiyete göre araştırmaktır.

Metod: 2009-2017 yılları için bulaşıcı hastalık mortalite verileri Türkiye Ulusal İstatistik Enstitüsü ölüm veri tabanından alındı. Yaşa göre standartlaştırılmış mortalite oranları, WHO standart popülasyonuna göre doğrudan yöntem kullanılarak hesaplandı. Joinpoint Regresyon Analizi, zaman içinde eğilimlerdeki önemli değişiklikleri belirlemek için kullanıldı. Bu yöntem ile, değişim noktaları arasındaki yıllık yüzde değişim (APC) ve dikkate alınan dönemdeki ortalama yıllık yüzde değişim tahmin edilmektedir. Ayrıca, tahmin edilen bu değerler için % 95 güven aralıkları hesaplanmaktadır.

Sonuçlar: 2009-2017 yılları arasında Türkiye'de enfeksiyon hastalıkları nedeniyle yaklaşık 136.000 kişi hayatını kaybetmiştir. Bu kayıpların çoğu (78.000'den fazla ölüm, neredeyse% 58'i) pnömoni tanısı almış hastalardır. Enfeksiyon hastalıklarından kaynaklanan yaşa göre standartlaştırılmış mortalite oranı 9.8(2009)'den 26.3(2017)'e yükselmiştir. Erkeklerde yaşa göre standartlaştırılmış bağırsak enfeksiyon mortalite oranı 2013 yılına kadar her yıl önemli ölçüde artmıştır (APC:% 100.1, $p < 0.05$). Benzer durum kadınlarda da gözlenmiştir (APC:71.2%, $p<0.05$). 2009-2017 döneminde, hem erkek hem de kadınlarda, özellikle 2011 yılından itibaren, pnömoni, sepsis, HIV ve grip hastalıklarına ait yaşa göre standartlaştırılmış mortalite oranlarında anlamlı artışlar gözlenmiştir. Yaş gruplarına göre değerlendirildiğinde, dönem boyunca grip mortalitesi, tüm yaş gruplarında önemli ölçüde artmıştır. Pnömoni mortalite oranlarında 0-4 ve 25-44 yaş gruplarında 2011 yılına kadar azalma ve 2011'den dönem sonuna kadar önemli artma gözlenmiştir. Ayrıca, diğer yaş gruplarında dönem boyunca önemli artışlar yaşanmıştır. Sepsis mortalite oranına göre, 25-44 yaş grubu hariç tüm yaş gruplarında önemli artışlar gözlenmiş, bu artışlar özellikle 2012 yılından sonra ortaya çıkmıştır.

Tartışma: Enfeksiyonla ilişkili ölüm oranlarını azaltmak için yaşlılar için koruyucu önlemler alınmalı ve grip, zatürree ve sepsis yaşlı nüfusta önleyici politikaların odak noktası olmalıdır.

Anahtar Kelimeler: Enfeksiyon Hastalıkları, Mortalite, Joinpoint Regresyon Analizi

K9

An Overview of Trends in Infection-Related Disease Mortality in Turkey Between 2009-2017

Nurhan Doğan[1], Neşe Demirtürk[2], İsmet Doğan[1]

[1] Afyonkarahisar Health Sciences University Faculty of Medicine, Department of Biostatistics and Medical Informatics, Afyonkarahisar
[2] Afyonkarahisar Health Sciences University Faculty of Medicine, Department of Infectious Diseases, Afyonkarahisar

Objective: The main purpose of this study is to investigate the change in mortality rate of infectious diseases in 2009-2017 period according to age and gender.

Method: Infectious disease mortality data for 2009-2017 were obtained from the Turkey National Institute of Statistics death database. The age-standardized mortality rates were calculated using the direct method, according to the WHO standard population. Joinpoint Regression Analysis was used to identify significant changes in trends over time. This analysis calculates the annual percentage (APC) change in rates between change points, and it also estimates the average annual percentage change in the whole period studied. Besides, calculates 95% confidence intervals.

Results: Between the years 2009-2017, about 136,000 people died due to infectious diseases in Turkey and most of them (more than 78,000 deaths, nearly 58% of all) were patient with pneumonia The age-standardized mortality rate from infectious diseases increased from 9.8/100,000 population in 2009 to 26.3/100,000 population in 2017. The age-standardized intestinal infectious mortality rate in men was increased significantly each year until 2013 (APC:100.1%, $p<0.05$). A similar situation was valid in female (APC:71.2%, $p<0.05$). In the 2009-2017 period, statistically significant increases in age-standardized mortality rates for pneumonia, sepsis, HIV and influenza disease were observed in both men and women, especially after 2011.When evaluated according to age groups, influenza mortality increased significantly in all age groups throughout the period. Pneumonia mortality rates decreased in the 0-4 and 25-44 age groups until 2011 and significant increase from 2011 to the end of the period. In addition, significant increases were occurred in other age groups during the period. While pneumonia disease decreased in the 0-4 and 25-44 age groups until 2011, significant increases were observed after 2011, while other age groups increased significantly during the period. According to the mortality rate of sepsis, significant increases were observed in all age groups except 25-44 age group and these increases occurred especially after 2012.

Conclusion: To reduce the mortality rates associated with infection, protective measures should be taken for the elderly. Influenza, pneumonia and sepsis should be the focus of preventive policies in this population.

Keywords: Infectious diseases, Mortality, Joinpoint Regression Analysis

TAM METİN BİLDİRİLER
FULL TEXT ORAL PRESENTATIONS

T1

Gerçek Veri ve Benzetim Çalışmalarıyla Yarı En Küçük Kareler Regresyonu Yöntemi

Erdoğan ASAR[1], Erdem KARABULUT[2]

[1] Türkiye İstatistik Kurumu

[2] Hacettepe Üniversitesi Tıp Fakültesi, Biyoistatistik Anabilim Dalı

ÖZET

Amaç: Bu çalışmada, Genelleştirilmiş Kestirim Eşitliklerinin (GEE) genişletilmesi ile elde edilen Yarı En Küçük Kareler Regresyonunun (QLS) tanıtılması ve yapılan uygulamalardan elde edilen QLS ve GEE sonuçlarının karşılaştırılması amaçlanmıştır.

Yöntem: QLS, ülkemizde yaygın olarak kullanılmayan bir yöntemdir. Bu yöntem iki aşamalı bir hesaplamaya dayalı ilişkili bir veri analiz yöntemidir. Çalışmanın amaçları çerçevesinde sağlık alanına ilişkin olarak gerçek veri ve benzetim çalışması ile elde edilen veriler ile analizler yapılmış, sonuçlar değerlendirilmiştir. Gerçek veri ile yapılan çalışmada; ortodontik tedavi görmüş 38 hastanın 114 gözlemi kullanılarak, 4 sonuç değişkeni üzerinde zaman ve grup değişkenlerinin etkileri araştırılmıştır. Benzetim çalışmasında, belirlenen 3 korelasyon yapısı ve 3 farklı değeri için 1000 tekrarlı 9 veri seti üretilmiş, bu veri setleri üzerinde 4'er adet çalışan korelasyon yapısı uygulanarak elde edilen 36 durumlu sonuçlar değerlendirilmiştir.

Bulgu ve Sonuçlar: Çalışmalar sonucunda, benzetim çalışmasına göre; kestirimlerin etkinliği bakımından genel olarak QLS'nin GEE'ye göre üstünlük gösterdiği, GEE ve QLS sonuçları karşılaştırmalarının daha iyi yapılabilmesi için "Markov" çalışan korelasyon yapısının QLS'de olduğu gibi GEE'de de uygulanabilir olması gerektiği, gerçek veri ile yapılan uygulama kapsamında; QLS'de kullanılan çalışan korelasyon yapıları içinde en yüksek korelasyon değerinin "Markov" çalışan korelasyon yapısı ile elde edildiği, bütün sonuç değişkenleri üzerinde grup değişkeni olarak yer alan tek çene cerrahisi ve çift çene cerrahisinin bir öneminin olmadığı, hem gerçek veri hem de benzetim ile elde edilen veri setlerinde; "Tri-diagonal" çalışan korelasyon yapısının uygulanmasında GEE için yakınsama sorunu oluşurken, QLS'de bu sorunun oluşmadığı sonuçları elde edilmiştir.

Anahtar Kelimeler: Regresyon, Genelleştirilmiş Kestirim Eşitlikleri, Yarı En Küçük Kareler Regresyonu, Markov Korelasyon Yapısı.

1.GİRİŞ

Araştırmacılar, gözlem birimlerinden (birim) kaynaklanan hataları azaltmak, gözlem birimlerinin zaman içindeki değişimlerini görmek vb. amaçlar için aynı gözlem birimi üzerinde birden fazla ölçüm (tekrarlı ölçümler) yapabilmektedirler. Boylamsal veri olarak adlandırılan bu tür veriler, kümelenmiş bir veri yapısındadır. Kümeler, tek bir gözlem biriminin farklı zaman dilimlerinde ya da farklı durum ve koşullarda tekrarlanan ölçümlerinden oluşur. Tekrarlı ölçümlü verilerin veri derlemesi zor, zaman alıcı ve maliyet bakımından da yüksektir. Ölçümler arasındaki sürelerin eşit olması beklenir, fakat bunun her zaman gerçekleşmesi olası değildir. Ayrıca tekrarlı ölçümler söz konusu olduğunda bağımsızlık varsayımı sağlanamamaktadır. Küme içindeki gözlemler bir korelasyona sahiptirler ve bu korelasyon analizde göz ardı edilmemelidir. Geleneksel regresyon modellerinde, hataların dağılımının normal olması, bağımlı değişkeninin alt kümelerinin varyanslarının homojen olması gibi varsayımların sağlanması gerekmektedir. Oysaki boylamsal veriler söz konusu olduğunda bu varsayımlar sağlanamayabilir. Bu yüzden boylamsal verilerin analizi, klasik yaklaşımlara dayanan modeller ile yapılamamaktadır. Bu tür verileri analiz etmek

için özel modeller geliştirilmiştir (1, 2, 4-7). Regresyon analizi, değişkenler arası bağımlılık yapısının incelenmesi ile ilgili yöntemlerden biridir. Bu analizin en önemli unsuru da değişkenler arası fonksiyonel ilişkiyi gösteren regresyon denklemi ya da başka bir ifadeyle de regresyon modelidir.

Genelleştirilmiş Doğrusal Modeller (Generalized Linear Models - GLM), geleneksel regresyon modellerinin bir uzantısıdır. **Genelleştirilmiş Kestirim Eşitlikleri (Generalized Estimating Equations - GEE)** ise GLM'nin genişletilmiş şeklidir ve özellikle Biyoistatistik alanında boylamsal verilerin modellenmesinde yaygın olarak kullanılan yöntemlerden biridir. GLM ve GEE'de regresyon modeline ilişkin kestirimler, yanıt değişkeninin üstel dağılım ailesinden geldiği durumlarda yapılır. GEE, ilişkili verilerin analizine olanak sağlamaktadır ve başka bir ifadeyle de regresyon parametrelerinin kestiriminde küme içi korelasyonu dikkate almaktadır. Bu durum GLM ile arasındaki en önemli farklılıktır.

Yarı En Küçük Kareler Regresyonu (Quasi-Least Squares Regression - QLS), GEE yaklaşımı çerçevesinde Shults ve Chaganty tarafından 1996-1999 yılları arasında geliştirilen bir ilişkili veri analiz yöntemidir. Bu yöntemde, korelasyon parametrelerinin kestirimi iki aşamalı olarak yapılmaktadır. QLS ile regresyon modeli kurulurken, tekrarlı ölçümler arasındaki ilişkinin modele yansıtılmasında kullanılabilecek korelasyon yapılarından olan Markov korelasyon yapısının, tekrarlı ölçümlerin eşit olmayan zaman aralıklarında yapıldığı ve eksik gözlemlerin olduğu durumlarda da kullanılabilmesi QLS'ye GEE'ye göre üstünlük sağlamaktadır (1, 6, 8-10).

2. METODOLOJİ

"QLS ve GEE'nin klasik yaklaşımlardan farkı, aynı örnek (gözlem birimi) üzerinde yapılan tekrarlı ölçümler arasındaki korelasyonu göz önüne almasıdır.

2.1. Çalışan Korelasyon Matris Yapıları

Korelasyon (ilişki) katsayıları, değişkenler arasındaki ilişkinin kuvveti (derecesi) ve yönü hakkında bilgi veren ölçülerdir (25). Çalışan korelasyon yapısı ($R_i(\alpha)$), kestirimle elde edilen "α" korelasyon katsayısına bağlı olarak i. gözlem biriminin ölçümleri arasındaki ilişkinin örüntüsünü tanımlamaktadır. Bu yapı ilişkili verinin modellenmesinde kullanılmaktadır. Başka bir ifadeyle $R_i(\alpha)$, "i" gözlem birimi üzerindeki ölçümler arasındaki korelasyonu içeren matristir (1, 17). Araştırmacılar, üzerinde çalıştıkları verinin durumuna göre uygun çalışan korelasyon yapısını belirlemeye çalışırlar. Çalışan korelasyon yapısının yanlış belirlenmesi durumunda, regresyon parametre kestirimlerinin etkinliği azalacaktır (2). Bu bakımdan, çalışan korelasyon yapılarının özelliklerini bilmek araştırmacılar için önem taşımaktadır.

Alanyazında geçen bazı çalışan korelasyon yapıları aşağıda belirtilmiş ve açıklanmaya çalışılmıştır (1- 5, 15, 17, 18).

- **Değiştirilebilir Korelasyon Yapısı**

"Değiştirilebilir çalışan korelasyon yapısı" küme içi ölçümler (aynı birime ait tekrarlı ölçümler) arasındaki korelasyonun tüm ölçümler arasında aynı olduğu varsayımına dayanmaktadır. Bu korelasyon yapısı kesitsel çalışmalardan elde edilen kümelenmiş veriler için uygundur.

$$j = k \quad \text{ise} \qquad Corr(y_{ij}, y_{ik}) = 1$$

$$j \neq k \quad \text{ise} \qquad Corr(y_{ij}, y_{ik}) = \alpha$$

Bu yapıda elde edilen korelasyon kestiriminin değer aralığı $(-1/(n_{max}-1), 1)$ aralığıdır. Burada i=1, 2, …, m küme olmak üzere, n_{max}; n'nin aldığı en yüksek değerdir.

- **Yapılandırılmamış Korelasyon Yapısı**

Bu korelasyon yapısında, korelasyon için varsayılan bir örüntü yoktur. Bu yapıda ölçümlerin birbirleriyle olan korelasyonu dikkate alınmaktadır.

$j = k$ ise $Corr(y_{ij}, y_{ik}) = 1$

$j \neq k$ ise $Corr(y_{ij}, y_{ik}) = \alpha_{jk}$

Bu korelasyon yapısı birim (küme) sayısının fazla ve ölçümler arasındaki korelasyonun aynı olduğu düşünülen veri setleri için kullanımı uygundur. Birimler (kümeler) üzerinde yapılan ölçüm sayısı fazla olduğunda korelasyon matrisinin boyutu büyük olmaktadır. Bu durumda, parametre kestirimlerinde yakınsama sorununun sıklıkla yaşanabilmektedir. Bu da korelasyon yapısının uygulanmasında çekinceler oluşturmasına neden olmaktadır.

- **Tri-diagonal Korelasyon Yapısı**

Bu yapıda, bir birimin ölçümleri arasındaki korelasyon ölçüm sıraları arasındaki farka göre belirlenmektedir;

$j = k$ ise $Corr(y_{ij}, y_{ik}) = 1$

$|j - k|$=1 ise $Corr(y_{ij}, y_{ik}) = \alpha$

diğer durumlarda $Corr(y_{ij}, y_{ik}) = 0$.

"Stationary 1-dependent" olarak da adlandırılan bu korelasyon yapısı, uygulamada az kullanılır. Korelasyonun alabileceği değer ($-1/c_{max}$, $1/c_{max}$) arasındadır. Bu aralık yaklaşık olarak ($-1/2$ ile $1/2$) arasında değişmektedir,

$$C_{max} = 2sin\left(\frac{\pi|n_{max}-1|}{2[n_{max}+1]}\right) .$$

- **Birinci Dereceden Otoregresif Korelasyon Yapısı (AR-1)**

AR-1, *s*ağlık alanında zamana bağlı tekrarlı ölçümler olduğunda sıklıkla kullanılan bir korelasyon yapısıdır. Bu korelasyon yapısında ilişkinin örüntüsü ölçüm sırasına bağlıdır. Çalışmadaki ölçümlerin zamana bağlı olduğu ve eşit zaman aralıklarında yapıldığı durumlarda kullanılması uygun olan bir korelasyon yapısıdır. AR-1'de korelasyon, (-1,1) aralığında değer alır. Ölçümler arasındaki ilişkinin örüntüsü;

$j = k$ ise $Corr(y_{ij}, y_{ik}) = 1$

$j \neq k$ ise $Corr(y_{ij}, y_{ik}) = \alpha^{|j-k|}$

olarak ifade edilir. Buradan da anlaşılabileceği üzere, bu örüntünün temeli; yakın zaman aralıklarındaki ölçümler arasında yüksek korelasyon, uzun zaman aralıklarıyla elde edilen ölçümler arasında da daha düşük korelasyon olduğu esasına dayanmaktadır.

- **Markov Korelasyon Yapısı**

Markov korelasyon yapısı, AR-1'in zaman aralıkları eşit olmayan ölçümler için genelleştirilmiş bir halidir. Eşit aralıklı olmayan düzensiz ölçümler söz konusu olduğunda kullanılması uygun olan bir korelasyon yapısıdır. Bu çerçevede ölçümler arasındaki ilişkinin örüntüsü, t ölçüm yapılan gerçek zaman olmak üzere;

$j = k$ ise $Corr(y_{ij}, y_{ik}) = 1$

$j \neq k$ ise $Corr(y_{ij}, y_{ik}) = \alpha^{|t_{ij}-t_{ik}|}$

GEE için uygulama olanağı olmayan Markov korelasyon yapısı QLS için uygulanabilmektedir. Bu uygulama üstünlüğü QLS'ye bir ayrıcalık tanımaktadır. Bu korelasyon yapısının değer aralığı (-1,1)'dir.

2.2. Genelleştirilmiş Kestirim Eşitliği (GEE)

GEE, Liang ve Zeger (1986) tarafından sonuç değişkenin tekrarlı ölçümleri arasındaki korelasyonu göz önüne alarak GLM'den geliştirilmiş bir kestirim yöntemidir. Dengeli olmayan verilerde de kullanılabilen bu yöntem, boylamsal verilerin etkin ve yansız regresyon tahminlerini üretmek amacıyla geliştirilmiştir. Geleneksel regresyon modellerindeki varsayımların sağlanamadığı durumlarda uygulanabilmektedir. GEE'yi GLM'den ayıran en önemli özellik; regresyon parametrelerinin kestiriminde gözlem birimlerinin her biri (küme) içindeki tekrarlı ölçümlere ilişkin korelasyonu dikkate almasıdır. GEE ile regresyon parametrelerinin kestiriminde, gözlem birim içi ölçümlere ilişkin korelasyon göz önüne alınarak model kurulur ve kestirimler elde edilir. Ölçümler arasındaki ilişkinin örüntüsünü veren korelasyon matrisi $Corr(Y_i)$ ile gösterilir.

GLM kestirim eşitliği yardımıyla GEE kestirim eşitliğine aşağıdaki gibi ulaşılabilir:

i: 1,2, … … … , m gözlem birimlerini,

j: 1,2, … … … . , n_i ölçümleri göstermek üzere;

$Y_i = \left(y_{i1}, \ldots \ldots \ldots . , y_{in_i}\right)'$ her bir "i" gözlem biriminden elde edilen ölçümü,

$x_{ij} = \left(x_{ij1}, \ldots \ldots \ldots . , x_{ijp}\right)'$ y_{ij} ile ilişkili ortak değişkenleri göstersin.

$E(Y_i) = \left(\mu_{i1},\ \mu_{i2}, \ldots \ldots . , \mu_{in_i}\right)' = \mu_i$ 'dir.

$\emptyset A_i$; j. diagonal elemanı y_{ij}'nin varyansı olan diagonal matrisi göstersin;

$\emptyset A_i = diag\left(var(y_{i1}), \ldots \ldots , var\left(y_{in_i}\right)\right) = \emptyset diag(h(\mu_{i1}), \ldots . , h\left(\mu_{in_i}\right))$.

$D_i = \frac{\partial \mu_i}{\partial \beta} = \left(\frac{\partial \mu_{ij}}{\partial \beta_k}\right)$ olsun.

Bu eşitlikler çerçevesinde GLM için regresyon parametresi kestirim eşitliği aşağıdaki biçime dönüşmüştür:

$$\sum_{i:1}^{m} D_i{}' A_i^{-1}(Y_i - \mu_i) = 0. \qquad (2.1)$$

Bu eşitlik, her gözlem birimi için korelasyon matrislerinin bir fonksiyonu olarak ifade edilebilecektir. GLM'de ölçümler bağımsızdır. Bu nedenle, bağımsız olan "i" gözlem birimine ilişkin korelasyon matrisi $Corr(Y_i)$, birim matris (I_{n_i} ; $n_i x n_i$) olacaktır. $Corr(Y_i)$ i tane gözlem birimindeki n_i bağımsız ölçümler arasındaki ilişkinin örüntüsünü verecektir (i: 1,2, … . . , n_i). Buradan aşağıdaki eşitlik yazılabilir;

$$A_i = A^{1/2} I_{n_i} A^{1/2} = A^{1/2} Corr(Y_i) A^{1/2}$$

GLM kestirim eşitliği de aşağıdaki şekli almış olur (1-3, 5, 17, 18);

$$\sum_{i:1}^{m} D_i{}' A_i^{-\frac{1}{2}} Corr(Y_i\ )^{-1} A_i^{-\frac{1}{2}}((Y_i - \mu_i) = 0. \qquad (2.2)$$

Bu eşitlikte yer alan korelasyon matrisi $Corr(Y_i)$'nin yerini, *"α"* korelasyon parametresine bağlı olan $R_i(\alpha)$ çalışan korelasyon matrisi (working correlation matrix) alacaktır. Aşağıdaki kısımda ele alınacak olan $R_i(\alpha)$'nin kestirimler için dikkate alınması ile ölçümler arasındaki ilişkinin de modelde yer alması sağlanmış olmaktadır. Bu bilgi çerçevesinde GEE kestirimi aşağıdaki eşitliğe dönüşmüştür (1, 2, 17, 19-24).

$$\sum_{i:1}^{m} D_i' A_i^{-\frac{1}{2}} R_i^{-1}(\alpha) A_i^{-\frac{1}{2}} (Y_i - \mu_i) = 0 \quad . \qquad (2.3)$$

2.3. Yarı En Küçük Kareler Regresyonu (QLS)

Yarı En Küçük Kareler Regresyonu, bir ilişkili veri analiz yöntemidir. Bu yöntem, GEE çerçevesinde geliştirilmiştir. Burada korelasyon ve regresyon parametreleri aşamalı kestirilmektedir (8). Modellemede, eşit aralıklı olmayan ölçümlerin kullanımına olanak sağlayan Markov korelasyon yapısı kullanılabilmektedir. Bu yönteme ilişkin ilk çalışmalar, 1996-1999 yılları arasında Shults ve Chaganty tarafından bir seri akademik çalışmalarla yapılmıştır. İlk olarak dengeli verilerde korelasyon parametresinin I. Aşama kestirimi ile ilgili çalışmalar yapılmıştır. Daha sonra I. Aşama parametre kestirimi, eşit zaman aralıklı ölçümler için dengeli olmayan verilerde kullanımına uygun hale getirilmiştir. Bu çalışmalardan sonra da korelasyon parametrelerinin kestirimi için II. Aşama geliştirilmiştir (1, 7, 8, 11-13, 14, 16, 17, 26).

QLS'de Regresyon ve Korelasyon Parametrelerinin Kestirim İşlem Süreci

QLS ile parametre kestirimleri yapılırken iki aşamalı bir süreç izlenir. Bu süreç yakınsama temeline dayanmaktadır (7, 11, 12, 17, 26);

- **I. Aşama'da** ilk olarak, çalışan korelasyon yapısının parametresi olan "α" için bir başlangıç değeri belirlenir. Alanyazındaki çalışmalarda sıfır ile başlandığı görülmüştür; $\hat{\alpha} = 0$.
- Daha sonra I. Aşama için A ve B adımları sıra ile uygulanır.
  - **A adımında;**
    - $\boldsymbol{\beta}$ **regresyon parametresinin kestirimi** GEE kestirim eşitliğinin (Eşitlik 2.3) çözümü ile elde edilir;
    $$\sum_{i:1}^{m} D_i{}' A_i^{-\frac{1}{2}} R_i^{-1}(\hat{\alpha}) A_i^{-\frac{1}{2}} (Y_i - \mu_i) = 0 .$$
    - **Pearson artıkları** hesaplanır; i. gözlemin j. Pearson artığı;
    $$z_{ij} = \left(y_{ij} - \mu_{ij}\right) \Big/ \sqrt{h(\mu_{ij})}$$
  - **B adımında; "$\hat{\alpha}$ değeri"** elde edilen β kestirim değeri ile güncellenir;
  $$\sum_{i:1}^{m} Z_i{}'(\hat{\beta}) \frac{\partial R_i^{-1}(\hat{\alpha})}{\partial \hat{\alpha}} Z_i(\hat{\beta}) = 0$$
  - B adımından sonra **A Adımına dönülür**. Bu işlemler α ve β parametre kestirimleri bir değere yakınsayana kadar devam ettirilir. Yakınsama olduğu görüldüğünde ikinci Aşamaya geçilir.
- **II. Aşamada** iki alt süreç vardır;
  - İlk olarak I. Aşama'da elde edilen $\hat{\boldsymbol{\alpha}}$ kullanılarak α 'nın son kestirimi (α_{QLS}) elde edilir;
  $$\sum_{i:1}^{m} trace\left(\frac{\partial R_i^{-1}(\hat{\alpha})}{\partial \hat{\alpha}} R_i(\hat{\alpha})\right) = 0$$
  - I. Aşama'nın A adımına dönülür. Korelasyon parametresinin II. Aşama kestirimi olan α_{QLS} kullanılarak $\boldsymbol{\beta}$ **'nın son kestirimi** elde edilir;
  $$\sum_{i:1}^{m} D_i' A_i^{-\frac{1}{2}} R_i^{-1}(\alpha_{QLS}) A_i^{-\frac{1}{2}} (Y_i - \mu_i) = 0$$

3.GEREÇ ve YÖNTEM

3.1. Gerçek Veri

Çalışmanın amaçları kapsamında, diş hekimliği alanına ilişkin veri kullanılarak bir uygulama yapılmıştır.

- **Kapsam**

Uygulamada, 2000-2018 yılları arasında Hacettepe Üniversitesi Diş Hekimliği Fakültesi Ortodonti Anabilim Dalında ortodontik tedavi gören,"İskeletsel Sınıf III" nedeniyle tek çene (mandibuler geriletme) veya çift çene (maksiller ilerletme ve mandibuler geriletme) cerrahisi uygulanmış, 18 yaş ve üzeri 34 hastanın lateral sefalometrik radyografları üzerinden yapılan 114 ölçüm kullanılmıştır. Herhangi bir sendromu bulunan bireyler çalışmada kapsam dışı tutulmuştur.

- **Veri Derleme Yöntemi**

Veriler, hastalara ait daha önceki dönemlerde tedavi sürecinde çekilmiş görüntüler üzerinden ölçümler yapılarak elde edilmiştir. Bu çalışma çerçevesinde veri derlenebilmesi için Etik Kurul onayı alınmıştır. Çalışma kapsamına alınan 38 hastanın lateral sefalometrik radyografları üzerinden aşağıda belirtilen dört değişkene ilişkin teknik ölçümler yapılmıştır;

- cme: C noktası menton mesafesi (mm),
- sda: Servikal düzlem açısı (°),
- cpogp: C noktasından pogonion dikmesine olan mesafe (mm),
- sfda: Servikal düzlem-fasiyal düzlem arasındaki açı (°).

Veride yukarıdaki belirtilen değişkenlere ilişkin 114 ölçüm yer almaktadır. Bu ölçümlere ek olarak hastalara uygulanan cerrahi yöntem ve radyografların çekim tarihleri de kaydedilmiştir. Ölçümlerin yapıldığı radyografların çekim tarihleri "zaman" değişkeni olarak alınmıştır. Zaman birimi "ay" olarak kullanılmıştır. İlk radyografın çekildiği zaman "0" olarak alınan "referans" ay'dır. Diğer ölçümler ise ilk ölçümden sonraki aylardaki radyograflardan yapılan ölçümler olup, 1-23 ay aralığında elde edilmiştir. Hastalara uygulanan cerrahi yöntem ise "grup" değişkeni olarak alınmıştır. İki grup söz konusudur;

- Grup 1: Tek çene cerrahisi uygulanan hastalar,
- Grup 2: Çift çene cerrahisi uygulanan hastalar.

Grup 1'de 14 hastadan (kümeden) elde edilen 42 ölçüm (gözlem) bulunurken, Grup 2'de ise 24 hastanın 72 ölçümü yer almaktadır.

Elde edilen veride;

- "zaman" ve "grup" değişkenleri bağımsız değişkenler,
- "cme", "cpogp", "sda", "sfda" değişkenleri sonuç değişkenleri

olarak belirlenmiştir. Her bir sonuç değişkeni için bağımsız değişkenler dikkate alınarak regresyon modeli kurulmuştur. Ayrıca, sonuç değişkeni üzerinde zaman ve grup etkileşiminin önemli olup-olmadığı incelenmiştir. Etkileşim (zaman×grup) teriminin, GEE ve QLS yöntemlerinin her ikisinde de sadece "cme" değişkeninde "Değiştirilebilir" çalışan korelasyon yapısı kullanılarak tasarlanan regresyon modellerinde önemli olduğu görülmüştür. Bu nedenle, regresyon modelleri kurulurken "cme" değişkeni için "Değiştirilebilir" çalışan korelasyon yapısı seçildiğinde etkileşim terimi modelde yer almıştır.

Y; sonuç değişkeni,

β' lar regresyon katsayıları olmak üzere her bir değişken için:

$$Y_{cme} = \beta_0 + \beta_1 \times zaman + \beta_2 \times grup$$

"cme" değişkeninde "Değiştirilebilir" çalışan korelasyon yapısı için;

$$Y_{cme} = \beta_0 + \beta_1 \times zaman + \beta_2 \times grup + \beta_3 \times (zaman \times grup)$$

$$Y_{cpogp} = \beta_0 + \beta_1 \times zaman + \beta_2 \times grup$$

$$Y_{sda} = \beta_0 + \beta_1 \times \text{zaman} + \beta_2 \times \text{grup}$$
$$Y_{sfda} = \beta_0 + \beta_1 \times \text{zaman} + \beta_2 \times \text{grup}$$

Bu aşamadan sonra veri analiz yönteminin belirlenmesine geçilmiştir. Veri incelendiğinde; aynı gözlem birimi (hasta) üzerinde tekrarlı ölçümler vardır, ölçüm aralıkları eşit değildir, eksik ölçümler söz konusudur. Verinin bu yapısı nedeniyle, veri analiz yöntemi olarak GEE ya da QLS kullanılabilecektir. Veri analizi, Stata (Sürüm 14.1) yazılımı ile yapılmıştır. GEE ve QLS için belirlenen çalışan korelasyon yapılarına göre regresyon modelleri kurulmuş ve yorumlar yapılmıştır.

GEE için; Sonuç değişkenleri üzerinde etkili olan bağımsız değişkenlerin belirlenmesinde "xtgee" komutu, "Değiştirilebilir"", "AR-1" ve "Tri-diagonal" çalışan korelasyon yapılarına göre çalışan korelasyon matrisinin elde edilmesinde "xtcorr" komutu kullanılmıştır.

QLS için; Sonuç değişkenleri üzerinde etkili olan bağımsız değişkenlerin belirlenmesinde "xtqls" komutu, "Değiştirilebilir", "AR-1", "Tri-diagonal" ve "Markov" çalışan korelasyon yapılarına göre çalışan korelasyon matrisinin elde edilmesinde "xtcorr" komutu kullanılmıştır.

3.2. Benzetim Çalışması

Bu çalışmanın amaçlarından biri de "QLS ile yapılan regresyon parametre kestirimlerinin GEE'den elde edilen kestirimlerden daha etkin kestirimler olduğunun gösterilmesi"dir. Bu amaç çerçevesinde bir benzetim çalışması yapılmıştır.

- **Veri Üretimi**

Benzetim ile dokuz veri seti üretilmiş, otuz altı durumlu sonuçlar elde edilmiştir. Benzetim çalışmasında R (sürüm 3.2.5) yazılımı kullanılarak iki aşamadan oluşan fonksiyonlar yazılmıştır. Bu fonksiyonlar içerisinde istenen parametreler girildikten sonra veri üretimi gerçekleştirilmiştir (28).

- İlk olarak, üretilecek veri setinin gerçek korelasyon değerlerine ve yapılarına karar verilmiştir; korelasyon değerleri olarak α= 0,5, α= 0,7 ve α= 0,9 olarak alınmıştır, her bir korelasyon değerine göre "Değiştirilebilir", "AR-1" ve "Markov" korelasyon yapılarına göre veri üretilmiştir.
- Her bir veri setinin 1000 tekrarlı olarak üretilmesi istenmiştir.
- Birimlerin ölçüm sayısı 4 olarak alınmıştır. Zaman değişkeni için 4 tekrara ilişkin veriler ölçüm aralıkları eşit olmayacak şekilde sırasıyla; 0, 1-3, 4-8, 9-15 aralıklarında üretilmiştir.
- Eksik veri oranı %5 olarak alınmıştır.
- Verinin 2 gruptan oluşması ve bu grupların dengeli bir yapıda olarak 20'şer birimli olması istenmiştir. Böylece, her bir veri setinde en fazla 160 gözlemli bir veri oluşmuştur.
- Üretilen veride sonuç değişkeninin dağılımının normal olması istenmiştir.
- Sonuç değişkeni ile bağımsız değişkenler arasındaki matematiksel bağıntının kurulmasını sağlayan regresyon katsayılarının belirlenmesinde, gerçek verideki değerler göz önüne alınmıştır.
- Yukarıda belirtilen parametreler, R'de yazılan fonksiyona girilerek her bir senaryo/durum için 1000 adet farklı veri seti elde edilmiştir (30).

- **Veri Analizi**

R yazılımı kullanılarak benzetim çalışması ile üretilen 1000 adet farklı her veri seti için, belirlenen her bir çalışan korelasyon yapıları dikkate alınıp, GEE ve QLS'ye göre regresyon modelleri kurulmuştur. Her bir veri seti için, Y sonuç değişkeni, zaman ve grup bağımsız değişkenler, β' lar regresyon katsayıları olmak üzere kurulan model aşağıda verilmiştir;

$$Y = \beta_0 + \beta_1 \times \text{zaman} + \beta_2 \times \text{grup}$$

Modelin kurulması ve analizlere altlık oluşturacak sonuçlar Stata (sürüm 14.1) yazılımı kullanılarak elde edilmiştir. 1000 adet farklı veriden oluşan her bir veri setinden elde edilen sonuçlar, hesaplama ve analizlere kolaylık sağlamak için alt alta getirilmiştir. Bu kısımda veri analizinde kullanılan komutlar, Bölüm 3.1'de belirtilen gerçek veri uygulamasındakilere benzerdir. 1000 adet farklı veri setleri ile çalışıldığından dolayı, modelin kurulması ve sonuçların alınması için ek bazı komutlar ile mevcut komutlar güncellenerek tamamlanmıştır. Kurulan regresyon modelleri ile regresyon katsayıları, hata kareler ortalaması (MSE) ve p değerleri vb. elde edilmiştir (27).

- **Regresyon Katsayılarının Etkinlik Hesabı**

Regresyon analizinin amaçları arasında regresyon katsayılarının kestirimini yapmak da yer almaktadır. Kestirimler yansız, tutarlı ve küçük varyansa sahip olmalıdır (29). Aslında bu kestirimler regresyon çalışmalarının temelini oluşturmaktadır. Çeşitli veri analizi yöntemleri ile elde edilen kestirimlerin hangilerinin daha kabul edilebilir olduğu konusu bilim insanlarının kafasını meşgul etmiştir. Regresyon katsayılarının etkinlik çalışması ile bu sorunun yanıtı bulunabilmektedir. Çalışmamız kapsamında, GEE ve QLS için kurulan regresyon modelleri ile elde edilen bilgiler kullanılarak, 1000 tekrarlı veri setlerinin her biri için regresyon katsayılarının etkinlik hesapları yapılmıştır. Elde edilen değerlere göre, GEE ve QLS ile kestirilen regresyon katsayılarının hangilerinin daha iyi kestirim olduklarının karşılaştırması olanağı sağlanmıştır. Etkinlik hesabı için kullanılan ölçütler aşağıda verilmiştir (1, 30). R: Tekrar (Benzetim) sayısı olmak üzere;

- Yanlılık:

$$\frac{\sum_{i=1}^{R}(\beta-\widehat{\beta}_i)}{R}$$

- Hata Kareler Ortalaması (MSE):

$$\frac{\sum_{i=1}^{R}(\beta-\widehat{\beta}_i)^2}{R}$$

- Etkinlik:

$$\frac{MSE_{GEE}}{MSE_{QLS}}$$

Bu oran "1"den büyükse; MSE_{QLS}'nin MSE_{GEE}'den daha küçük olduğu anlaşılır. Dolayısıyla, QLS kestiriminin GEE kestirimine göre daha etkin, daha üstün bir kestirim olduğu sonucuna varılır.

4.BULGULAR

4.1.Gerçek Veri Setine İlişkin Bulgular

Gerçek veri seti kullanılarak GEE ve QLS ile analizleri yapılan çalışmada, "cme", "cpogp", "sda", "sfda" sonuç değişkenleri üzerinde "zaman" ve "grup" bağımsız değişkenlerinin anlamlı bir etkisinin olup olmadığı ortaya konan amaçlar çerçevesinde araştırılmıştır. GEE ve QLS'ye göre tüm yanıt değişkenleri için oluşturulan regresyon modellerine ilişkin sonuçların bir bütün halinde görülebileceği tablo aşağıda verilmiştir (Tablo 4.1). Aşağıda verilen, genel bir fikir edinmemizi sağlayan bu tablodan çıkarılacak üç önemli sonuç görülmektedir;

1- "cme" ve "cpogp" yanıt değişkenleri üzerinde "zaman" bağımsız değişkeni "Tri-diagonal" korelasyon yapısı dışında önemli, "grup" bağımsız değişkeni ise tüm korelasyon yapılarında önemsizdir,
2- Zaman ve grup etkileşimi sadece "cme" yanıt değişkeninde "Değiştirilebilir" korelasyon yapısı için anlamlıdır,

3- "sda" ve "sfda" yanıt değişkenleri için tüm korelasyon yapılarında her iki bağımsız değişken de önemsiz bir etkiye sahiptir.

Tablo 4.1. GEE ve QLS yöntemlerine göre kurulan regresyon modeli sonuçları

		Regresyon Katsayıları								Standart Hata								p değeri							
	Çalışan	GEE				QLS				GEE				QLS				GEE				QLS			
Sonuç	Korelasyon				zaman x				zaman x				zaman x				zaman x				zaman x				zaman x
Değişkeni	Yapısı	sabit	zaman	grup	grup	sabit	zaman	grup	grup	sabit	zaman	grup	grup	sabit	zaman	grup	grup	sabit	zaman	grup	grup	sabit	zaman	grup	grup
cme	Değiştirilebilir	38,658	-0,490	-1,198	0,168	38,650	-0,488	-1,194	0,166	4,775	0,118	2,807	0,070	4,754	0,141	2,794	0,084	**0,000**	**0,000**	0,670	**0,016**	**0,000**	**0,001**	0,669	0,047
	AR 1	37,686	-0,152	-0,450	(***)	37,673	-0,154	-0,458	(***)	4,744	0,037	2,786	(***)	4,684	0,048	2,749	(***)	**0,000**	**0,000**	0,870	(***)	**0,000**	**0,001**	0,868	(***)
	Tri-diagonal	(*)	(*)	(*)	(*)	39,008	-0,025	-0,914	(***)	(*)	(*)	(*)	(*)	3,105	0,097	1,780	(***)	(*)	(*)	(*)	(*)	**0,000**	0,798	0,608	(***)
	Markov	(**)	(**)	(**)	(**)	38,606	-0,295	-0,539	(***)	(**)	(**)	(**)	(**)	4,605	0,076	2,694	(***)	(**)	(**)	(**)	(**)	**0,000**	**0,000**	0,841	(***)
sda	Değiştirilebilir	119,832	-0,162	3,937	(***)	119,831	-0,162	3,937	(***)	8,532	0,090	5,008	(***)	8,517	0,092	4,999	(***)	**0,000**	0,072	0,432	(***)	**0,000**	0,079	0,431	(***)
	AR 1	120,705	-0,179	3,470	(***)	120,629	-0,174	3,497	(***)	8,498	0,098	4,986	(***)	8,323	0,120	4,878	(***)	**0,000**	0,069	0,486	(***)	**0,000**	0,146	0,473	(***)
	Tri-diagonal	(*)	(*)	(*)	(*)	121,468	-0,062	2,468	(***)	(*)	(*)	(*)	(*)	6,049	0,193	3,481	(***)	(*)	(*)	(*)	(*)	**0,000**	0,748	0,478	(***)
	Markov	(**)	(**)	(**)	(**)	119,077	-0,108	4,301	(***)	(**)	(**)	(**)	(**)	7,931	0,190	4,612	(***)	(**)	(**)	(**)	(**)	**0,000**	0,571	0,351	(***)
cpogp	Değiştirilebilir	52,677	-0,155	0,234	(***)	52,670	-0,156	0,232	(***)	3,525	0,026	2,070	(***)	3,493	0,032	2,051	(***)	**0,000**	**0,000**	0,910	(***)	**0,000**	**0,000**	0,910	(***)
	AR 1	52,726	-0,099	0,299	(***)	52,731	-0,102	0,294	(***)	3,504	0,034	2,057	(***)	3,469	0,039	2,035	(***)	**0,000**	**0,004**	0,884	(***)	**0,000**	**0,009**	0,885	(***)
	Tri-diagonal	(*)	(*)	(*)	(*)	52,278	-0,073	0,381	(***)	(*)	(*)	(*)	(*)	2,485	0,079	1,429	(***)	(*)	(*)	(*)	(*)	**0,000**	0,353	0,790	(***)
	Markov	(**)	(**)	(**)	(**)	53,643	-0,210	0,061	(***)	(**)	(**)	(**)	(**)	3,364	0,063	1,964	(***)	(**)	(**)	(**)	(**)	**0,000**	**0,001**	0,975	(***)
sfda	Değiştirilebilir	89,100	0,059	0,094	(***)	89,098	0,059	0,094	(***)	4,221	0,047	2,477	(***)	4,190	0,051	2,458	(***)	**0,000**	0,208	0,970	(***)	**0,000**	0,241	0,969	(***)
	AR 1	88,780	0,017	0,217	(***)	88,782	0,017	0,216	(***)	4,135	0,060	2,424	(***)	4,124	0,061	2,417	(***)	**0,000**	0,781	0,929	(***)	**0,000**	0,777	0,929	(***)
	Tri-diagonal	(*)	(*)	(*)	(*)	87,892	0,031	0,370	(***)	(*)	(*)	(*)	(*)	3,121	0,102	1,798	(***)	(*)	(*)	(*)	(*)	**0,000**	0,764	0,837	(***)
	Markov	(**)	(**)	(**)	(**)	87,880	0,133	0,434	(***)	(**)	(**)	(**)	(**)	3,919	0,096	2,278	(***)	(**)	(**)	(**)	(**)	**0,000**	0,166	0,849	(***)

Notlar:1- (*): GEE'de yakınsama olmadığından dolayı hesaplanamıyor.

2- (**): GEE'de uygulanamıyor.

3- (***): Sonuç değişkeni üzerinde zaman ve grup etkileşimi önemli olmadığından dolayı regresyon modeline alınmamıştır.

Kullanılan çalışan korelasyon yapılarına göre elde edilen korelasyon değerlerine ilişkin düzenlenen tablo aşağıda verilmiştir (Tablo 4.2). Bu tabloya göz atıldığında; her bir sonuç değişkeni itibarıyla, seçilen çalışan korelasyon yapılarına göre elde edilen korelasyon değerleri arasında en yüksek değerin, Markov korelasyon yapısı kullanılarak elde edilen değer olduğu görülmektedir.

Tablo 4.2. Sonuç değişkenleri ve çalışan korelasyon yapılarına göre QLS'de elde edilen korelasyon değerleri.

Sonuç	**Çalışan Korelasyon Yapısı**			
Değişkeni	Değiştirilebilir	AR-1	Tri-diagonal	Markov
cme	0,923	0,931	0,660	**0,980**
sda	0,883	0,867	0,628	**0,954**
cpogp	0,916	0,915	0,635	**0,973**
sfda	0,856	0,858	0,603	**0,951**

4.2. Benzetim Çalışması Bulguları

GEE ve QLS ile elde edilen $\beta_1, \beta_2, \beta_0$ kestirimlerinin etkinlik hesabı, 1000'er tekrarlı 9 veri seti için kullanılan çalışan korelasyon yapıları göz önüne alınarak yapılmıştır. Elde edilen hesaplamalara göre:

$\hat{\beta}_1$: Modelde yer alan "zaman" bağımsız değişkenine ilişkin katsayı kestirimi, $\hat{\beta}_2$: Modelde yer alan "grup" bağımsız değişkenine ilişkin katsayı kestirimi, $\hat{\beta}_0$: Modelde yer alan sabite ilişkin katsayı kestirimi olmak üzere;

- Genel olarak, QLS kestirimlerinin GEE kestirimlerine göre daha etkin olduğu görülmektedir; MSE_{GEE} / MSE_{QLS} > 1.
- Etkinlik değerleri birbirlerine yakın ve "1" civarındadır.
- **"Değiştirilebilir"** gerçek korelasyon yapısı ile üretilen veriler için;
  - Veri üretiminde kullanılan tüm korelasyon değerleri için (0,5, 0,7, 0,9), "AR-1" çalışan korelasyon yapısı kullanılarak elde edilen β_1 ve β_2 kestirimlerinin tamamında QLS kestirimleri GEE'ye göre üstündür,
  - Tüm korelasyon değerleri için "Değiştirilebilir" çalışan korelasyon yapısı için β_2 kestirimlerinin tamamında QLS kestirimleri GEE'ye göre daha etkin kestirimlerdir,
  - Tüm gerçek korelasyon değerlerinde, gerek GEE gerekse QLS kestirimlerine ait "hata kareler ortalaması" en küçük kestirimler, "Değiştirilebilir" çalışan korelasyon yapısına aittir.
- **"AR-1"** gerçek korelasyon yapısı ile üretilen veriler için;
  - Korelasyon değeri 0,9 için, β_1 kestirimlerinde GEE üstündür,
  - β_2 kestirimlerinde genel olarak QLS daha iyidir,
  - Tüm gerçek korelasyon değerlerinde, GEE ve QLS kestirimlerine ait "hata kareler ortalaması" en küçük kestirimler, genel olarak "AR-1" çalışan korelasyon yapısına aittir.
- "**Markov"** gerçek korelasyon yapısı ile üretilen veriler için;
  - Tüm gerçek korelasyon değerleri için β_1 kestirimlerinde "Değiştirilebilir" çalışan korelasyon yapısında GEE kestirimleri, AR-1 çalışan korelasyon yapısı için ise QLS kestirimleri daha iyi performanslı kestirimler vermiştir. β_2 kestirimlerinde ise "AR-1" gerçek korelasyon yapısında GEE kestirimleri QLS kestirimlerine göre daha üstündür, "Değiştirilebilir" çalışan korelasyon yapısında ise değişkenlik göstermektedir.
  - Tüm gerçek korelasyon değerlerinde, GEE ve QLS kestirimlerine ait "hata kareler ortalaması" en küçük kestirimler genel olarak "Markov " ve "AR-1" çalışan korelasyon yapılarına aittir.

5.TARTIŞMA

5.1. Gerçek veri çalışmasında;

- "cme" ve "cpogp" sonuç değişkenleri üzerinde, üç farklı eşit olmayan zaman aralıklarında yapılan ölçümleri içeren "zaman" değişkeninin önemli bir etkiye sahip olduğu görülmüştür. "İskeletsel sınıf III" malokluzyon nedeniyle tek çene cerrahisi ve çift çene cerrahisi yapılan hastaların sınıflandırıldığı "grup" değişkeninin ise bu sonuç değişkenleri üzerinde anlamlı bir etkisi yoktur.
- "sda" ve "sfda" yanıt değişkenlerinde, gerek "zaman" değişkeninin gerekse "grup" değişkeninin etkisi önemsizdir.
- Bütün sonuç değişkenleri için, **QLS'de** "Markov" çalışan korelasyon yapısı ile elde edilen korelasyon değeri, "Değiştirilebilir", "AR-1" ve "Tri-diagonal" çalışan korelasyon yapılarına göre elde edilen korelasyon değerlerinden daha yüksek çıkmıştır;

- "cme" değişkeni için "Markov", "Değiştirilebilir", "AR-1" ve "Tri-diagonal" değerleri sırasıyla şöyledir; 0,980, 0,923, 0,931, 0,660.
- "cpogp" için aynı sırayla; 0,973, 0,916, 0,915, 0,635.
- "sda" için yine aynı sırayla; 0,954, 0,883, 0,867, 0,628.
- "sfda" için yine aynı sırayla; 0,951, 0,856, 0,858, 0,603.

Alanyazında bu duruma benzer sonuçların alındığı çalışmalar aşağıda verilmiştir:

- 2006 yılında Schults ve arkadaşlarının çocuklarda böbrek nakli sonrasında görülen obezite ile ilgili yaptıkları araştırmada, "Markov", "Değiştirilebilir" ve "AR-1" çalışan korelasyon yapıları için elde edilen korelasyon değerleri sırasıyla; 0,918, 0,507, 0,699 olarak bulunmuştur (15).
- Xie ve Shults tarafından 2009 yılında yapılan, farelerin kan basıncının tekrarlı ölçümleri ile ilgili çalışmalarında da "Markov", "Değiştirilebilir", "AR-1" ve "Tri-diagonal" çalışan korelasyon yapıları için elde edilen korelasyon değerleri sırasıyla; 0,943, 0,705, 0,776, 0,522 olarak elde edilmiştir (8).
- Shults ve Hilbe tarafından yazılan, 2014 yılında yayınlanan "Quasi-Least Squares Regression" adlı kitapta yer alan obezite ile ilgili çalışmada, elde edilen korelasyon değerleri içinde en yüksek olan "Markov" korelasyon yapısı kullanılarak elde edilen değerdir. "Markov", "Değiştirilebilir", "AR-1" ve "Tri-diagonal" çalışan korelasyon yapıları için elde edilen korelasyon değerleri sırasıyla; 0,927, 0,714, 0,856, 0,562 olarak hesaplanmıştır (1).

- **GEE yönteminde** bütün sonuç değişenleri için "Tri-diagonal" çalışan korelasyon yapısı kullanıldığında, parametre kestirimleri yakınsama sağlanamaması nedeniyle elde edilememiştir. Bunun tersine, QLS yönteminde ise bütün yanıt değişkenleri için "Tri-diagonal" çalışan korelasyon yapısı için parametre kestirimleri elde edilmiştir. Elde edilen korelasyon parametreleri, bağımlı değişkenler olan "cme", "cpogp", "sda" ve "sfda" için sırasıyla; 0,660, 0,635, 0,628, 0,603'tür. Bu değerlerin tamamı "Tri-diagonal" yapının alabileceği değer aralığının (yaklaşık -1/2 ile 1/2 arasındadır) dışındadır.

  Alanyazında benzer sonuçlar alınan çalışmalar aşağıda sıralanmıştır:

  - Schults ve arkadaşları 2006 yılında yaptıkları araştırmada, GEE yöntemiyle "Tri-diagonal" çalışan korelasyon yapısı kullanılarak elde edilen korelasyon değerini 0,826, QLS ile elde edilen korelasyon değeri ise 0,518 olarak bulmuşlardır. GEE için elde edilen değer "Tri-diagonal" yapının alabileceği değer aralığının oldukça dışındayken, QLS ile elde edilen değer bu aralığın üst sınırı civarındadır (15).
  - Benzer bir durum da, Xie ve Shults'un 2009 yılında yaptıkları çalışmada ortaya çıkmıştır. GEE ile "Tri-diagonal" çalışan korelasyon yapısı kullanılarak elde edilen korelasyon değeri 0,742, QLS ile elde edilen değeri ise 0,522 olarak bulunmuştur (8).

5.2. Benzetim çalışmasında;

- Genel olarak, QLS ile yapılan regresyon parametre kestirimlerinin, GEE ile elde edilen kestirimlerden daha etkin kestirimler olduğu benzetim çalışması ile ortaya konulmuştur. Bununla birlikte etkinlik değerleri "1" değerine yakındır. Alanyazında, QLS ile elde edilen regresyon katsayılarının kestirimlerine ilişkin etkinlik çalışmalarına az rastlanılmıştır. Shults ve Hilbe'nin yazarı olduğu ve 2014 yılında

yayınlanan "Quasi-Least Squares Regression" adlı kitabında yer alan bir benzetim çalışmasında, düşük korelasyon değerlerine göre üretilen verilerde, "Yapılandırılmamış" çalışan korelasyon yapısı kullanılarak elde edilen regresyon katsayılarının etkinliği GEE ve QLS yöntemleri için eşit çıkmıştır, diğer bir deyişle etkinlik değerleri 1'dir. "Değiştirilebilir" korelasyon yapısı için ise etkinlik değerleri 1'e yakın olup, QLS'nin daha etkin olduğu sonucuna ulaşılmıştır. Yüksek korelasyon değerlerinde ise, gerek "Yapılandırılmamış" gerekse "Değiştirilebilir" korelasyon yapıları kullanılarak elde edilen regresyon parametrelerinin QLS kestirimlerinin GEE kestirimlerine göre daha etkin olduğu belirtilmektedir (1).

- Ayrıca, GEE'de "Markov" çalışan korelasyon yapısının uygulanamaması ve "Tri-diagonal" yapıda yakınsama sorununun yaşanmasından dolayı bütün benzetim durumlarında GEE ve QLS kestirimlerinin etkinlikleri karşılaştırılamamıştır.

5.3. Gerçek veri ve benzetim çalışmalarında ortak olarak;

- Gerçek veri ve benzetim ile elde edilen tüm veri setlerinde yapılan bütün parametre kestirimlerinde, GEE'de "Tri-diagonal" çalışan korelasyon yapısı için yakınsama sorunu ile karşılaşılmıştır. Oysaki, QLS'de tüm çalışan korelasyon yapılarında yakınsamada başarısızlık olmamıştır.
- GEE'de "Markov" çalışan korelasyon yapısı kullanılamazken, QLS'de tekrarlı ölçümlerin eşit zaman aralıklarında yapılmadığı ve eksik ölçümlerin olduğu verilerde "Markov "çalışan korelasyon yapısı kullanılabilmektedir. Bu durumun, bu tür verilerin analizinde QLS'ye GEE karşısında üstünlük sağladığı görülmüştür.

6.SONUÇ ve ÖNERİLER

QLS ve GEE'ye dayalı olarak yapılan uygulamalara ait sonuç ve öneriler aşağıda verilmiştir:

- GEE'de parametrelerin kestiriminde yakınsama sorunu "Tri-diagonal" çalışan korelasyon yapısı için gerek gerçek veride gerekse benzetim çalışmasında görülmüştür. QLS için ise, kestirimlerin yapılmasında kullanılan sürece bağlı olarak tüm durumlarda yakınsamada başarısızlık yaşanmamıştır.
- "Tri-diagonal" çalışan korelasyon yapısında GEE için yakınsama sorunu oluşması, aynı yapıdaki QLS sonuçları ile GEE sonuçlarının karşılaştırılmasının yapılamamasına neden olmuştur.
- "Markov" çalışan korelasyon yapısının QLS için uygulanabilmesi, GEE için ise uygulama olanağının olmaması, QLS ve GEE sonuçlarının bu yapıda karşılaştırılmasını engellemiştir.
- Gerçek veri ile yapılan uygulama kapsamında,
  - QLS'de kullanılan çalışan korelasyon yapıları olan, "Değiştirilebilir", "AR-1", "Tri-diagonal" ve "Markov" çalışan korelasyon yapıları içinde en yüksek korelasyon değeri "Markov" korelasyon yapısı ile elde edilmiştir.
  - Bütün sonuç değişkenleri üzerinde "grup" değişkeni olarak yer alan tek çene cerrahisi ve çift çene cerrahisinin bir önemi yoktur.
  - "cme" ve "cpogp" sonuç değişkenleri üzerinde ölçümlerin zamanlarını içeren "zaman" değişkeninin etkisi önemliyken, diğer yanıt değişkenleri olan "sda" ve "sfda" için önemsiz olduğu anlaşılmıştır.
- Benzetim çalışması ile kestirimlerin etkinliği bakımından QLS'nin GEE'ye göre üstünlük gösterdiği görülmüştür.
- GEE ve QLS sonuçlarının karşılaştırmalarının daha iyi yapılabilmesi için;

- "Markov" çalışan korelasyon yapısının, QLS'de olduğu gibi GEE'de de uygulanabilir olmalıdır,
- En uygun çalışan korelasyon yapısının belirlenmesinde kullanılan "Yarı En Küçük Kareler Bilgi Kriterinin (QIC)" GEE'de olduğu gibi QLS'de de uygulanabilir olması gereklidir.

- Bundan sonra yapılacak çalışmalar olarak aşağıdaki çalışma konuları önerilebilir;
  - GEE için "Markov" korelasyon yapısının kullanılabilmesi için yazılımların geliştirilerek araştırmacılara sunulması çok yararlı olacaktır,
  - Bu çalışmanın farklı sağlık alanlarında da uygulanması uygulamanın çeşitliliği açısından zenginlik katacaktır.

7.KAYNAKLAR

1. Shults J, Hilbe JM. Quasi-Least Squares Regression. New York: Taylor&Francis Group; 2014.
2. Hedeker D, Gibbons RD. Longitudinal Data Analysis. New Jersey: John Wiley & Sons, Inc.; 2006.
3. Wang M. Generalized Estimating Equations in Longitudinal Data Analysis: A Review and Recent Developments. Advances in Statistics. 2014; 303728: 1-11.
4. Hardin JW, Hilbe JM. Generalized Estimating Equations. Second Edition. New York: Taylor&Francis Group; 2013.
5. Agresti A. An Introduction to Categorical Data Analysis. Second Edition. New Jersey: John Wiley & Sons, Inc.; 2007.
6. Alpar R. Çok Değişkenli İstatistiksel Yöntemler. 5.Baskı. Ankara: Detay Yayıncılık; 2017.
7. Kim H, Shults J. %QLS SAS Macro: A SAS Macro for Analysis of Correlated Data Using Quasi-Least Squares. Journal of Statistical Software. 2010; 35(2): 1-22.
8. Xie J, Shults J. Implementation of quasi-least squares With the R package qlspack. UPenn Biostatistics Working Papers. 2009; 32: 1-19.
9. Smith T, Smith B. Proc Genmod with GEE to Analyze Correlated Outcomes Data Using SAS [Internet]. 2018 [Erişim Tarihi 29 Mart 2019]. Erişim adresi: https://www.lexjansen.com/wuss/2006/tutorials/TUT-Smith.pdf.
10. Aydın D. Uygulamalı Regresyon Analizi. Ankara: Nobel; 2014.
11. Chaganty NR. An alternative approach to the analysis of longitudinal data via generalized estimating equations. Journal of Statistical Planning and Inference. 1997; 63: 39-54.
12. Shults J, Chaganty NR. Analysis of Serially Correlated Data Using Quasi-Least Squares. Biometrics. 1998; 54(4): 1622-1630.
13. Chaganty NR, Shults J. On eliminating the asymptotic bias in the quasi-least squares estimate of the correlation parameter. Journal of Statistical Planning and Inference. 1999; 76: 145-161.
14. Shults J, Morrow AL. Use of Quasi-Least Squares to Adjust for Two Levels of Correlation. Biometrics. 2002; 58(3): 521-530.
15. Shults J, Ratcliffe SJ, Leonard M. Improved generalized estimating equation analysis via xtqls for implementation of quasi-least squares in Stata. Upenn Biostatistics Working Papers. 2006; 13: 1-21.
16. Shults J. Quasi-least squares fitting. WIREs Comput Stat. 2015; 7: 194-204.
17. Davis CS. Statistical Methods fort he Analysis of Repeated Measurements. New York: Springer-Verlag; 2002.
18. Ziegler A. Generalized Estimating Equations. New York: Springer; 2011.
19. Liu T, Bai Z, Zhang B. Weighted estimation equation: modified GEE in longitudinal data analysis. Front Math China. 2014; 9(2): 329-353.

20. Pan W, Connett JE. Selecting The Working Correlation Structure In Generalized Estimating Equations With Application To The Lung Health Study. Statistica Sinica. 2002; 12(2): 475-490.
21. Barnett AG, Koper N, Dobson AJ, Schmiegelow F, Manseau M. Using information criteria to select the correct variance-covariance structure for longitudinal data in ecology. Methods in Ecology and Evolution. 2010; 1: 15-24.
22. Pan W. Akaike's Information Criterion in Generalized Estimating Equations. Biometrics. 2001; 57: 120-125.
23. Hin LY, Wang YG. Working-correlation-structure identification in generalized estimating equations. Statist. Med. 2009; 28(4): 642–658.
24. Rotnitzky A, Jewell NP. Hypothesis testing of regression parameters in semiparametric generalized linear models for cluster correlated data. Biometrika. 1990;77(3): 485-497.
25. Alpar R. Uygulamalı İstatistik ve Geçerlik-Güvenirlik. 4.Baskı. Ankara: Detay Yayıncılık; 2016.
26. Shi G, Chaganty R. Application of Quasi-Least Squares to Analyse Replicated Autoregressive Time Series Regression Models. Journal of Applied Statistics. 2004; 31(10): 1147-1156.
27- StataCorp. 2015. Stata Statistical Software: Release 14.1. Texas: StataCorp LP.
28. R Core Team. The R (3.2.5) Project for Statistical Computing [Internet]. 2016 [Erişim Tarihi 10 Eylül 2018]. Erişim adresi: http://www.R-project.org/.
29. Muluk Z, Toktamış Ö, Karaağaoğlu E, Kurt S. Deney Düzenlemede İstatistiksel Yöntemler (Hicks CR.'ın yazarı olduğu kitabın 2.baskıdan çevirisi). Ankara: Gazi; 2009.
30. Erdoğan BD. Çoklu Atama Yöntemlerinin Rasch Modelleri İçin Performansının Benzetim Çalışması ile İncelenmesi [Doktora tezi]. Ankara: Ankara Üniversitesi; 2012.

T2

Risk Kestiriminde Kestirim Eğrisi ve Gizli Sınıf Kümeleme Analizi'nin Birlikte Kullanımı

Didem DERİCİ YILDIRIM[1], Merve TÜRKEGÜN[1], Bahar TAŞDELEN[1], Tuğba ARIKOĞLU[2], Semanur KUYUCU[2]

[1] Mersin Üniversitesi Tıp Fakültesi, Biyoistatistik Anabilim Dalı, MERSİN

[2] Mersin Üniversitesi Tıp Fakültesi, Çocuk Sağlığı ve Hastalıkları Anabilim Dalı, MERSİN

1. Giriş

Kişiselleştirilmiş tıp (KT, Personalized Medicine-PM), hastaya verilen tedaviyi yaş ve cinsiyet gibi demografik bilgilerin yanı sıra genetik ya da genetik olmayan biyomarkerlara göre şekillendirmeyi amaçlayan ve son yıllarda yaygınlaşan bir araştırma alanıdır. Doğru hastaya, doğru tedavinin, doğru zamanda uygulanabilmesini sağlayarak klinisyenlere yeni yöntem, bilgi ve tedavi seçeneklerini sunmak amacıyla Kestirim Eğrileri'nden (KE) sıklıkla yararlanılmaktadır. KE'nin temel amacı, hasta popülasyonunu, her bir hasta için uygun tedavileri içeren alt gruplara ayıracak bir biyomarker ya da biyomarker kombinasyonlarını belirlemektir. Biyomarkerlar hastalık için biyolojik bir durumun ya da sürecin göstergesidir. Bu nedenle,

- Hastalık riskini belirlemek,
- Hastalığın sonuç değişkeni ile tanı ve prognoz arasında bağlantı oluşturabilmek,
- Hastalığın altında yatan nedenlerin anlaşılmasına yardımcı olmak,
- Uygulanan tedaviye verilen hasta yanıtını tahmin etmek ve değerlendirmek,
- Klinik karar almak ve tedaviyi tabakalaştırmak amacıyla kişiselleştirilmiş tıp alanında kullanılmakta ve önemli bir yer tutmaktadır (1,2).

Özellikle hastalık riski hakkında bilgi veren biyomarkerların seçimi kronik hastalıkların tedavisinde önemli bir yere sahiptir. Tanısal bir marker yardımıyla hasta ve sağlıklı bireyleri sınıflandırabilen ROC eğrisi, Lojistik regresyon ve benzeri klasik yöntemler, bireylerin gerçek durumunu gösteren "Referans Test" olmadığı durumda tanı koymada ve hastalık riski belirlemede yetersizdir. Aynı zamanda klasik sınıflama ölçütleri kişinin hastalık durumunu tanımladıkları için hastalığa yakalanma riski hakkında bilgi vermemektedirler (3).

Bu nedenle, bu çalışmada literatürde birey odaklı sınıflama yaklaşımlarından biri olarak geçen Gizli Sınıf Kümeleme Analizi (GSKA) ile yine her bir birey için riskin hesaplanabildiği Kestirim Eğrisi (KE) yönteminin birlikte kullanımı araştırılacak olup, kronik bir hastalık olan astım tanısı koymada klinik öneme sahip parametrelerden yararlanılacaktır.

2. Gereç ve Yöntem

Çalışmaya Mersin Üniversitesi Sağlık Araştırma ve Uygulama Merkezi, Çocuk Allerji Kliniği'ne 2016-2017 yılları arasında başvuran 6-24 yaş arasında, astım tanısı almış 60 birey ve sağlıklı 51 birey dahil edilmiştir. Astım tanısı, astım için geliştirilmiş küresel kılavuzlara (Global Initiative for Asthma, GINA) göre tekrarlayan öksürük, hışıltı ve nefes darlığı öyküsü ile birlikte hava akımı kısıtlaması veya inhale kortikosteroidlere/bronkodilatörlere olumlu cevap veren bireylere konulmaktadır. Çalışmaya alınan hastalar bu kriterlere göre astım tanısı almış olup, bir yıldır takibi yapılan hastalardır. Çalışmamızda bu kriterlerin yanı sıra astımı predikte etmede önemli değişkenler olan Exhaled nitric oxide (FeNO), Total IgE ve Asymmetric Dimethylarginine (ADMA) parametrelerinden yararlanılacaktır.

Bu parametreler için öncelikle yaş ve cinsiyet düzeltmeli lojistik regresyon analizi ile kestirim eğrisi için düşük ve yüksek risk değerleri elde edilecektir. Sırasıyla bağımlı değişken olarak kesin tanı grubu ile Gizli Sınıf Kümeleme Analizi sonucunda elde edilen gruplar alınarak kestirim eğrileri çizilecektir. Univariate analizler için STATISTICA Version 13.3.1. paket programı kullanılmıştır. Kestirim eğrileri Stata/MP 11.0 programında "predcurve'' package kullanılarak çizilmiştir. Gizli Sınıf Kümeleme Analizi ise Latent Gold 5.1 paket programında gerçekleştirilmiştir (4).

2.1. Gizli Sınıf Kümeleme Analizi (GSKA)

Bir popülasyon içerisindeki homojen alt grupların bilinmediği durumda benzer yapıda bireyleri sınıflandırırken geleneksel yöntemlerden kümeleme analizi kullanılabilir. Ancak bu tekniğin uygulanabilmesi için bazı varsayımların sağlanması gerekmektedir. Bu varsayımlar sürekli değişkenler için çok değişkenli normallik, varyansların homojenliği ve doğrusallık olabilir. Varsayımların sağlanmadığı durumlarda kümeleme analizi yerine kullanılabilen Gizli Sınıf Kümeleme Analizi (GSKA) son yıllarda oldukça popüler hale gelmiştir. Kümeleme analizi ile GSKA arasındaki en önemli fark, varsayımlardan ziyade modele dayalı sınıflandırma yaklaşımına sahip olmasıdır. En basit GSKA modeli, y_i gözlenen değişken setlerine ait bireyin skoru, K gizli küme sayısı, π_k bir bireyin k. gizli kümeye ait olma olasılığı olmak üzere Eşitlik 1'de ki gibidir (5).

$$\boldsymbol{f}(\boldsymbol{y_i}|\boldsymbol{\theta}) = \sum_{\boldsymbol{k=1}}^{\boldsymbol{K}} \boldsymbol{\pi_k}\, \boldsymbol{f_k}\,(\boldsymbol{y_i}|\boldsymbol{\theta_k}) \qquad \text{Eşitlik 1}$$

Analiz sırasında, model parametreleri tahmin edildikten sonra, modele ait uyum iyiliği kontrolü yapılarak analize dahil edilen bireyler sınıflandırılır. Modelin parametreleri, her bir sınıfta yer alma olasılıklarıdır. Aynı zamanda gizli sınıf kümeleme analizi ile her bir birey için sınıf olasılıkları tahmin edilebilmektedir. Bu değerler sonsal (posterior) olasılıklar olarak adlandırılmaktadır (6).

2.2. Kestirim Eğrisi (KE)

Kestirim eğrisi; bir biyomarkerın risk tahmin kapasitesini modelleyerek, seçilen kohort için modelle ilişkili tahmini risk düzeylerinin dağılımını görsel olarak verir. Kohortun seçildiği popülasyonun ve ayrıca her bir bireyin hastalık riskini belirler. Bireylerin beklenen mutlak hastalık risk değerlerine karşın mutlak risk olasılıklarının dağılımını verir.

D hastalıkla ilgili binary yapıdaki sonuç değişkeni (D= 1 hastalık var, D=0 hastalık yok). Y ise sürekli yapıdaki biyomarker değeri olsun; Y biyomarkerı için risk Eşitlik 2 ile hesaplanır.

$$Risk(Y) = P(D = 1|Y = y) \qquad \text{Eşitlik 2}$$

Y biyomarkerı için kestirim eğrisi; x ekseninde risk yüzdeliğine (v) karşılık, y ekseninde hastalık risklerini gösteren eğridir. X ekseni biyomarkerın kümülatif dağılım fonksiyonu "$v = F(y)$"yi göstermek üzere; Kestirim Eğrisi v 'ye karşılık $R(v)$ 'nin eğrisidir ve Eşitlik 3 ile gösterilir (7).

$$R(v) = P\big(D = 1|Y = F^{-1}(v)\big) \qquad \text{Eşitlik 3}$$

Risk yüzdeliği $R(v)$'nin ters fonksiyonu $R^{-1}(p)$, $risk\,(Y)$ 'nin popülasyondaki kümülatif dağılımını ifade eder. Riski, p'ye eşit ya da küçük olan popülasyonun oranıdır. ($R^{-1}(p) = P[risk\,(Y) < p]$). Ters fonksiyonların elde edilebilmesi için öncelikle hastalığın düşük ve yüksek risk eşiklerinin belirlenmesi gerekir. Bu eşikler bireylerin hastalık seyrine, uygulanacak tedavi şekline ve genel prevalansa göre belirlenmektedir. p_L düşük risk, p_H ise yüksek risk eşiğini göstermek üzere; düşük riskli popülasyonun oranı $R^{-1}(p_L)$ ve yüksek riskli popülasyonun oranı $1 - R^{-1}(p_H)$'dir. Böylece hasta popülasyonunun ilgili biyomarkera göre yüzde ne kadarının düşük riskin altında ya da yüksek riskin üstünde olduğu belirlenir. Düşük ve yüksek risk eşikleri

arasında kalan riskler belirsiz olup; $R^{-1}(p_H) - R^{-1}(p_L)$ ile hesaplanır (7,8). Bu grubun risk sınıfı belli olmadığı için bu risk sınıfı "*Gri Bölge*" olarak tanımlanabilir.

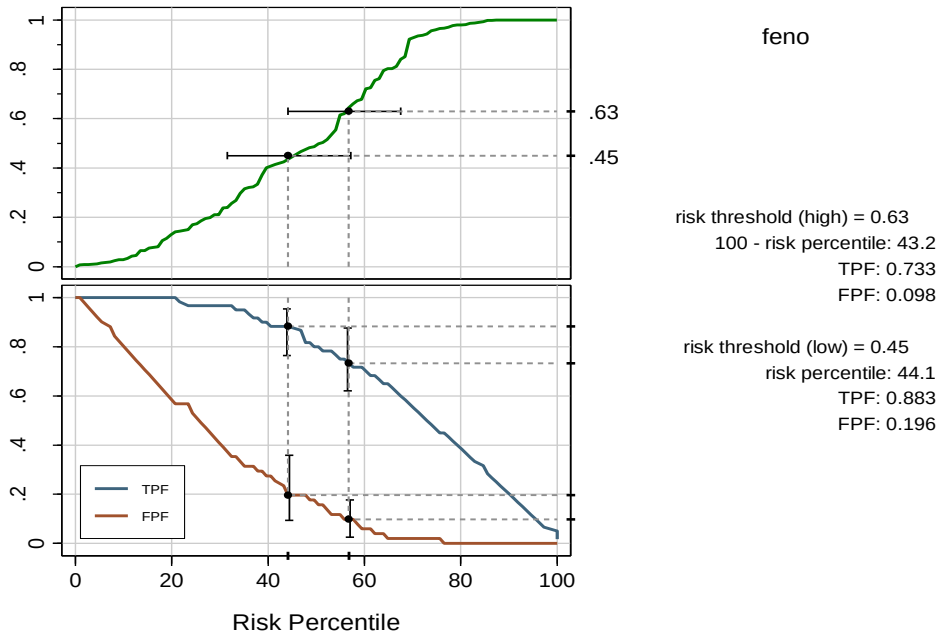
3. Bulgular

Çalışmaya katılanların yaş ortalaması 13.00±4.53 olup, %54.1'i erkeklerden %45.9'u kızlardan oluşmaktadır. Astım bireyler ile sağlıklı bireylerin cinsiyet dağılımı istatistiksel olarak farklıdır (p=0.004). Astım grubunun yaş ortalaması 11.50±3.19 sağlıklı grubun ise 14.76±5.22'dir. İki grup arasında yaş bakımından istatistiksel olarak anlamlı fark vardır (p<0.001). Yaş ve cinsiyet bakımından benzer gruplar oluşturmak örnek genişliğini azaltacağından, yaş ve cinsiyet düzeltmeli modelleme (age and sex-adjusted modelling) yapılacaktır.

Kestirim eğrisi modellemesi için, tanı grubu bağımlı değişken, biyomarkerlar bağımsız değişkenler ve yaş ve cinsiyet kovaryat değişkenler olmak üzere, multiple lojistik regreyon modeli kullanılarak her bir hasta için risk kestirimleri (PredRisk) elde edilmiş ve bu değerlerin ortalaması ve %95 güven aralığı hesaplanmıştır. Güven aralığının alt sınırı düşük risk eşiği ve üst sınırı ise yüksek risk eşiği olarak belirlenmiştir. Böylelikle "Risk Grup"ları oluşturulmuştur. Çalışmada incelenen biyomarkerlar bakımından 6-24 yaş aralığında bir bireye astım tanısı koymak için düşük risk eşiği 0.45, yüksek risk eşiği ise 0.63 olarak hesaplanmıştır. Çalışmaya katılanların 0.18'nin ise riski ne yüksek risk eşiğinin üstünde ne de düşük risk eşiğinin altındadır.

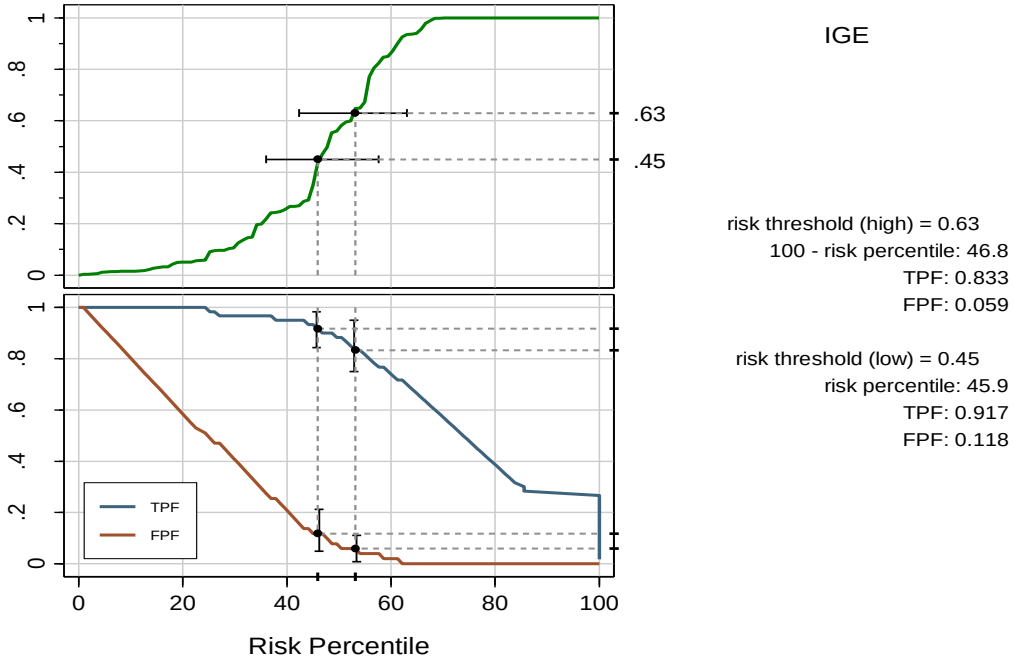
Marker seçimi yaparken, sınıflama performans istatistikleri True Positive Fraction (TPF) ve False Positive Fraction (FPF) ile değerlendirilebilir. Risk eşik değerleri için hesaplanan TPF; ilgili biyomarkera göre Kestirim eğrisinin hastaları tespit etme oranı iken FPF ise Kestirim eğrisinin sağlıklı bir bireyi yanlışlıkla hasta olarak tespit etme oranı olarak tanımlanabilir (9,10). Yaş ve cinsiyete göre yapılan düzeltmelerden sonra elde edilen ≥0.60 TPF ve ≤0.30 FPF oranları ve bunların %95 güven aralıkları dikkate alınmıştır. "Sıfır" değerini içermeyen güven aralığına sahip biyomarker(lar) istatistiksel olarak anlamlı kabul edilmiştir.

TPF ve FPF sonuçlarına göre düşük riskli hastaları sınıflandırmada kullanılacak en iyi biyomarker Feno'dur. TPF (0.88.[95%CI=0.765-0.954]) ve FPF (0.19.[95%CI=0.094-0.359])'ye sahip olan Feno için düşük riskli olarak tahmin edilen hastalar kohortun %44.1'dir. Buna göre kohortun %44.1'nin astım olma riski düşük risk eşiğinde ya da bu risk eşiğinin altındadır. Feno'ya ait kestirim eğrisi Şekil 1'de verilmiştir. Aynı zamanda Feno'ya göre kohortun %43.2'sinin astım olma riski yüksek risk eşiğinde ya da üstündedir. Feno astım olma riski yüksek olan hastaların %73'ünü doğru tespit etmiştir. Hastaların %12.7'si ise gri bölgede yer aldığı düşük veya yüksek riskli gruba sınıflandırılamadığı söylenebilir.



Şekil 1. Feno için Kestirim Eğrisi

TPF ve FPF sonuçlarına göre yüksek riskli hastaları sınıflandırmada kullanılacak en iyi marker Total IgE'dir. Yüksek risk için TPF=0.917 [0.741-0.950] ve FPF=0.118[0.010-0.110] olarak hesaplanmıştır. Kestirim Eğrisi Şekil 2'de verilmiştir. Total IgE'ye göre hastaların %45.9'unun astım olma riski yüksek risk eşiğinde ya da üstündedir. Ayrıca astım olma riski düşük olan hastaları da tahmin etmede de oldukça başarılı olan Total IgE için hesaplanan TPF 0.830. FPF ise 0.059'dur. Total IgE kohortun %46.8'ni astım olma riskini düşük risk eşiğinde ya da altında tahmin etmiştir. Total IgE için riski gri bölgeye düşenler kohortun %6.4'dür.



Şekil 2. Total IgE için Kestirim Eğrisi

ADMA biyomarkerının yüksek TPF oranlarına rağmen 0.30'dan büyük elde edilen FPF değerleri yüzünden düşük ve yüksek riskli hastaları sınıflandırmada kullanılamayacağı tespit

edilmiştir. Gri bölgeye düşen hastalar hem yüksek riskli hem de düşük riskli olarak değerlendirmeye alındığında oluşturulan yeni gruplarda ilgili marker ortalamaları gruplar arasında istatistiksel olarak anlamlı farklılık göstermiştir.

Referans test ile kesin tanı koymanın zor olduğu durumlarda, elde edilen ve hastalık üzerinde etkisi olduğu düşünülen parametreler ile gizli sınıf kümeleme analizi yapılmış olup, elde edilen iki sınıf referans test sonucu olarak alınarak, KE için TPF ve FPF sonuçları elde edilmiştir. GSKA sonucunda bireylerin %40.5'i birinci örtük sınıfa, 59.5'i ise ikinci örtük sınıfa atanmıştır. Parametrelerin kesin tanı karşısında risk gruplarını ayırmadaki başarısı Tablo 1'de, GSKA sonucu elde edilen gruplar karşısında risk gruplarını ayırmadaki başarısı ise Tablo 2'de verilmiştir.

Tablo 1. Kesin tanı gruplarına göre kestirim eğrisi sonuçları

		Düşük Risk Eşik Değeri = 0.45				**Yüksek Risk Eşik Değeri =0.63**				
	TPF	**% 95 GA**	**FPF**	**Risk %**	**%95 GA**	**TPF**	**%95 GA**	**FPF**	**%95 GA**	**Risk %**
Feno	0.883	0.765-0.954	0.19	44.1	0.094-0.359	0.730	0.620-0.870	0.090	0.025-0.177	43.2
IgE	0.917	0.840-0.980	0.118	45.9	0.04-0.204	0.833	0.741-0.950	0.059	0.00-0.110	46.8
ADMA	0.867	0.697-0.954	0.510	30.6	0.264-0.705	0.583	0.382-0.783	0.235	0.097-0.413	41.4

IgE astımın tanısında kullanılan referans testlerle birlikte kullanıldığında yüksek riskli hastaları ayırt etmede başarıyı artırmaktadır.

Tablo 2. GSKA ile elde edilen gruplara göre kestirim eğrisi sonuçları

	Düşük Risk Eşik Değeri = 0.45					**Yüksek Risk Eşik Değeri =0.63**				
	TPF	**% 95 GA**	**FPF**	**%95 GA**	**Risk %**	**TPF**	**%95 GA**	**FPF**	**%95 GA**	**Risk %**
Feno	0.822	0.727-0.951	0.106	0.031-0.185	61.3	0.778	0.667-0.906	0.030	0.000-0.078	32.4
IgE	0.689	0.551-0.870	0.121	0.058-0.262	65.5	0.578	0.05-0.734	0.091	0.006-0.123	27.9
ADMA	0.733	0.530-0.910	0.333	0.193-0.490	51.4	0.311	0.098-0.652	0.121	0.000-0.220	18.9

Kesin tanı olmadan ADMA astım riski düşük olanları kestirmede anlamlı bir parametre değildir.

4. Tartışma ve Sonuç

Biyomarkerlardan yola çıkarak bir hastalık için riski hesaplamamızı sağlayan kestirim eğrisi yöntemi, klinik alanda tanı koymakta oldukça güçlük çekilen, birçok hastalık ile benzer semptomlar gösteren ve karıştırıcı faktörlerin fazla olması gibi durumlar için oldukça kullanışlı bir yöntemdir. Aynı zamanda kesin tanının olmadığı bir durumda, literatürde sıklıkla tanı gruplarını belirlemede kullanılan bir yöntem olan gizli sınıf kümeleme analizi ile de kestirim eğrisi elde edilebilir. Yüksek riskli hastaları tespit etmede, tanı testlerine yardımcı biyomarker olarak IgE nin kullanılması önerilebilir. Ancak IgE'nin astım tanısı koyduran testler uygulanmadan kullanıldığında, kestirim gücünün düşük olduğuna dikkat çekmek gerekir. FeNO kesin tanı konmadan bile hem astım tanısı alma riski düşük bireyleri hem de astım tanısı alma riski yüksek bireyleri kestirimde başarılı bir parametredir.

Kaynaklar

1. Li W., Chen C., Li X,. Beckman R.A. Estimation of Treatment Effect in Two Stage Confirmatory Oncology Trials of Personalized Medicines. Statist. Med. 36:1843–1861, 2017.
2. ZhaoY., Zeng D. Recent Development on Statistical Methods for Personalized Medicine Discovery. Front. Med. 7(1): 102–110, 2013.
3. Smeden M., Naaktgeboren C.A., Reitsma J.B., Moons K., Groot J. "Latent Class Models in Diagnostic Studies When There is No Reference Standard—A Systematic Review" Am J Epidemiol, 179(4):423-431, 2014.
4. Vermunt, J.K., Madigson, J. Latent Gold 5.1 Belmont, Massachusetts, USA: Statistical Innovations Inc**.** (2013).
5. Vermunt, J.K., Magidson, J. Latent class cluster analysis. In J.A. Hagenaars & A.L. McCutcheon (eds.), Advances in Latent Class Analysis. Cambridge University Press. (2002).
6. Goddman L.A. Exploratory Latent Structure Analysis Using Both Identifiable and Unidentifiable Models. Biometrika, 61: 215-231, (1974).
7. Huang Y., Pepe M.S., Feng Z. Evaluating the Predictiveness of a Continuous Marker. Biometrics 63(4): 1181–1188, (2007).
8. Huang Y., Pepe M.S. Assessing risk prediction models in case–control studies using semiparametric and nonparametric methods. Statistics in Medicine 29:1391-1410,(2010).
9. Pepe M.S., Feng Z., Huang Y., Longton G., Prentice R., Thompson I.M., Zheng Y. Integrating the Predictiveness of a Marker with Its Performance as a Classifier. American Journal of Epidemiology 167:362–368, (2008).
10. Pepe M.S., Janes H., Longton G.,Leisenring W., Newcomb P. Limitations of the Odds Ratio in Gauging the Performance of a Diagnostic, Prognostic, or Screening Marker. American Journal of Epidemiology 159(9):882-890, (2004).

T2

Usage of Predictiveness Curve and Latent Cluster Analysis with Together for Risk Estimation

Didem DERICI YILDIRIM[1], Merve TURKEGUN[1], Bahar TASDELEN[1], Tugba ARIKOGLU[2], Semanur KUYUCU[2]

[1] Mersin University Faculty of Medicine, Biostatistics and Medical Informatics, MERSIN

[2] Mersin University Faculty of Medicine, Pediatrics, MERSIN

1. Introduction

Personalized Medicine (PM) has became popular in recent years. It is aimed to plan the individual treatment according to genetic or other biomarkers as well as the demographic factors such as age and gender. Predictiveness Curve (PC) is frequently used to provide information to clinicians about new treatment methods by receiving appropriate treatment for each patient in the personalized medicine field. The main purpose of PC is to identify a biomarker or biomarker combination that distinguish population into sub-groups in terms of suitable treatment for each patient. Biomarkers are indicators of a biological condition or process for the disease. And so, this method is used in personalized medicine because of the listed reasons below.

- Determining the risk of disease
- Establishing a relationship between outcome and diagnosis/prognosis
- Understanding reasons of disease
- Estimating and evaluating patient response to treatment
- Making clinical decision
- Stratification of treatment (1,2)

Especially the selection of biomarkers which provide information about disease risk has an important role in the treatment of chronic diseases. Classical methods such as ROC curve, logistic regression and similar methods that can classify people as patients and healthy using a diagnostic marker are inadequate in terms of diagnosing and determining the risk of disease if there is no gold standard test. Besides, the criteria of classical methods do not provide information about the risk of disease occurrence because they just define the individual's disease status in the moment (3).

For this reason, this study is aimed to use the Latent Class Cluster Analysis (LC Cluster) which is a person-centered method and PC which is calculated the risk for each individual with together for diagnosis of asthma utilizing the clinical parameters for this disease.

2. Material & Methods

The data was collected from Mersin University Faculty of Medicine Research and Training Hospital, Department of Pediatric Allergy and Immunology. The children and young adults attending to the department with asthma and aged 6–24 years were enrolled. Sixty patients with asthma and fifty one healthy people were included to the study. A definitive diagnosis of asthma was made according to Global Initiative for Asthma (GINA) guidelines,that is a history of recurrent cough, wheezing, shortness of breath and chest tightness together with reversibility of airflow limitation or favorable response to inhaled corticosteroids and/or bronchodilators. They had been followed up in our clinic for at least one year. In our study, significant parameters of asthma

prediction such as Exhaled nitric oxide (FeNO), Total IgE and Asymmetric Dimethylarginine (ADMA) were evaluated.

Initially, the low and high thresholds were obtained from age and gender adjusted logistic regression. Then, the outcome variable was selected as diagnosis group and latent groups obtained by LC Cluster, respectively. We performed univariate analysis using STATISTICA 13.3.1 (TIBCO, CA, USA). The "predcurve'' package of Stata/MP 11.0 and Latent Gold 5.1 were used for PC and LC Cluster, respectively (4).

2.1. Latent Class Cluster Analysis (LC Cluster)

To identify the unknown homogeneous subgroups in a population, traditional methods such as clustering analysis from can be used to classify individuals of similar structure. However, there are some assumptions to use this method. These assumptions are multivariate normality, variance homogeneity and linearity. If assumptions are not provided, there is an alternative and popular method called Latent Class Cluster Analysis is used. The main difference of two methods is that LC Cluster has a model based clustering approach. The main LC Cluster model is given in Equation 1. In the model, y_i is the score of observed variables set, the latent number of cluster is K, π_k is the probability of being cluster k (5).

$$\boldsymbol{f(y_i|\theta) = \sum_{k=1}^{K} \pi_k f_k (y_i|\theta_k)} \quad \text{Equation 1}$$

In analysis, individuals are classified utilizing the best model according to goodness of fit statistics after parameter estimation. The parameters of model are the class membership probabilities. Besides, Class membership for each individual can be estimated with his method. These values are called as posterior probabilities (6).

2.2. Predictiveness Curve (PC)

Predictiveness Curve provides a visual conception of the estimated risk levels associated with the model obtained from selected cohort. It determines both risk of selected cohort and disease risk for each individual. A graph that shows the distribution of absolute risk probabilities against expected disease risk values is obtained.

D is a binary outcome related with disease (D=1, disease/crisis is present, D=0, disease/crisis is not present) and Y is a continuous variable; risk is calculated by Equation 2 for the Y biomarker;

$$(Y)=(D=1|Y=y) \quad \text{Equation 2}$$

Predictiveness Curve for Y biomarker is the curve of risk percentage (v) in the x axis, versus risk of disease in the y axis. The x axis is the biomarker's cumulative distribution function ''v=F(y)'' is PC is curve of (v) versus to v ', and is indicated by Equation 3 (7).

$$(v)=(D=1|Y=F-1(v)) \quad \text{Equation 3}$$

The Predictiveness Curve can be interpreted with risk percentages (v) or inverse functions. The biomarker, which has higher variability for (v), predicts the risk more successfully compared to others. Inverse functions are interpreted easier than the risk percentages. It enables the comparison of two biomarkers for disease risk as low or high risk. Because PC, by using the inverse functions, creates a common scale independent from their scale types for the markers. The inverse function of $R(v)$ is $R-1(p)$ and explains the cumulative distribution of Y. Risk is calculated by the ratio of the population equal or smaller than p. $(R-1(p)=P[risk\ (Y)<p])$

Low and high-risk thresholds of the disease risk should be determined in order to obtain inverse functions. These thresholds are detected by the prognose of disease, treatment and prevalence. pL displays "Low Risk" and pH displays "High Risk" thresholds.$1-R-1(pH)$ is the ratio of with high-risk population and $R-1(pL)$ is with low-risk population. Thus, the risk ratio for disease/crisis can be calculated for selected cohort. The risk values between the low and high-risk

thresholds are unclear for classification ($R-1(pH)-R-1(pL)$) (7,8). These patients can be classified as "Grey Area'.

3. Results

The mean age of participants was 13.00±4.53 and the majority of sample was boys with the percentage of 54.1. There was a significant relationship between gender and group (p=0.004). Besides, age was statistically different in asthma group (11.50±3.19) in comparison with healthy (14.76±5.22) group (p<0.001). Therefore, age and gender are considered as covariates of models.

In this study, multiple logistic regression model including diagnosis group as dependent variable and biomarkers as independent variables was used. This model was adjusted by age and gender. The mean and 95% CI of risk was calculated for each patient. The lower limit of confidence interval was defined as low risk threshold and the upper limit as high risk threshold. Risk groups were obtained. Low risk threshold for asthma risk was calculated as 0.45 and high risk threshold was 0.63. The Grey area threshold was 0.18.

For selection of biomarkers, True Positive Fraction (TPF) and False Positive Fraction (FPF) were used. TPF was the ratio of detection patients accurately, FRP was the ratio of detection patients false (9,10). The thresholds had been ≥0.60 for TPF and ≤0.30 for FPF after the adjustment of age and gender. Confidence intervals which zero was not included were accepted as statistical significant biomarkers.

According to TPF and FPF, Feno is the best biomarker to classify low-risk patients.TPF (0.88.[95%CI=0.765-0.954]) and FPF (0.19.[95%CI=0.094-0.359]) were calculated for FeNO. The percentage of low risk patients in cohort was 44.1, and high risk percentage was 43.2. The PC of Feno was given in Figure 1. Feno was predicted 73% of individuals with high risk truly. The grey area percentage was 12.7 and these individuals could not be classified.

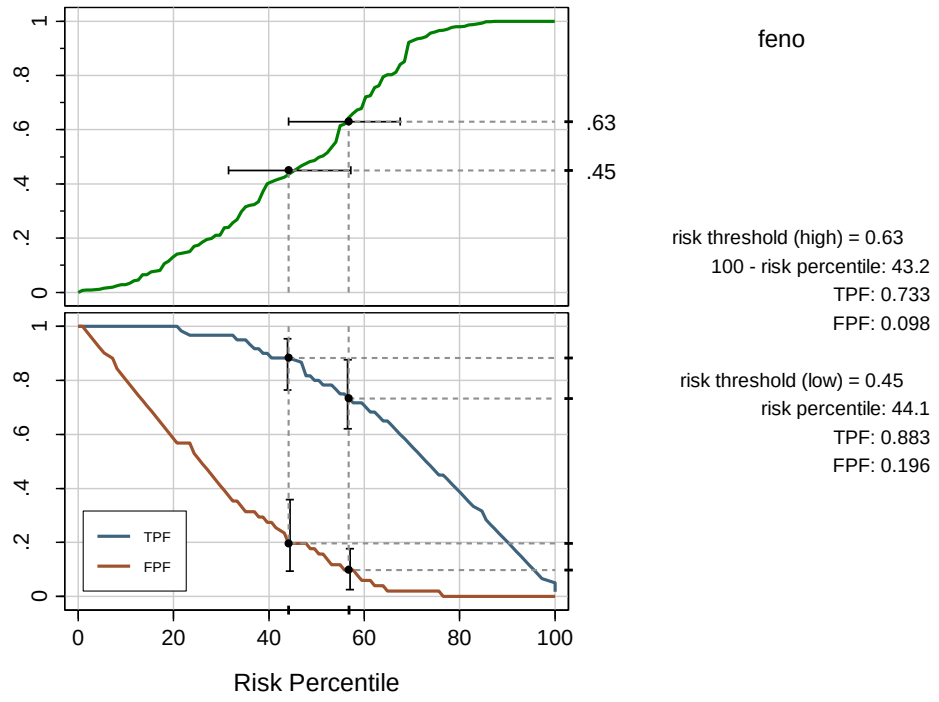


Figure 1. The PC for Feno

Total IgE is the best biomarker to classify high-risk patients. TPF=0.917 [0.741-0.950] ve FPF=0.118[0.010-0.110] were calculated for Total IgE. The PC of IgE was given in Figure 2. The percentage of high risk patients in cohort was 45.9. Feno was predicted 46.8% of individuals with low risk truly. The grey area percentage was 6.4 and these individuals could not be classified.

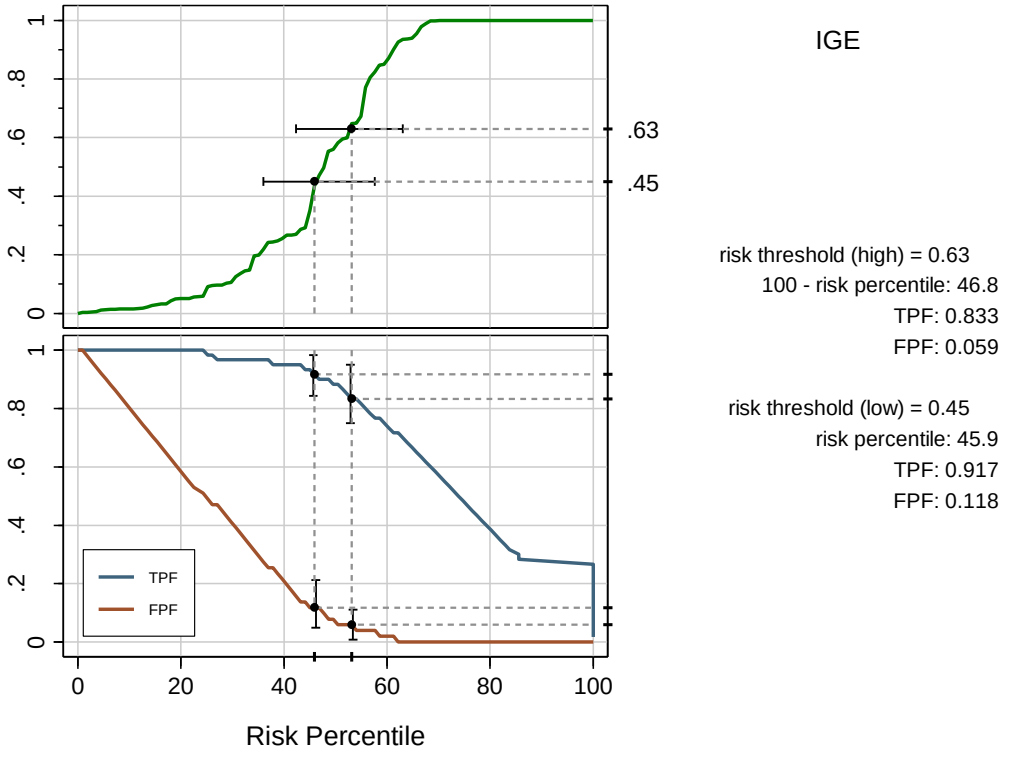


Figure 2. The PC for Total IgE

According to PC, ADMA was not as significant as FeNO and Total IgE for distinguishing low and high risk. The patients in gray area were evaluated as both high-risk and low-risk, the mean of biomarkers were significantly different for both situation. In cases where it was difficult to make a definitive diagnosis with a reference test, LC Cluster analysis was done with significant clinical parameters, two classes were taken as a reference test, TPF and FPF results were obtained for PC. In LC Cluster, 40.5 percentage of cohort was in the first latent class and 59.5 percentage of cohort was in the second latent class. The performance of parameters against to diagnosis group were given in Table 1 and against to LC Clusters were given in Table 2.

Table 1. PC results for diagnosis groups

		Low Risk threshold = 0.45				**High Risk threshold =0.63**				
	TPF	**% 95 GA**	**FPF**	**Risk %**	**%95 GA**	**TPF**	**%95 GA**	**FPF**	**%95 GA**	**Risk %**
Feno	0.883	0.765-0.954	0.19	44.1	0.094-0.359	0.730	0.620-0.870	0.090	0.025-0.177	43.2
IgE	0.917	0.840-0.980	0.118	45.9	0.04-0.204	0.833	0.741-0.950	0.059	0.00-0.110	46.8

ADMA	0.867	0.697-0.954	0.510	30.6	0.264-0.705	0.583	0.382-0.783	0.235	0.097-0.413	41.4

IgE increased the success in differentiating high risk patients in conjunction with reference test for the diagnosis of asthma. Without definitive diagnosis, ADMA was not a significant parameter to predict asthma risk.

Table 2. PC results for LC Clusters

	Low Risk threshold = 0.45					**High Risk threshold =0.63**				
	TPF	**% 95 GA**	**FPF**	**%95 GA**	**Risk %**	**TPF**	**%95 GA**	**FPF**	**%95 GA**	**Risk %**
Feno	0.822	0.727-0.951	0.106	0.031-0.185	61.3	0.778	0.667-0.906	0.030	0.000-0.078	32.4
IgE	0.689	0.551-0.870	0.121	0.058-0.262	65.5	0.578	0.05-0.734	0.091	0.006-0.123	27.9
ADMA	0.733	0.530-0.910	0.333	0.193-0.490	51.4	0.311	0.098-0.652	0.121	0.000-0.220	18.9

4. Discussion

The Predictiveness Curve, which enables us to calculate the risk for a disease based on biomarkers, was a very useful method because of the difficulty in diagnose, determining complex symptoms and confounding factors. Also, in the absence of definite diagnosis, the predictiveness curve could be drawn with LC clusters to determine risk groups. IgE level was predictive in asthma, and it might be used to especially differentiate high risk patients. However, the predictive power of IgE was low when there was no definitive diagnosis of asthma. FeNO was a successful parameter in estimating both individuals with low risk of asthma and individuals with high risk of asthma, even without definitive diagnosis.

References

1. Li W., Chen C., Li X,. Beckman R.A. Estimation of Treatment Effect in Two Stage Confirmatory Oncology Trials of Personalized Medicines. Statist. Med.2017; 36:1843–1861. doi: 10.1002/sim.7272
2. ZhaoY., Zeng D. Recent Development on Statistical Methods for Personalized Medicine Discovery. *Front. Med*. 2013; 7[1]: 102–110. Doi:10.1007/s11684-013-0245-7
3. Smeden M., Naaktgeboren C.A., Reitsma J.B., Moons K., Groot J. "Latent Class Models in Diagnostic Studies When There is No Reference Standard—A Systematic Review" Am J Epidemiol, 179(4):423-431, 2014.
4. Vermunt, J.K., Madigson, J. Latent Gold 5.1 Belmont, Massachusetts, USA: Statistical Innovations Inc**.** (2013).

5. Vermunt, J.K., Magidson, J. Latent class cluster analysis. In J.A. Hagenaars & A.L. McCutcheon (eds.), Advances in Latent Class Analysis. Cambridge University Press. (2002).
6. Goddman L.A. Exploratory Latent Structure Analysis Using Both Identifiable and Unidentifiable Models. Biometrika, 61: 215-231, (1974).
7. Huang Y., Pepe M.S., Feng Z. Evaluating the Predictiveness of a Continuous Marker. Biometrics 63(4): 1181–1188, (2007).
8. Huang Y., Pepe M.S. Assessing risk prediction models in case–control studies using semiparametric and nonparametric methods. Statistics in Medicine 29:1391-1410,(2010).
9. Pepe M.S., Feng Z., Huang Y., Longton G., Prentice R., Thompson I.M., Zheng Y. Integrating the Predictiveness of a Marker with Its Performance as a Classifier. American Journal of Epidemiology 167:362–368, (2008).
10. Pepe M.S., Janes H., Longton G.,Leisenring W., Newcomb P. Limitations of the Odds Ratio in Gauging the Performance of a Diagnostic, Prognostic, or Screening Marker. American Journal of Epidemiology 159(9):882-890, (2004).

T3

Aynı Veri Setleri Üzerinde Bazı Etki Büyüklüğü Hesaplama Yöntemlerinin Tahmin Başarılarının Karşılaştırılması

Duygu KORKMAZ[1,2], İlker ÜNAL[1]

[1] Çukurova Üniversitesi Tıp Fakültesi Biyoistatistik Anabilim Dalı, Adana

[2] Yüzüncü Yıl Üniversitesi Tıp Fakültesi Tıp Eğitimi Anabilim Dalı, Van

e-mail: duygukorkmaz@yyu.edu.tr

Giriş: İstatistiksel analiz sonuçlarını rapor ederken, gruplar arasında gözlenen farklılıkların istatistiksel olarak anlamlı olmasının yanı sıra, bu gözlenen farklılıkların pratik öneme sahip olduğunu gösterecek bazı etki büyüklüklerini sunmak da çok önemlidir. Etki büyüklüğü hesaplama yöntemleri arasında en sık kullanılan ve en iyi bilinenleri Eta kare ve Cohen'dir. Kullanılan diğer bazı yöntemler ise Epsilon kare ve Omega karedir.
Amaç: Bu çalışmada amaç, sık kullanılan dört etki büyüklüğü hesaplama yöntemlerinin popülasyon değerini tahmin etme başarılarının farklı senaryolarda üretilen verilerle karşılaştırılmasıdır.
Yöntem: Farklı grup sayısı (3, 4, 5, 10) ve farklı etki büyüklüğü derecesi (küçük, orta, büyük) senaryolarında 1.000.000'luk veri setleri, popülasyon veri setleri olarak üretilmiştir. Bu veri setlerinden istenilen örnek büyüklüğünde (5, 10, 20, 30, 50, 100) örneklemler çekilmiştir. Çekilen bu örneklemlerden Eta kare, Cohen, Epsilon kare ve Omega kare yöntemleri ile hesaplanan etki büyüklüklerinin popülasyon etki büyüklüklerini tahmin etme başarısı karşılaştırılmıştır.
Bulgular: Genel olarak Omega kare ve Epsilon kare hesaplama yöntemleri popülasyon parametre tahmininde Eta kare ve Cohen hesaplama yöntemlerinden daha başarılı sonuçlar vermiştir. Küçük orta ve büyük etki büyüklüklerini tahmin etme başarısı en iyi olan yöntem Omega kare olarak belirlenmiştir. Grup sayısı arttıkça (5,10) Epsilon kare tahminleri de Omega kareye benzer şekilde tüm örneklerde (5,10,20,30,50,100) en iyi tahminlere ulaşmıştır. Eta kare ve Cohen f tahminleri ise gözlem sayıları küçük olduğunda başarısız sonuçlar vermiştir.
Sonuç: Sık kullanılmasına rağmen Eta kare ve Cohen f'in, popülasyona ait etki büyüklüğünün gerçek değerini tahmin etme başarısının Omega kare ve Epsilon kare kadar iyi olmadığı görülmüştür. Grup sayısı 10 olduğunda ise çalışmaya alınan etki büyüklüğü hesaplama yöntemlerinin hepsinde tahmin başarısı oldukça düşmektedir (gerçek değerden uzaklaşmaktadırlar).Etki büyüklüğünün derecesi tahminlerdeki başarılar üzerinde etkili değildir. Tüm senaryolar dikkate alındığında Omega ve Epsilon karenin diğer etki büyüklüğü hesap yöntemlerinden daha yansız ve birbirlerine yakın sonuçlar verdiği görülmüştür.

Anahtar Kelimeler: Etki büyüklüğü, Eta kare, Cohen, Epsilon kare, Omega kare

1. Giriş

Çalışmalarda genellikle araştırmacılar gruplar arası farkın istatistiksel olarak önemli olup olmadığı ile ilgilenirken, klinik olarak önemliliği göz ardı etmektedir. Pratikte, istatistiksel anlamlılık ölçütü olarak kullanılan " P" değeri ne kadar küçükse, araştırılan özelliğin o kadar etkili ya da daha güçlü olduğuna inanılır. Oysa "P" değeri okuyucuya bir etkinin olup olmadığını bildirebilirken, etkinin büyüklüğü hakkında bir fikir vermez. Bununla birlikte, istatistiksel

anlamlılık, çalışılan örneklemin büyüklüğünden etkilenir. Küçük farklılıklar büyük örneklerde anlamlı olarak bulunabileceği gibi klinik olarak bir öneme sahip olmayabilir. Etki büyüklüğü, klinik olarak anlamlı farklılığın ortaya konması için ilgilenilen sonuç değişkenine göre ortalama ya da oranlar arasındaki beklenen farklılık olarak ifade edilebilir. Çalışmaları raporlama ve yorumlamada istatistiksel önemliliğin ("P" değeri) yanı sıra, klinik önemliliği (etki büyüklüğü) belirtmek son derece önemlidir (Sulliyan, Fein, 2012). Buna bağlı olarak son zamanlarda, bilimsel dergilerin büyük bir çoğunluğu, istatistiksel analiz sonuçlarını belirtirken bazı etki büyüklüğü ölçütlerini rapor etmeyi talep etmektedir (Skidmore ve Thompson, 2013; Okada, 2013). Bu çalışmada etki büyüklüğü değerini doğru bir şekilde sunabilmek için grup karşılaştırmalarında kullanılan bazı etki büyüklüğü yöntemlerinin (Eta kare, Omega kare, Epsilon kare, Cohen f) farklı senaryolar altında (örneklem sayısı ve grup sayısını değiştirerek) popülasyona ilişkin küçük, orta ve büyük etki büyüklüklerinin tahmin başarılarının karşılaştırılması amaçlanmıştır.

2. Gereç ve Yöntem

2.1 Veri üretme ve analiz

Popülasyondan elde edilen etki büyüklüğü değerlerini örnekten hesaplanan etki büyüklüğü değerlerinin tahmin başarılarını incelemek amacı ile 1000000 gözlemden oluşan farklı grup sayılarında (3,4,5,10) ve farklı etki büyüklüklerinde (küçük, orta, büyük) veri üretilmiştir. Cohen'e (1998) göre etki büyüklüğü Eta kare, Omega kare ve Epsilon kare için ≤0.01 küçük, 0.01-0.06 orta, ≥0.06 büyük olarak kabul edilmiştir (Cohen,2013). Cohen f için ise ≤0.10 küçük, 0.10-0.25 orta, ≥0.25 büyük olarak kabul edilmiştir. Popülasyon olarak kabul üretilen 1000000'luk veri setleri üzerinde tek yönlü varyans analizi yapılmıştır. Yapılan analiz sonrasında Eta kare (η^2), Epsilon kare (ϵ^2), Omega kare ω^2 ve Cohen f yöntemleri ile etki büyüklüğünün parametre değerleri hesaplanmıştır. Daha sonra popülasyondan istenen örnek büyüklüğünde (5, 10, 20,30,50,100) örneklemler çekilmiştir. Çekilen bu örneklemlere ilişkin tek yönlü Varyans Analizi yapıldıktan sonra η^2, ϵ^2, ω^2, Cohen f yöntemleri için parametre tahminleri elde edilmiştir. Bu işlem her senaryo için biner kere tekrar edilmiş ve yapılan tahminlerin ortalama değeri ve %95 güven aralıkları hesaplanmıştır. Tüm analizler "R" paket programı kullanılarak yapılmıştır.

2.2 Etki Büyüklüğü Hesaplama Yöntemleri

Eta kare

Kelley (1935) tarafından önerilen Eta kare:

$$\hat{\eta}^2 = \frac{SS_{Effect}}{SS_{Toplam}} \quad (1)$$

SS_{effect}: İlgilenilen değişkene ait kareler toplamı

Epsilon kare

Kelley (1935) tarafından önerilen Epsilon kare :

$$\hat{\epsilon}^2 = \frac{SS_{Effect} - df_{Effect} MS_{Error}}{SS_{Total}} \quad (2)$$

Omega kare

Hays (1963) tarafından önerilen Omega kare:

$$\widehat{\omega}^2 = \frac{SS_{Effect} - df_{Effect} MS_{Error}}{SS_{Total} + MS_{Error}} \quad (3)$$

Cohen f

Cohen (1973) tarafından önerilen Cohen f:

$$\hat{f} = \sqrt{\frac{\eta^2}{1 - \eta^2}} \quad (4)$$

3. Bulgular

Küçük, Orta ve Büyük senaryolarda, k=3,4,5,10 ve n=5,10,20,30,50,100 değerlerine ilişkin tablolar aşağıda verilmiştir. Tablo 1'de etki büyüklüğü hesaplama yöntemlerinin 3 gruplu senaryo için farklı örneklem büyüklüğüne ve değişen etki büyüklüğü düzeylerine göre elde edilen ortalama değerleri ve %95 güven aralıkları verilmiştir. Bu tabloda verilen ortalama değerlerinin popülasyon parametre değerinden uzaklığına göre incelendiğinde;

- Etki büyüklüğünün küçük olduğu senaryoda tüm n değerleri için Omega kare ve Epsilon kare; n=50,100 olduğunda ise Eta kare tahminlerde başarılıdır.
- Etki büyüklüğünün orta olduğu senaryoda tüm n değerleri için Omega kare ve Epsilon kare; n=50,100 olduğunda ise Cohen f tahminlerde başarılıdır.
- Etki büyüklüğünün büyük olduğu senaryoda tüm n değerleri için Omega kare ve Epsilon kare; n=30,50,100 olduğunda Cohen f; n=50,100 olduğunda ise Eta kare tahminlerde başarılıdır.

Tablo 2'de etki büyüklüğü hesaplama yöntemlerinin 4 gruplu senaryo için farklı örneklem büyüklüğüne ve değişen etki büyüklüğü düzeylerine göre elde edilen ortalama değerleri ve %95 güven aralıkları verilmiştir. Bu tabloda verilen ortalama değerlerinin popülasyon parametre değerinden uzaklığına göre incelendiğinde;

- Etki büyüklüğünün küçük olduğu senaryoda tüm n değerleri için Omega kare; n=10,20,30,50,100 iken Epsilon kare; n=20,30,50,100 iken Eta kare tahminlerde başarılıdır.
- Etki büyüklüğünün orta olduğu senaryoda tüm n değerleri için Omega kare; n=10,20,30,50,100 iken Epsilon kare; n=100 için Eta kare ve Cohen f tahminlerde başarılıdır.
- Etki büyüklüğünün büyük olduğu senaryoda tüm n değerleri için Omega kare; n=10,20,30,50,100 iken Epsilon kare; n=30,50 ve 100 iken Eta kare ve Cohen f tahminlerde başarılıdır.

Tablo 3'te etki büyüklüğü hesaplama yöntemlerinin 5 gruplu senaryo için farklı örneklem büyüklüğüne ve değişen etki büyüklüğü düzeylerine göre elde edilen ortalama değerleri ve %95 güven aralıkları verilmiştir. Bu tabloda verilen ortalama değerlerinin popülasyon parametre değerinden uzaklığına göre incelendiğinde;

- Etki büyüklüğünün küçük olduğu senaryoda tüm n değerleri için Omega kare ve Epsilon kare; n=20,30,50,100 iken Eta kare tahminlerde başarılıdır.
- Etki büyüklüğünün orta olduğu senaryoda tüm n değerleri için Omega kare ve Epsilon kare; n=30,50,100 iken Cohen f; n=50,100 iken Eta kare tahminlerde başarılıdır.
- Etki büyüklüğünün büyük olduğu senaryoda tüm n değerleri için Omega kare ve Epsilon kare; n=20,30,50,100 iken Cohen f; n=30,50,100 iken Eta kare tahminlerde başarılıdır.

Tablo 4'te etki büyüklüğü hesaplama yöntemlerinin 10 gruplu senaryo için farklı örneklem büyüklüğüne ve değişen etki büyüklüğü düzeylerine göre elde edilen ortalama değerleri ve %95 güven aralıkları verilmiştir. Bu tabloda verilen ortalama değerlerinin popülasyon parametre değerinden uzaklığına göre incelendiğinde;

- Etki büyüklüğünün küçük olduğu senaryoda tüm n değerleri için Omega kare ve Epsilon kare; n=10,20,30,50,100 iken Eta kare tahminlerde başarılıdır.
- Etki büyüklüğünün orta olduğu senaryoda tüm n değerleri için Omega kare ve Epsilon kare; n=20,30,50,100 iken Eta kare tahminlerde başarılıdır.
- Etki büyüklüğünün büyük olduğu senaryoda tüm n değerleri için Omega kare ve Epsilon kare; n=20,30,50,100 iken Eta kare tahminlerde başarılıdır.

Tablo 1. Grup sayısı 3 olduğu durumda etki büyüklüğüne ve örneklem büyüklüğüne göre etki büyüklüğü hesaplama yöntemlerinin tahmin ortalamaları ve %95 güven aralıkları

Küçük etki büyüklüğü (k=3)				
	Eta kare	Omega kare	Epsilon kare	Cohen f
Popülasyon	0,004	0,004	0,004	0,063
N = 5	0,101 (0-0,346)	0,087 (0-0,281)	0,092 (0-0,294)	0,294 (0-0,768)
N = 10	0,049 (0-0,168)	0,042 (0-0,134)	0,043 (0-0,138)	0,191 (0-0,472)
N = 20	0,026 (0-0,093)	0,022 (0-0,076)	0,023 (0-0,077)	0,133 (0-0,331)
N = 30	0,02 (0-0,073)	0,017 (0-0,061)	0,017 (0-0,062)	0,115 (0-0,288)
N = 50	0,013 (0-0,049)	0,011 (0-0,042)	0,011 (0-0,042)	0,094 (0-0,233)
N = 100	0,009 (0-0,029)	0,007 (0-0,025)	0,007 (0-0,025)	0,076 (0-0,182)
Orta etki büyüklüğü (k=3)				
	Eta kare	Omega kare	Epsilon kare	Cohen f
Popülasyon	0,035	0,035	0,035	0,191
N = 5	0,13 (0-0,4)	0,104 (0-0,327)	0,11 (0-0,341)	0,35 (0-0,865)
N = 10	0,08 (0-0,253)	0,064 (0-0,214)	0,066 (0-0,22)	0,257 (0-0,613)
N = 20	0,057 (0-0,175)	0,047 (0-0,156)	0,048 (0-0,158)	0,215 (0-0,489)
N = 30	0,05 (0-0,141)	0,042 (0-0,127)	0,042 (0-0,128)	0,204 (0-0,431)
N = 50	0,042 (0-0,108)	0,037 (0-0,1)	0,037 (0-0,101)	0,193 (0,013-0,372)
N = 100	0,04 (0-0,087)	0,036 (0-0,084)	0,036 (0-0,084)	0,194 (0,062-0,325)
Büyük etki büyüklüğü (k=3)				
	Eta kare	Omega kare	Epsilon kare	Cohen f
Popülasyon	0,115	0,115	0,115	0,360
N = 5	0,177 (0-0,483)	0,138 (0-0,406)	0,145 (0-0,423)	0,442 (0-1,004)
N = 10	0,14 (0-0,347)	0,113 (0-0,308)	0,116 (0-0,316)	0,382 (0,002-0,761)
N = 20	0,128 (0-0,273)	0,112 (0-0,256)	0,114 (0-0,26)	0,371 (0,107-0,636)
N = 30	0,121 (0,005-0,236)	0,11 (0-0,225)	0,111 (0-0,227)	0,362 (0,153-0,571)
N = 50	0,119 (0,023-0,214)	0,112 (0,017-0,208)	0,113 (0,017-0,209)	0,362 (0,19-0,534)
N = 100	0,118 (0,054-0,182)	0,114 (0,05-0,178)	0,115 (0,05-0,179)	0,363 (0,249-0,477)

Tablo 2. Grup sayısı 4 olduğu durumda etki büyüklüğüne ve örneklem büyüklüğüne göre etki büyüklüğü hesaplama yöntemlerinin tahmin ortalamaları ve %95 güven aralıkları

Küçük etki büyüklüğü (k=4)				
	Eta kare	Omega kare	Epsilon kare	Cohen f
Popülasyon	0,003	0,003	0,003	0,058
N = 5	0,081 (0-0,275)	0,069 (0-0,223)	0,072 (0-0,231)	0,253 (0-0,648)
N = 10	0,039 (0-0,141)	0,034 (0-0,116)	0,034 (0-0,118)	0,168 (0-0,419)
N = 20	0,022 (0-0,08)	0,019 (0-0,068)	0,019 (0-0,069)	0,123 (0-0,305)
N = 30	0,016 (0-0,056)	0,013 (0-0,047)	0,013 (0-0,048)	0,104 (0-0,252)
N = 50	0,011 (0-0,04)	0,009 (0-0,035)	0,009 (0-0,035)	0,085 (0-0,21)
N = 100	0,007 (0-0,024)	0,006 (0-0,021)	0,006 (0-0,021)	0,069 (0-0,165)
Orta etki büyüklüğü (k=4)				
	Eta kare	Omega kare	Epsilon kare	Cohen f
Popülasyon	0,038	0,038	0,038	0,199
N = 5	0,084 (0-0,267)	0,068 (0-0,212)	0,071 (0-0,22)	0,263 (0-0,637)
N = 10	0,062 (0-0,194)	0,049 (0-0,165)	0,05 (0-0,168)	0,224 (0-0,515)
N = 20	0,048 (0-0,13)	0,039 (0-0,115)	0,039 (0-0,116)	0,204 (0-0,411)
N = 30	0,045 (0-0,111)	0,038 (0-0,101)	0,038 (0-0,102)	0,202 (0,028-0,377)
N = 50	0,044 (0-0,095)	0,039 (0-0,09)	0,039 (0-0,09)	0,204 (0,066-0,343)
N = 100	0,041 (0,005-0,077)	0,038 (0,002-0,075)	0,039 (0,002-0,075)	0,201 (0,103-0,3)
Büyük etki büyüklüğü (k=4)				
	Eta kare	Omega kare	Epsilon kare	Cohen f
Popülasyon	0,128	0,128	0,128	0,382
N = 5	0,165 (0-0,418)	0,128 (0-0,36)	0,133 (0-0,372)	0,423 (0-0,882)
N = 10	0,149 (0-0,337)	0,126 (0-0,311)	0,128 (0-0,317)	0,405 (0,073-0,736)
N = 20	0,138 (0,009-0,267)	0,125 (0-0,255)	0,127 (0-0,257)	0,392 (0,167-0,617)
N = 30	0,131 (0,028-0,234)	0,123 (0,02-0,226)	0,124 (0,02-0,228)	0,383 (0,204-0,563)
N = 50	0,131 (0,051-0,211)	0,126 (0,046-0,206)	0,126 (0,046-0,207)	0,385 (0,246-0,524)
N = 100	0,13 (0,075-0,184)	0,127 (0,072-0,182)	0,127 (0,073-0,182)	0,384 (0,29-0,479)

Tablo 3. Grup sayısı 5 olduğu durumda etki büyüklüğüne ve örneklem büyüklüğüne göre etki büyüklüğü hesaplama yöntemlerinin tahmin ortalamaları ve %95 güven aralıkları

Küçük etki büyüklüğü (k=5)				
	Eta kare	Omega kare	Epsilon kare	Cohen f
Popülasyon	0,004	0,004	0,004	0,063
N = 5	0,075 (0-0,245)	0,062 (0-0,2)	0,064 (0-0,206)	0,243 (0-0,601)
N = 10	0,039 (0-0,132)	0,033 (0-0,11)	0,034 (0-0,112)	0,169 (0-0,411)
N = 20	0,021 (0-0,073)	0,018 (0-0,061)	0,018 (0-0,062)	0,121 (0-0,293)
N = 30	0,016 (0-0,056)	0,013 (0-0,049)	0,014 (0-0,049)	0,102 (0-0,253)
N = 50	0,01 (0-0,035)	0,008 (0-0,031)	0,008 (0-0,031)	0,083 (0-0,199)
N = 100	0,008 (0-0,026)	0,007 (0-0,024)	0,007 (0-0,024)	0,073 (0-0,172)
Orta etki büyüklüğü (k=5)				
	Eta kare	Omega kare	Epsilon kare	Cohen f
Popülasyon	0,041	0,041	0,041	0,206
N = 5	0,097 (0-0,288)	0,077 (0-0,239)	0,08 (0-0,247)	0,29 (0-0,674)
N = 10	0,067 (0-0,206)	0,056 (0-0,182)	0,057 (0-0,185)	0,233 (0-0,541)
N = 20	0,053 (0-0,146)	0,045 (0-0,134)	0,046 (0-0,135)	0,213 (0-0,441)
N = 30	0,049 (0-0,128)	0,044 (0-0,12)	0,044 (0-0,121)	0,209 (0,011-0,408)
N = 50	0,046 (0-0,103)	0,042 (0-0,099)	0,042 (0-0,1)	0,208 (0,058-0,359)
N = 100	0,044 (0,004-0,084)	0,042 (0,002-0,082)	0,042 (0,002-0,082)	0,208 (0,102-0,314)
Büyük etki büyüklüğü (k=5)				
	Eta kare	Omega kare	Epsilon kare	Cohen f
Popülasyon	0,126	0,126	0,126	0,379
N = 5	0,152 (0-0,366)	0,12 (0-0,32)	0,124 (0-0,329)	0,405 (0,019-0,79)
N = 10	0,138 (0-0,289)	0,119 (0-0,269)	0,121 (0-0,273)	0,39 (0,12-0,659)
N = 20	0,131 (0,016-0,246)	0,121 (0,006-0,236)	0,122 (0,006-0,238)	0,381 (0,18-0,583)
N = 30	0,128 (0,039-0,217)	0,122 (0,032-0,211)	0,122 (0,032-0,212)	0,379 (0,223-0,536)
N = 50	0,129 (0,06-0,198)	0,125 (0,056-0,194)	0,125 (0,056-0,195)	0,382 (0,263-0,501)
N = 100	0,128 (0,078-0,177)	0,126 (0,076-0,175)	0,126 (0,077-0,175)	0,381 (0,296-0,467)

Tablo 4. Grup sayısı 10 olduğu durumda etki büyüklüğüne ve örneklem büyüklüğüne göre etki büyüklüğü hesaplama yöntemlerinin tahmin ortalamaları ve %95 güven aralıkları

Küçük etki büyüklüğü (k=10)				
	Eta kare	Omega kare	Epsilon kare	Cohen f
Popülasyon	0,003	0,003	0,003	0,058
N = 5	0,038 (0-0,132)	0,032 (0-0,11)	0,032 (0-0,112)	0,166 (0-0,405)
N = 10	0,02 (0-0,073)	0,017 (0-0,062)	0,017 (0-0,063)	0,116 (0-0,29)
N = 20	0,012 (0-0,041)	0,01 (0-0,035)	0,01 (0-0,035)	0,091 (0-0,219)
N = 30	0,01 (0-0,033)	0,008 (0-0,029)	0,008 (0-0,029)	0,082 (0-0,192)
N = 50	0,007 (0-0,024)	0,006 (0-0,022)	0,006 (0-0,022)	0,069 (0-0,164)
N = 100	0,006 (0-0,016)	0,005 (0-0,015)	0,005 (0-0,015)	0,065 (0-0,138)
Orta etki büyüklüğü (k=10)				
	Eta kare	Omega kare	Epsilon kare	Cohen f
Popülasyon	0,036	0,036	0,036	0,193
N = 5	0,058 (0-0,165)	0,046 (0-0,141)	0,046 (0-0,143)	0,219 (0-0,475)
N = 10	0,05 (0-0,122)	0,041 (0-0,111)	0,041 (0-0,112)	0,212 (0,028-0,396)
N = 20	0,045 (0-0,098)	0,041 (0-0,093)	0,041 (0-0,094)	0,209 (0,07-0,347)
N = 30	0,043 (0,003-0,084)	0,04 (0-0,081)	0,04 (0-0,081)	0,207 (0,1-0,313)
N = 50	0,043 (0,010-0,076)	0,041 (0,008-0,074)	0,041 (0,008-0,075)	0,208 (0,12-0,296)
N = 100	0,041 (0,019-0,063)	0,04 (0,018-0,062)	0,04 (0,018-0,063)	0,205 (0,146-0,264)
Büyük etki büyüklüğü (k=10)				
	Eta kare	Omega kare	Epsilon kare	Cohen f
Popülasyon	0,122	0,122	0,122	0,373
N = 5	0,150 (0,004-0,295)	0,13 (0-0,275)	0,132 (0-0,28)	0,411 (0,16-0,663)
N = 10	0,146 (0,046-0,245)	0,136 (0,036-0,236)	0,137 (0,037-0,237)	0,409 (0,24-0,579)
N = 20	0,140 (0,070-0,210)	0,135 (0,065-0,206)	0,136 (0,065-0,206)	0,401 (0,283-0,52)
N = 30	0,137 (0,081-0,194)	0,134 (0,078-0,190)	0,134 (0,078-0,191)	0,398 (0,303-0,493)
N = 50	0,139 (0,095-0,182)	0,137 (0,093-0,180)	0,137 (0,093-0,18)	0,4 (0,327-0,474)
N = 100	0,137 (0,106-0,168)	0,136 (0,105-0,167)	0,136 (0,105-0,167)	0,398 (0,346-0,45)

4. Tartışma ve Sonuç

Grup karşılaştırmalarında yaygın olarak kullanılan varyans analizi yöntemi (ANOVA-F testi) ile elde edilen "P" değeri sadece istatistiksel önemlilik hakkında bilgi vermektedir. İlgilenilen faktörlerin pratik/klinik önemliliği ve hakkında bilgi verme yönünden zayıf ve eksik kalmaktadır. Varyans analizi sonuçları rapor edilirken "P" değerinin yanında geliştirilmiş çok sayıda etki büyüklüğü ölçüsü vardır. En doğru etki büyüklüğünü rapor etmek için etki büyüklüğü yöntemlerinin tahmin başarısı incelenmelidir.

Keselman (1975) tek yönlü varyans analizinde hangi etki büyüklüğünün en yansız etki büyüklüğü olduğunu göstermek için Eta kare, Omega kare ve Epsilon kare'yi karşılaştırmıştır. 5000 simulasyon denemesi sonucunda Eta-Kare'nin daha düşük standart sapmalı olması nedeniyle daha yansız bir tahmin edici olduğunu belirtmiştir. Okada (2013) ise aynı çalışmayı 1000000

simulasyon denemesi ile yapmıştır ve en yansız sonuçları Omega kare'nin verdiğini belirtmiştir. Yiğit ve Mendeş (2016) yaptıkları çalışmada çeşitli senaryolar altında Omega-kare ve Epsilon-kare'nin daha güvenilir sonuçlar verdiğini bildirmişlerdir. Buradan hareketle ele alınan etki büyüklüğü ölçülerinin farklı senaryolar altında doğru tahminlerine ilişkin yapılan çalışmaların çelişkilerinden dolayı bu çalışmada Eta kare, Omega kare, Epsilon kare ve Cohen f değerlerinin tahmin başarısı incelenmiştir.

Sık kullanılmasına rağmen Eta kare ve Cohen'in, popülasyona ait etki büyüklüğünün gerçek değerini tahmin etme başarısının Omega kare ve Epsilon kare kadar iyi olmadığı görülmüştür. Eta karedeki yanlılığın sebebi eşitlik (1) incelendiğinde dikkat edileceği üzere hesap yapılırken sadece SS_{Effect} ile ilgilenmekten kaynaklanmaktadır. Eşitlik (2)'de Eta karenin payından etkiye ait serbestlik derecesi ile hata kareler ortalamasının çarpımı çıkarılmıştır ve Epsilon karenin paydasına eklenen MS_{Error} ile de Omega kare değeri elde edilmiştir. Burada yapılan düzeltmeler örnek seçiminden kaynaklı hatayı azaltmaktadır. Grup sayısı 10 olduğunda ise çalışmaya alınan etki büyüklüğü hesaplama yöntemlerinin hepsinde tahmin başarısı oldukça düşmektedir (gerçek değerden uzaklaşmaktadır). Tüm senaryolar dikkate alındığında ise Omega ve Epsilon karenin birbirine yakın ve diğer etki büyüklüğü hesap yöntemlerinden daha yansız sonuçlar verdiği görülmüştür.

Kaynaklar

1. Sullivan, G. M., & Feinn, R. (2012). Using effect size—or why the P value is not enough. *Journal of graduate medical education*, *4*(3), 279-282.
2. Skidmore, S. T., & Thompson, B. (2013). Bias and precision of some classical ANOVA effect sizes when assumptions are violated. *Behavior research methods*, *45*(2), 536-546.
3. Cohen, J. (2013). *Statistical power analysis for the behavioral sciences*. Routledge.
4. Kelley, T. L. (1935). An Unbiased Correlation Ratio Measure, Proceedings of the National Academy of Sciences of the United States of America, 21, 9, 554–559.
5. Hays, W. (1963). STATISTICS FOR PSYCHOLOGISTS, Rinehart Winston, New York.
6. Cohen, J. (1973). Eta-Squared and Partial Eta-Squared in Fixed Factor Anova Designs, Educational and Psychological Measurement, 33, 1, 107–112.
7. Keselman, H. J. (1975). A Monte Carlo investigation of three estimates of treatment magnitude: Epsilon squared, eta squared, and omega squared. *Canadian Psychological Review/Psychologie Canadienne*, *16*(1), 44.
8. Okada, K. (2013). Is omega squared less biased? A comparison of three major effect size indices in one-way ANOVA. *Behaviormetrika*, *40*(2), 129-147.
9. Yigit, S., & Mendes, M. (2018). Which effect size measure is appropriate for one-way and two-way ANOVA models? A Monte Carlo simulation study. Revstat Statistical Journal, 16, 295-313.

T4

Randomize Olmayan Klinik Çalışmalarda En Uygun Eşleştirme Analizi için Makine Öğrenme Algoritmaları ile Yeni Propensity Skor Tahmin Modellerinin Geliştirilmesi ve R Shiny ile İnteraktif Bir Web Uygulaması

Development of New Propensity Score Estimation Models with Machine Learning Algorithms for Optimal Matching Analysis in Non-Randomized Clinical Trials and an Interactive Web Application with R Shiny

Emre DEMİR[1], Serdal Kenan KÖSE[2]

[1]*Hitit Üniversitesi, Tıp Fakültesi, Biyoistatistik Anabilim Dalı, Çorum*

[2]*Ankara Üniversitesi, Tıp Fakültesi, Biyoistatistik Anabilim Dalı, Ankara*

emredemir82@gmail.com

ÖZET

Amaç: Deneysel araştırmalarda deney sonucunu etkileyebilecek karıştırıcı etkenler randomizasyon ile kontrol edilirken gözlemsel çalışmalarda deneklerin araştırma gruplarına atanma olasılığı eşit olmadığı için karıştırıcı değişkenler göz ardı edilerek ulaşılan sonuçlarda yanlılık bulunmaktadır. Bu nedenle gözlemsel çalışmalarda bir tedavinin nedensel etkisi araştırılırken yaygın olarak Propensity skor (PS)'a dayalı eşleştirme analizi kullanılmaktadır. Bu çalışmada PS tahmini için klasik yöntem olan Lojistik Regresyon (LR)'a alternatif olarak Makine öğrenme (ML) ve bazı yaygın kullanılan yöntemlerin başarısı benzetim çalışması ile değerlendirilerek yeni bir kombine PS tahmin modeli geliştirmek amaçlanmıştır.

Yöntem: Birinci benzetim çalışmasında (10000 benzetim) etkileşim etkilerinin ve doğrusal olmayan ilişkilerin bulunduğu modellerde LR ile rekabet edebilecek algoritmaların belirlenmesi için öncelikle LR için avantajlı olan modelde (Senaryo A: sadece temel etki ve toplamsal model) n=500 ve n=1000 örneklem büyüklüklerinde üretilen veri seti kullanılarak 21 farklı tahminci ve LR ile PS'ler hesaplanmıştır. PS'ler arasındaki uyumu değerlendirmek için sınıf içi korelasyon katsayısı kullanılmıştır. İkinci benzetim çalışmasında (1000 benzetim) A senaryosuna ek olarak etkileşim etkilerinin ve kuadratik terimlerinde bulunduğu B ve C senaryolarında ilk benzetim sonucunda başarılı bulunan 12 algoritma ve LR ile elde edilen PS'ler kullanılarak eşleştirme analizleri yapılmıştır. Elde edilen denge (Tedavi grubunda ortalama tedavi etkisi, yanlılık, standardize edilmiş ortalamalar arasındaki farklar, varyans oranları ve kolmogorov smirnov değerleri) sonuçlarına göre başarılı tahminciler ile yeni bir kombine tahminci model oluşturularak üçüncü benzetim çalışması yapılmıştır. Üçüncü benzetim çalışmasında (1000 benzetim) üç farklı senaryoda kombine tahminci ve LR ile elde edilen PS'ler kullanılarak yapılan eşleştirme analizi denge sonuçları karşılaştırılmıştır. PS tahmini ve eşleştirme analizi için interaktif web uygulaması R Shiny paketi ile geliştirilmiştir.

Bulgular: Birinci benzetim sonucunda LR ile en uyumlu algoritmalar BAYESGLM, GLMNET, GAM, STEPAIC, DBARTS, GBM, EARTH, QDA, KSVM, CFOREST, NNET ve IPREDBAGG olarak belirlenmiştir. İkinci benzetim sonucunda 1000 örneklem büyüklüğünde A senaryosunda LR ile

rekabet edebilen, B ve C senaryolarında LR'den daha başarılı olan KSVM, NNET, DBARTS, QDA, EARTH ve STEPAIC algoritmaları ile kombine tahmin modeli oluşturulmuştur. Üçüncü benzetim sonucunda sırasıyla A, B ve C senaryolarında LR %0,496; %10,160; %17,762 ve kombine tahminci %0,527; %7,405; %7,672 yanlılık ile kestirimde bulunmuştur.

Sonuç: Denge sonuçlarına göre n=500 ve n=1000 örneklem büyüklüğünde kombine tahmin modelinin senaryo zorlaştıkça LR'ye göre daha az yanlılıkla tedavi etkisini tahmin ettiği belirlenmiştir. Özellikle n=1000 örneklem büyüklüğünde en zor senaryo olan C senaryosunda kombine tahmincinin LR'den %10,09 daha az yanlı kestirimde bulunması ve LR'nin yüksek bir yanlılık kestiriminde bulunması ML veya diğer algoritmaların PS tahmininde kullanılmasının yanlılığı azaltarak gözlemsel çalışmaların güvenirliğini arttıracağını göstermektedir.

Anahtar Kelimeler: Propensity skor, eşleştirme, lojistik regresyon, makine öğrenme, gözlemsel çalışmalar

1. GİRİŞ

Randomizasyon, hastaların deney gruplarına şansa bağlı olarak atanmasını sağlayarak seçim yanlılığının oluşmasını engellemek ve klinik denemelerde etkinliği araştırılan tedavi dışındaki tüm ortak değişkenler/faktörler için grupların benzer dağılım özelliklerine sahip olmasını sağlamak amacıyla yapılır [1]. Deneysel araştırmalarda deney sonucunu etkileyebilecek karıştırıcı etkenler randomizasyon ile kontrol edilirken gözlemsel çalışmalarda deneklerin vaka ve kontrol gruplarına atanma olasılığı eşit olmadığı için karıştırıcı ortak değişkenler göz ardı edilerek ulaşılan sonuçlarda yanlılık bulunmaktadır [2-3]. Bu nedenle gözlemsel çalışmalarda bir tedavinin sonuç üzerindeki nedensel etkisi araştırılırken yaygın olarak Propensity skor (PS)'a dayalı eşleştirme analizi kullanılmaktadır [4-6]. Eşleştirmenin amacı kontrol grubu içerisinden tedavi grubundaki bireylerin sahip olduğu ortak değişken değerlerine benzer özellikteki bireyleri seçmektir. Geleneksel yöntemlerde (tabakalara ayırma, kısıtlama, kovaryans düzeltmesi) yanlılığı gidermek için sınırlı sayıda ortak değişken kullanabilmekte iken PS çok sayıda ortak değişken bilgisinin özet ölçüsünü yansıtmaktadır [7].

Bu çalışmada deneklerin tedavi gruplara atanma olasılığı ile ortak değişkenler arasında ana etkiler dışında etkileşim etkilerinin ve doğrusal olmayan ilişkilerin bulunduğu modellerde PS tahmini için klasik yöntem olan Lojistik Regresyon (LR)'a alternatif olarak Makine öğrenme (ML) ve bazı yaygın kullanılan yöntemlerin başarısını benzetim çalışması ile değerlendirmek suretiyle yeni bir kombine (ensemble) PS tahmin modeli geliştirmek amaçlanmıştır. Ayrıca R Shiny paketinde klasik model ve geliştirilen kombine model ile PS'ye dayalı eşleştirme analizi yapabilen interaktif bir web uygulamasının geliştirilmesi amaçlanmıştır.

2. YÖNTEM

2.1. Nedensel Çıkarım

Diğer tüm şartların eşit olduğu durumlarda A eyleminin olduğu ve olmadığı iki durum arasında karşılaştırma yapıldığında eğer iki sonuç farklıysa, A eyleminin sonuç üzerinde nedensel bir etkiye sahip olduğu söylenecektir. Epidemiyolojide, A genellikle maruziyet veya tedavi olarak adlandırılmaktadır [8].

2.2. Propensity Skor

Rosenbaum ve Rubin (1983), PS'yi, gözlenen ortak değişkenlere göre belirli bir tedavi grubuna atanmanın koşullu olasılığı olarak tanımlamıştır. e(x); PS'yi, Ti; tedavi (Ti = 1) ve kontrol (Ti = 0) grubunu göstermek üzere, ortak değişkenlere (x) bağlı olarak tedaviye atanma olasılığı eşitlik 1'deki gibi ifade edilmiştir [4].

$$e(x) = P(T = 1|x),\ P(T_1, \dots, T_n|x_1, \dots, x_n) = \prod_{i=1}^{N} e(x_i)^{T_i}\ \big(1 - e(x_i)\big)^{1-T_i} \qquad (1)$$

Randomize çalışmalarda Ti rastgele bir mekanizmaya göre belirlenen bir dağılıma sahip olduğu için gözlemsel çalışmalardan farklılaşmaktadır. Çünkü randomize çalışmalarda PS, kabul edilen bir olasılığa dayalı bilinen bir fonksiyondur [6]. Gözlemsel çalışmaların neredeyse tamamında ise PS fonksiyonu belirli bir olasılığa dayalı olmadığı için bilinmemektedir. Ancak bir lojit model kullanılarak gözlenen ortak değişkenlere bağlı olarak tahmin edilebileceği Rosenbaum ve Rubin (1983) tarafından gösterilmiştir [4].

2.3. Eşleştirme (Matching)

Eşleştirme, kontrol grubu içerisinden tedavi grubunun ortak değişken dağılımına en benzer olan grubu belirlemek için kullanılan bir örnekleme yöntemidir [4-6]. PS' ye dayalı eşleştirme analizinde kontrol grubunda propensity skoru P_j olan j'inci gözlem tedavi grubunda propensity skoru P_i olan bir gözlem ile eşleşir. Burada propensity skorları arasındaki farkı en küçük olan P_i ve P_j seçilir. Propensity skorları yakın olmayan gözlemlerin eşleştirilmesi problemine karşı "kabul edilebilir fark (ε=caliper)" sınırlaması önerilmiştir. En yakın eşleştirme $\left|P_i - P_j\right| < \varepsilon$ olacak şekilde önceden belirlenen bir toleransa göre (ε) gerçekleşir [9].

2.3.1. Ortalama Farklar (OF), Standardize Edilmiş Ortalama Farklar (SOF) ve Varyans Oranları (VO)

$\bar{x}_t$ tedavi, $\bar{x}_c$ kontrol grubundaki herhangi bir ortak değişken ortalamasını, s_t^2 tedavi, s_c^2 kontrol grubundaki herhangi bir ortak değişken varyansını göstermek üzere sürekli değişkenler için OF, SOF ve VO;

$$OF = \bar{x}_t - \bar{x}_c, SOF = \frac{\bar{x}_t - \bar{x}_c}{\sqrt{\frac{s_t^2 + s_c^2}{2}}}, \quad VO = \frac{s_t^2}{s_c^2} \qquad (2)$$

$\bar{p}_t$ tedavi, $\bar{p}_c$ kontrol gurubunun prevalansını göstermek üzere kategorik değişkenler için OF, SOF ve VO;

$$OF = \bar{p}_t - \bar{p}_c, \quad SOF = \frac{\bar{p}_t - \bar{p}_c}{\sqrt{\frac{\bar{p}_t(1-\bar{p}_t) + \bar{p}_c(1-\bar{p}_c)}{2}}}, VO = \frac{p_t*(1-p_t)}{p_c*(1-p_c)} \qquad (3)$$

olarak tanımlanmıştır [10,11].

SOF örneklem büyüklüğünden etkilenmemekte ve farklı birimde ölçülen değişkenlerin göreceli olarak karşılaştırılmasına olanak sağlamaktadır. Mutlak SOF'un 0 olması grupların benzer olduğunu ve dengenin sağlandığını göstermektedir (<0,1 dengenin iyi olduğunu gösterir). Varyans oranının 1'e yakın olması dengenin iyi olduğunu gösterir (1 olması tam dengeyi göstermektedir).

2.3.2. Kolmogorov Smirnov (KS) Test İstatistiği

KS istatistiği tedavi gruplarındaki örneklemlerin ağırlıklandırılmış deneysel dağılım fonksiyonlarına dayanmaktadır. Bu test tedavi ve kontrol grupları arasında ortak değişkenlere göre dağılımın uyuşup uyuşmadığını belirlemeye yardımcı olmaktadır [12]. KS sadece ortalamayı değil tüm dağılımı karşılaştırmaktadır ancak SOF'tan farklı olarak istatistik değeri örneklem büyüklüğüne bağlıdır. KS istatistiğinin 0'a yakın olması dengenin sağlandığını göstermektedir.

2.3.3. Ortalama Tedavi Etkisi (ATE) ve Tedavi Grubunda Ortalama Tedavi Etkisi (ATT)

Y sonuç değişkeni, T_i iki düzeyli bir tedavi değişkeni ve bir birey tedavi aldığı durumda T=1 ve almadığı durumda T=0 olmak üzere, i. birey için (i = 1, …, n) tedavi etkisi aşağıdaki gibi tanımlanmıştır [13].

$$\tau_i = Y_i(1) - Y_i(0) \quad (4)$$

Ancak gözlemsel çalışmalarda her birey için sadece bir potansiyel sonuç değeri olacağı için bu durumda temel bir problem oluşmaktadır ve gözlenemeyen sonuç değişkeni karşı olgu (counter-factual) olarak adlandırılmaktadır [14]. Bu yüzden i. birey için tedavi etkisi hesaplanması mümkün olmadığı için araştırmalarda ATE ve ATT kavramlarına odaklanılmıştır. Eğer bireyler tedavi ve kontrol gruplarına rastgele atanmışsa ve tedavi sonrası gruplar arasında anlamlı fark bulunmuşsa farklılığa tedavinin yol açtığı söylenir ve T∈ {0,1} ve Y sonuç değişkeni olmak üzere ATE aşağıdaki eşitlikteki gibi tanımlanabilir [11].

$$E(Y \mid T=1) - E(Y \mid T=0) \quad (5)$$

Gözlemsel çalışmalarda ise ATT eşitlik 6'daki gibi tanımlanabilir [11]

$$E(Y(1) \mid T=1) - E(Y(0) \mid T=1) \quad (6)$$

Gözlemsel çalışmalarda tedavi almayan bireylerin sonuç değişkenlerinin ortalamasını kullanmak anlamsız olacaktır. Çünkü bir bireyin tedavi almasını etkileyen sebepler sonuç değişkenini de etkilemiştir. Yani tedavi verilmeden öncede tedavi ve kontrol grupları arasında fark olması muhtemeldir [15]. Randomizasyonun yapıldığı ve sonuç değişkeninden bağımsız olarak tüm bireylerin tedavi grubuna atanmasında eşit olasılığa sahip olduğu bir deneysel çalışmada ATE, ATT' ye eşit olacaktır. Çünkü randomizasyondan dolayı kontrol grubu, tedavi grubunun gözlemlenemeyen karşı olgularının yerine geçecektir. Randomizasyonun yapılamadığı çalışmalarda ise ATE≠ATT olacaktır. Çünkü bazı bireylerin tedavi grubuna atanma olasılığı sıfırdır. Bu yüzden gözlemsel çalışmalarda tedavi grubundaki bireylerin PS'sine en yakın olan kontrol grubundaki bireyler seçilerek tedavi atama mekanizmasının bağımsızlığı ile ilgili kritik varsayım teorik olarak tatmin edilmiş olur [6].

2.4. Benzetim Çalışması

İki düzeyli tedavi değişkenini sınıflamak ve PS değerlerini hesaplamak için 21 farklı algoritma kullanılmıştır. Sadece ana etkilerin olmadığı modellerde daha başarılı olacağı düşünülen ML algoritmalarının sadece temel etkiler olan modellerde de en az LR kadar başarılı olması gerekmektedir. Bu yüzden PS tahmininde kullanılacak en iyi kombine algoritma içerisinde

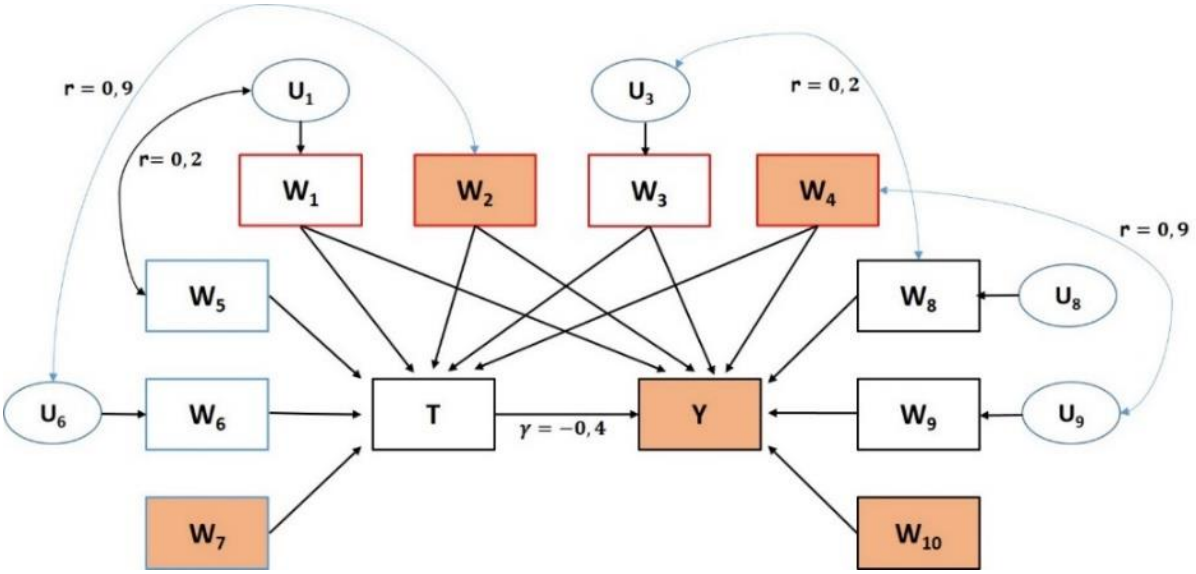
bulunacak aday algoritmaları belirleyebilmek için öncelikle sadece A senaryosu kullanılarak orta (n=500) ve büyük (n=1000) olmak üzere 2 farklı örneklem büyüklüğünde veri seti üretildi. A senaryosu ile üretilen 2 farklı örneklem büyüklüğündeki veri seti kullanılarak 21 farklı tahminci yöntemin her biri için 10000 benzetim ile PS'ler hesaplandı. 21 farklı benzetim içerisinde her bir yöntem ile birlikte LR kullanılarak da PS hesaplandı. Her bir benzetim çalışması içerisinde ML yöntemleri ve LR ile hesaplanan PS'ler arasındaki uyum istatistikleri hesaplandı. Birinci benzetim çalışması için toplamda 420,000 kez PS hesaplandı. PS'ler arasındaki uyumu değerlendirmek için Sınıf içi korelasyon katsayısı (SKK) kullanıldı.

Araştırmanın ikinci bölümünde A senaryosuna ek olarak etkileşim etkilerinin ve kuadratik terimlerinde bulunduğu B ve C senaryoları için 500 ve 1000 örneklem büyüklüğünde veri üretilerek ilk benzetim sonucunda başarılı uyum düzeyine sahip bireysel algoritmalar ve LR ile elde edilen PS'ler kullanılarak eşleştirme analizleri gerçekleştirildi. Üretilen 2 farklı örneklem büyüklüğündeki veri setleri için her bir benzetim içerisinde LR ve başarılı bulunan 12 tahmincinin her biri ile 1000 benzetim kullanılarak PS'ler hesaplandı ve bu PS'ler ile eşleştirme analizleri yapıldı. Toplamda 26,000 kez hesaplanan PS'ler ile 26,000 eşleştirme analizi gerçekleştirildi. Eşleştirme analizleri sonucunda elde edilen denge değerlendirmesi sonuçlarına göre tahminciler karşılaştırıldı. Denge açısından başarılı olarak değerlendirilen tahminciler ile PS hesabı için yeni bir kombine tahminci model oluşturularak son bir benzetim çalışması yapıldı. Son benzetim çalışmasında kombine tahminci ve klasik model olan LR ile iki farklı örneklem büyüklüğünde 1000 benzetim yapılarak elde edilen PS'ler ile yapılan eşleştirme analizi denge sonuçları karşılaştırıldı. Toplamda 12,000 kez hesaplanan PS'ler ile 12,000 eşleştirme analizi gerçekleştirildi. Denge değerlendirmeleri için tüm benzetimlerde ATT ve yanlılık değeri, SOF, OF, VO ve KS değerleri kullanıldı.

Benzetim ve uygulamalarda ML algoritmaları ve diğer tahminciler için R paket programında açık kaynak kodlu SuperLearner ve alt kütüphaneleri kullanıldı [16]. SuperLearner paketi, Polley ve Van Der Laan (2010), tarafından parametrik regresyon modelleri ve birçok ML algoritması içerisinden en başarılı kestirim elde eden yöntemin veya kombine yöntemlerin seçimi için geliştirilmiştir. Paket V-fold cross-validation kullanarak hesapladığı kayıp değer fonksiyonuna göre birçok aday ML algoritmaların doğrusal bileşimini ağırlıklandırarak yeni bir kombine tahminci (ensemble learning) sağlamaktadır. Eşleştirme analizi ve PS tahmini Sekhon (2011) tarafından geliştirilen Matching kütüphanesi kullanılarak yapılmıştır [17]. Denge değerlendirmesi için kullanılan grafik çizimleri ise "ggplot2" kütüphanesi kullanılarak yapılmıştır. Kestirilen PS'lerin uyum analizleri için "irr" kütüphanesi kullanılmıştır [18]. Araştırmamızda eşleştirme analizleri için PS'ye dayalı Nearest neighbor matching yöntemi kullanılmıştır. Eşleştirme analizi uygulamalarında kontrol grubundaki bir deneğin tedavi grubundaki birden fazla denek ile eşleşmesine izin veren "with replacement" seçeneği kullanılmıştır.

2.5. Benzetim veri seti

Monte Carlo Benzetim için Setoguchi ve ark. (2008)'nın önerdiği ve birçok araştırmada kullanılan benzetim veri seti kullanılmıştır (Şekil 1) [19-23]. Setoguchi ve ark. (2008)'nın önerdiği veri setinde tedavi etkisini gösteren sonuç değişkeni (Y_i), tedavi ve ortak değişkenlerin bir doğrusal kombinasyonu olacak şekilde Lee ve ark. (2010) tarafından değiştirilmiştir. Çalışmamızda da bu düzeltme yapılmıştır.



Şekil 1. Benzetim veri seti

Veri setinde tedavi ve kontrol gruplarını gösteren 1 kategorik değişken (T), tedavi etkisini gösteren 1 sürekli sonuç değişkeni (Y), 4 farklı karıştırıcı değişken (W_1-W_4), sadece tedavi değişkeni ile ilişkili 3 farklı ortak değişken (W_5-W_7) ve sadece sonuç değişkeni ile ilişkili 3 farklı ortak değişken (effect modificator; W_8-W_{10}) bulunmaktadır. Bu değişkenlerden W_1, W_3, W_5, W_6, W_8, W_9 iki düzeyli (0,1) kategorik değişken ve W_2, W_4, W_7, W_{10} sürekli değişken olarak üretilmiştir. Değişkenlerden W_1-W_4, W_7 ve W_{10} standart normal dağılımdan üretilmiştir ($\mu = 0, \sigma = 1$). Daha sonra W_5, W_6, W_8 ve W_9 üretilmiştir. W_1 ile W_5 ve W_3 ile W_8 arasında 0,2 korelasyon, W_2 ile W_6 ve W_4 ile W_9 arasında 0,9 korelasyon bulunacak şekilde üretilmiştir. Son olarak W_1, W_3, W_5, W_6, W_8, W_9 değişkenleri içerisindeki her bir veri genel ortalama ile karşılaştırılarak ortalamadan büyük olanlar 1, küçük olanlar 0 olacak şekilde kategorik değişkene dönüştürülmüştür. Sonuç değişkeni (Y), tedavi değişkeni ve ortak değişkenlerin bir doğrusal kombinasyonu olarak üretilmiştir ($Y = a_0 + a_i W_i + \gamma T$). γ tedavi etkisi olmak üzere -0,4 olarak üretilmiştir. Tedavi değişkeni iki düzeyli kategorik bir değişken olarak ortak değişkenlerin bir fonksiyonu olacak şekilde logit model ($P(T=1|W_i) = 1/(1+\exp(-\beta \cdot f(W_i)))$) kullanılarak üretilmiştir. Tedavi alma veya kontrol grubunda olma olasılığını belirleyecek bu modelde $f(W_i)$ senaryolara göre ortalama maruziyet olasılığı yaklaşık 0,5 olacak şekilde Bernoulli dağılımından üretilmiştir.

3. BULGULAR

3.1. Birinci Benzetim sonuçları; Lojistik regresyon ile en uyumlu algoritmalarının belirlenmesi

A senaryosu ile üretilen n=500 ve n=1000 veri seti kullanılarak LR ve 21 farklı ML algoritması ile 10000 benzetim sonucunda hesaplanan PS değerleri arasındaki SKK değerleri güven aralıkları ile birlikte Tablo 1'de verilmiştir.

Tablo 1. Farklı algoritmalarla kestirilen PS'ler arasındaki SKK uyum değerleri

Algoritmalar	**SKK**	
	n=500	**n=1000**
BAYESGLM (Bayesci genelleştirilmiş doğrusal model)	1.000 (1.000-1.000)	1.000 (1.000 - 1.000)
CARET (Sınıflama ve regresyon ağaçları)	0.619 (0.562-0.670)	0.657 (0.620-0.690)
CARET.RPART (Özyinelemeli parçalama)	0.549 (0.485-0.606)	0.546 (0.502-0.588)
CFOREST (Koşullu rasgele ormanlar)	0.877 (0.855-0.896)	0.896 (0.883-0.907)
DBARTS (Kesikli bayesci eklemeli regresyon ağaçları)	0.956 (0.948-0.963)	0.966 (0.962-0.970)
EARTH (Çok değişkenli uyarlanabilir regresyon eğrileri)	0.894 (0.875-0.910)	0.943 (0.936-0.950)
GBM (Boosted model)	0.909 (0.892-0.923)	0.943 (0.936-0.950)
GAM (Genelleştirilmiş toplamsal model)	0.993 (0.992-0.995)	0.997 (0.996-0.997)
GLMNET (Cezalandırılmış, genelleştirilmiş lineer modeller)	0.995 (0.993-0.995)	0.998 (0.998-0.998)
IPREDBAGG (Geliştirilmiş tahmin modelleri)	0.824 (0.794-0.851)	0.840 (0.821-0.857)
KERNELKNN (Çekirdek k en yakın komşu algoritması)	0.751 (0.710-0.787)	0.769 (0.743-0.794)
KNN (k en yakın komşuluk)	0.752 (0.712-0.788)	0.769 (0.743-0.794)
KSVM (Çekirdek destek vektör makineleri)	0.910 (0.894-0.924)	0.929 (0.921-0.937)
NNET (Yapay sinir ağları)	0.791 (0.757-0.822)	0.885 (0.871-0.897)
QDA (Karesel (kuadratik) diskriminant analizi)	0.876 (0.854-0.895)	0.930 (0.921 0.938)
RANDOMFOREST (Rasgele ormanlar)	0.563 (0.501-0.620)	0.564 (0.521-0.605)
RPART (Özyinelemeli parçalama regresyon ağaçları)	0.634 (0.579-0.684)	0.663 (0.627-0.696)
RPARTPRUNE (Budamalı özyinelemeli regresyon ağaçları)	0.602 (0.543-0.655)	0.641 (0.603-0.676)
STEPAIC (Adımsal Akaike bilgi kriteri)	0.973 (0.968-0.977)	0.988 (0.986-0.989)
SVM (Destek vektör makineleri)	0.614 (0.557-0.665)	0.624 (0.585-0.660)
XGBOOST (Ölçeklenebilir gradient boosting)	0.559 (0.496-0.616)	0.555 (0.510 - 0.596)

SKK: Sınıf içi korelasyon katsayısı.

SKK uyum katsayıları güven aralığında 0.80 ve üzeri olan BAYESGLM, GLMNET, GAM, STEPAIC, DBARTS, GBM, EARTH, QDA, KSVM, CFOREST, NNET ve IPREDBAGG algoritmaları LR ile uyumlu algoritmalar olarak bulunmuştur. İkinci ve üçüncü benzetim çalışmasında sadece LR ile uyumlu bulunan algoritmalar A, B ve C senaryoları için gerçekleştirilen benzetimlere dahil edilmiştir.

3.2. İkinci Benzetim Sonuçları; Kombine tahminci model için tahminci aday algoritmalarının belirlenmesi

3.2.1. n=500 Örneklem Büyüklüğü için Benzetim Sonuçları

İkinci benzetim çalışmasında LR ve LR ile uyumlu olarak belirlenen 12 farklı tahminci algoritma ile 1000 benzetim kullanılarak hesaplanan PS'lere dayalı olarak yapılan eşleştirme analizi sonucunda, senaryo A, B ve C için n=500 örneklem büyüklüğü ile elde edilen ATT, standart hataları ve yanlılık değerleri, SOF, OF, VO ve KS test istatistiği değerleri Tablo 2'de sunuldu.

Tablo 2. Denge sonuçları (n=500; senaryo A, B ve C)

Algoritmalar	Senaryo A				Senaryo B				Senaryo C			
	ATT	AI SH	SH	Y(%)	ATT	AI SH	SH	Y(%)	ATT	AI SH	SH	Y(%)
LR	-0.402	0.108	0.061	0.494	-0.436	0.109	0.065	9.087	-0.466	0.101	0.061	16.523
KSVM	-0.432	0.137	0.060	8.042	-0.446	0.137	0.062	11.496	-0.473	0.153	0.057	18.185
BAYESGLM	-0.397	0.108	0.061	-0.689	-0.446	0.110	0.065	11.536	-0.467	0.101	0.060	16.749
GLMNET	-0.398	0.107	0.061	-0.575	-0.444	0.110	0.065	10.957	-0.466	0.101	0.060	16.539
NNET	-0.390	0.103	0.060	-2.463	-0.413	0.104	0.063	3.258	-0.439	0.103	0.059	9.847
DBARTS	-0.420	0.125	0.061	4.983	-0.439	0.132	0.063	9.740	-0.459	0.151	0.058	14.868
QDA	-0.388	0.123	0.061	-2.876	-0.397	0.126	0.063	-0.859	-0.402	0.152	0.057	0.459
GAM	-0.400	0.110	0.061	-0.078	-0.428	0.114	0.064	6.876	-0.481	0.129	0.059	20.126
EARTH	-0.391	0.113	0.061	-2.272	-0.415	0.118	0.063	3.783	-0.441	0.140	0.058	10.199
STEPAIC	-0.393	0.106	0.061	-1.755	-0.423	0.110	0.065	5.692	-0.442	0.099	0.060	10.552
IPREDBAGG	-0.566	0.213	0.059	41.421	-0.565	0.214	0.060	41.275	-0.509	0.229	0.056	27.198
GBM	-0.489	0.132	0.060	22.150	-0.509	0.141	0.062	27.174	-0.507	0.178	0.057	26.759
CFOREST	-0.537	0.180	0.059	34.336	-0.550	0.175	0.061	37.607	-0.550	0.200	0.056	37.485
	Senaryo A				**Senaryo B**				**Senaryo C**			
	SOF	**OF**	**VO**	**KS**	**SOF**	**OF**	**VO**	**KS**	**SOF**	**OF**	**VO**	**KS**
LR	0.005	0.005	1.034	0.111	0.026	0.018	1.001	0.118	0.057	0.045	1.097	0.145
KSVM	0.039	0.025	1.124	0.149	0.045	0.029	1.120	0.151	0.101	0.069	1.242	0.186
BAYESGLM	0.004	0.003	1.034	0.111	0.034	0.025	1.001	0.120	0.054	0.043	1.095	0.144
GLMNET	0.010	0.007	1.031	0.110	0.031	0.021	1.004	0.119	0.055	0.044	1.097	0.145
NNET	0.023	0.016	1.043	0.104	0.028	0.020	1.020	0.114	0.043	0.029	1.082	0.134
DBARTS	0.045	0.026	1.077	0.120	0.055	0.033	0.021	0.128	0.105	0.067	1.135	0.158
QDA	0.009	0.008	1.047	0.124	0.011	0.008	1.054	0.130	0.026	0.019	1.130	0.166
GAM	0.004	0.003	1.037	0.109	0.026	0.017	1.024	0.114	0.083	0.058	1.094	0.146
EARTH	0.028	0.020	1.043	0.117	0.023	0.016	1.033	0.124	0.029	0.023	1.087	0.150
STEPAIC	0.025	0.017	1.034	0.113	0.022	0.015	0.999	0.121	0.046	0.037	1.092	0.146
IPREDBAGG	0.147	0.115	1.259	0.256	0.142	0.110	1.265	0.257	0.111	0.086	1.253	0.267
GBM	0.073	0.055	1.083	0.142	0.079	0.061	1.097	0.153	0.070	0.050	1.161	0.188
CFOREST	0.127	0.092	1.197	0.207	0.135	0.094	1.175	0.205	0.176	0.113	1.307	0.231

ATT: Tedavi grubunda ortalama tedavi etkisi, SH: Standart hata, AI SH: Abadie Imbens standart hata, SOF: Standardize edilmiş ortalama farklar, OF: Ortalama farklar, VO: Varyans oranı, KS: Kolmogorov Smirnov test istatistiği, Y: Yanlılık.

Denge sonuçlarına göre A senaryosunda yanlılık değeri %10'dan daha az olan ve diğer denge ölçütlerine göre başarılı olan KSVM, BAYESGLM, GLMNET, NNET, DBARTS, QDA, GAM, EARTH, STEPAIC algoritmaları kombine tahminci için aday algoritma olarak belirlendi. B senaryosunda LR'nin yanlılık değeri olan %9,087'den daha az olan ve diğer denge ölçütlerine göre başarılı olan NNET, QDA, GAM, EARTH ve STEPAIC algoritmaları kombine tahminci için aday algoritma olarak belirlendi. C senaryosunda LR'nin yanlılık değeri olan %16,52'den daha az olan ve diğer denge ölçütlerine göre başarılı olan NNET, DBARTS, QDA, EARTH ve STEPAIC algoritmaları kombine tahminci için aday algoritma olarak belirlendi. A, B ve C senaryolarının tamamında başarılı olan NNET, QDA, EARTH ve STEPAIC algoritmaları n=500 örneklem büyüklüğü için Kombine tahminci içerisinde yer almasına karar verildi.

3.2.2. n=1000 Örneklem Büyüklüğü için Benzetim Sonuçları

Üçüncü benzetim çalışmasında senaryo A, B ve C için n=1000 örneklem büyüklüğü ile elde edilen ATT, standart hataları ve yanlılık değerleri, SOF, OF, VO ve KS test istatistiği değerleri Tablo 3'de sunuldu.

Tablo 3. Denge sonuçları (n=1000; senaryo A, B ve C)

Algoritmalar	Senaryo A				Senaryo B				Senaryo C			
	ATT	AI SH	SH	Y(%)	ATT	AI SH	SH	Y(%)	ATT	AI SH	SH	Y(%)
LR	-0.401	0.076	0.043	0.226	-0.437	0.077	0.046	9.268	-0.465	0.070	0.043	16.37
KSVM	-0.418	0.095	0.043	4.500	-0.427	0.095	0.044	6.701	-0.450	0.111	0.040	12.41
BAYESGLM	-0.398	0.075	0.043	-0.469	-0.441	0.077	0.046	10.248	-0.467	0.070	0.043	16.72
GLMNET	-0.3996	0.075	0.043	-0.101	-0.441	0.077	0.046	10.358	-0.467	0.069	0.043	16.74
NNET	-0.390	0.073	0.043	-2.612	-0.420	0.075	0.045	4.924	-0.460	0.075	0.042	14.89
DBARTS	-0.410	0.085	0.043	2.609	-0.430	0.089	0.045	7.573	-0.439	0.111	0.041	9.715
QDA	-0.399	0.081	0.043	-0.291	-0.407	0.084	0.044	1.640	-0.395	0.106	0.041	-1.344
GAM	-0.399	0.076	0.043	-0.336	-0.434	0.079	0.045	8.578	-0.480	0.091	0.042	19.98
EARTH	-0.390	0.079	0.043	-2.483	-0.414	0.081	0.045	3.613	-0.435	0.100	0.041	8.758
STEPAIC	-0.387	0.075	0.043	-3.143	-0.412	0.077	0.046	2.964	-0.446	0.069	0.043	11.622
IPREDBAGG	-0.499	0.101	0.042	24.72	-0.527	0.107	0.043	31.630	-0.475	0.119	0.041	18.77
GBM	-0.466	0.088	0.043	16.46	-0.476	0.093	0.044	19.110	-0.470	0.119	0.041	17.40
CFOREST	-0.545	0.156	0.042	36.17	-0.543	0.151	0.043	35.792	-0.516	0.187	0.039	28.99
	Senaryo A				**Senaryo B**				**Senaryo C**			
	SOF	**OF**	**VO**	**KS**	**SOF**	**OF**	**VO**	**KS**	**SOF**	**OF**	**VO**	**KS**
LR	0.003	0.003	1.017	0.073	0.029	0.021	0.985	0.081	0.055	0.044	1.083	0.115
KSVM	0.024	0.015	1.077	0.097	0.030	0.019	1.071	0.099	0.079	0.053	1.165	0.129
BAYESGLM	0.003	0.002	1.018	0.072	0.032	0.023	0.984	0.082	0.057	0.045	1.082	0.116
GLMNET	0.004	0.003	1.019	0.072	0.028	0.020	0.984	0.081	0.056	0.045	1.084	0.117
NNET	0.015	0.011	1.025	0.077	0.021	0.014	1.001	0.084	0.051	0.037	1.049	0.114
DBARTS	0.032	0.019	1.044	0.073	0.040	0.024	1.038	0.078	0.073	0.046	1.080	0.103
QDA	0.004	0.003	1.020	0.075	0.010	0.007	1.023	0.079	0.027	0.019	1.091	0.106
GAM	0.002	0.001	1.017	0.069	0.027	0.018	1.006	0.074	0.084	0.058	1.069	0.104
EARTH	0.014	0.010	1.021	0.074	0.014	0.010	1.016	0.077	0.027	0.022	1.054	0.101
STEPAIC	0.011	0.008	1.015	0.075	0.018	0.012	0.983	0.082	0.045	0.037	1.081	0.117
IPREDBAGG	0.110	0.078	1.079	0.127	0.110	0.079	1.086	0.132	0.084	0.062	1.066	0.131
GBM	0.051	0.040	1.046	0.090	0.057	0.043	1.056	0.096	0.054	0.037	1.090	0.115
CFOREST	0.154	0.111	1.177	0.178	0.153	0.105	1.165	0.171	0.171	0.105	1.296	0.192

Denge sonuçlarına göre A senaryosunda yanlılık değeri %10'dan daha az olan ve diğer denge ölçütlerine göre başarılı olan KSVM, BAYESGLM, GLMNET, NNET, DBARTS, QDA, GAM, EARTH, STEPAIC algoritmaları kombine tahminci için aday algoritma olarak belirlendi. B senaryosunda

LR'nin yanlılık değeri olan %9,268'den daha az olan ve diğer denge ölçütlerine göre başarılı olan KSVM, NNET, DBARTS, QDA, GAM, EARTH ve STEPAIC algoritmaları kombine tahminci için aday algoritma olarak belirlendi. C senaryosunda LR'nin yanlılık değeri olan %16,37'den daha az olan ve diğer denge ölçütlerine göre başarılı olan KSVM, NNET, DBARTS, QDA, EARTH ve STEPAIC algoritmaları kombine tahminci için aday algoritma olarak belirlendi. A, B ve C senaryolarının tamamında başarılı olan KSVM, NNET, DBARTS, QDA, EARTH ve STEPAIC algoritmaları n=1000 örneklem büyüklüğü için Kombine tahminci içerişinde yer almasına karar verildi.

3.3. Üçüncü Benzetim Sonuçları; LR ve kombine modelin karşılaştırılması

PS tahmini için geliştirilecek kombine tahminci model içerisine n=500 ve n=1000 örneklem büyüklükleri için gerçekleştirilen ikinci benzetim sonuçlarında başarılı bulunan algoritmalar dahil edilerek (n=500: NNET, QDA, EARTH ve STEPAIC; n=1000: KSVM, NNET, DBARTS, QDA, EARTH ve STEPAIC) oluşturulan kombine tahminci ile A, B ve C senaryolarında 1000 benzetim ile elde edilen PS'ler kullanılarak eşleştirme analizi yapılmış ve elde edilen denge sonuçları Tablo 4'te verilmiştir.

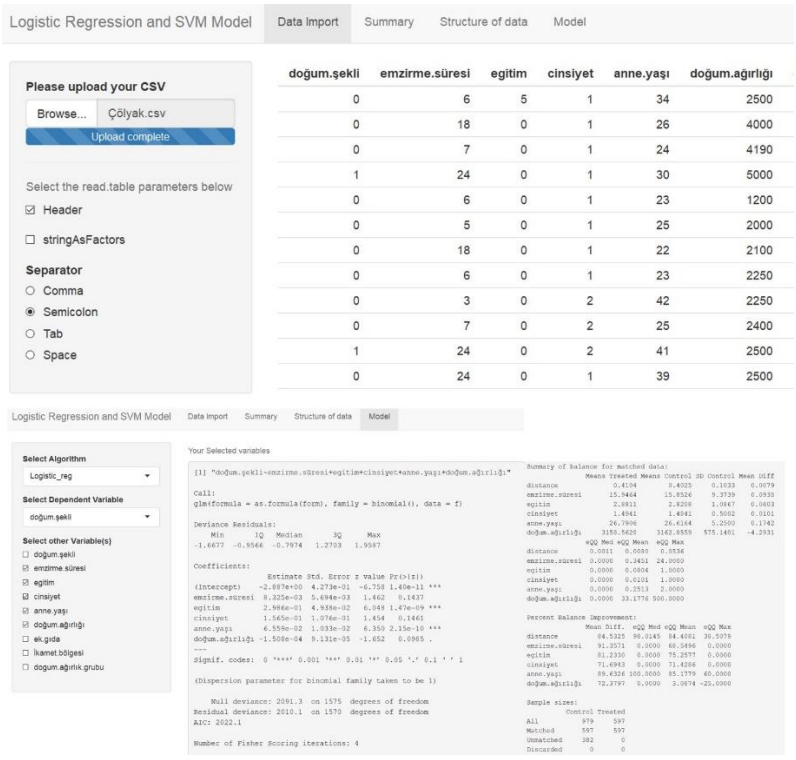
Tablo 4. LR ve kombine tahminci için özet denge sonuçları (n=500 ve 1000; senaryo A, B ve C)

	n=500						n=1000					
	Senaryo A		Senaryo B		Senaryo C		Senaryo A		Senaryo B		Senaryo C	
	LR	KTM	LR	KTM	LR	KTM	LR	KTM	LR	KTM	LR	KTM
ATT	-0.401	-0.411	-0.440	-0.434	-0.467	-0.446	-0.398	-0.402	-0.441	-0.430	-0.471	-0.431
AI SH	0.108	0.112	0.110	0.119	0.101	0.150	0.075	0.078	0.077	0.084	0.070	0.111
SHE	0.061	0.061	0.065	0.064	0.060	0.058	0.043	0.043	0.046	0.045	0.043	0.041
Yanlılık	0.276	2.814	9.914	8.552	16.856	11.478	-0.496	0.527	10.160	7.405	17.762	7.672
SOF (MO)	0.005	0.009	0.029	0.029	0.055	0.063	0.003	0.009	0.030	0.029	0.058	0.061
OF (MO)	0.004	0.006	0.021	0.020	0.044	0.041	0.003	0.005	0.021	0.019	0.046	0.039
VO (MO)	1.033	1.044	1.000	1.029	1.100	1.153	1.016	1.026	0.985	1.021	1.082	1.098
KS (MO)	0.111	0.115	0.119	0.125	0.145	0.164	0.072	0.073	0.080	0.081	0.116	0.106

LR: Lojistik Regresyon, KTM: Kombine tahmin modeli, ATT: Tedavi grubunda ortalama tedavi etkisi, SH: Standart hata, AI SH: Abadie Imbens standart hata, SOF: Standardize edilmiş ortalama farklar, OF: Ortalama farklar, VO: Varyans oranı, KS: Kolmogorov Smirnov test istatistiği, MO: Mutlak ortalama

3.4. R Shiny ile İnteraktif Bir Web Uygulaması

Çalışmamızda elde edilen sonuçlar dikkate alınarak PS tahmini ve eşleştirme analizi için R Shiny paketi ile interaktif bir web uygulaması geliştirilmiştir. Uygulama ile LR, birçok farklı ML algoritması ve kombine tahminci ile PS hesaplanarak eşleştirme analizi yapılabilmektedir. Ayrıca uygulama ile eşleştirme analizi sonrasında denge istatistikleri elde edilebilmektedir. Uygulamanın geliştirilmesine devam edilmekte olup yeni algoritmalar ve denge değerlendirmesi için grafiksel yaklaşımlar eklenerek yazılımın daha da güçlendirilmesi amaçlanmaktadır. Modülün örnek ekran görüntüsü Şekil 2'de gösterilmiştir.



Şekil 2. R Shiny interaktif web uygulaması

4. TARTIŞMA

McCaffrey ve ark. (2004), PS tahmininde Genelleştirilmiş Boosted Modelleri (GBM) kullanarak başarılı sonuçlar elde etmişlerdir [24]. Daha sonra Setoguchi ve ark. (2008), Lee ve ark. (2010) özellikle modelde sadece ana etkiler olmadığında ve doğrusal olmayan ilişkiler mevcut olduğunda ML Algoritmaları ile PS tahmininde başarılı sonuçlar elde etmişlerdir. Setoguchi ve ark. (2008) yaptıkları benzetim çalışmasında B (hafif doğrusal olmayan model) ve G (orta düzeyde doğrusal ve toplamsal olmayan model) senaryolarında LR'nin NNET'e göre daha yanlı tahminler ürettiği sonuçlarını raporlamışlardır. LR'nin sadece temel etkiler olan senaryoda yansız tahminler ile güçlü bir performansa sahip olduğunu ifade etmişlerdir [19].

Lee ve ark. (2010) yaptıkları benzetim çalışmasında farklı örneklem büyüklüklüğü ile belirlenen senaryoların tamamında bagged CART, random forests ve boosted CART'ın yüksek düzeyde performans sağladığını ve random forests ile boosted CART ML yöntemlerinin PS tahmininde kullanımlarının yaygınlaştırılarak daha ileriye götürülmesi gerektiğini ifade etmişlerdir. Ayrıca ML algoritmalarının büyük verilerin yanında n=500 örneklem büyüklüğünde de PS tahmininde kullanılabileceği sonucuna ulaşmışlardır [20]. Pirracchio ve ark. (2014) etkileşim ve doğrusal olmayan karmaşık ilişkilerin olduğu modelerde 15 farklı ML algoritmasından oluşan Super Learner tahmincisinin LR'ye göre daha yansız tahminler ürettiği sonucuna ulaşmışlardır [23].

Ancak LR dışında ML veya Bayesci genelleştirilmiş doğrusal modeller, Kuadratik Diskriminant Analizi, Adımsal Akaike Bilgi Kriteri gibi farklı tahmincilerin kullanıldığı çalışmalar yeterli düzeyde değildir. Özellikle LR ile ML yöntemlerinin birlikte kullanılarak daha başarılı eşleştirme sonuçlarına ulaşılabileceği Pirracchio ve ark. (2014) tarafından gösterilmiştir. Ancak algoritmaların birlikte kullanılması ile ilgili literatürde Pirracchio ve ark.'nın (2014) çalışması dışında başka çalışmaya rastlanmamıştır. Bu çalışmada da hangi ML algoritmalarının PS

hesabında güçlü olduğuna dair ön bilgi olmadığı için kombine modeller içerisindeki çok sayıdaki algoritmalar rastgele seçilmiş olup bulgular tek bir kombine tahminci ile elde edilmiştir.

Çalışmamızda benzetimlerde elde edilen eşleştirme sonuçları incelendiğinde her iki örneklem büyüklüğünde de model zorluluk derecesi arttıkça LR'nin yanlılık değerinin arttığı görülmektedir. LR, 500 ve 1000 örneklem büyüklüğünde A, B ve C senaryolarında neredeyse benzer yanlılık değeri ile kestirimlerde bulunmuştur. Ancak kombine tahminci literatürdeki araştırmalara benzer olarak n=1000 büyüklüğünde A, B ve C senaryolarının tamamında n=500 örneklem büyüklüğüne göre daha yansız kestirimlerde bulunmuştur. LR ile kombine algoritma sonuçları karşılaştırıldığında A senaryosunda benzer kestirimlerde bulunduğu görülmektedir. Buda kombine tahmincinin LR ile rekabet edebildiğini göstermiştir. B ve C senaryolarında ise kombine tahminci daha başarılı kestirimlerde bulunmuştur. Özellikle n=1000 örneklem büyüklüğünde en zor senaryo olan C senaryosunda kombine tahmincinin LR'den %10,09 daha az yanlı kestirimde bulunması ve LR'nin %17,76 gibi yüksek bir yanlılık kestiriminde bulunması karmaşık modellerde ML veya diğer algoritmaların LR'ye alternatif veya daha başarılı algoritma olarak PS tahmininde kullanılmasının yanlılığı azaltacağını göstermektedir. Ayrıca belirlenen algoritmaların özellikle büyük örneklemlerde daha başarılı kestirimlerde bulunduğu sonucuna ulaşılmıştır.

Denge göstergeleri incelendiğinde her iki örneklem büyüklüğünde SOF, OF, VO ve KS değerleri LR ve kombine algoritma ile neredeyse benzer olarak bulunmuştur. Ancak özellikle C senaryosunda Yapay Sinir Ağları, Kuadratik Diskriminant Analizi, Çok Değişkenli Uyarlanabilir Regresyon Eğrileri ve Adımsal Akaike Bilgi Kriteri yöntemleri LR'ye göre benzer VO ve KS değerleri elde etmiş olsa da, SOF değerlerinde daha başarılı eşleştirme analizi gerçekleştirmiştir.

Araştırmamızın birinci kısıtlılığı benzetim sonuçlarının bizim ürettiğimiz ilişkilere dayalı olan benzetim verisine dayanmasıdır. Ancak Setoguchi (2008) üretilen bu veri yapısının hem klinik farmakoloji hem de epidemiyolojiyi kapsayan ve çok sayıda insanda ilaçların kullanımının ve etkilerinin incelendiği Farmakokepidemiyolojik çalışmalardaki yaygın olan veri yapısını modellediğini ifade etmiştir. Üretilen bu veride birçok sürekli ve kategorik karıştırıcı değişken bulunması ve gerçek tedavi etkisi araştırmalarından yola çıkarak üretilmesi bu kısıtlılık için bir cevap oluşturmaktadır. İkinci bir kısıtlılık ise; Setoguchi ve ark. (2008), Lee ve ark. (2010), Pirracchio ve ark. (2014)'nın çalışmalarında olduğu gibi bu çalışmada da, her bir algoritmanın yalnızca varsayılan özellikleri kullanılmıştır. Çünkü PS tahmininde ML algoritmaları için birçok araştırmacının kullanacağı seçeneğin varsayılan ayarlar olacağı düşünülmektedir. Herhangi bir algoritmanın, ML uzmanı bir kullanıcı tarafından uygulandığında daha iyi performans göstermesi olasıdır. Ayrıca sonuç değişkeni olarak kabul edilen tedavi ve kontrol grubunu gösteren tedavi değişkeni ile ortak değişkenler arasındaki etkileşimler analiz edilerek etkileşimlere sahip modellerin olduğu LR kullanılarak elde edilen eşleştirme dengesinin sonuçları yaygın olarak kullanılan basit ana etkiler modelinden daha iyi performans gösterebilecektir. Bazı araştırmacılar etkileşimleri kontrol etmeyi önermiş olsalar da uygulamada yaygın olarak dikkate alınmamaktadır [20].

Bu çalışmanın diğer çalışmalardan bazı güçlü yanları mevcuttur. Bunlardan ilki kombine modeller içerisinde veya literatürde LR ile karşılaştırılan ML çalışmalarında ilk kez LR ile en uyumlu kestirimlerde bulunan algoritmalar bu çalışma ile belirlenmiştir. İkincisi ise PS hesabında bazı algoritmalar bireysel olarak ilk kez bu çalışmada kullanılmıştır. Özellikle BAYESGLM, CFOREST, DBARTS, EARTH, GAM, IPREDBAGG, GLMNET, KERNELKNN, KSVM, QDA, STEPAIC, SVM,

XGBOOST algoritmaları ile PS tahmini yapılan bir çalışmaya literatürde rastlanmamıştır. Literatürde özellikle random forest, GBM gibi ML algoritmalarına odaklanılmıştır.

5. SONUÇ

Gözlemsel çalışmalarda elde edilen ATT ve ATE değerleri doğrudan eşleştirme analizinde kullanılan PS'lere ve dolayısıyla eşleştirme başarısına bağlıdır. Eşleştirme ise PS'ye göre yapılmaktadır. PS'nin vaka ve kontrol grubundaki deneklere verilen ortak değişkenlerin benzerliğine bağlı bir özet istatistik olduğu düşünüldüğünde her çalışmada LR ile PS hesaplanması durumunda tedavi veya kontrol grubuna atanma olasılığında doğrusal olmayan ilişkiler veya etkileşimlerin bulunması halinde yanlı sonuçlar elde edilmiş olacaktır. Elde ettiğimiz sonuçlara göre gözlemsel çalışmalarda LR'ye ek olarak ML algoritmaları veya diğer algoritmaların da PS hesaplanmasında kullanılarak denge değerlendirmesinin yapılması ve denge başarısına göre yansız ATT veya ATE nin hesaplanması önerilebilir. Ayrıca parametrik yöntemler (LR) tedavi gruplarına atanma olasılığı ile ortak değişkenler arasındaki ilişkinin fonksiyonel yapısı ile ilgili varsayımlar gerektirir. ML yöntemlerinde doğrusallık veya etkileşimler incelenmeden doğrudan ortak değişkenler seçilerek skorlar hesaplanabilmektedir. Randomize klinik çalışmaların yüksek maliyeti ve etik kurul izninde yaşanan zorluklar birçok araştırmacının daha düşük maliyetli olan gözlemsel çalışmaları tercih etmesine neden olmaktadır. Bu nedenle, PS özellikle tıp alanındaki istatistiksel analizlerde yaygın olarak kullanılmaktadır. PS hesaplamak ve eşleştirme analizinde gruplar arası dengeyi sağlamak için son yıllarda kullanımı artmaya başlayan ML yaklaşımlarının gözlemsel çalışmaların güvenirliğini arttıracağı düşünülmektedir.

TEŞEKKÜR

Değerli katkılarından dolayı Prof. Dr. Atilla Halil ELHAN, Prof. Dr. Erdem KARABULUT, Prof. Dr. Yasemin YAVUZ, Prof. Dr. Mehtap AKÇİL OK, Doç. Dr. Derya GÖKMEN ve Doç. Dr. Beyza DOĞANAY ERDOĞAN'a teşekkür ederiz.

Bu çalışma; Ankara Üniversitesi, Sağlık Bilimleri Enstitüsü tarafından kabul edilen birinci yazara ait Doktora Tez çalışmasından üretilmiştir.

KAYNAKLAR

1. Rosenberger WF, Lachin JM. Randomization in Clinical Trials: Theory and Practice. John Wiley, New York; 2002.
2. Rubin DB. Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology.* 1974;66:688-701.
3. Hill HA, Kleinbaum DG. Bias in observational studies. *Encyclopedia of Biostatistics,* John Wiley and Sons, Chichester; 2005.
4. Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. *Biometrika.* 1983;70:41-55.
5. Rosenbaum PR, Rubin DB. Reducing bias in observational studies using subclassification on the propensity score. *Journal of the American Statistical Association.* 1984;79:516-524.
6. D'agostino RB JR. Tutorial in biostatistics: propensıty score methods for bias reduction in the comparıson of a treatment to a non-randomized control group. *Statistics in Medicine.* 1998;17:2265-2281.
7. Kaspar EÇ, Bekiroğlu N, Genceli M (2010). Gözleme dayalı çalışmalarda eğilim skoru (propensity score) ve tıp bilimleri'nde bir uygulama. Türkiye Klinikleri Biyoistatistik Dergisi, 2: 1-10.

8. Hernan MA. A definition of causal effect for epidemiological research. *J Epidemiol Community Health.* 2004;58:265-271.
9. Cochran WG, Rubin DB. Controlling bias in observational studies: a review. *Sankhya, Ser. A.* 1973;35:417-46.
10. Austin PC. Balance diagnostics for comparing the distribution of baseline covariates between treatment groups in propensity-score matched samples. *Statistics in Medicine.* 2009;28:3083-3107.
11. Austin PC. An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate Behav. Res.* 2011;46:399-424.
12. Demir O, Dolgun A, Etikan İ, Kuyucu YE, Saraçbaşi O. Simulation study on performance of balance metrics in propensity score weighting method. *J Contemp. Med*. 2017;7:265-277.
13. Caliendo M, Kopeinig S. Some practical guidance for the implementation of propensity score matching. *Journal of Economic Surveys.* 2008;22:31-72.
14. Heckman J. The scientific model of causality. *Sociological Methodology*, 35:1-97.
15. Imbens GW. Nonparametric estimation of average treatment effects under exogeneity: a review, *The Review of Economics and Statistics.* 2004;86:4-29.
16. Polley EC, Van Der Laan MJ. Super learner in prediction. U.C. Berkeley Division of Biostatistics Workihong Paper Series; 2010. 226 p.
17. Sekhon, JS. Multivariate and propensity score matching software with automated balance optimization: the matching package for R. *Journal of Statistical Software.* 2011;42:1-52.
18. Gamer M, Lemon J, Singh P. Various coefficients of interrater reliability and agreement. Package 'irr', version 0.84; 2015.
19. Setoguchi S, Schneeweiss S, Brookhart MA, Glynn RJ, Cook EF. Evaluating uses of data mining techniques in propensity score estimation: a simulation study. *Pharmacoepidemiology and drug safety.* 2008;17:546-55.
20. Lee BK, Lessler J, Stuart EA. Improving propensity score weighting using machine learning. *Statistics in Medicine.* 2010;29:337-346.
21. Lee BK, Lessler J, Stuart EA. Weight trimming and propensity score weighting. *PLoS ONE.* 2011;6:e18174.
22. Leacy FP, Stuart EA. On the joint use of propensity and prognostic scores in estimation of the average treatment effect on the treated: a simulation study. *Statistics in Medicine*. 2014;33:3488-3508.
23. Pirracchio R, Petersen M, Van Der Laan M. Improving propensity score estimators' robustness to model misspecification using super learner. *American Journal of Epidemiology.* 2014; 181:108-19.
24. Mccaffrey DF, Ridgeway G, Morral AR. Propensity score estimation with boosted regression for evaluating causal effects in observational studies. *Psychological Methods.* 2004;9:403-425.

T5

TIBBİ GÖRÜNTÜ İŞLEME İLE TANI KOYMADA VERİ MADENCİLİĞİ YÖNTEMLERİNİN SINIFLAMA PERFORMANSLARININ KARŞILAŞTIRILMASI

Hanife AVCI[1], Jale KARAKAYA[1]

[1] Hacettepe Universitesi Tıp Fakültesi Biyoistatistik Anabilim Dalı, ANKARA

ÖZET

Dünyada ve ülkemizde en çok görülen ölüm nedenlerinden birisi kanserdir. Kanserlerin pek çok türü bulunmaktadır. Kadınlarda en sık görülen kanser türü ise meme kanseridir. Kanserde erken tanı koymak, ölüm oranını azaltması açısından oldukça önemlidir. Meme kanserinin araştırılmasında, radyologlara kolaylık sağlamak ve birden fazla radyoloğun incelemesi gereken görüntülerde iş yükünü azaltmak amacıyla görüntü işleme teknikleri tercih edilebilir. Tıbbi görüntüleme teknikleri, kanserin saptanmasında ve sınıflandırılmasında yaygın olarak kullanılmaktadır. Her bir görüntünün incelenmesi radyoloğun büyük ölçüde zamanını alır. Tanı koyma aşamasında hekimlere yardımcı olmak için bilgisayar destekli tanı modelleri (Computer-aided diagnostic models, CADx) önerilmiştir [1]. Bu yazıda, meme kanseri sınıflandırması ve tanı sürecinde doktorlara yardımcı olmak amacıyla tıbbi görüntü işleme tekniklerinden yararlanılmıştır. Daha sonra klasik veri madenciliği yöntemlerinden Rastgele Orman (Random Forest), Yapay Sinir Ağları ve Destek Vektör Makineleri algoritmalarının sınıflama performansları karşılaştırılmıştır.

Anahtar Kelimeler: Tıbbi Görüntüleme, Görüntü İşleme, Veri Madenciliği, Yapay Sinir Ağları

1. GİRİŞ

Son yıllarda meme kanseri, kadınlar arasında en önemli ölüm nedenlerinden ve kadınlarda en sık teşhis edilen kanserlerden biridir. Meme kanseri, memenin hücre dokuları anormal hale geldiğinde kontrolsüz olarak çoğalmasıyla ortaya çıkar. Bu anormal hücreler, bir tümör haline gelen büyük dokuları oluşturur. Tümörler erken teşhis edilirse başarılı bir şekilde tedavi edilebilir. Rezonans manyetik görüntüleme (MRG), ultrasonografi ve mamografi gibi meme kanseri taramasında birden fazla tanı yöntemi kullanılmaktadır [2]. Mamografi, meme kanserinin erken teşhis edilmesinde tarama testi olarak yaygın bir şekilde kullanılmaktadır. Çünkü, mamografi ile elde edilen görüntüler yardımıyla, memede elle muayene ile saptanamayacak kadar küçük değişiklikler bile erkenden saptanabilir. Mikrokalsifikasyonlar tarama yöntemleri kullanılarak tespit edilebilen en erken meme kanseri belirtileridir. Genel olarak meme dokularında kitlelerin tespiti, sadece boyut ve şekildeki büyük farklılıklar nedeniyle değil, aynı zamanda kitlelerin mamografiyi kullanırken sıklıkla kötü görüntü kontrastı göstermesi nedeniyle, mikro kireçlenme tespiti ile karşılaştırıldığında daha zordur. Bu nedenle, bilgisayar destekli tanı modelleri (CADx) radyologlar için iyi huylu ve kötü huylu kitleleri ayırt etmede faydalı olabilirler [3]. Bu incelemeyi yapan radyologların işini kolaylaştırmak ve inceleme süresini kısaltmak amacıyla geliştirilen birçok görüntü işleme algoritması mevcuttur.

2. GÖRÜNTÜ İŞLEME

Görüntü işleme, bir görüntüyü bir formdan diğerine dönüştürme yöntemi olarak tanımlanabilir. Görüntü işleme ile dijital görüntüler üzerinde çeşitli işlemler uygulanarak yeni görüntüler elde edilir. Bu dijital görüntü verilerine, amaca uygun bilgisayar ve yazılımlar sayesinde matematiksel algoritmalar uygulanır. Özetle görüntü işleme, elde edilen görüntülerden amaçlanan hedefe uygun sayısal değerler elde etme işlemidir. Bilgisayar ve görüntü işleme programları sayesinde görüntü üzerinde analiz yaparken keskinleştirme, sıkıştırma, bulanıklaştırma, nesne sınırlarını belirginleştirme, gürültü giderme, maskeleme, görüntüyü bölütleme, kontrast artırma vb. işlemlerinin tümü pikseller üzerine uygulanır [4]. Sayısal görüntü işleme teknikleri sayısal görüntünün kalitesini artırmak amaçlı uygulanır. Görüntü işlemede kullanılan çözünürlük, parlaklık (kontrast), gürültü gibi bazı temel kavramlar mevcuttur. Görüntülerdeki piksel sayısının fazla olması, görüntünün çözünürlüğünü artırmaktadır. Bir anlamda çözünürlük bir kalite göstergesidir. Parlaklık, görüntüdeki en karanlık bölge (siyah) ile en parlak bölgenin (beyaz) arasındaki fark olarak tanımlanabilir. Gürültü, görüntü gönderilirken ya da alınırken görüntü kalitesini düşüren, görüntü üzerindeki bozukluklar, lekeler kısaca görüntü kirliliği olarak tanımlanabilir. Bazen elektronik kaynaklardan ya da kamera gibi cihazlardan dolayı kaynaklanan bu gürültülerin görüntüler üzerinden kaldırılması gerekir.

Görüntü işleme, çok çeşitli bilimsel, sanatsal ve ticari uygulamalar için kullanılır. Son yıllarda özellikle sağlık alanında görüntü işleme yönteminden sıklıkla yararlanılmaktadır. Görüntü işleme, tanı amaçlı radyoloji de dahil olmak üzere görüntüleri kullanan birçok disiplin için gereklidir [5]. Gelişen teknoloji ile birlikte sağlık hizmeti vermekte olan kurumlar, hizmet sunumlarındaki verimliliği arttırmak, maliyetleri düşürmek, hasta bakımını geliştirmek ve hastalara hak ettikleri hizmetleri zamanında verebilmek için bilgisayar tabanlı bilgi sistemlerine yönelmektedirler. Görüntü işleme 2 aşamadan oluşmaktadır. Görüntüler üzerinde uygulanan birinci aşama görüntü ön işleme, ikinci aşama ise özellik çıkarımıdır. Görüntü ön işleme aşamasında, görüntüler üzerinde görüntüyü keskinleştirme, kontrastı artırma, normalleştirme, histogram eşitleme gibi adımlar uygulanır. Bu işlemler, görüntülerin parlaklığını ve kontrastını arttırmak, görüntü üzerindeki gürültülerin giderilmesi, görüntüde yer alan gereksiz ayrıntıların azaltılması, görüntüde mevcut olan kenarları ve ince ayrıntıları vurgulamak amacıyla gerçekleştirilir.

2.1 Görüntü Ön İşleme

Bilgisayara yüklenen görüntüler üzerinde işlem yapabilmek için görüntü keskinleştirme, kontrastı artırmak, normalleştirme ve histogram eşitleme gibi aşamaların görüntü üzerine uygulanması gerekir.

2.1.1 Görüntü Keskinleştirme

Görüntüde mecvut olan gürültünün etkisini azaltmak için görüntü netleştirme yapılır ve keskinleştirme teknikleri uygulanır. Bu adım, kontrastı artırmak ve görüntüde mevcut olan kenarları ve ince ayrıntıları vurgulamak için gerçekleştirilir.

2.1.2 Kontrastı Artırmak

Bu işlem, çok çeşitli gri seviyeleri kapsayacak şekilde görüntü yoğunluğu değeri dağılımını değiştirir. Pikseller arasındaki kontrastı artırmak nesneleri daha iyi ayırt etmeyi sağlar.

2.1.3 Normalleştirme

Normalleştirme, görüntünün piksel değer aralığını değiştiren bir işlemdir. Bu nedenle aralık, veri türü için maksimum aralığa veya görüntülerin yüzdelik gösterimi için 0-1' e eşittir. Maksimum aralık 8 bit görüntüler için 0-255 ve 16 bit görüntüler için 0-65535' tir.

2.1.4 Histogram Eşitleme

Histogram, bir görüntüdeki renk değerlerinin sayılarını gösteren bir grafik olarak ifade edilebilir. X ekseni, olası gri değerleri temsil eder ve Y ekseni her gri değer için bulunan piksel sayısını gösterir. Histogram eşitleme ise bir görüntüdeki renk değerlerinin belli bir yerde kümelenmiş olmasından kaynaklanan renk dağılımı bozukluğunu gidermek için kullanılan bir yöntemdir. Götüntü işlemede histogramlar kullanılarak görüntü daha belirgin bir hale getirilebilmektedir. Bir görüntünün parlaklığını ve kontrastını arttırmak için histogram eşitleme yapılır.

2.2 Özellik Çıkarımı

Bir görüntüyü bütün bir bilgi olarak değilde bu görüntüyü oluşturan nesnelerin bazılarının çıkarılıp işleme hazır hale getirilmesidir. Başka bir deyişle, özellik çıkarma işlemi bir boyut azaltma işlemidir. Buna göre karmaşık olan bir verinin boyutları azaltılarak daha basit bir problem haline indirgenir. Özellik çıkarma, önemli verileri tanımlamaya yardımcı olan ve görüntüdeki gereksiz ayrıntıları ortadan kaldırmak için faydalı özellikleri tespit etmeyi amaçlar.

Özellik çıkarma işlemi için genellikle istatistiksel özellik çıkarma metodu kullanılmaktadır. Bu yöntem, belirli bir bölgedeki nesnenin özelliklerinin tespit edilebilmesi için kullanılan en temel özellik çıkarma yöntemlerinden birisidir. Ön işleme aşamasından sonra ise, görüntülere ait özellik çıkarımı işlemi gerçekleştirilir.

2.2.1 Gri Seviye Eş Oluşum Matrisi

Görüntüdeki lezyonu tanımlamak için en basit yaklaşımlardan biri, bir görüntü veya bölgenin yoğunluk histogramının istatistiksel özelliklerini kullanmaktır. Ancak hesaplamada sadece histogramları kullanmak, görüntüdeki piksellerin birbirlerine göre göreceli konumu ile ilgili bilgi vermemektedir. Bu nedenle eş oluşum matrisi gibi istatistiksel bir yaklaşım kullanmak, görüntüdeki komşu piksellerin göreceli konumu hakkında değerli bilgiler sağlamamıza yardımcı olacaktır.

Gri seviye eş-oluşum matrisi ile görüntülerden, otokorelasyon (autocorrelation), kontrast (contrast), karşıtlık (dissimilarity), enerji (energy), entropi (entrophy), maksimum olasılık (maximum probability), toplam entropi (sum entrophy), fark varyansı (difference variance), fark entropisi (difference entrophy), bilgi korelasyon ölçümü 1 (information measure of correlation 1), bilgi korelasyon ölçümü 2 (information measure of correlation 2), normalleştirilmiş ters fark (inverse difference normalized) ve normalleştirilmiş ters fark momenti (inverse difference moment normalized) gibi özellikler elde edilebilir. Elden edilen özelliklerin matematiksel formüllerine [13], [14] ve [15] kaynaklarından ulaşılabilir.

3. VERİ MADENCİLİĞİ

Bilimsel teknolojilerdeki ilerlemeler sağlık alanına şimdiye dek görülmemiş büyüklükte karmaşık veriler getirmiştir. Bu karmaşık veri yığınından, ihtiyacı karşılayacak değerli verilerin elde edilmesi işlemine Veri Madenciliği denilmektedir. Veri madenciliği, büyük veri kümelerinden, veriler arasındaki benzer eğilimleri keşfetmeyi amaçlayan, bir anlamda veritabanlarından bilginin

keşfedilmesini sağlayan ve gelecekle ilgili doğru tahminlerde bulunmamıza yardımcı olan bir yöntemler bütünüdür. Bilişim biliminden gelen veri madenciliği ve makine öğrenmesi yöntemleri, diğer alanlardaki benzer problemlerde iyi performans gösteren güvenilir yöntemler sunar. Geniş anlamda klasik veri madenciliği için birçok algoritma mevcuttur. Bu algoritmaların bazıları kestirim ve bazıları da sınıflandırma amaçlı kullanılmaktadır.

Veri madenciliği algoritmaları özellikle değişken sayısının fazla olduğu karmaşık veri kümeleri için faydalıdır. Makine öğrenmesi algoritmaları genel olarak Danışmanlı Öğrenme (Supervised Learning) ve Danışmansız Öğrenme (Unsupervised Learning) olmak üzere ikiye ayrılır. Danışmanlı öğrenme algoritmaları, regresyon ve sınıflandırma problemleri için kullanılmaktadır [6]. Danışmanlı öğrenmede sınıflandırma için kullanılan algoritmalar; Karar ağaçları, Naive Bayes, K-En Yakın Komşu, Yapay Sinir Ağları, Rastgele Orman (random forest) ve Destek Vektör Makinesi algoritmalarıdır. Regresyon için kullanılan algoritmalar, Lineer Regresyon ve Lojistik Regresyon algoritmalarıdır. Danışmansız öğrenme algoritmaları ise, kümeleme, birliktelik ve boyut indirgeme gibi amaçlarla kullanılmaktadır. Danışmansız öğrenmede kümeleme ve birliktelik kuralları için kullanılan algoritmalar; K-ortalamalar, K-medoid ve Apriori (birliktelik kuralları) algoritmalarıdır. Makine öğrenmesinde kullanılan algoritmalar genel olarak eğitim aşaması (training phase) ve test aşaması (test phase) şeklinde iki aşamadan oluşmaktadır. Eğitim verisi, algoritmanın öğrenmesi için sunulan bir alt küme dizisidir. Algoritma bu veriyi inceleyerek tahminlerde bulunarak bir model kurar. Test verisi ise, algoritmanın eğitim verisinden yararlanarak kurduğu modelin ne kadar gerçeğe uygun olduğunu test etmek amacıyla kullanılan veri setidir. Veriler, genellikle eğitim ve test setlerine %80 ve %20 oranında ayrılmaktadır. Veri setinin %80'lik kısmı eğitilip %20'lik kısmı üzerinde test edilir. Böylece, eğitim seti içerdiği örneklerle test seti dediğimiz veri setinin kalan kısmındaki sonuçları doğru tahmin etmeye çalışır.

Bu bilgilerden yola çıkarak; bu çalışmada amaç, hastalara ilişkin görüntüler üzerinde uygulanacak görüntü işleme teknikleri ile klasik veri madenciliği algoritmalarının sınıflama performanslarını tıp alanında bir uygulama yaparak karşılaştırmak ve böylece doğru ve hızlı tanı konulmasında yardımcı olmaktır. Böylece hem görüntü işleme sonrasında elde edilen özelliklerin hem de bu özelliklerin birlikte incelendiği veri madenciliği yöntemlerinin sınıflandırmadaki performansı değerlendirilmiş olacaktır.

4. GEREÇ VE YÖNTEM

Çalışmada açık erişimli bir veri tabanından (The mini-MIAS database of mammograms) elde edilen 40 tane mamografi görüntüsü kullanılmıştır. Bu 40 görüntünün 20 tanesi normal, 20 tanesi de kötü huylu kitleye sahip bireylere aittir. Görüntülerin işlenmesi için MATLAB-Image Processing Toolbox ve Veri Madenciliği sınıflandırma için Orange yazılımları kullanılmıştır.

- Birinci aşamada, görüntülerin iyileştirilmesi için görüntü ön işleme adımları (keskinleştirme, kontrastı artırma, normalleştirme ve histogram eşitleme) uygulanmıştır.
- İkinci aşamada, görüntü üzerinde segmentasyon (bölütleme) işlemi uygulanarak kanserli kısmın görüntüden çıkarılması sağlanmıştır.
- Üçüncü aşamada da gri seviye eş-oluşum matrisi (GLCM- Gray Level Co-occurence Matrix) ile mamogram görüntülerinin özellik çıkarımı yapılmıştır. Ayrıca bu aşamada her özellik için grup karşılaştırılmasında t testi ve sınıflama performansları için ROC Analizi yapılmıştır.
- Son aşama olan sınıflandırma kısmında Klasik Veri Madenciliği yöntemlerinden Destek Vektör Makinesi, Rastgele Orman (random forest) ve Yapay Sinir Ağları algoritmalarının görüntülerden

elde edilen özellikleri kullanarak gruplar normal ve kötü huylu kanser olarak sınıflandırılmış ve performansları karşılaştırılmıştır.

5. BULGULAR

Mamogram görüntüleri The mini- MIAS database of mammograms veri tabanından elde edilmiştir. Kesin yöntemle 20 Normal 20 Malign olarak belirlenen gruplar gri seviye eş oluşum matrisinden elde edilen özellikler açısından karşılaştırılmıştır.

Tablo 1. Normal ve Malign Grup Karşılaştırması

Özellikler	Normal (n=20)		Malign (n=20)		p değeri
	Art. ortalama	St.Sapma	Art. ortalama	St. Sapma	
otokorelasyon	15020.28	1922.21	13920.18	1640.87	0.059
kontrast	11538.78	2870.91	9439.69	3426.14	0.042
karşıtlık	86.11	12.47	76.28	14.61	0.028
enerji	2.48E-05	6.54E-06	3.23E-05	6.72E-06	0.001
entropi	10.90	0.12	10.69	0.12	<0,001
maksimum olasılık	9.38E-05	2.48E-05	5.93E-05	2.54E-05	<0,001
toplam entropi	6.00	0.07	5.88	0.11	<0,001
fark varyansı	11538.78	2870.91	9439.69	3426.14	0.042
fark entropisi	5.34	0.11	5.22	0.12	0.002
bilgi korelasyon ölçümü 1	-0.02	0.01	-0.04	0.02	0.001
bilgi korelasyon ölçümü 2	0.39	0.13	0.54	0.13	<0,001
normalleştirilmiş ters fark	0.77	0.03	0.79	0.03	0.024
normalleştirilmiş ters fark momenti	0.87	0.03	0.89	0.03	0.032

Normal ve kötü huylu (malign) kanser dokusuna sahip kişilerin mamogram görüntüleri arasında iki ortalama arasındaki farkın anlamlılık testi ile Tablo 1.'deki özellikler bakımından fark bulunmuştur. Hesaplanan özellikler arasından $p<0.05$ ve anlamlılığa yakın olan özellikler sınıflama başarısını incelemek için değerlendirmeye alınmıştır.

Matrisden elde edilen özelliklerin her birinin sınıflama performansı ROC Analizi ile incelenmiştir.

Tablo 2. Özelliklerin ROC Analizi Sonuçları

Özellikler	Eğri altında kalan Alan	Standart Hata	p değeri
otokorelasyon	0.643	0.089	0.123
kontrast	0.728	0.082	0.014
karşıtlık	0.723	0.082	0.016
enerji	0.813	0.068	0.001

entropi	0.885	0.053	0.000
maksimum olasılık	0.848	0.070	0.000
toplam entropi	0.815	0.071	0.001
fark varyansı	0.728	0.082	0.014
fark entropisi	0.780	0.072	0.002
bilgi korelasyon ölçümü 1	0.818	0.066	0.001
bilgi korelasyon ölçümü 2	0.813	0.067	0.001
normalleştirilmiş ters fark	0.718	0.082	0.019
normalleştirilmiş ters fark momenti	0.728	0.082	0.014

Bu özelliklerin sınıflandırmadaki performanslarına ait eğri altında kalan alanlar ve standart hata sonuçları Tablo 2'de gösterilmiştir. Sonuçlar incelendiğinde, eğri altında kalan alanlar açısından 0.643 ile 0.885 arasında değişen değerler görülmüştür. Otokorelasyon özelliği dışında tüm özelliklerin sınıflama performansları anlamlılık göstermiştir.

Diğer bir amacımız, bu değişkenlerin hepsini kullanarak daha yüksek bir sınıflandırma performansı elde etmektir. Bu nedenle, bu özelliklerin birlikte kullanıldığı veri madenciliği yöntemlerinin performansları karşılaştırılmıştır.

Tablo 3. Veri Madenciliği Yöntemlerinin Sınıflama Performanslarının Karşılaştırılması

	SINIFLAMA YÖNTEMLERİ		
Sınıflama Yöntemleri İçin Performans Ölçüleri	**Destek Vektör Makineleri (Support Vector Machine)**	**Rastgele Orman (Random Forest)**	**Yapay Sinir Ağları (Artifical Neural Network)**
Doğruluk Oranı	0.80	0.80	0.725
Duyarlılık	0.70	0.85	0.75
Seçicilik	0.90	0.75	0.70
F1 Ölçüsü	0.80	0.79	0.725
AUC	0.90	0.810	0.765

Tablo 3.'de üç farklı sınıflama yönteminin performansları verilmiştir. Genel olarak, kullanılan algoritmalar arasında Destek Vektör Makinesi algoritmasının doğruluk oranı, seçicilik, F1 ölçüsü ve eğri altında kalan alan (AUC) değeri Rastgele Orman ve Yapay Sinir Ağlarına göre daha yüksek çıkmıştır. En yüksek duyarlılık değeri ise Rastgele Orman algoritmasına aittir. Elden edilen sonuçlara göre sınıflandırma performansı en iyi algoritma genel olarak Destek Vektör Makinesi olarak belirlenmiş ve literatürle de benzer sonuçlar elde edilmiştir.

6. TARTIŞMA VE SONUÇ

Bu çalışmanın temel avantajı, bir mamogram görüntüsünden çıkarılan özellikler yardımı ile, kanserle ilişkili çeşitli konular arasındaki morfolojik farklılıkların vurgulanması ve elde edilen özelliklerin farklılıkları göstermesidir. Mamogram görüntüleri üzerinde görüntü işleme tekniklerinin kullanılması, radyoloğa kolaylık sağlaması açısından son derece önemlidir. Elde edilen sonuçlara göre, kanserin erken teşhisine yardımcı olabileceği düşünülebilir.

Ancak çalışmanın bazı kısıtlılıkları mevcuttur. Görüntü işleme aşamasında farklı ön işleme adımları uygulanabilir. Bu işlemler sonrasında farklı özelliklerin sınıflama performansları incelenebilir. Çalışmada açık erişimli veri tabanından elde edilen sadece 40 kişiye ait görüntü üzerinde görüntü işleme adımları gerçekleştirilmiştir. Kullanılan algoritmalar daha büyük veri setlerinde de uygulanabilir. Bunun yanı sıra, gerçek veri setlerinden elde edilen görüntüler ve klinik bulgular da çalışmaya dahil edilebilir.

7. KAYNAKLAR

1. T. Ayer, M. US Ayvacı, Ze Xiu Liu, O. Alagoz and E. Burnside. Computer-aided diagnostic models in breast cancer screening. Imaging Medicine 2010 June 1,313-323. doi: 10.2217/IIM.10.24

2. P. Gırı, K. Saravanakumar. Oriental Journal Of Computer Science & Technology. Breast Cancer Detection using Image Processing Techniques. doi: 10.13005/ojcst/10.02.19, 2017

3. M. M. Mehdy, P. Y. Ng, E. F. Shair, N. I. Md Saleh and C. Gomes. Artificial Neural Networks in İmage Processing for Early Detection of Breast Cancer. Hindawi, Computational and Mathematical Methods in Medicine, Volume 2017, 29 pages, doi.org/10.1155/2017/378195

4. Usame Ömer OSMANOĞLU, Görüntü İşleme ve Analizinin Tıpta Kullanımı ve Bir Uygulama (Yüksek Lisans Tezi)

5. J. Diz, G. Marreiros, A. Freitas. Applying Data Mining Techniques to Improve Breast Cancer Diagnosis. doi: 10.1007/s10916-016-0561-y,2016

6. Pang-Ning Tan, Michael Steinbach, Vipin Kumar. Introduction to Data Mining

7. H. Li, S. Zhuang, D. Li, J. Zhao, Y. Ma. Bening and malignant classification of mammogram images based on deep learning. Biomedical Signal Processing and Control 2019. doi:10.1016/j.bspc.2019.02.017

8. S. Min, B. Lee, S. Yonn. Deep learning in bioinformatics. Briefings in Bioinformatics, 2017, 851-869. doi: 10.1093/bib/bbw068

9. D. Selvathi, A. Aarthy Poornila. Deep Learning Techniques for Breast Cancer Detection Using Medical Image Analysis. doi: 10.1007/978-3-319-61316-1_8

10. Thurman Gillespy and Alan H. Rowberg. Displaying Radiologic Images on Personal Computers. Journal of Digital Imaging. doi: 10.1007/bf03168488

11. J. Li, L. Szekely, L. Eriksson, B. Heddson, A. Sundbom, K. Czene, P. Hall and K. Humphreys. High-throughput mammographic-density measurement: a tool for risk prediction of breast cancer. doi: 10.1186/bcr3238, 2012

12. P. Hamet, J. Tremblay. Metabolism Clinical and Experimental. Artificial intelligence in medicine. doi: 10.1016/j.metabol.2017.01.011

13. R. M. Haralick, K. Shanmugam, and I. Dinstein, Textural Features of Image Classification, IEEE Transactions on Systems, Man and Cybernetics, vol. SMC-3, no. 6, Nov. 1973

14. L. Soh and C. Tsatsoulis, Texture Analysis of SAR Sea Ice Imagery Using Gray Level Co-Occurrence Matrices, IEEE Transactions on Geoscience and Remote Sensing, vol. 37, no. 2, March 1999.

15. D A. Clausi, An analysis of co-occurrence texture statistics as a function of grey level quantization, Can. J. Remote Sensing, vol. 28, no.1, pp. 45-62, 2002

T6

Dairesel(Circular) Verinin Tanıtılması Gerçek Veri Seti Üzerine Uygulama

Hülya BİNOKAY[1], Yaşar SERTDEMİR[1]

[1]Çukurova Üniversitesi Tıp Fakültesi Biyoistatistik Anabilim Dalı, Adana

Email: *hbinokay@cu.edu.tr*

ÖZET

Amaç: Dairesel veri, belirli bir başlangıç noktası ve dönüş yönü (saat yönü veya tersi) olan birim çemberin çevresi üzerinde nokta olarak gösterilebilen gözlemler topluluğuna denir. Olaylar periyodik olarak gerçekleştiğinde (haftanın günlerinde ya da günün saatlerinde) gözlenen olayların oluşturduğu verilerdir. Periyodik yapısından dolayı doğrusal verilerden farklıdır ayrıca dairesel veriler aralıklı ölçek içerisinde değerlendirilmektedir. Bu veriler yönlü ya da yönsel veriler, açısal veriler olarak da bilinmektedir.

Sağlık alanında hastalıkların ortaya çıkma zamanları ve ölümlerin zamana göre dağılımı dairesel veriyi oluşturmaktadır. Örneğin, gün içinde acil servisine başvuru saatleri, hastaların acil servise ulaştırılma süreleri, epilepsi hastasının nöbet geçirme zamanları ve yıl içerisinde acile alerjik astım atağı ile başvuran hastaların başvuru zamanları dairesel veri olarak değerlendirilebilir. Bu veriler doğrusal yapıda olan verilerden farklı tanımlayıcı istatistik formüllere ve dağılımlara sahiptir.

Bu çalışmada, gerçek bir veri seti üzerinden, dairesel verilerin tanımlayıcı istatistiklerinin hesaplanması, önemlilik testlerinin incelenmesi ve görselleştirilmesi amaçlanmıştır.

Yöntem: Dairesel verilerin analizinde kullanılan hesaplamaların gösterilmesi amacıyla bir ilimizdeki özel ve devlet hastanelerine 2016 yılında koah şikayeti ile başvuran hastaların başvuru tarihleri, cinsiyetleri ve kükürt değerlerinin bulunduğu veri seti kullanılmıştır. Cinsiyet gruplarına göre Başvuruların ve havadaki kükürt değerinin yüksek olduğu günlere karşılık gelen açıların önemlilik testleri ve tanımlayıcı istatistikleri hesaplanmıştır. Düzgün dağılıma uygunluğu belirlemek için Rayleigh, testi ve gruplar arası farklar için Watson Wheeler veya Williams testleri kullanılmıştır. Analizler için R 3.3.3 paket programı ve görselleştirmeler için NCSS ve SPSS paket programları kullanılmıştır.

Bulgular ve Sonuç:

	Başvuruların yüksek olduğu günler		**Kükürt gruplarına göre**		**Havadaki Kükürt değerinin yüksek olduğu durum**	
Tanımlayıcı İstatistikler	**Erkek (n=11079)**	**Kadın (n=4397)**	**So2≥20 (n=1410)**	**so2<20 (n=14993)**	**Erkek (n=1025)**	**Kadın (n=385)**
Ortalama yön ($\bar{\theta}$)	35.65^0	55.33^0	1.41^0	93.01^0	1.30^0	1.69^0
Ortalama vektör uzunluk ($\bar{R}$)	0.10	0.11	0.87	0.06	0.87	0.86
	p= 0.039		**p<0.001**		**p<0.001**	
Dairesel varyans (V)	0.90	0.89	0.13	0.94	0.13	0.14
Standart sapma (v)	2.12	2.10	0.54	2.36	0.54	0.55
Von Mises yoğunlaşma parametresi (κ)	0.21	0.22	4.01	0.12	4.05	3.93

Başvuruların yüksek olduğu günler göz önüne alındığında koah hastası olan erkek ve kadınların o rtalama başvuru zamanı kış aylarında olmaktadır. Cinsiyete göre bakıldığında başvuru tarihleri a çısından fark gözlenmiştir.
Havadaki kükürt değerinin yüksek olduğu durum incelendiğinde koah hastası olan erkek ve kadı nların ortalama başvuru zamanı kış aylarında olmaktadır. Cinsiyete göre bakıldığında başvuru ta rihleri açısından fark gözlenmiştir. Kükürdün yüksek olduğu grupta ortalama başvuru zamanı kı ş aylarında iken düşük olduğu grupta nisan aylarındadır. Düşük ve yüksek kükürt grupları için ba şvuru tarihleri bakımından farklılık gözlenmiştir. Koah atakları daha çok kış aylarında oluşmakta dır ve havadaki kükürt değerlerinin yüksek olduğu günlerde ataklar artmaktadır. Erkeklerde kad ınlara göre daha çok görülmektedir. Sonuç olarak mevsimsel olarak meydana gelen hastalıkların dairesel veriler için uygulanabilirliği gözlenmiştir.

Anahtar Kelimeler: Dairesel veri, Von Mises, Doğrusal veri.

1.GİRİŞ

Dairesel veri, belirli bir başlangıç noktası ve dönüş yönü (saat yönü veya tersi) olan birim çemberin çevresi üzerinde nokta olarak gösterilebilen gözlemler topluluğuna denir. Olaylar periyodik olarak gerçekleştiğinde (haftanın günlerinde ya da günün saatlerinde) gözlenen olayların oluşturduğu verilerdir. Periyodik yapısından dolayı doğrusal verilerden farklıdır ayrıca dairesel veriler aralıklı ölçek içerisinde değerlendirilmektedir. Bu veriler yönlü ya da yönsel veriler, açısal veriler olarak da bilinmektedir. (Arthur Pewsey 2013).

Dairesel veriler meteoroloji, fizik, psikoloji, tıp ve biyoloji gibi bilimin birçok alanında ortaya çıkmaktadır. Meteorolojide rüzgar yönü, biyolojide hayvanların yön hareketleri, okyanus akım yönü, kemik kırılma düzleminin yönü gibi durumlarda da dairesel veriler oluşmaktadır. Dairesel verileri sadece yönlere dayalı veriler olarak düşünmek doğru değildir. (Fisher (1993) and Mardia and Jupp (1999)).

Yıl ya da gün içerisinde tekrar eden her olay için kullanılmaktadır. Bu olaylar 24 dilimlik bir saat ya da olayların tarihleri ile ifade edilebilir. Bu saat ve tarihler açıya dönüştürülerek işlem yapılmaktadır. Bu açılar derece veya radyan cinsinden ifade edilmektedir.

Sağlık alanında hastalıkların ortaya çıkma zamanları ve ölümlerin zamana göre dağılımı dairesel veriyi oluşturmaktadır. Örneğin, gün içinde acil servisine başvuru saatleri, hastaların acil servise ulaştırılma süreleri, epilepsi hastasının nöbet geçirme zamanları ve yıl içerisinde acile alerjik astım atağı ile başvuran hastaların başvuru zamanları dairesel veri olarak değerlendirilebilir. Bu veriler doğrusal yapıda olan verilerden farklı tanımlayıcı istatistik formüllere ve dağılımlara sahiptir. En sık kullanılan dağılımlar ise Von Mises diğer adıyla dairesel normal dağılım ve Düzgün (Uniform) dır.

Doğrusal verilerde olduğu gibi dairesel veri yapılarında da normallik varsayımı vardır. Bu normallik varsayımınn test edilmesinde Von Mises dağılımı kullanılmaktadır. Von Mises dağılımının yoğunlaşma parametre değeri büyüdükçe normal dağılıma yaklaşır. Gruplar arasında farklılığı araştırırken normal dağılım varsayımı sağlanmalı ve karşılaştırılacak grupların aynı konsantrasyon parametresine sahip olduğu belirlendikten sonra gruplar arası farklılık Watson Williams testi ile belirlenir. Normalik ve/veya konsantrasyon parametreleri varsayımları sağlanamadığı durumlarda Watson Wheeler testi kullanılır.

2. YÖNTEM

Çalışmamızda başvuruların yüksek olduğu günler ve havadaki kükürt değerinin yüksek olduğu durum göz önüne alındığında cinsiyet gruplarına göre; açıya dönüştürülen başvuru tarihlerinin tanımlayıcı istatistikleri(Ortalama yön, Ortalama vektör uzunluk, Dairesel varyans, Dairesel standart sapma, Von Mises yoğunlaşma parametresi) hesaplanmıştır. Aynı hesaplamalar kükürt gruplarına göre de yapılmıştır. Düzgün dağılıma uygunluğu belirlemek için Rayleigh testi ve gruplar arası farklar için normal dağılıma uygunluğu durumunda Watson Williams aksi halde Watson Wheeler testi kullanılmıştır.

2.1. Gerçek veri çalışması

Bir ilimizdeki özel ve devlet hastanelerine 2016 yıllında koah şikayeti ile başvuran hastaların başvuru tarihleri, cinsiyetleri ve havadaki kükürt değerlerinin bulunduğu gerçek bir veri seti kullanılmıştır.

2.2.Dairesel tanımlayıcı istatistikler

Düzlemde herhangi bir nokta, dik koordinat sisteminde (x,y) ve kutupsal koordinat sisteminde (r,θ) ile gösterilmektedir. Dairesel veriler açılar ile ifade edildiğinden koordinat sistemi üzerinde sinüs ve cosinüs trigonometrik fonksiyonları kullanılmaktadır. Buradaki dönüşüm;

cos θ= $\frac{x}{r}$ sin θ= $\frac{y}{r}$ şeklindedir.

Birim çember üzerinde noktaların vektörel uzunlukları göz ardı edildiğinden açısal değeri ile işlem yapılmaktadır.

2.2.1. Dairesel ortalama yön

$\theta_1, \theta_2, \ldots, \theta_n$ dairesel veri seti için ortalama yön($\bar{\theta}$)

S=$\sum_{i=1}^{n} sin\,\theta_i$ ve C= $\sum_{i=1}^{n} cos\,\theta_i$

$$\bar{\theta}=\begin{cases} tan^{-1}(S/C), & S \geq 0, C > 0 \\ tan^{-1}(S/C)+\pi & C < 0 \\ tan^{-1}(S/C)+2\pi & S < 0, C \geq 0 \end{cases} \quad (1)$$

şeklindedir. (Fisher 1993).

2.2.2 Ortalama vektör uzunluk

Vektör uzunluğu R olmak üzere;

R=||R|| = $\sqrt{C^2+S^2}$ R∈[0,n] ortalama vektör uzunluğu $\bar{R}$= R/n $\bar{R}$ ∈[0,1] olarak hesaplanır. Ortalama vektör uzunluğu açıların ortalama yön etrafında ne kadar yoğunlaştığını gösteren bir ölçüdür. $\bar{R}$ 1'e yakınsa daire üzerindeki bütün açıların yönlerinin benzer olduğunu gösterir. $\bar{R}$ 0'a yakınsa bütün açılar daire üzerine yayılmıştır.(Mardia ve Jupp, 2000)

2.2.3. Dairesel Varyans

Dairesel varyans açılardaki ortalama yöne ilişkin değişimi ölçer. Ortalama vektör uzunluğu açıların ortalama yön etrafında ne kadar yoğunlaştığını gösteren bir ölçüm olduğundan ortalama yönün etrafında bir dağılım ölçüsü olan varyans ile ortalama vektör uzunluğu arasında ilişki vardır.

V=1-$\bar{R}$ şeklinde hesaplanır. (2)

Küçük dairesel varyans değeri açıların daire üzerinde yoğunlaştığını ifade eder.(Berens P)

Standart sapma ise varyansa uygulanan dönüşüm ile elde edilir.

$$\upsilon=\sqrt{-2\ln\bar{R}} \qquad 0<\upsilon<\infty \quad (3)$$

Dairesel verilerde doğrusal verilerde olduğu gibi varyans küçüldükçe dağılım homojen hale gelir ancak doğrusal varyansın aksine dairesel varyans 0 ile 1 arasında değer alır.(Fısher 1993)

Tüm açılar daire üzerinde aynı yönde ise yani düzgün dağılmamışsa, ortalama vektör uzunluk 1'e yakın olacak buna bağlı olarak varyans minimum olacaktır.

Tüm açılar dairenin çevresine eşit olarak dağılmışsa yani düzgün dağılıma uygunsa, ortalama vektör uzunluk 0'a yakın olacaktır ve buna bağlı olarak varyansda 1'e yakın olacaktır

2.2.4. Konsantrasyon Parametresi

Veri setindeki açı değerlerinin daire üzerinde homojen dağılıp dağılmadığını veya ortalama yönünde bir konsantrasyon gösterip göstermediğini belirler. κ ile gösterilir. Bu değer 0'a yakın olduğunda açıların daire üzerinde düzgün bir şekilde dağıldığını, 2'den büyük olduğunda ortalama yönde yoğunlaştığını söyleyebiliriz. Konsantrasyon parametresi için uygun bir yaklaşım;

$$\kappa=\begin{cases} 2\bar{R}+\bar{R}^3+5\bar{R}^5/6 & \bar{R}<0.53 \\ -0.4+1.39\bar{R}+\frac{0.43}{1-\bar{R}} & 0.53\le\bar{R}<0.85 \\ 1/(\bar{R}^3-4\bar{R}^3+3\bar{R}) & \bar{R}\ge 0.85 \end{cases} \quad (4)$$

Şeklindedir.(Gaile and Burt 1980)

2.2. Dairesel Olasılık Dağılımları

Dairesel verilerdeki olasılık dağılımları, birim çemberin çevresi üzerinde toplanmış olan olasılık dağılımlarıdır. Dağılımdaki her bir açı çember üzerinde bir nokta ile gösterildiğinden her bir olasılık bir açıyı belirtmektedir. Radyan cinsinden rastgele değişken olan θ'nın değişim aralığı [0,2π) aralığındadır. En önemli dağılımlar düzgün(uniform) ve dairesel normal dağılım olarak ifade edilen Von Mises dağılımıdır.(Jammalamadaka and gupta, 2001)

2.2.1 Düzgün Dağılım

Dairesel düzgün dağılım açıların daire etrafında düzgün bir şekilde dağıldığını ifade eder. Bütün açılar daire üzerinde eşit olasılığa sahiptir. Daire çevresinde açılar eşit olarak dağıldığından ortalama vektör uzunluk 0'a varyans 1'e yakındır. Olasılık yoğunluk fonksiyonu

$$f(\theta)=\frac{1}{2\pi} \quad 0\le\theta<2\pi \quad (5)$$

2.2.2. Dairesel Normal Dağılım(Von Mises)

Von Mises(1918) tarafından önerilen bir dağılımdır. Dairesel rastgele değişken θ normal dağılıma sahipse dağılım Von Mises dağılımı olarak ele alınmaktadır. Olasılık yoğunluk fonksiyonu

$$f(\theta;\mu,\kappa)=\frac{1}{2\pi I_o\kappa}e^{\kappa\cos(\theta-\mu)} \quad (6)$$

$0\le\theta<2\pi$

$0\le\kappa<\infty$

$-\pi\le\mu\le\pi$ şeklindedir. VM(μ, κ) dairesel normal dağılıma sahiptir. Eşitlik 6 da yer alan $I_o(\kappa)$ fonksiyonu,

$$I_o(\kappa)=\frac{1}{2\pi}\int_0^{2\pi}e^{\kappa\cos(\theta-\mu)}d\ \theta \quad (7)$$

olarak tanımlanan ve μ ortalama yönü, κ yoğunlaşma parametresini göstermek üzere Bessel fonksiyonunu ifade eder.(Von Mises 1918) Yoğunlaşma parametresi ne kadar büyük olursa ortalama yön etrafında kümelenme artacaktır. Yoğunlaşma parametre değeri büyüdükçe normal dağılıma yaklaşacaktır.(Gaile and Burt 1980)

2.3. Önemlilik testleri

En çok kullanılan olasılık dağılımlarından biri düzgün dağılımdır. Düzgün dağılım açıların daire üzerinde dağıldığını mı yoksa ortalama yöne doğru yoğunlaşıp yoğunlaşmadığını gösterir. Bu durum için kullanılan testler vardır. Bu testlerden en çok kullanılan ise Rayleigh testidir.

Rayleigh testi Mardia ve Jupp(2000) tarafından ele alınmıştır. R vektörü büyüdüğünde, ortalama yön 0'dan uzaklaştığında düzgün dağılımı red etme fikrine dayanır.(Fisher, Lewis ve Embleton, 1993) test istatistiği $2n\bar{R}^2$

2.3.1 Watson Williams Testi

Test edilen sıfır hipotezi H_0: $\mu_1=\mu_2$ olmak üzere $\mu_1$ve μ_2 ortalama yönler arasında farklılık olup olmadığını test eder. Watson ve Williams tarafından önerilmiştir. Normallik varsayımının sağlanması durumunda kullanılmaktadır.(Stephens 1972)

Kullanılan test istatistiği;

$F=K\frac{(N-2)(R_1+R_2-R)}{N-R_1-R_2}$ $N= n_1+n_2$

R rayleigh tetsinden elde edilir. K grup sayısıdır. $F_{\alpha,1,N-2}$ tablo değeri ile kıyaslanır.

2.3.2 Watson Wheeler Testi

Parametrik olmayan bir yöntemdir. Watson ve Wheeler tarafından önerilmiştir. Sıfır hipotezi H_0: $\mu_1= \mu_2$ test edilir. Her bir a açısı sıralanır ve her biri için dairesel sıralama denilen bir hesaplama yapılır.

$d=\frac{(360^0)(açının\ sırası)}{N}$

tüm açılar daire etrafına eşit bir şekilde yerleştirilir.

$C_i=\sum_{j=1}^{ni} cosdj$ $S_i=\sum_{j=1}^{ni} sindj$ i iki gruba karşılık gelir.

Bu durumda test istatistiği

$W=\frac{2(N-1)(C_i{}^2+S_i{}^2)}{n1n2}$ (Batschelet, 1981)

3. BULGULAR

Yapılan çalışmada elde edilen bulgular Koah şikayeti ile başvuran hastalar için tanımlayıcı istatistikler ve önemlilik testleri olarak verilmiştir.

Başvuruların yüksek olduğu günler için elde edilen bulgular tablo 1'de gösterilmiştir.

Erkek ve Kadın hastaların ortalama başvuru zamanı kış aylarında olmaktadır. Ortalama vektör uzunluğu ve yoğunlaşma parametresi her iki grupta da 1 değerinden uzakta olduğundan daire üzerinde düzgün dağıldığını söyleyebiliriz yani erkek ve kadın hastaların başvuru zamanları yılın her gününde oluşabilir. Yapılan düzgün dağılıma uygunluk testi sonucunda da grupların düzgün dağıldığı gözlenmiştir. Cinsiyet gruplarına göre yoğunlaşma parametreleri 1'den uzakta olduğu için normallik varsayımı sağlanmadığından Watson Wheeler testi uygulanmıştır ve cinsiyet grupları için ortalama başvuru zamanları arasında fark gözlenmiştir. (tablo 1)

Tablo 1 Cinsiyet gruplarına göre tanımlayıcı istatistikler ve önemlilik testleri

Tanımlayıcı İstatistikler	**Erkek(n=11079)**	**Kadın(n=4397)**
Ortalama yön	35.65^0	55.33^0
Ortalama vektör uzunluk	0.10	0.11
Dairesel varyans	0.90	0.89
Dairesel standart sapma	2.12	2.10
Von Mises yoğunlaşma parametresi	0.21	0.22
Rayleigh testi	p_{erkek}=0.074	$P_{kadın}$=0.289
Watson Wheeler	p=0.038	

Kükürdün yüksek olduğu grupta ortalama başvuru zamanı kış aylarında iken düşük olduğu grupta nisan aylarındadır. Ortalama vektör uzunluğu kükürdün yüksek olduğu grupta 1'e yakın ve yoğunlaşma parametresi 1'den büyüktür.

Kükürdün düşük olduğu grupta ortalama vektör uzunluğu 0'a yakın ve yoğunlaşma parametresi 1'den uzaktadır. Bu durumda kükürdün yüksek olduğu grup düzgün dağılıma sahip değilken kükürdün düşük olduğu grup daire üzerinde düzgün dağılmaktadır. Yapılan düzgün dağılıma uygunluk testi sonucunda da, kükürdün düşük olduğu grubun düzgün dağıldığı fakat kükürdün yüksek olduğu grubun düzgün dağılmadığı gözlenmiştir.

Kükürdün yüksek olduğu grup için yoğunlaşma parametre değeri 1'den büyük olduğu için normal dağılıma uyduğu ancak kükürdün düşük olduğu grup için yoğunlaşma parametre değeri 1'den küçük olduğundan normal dağılıma uymadığı gözlenmiştir. Konsantrasyon parametreleri incelendiğinde fark gözlenmiştir($p<0.001$). Bu durumda düşük ve yüksek kükürt grupları için Watson Wheeler testi uygulanmış ve başvuru zamanları için anlamlı fark gözlenmiştir(tablo 2).

Tablo 2 Kükürt gruplarına göre tanımlayıcı istatistikler ve önemlilik testleri.

Tanımlayıcı İstatistikler	**So2≥20 (n=1410)**	**so2<20 (n=14993)**
Ortalama yön	1.41^0	93.01^0
Ortalama vektör uzunluk	0.87	0.06
Dairesel varyans	0.13	0.94
Dairesel standart sapma	0.54	2.36
Von Mises yoğunlaşma parametresi	4.01	0.12
Rayleigh testi	$P_{yüksek}<0.001$	$P_{düşük=}0.122$
Watson Wheeler	P<0.001	

Havadaki kükürt değerinin yüksek olduğu durum göz önüne alındığında;

Erkek ve kadın hastaların ortalama başvuru zamanı kış aylarında olmaktadır. Ortalama vektör uzunluğu her iki grupta da 1'e yakın olduğundan daire üzerinde düzgün dağılmadığını söyleyebiliriz yani erkek ve kadın hastaların başvuru zamanları yılın belirli bir döneminde yoğunlaştığı söylenebilir. Yapılan düzgün dağılıma uygunluk testi sonucunda da cinsiyet gruplarının düzgün dağılmadığı gözlenmiştir. Cinsiyet grupları aynı yoğunlaşma parametresine sahiptir ve konsantrasyon parametreleri 1'den büyüktür. Bu durumda cinsiyet grupları için Watson Williams testi uygulanmış ve başvuru zamanları için anlamlı fark gözlenmiştir.(tablo 3)

Tablo 3 Cinsiyet gruplarına göre tanımlayıcı istatistikler ve önemlilik testleri

Tanımlayıcı İstatistikler	**Erkek (n=1025)**	**Kadın (n=385)**
Ortalama yön	1.30^0	1.69^0
Ortalama vektör uzunluk	0.87	0.86
Dairesel varyans	0.13	0.14
Dairesel standart sapma	0.54	0.55
Von Mises yoğunlaşma parametresi	4.05	3.93
Rayleigh testi	$P_{erkek}<0.001$	$P_{kadın}<0.001$
Watson Wheeler	$p<0.001$	

4. TARTIŞMA VE SONUÇ

Bilimsel araştırmalarda verinin yapısına göre analiz yöntemleri farklılık göstermektedir. Bu çalışmada dairesel veriler için tanımlayıcı istatistikler ve önemlilik testleri, gerçek bir veri seti kullanılarak tanıtılmıştır. Bu veri setinde erkek ve kadın KOAH hastalarının başvuru tarihleri göz önüne alındığında, kükürdün yüksek olduğu günlerde erkek ve kadın hastalarının ortalama başvuru zamanının ocak ayına, kükürdün düşük olduğu günlerde ise ortalama yön nisan aylarında olmaktadır

Doğrusal verilere uygulanan aritmetik ortalama hesaplansaydı, ortalama başvuru zamanı yaz aylarında gözlenecekti. Bu durum göz önüne alındığında dairesel tipteki verilere, doğrusal veri gibi yaklaşmanın sakıncaları olduğu gözlenmiştir.

Koah hastalığı olan kişiler için akut alevlenmesi genellikle kış aylarında ortaya çıktığı için bizim sonucumuzu destekler niteliktedir.

Sonuç olarak mevsimsel olarak meydana gelen hastalıkların dairesel veriler için uygulanabilirliği gösterilmiştir.

5. KAYNAKÇA

1. Pewsey A,Neuhauser M, Ruxton GD (2013). Circular statistics in R
2. Fisher NI. Statistical Analysis of Circular Data. Cambridge University Press, New York
3. Mardia KV. Statistics of Directional Data. Academic Press. Inc., London
4. Berens P. CircSat: A MATLAB Toolbox for Circular Statistics. Journal of Statistical Software
5. Gaile GL, Burt JE. Directional Statistics. Concepts and Techniques in Modern Geography, London,. (1980)
6. Jammalamadaka SR, SenGupta A. Topics in Circular Statistics. London: World Scientific; 2001

7. Stephens, M. A. (1972). Multisample tests for the von Mises distribution. Journal of the American Statistical Society, 67, 456–61
8. Batshcelet, E. 1981. Circular Statistics in Biology. Academic Press, New York.
9. Fisher, N., Lewis, T., and Embleton, B. (1993). Statistical analysis of spherical data. Cambridge University Press.
10. Mardia, K. and Jupp, P. (2009). Directional statistics, volume 494. Wiley
11. von Mises, R. (1918). Über die 'Ganzzahligkeit' der Atomgewichte und verwandte Fragen. *Physikalische Zeitschrift*, 19, 490–500.

T7

COMPARATIVE ANALYSIS OF MNA SCORES WITH LINEAR MIXED EFFECT MODELS AND LINEAR REGRESSION MODELS IN PATIENTS WITH PEG FEED

Neslihan GÖKMEN[1], Arzu BAYGÜL[2], Aydın ERAR[3]

[1] Istanbul Technical University Basic Sciences, Istanbul
[2] Koç University Biostatistics, Istanbul
[3] Mimar Sinan Fine Arts University, Statistics, Istanbul

Mixed effects models are powerful tools for longitudinal data analysis. It incorporates the time points of the measurements as a continuous variable and characterized by the ability to model the mean response level as a combination of population characteristics and individual-specific effects shared by all individuals. Therefore, it is frequently used in health studies.

Aim: The aim of this study is to present the performance evaluation of linear mixed effect models and classical linear regression models used in longitudinal data analysis based on the health data. For this purpose, it was aimed to determine the parameters affecting the MNA (Mini Nutritional Assessment) score in PEG feed patients.

Method: The MNA scores,which were measured at baseline, 3rd months, 6th months and 12th months after PEG implantation from 53 patients included in the study .After PEG implantation, survival time was recorded on a daily basis and mortality was included in the analysis as a group variable. Linear mixed effect models were estimated by using the restricted maximum likelihood estimation method, which is the best linear unbiased estimator. Least squares estimation was used in linear regression modeling. Linear mixed effect models with fixed and random effects were modeled as univariate and multivariate. In the selection of variables, the significance of the model was evaluated by ANOVA test. Information criteria (AIC,BIC-Akaike, Bayesian information criterion) were used to select the appropriate model.

Results: Univariate model with random effects performance of four different models was found to be higher than the others. While the mean square errors and confidence interval width were lower, the model significance was high. The random effects constant varies by measurement and group, and the inclusion of individual effects in the model also strengthens the model performance. MNA score is the most important factor affecting mortality of patients are said to have shown a decrease of 12.2 units compared to those living for MNA levels.

Conclusion: According to the findings and application results in the literature; the performance of the linear mixed effect model was found to be higher than the multiple regression model. Since it is not possible to provide multiple linear regression assumptions in longitudinal and unstable data, it is preferable to use linear mixed effect models.

Keywords: linear mixed effects models, fixed effect, random effect, multiple linear regression models

COMPARATIVE ANALYSIS OF MNA SCORES WITH LINEAR MIXED EFFECT MODELS AND LINEAR REGRESSION MODELS IN PATIENTS WITH PEG FEED

1. INTRODUCTION

Mixed effect models are powerful tools in longitudinal data analysis (Wu and Zhang, 2006). This study is carried out to find the best fitted model for a longitudinal case in health. In the first part, the structure and advantages of linear mixed effects models are mentioned. Secondly, the structure of the data and application tools are explained. In the third section, the findings are given and finally, the conclusions and discussion section is located.

Mixed effect models are also referred to by other names such as random effect models, random coefficient models, hierarchical models, and multilevel models (Armitage and Colton, 1998). The variance analysis method studied by Fisher in 1937 is known as separating the sum of squares from the mean into components. However, since variance analysis does not give good results in unstable data, methods based on the theory of linear models have emerged for variance analysis in recent years. Restricted maximum likelihood (REML) estimator, developed by Patterson and Thompson in 1971, is the best linear unbiased estimator (BLUE) used for cross-sectional random effects in mixed effect models.

There are two classical approaches commonly used in the analysis of longitudinal studies. The first is based on repeated measurements ANOVA, and the second is based on multivariate ANOVA (MANOVA). It is assumed that the homogeneous errors between the groups are normally distributed in both models. For both models, the primary objective is to compare group means, and none of these models is informative about individual change curves (eg, individual-specific trends). In addition, the time points at which repeated measurements were taken are the same for each individual and the same number of repeated measurements is required from each individual. Such a data structure is balanced. This prevents analysis of unbalanced designs in which different individuals are measured under different conditions. Especially in cases where the experimental units are human, it may not be possible to have the same number of observations taken from the units, even if the study is designed with attentive care. During the experiment, there may be individuals leaving the experiment (Doğanay, 2007). In addition, in the variance analysis methods for both univariate and multivariate repeated measurements, the variance covariance matrix of repeated measurements is assumed to be the same for all groups. However, this assumption is not very accurate in practice. For example; individual monitoring from individuals with high systolic blood pressure tend to vary more than individuals with low systolic blood pressure. That is, the variance of within-subject error will be higher in the group with high systolic blood pressure. This will be reflected in the variance of repeated observations. Large response values will have large variance, small response values will have small variance. In such a case, it would be wrong to assume that observations in different groups have the same variance covariance matrix.

The classification chart for some of the models are given Figure 1. While general linear models are divided into mixed and generalized linear models, mixed and generalized linear models combine as generalized linear models.

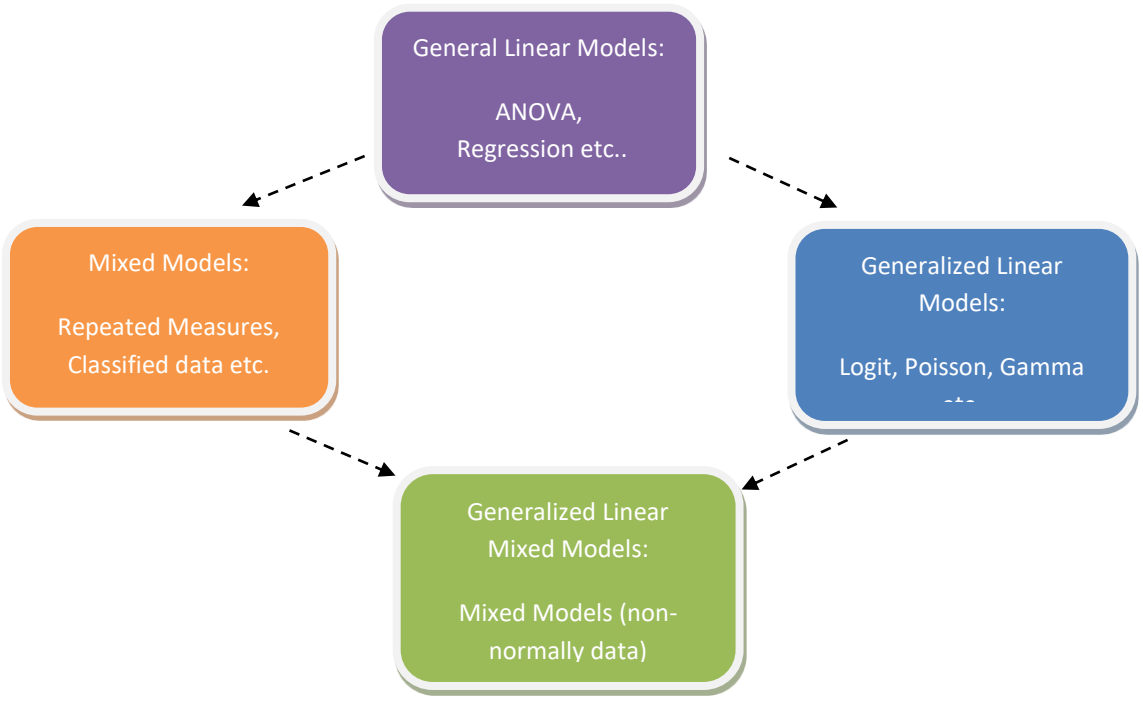


Figure 1. Classification of models

Linear mixed models, in addition to the fixed effects of linear estimators, generalized linear effects with random effects is a state of models (Koç&Cengiz, 2012).

1.1 Fixed and Random Effects

Fixed effects are selected from the specific interest levels of a factor. We can use all levels or a subset of levels. Only those levels can be deduced. If levels are predetermined, they should be considered as fixed effects.

Random effects are levels of a factor selected from the probability distribution. Random effects are interested in causing the variance in the dependent variable. Random effects are interested in causing the variance in the dependent variable. Rather than dealing with individual instruments and fixed factor levels, we are interested in the variance of instruments between a random factor level. The randomly selected from a larger population of active levels of these effects can be considered as random effect. At mixed-effect models, fixed effects are taken as population effects and random effects are taken as individual effects that show deviations from the population values. Measurements that share the same level of random effect are modeled in relation (Vittinghoff et al., 2005).

1.2 Advantages of Linear Mixed Effects Models

Subjects do not have to have the same number of measurements, so subjects with incomplete measurements may also be included in the analysis. Another important feature is that mixed effect models add time points to the analysis as a continuous variable. For this reason, it is no longer necessary to have the measurements taken at the same time points. Both time-independent and time-dependent common variables can be included in the model. In this case, the change in response variable can be defined both by the invariant characteristics of the

individual (such as gender, race) and by characteristics that change over time (such as cholesterol level, age) (Hedeker and Gibbons, 2006). Since some of the regression parameters differ from individual to individual, these models take into account the sources of natural heterogeneity within the population. That is, individuals in the population have their own average response profiles and regression parameters considered to be random. A distinctive feature of linear mixed-effect models is that it can model the mean response level as a combination of population characteristics (population covariates, ie. β) shared by all individuals and individual-specific effects (ie. ν). While population characteristics are constant effects, individual-specific effects are random effects. For example; estimation of individual growth curves over time can be done with the help of linear mixed effect models. In a clinical trial, it may be desirable to make individual inferences rather than the overall mean response of the subjects in the study. Such estimates can be used in the clinical study to identify subjects who do not respond well to the treatment they receive in the assigned group. A striking feature of the mixed effect models is that it is not necessary for repeated observations from individuals to be taken under the same number and / or under the same conditions. Therefore, these models are suitable for the analysis of repetitive measurement data in unbalanced structure (Fitzmaurice et al., 2004).

1.3 Structure of Linear Mixed Effects Models

Linear mixed effect model is given below.

$$y_i = X_i\beta + Z_i\nu_i + e_i$$
$$\nu_i \sim N(0, G)$$
$$e_i \sim N(0, R_i)$$
$$\nu_1, \ldots, \nu_N$$
$$e_1, \ldots, e_N$$

β : (p×1) dimensional population parameters (constant effects) vector

ν_i : (qx1) dimensional individual-specific parameters (random effects) vector

X_i: ($n_i$xp) dimensional covariate matrix

Z_i: ($n_i$xq) dimensional covariate matrix can be defined as q ≤ p.

Z_i is also known as the design matrix that binds the vector of random effects (ν_i). In fact, the columns of Z_i are a subset of the columns of X_i. The model, which parameters will vary from individual to individual can be determined by the columns of X_i corresponding to Z_i. That is, if any parameter in the β vector matches which column in X_i, the corresponding column in Z_i, the design matrix for random effects, can be generated (random) between subjects. Random effects in the model are assumed to have a multivariate normal distribution with ν_i , zero mean and G covariance matrix (Kuznetsova, Brockhof& Christensen,2017).

Linear mixed effect model, the vector of β regression parameters (fixed effects) is assumed to be the same for all individuals and has population-averaged interpretations. The average of the change profiles in the responses of all individuals would give the average change profile for the population and this change profile would be assumed to be the same for all individuals. The random effects vector vi contains individual regression parameters unlike β. These are random effects and, when combined with fixed effects, give the average response profile of any individual.

1.4 Restricted Maximum Likelihood Estimation (REML)

The mean response vector y_i is assumed to show a multivariate normal distribution with $E(y_i)=X_i\beta$ mean and $Cov(y_i) =\Sigma_i=\Sigma_i(\theta)$ covariance matrix. Here, θ is the vector of covariance parameters forming q × 1 dimensional which is composing Σ_i.

The main idea of RMLE is to separate the part of the data used in the prediction of Σ_i, and the part used in the prediction of β (Fitzmaurice et al., 2004). Thus, Σ_i 's estimation will be based on only the relevant portion of the available data. That is, the basic idea in RMLE is to remove β from likelihood. Thus, the estimator is defined only in terms of Σ_i. There are several ways to do this. One of the possible means for achieving restricted likelihood; data is the transformation of observations into a set of linear combinations that do not depend on β. For example, after estimating β using the least squares, errors can be used as data for estimating Σ_i (or θ). The likelihood for these errors will depend on θ (Morrel,1998). By making the maximum likelihood of error shown Figure 2, the estimation of θ is obtained by considering the fact that β is also predicted. Note that the additional term in the REML can be shown in the form of the covariance of $\hat{\beta}$.

$$-\frac{1}{2}\sum_{i=1}^{N}\log|\Sigma_i|-\frac{1}{2}\sum_{i=1}^{N}(y_i-X_i\hat{\beta})'\Sigma_i^{-1}(y_i-X_i\hat{\beta})-\frac{1}{2}\log\left|\sum_{i=1}^{N}X_i'\Sigma_i^{-1}X_i\right|$$

$$\Downarrow$$

$$-\frac{1}{2}\log\left|\sum_{i=1}^{N}X_i'\Sigma_i^{-1}X_i\right| = \frac{1}{2}\log\left|\left(\sum_{i=1}^{N}X_i'\Sigma_i^{-1}X_i\right)^{-1}\right|$$

$$= \log\left|Cov(\hat{\beta})\right|^{1/2}$$

Figure 2. Restricted Maximum Likelihood Estimation (Doğanay, 2007)

The maximum likelihood estimation is also corrected to obtain the restricted maximum likelihood estimation.

2. MATERIAL & METHOD

53 patients with PEG implantation were included in the study. The following variables were included in the study in order to evaluate the MNA (Mini Nutritional Assessment) measurements (0. months, 3. months, 6. months, 12. months) according to the mortality of PEG implanted patients. MNA score measurements were taken as dependent variables in the study. The group variable mortality is encoded as lives: 0 and is ex: 1. Time points where repeated measurements were taken were analyzed on a day-by-day basis and without actual coding. Age at baseline and duration of survival throughout the study after PEG implantation were taken as continuous variables. Descriptive statistics such as mean, median (Med.), standard deviation (Std.Dev.), minmum (Min), maximum (Max) were used to define the demographics. Univariate analysis (Mann-Whitney U test) was performed to show significant difference between independent groups distributed non-normally and Wilcoxon Signed Rank test was used dependent non-normally measurements. The linear mixed effects model was predicted by using RMLE, whereas least square estimation was used for predicting multiple linear regression. ANOVA test was utilized in model selection part and information criterias such as AIC, BIC were evaluated to determine the appropriate model. Analyzes were performed by loading lme4, rlm packages in R.

3. RESULTS

The demographics could be seen in Table 1. The average age of the patients included in the study was 78.5±13.4 years and the mean survival time was 148.9±145.5 days. 75.5% (n=40) of the study participants have died at the end of the study.

Table 1. Distributions of demographics

	N	**Med (Min-Max)**	**Mean±Std.Dev.**
Age	53	83 (18-91)	78,5±13,4
Duration of survival time (days)	53	73 (2-365)	148,9±145,5
		N	**%**
Mortality	**Alive**	13	24,5
	Ex	40	75,5

According to Table 2, the significant difference was found between alive and ex in terms of MNA Baseline scores. The alive ones had higher average of MNA score than Ex ones (Mann-Whitney U $p<0.05$). This univariate analysis shows only the effect of the group, and the effect of measurements should be examined.

Table 2. Comparison of mortality according to MNA

	Alive (n=13)	**Ex (n=40)**	
	Mean±Std.Dev. **Med (Min-Max)**	**Mean±Std.Dev.** **Med (Min-Max)**	**p***
MNA Baseline	14.8±15.5 15.5 (7-21.5)	11.4±11 11 (3-22)	**0.048**
MNA 3rd Month	16±17.5 17.5 (8.5-22.5)	15.3±15.5 15.5 (8-20.5)	0.515
MNA 6th Month	21.0±21 3.6 (11-22.5)	17.0±17 2.6 (13-19.5)	0.101
MNA 12th Month	20.3±21.5 21.5 (13.5-24)	-	-

**Mann-Whitney U*

When the effect between measurements is examined, it was seen that the statistically significant difference was found between paired measurements of MNA scores (Wilcoxon Signed Rank test $p<0.05$). MNA score increases as follow-up time increases.

Table 3. Comparison of MNA score measurements

MNA Baseline	**MNA 3rd Month**	**MNA 6th Month**	**MNA 12th Month**	**p^1**	**p^2**	**p^3**
Mean±Std.Dev. **Med (Min-Max)**	**Mean±Std.Dev.** **Med (Min-Max)**	**Mean±Std.Dev.** **Med (Min-Max)**	**Mean±Std.Dev.** **Med (Min-Max)**			
12.2±5.5 13 (3-22)	15.7±3.9 16 (8-22.5)	18.1±3.5 17.3 (11-22.5)	20.3±3.6 21.5 (13.5-24)	**0.003**	**<0.001**	**0.001**

[1]0 vs. 3rd , [2]0 vs. 6th, [3]0 vs. 12th, Wilcoxon test

It is appropriate to use linear mixed effect models to see the effect of time and group together. In addition, it is seen that the distributions are far from normal even though there are no outliers in terms of group and measurement. This also leads us to construct linear mixed effect models.

In the modeling phase, firstly linear regression model was used to obtain classical least squares estimations. MNA score was taken as the dependent variable and Group (Ex vs. Alive) was taken as the independent variable. The model was statistically significant ($p <0.001$). However, the

determination coefficient is quite low. Ex can be said to reduce the MNA score 4,973 times (Table 4).

Table 4. The summary of linear regression model

	R^2	**Adjusted** R^2	**p**	**F**	**MSE**
Model	0.206	0.199	**<0,001**	27.78	24.231
	Unstandardized β	**Standard Erros**	**Standardized** β	**t**	**p**
Constant	17.5	0.683		25.636	**<0.001**
Group	-4.973	0.952	-0.454	-5.223	**<0.001**

The normal P-P graph of the residuals is shown in Figure 3. Assumptions are observed. However, since the effect of differences due to measurement times and individual characteristics is not considered in this model, linear mixed effect models will be tried. When the Q-Q plot is examined, the normal distribution assumption is not provided and the transition to linear mixed effect models can be made as mentioned before.

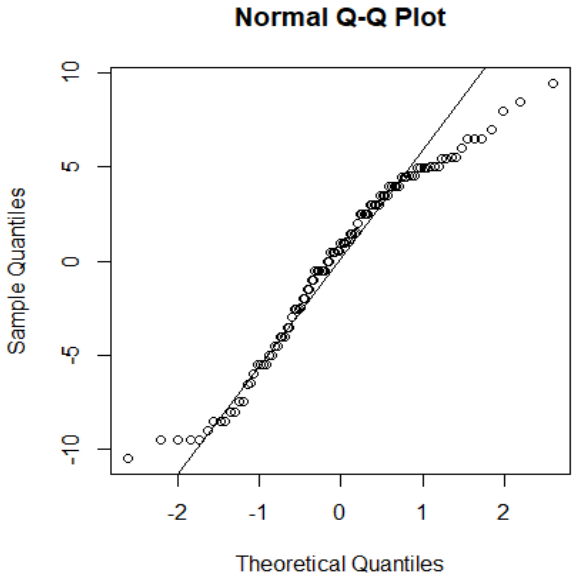


Figure 3. The normal P-P graph of the residuals (least squares)

Linear mixed effects model was used to obtain restricted maximum likelihood estimations. The dependent variable was taken as the MNA score. The fixed effect was taken as Group (Ex vs. Alive), while the random effects were taken as individuals (ID) and measurement period (months). The model was found statistically significant (p<0.001). As seen in Table 5, it can be said that ex group decreases MNA score by 3.610 times. Variability between measurements was calculated as 5.214, inter-individual variability was 23.259 and residual variance was 2.560. The correlation between fixed effects was -0.651.

Table 5. The summary of linear mixed effects model

Random Effects	**Variance**	**Standard Error**	
ID	23,259	4,823	
Months	5,214	2,283	
Residuals	2,560	1,600	
Fixed Effects	**Estimation**	**Standard Error**	**t**
Constant	17,5	1,773	9,873
Group	-3,610	1,593	-2,267
Correlation Between Fixed Effects	-0,651		

Since the normality assumption is not considered in mixed models, the model is valid even if the assumption is not provided according to the Q-Q plot. There is now a random distribution in the plot against the predicted. There is no strong evidence for heteroscedasticity (Figure 4).

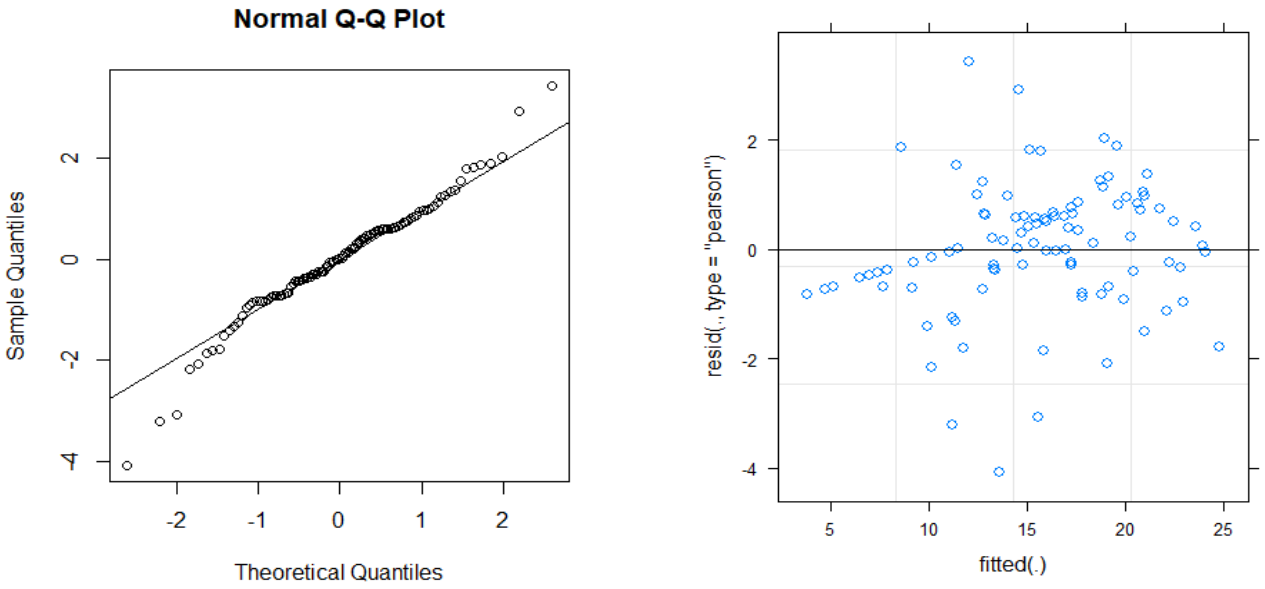


Figure 4. Resiudal graphs (RMLE)

Survival time was evaluated considering multiple fixed effects. When the survival time was considered, the correlation between the average of mna was found to be positively significant (r=0.417, p=0.002). Therefore, it was added to the model as a fixed effect. The model was statistically significant (p <0.001). As seen in table 6, the group ex decreases the MNA score by 0,602 times and and the 1-unit change in the time elapsed after the peg insertion increases the MNA score by 0.011 times. Variability between measurements was calculated as 5.147, inter-individual variability as 23.057 and residual variance as 2.564. The intra-group correlation was -0.912, -0.884 on days, and the correlation between groups and days was 0.851.

Table 6. The summary of linear mixed effects model (with survival time)

Random Effects	Variance	Standard Error	
ID	23.057	4.802	
Months	5.147	2.269	
Residuals	2.564	1.601	
Fixed Effects	**Estimation**	**Standard Error**	**t**
Constant	13.59	3.77	3.605
Group	-0.602	3.02	-0.200
Days after PEG insertion (Survival time)	0.011	0.01	1.171
Correlation Between Fixed Effects	**(Intra)**	**Group**	
Group	-0.912		
Days after PEG insertion (Survival time)	-0.884	0.851	

Since the normality assumption is not considered in mixed models, the model is valid even if the assumption is not provided according to the Q-Q plot. There is now a random distribution in the plot against the predicted. There is no strong evidence for heteroscedasticity (Figure 5).

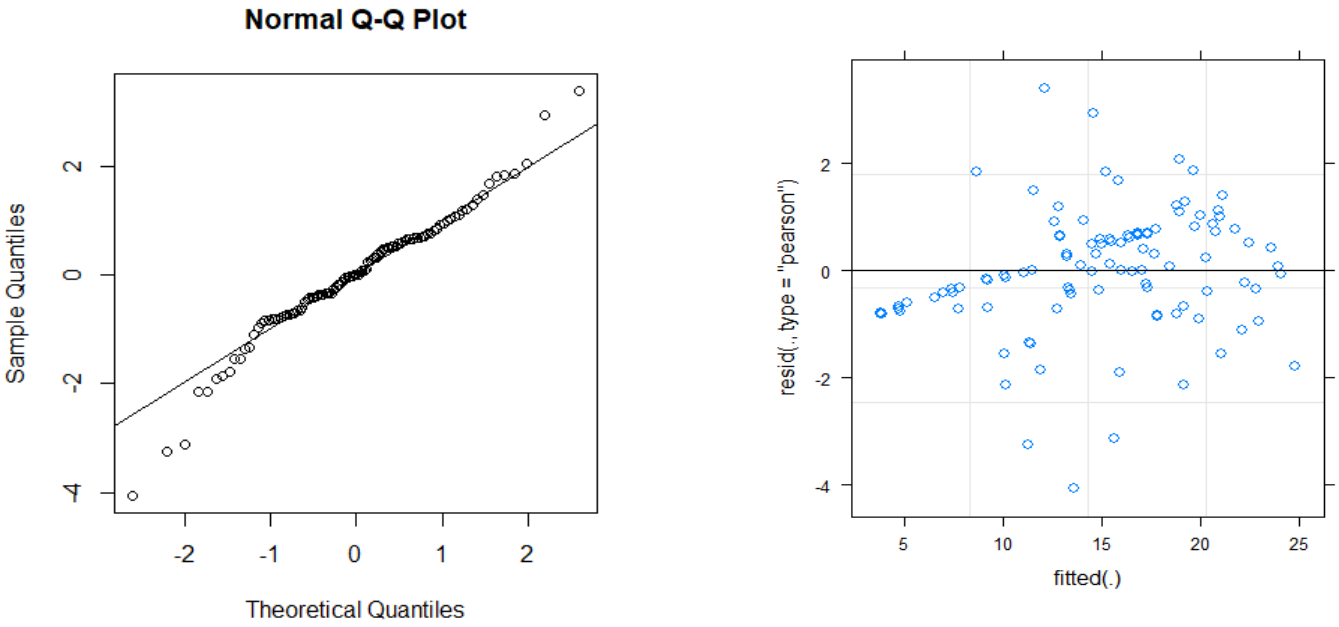


Figure 5. Resiudal graphs (RMLE-with survival time)

After insertion of PEG, the significance of the day variable was examined by ANOVA test and model 1 was found to be more appropriate since p> 0.05. When the information criteria are examined, it can be said that model 1's AIC and BIC values are lower (Table 7). Therefore, Model 1 is included in the comparative results in the following results.

Table 7. The summary of linear mixed effects model (with survival time)

	Df	**AIC**	**BIC**	X^2	**p**
Model 1	5	572.21	585.58		
Model 2	6	572.79	588.82	1.4257	0.232

According to the comparative results in Table 8, although the group variable is more important when estimating with classical regression method, the high variance of random effects is an important evidence for choosing the linear mixed effect model. It can be said that passing the group reduces the MNA score by 3,610 times. The fact that the starting points are different in terms of time and individuals also shows the accuracy of selecting mixed models.

Table 8. The summary of linear mixed effects model (with survival time)

Group	**Estimation**	**Standard Error**	**t**
Linear Regression Model (LS)	-4,973	0,952	-5,223
Linear Mixed Effects Model (Model 1-RMLE)	-3,610	1,593	-2,267

4. DISCUSSION AND CONCLUSION

The feed tube used in the elderly in nursing homes, and can make it difficult to score the MNA. For questions about dietary intake, which is a problem for MNA, the elderly who consume nutritional supplements to score MNA, it will be difficult to score points when nutrients are fed parenterally or enterally (Langkamp-Henken, 2006). However, early detection of poor nutritional status is important to allow appropriate interventions that may have a significant impact on patient morbidity and mortality. In a similar study, those with a MNA score below 17 were found to be at 3.03 (OR) times the risk of mortality (Chan et.al,2010). This study was based on the determination of MNA score levels in peg implanted patients and modeling of the factors affecting MNA score. Since the data are collected longitudinally, it can be said that the mixed effect models established considering the individual and temporal effects are suitable for containing more information. In the estimation made with linear mixed models, it was observed that MNA score had a 3.61 units effect on death, close to the literature.

The first limitation of the study was assuming the model would fit linearly to the data. Non-linear models can be added comparatively in the following studies. The second limitation of the study was having limited demographics information. The research can be expanded by adding gender, drug utilization, chronic diseases.

REFERENCES

1. Wu, H., Zhang, J. (2006). Nonparametric Regression Methods for Longitudinal Data Analysis, John Wiley & Sons, Inc.
2. Armitage, P., Colton, T. (1998). Encyclopedia of Biostatistics, Wiley, New York , 4: 2657-2662.
3. Patterson, H.D., Thompson, R. (1971). Recovery of inter-block information when block sizes are unequal, Biometrica, 58(3): 545-554.
4. Doğanay, B. (2007). "Uzunlamasına Çalışmaların Analizinde Karma Etki Modelleri", Ankara Üniversitesi Sağlik Bilimleri Enstitüsü Biyoistatistik Anabilim Dali Yüksek Lisans Tezi, Ankara.
5. Koç, T.,Cengiz, M.A. (2012). Genelleştirilmiş Lineer Karma Modellerde Tahmin Yöntemlerinin Uygulamalı Karşılaştırılması. Karaelmas Fen ve Mühendislik Dergisi, 2 (2): 47-52.
6. Vittinghoff, E., Shiboski, S.C., Glidden, D.V., Mcculloch, C.E. (2005), Regression Methods in Biostatistics: Linear, Logistics, Survival and Repeated Measrures Models, Springer Science&Business Media Inc.
7. Hedeker D., Gibbons R.D. (2006). Longitudinal Data Analysis, John Wiley & Sons, Inc., Hoboken, New Jersey.
8. Fitzmaurice, G., Laird, N., Ware, J. (2004). Applied Longitudinal Analysis, Wiley Series in Probabilty and Statistics.
9. Kuznetsova, A., Brockhoff, P.B., Christensen, R.H.B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models, Journal of Statistical Software, 82 (13): 1-26.
10. Morrel, C. H. (1998). Likelihood ratio testing of variance components in the linear mixedeffects model using restricted maximum likelihood. Biometrics 54: 1560–1568.
11. Langkamp-Henken, B. (2006). Usefulness of the MNA® in the Long-Term And Acute-Care Settings Within the United States, The Journal of Nutrition, Health & Aging. 10 (6):502-509.
12. Chan, M., Lim, Y.P., Ernest, A., Tan, T.L. (2010). Nutritional assessment in an Asian nursing home and its association with mortality. The Journal of Nutrition, Health & Aging. 14 (1): 23-28.

T8

ULSAM Veri Setinde Tek Nükleotid Polimorfizmi Verileri ile Makine Öğrenimi Yöntemlerinin Kullanımı

R. Onur ÖZTORNACI[1], Bahar TAŞDELEN, Andrew P. MORRIS[2], Hamzah SYED[3]

[1] Mersin University Faculty of Medicine, Department of Biostatistics and Medical Informatics, Mersin, Turkey

[2] School of Biological Sciences, Division of Evolution & Genomic Sciences, University Of Manchester, Manchester, UK

[3] Faculty of Pop Health Sciences, ICH Genetics & Genomic Medicine Prog., University College London, London, UK

Giriş

Genom-boyu ilişki çalışmaları (GWAS) hastalığın genetik yapısını araştırmak için en yaygın kullanılan stratejidir. GWAS metodolojisi, genomdaki Tek Nükleotid Polimorfizmlerine (SNP'ler) göre hastalar ve sağlıklı bireyler arasındaki farkları inceler. Bir hastalıkla ilgili SNP'leri tespit etmeye çalışırken istatistiksel önem eşiğinin tanımlanması esastır. Bu nedenle, büyük eşik değerleri kullanmak, yanlış-pozitif SNP ilişkilerinin sayısını potansiyel olarak artırabilir. Alternatif olarak, küçük eşik değerleri kullanmak, gerçek pozitif SNP ilişkilerinin miktarını azaltabilir. Her iki durumda da yanıltıcı sonuçlar üretilebilir. GWAS için, 5 x10e-8 p-değeri eşiğini kullanmak standart bir uygulama haline gelmiştir [1]. Makine öğrenimi (ML) algoritmalarının uygulanması bu durumda bazı avantajlara sahiptir [2]; i) Çok sayıda muhtemel korelasyon değişkeninin birlikte modellenebileceği durum ve kontrollerin sınıflandırılması için; ii) bir önem eşiği gerekli değildir; iii) gen-gen etkileşimleri değerlendirilebilir ve iv) normallik ve varyansların homojenliğine dayalı varsayımlar gerekli değildir [1].

Yöntem

Genom-Boyu İlişki Analizi (GWAS)

GWAS belirli bir hastalığa neden olduğu düşünülen genleri ve genom bölgelerini belirlemek için, hastalık durumu ve genetik varyasyon arasındaki ilişkiyi inceleyen çalışmalardır. Yola çıkarken belirli bir hipotez yoktur, tüm genom incelenerek hastalıkla ilişkili olduğu düşünülen bölgeler tespit edilir. Aynı anda tüm kromozomlara dolayısıyla da birçok gene bakılarak Gen-gen ve Gen-Çevre etkileşimleri incelenir [2, 3]. GWAS uygulamaları esnasında dosya türleri önem arz etmektedir. Genel olarak kullanılan dosya uzantıları ve yapıları; .bed= Genotip, .bim= Bytessence InstallMaker; genel olarak proje haritalama (bilgileri depolayan) .fam= Fenotip dosyalarıdır. GWAS'ı i) Verilerin Temizlenmesi (Kalite Kongrol), ii) Hastalıkla İlişki Tespiti İçin Model Oluşturma ve iii) Sonuçların Manhattan Plot ve QQ plot Yardmıyla Gösterimi olarak üç ana başlıkta toplamak mümkündür.

Makine Öğrenimi Yöntemleri

Random Forest

2001 yılında Breiman tarafından bulunmuştur. Oluşturulan birçok rastgele karar ağacının kombinasyonudur. Ağaçlar inşa edilirken bootstrap örneklem seçimi kullanılır [4]. Sınıflamada Gini katsayısı kullanılır ve oluşturulan ağaçlar için budama yapılmaz. Fazla sayıda değişken içeren veri setleri için de kullanılabilir yöntemdir, ancak; iyi donanıma sahip yüksek geçici belleği olan bilgisayarlarda kullanılması gerekliliği bir dezavantaj olarak görülebilir [5].

C5 Algoritması

1994 yılında Quinlan tarafından C4.5 (J48) algoritmasının daha hızlı versiyonudur. Daha etkin bellek kullanabilen ve daha küçük karar ağaçları üretir. Winnowing özelliği ile budama hatalarını azaltır. Aşırı öğrenme (over-fitting) probleminin üstesinden gelebilen bir algoritmadır. Linux/Unix için tasarlanmış See5 algoritmasının Windows için geliştirilmiş halidir [6].

Destek Vektör Makineleri

Destek vektör makineleri, ilk olarak 1992 yılında Vladmir Vapnik ve ark. tarafından sunulmuştur. Bir düzlem üzerinde bulunan gruplar arasında sınır çizerek grupları birbirlerinden ayırmak mümkündür. Sınırın çizileceği yer iki grubun da üyelerine en uzak olan yer olmalıdır. Destek vektör makineleri bu sınırın nasıl çizileceğini belirler [7].

ULSAM Veri Seti

ULSAM (The Uppsala Longitudinal Study of Adult Men), İsveç'in Uppsala bölgesindeki sağlıklı yaşlı erkeklerin araştırılmasıdır (Lithell ve ark. 2000). 1970'te Uppsala'da yaşayan 50 yaşındaki erkeklerin tümü ile başlatılmıştır ve belirli aralıkla ölçümlere devam edilmiştir. Katılımcılara 70 yaşındayken insülin direncini ölçmek için bir glikoz tolerans testi ve insülin kelepçesi verilmiştir ve tip-II diyabet vaka durumu, doktor tanılı hastalık veya tam açlık şekeri> 6.1mmol /l olarak belirlenerek, bu esaslara uygun şekilde hasta ve sağlıklı bireylerin ayırımı yapılmıştır. ULSAM kohortu, Illumina 2.5 milyon HumanOmni SNP BeadChip ve Illumina CardioMetabochip ile genotiplendirildi. Avrupa kökenli olmayan atalar dışlandı. SNP kalite kontrol için; HWE popülasyon dengesi için p <10e-6 değeri seçildi. MAF <%1 olarak belirlendi. Linkage-Disequilibrium (LD) incelenmesi ile budama yapılarak; çok boyutlu ölçeklendirme (MDS) ile popülasyon yapısını ayarlamak üzere temel bileşenler elde edildi. Üzerinde çalıştığımız veri seti, 1116 örnek ve 692483 SNP'den oluşmaktadır.

Makine Öğrenimi İçin Veri Hazırlığı

GWAS verilerinde farklı yöntemler kullanarak makine öğrenimi için değişken seçimi yapılabilir (feature selection) [8]. Univariate değişken seçimi ile p değeri 0.25 altında kalan 33656 SNP seçerek, makine öğrenimi yöntemlerini uyguladık [9]. PLINK programının clumping yönteminde farklı parametreler kullanarak 2611 SNP seçerek, makine öğrenimi yöntemlerini uyguladık [10, 11]. Bu çalışmada p1 için 0.01, p2 için 0.001, r2 için 0.2 ve kb için ise 10 parametrelerini kullandık. Tüm makine öğrenimi yöntemleri için 10 katlı çapraz geçerlilik yapılıaraki veri seti %30 test %70 eğitim olarak kullanılıp, eğitimde kullanılan veriler test setinde kullanılmamıştır.

Bulgular

Hasta ve Sağlıklı bireyler arasında; yaş istatistiksel olarak anlamlı farka sahip değilken (p=0.409), vücut kitle indeksi (VKI) istatistiksel olarak anlamlı farka sahiptir (p<0.001) [Tablo1].

Gruplar / Değişkenler	Sağlıklı (N=951) Ort±St.Sap.	Hasta (N=165) Ort±St.Sap.	t	p Değeri
VKI	25,97±3,22	27,93±3,94	-6.035	<0.001
Yaş	70,99±0,63	71,03±0,66	-0.826	0.409

Ort.= Ortalam, St.Sap.= Standart Sapma, t= Bağımsız İki Grup t-testi İstatistiği

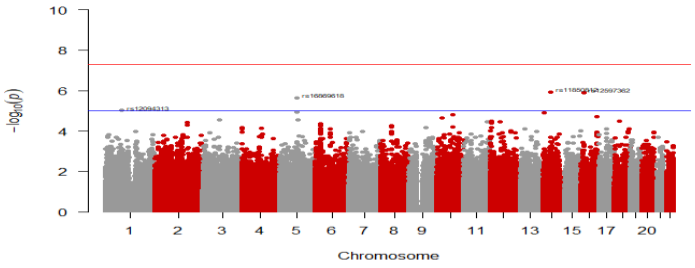
GWAS Bulguları

Lojistik regresyon sonuçlarından; ***rs11850812*** (OR: 3.64, p-değeri: 1.23E-06, kromozom 14), ***rs12597362*** (OR: 3.44, p değeri: 1.28E-06, kromozom 16), ***rs16869618*** (OR: 3.09, p-değeri: 2.34E-06, kromozom 5), ***rs12094313*** (OR:4.29, p-değeri: 9.68E-06, kromozom 1), ***rs11163137*** (OR: 4.29, p-değeri: 9.68E-06, koromozom 1), ***rs12520306*** (OR: 2.73, p-değeri: 1.14E-05, kromozom 1) tip II diyabet ile yüksek oranda ilişkilidir (Tablo 2).

Tablo 2: GWAS Sonuçlarına Göre Tip II Diyabet ile İlişkili İlk 5 SNP (Anlamlılık Düzeyi p <10e-5)

CHR	SNP	A1	OR	SE	L95	U95	Wald	P
14	rs11850812	T	3,646	0,267	2,162	6,148	4,851	1,23E-06
16	rs12597362	A	3,444	0,255	2,088	5,682	4,842	1,28E-06
5	rs16869618	C	3,093	0,239	1,936	4,942	4,722	2,34E-06
1	rs12094313	G	4,297	0,330	2,253	8,198	4,424	9,68E-06
1	rs11163137	A	4,297	0,330	2,253	8,198	4,424	9,68E-06

CHR: Kromozom Numarası, SNP: Tek nükleotid Polimorfizmi, A1: Minor Allele İsmi, Wald: Lojistik Regresyon Additive Model, OR: Odds Ratio, SE: OR için Standart Hata, L95- U95: OR Güven Aralıkları



Grafik 1: Manhattan Plot x-ekseni kromozom numaralı, y-ekseni -log(p) değerlerini ifade eder. Kromozom 1, 5, 14, 16 tip II diyabetle ilişkilidir.

Makine Öğrenimi Bulguları

Üç farklı makine öğrenme yöntemi arasında, ki-kare testi kullanarak SNP seçimi yapıldığında en iyi sınıflandırma algoritması, DVM (radyan kernel kullanılarak %99 doğruluk sınıflandırma oranı (accuracy): 0.99, Sensitivite: 0.95, Spesifite: 0.95) olduğu dikkat çekmektedir (Tablo 3-4). RF ve C5 doğru sınıflandırma oranlarının eşit olduğu görülmektedir (doğruluk sınıflandırma oranı (accuracy): 0.85) Üç makine öğrenimi yöntemi için de Sensitivite değerlerinin birbirlerine yakın ve 0.90'ın üzerinde olduğu görülmüştür.

Sonuç

Farklı özellik seçim yöntemlerine sahip ML modelleri seçerken ve kullanırken araştırmacı, ilgili hastalığın etkisini dikkate almalıdır. Modellerin özgüllüğü ve hassasiyeti farklı hastalık için farklı sonuçlar gösterebilmektedir. Bu çalışmada, tip II diyabetle ilişkili olabileceği düşünülen SNP'ler belirlenmiş ve de ML yöntemleri arasında da DVM'nin en iyi sınıflama sonuçlarına sahip olduğu gösterilmiştir. Bu çalışmanın sonuçları Ki-kare testinin clumping yöntemime iyi bir alternatif olarak değişken seçiminde kullanılabileceğini düşündürtmektedir. Hem klasik istatistiksel yaklaşımlar hem de ML yaklaşımları 14, 16, 5 ve 1 kromozomlar içerisinde bulunan tip-II diyabet ile ilişkili SNP'leri açığa çıkartmıştır. ML yöntemleri GWAS verileri için validasyon yöntemi olarak düşünülebileceği sonucuna da ulaşılmıştır.

Tablo 3 Makine Öğrenimi Sonuçları (33656 SNP)

Sınıflama Yöntemi		Gerçek Sınıf		PPV*	NPV**	Sensitivite	Spesifite	Accuracy
Destek Vektör Makineleri		Vaka	Kontrol					
Tahmini Sınıf	Vaka	285	2	0,99	1,00	1,00	0,95	0,99
	Kontrol	0	47					
Random Forest		Vaka	Kontrol					
Tahmini Sınıf	Vaka	285	49	0,85	-	1	0,00	0,85
	Kontrol	0	0					
C5		Vaka	Kontrol					
Tahmini Sınıf	Vaka	285	48	0,85	1,00	1,00	0,02	0,85
	Kontrol	0	1					

Tablo 4. Makine Öğrenimi Sonuçları (2611 SNP)

Sınıflama Yöntemi		Gerçek Sınıf		PPV*	NPV**	Sensitivite	Spesifite	Accuracy
Destek Vektör Makinleri		Vaka	Kontrol					
Tahmini Sınıf	Vaka	284	0	1,00	0,98	0,99	1,00	0,99
	Kontrol	1	49					
Random Forest		Vaka	Kontrol					
Tahmini Sınıf	Vaka	285	49	0,85	--	1,00	0,00	0,85
	Kontrol	0	0					
C5		Vaka	Kontrol					
Tahmini Sınıf	Vaka	285	45	0,86	1,00	1,00	0,08	0,86
	Kontrol	0	4					

Kaynaklar

1. Fadista, J., Manning, A. K., Florez, J. C., & Groop, L. (2016). The (in) famous GWAS P-value threshold revisited and updated for low-frequency variants. European Journal of Human Genetics, 24(8), 1202-5. doi: 10.1038/ejhg.2015.269. Epub 2016 Jan 6.

2. Coşgun, E., Kaya, A., Keçeli, A. S., Aksarı, Y., & Karaağaoğlu, E. Gene 3E: Genetik Veri Madenciliği için Yeni Analiz Aracı.

3. Battaloğlu, E., & Başak, A. N. (2010). Kompleks Hastalık Genetiği: Güncel Kavramlar ve Nörolojik Hastalıkların Tanısında Kullanılan Genomik Yöntemler. *Klinik Gelişim Dergisi*, *23*, 128-33.

4. L. Breiman. Random forests. Machine Learning, 45:5–32, 2001.

5. Chen, X., & Ishwaran, H. (2012). Random forests for genomic data analysis. *Genomics*, *99*(6), 323-329.

6. Pandya R., Pandya J., "C5.0 Algorithm to Improved Decision Tree with Feature Selection and Reduced Error Pruning", International Journal of Computer Applications (0975 – 8887), Volume 117 – No. 16, 2015

7. Şeker S.E.," İş Zekası Ve Veri Madenciliği",Cinius Yayınları, 1.Baskı , İstanbul, 2013, ISBN: 978-605-127-671-7

8. Xing, E. P., Jordan, M. I., & Karp, R. M. (2001, June). Feature selection for high-dimensional genomic microarray data. In *ICML*(Vol. 1, pp. 601-608)

9. Alpar, R. (2011). Çok Değişkenli İstatistiksel Yöntemler, Ankara: Detay Yayıncılık

10. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., ... & Sham, P. C. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics*, *81*(3), 559-575.

11. Draisma, H. H., Pool, R., Kobl, M., Jansen, R., Petersen, A. K., & Vaarhorst, A. A. *Genome-wide association study identifies novel genetic variants contributing to variation in blood metabolite levels. Nat Commun. 2015; 6: 7208*. Epub 2015/06/13. https://doi. org/10.1038/ncomms8208 PMID: 26068415.

T9

Türkiye'de Bebek Ölüm Riskinin Mekan-zamansal Bayesci Modeller ile İncelenmesi

Sade KILIÇ YILDIRIM[1], Celal Reha ALPAR[2] , Erdem KARABULUT[2] , Atilla Halil ELHAN[3]

[1]Türkiye İstatistik Kurumu, Ankara
[2]Hacettepe Üniversitesi Tıp Fakültesi Biyoistatistik Anabilim Dalı, Ankara
[2]Ankara Üniversitesi Tıp Fakültesi Biyoistatistik Anabilim Dalı, Ankara

Amaç: Bu çalışmada 2009 ve 2017 yılları arasında Türkiye'nin her ilinde bebek ölüm riskini. mekan. zaman ve mekan-zaman etkileşimi kavramlarını içererek belirlemek, risk haritaları elde etmek ve etkileyen risk faktörlerini incelemek amaçlanmıştır.

Yöntem:Bebek ölümlerinin sayısının Türkiye'de 2009 ve 2017 yılları arasında her il için Poisson dağılımına sahip olduğu varsayılmaktadır. Bağlantı fonksiyonu (log) ile dağılımın bilinmeyen parametresi bebek ölüm riski, spatio-temporal Bayesian modelleriyle modellenmiştir. Modeller, R yazılımında INLA (entegre iç içe Laplace yaklaşımları) paketi ile uygulanır. En iyi model sapma bilgi kriterine (DIC) göre belirlenir. İlin kişi başına gayri safi yurtiçi hasılası ve anne yaşının bebek ölüm riski üzerindeki etkileri, en iyi mekan-zamansal model ile ve mekan ve zaman kavramı olmadan doğrusal model ile incelenmiştir.

Bulgular: 2009'dan 2017'ye kadar, yapılandırılmamış mekansal ve yapılandırılmış zamansal etkileşim rastgele etkisine sahip en iyi model için, 1'den büyük riski olan il sayısı azalmıştır. İlin kişi başına gayri safi yurtiçi hasılası doğrusal model (RR (Göreli risk) = 0.881;% 95 Cl (güvenirlik aralığı) (0.875; 0.887)) ve mekan-zamansal model (RR = 0.907;% 95 Cl (0.841; 0.988) için bebek ölümü riski üzerinde azaltıcı sahiptir. Doğrusal model için anne yaşının 20 yaşından küçük olması (RR = 1.021;% 95 Cl (1.018; 1.024) ve 40 yaşından büyük (RR = 1.08;% 95 Cl (1.073; 1.087)) olması riski üzerinde artırıcı etkiye sahiptir, ancak mekan-zamansal model için risk üzerinde bir etkisi yoktur. Risk faktörlerinin bulunduğu zaman-mekansal model ve doğrusal modelin sapma bilgi kriteri (DIC) sırasıyla 5753.27 ve 8558.05'tir. Mekansal ve zamansal etkileşim etkisinin ve zamansal etkinin, mekansal-zamansal model tarafından açıklanan riskteki değişkenliğine olan katkısı sırasıyla 52.61 % and 29.41 % dir , modele risk faktörlerinin eklenmesiyle bu etkilerin katkısı sırasıyla 59.70 % ve 14.89% dir.

Sonuç: Zamansal etkinin, risk faktörleri eklenen mekan-zamansal model ile modellenen riskteki değişkenliğe katkısı, risk faktörü olmayan modele göre belirgin şekilde azalır, ancak etkileşim etkisi artar. Mekan-zamansal modelde anne yaşı risk üzerinde etkili değildir, ancak sapma bilgi kriteri (DIC) nedeniyle doğrusal modelden daha çok tercih edilir. Bu nedenle bebek ölüm riski yalnızca etkileyen faktörler açısından incelenmemeli, etkileyen risk faktörleri ile birlikte zaman, mekan ve zaman-mekan etkileşiminide göz önünde bulundurmak çok önemlidir.

Anahtar Kelime: entegre iç içe Laplace yaklaşımları, mekan-zamansal Bayesci model, bebek ölüm riski.

1.GENEL BİLGİLER

Mekan-zamansal modeller sağlık alanındaki (ölüm, insidans, intihar vb.) Risk değerlendirmesinde kullanılmaktadır. Bu modeller, mekanın etkisini, zamanın etkisini ve belirli faktörlerin risk üzerindeki etkilerini belirlemek ve riski tahmin etmek için mekam-zaman

etkileşimlerinin etkisini içerir. Coğrafi konum ve zamanı dikkate almak önemlidir. Sağlık alanındaki riskin belirlenmesinde; birbirine yakın alanlarda benzer riskler olabilir. Aynı yapı zaman içinde geçerli olabilir. Risk haritaları mekansal ve mekansal-zamansal risk farklarını ortaya çıkarmak için karşılaştırılmıştır .

Alan ve zaman içeren verilere erişebilmek, mekansal-zamansal modellerin verilere uygulanmasını sağlar. Verilerdeki mekansal, zamansal ve mekansal-zamansal yapılar modeldeki Gaussian Markov rastgele alanları ile birlikte verilmiştir. Tahminlerdeki belirsizlik Bayesci yaklaşımlar tarafından dikkate alındığında, bu modellere Bayesian mekan--zamansal denir.

Sparks (2015), ABD'de etnik gruplar arasındaki 2000-2008 yılları arasındaki sindirim ve solunum kanseri görülme riskindeki eşitsizlikleri Bayesci hiyerarşik mekan-zamansal modeller ile incelemiştir. Tüm modeller, bu eşitsizliklere neden olabilecek sağlık hizmetlerine farklı erişim, sosyoekonomik durumdaki farklılıklar ve çalışma alanı gibi risk faktörleri de dikkate alınarak INLA (Integrated Nested Laplace Approximation) yöntemi ile uygulanmıştır. Genel olarak, Hispaniklerin Teksas'taki popülasyonu için bu kanserlerin her ikisinde de önemli bir dezavantajı olduğu ve bu eşitsizlik mekânsal açıdan büyük ölçüde farklılık gösterdiği belirlenmiştir. Sindirim ve solunum kanserlerinde Hispaniklerle Hispanik olmayanlar arasındaki en büyük eşitsizlik, Teksas'ın doğusunda ortaya çıktığı ve bu yapının 2000'de ortaya çıkıp 2008 yılına kadar devam ettiği saptanmıştır (1). Buna ek olarak Brezilya'da 2000-2016 yılları arasında insan leptospiroz hastalığına ait morbidite ve mortalite riski, Santos Baquero ve ark. (2018) tarafından mekan-zamansal açıdan incelenmiştir. Bayesian mekan-zamansal modeller INLA yöntemi ile uygulanmıştır. Sonuç olarak; morbidite riskinde ana etkinin mekânsal etki ile olduğu, mortalite riskinde ise mekan-zamansal etkileşiminin ana etki olduğu belirlenmiştir. Toprak nemi, yağış, yoksulluk ve kentsel hanehalkı oranındaki azalma, risk faktörü olarak modele alınarak incelenmiştir (2). Martínez-Bello ve ark. (2018) tarafından yapılan çalışmada ise Kolombiya'nın bir bölgesi ve bir şehrinde Bayesian mekan-zamansal etkileşim modelleri kullanılarak 2015-2016 zika salgını döneminde dengue (sivrisineklerin bulaştırdığı viral bir hastalık) ve Zika virüsü hastalığının (ZVD) paralel göreceli riski tahmin edilmiştir. Modelleme de epidemiyolojik hafta zaman ölçüsü olarak kullanılıp INLA yöntemi uygulanmıştır (3).

2. YÖNTEM

Türkiye' de her il (i=(1,2,….,81)) ve yıl (t=2009, 2010,……., 2017) için canlı doğan bebek sayısı R_{it} olmak üzere her il ve yıl için beklenen bebek ölüm sayısı E_{it}

$$E_{it} = R_{it}\frac{\sum_{it} y_{it}}{\sum_{it} R_{it}}$$

olarak ifade edilmektedir.

Türkiye'de her il ve yıl için bebek ölüm sayısının y_{it} Poisson dağılımı gösterdiği varsayılmaktadır.

$$y_{it}|\theta_{it} \sim Poisson(E_{it}\theta_{it})$$

Burada her il ve yıl için bebek ölüm riski (θ_{it}); Poisson dağılımının bilinmeyen parametresidir. Log bağlantı fonksiyonu ile ifade edilen bebek ölüm riski; $\log(\theta_{it})$; mekan-

zamansal Bayesci modeller kullanılarak tahmin edilmeye çalışılacaktır. Parametrik ve parametrik olmayan mekan-zamansal modeller R yazılımında INLA(paketi ile uygulanır.

Gizli Gaussian model (Latent Gaussian model)

Gözlemlerin $\{y_1, \ldots\ldots, y_n\}$; y veri kümesinde bulunduğu varsayılsın. Gözlemin (y_i) (i = 1,..., n), üssel ailenin (Gauss, gama, üssel, weibull, cox, binom, Poisson) dağılımlarından birine sahip olduğu varsayıldığında; Belirli bir bağlantı(link) fonksiyonu ile dağılımın bilinmeyen parametresi (η_i) , zamansal rastgele etki, mekansal rasgele etki, mekan -zamansal rastgele etkiler dahil edilerek yapılandırılmış toplamsal regresyon modeli ile aşağıdaki gibi modellenmiştir.

$$\eta_i = \beta_0 + \sum_{m=1}^{M} \beta_m x_{mi} + \sum_{l=1}^{L} f_l(z_{li}) + \epsilon_i$$

Yapısal toplamsal regresyon modellerinde bu mekansal, zamansal ve mekan-zamansal yapıların modellenmesinde gizli gaussian markov rastgele alanları kullanıldığında Gizli Gaussian modeller elde edilir. f_l; doğrusal olmayan etkileri(zamansal, mekansal, zamansal mekansal etkileşim etkisi), β_i; x_m değişkenine ait doğrusal etkiyi, β_0 sabit olmak üzere bu bileşenler $x = \{\eta_i, \beta_0, \beta_m, f_l\}$ gizli rasgele alanı altında toplanır. x in her bir elemanına Gaussian önselleri atanarak gizli Gauss modeli elde edilmektedir.

Gizli Gauss modelleri üç aşamalı hiyerarşik bir model olarak kabul edilir:

Gözlenen veriler (y): θ; bu gizli sürecin parametresi olan hiperparametre olarak adlandırılır. Gözlenen veri kümesinin $y = \{y_1, \ldots\ldots, y_n\}$ dağılımı; her gözlemin (y_i) her Gaussian Markov Rastgele alanı (x_i) ve hiperparametre (θ_i)verildiğinde koşullu olarak bağımsız olduğu varsayıma göre hesaplanır. Gözlenen veri setinin dağılımı olabilirlik fonksiyonu ile aşağıdaki gibi ifade edilir:

$$\pi(y\,|x, \theta_1\,) = \prod \pi(y_i|x_i, \theta_1)$$

Gauss Markov Rastgele Alanı (GMRF) (x): Gaussian Markov rastgele alanı x, ortalaması μ ve kovaryansı (ters pericision matrisi) $\Sigma = Q^{-1}$ olan bir Gauss (Normal) dağılımına sahip rastgele bir vektördür.

$$x|\theta_2 \sim \mathcal{N}(\mu, Q^{-1}(\theta_2))$$

x in dağılımının sıfır ortalama ve $Q^{-1}(\theta_2)$ kovaryans ile Gauss olduğu varsayılmıştır. x in her bir bileşeninin kesinlik(precision) matrisi, bileşenin komşuluk yapısına ve bilinmeyen $1/\sigma^2$ ye bağlıdır.

Hiperparametreler (θ): Hiperparametrelere önseller atanır..

$$\theta \sim \pi\,(\theta)$$

Entegre İç içe Laplace Yaklaşımı (INLA)

Gauss Markov rastgele alanının ve hiperparametrelerin ortak posterior dağılımı şu şekilde hesaplanır :

$$\pi(x,\theta|y)\ \alpha\ \ \pi(x,\theta)\,\pi(y|x,\theta)\alpha\ \ \pi(x|\theta)\pi(\theta)\ \prod_{i=1}^{n}\pi(y_i|x_i,\theta)$$

$$\alpha\,\pi(\theta)\ |Q(\theta)|^{n/2}exp\left[-\frac{1}{2}x^T Q\ (\theta)\ x+\sum_{i=1}^{n}\log\{\,\pi(y_i|x_i,\theta)\right]$$

Aşağıdaki integrallerle hiperparametrelerin ve Gaussian Markov Random alanının posterior dağılımları hesaplanmaktadır.

$$\pi(x_i|y)=\int\pi\ (x_i|\theta,y)\,\pi\ (\theta|y)\,d\theta\quad i=1,2,\dots.,n$$
$$\pi\big(\theta_j|y\big)=\int\pi\ (\theta|y)\,d\theta_{-j}\quad j=1,2,\dots.,m$$

Gauss Markov Rastgele alanının posterior dağılımı, INLA tarafından yaklaşık posterior dağılımı şu şekilde olur:

$$\tilde{\pi}(x_i|y)=\sum\tilde{\pi}\big(x_i\big|\theta^{(k)},y\big)\ \tilde{\pi}\ \big(\theta^{(k)}\big|y\big)\ \triangle\theta^{(k)}$$

INLA yöntemi yaklaşık posterior dağılımları elde etmek için iki ana bölümde incelenmiştir:

- Yaklaşık $\pi(\theta|y)\big(\tilde{\pi}(\theta|y)\big)$ elde edilir $\tilde{\pi}(\theta|y)$ iyi $\theta^{(k)}$ noktaları bulmak için incelenir ve alan ağırlıklandırmaları $\triangle\theta^{(k)}$ hesaplanır.
- Yaklaşık $\pi(x|\theta,y)\ \big(\tilde{\pi}(x|\theta,y)\big)$ elde edilir ve iyi noktalar $\theta^{(k)}$ için $\big(\tilde{\pi}(x|\theta,y)\big)$değerleri hesaplanır.

Mekan-zamansal modeller:

u_i; yapılandırılmış mekan etkisi, v_i; yapılandırılmamış mekan etkisi olmak üzere mekan-zamansal modeller parametrik ve parametrik olmayan olmak üzere iki başlık altında incelenecektir. Mekansal yapısal etki ve mekansal yapılandırılmamış etki olarak iki rastgele etki içeren BYM (Besag-York-Mollie) modeliyle incelenecektir. Yapılandırılmış mekansal tesadüf etkisi, ilin komşu yapısına dayanarak belirlenir. Alanlar arasındaki heterojenite etkileri, yapılandırılmamış mekansal rastgele etki ile yansıtılmaktadır

Parametrik model:

Model1: Tüm alanları kapsayan genel zaman eğilimi (β) ile her alana özel eğilim ile genel eğilim arasındaki farkı belirleyen (δ_i) yı içerir.Eğer (δ_i) ; 0 dan küçük ise genel eğilim alana özel eğilimin altında; 0 dan büyük ise genel eğilim alana özel eğilimin üstündedir.

$$log(\theta_{it})=b_0+u_i+v_i+(\beta+\delta_i)*t$$

Parametrik olmayan modeller: Yapılandırılmış γ_t ve yapılandırılmamış ϕ_t zaman etkisi içerir. Yapılandırılmış zaman etkisi rastgele yürüyüş (random walk) modeli 1 ve rastgele yürüyüş modeli 2 ile modelenebilmektedir.

Model 2:

$$log(\theta_{it}) = b_0 + u_i + v_i + \gamma_t + \phi_t$$

Model 3-4-5-6: Model 2 ' e zaman mekan etkileşiminin δ_{it} eklenmesi ile aşağıdaki gibi ifade edilir.

$$log(\theta_{it}) = b_0 + u_i + v_i + \gamma_t + \phi_t + \delta_{it}$$

Model 3: Yapılandırılmamış mekan ve zaman etkileşimini içermektedir.

Model 4: Yapılandırılmamış mekan ve yapılandırılmış zaman etkileşimini içermektedir.

Model 5: Yapılandırlmamış zaman ve yapılandırılmış mekan etkileşimini içermektedir.

Model 6: Yapılandırılmış mekan ve zaman etkileşimini içermektedir.

En iyi model sapma bilgi kriterine (DIC) göre belirlenir, Mekansal, zamansal ,mekan ve zaman etkileşiminin etkilerine ait sonsal dağılımlar bu modelden elde edilir.

$$log(\theta_{it}) = b_0 + u_i + v_i + \gamma_t + \phi_t + \delta_{it} + b_1(kişibaşıgsyh/10000)_{it+}b_2(20\ yaş\ altı\ anne\ yüzde)_{it} + b_3(40\ yaş\ üstü\ anne\ yüzde)_{it}$$

Annelerin yaşı, il başına düşen gayri safi yurtiçi hasıla en iyi modelde eklenerek bebek ölüm riski üzerindeki etkileri belirlenmiştir. Anne yaşı, 20 yaşından küçük, 40 yaşından büyük annelerin yüzdesine göre iki değişkenle incelenmiştir..

$$log(\theta_{it}) = b_0 + u_i + v_i + \gamma_t + \phi_t + \delta_{it} + b_1(kişibaşıgsyh/10000)_{it+}b_2(20\ yaş\ altı\ anne\ yüzde)_{it} + b_3(40\ yaş\ üstü\ anne\ yüzde)_{it}$$

Diğer taraftan anne yaşı, ilin gayri safi yurtiçi hasıla bebek ölüm oranı riskine etkisi mekansal, zamansal ve mekansal ve zamansal etki etkileşimi olmadan doğrusal model ile değerlendirilmektedir.

$$log(\theta_{it}) = b_0 + b_1(kişibaşıgsyh/10000)_{it+}b_2(20\ yaş\ altı\ anne\ yüzde)_{it} + b_3(40\ yaş\ üstü\ anne\ yüzde)_{it}$$

3.BULGULAR

En düşük sapma bilgi kriterine sahip, yapılandırılmamış mekansal ve yapılandırılmış zaman (rastgele yürüyüş 1) rastgele etki etkileşimi etkisi olan model, en iyi modeldir.

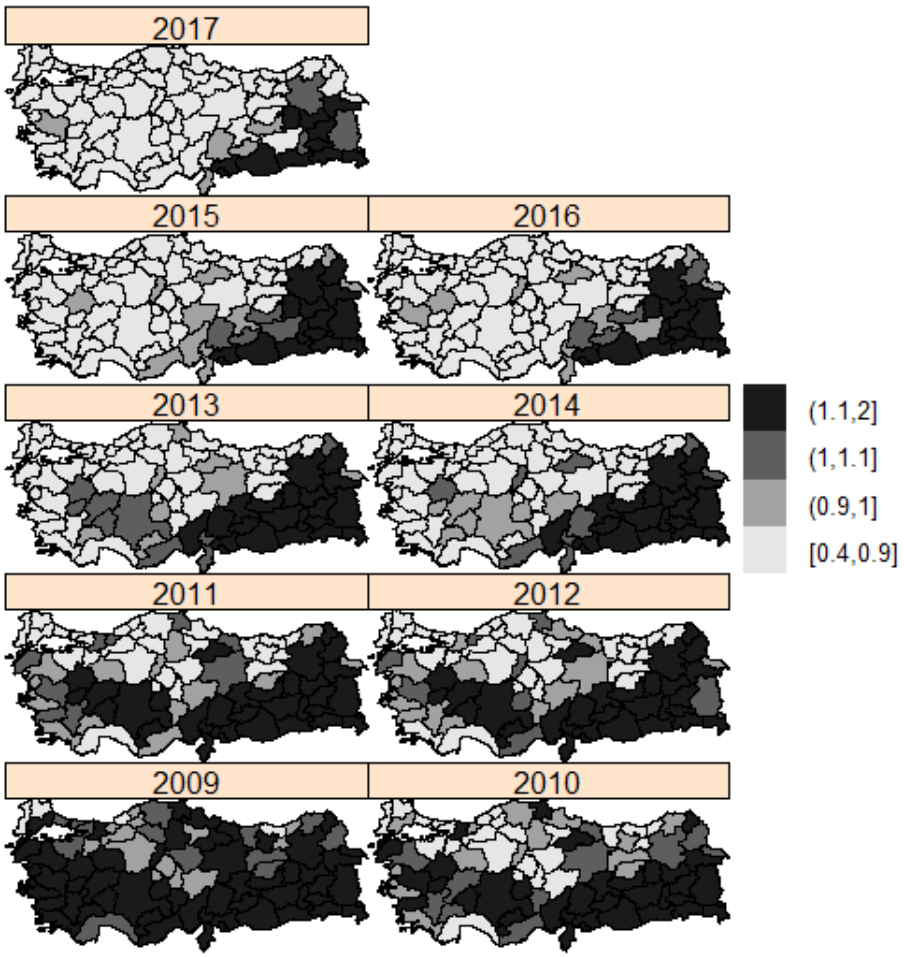
Model	Sapma Bilgi Kriteri (DIC)	
Model 1	6324,65	
	Rastgele yürüyüş1	Rastgele yürüyüş 2
Model 2	6655,98	6656,07
Model 3	5772,6	5772,79
Model 4	5748,76	5848,18

Model 5	5768,98	5768,99
Model 6	5755.28	5827.76

En iyi model tarafından açıklanan mekan-zamansal etkinin bebek ölüm riskindeki değişkenliğe yüzde katkısı% 52.61'dir. % 29.1 ile; yapılandırılmış zaman etkisi mekansal-zamansal etkileşim izler.

Rasgele etki	*Yüzde*
Yapılandırılmamış mekan etkisi	8,01
Yapılandırılmış mekan etkisi	9,39
Yapılandırılmış zaman etkisi	29,41
Yapılandırılmamış zaman etkisi	1,03
Yapılandırlmamış mekan yapılandırlmış zaman etkileşimi etkisi	52,16

İllerin çoğu 2009'da 1'in üzerinde riske sahiptir.. 2009'dan 2017'ye kadar 1'in üzerinde riski olan illerin sayısı azalmıştır. Ölüm riski 1'in üzerinde olan iller. 2017'de doğu ve güneydoğu Anadolu bölgelerinde kümelenmiştir.



Doğrusal modelin sapma bilgi kriteri 8558.05'tir. Doğrusal model için kişi başına düşen ilin gayri safi yurtiçi hasıla bebek ölüm riskini azaltıcı yönde etkilemektedir. Kişi başına gayri safi milli hasılada her 1000 TL'lik artış. bebek ölüm riskini yüzde 12 azaltmaktadır. Yaşları 20 yaşın altındaki annelerin yüzdesi ve yaşları 40 yaşın altındaki annelerin bebek ölümü riski üzerinde artırıcı etkisi vardır. 20 yaşından küçük annelerin yüzdesindeki her bir birim artış bebek ölüm riskini yüzde 2 artırıyor. 40 yaşından büyük annelerin yüzdesindeki her bir birim artış bebek ölüm riskini yüzde 8 artırıyor.

	ortalama	Standart sapma	RR	95 % Cl
gsonbin	-0,127	0,003	0,881	(0,875;0,887)
X20altyü	0,021	0,002	1,021	(1,018;1,024)
X40üstyü	0,077	0,003	1,08	(1,074;1,087)

Risk faktörleri eklenmiş mekan-zamansal modelin sapma bilgi kriteri 5753.27'dir. İlin kişi başına gayri safi milli hasılada her 1000 TL'lik artış bebek ölüm riski üzerinde azaltıcı etki

yaratmaktadır. Kişi başına gayri safi yurtiçi hasılada 1000 TL'lik bir artış. bebek ölüm riskini yüzde 9 azaltır. Yaşları 20'den küçük olan annelerin yüzdesi (RR = 1.014 ((0.989; 1.039). 40 yaşın üzerindeki annelerin yüzdesi (RR = 1.002 (0.954; 1.051) bebek ölüm riskini etkilememektedir.

Parameter	Ortalama	Standart sapma	RR	95 % Cl
gsonbin	-0,097	0,041	0,907	(0,841;0,988)
X20altyü	0,014	0,013	1,014	(0,989;1,039)
X40üstyü	0,002	0,025	1,002	(0,954;1,051)

Risk faktörleri eklenmiş mekan-zamansal model tarafından açıklanan bebek ölüm riskindeki değişkenliğe olan mekan zaman etkileşim etkisinin katkısı% 59.70'tir. Bunu % 14.89 yapılandırılmış zaman etkisi izler.

Rasgele etki	*Yüzde*
Yapılandırılmamış mekan etkisi	11,24
Yapılandırılmış mekan etkisi	11,38
Yapılandırılmış zaman etkisi	14,89
Yapılandırılmamış zaman etkisi	2,77
Yapılandırlmamış mekan yapılandırlmış zaman etkileşimi etkisi	59,70

4. SONUÇ

Zamansal etkinin. risk faktörleri eklenen mekan-zamansal model ile modellenen riskteki değişkenliğe katkısı. risk faktörü olmayan modele göre belirgin şekilde azalır. ancak etkileşim etkisi artar.

Mekan-zamansal modelde anne yaşı risk üzerinde etkili değildir. ancak sapma bilgi kriteri (DIC) nedeniyle doğrusal modelden daha çok tercih edilir.

Bu nedenle bebek ölüm riski yalnızca etkileyen faktörler açısından incelenmemeli. etkileyen risk faktörleri ile birlikte zaman. mekan ve zaman-mekan etkileşiminide göz önünde bulundurmak çok önemlidir.

KAYNAKLAR

1. Shaddick G.. James V. Zidek JV. Spatio-Temporal Methods in Environmental Epidemiology. Boca Raton. FL: CRC Press Taylor & Francis Group. LLC; 2016.

2. Blangiardo M.. Michela C. Spatial and Spatio-temporal Bayesian Models with R-INLA. United Kingdom: John Wiley & Sons. Ltd; 2015.

3. Blangiardo M.. Cameletti M.. Baio G.. Rue H. Spatial and Spatio-Temporal models with R-INLA. Spatial and Spatio-temporal Epidemiology. 2013;7:39–55

4. Martino S. H. Rue. Case Studies in Bayesian Computation using INLA [Internet]. 2011[Erişim Tarihi 26 Mart 2018]. Erişim adresi:https://inla.r-inla-download.org/r-inla.org/papers/martino-rue-book.pdf

5. Gelman A.. Carlin JB.. Stern HS.. Dunson DB.. Vehtari A.. Rubin DB. Bayesian Data Analysis.3 rd. ed. Boca Raton. FL: CRC Press Taylor & Francis Group. LLC; 2015.

6. Ambrosius WT . Topics in Biostatistics . New Jersey: Humana Press Inc.;2007.

7. Rue H. Riebler A.. Sorbye S.. Illian JB.. Simpson DP.. Lindgren FK. Bayesian Computing with INLA: A Review [Internet]. 2016 [Erişim tarihi 30 Eylül 2017]. Erişim adresi: https://arxiv.org/pdf/1604.00860.pdf

8. R Havard. Martino S. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. Journal of the Royal Statistical Society. B. 2009; 71.Part 2:319-392

9. Rue H.. Held L. Gaussian Markov Random Fields Theory and Applications. Boca Raton. FL: Chapman &Hall/CRC; 2005.

10. Havard Rue. Sara Martino. Approximate Bayesian Inference for Hierarchical Gaussian Markov Random Fields Models[Internet].2006 [Erişim tarihi 14 Nisan 2018]. Erişim adresi:https://pdfs.semanticscholar.org/8fe6/6217f4b69b86f63b04be494f3828c8b27c9f.pdf

11. Maiti T. Hierarchical Bayes estimation of mortality rates for disease mapping. Journal of Statistical Planing and Inference. 1998; 69: 339-348.

12. Meza JL. Empirical Bayes estimation smoothing of relative risks in disease mapping. Journal of Statistical Planing and Inference.2003;112: 43-62.

13. Bakka H.. Rue H.. Fuglstad GA.. Riebler A.. Bolin D.. Illian J. et al. Spatial modelling with R-INLA: A review [Internet].2018[Erişim tarihi 26 Şubat 2019]. Erişim adresi: https://arxiv.org/pdf/1802.06350.pdf

14. Langford IH.. Leyland AH.. Rasbash J.. Goldstein H. Multilevel Modelling of the Geographical Distributions of Diseases. Journal of Royal Statistical Society. Series C. 1999; 48: 253-268.

15. Rue H.. Martino S.. Chopin N. Approximate Bayesian Inference for Latent Gaussian Models Using Integrated Nested Laplace Approximations [Internet]. 2007 [Erişim tarihi 12 Nisan 2018]. Erişim adresi:
https://pdfs.semanticscholar.org/957e/9268da6c3638cb3a7613f2d7c4c77edac15a.pdf

16. Martino S. Approximate Bayesian Inference for Multivariate Stochastic Volatility Models[Internet].2007[Erişim tarihi 28 Mart 2019]. Erişim adresi:https://pdfs.semanticscholar.org/637c/9b705beb27c898775f71ea1aef39b7041d14.pdf?_ga=2.64934703.1034824228.1553421122-45219249.1553421121

17. Pietiläinen V. Approximations for Integration Over the Hyperparameters in Gaussian Processes.[Master of Science].Espoo:Aalto University; 2010.

18. Knorr-Held L. Bayesian Modelling of Inseparaable Space-Time Variation in Disease Risk[Internet].1999[Erişim tarihi 20 Kasım 2018].Erişim adresi: https://pdfs.semanticscholar.org/e444/80d0e7754bcc703c2d866a9b9c115d43e7de.pdf

19. Santos Baquero O.. Gustavo Machado. Spatiotemporal dynamics and risk factors for human Leptospirosis in Brazil. SCIENTIFIC REPORTS.(2018)8:15170 |DOI:10.1038/s41598-018-33381-3

20. Waruru A..Achia TNO.. Muttai H.. Ng'ang'a L.. Zielinski-Gutierrez E.. Ochanda B. et al. Spatial- temporal trend for mother-to child transmission of HIV up to infancy and during pre-Option B+ in western Kenya. 2007-2013. (2018) PeerJ 6:e4427; DOI 10.7717/peerj.4427

21. Helbich M.. Plener PL.. Hartung S.. Blüml V. Spatiotemporal Suicide Risk in Germany: A Longitudinal Study 2007-2011. SCIENTIFIC REPORTS. 7:7673 | DOI:10.1038/s41598-017-08117-4

22. Librero J.. Ibañez B.. Martinez-Lizaga N.. Peiro S.. Bernal-Delgadao E. et al. Applying spatio-temporal models to assess variations across health care areas and regions: Lessons from the decentralized Spanish National Health System. 2017 PLOS ONE | DOI:10.1371/journal.pone.0170480

23. Ma Y.. Zhang T.. Liu L.. Lv Q..Yin F.. Spatio-Temporal Pattern and Socio-Ekonomic Factors of Bacillary Dysentery at County Level in Sicuhan Province. China. SCIENTIFIC REPORTS(2015)| 5:15264 | DOI: 10.1038/srep15264
24. Sparks C. An examination of disparities in cancer incidence in Texas using Bayesian random coefficient models. (2015) PeerJ 3:e1283; DOI 10.7717/peerj.1

T10

Beyin EEG Ölçümlerini Bayesçi Sınıflandırma

Selim DÖNMEZ[1], Özer ÖZAYDIN[1]

[1]Osmangazi Üniversitesi Fen-Edebiyat Fakültesi, İstatistik Bölümü, Eskişehir

Özet:

Bu çalışmada, UCI(University of California Irvine) verisetlerinden epileptik nöbet verisinde epileptik nöbet geçirmediği halde incelenmiş sınıfların özgül yanları ortaya koyulmaya çalışılmıştır. Binom karma modelinde Bayesçi yöntemlerle parametre tahmin ederek, sözkonusu veriden azami verimlilik elde edilmeye çalışılmıştır. Medyan tarafından belirlenen veride 3 kategori kurduktan sonra medyan değişkeni türetilmiş ve lojistik regresyon modeli ile sınıflandırılmıştır. Bunun sonucunda çapraz tablo gösterdi ki veride tümör bulunan ve bulunmayan bölgeler arasında %97.2 başarı ile ayrım yapılırken, gözlerin çalışması sırasında elde edilen EEG ölçümlerinde %96.9'a varan bir başarı elde edilmiştir.

Anahtar Sözcükler: Karma Model, Bayesçi Yöntemler, Zaman Serileri, lojistik regresyon

Giriş

UCI verilerinden epilepsi nöbet tanısı, üzerinde çokça durulmuş bir veridir. 18 çalışmada kullanılan epilepsi verisi ilk defa Andrzejak vd.(2001) tarafından kullanılmıştır. Çalışmada dinamik beyin yapısının, bölgesel olarak nasıl çalıştığını gözlemlemek ve bölgeler arası, fizyolojik ve patolojik beyin durumları arasındaki farklılıkları araştırmak için elde edilmiştir. İlk elde edildiğinde veri, her biri 100 kişinin 23.5 saniyelik aralıklardan oluşan zaman serilerini içeren 5 dosya iken, Qiuyi Wu ve Ernest Fokoue tarafından UCI'ın veritabanına eklendiğinde 11500x178'lik boyutlara sahip her biri 1 saniyelik aralıklarla oluşan bir veriye dönüştürülmüştür. Andrzejak vd. (2001), bu veriyi analiz ederken doğrusal olmayan süreçlerin epilepsi nöbetlerinin güçlü belirteçlerinden biri olduğunu bulmuştur.

Kemal ve Güneş (2008), epilepsi verisini sınıflandırırıken özellik belirleme, boyut düşürme ve bulanık kümeleme yöntemlerinden oluşan bir metodoloji kullanmıştır. Çalışmada, belirtilen amaçlar için sırayla Welch yöntemi, temel bileşenler analizi ve bulanık altyapı yerleştirmesi kullanılmıştır. Bunun yanında çapraz onaylama ile sınıflandırma başarısı ölçülen metodoloji, deneme verisi olarak verinin %70'i ve %80'i kullanıldığında %100 oranında sınıflandırma yapmıştır.

Venema vd. (2006) çalışmasında 13 veri üzerinde iteratif olarak düzeltimiş Fourier dönüşümünü uygulamıştır. Bu algoritma aynı büyüklük dağılımına ve güç yörüngesine sahip yapay zaman serileri üretiyordu ancak yöntemin yeniliği, Fourier dönüşümünün farklı şekilde yapılmasıydı. Çalışılan 13 veriden epilepsi verisinde %100'e yakın başarı elde etmiştir.

Polat ve Güneş (2007), çalışmasında hibrit bir sistem kullanarak epileptik nöbet tespit etmeye çalışmışlardır. Bu sistemde karar ağacı sınıflayıcısını ve hızlı Fourier dönüştürücüsünü kullanırken, çapraz onaylama yöntemiyle %98.68 ve %98.72 oranında sınıflandırma gerçekleştirmişlerdir. Bu bahsedilenler haricinde çalışmalar için, http://epileptologie-bonn.de/cms/front_content.php?idcat=193&lang=3&changelang=3 adresine bakılabilir.

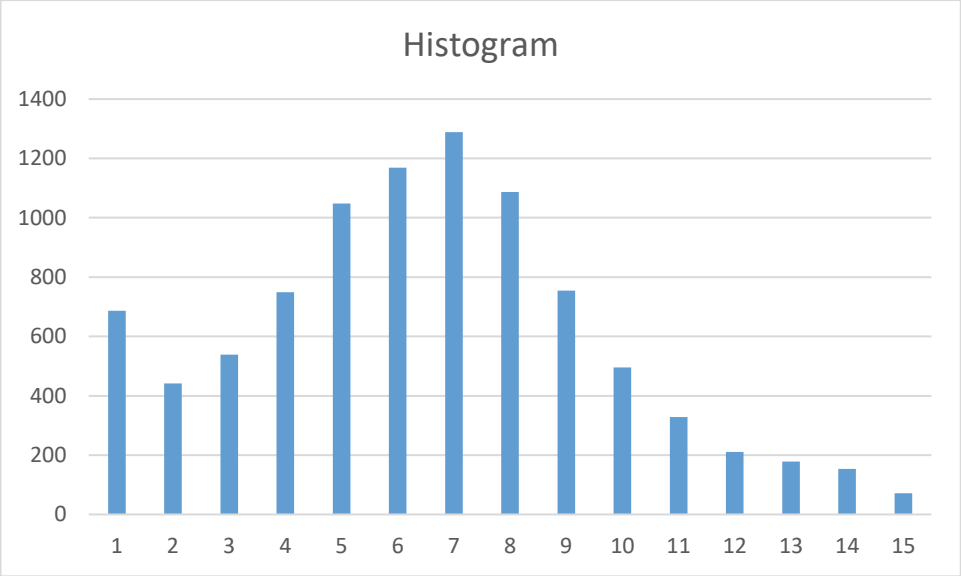
Veri ve Yöntem

EEG veriseti içerisinde epileptik nöbet geçirmeyen hastalar 4 gruba ayrılmışlardır. Bu gruplar beyninde tümör olan alanda EEG'si çekilenler, beyninde tümör olduğu halde beyninde sağlıklı alanda EEG'si çekilenler, gözleri kapalıyken EEG'si çekilenler, gözleri açıkken EEG'si çekilenler olarak kategorize edilmiştir. Bu kategoriler sırayla 2,3,4,5 olarak kodlanırken, orjinal veride epilepsi hastaları 1 olarak kodlanmıştır. Bu verideki bütün zaman serilerini, ön işlemeden geçirerek iki düzeyli kategorik zaman serilerine dönüştürdük. Bunu yapmamız şu şekilde oldu:

- 1. adım: Bütün zaman serilerinin medyanını bulunması
- 2. adım: Medyanlardan epileptik nöbet geçirenlerin ve geçirmeyenlerin ayrı ayrı sınıflara koyulması
- 3. adım: İki grupta medyanların medyanının alınması
- 4. adım: İki grup medyanının ortalamasının alınması
- 5.adım: Veride bu değere göre yüksek EEG ölçümlerinin 1 ve düşük olanların 0 değerini alması

4. adımda bu değer -5.25 olarak hesaplanmıştır. Bu çalışmada Çalışmadaki olası sorunların nasıl ele aldığımı görmek için sonuç ve tartışma kısmına bakın) olarak aldık. Orjinal veride 5 grup vardı ancak mevcut literatür çalışmasından dolayı epilepsi hastalarının zaman serileri gözardı edilmiştir. Böylece 4 grubun herbirinde 2300'er tane zaman serisi elde edilmiştir. Bunun neticesinde iki düzeyden 1 değeri alanların histogramı şöyledir:

Grafik 1: Epilepsi hastası olmayanların 1 değerlerinin histogramı



Sözkonusu grafik incelendiğinde, 2 ya da 3 bileşenli binom karma modeli işaret ettiği görülmektedir. Çalışmada 2 ya da 3 bileşenli karma model parametreleri, Bayesçi yönteme göre oluşturulan algoritma ile hesaplanmış ve zaman serilerinin kümeleri oluşturulmuştur. Bu kümeler oluşturulurken Matlab R2016a ve Matlab R2017'den faydalanılmıştır ve oluşturulan sınıflayıcı değişken(med değişkeni) kategorileri böyle elde edilmiştir. 1000 iterasyon ile işlem yapılırken önsel dağılım olarak beta dağılımı seçilmiş ve Frühwirth-Schnatter (2006)'dan faydalanılmıştır.

Bu yöntemde Markov zincirleri kullanılır. Binom karma modeldeki başarı olasılıkları Beta(1,1) önsel dağılım olarak seçilir. Her iterasyonda k. kümeye eleman atandıkça kümeye özgü a_k ile b_k (1) ve (2)'deki gibi güncellenir:

$$a_k = a_0 + \sum_{i:S_i=k} y_i \quad (1)$$

$$b_k = b_0 + \sum_{i:S_i=k} (T - y_i) \quad (2)$$

Burada T, zaman serisinin uzunluğunu ve S_i i. zaman serisinin girdiği kümeyi temsil etmektedir. Bu algoritmada önerilen dağılım, düzgün dağılım ve Beta dağılımından biri seçilerek her iterasyonda kümeler tekrar seçilmiştir. Sonuç olarak 1000 iterasyonda kategoriler atandığında çapraz tablo haline getirilmiş ve sonrasında lojistik regresyon kullanılarak sınıflar ayırdedilmiştir.

Bulgular

Med değişkeninin orjinal değişken'den bağımsız olup olmadığına dair hipotez testi yapılmıştır ve bu testten elde edilen p değeri, çapraz tablolama neticesinde 0.001'den küçük çıkmıştır. Dolayısıyla med değişkenine göre sınıflandırma geçerlidir. Çapraz tablo sonucu, Tablo 1'deki gibi çıkmıştır:

Tablo 1:Med değişkenine göre orjinal değişkenin çapraz tablosu

		med			
		1	2	3	toplam
Y	2	55	2207	38	2300
	3	72	2201	27	2300
	4	39	2197	64	2300
	5	104	2133	63	2300
	toplam	270	8738	192	9200

Bu çapraz tablo, kümelemedeki başarıyı belirlemiştir. Daha sonra lojistik regresyon ile sınıflandırma yapmak için kategori dereceleri, gölge değişken haline getirilmiştir. Lojistik regresyon yöntemiyle sınıfların ayrıştırılması yapılmıştır. Y değişkeninden 4 ve 5 değerli gözlemleri alınıp, lojistik regresyon modeli kurduğumuzda sonuç Tablo 2'deki gibi elde edilmiştir:

Tablo 2:Sınıf kategorilerinin katsayıları ve önemlilik dereceleri

	tahmin	standart hata	t değeri	p değeri
sabit terim	0.98083	0.18777	5.2237	<0.05
med2	-1.0104	0.19021	-5.3119	<0.05
med3	-0.99658	0.25837	-3.8572	0.000115

Burada med2, medyan sınıflandırmasında 2. sınıfa düşüp düşmediğini belirleyen bir gölge değişkendir.Y değişkeninden 2 ve 3 değerli gözlemler alındığı zaman oluşturulan veri ile lojistik regresyon yapıldığındaTablo 3'teki bulgulara erişilmiştir:

Tablo 3:Sınıf kategorilerinin katsayıları ve önemlilik dereceleri

	tahmin	standart hata	t değeri	p değeri
sabit terim	-0.38221	0.17031	-2.2442	0.024822
med2	0.38867	0.173	2.2466	0.024663
med3	0.58765	0.24665	2.3825	0.017193

Sonuç ve Tartışma

Bütün bulgulara göre kanser hastalarının EEG'si çekilirken tümör bölgesinin belirlenmesi çok önemlidir. Araştırmada kullanılan söz konusu yöntem %97.2'lik oranda bir ipucu verebilir. Gözlerin çalışması ise EEG sinyalleriyle anlaşılabilecek bir düzeyde tespit edilebilir. Bunun yanında veri işlemesi konusunda birtakım sorular ortaya çıkmasının muhtemel olabileceği düşünülebilir. Medyanlar ile yapılan ön işlemede tek bir zaman serisi oluşturularak medyan hesaplandığında bu değerin -7 olduğunu gördük ama hali hazırda epilepsi nöbeti geçirenler ve geçirmeyenler diye iki grup varken bu çok adil değildi. Çünkü bütün epilepsi nöbet yaşayanlar diğerlerinin %25'i kadardı ve ileride yapılacak çalışmalarda bu rahatlıkla bozulabilir. Bu nedenle iki farklı grubu tanımak sözkonusu zarardan dönmeye yardımcı olmuştur. Öte yandan Bayesçi kümeleme yapmak kolay bir iş değildir. Çünkü seçtiğiniz algoritmaya ve bilgisayar performansına bağlı olarak Bayesçi kümeleme kullanılır. Bu sebepten çalışmada 1000 iterasyon yapılmıştır ve ideal sayıda bileşene ulaşmaya yardımcı olmuştur. Esas önemli nokta, bu çalışmaların nörolojik anlamda neye denk geleceğidir çünkü önemli olan beynin yapısını kavramaktır.

Kaynakça

1. Adeli H, Ghosh-Dastidar S, Dadmehr N (2006). A wavelet-chaos methodology for analysis of EEGs and EEG Sub-Bands to detect seizures and epilepsy. IEEE Transactions on Biomedical Engineering, 54, 2, 205 – 211
2. Andrzejak, RG, Lehnertz, K, Rieke, C, Mormann, F, David, P, Elger, CE (2001). Indications of nonlinear deterministic and finite dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state, Phys. Rev. E, 64, 061907
3. Frühwirth-Schnatter, S (2006). Finite mixture and markov switching models. Springer Series,
4. Polat K, Gunes S (2007). Artificial immune recognition system with fuzzy resource allocation mechanism classifier, principal component analysis and FFT method based new hybrid automated identification system for classification of EEG signals. Expert Systems With Applications, 34, 3, 2039-2048

5. Srinivasan V, Eswaran C, Sriraam N (2005). Artificial Neural Network Based Epileptic Detection Using Time-Domain and Frequency-Domain Features. Journal of Medical Systems, 29, 647-660
6. Srinivasan V, Eswaran C, Sriraam N, H (2006). Approximate Entropy based Epileptic EEG detection using Artificial Neural Networks. IEEE Transactions on Information Technology in Biomedicine, 11, 3, 288-295
7. Venema V, Ament F, Simmer C (2006). A Stochastic Iterative Amplitude Adjusted Fourier Transform algorithm with improved accuracy. Nonlin. Processes Geophys., 13, 321–328

T11

Sınıflama Metotlarının Karşılaştırılmasını Kolaylaştırmak için İnteraktif Bir R Shiny Uygulaması: ClasComp

Uğur TOPRAK[1], Beyza DOĞANAY ERDOĞAN[2], Ersin ÖĞÜŞ[1]

[1]*Başkent Üniveristesi Tıp Fakültesi Biyoistatistik Anabilim Dalı, Ankara*
[2]*Ankara Üniveristesi Tıp Fakültesi Biyoistatistik Anabilim Dal, Ankara*

toprakugur@ymail.com

1. Giriş

Sınıflandırma bir karar verme problemi olarak, gözlemler ele alındığında yeni bir gözlemin ait olduğu kategorinin belirlenmesidir. Doğada ve çevremizde gördüğümüz tüm varlıkları, ister istemez, farkında olsak da olmasak da sınıflandırırız. Bu temel ve sıklıkla kullanılan teknik karar verme sürecinin temelini oluşturmaktadır ().

İstatistik ve makine öğrenmesi alanlarında, sınıflandırma, yeni bir gözlemin bir kategori kümesinden hangisine ait olduğunu, temel gözlemlerden ve bilinen kategorilerinden oluşan bir çalıştırma seti kullanarak bulunmasını ifade eder. En sık kullanım alanlarına örnek olarak, e-postaların gerekli ya da gereksiz olarak sınıflandırılması veya bir hastalık teşhisinin hastanın gözlemlenen karakteristiklerinden (cinsiyet, kan basıncı, çeşitli semptomların varlığı vb.) yararlanılarak bulunması verilebilir. Sınıflandırma, sahip olunan veri setinden gözle görülemeyecek örüntüleri çıkararak, bu örüntüler yardımıyla ilgilenilen ve bilinmeyen yeni durumlar için ait oldukları en uygun sınıfa atamayı yapmak üzere oluşturulan istatistiksel tahmin modelleridir ().

Basit bir sınıflandırma modeli; tahmin değişkenleri yardımıyla, hedef değişkeninin hesap edildiği bir fonksiyon olarak ifade edilmektedir. Tahmin değişkenleri ile hedef değişkenleri arasındaki ilişki veri seti üzerinde yapılan deneylerle bulunmaktadır. Bir sınıflandırma işleminde kullanılan veri temel olarak ikiye ayrılmaktadır; verilerin ilk bölümü istatistiksel modellerin kurulduğu eğitim verisi diğeri ise modelin daha önce karşılamadığı test verisidir. Sınıflandırma modeli öncelikle eğitim verisi ile eğitilip test verisi ile de değerlendirilmektedir. Test verilerinde elde edilen sınıflandırma başarısı modelin doğruluğunu vermektedir ().

Mevcut sınıflandırma yöntemlerinin kullanımına yönelik birçok paket program ve yazılım dilinin kullanılabildiği platform bulunmaktadır. Fakat, bu program ve platformlar ilgili alanda ki araştırmacılar tarafından kullanım ve raporlama zorluğu, kodlama becerileri gerekliliği, güncel yaklaşımlarım eksikliği gibi nedenlerden dolayı oldukça zaman alıcı ve karmaşık görülmektedir. Literatürde sınıflandırma metotlarının karşılaştırmasında kullanılan ücretsiz, önyükleme gerektirmeyen, platformdan bağımsız, herhangi bir kodlama becerisi gerektirmeyen, arayüz bakımından kullanıcı dostu olan, güncel metotların bulunduğu, ayrıca bu metotların hem tek tek hem de birlikte performanslarının değerlendirilip raporlandığı bir program ya da platform, basit ve birçok yönden eksik örnekleri bulunsa da gelişmiş bir örneği bulunmamaktadır. Bu çalışma ile hemen her alandan araştırmacıların sınıflandırma problemlerini ve sınıflandırma metotları arasındaki performans karşılaştırmalarını kolaylaştırmak üzere interaktif bir R Shiny web uygulaması geliştirmek amaçlanmıştır.

2. GEREÇ VE YÖNTEM

Dünyada veri kaynaklarının hızla artmasıyla birlikte analiz edilmesi gereken verilerin büyüklüğü de hızlı bir artış göstermiştir. Bu artış, verilerin analiz edilmesi gereken platformlarında çeşitlenmesine neden olmuştur. Günümüzde sınıflandırma metodu için birçok yaklaşım ortaya çıkmıştır. Bu yaklaşımlar, farklı paket programlar ve yazılımlar kullanılması nedeniyle ortak bir platformu paylaşamamaktadır. Aşağıda uygulamamız altında topladığımız bir kısmı yaygın olarak kullanılan bir kısmı ise yeni yaklaşımlara sahip sınıflandırma algoritmaları kısaca verilmiştir.

2.1. Lineer Diskriminant Analizi

Lineer diskriminant analizi (LDA), veri setinde bulunan verilerin kategorilere atanırken taşıdığı özelliklere göre ayrım yapmaktadır. Bir kategorik bağımlı değişken ile sayısal değerler alan bağımsız değişkenler arasında kullanılmaktadır. LDA, bağımsız değişkenlerin bağımlı değişkenleri etkilemelerine göre ya aynı ya da farklı gruplara sınıflandırılmasını gerçekleştirmektedir (). Fisher LDA için aşağıdaki skor fonksiyonunu tanımlamaktadır.

$$Z = \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_d x_d$$

$$S(\beta) = \frac{\beta^T \mu_1 - \beta^T \mu_2}{\beta^T C \beta} \quad \text{Skor fonksiyonu}$$

$$\Downarrow$$

$$S(\beta) = \frac{\bar{Z}_1 - \bar{Z}_2}{\text{Z'nin Gruptaki Varyansı}} \qquad (1)$$

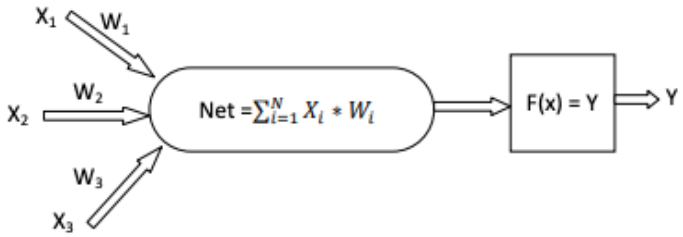
Burada amaç skor fonksiyonunu maksimize eden lineer katsayıları tahmin etmektir. β doğrusal model katsayıları, C kovaryans matrisler ve μ ortalama vektörleri ifade eder. En iyi diskriminant, iki grup arasında Mahalanobis mesafesini hesaplanarak belirlenmektedir. Mahalanobis mesafesinin üçten küçük olması yanlış sınıflandırma olasılığını oldukça küçük olduğunu anlamına gelmektedir.

$$\beta = C^{-1}(\mu_1 - \mu_2)$$

$$C = \frac{1}{n_1 + n_2}(n_1 C_1 + n_2 C_2) \qquad (2)$$

2.2. Yapay Sinir Ağları

Yapay sinir ağları (YSA); insan beynin esinlenerek, öğrenme sürecinin matematiksel olarak modellenmesi çalışmaları sonucu ortaya çıkmıştır. Temel olarak insan beynindeki birçok biyolojik sinir hücresinin birbirine bağlanmasına benzer bir yapıda, YSA; biyolojik sinir hücresinin girdi, işlem ve çıktı karakteristiğini taklit eden birçok basit, çoğunlukla adaptif işlem birimlerinin (yapay sinir hücrelerinin) farklı etki seviyelerinde, belirli bir işlem bütününü gerçekleştirmek üzere birbirine bağlanması ile oluşturulmuş istatistiksel model ağlarıdır (). Basit bir YSA yapısı aşağıda gösterilmiştir.

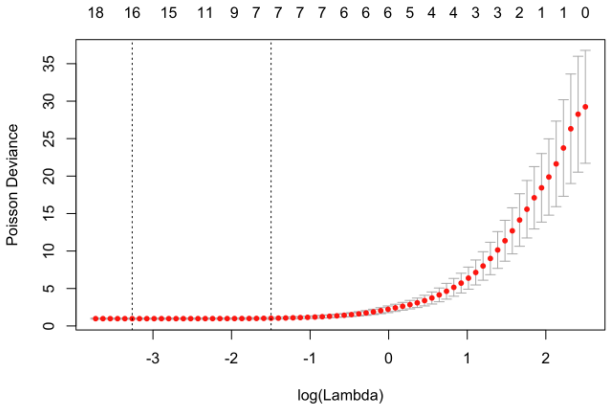


Şekil1. Yapay sinir hücresinin yapısı

Bir YSA modeli temel olarak üç yapıdan oluşmaktadır. Bunlar girdiler, ağırlar ve toplama ya da birleştirme fonksiyonudur. Girdiler nöronlara gelen verileri ifade etmektedir. Girdiler yapay sinir hücresine bir diğer hücreden gelebileceği gibi direk olarak dış dünyadan da gelebilmektedir. Bu girdilerden gelen veriler biyolojik sinir hücrelerinde olduğu gibi toplanmak üzere nöron çekirdeğine gönderilir. Ağırlıklar, yapay sinir hücresine gelen bilgiler girdiler üzerinden çekirdeğe ulaşmadan önce geldikleri bağlantıların ağırlığıyla çarpılarak çekirdeğe iletilmesi işlemdir. Bu sayede girdilerin üretilecek çıktı üzerindeki etkisi ayarlana bilinmektedir. Bu ağırlıkların değerleri pozitif, negatif veya sıfır değerler alabilmektedir. Ağırlığı sıfır olan girdilerin çıktı üzerinde herhangi bir etkisi olmamaktadır. Toplama fonksiyonu bir yapay sinir hücresine ağırlıklarla çarpılarak gelen girdileri toplayarak o hücrenin net girdisinin hesaplandığı aşamayı göstermektedir.

2.3. glmnet

gmlnet, genelleştirilmiş doğrusal modellerin konveks cezalandırma metotları ile geliştirilerek birçok problemin çözümünde kullanılan hızlı bir algoritmadır. İstatistiksel modelleri lasso ve farklı elastik net yöntemleriyle cezalandırarak başta doğrusal regresyon, lojistik regresyon, mutinomial regresyon ve ridge regresyon problemleri, ayrıca kayıp verinin yüksek olduğu durumlar için de sınıflandırma çözümleri getirebilmektedir ().



Şekil 2. glmnet ile çözümlenmiş bir Poisson modeline ait λ'nın optimize edilmesi

2.4. Rassal Orman (Random Forest)

Rassal orman (RO), ağaç tipi sınıflandırıcılar olarak tanımlanmaktadır (Breiman 2001). RO, tüm değişkenler arasından en iyi dalı kullanarak her bir düğümü dallara ayırmak yerine, her bir düğümde rastgele olarak seçilen değişkenler arasından en iyisini kullanarak her bir düğümü dallara ayırır. Her bir veri seti orijinal veri setinden yer değiştirmeli olarak üretilmektedir. Sonra rastgele özellik seçimi kullanılarak ağaçlar geliştirilmektedir.

RO öncelikle kullanıcı tarafından en az iki değişken tanımlanmasıyla başlatılmaktadır. Her bir düğümde m değişkenleri tüm değişkenler arasından rastgele olarak seçilip ve bu değişkenler arasından en iyi dal belirlenmektedir. Yeterli öngörü gücü ile yeterli miktarda düşük korelasyon sağlayan değişken sayısının seçimi son derece önem taşımaktadır (Horning 2010). Temelince çalışma algoritması olarak sınıflandırma ve regresyon ağaçları kullanmaktadır (Breiman 2001).

2.5. eXtreme Gradient Boosting (XGBoost)

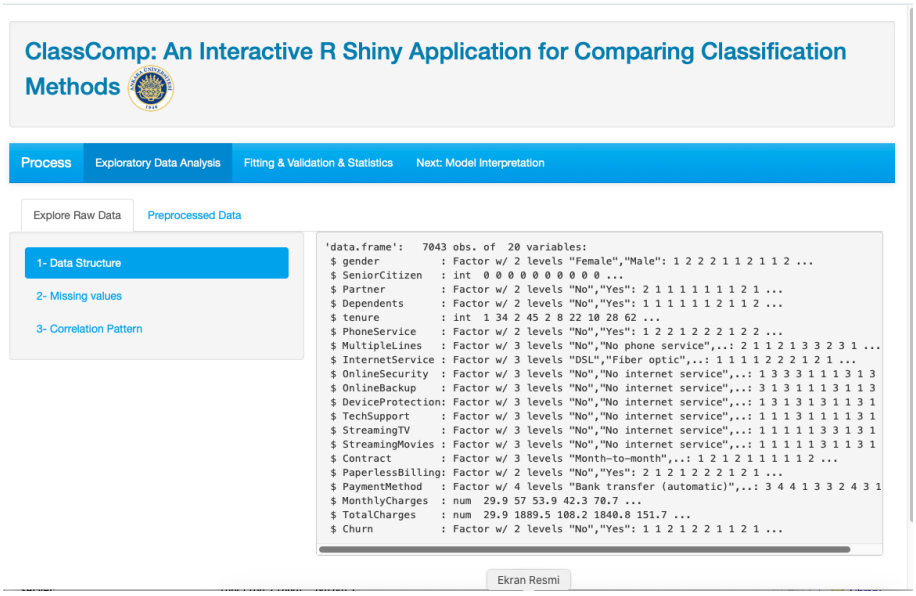
Boosting türü sınıflandırma yöntemleri arasında en iyi yöntemlerden biri olarak gösterilmektedir. XGBoost genellikle yüksek ölçekli verilerde gösterdiği yüksek sınıflama perfomansı ile öne çıkmaktadır. XGBoost, herhangi bir eğitim setine ihtiyaç duymadan, iteratif olarak veri setinden k tane ağaç oluşturup her iterasyonda bir önceki oluşan sınıflamada yanlış sınıflara atanan gözlemler için yeni modeller kurmaktadır. Hızlı çalışması, doğru sınıflandırma başarısının yüksek ve aşırı öğrenmeye karşı direnç olması XCBoost'un diğer yöntemlere karşı üstünlüğünü göstermektedir ().

2.6. ClassComp, R Shiny Uygulaması

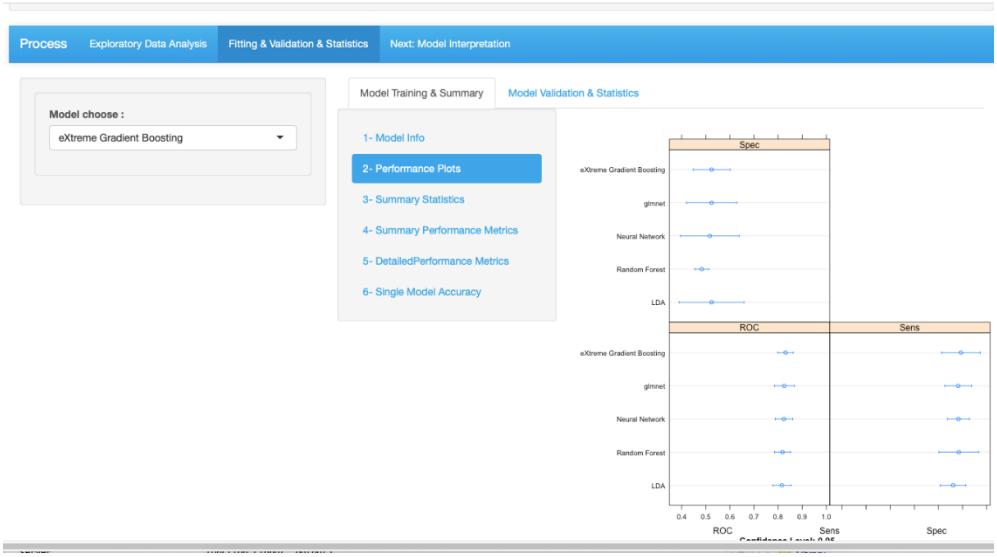
Geliştirdiğimiz R Shiny uygulaması, internet üzerinden çalışmaktadır. Geliştirme aşamasında R V 3.6.1kullanılmıştır. İnternete bağlanabilen tüm cihazlar üzerinde sorunsuz çalışmaktadır. Uygulamamızın en önemli özelliklerinden bir tanesi, herhangi bir

önyüklemeye ihtiyaç duymadan kullanılabilmesidir. Kullanıcı dostu ara yüzü menüler arasında kaybolmadan kolayca kullanılabilmektedir.

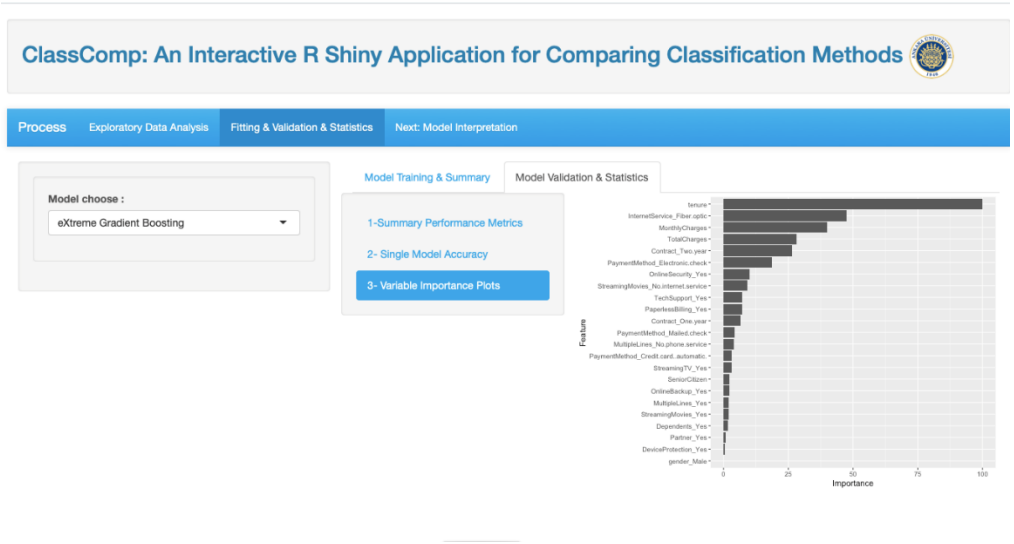
Uygulama ekranı ilk açıldığında verilerinizi yükleyebileceğiniz "upload" paneli yardımıyla verileriniz kolayca sisteme aktarılabilmektedir. Sonrasında otomatik olarak ham verileriniz için veri yapısı, kayıp veri durumu ve korelasyon analizi yapılmaktadır. Eğer veri setinizde bu durumlarla karşı karşıya iseniz yine otomatik olarak kayıp verilerin bulunduğu satırları silinmesini, yüksek ilişkili değişkenleriniz varsa bunlarında çıkarılması sağlanabilmektedir.



Model eğitim sekmesine tıklayıp ilgilendiğiniz her hangi bir modeli seçtiğinizde arka planda tüm algoritmalar sırasıyla çalışacak ve size eğitim setindeki başarı oranlarına dair performans ölçütlerini raporlayacaktır.



Model validasyon sekmesine tıkladığınızda ise modelinize otomatik olarak orijinal veri setinden ayırılan test seti verilecek gereke tek tek gerekse toplu olarak sınıflandırma sonuçları için rapoları elde edebileceksiniz.



3. TARTIŞMA

Geliştirdiğimiz uygulama ile yaptığımız örnek çalışmalarda diğer platformlara göre kullanıcı arayüzünün basit ve kullanışlı, çalışma performansının yüksek olması en büyük avantajlarıdır. Herhangi bir platforma bağlı kalmadan ve hiçbir kodlama bilgisi gerektirmeden kullanılabilecek bu uygulama araştırmacılar tarafından önemli bir araç olacaktır.

4. SONUÇ

Sınıflandırma problemlerinde sınıflayıcıların performans ölçütlerinin karşılaştırılması önemli noktalardan biridir. Veri setinin yapısına göre farklı performans gösteren bu algoritmaları bir arada kullanabilmek ve araştırmacıların kendi veri setlerini en iyi sınıflayan yöntemleri keşfetmesinde kullanılabilecek olan uygulamamız taşıdığı nitelikler itibariyle özel bir yere sahip olması beklenilmektedir. Uygulamamıza yakın gelecekte tüm karşılaştırmalı performans sonuçlarının APA formatında raporlanmasını sağlayacak bir modül üzerinde çalışmalara devam edilmektedir.

KAYNAKLAR

1. Breiman L., (2001), Random forests,machine learning, 2001 Kluwer Academic Publishers, 45(1), 5-32
2. Friedman, J., Hastie, T., & Tibshirani, R. (2009). glmnet: Lasso and elastic-net regularized generalized linear models. *R package version*, *1*(4).
3. Fan, J., Wang, X., Wu, L., Zhou, H., Zhang, F., Yu, X., ... & Xiang, Y. (2018). Comparison of Support Vector Machine and Extreme Gradient Boosting for predicting daily global solar

radiation using temperature and precipitation in humid subtropical climates: A case study in China. *Energy conversion and management*, *164*, 102-111.

4. Pal, M. (2005). Random forest classifier for remote sensing classification. *International Journal of Remote Sensing*, *26*(1), 217-222.
5. Horning N., (2010), RandomForests : An algorithm for image classification and generation of continuous fields data sets, International Conference on Geoinformatics for Spatial Infrastructure Development in Earth and Allied Sciences (GISIDEAS) 2010, 9-11 December, Hanoi,Vietnam
6. Harrington, P. (2012). *Machine learning in action*. Manning Publications Co..
7. Kotsiantis, S. B., Zaharakis, I., & Pintelas, P. (2007). Supervised machine learning: A review of classification techniques. *Emerging artificial intelligence applications in computer engineering*, *160*, 3-24.
8. Adeli, H. (1999). Machine Learning-Neural Networks, Genetic Algorithms and Fuzzy Systems. *Kybernetes*.

T12

TÜRKİYE'DEKİ TIP FAKÜLTELERİNDE MEZUNİYET ÖNCESİ VERİLEN BİYOİSTATİSTİK EĞİTİMİNİN İÇERİĞİNİN STANDARTLAŞTIRILMASI-BİR DELPHİ ÇALIŞMASI

YÜKSEL S[1], ALKAN A[1], DEMİR P[1], SANİSOĞLU S.Y[1], ERCAN İ[2], TÜRE M[3], KARAHAN S[4], ATEŞ C[5], ELMALI F[6], ERDOĞAN S[7], KAYA M.O[8], ŞAHİN B[9], YAVUZ Z[10], YILDIRIM D[7]

1. Ankara Yıldırım Beyazıt Üniversitesi Tıp Fakültesi Biyoistatistik Anabilim Dalı, ANKARA
2. Bursa Uludağ Üniversitesi Tıp Fakültesi Biyoistatistik Anabilim Dalı, BURSA
3. Adnan Menderes Üniversitesi Tıp Fakültesi Biyoistatistik Anabilim Dalı, AYDIN
4. Hacettepe Üniversitesi Tıp Fakültesi Biyoistatistik Anabilim Dalı, ANKARA
5. Yüzüncü Yıl Üniversitesi Tıp Fakültesi Biyoistatistik Anabilim Dalı, VAN
6. İzmir Kâtip Çelebi Üniversitesi Tıp Fakültesi Biyoistatistik Anabilim Dalı, İZMİR
7. Mersin Üniversitesi Tıp Fakültesi Biyoistatistik Anabilim Dalı, MERSİN
8. Fırat Üniversitesi Tıp Fakültesi Biyoistatistik ve Tıp Bilişimi AD, ELAZIĞ
9. Pamukkale Üniversitesi Tıp Fakültesi Biyoistatistik Anabilim Dalı, DENİZLİ
10. Ankara Üniversitesi Tıp Fakültesi Biyoistatistik Anabilim Dalı, ANKARA

Özet

Amaç: Bu çalışmanın amacı, Türkiye'de Tıp Fakültelerinde mezuniyet öncesi verilen biyoistatistik eğitiminin içeriğine ilişkin; ders adları, ders içerikleri, ders anlatma yöntemleri, ders tipi, ders süresi ve dersin verilmesi gereken sınıf düzeyi hakkında uzman fikirlerinden yararlanarak, biyoistatistik eğitiminin içeriği hakkında standardizasyon sağlamaktır.

Yöntem: Çalışmanın amacına yönelik, içeriğinde üç tur tanımlanmış Delphi yöntemi kullanılarak bir çevrimiçi anket çalışması yapılmıştır. Anketteki ders başlıkları ve ders içerikleri halihazırda farklı Tıp Fakültelerinde verilen ders başlıkları ve içerikleri dikkate alınarak hazırlanmıştır. Biyoistatistik Derneği'ne kayıtlı 153 öğretim elemanına, çalışmaya katılmaları için, davet e-postası gönderilmiştir. Çalışmayı kabul eden ve çalışmaya katılma koşullarını sağlayan 33 kişiye 1.tur anketi gönderilmiştir. 1.tur anketi, 24 ders başlığı ve 236 ders içeriğinin sorgulanmasına ilişkin yapılandırılmış anket formundan oluşturulmuştur. 1.tur anketini tamamlayan 23 kişiye 19 ders başlığı ve 78 ders içeriği, 1.turda fikir birliği sağlanmadığı için, 2.turda tekrar sorulmuştur ve 16 kişiden yanıt alınmıştır. 3.turda 16 ders başlığı ve 61 ders içeriği tekrar sorgulanmış ve nihai sonuçlar 16 kişi üzerinden elde edilmiştir.

Bulgular: Toplam 24 konu başlığından 22'sinde fikir birliğine varılmıştır. Kuramsal dağılımlar başlığı normal dağılımlar ve özellikleri olarak indirgenmiş, tek örneklem testlerinin ise hipotez testleri-2 başlığı altında anlatılmasına karar verilmiştir. Çoklu ikili lojistik regresyon analizinin ise müfredatta olması konusunda fikir birliği sağlanamamıştır. 236 ders içeriğinden 40'ında (%16,9) fikir birliğine varılamamıştır. Ders tipi ve dersin anlatılması gereken sınıf düzeyi sırasıyla iki ve bir ders başlığı hariç tüm dersler için belirlenmiştir. Teorik ve uygulama ders süreleri tanımlanmıştır.

Sonuç: Türkiye'de Tıp Fakültelerinde verilen Biyoistatistik ders müfredatının içeriği üzerine şimdiye kadar Türkiye çapında yapılan bir çalışma bulunmamaktadır. Bu çalışma ile, Biyoistatistik müfredatının içeriğine ilişkin net bir içerik çıkarılmıştır. Bu içeriğin, biyoistatistik çekirdek müfredatının geliştirilmesine katkı sağlayacağı düşünülmektedir. Fikir birliği sağlanmayan ders

içeriklerinin de sonuçlara yansımış olması, Biyoistatistik eğitiminde detaylı irdelenmesi gereken konuların olduğunun ispatıdır.

Anahtar kelimeler: Biyoistatistik müfredatı, tıp eğitimi, Delphi yöntemi

GİRİŞ

Tıp Fakültesinde çekirdek müfredatın geliştirilmesine ilişkin çalışmalar son zamanlarda artmıştır. Bunun belki en büyük sebebi, gelişen ve artan tıp ve sağlık literatürünün öğrencilerde aşırı bilgi yüklenmesini engellemektir. Bununla birlikte, uzmanlık alanlarının spesifikleşmesi ile önce ulusal sonra ise uluslararası düzeyde standardizasyonu sağlamak da tıp eğitim müfredatını sık sık güncelleme gereksinimini doğurmuştur. Bir tıp öğrencisinin uzmanlık eğitimi süresinde izleyeceği yolu belirlemesinde oldukça büyük etkisi olan çekirdek müfredat içerikleri gün geçtikçe önemini arttırmaktadır.

Ülkemizde yaklaşık 20 yıl önce Tıp Fakültelerinde Biyoistatistik ders içeriği, teorik istatistik bilgiler ile hazırlanmaktaydı. Bilgisayar kullanımının kısıtlı olduğu, laboratuvarların olmadığı bu dönemlerde tıp öğrencilerine temel istatistiki hesaplamaların ve analizlerin elle hesaplanmasının anlatılması, öğrencilerin teorik istatistik bilgi kazanımlarını gerçek veri setinde uygulayabilmelerine imkân tanımamıştır. Uygulamasına değinilmemiş istatistiki teorik bilgilerin öğrenci tarafından uzmanlık döneminde unutulması, lisans döneminde verilen biyoistatistik eğitiminin etkinliğinin sorgulanması gerekliliğini ortaya koymuştur. Gelişen teknoloji ve imkanlarla, halihazırda çoğu anabilim dalı öğrencilerine verdiği teorik istatistik bilgileri paket programlar ile desteklerken bazı anabilim dalları yıllar öncesini temel alan öğretme yöntemini kullanmaktadır. Biyoistatistik anabilim dallarının ders içeriklerine bakıldığında ise, her bir anabilim dalının farklı içeriklerle ders anlattığı görülmektedir. Bu nedenle ulusal düzeyde, Tıp fakültelerinde verilen biyoistatistik eğitimi içeriğindeki ders adları, ders içerikleri, ders anlatma yöntemleri, ders tipleri, ders süresi ve dersin verilmesi gereken sınıf düzeyi hakkında fikir birliğine ihtiyaç vardır. Biyoistatistik çekirdek müfredatının hazırlanmasında yardımcı olması adına çalışma tasarlanmış ve yürütülmüştür.

YÖNTEM

Bu çalışmada üç aşamalı Delphi yöntemi kullanılmıştır. Bir görüş birliği sağlama aracı olarak ifade edilen Delphi yöntemi bir probleme farklı açılardan bakan bireylerin ya da grupların yüz yüze gelmeden uzlaşmalarını amaçlayan bir yöntemdir. Delphi yöntemi; kararların güçlü kişi ya da gruplar tarafından etkilenme olasılığı bulunduğu durumlarda geçerli ve güvenilir sonuçlar elde etmek amacıyla kullanılabilir (Turoff,2001). Genel olarak; katılımda gizlilik, yapılandırılmış ya da yarı yapılandırılmış anketlerin ardışık olarak uygulanması, grup tepkisinin nitel ve nicel analizleri, analiz sonuçlarının katılımcılara geri bildirimi, katılımcılara her bir aşamada düşüncelerini yeniden şekillendirme ve karar verme fırsatının verilmesi, ardışık uygulamaların görüş birliği oluşana kadar devam ettirilmesi Delphi yönteminin en temel özellikleridir (Dalkey,1972; Sackman, 1975).

Çalışmanın verisi, Ankara Yıldırım Beyazıt Üniversitesi İstatistik Danışmanlık Uygulama ve Araştırma Merkezi tarafından alınan REDCap elektronik data toplama aracı ile toplanmış ve düzenlenmiştir. REDCap (Research Electronic Data Capture), araştırma çalışmaları için veri toplamayı desteklemek üzere tasarlanmış güvenli, web tabanlı bir yazılım platformudur. REDCap;

1) veri toplama için bir arayüz sağlar 2) veri manipülasyonu ve dışa aktarım süreçleri için denetleme sağlar 3) çalışma verisini, sıkça kullanılan istatistik paket programına aktarmayı sağlar 4) veri entegrasyonu ve dış kaynaklarla birlikte çalışabilirlik sürecini sağlar (Harris,2009; Harris 2019).

Delphi tekniğinin uygulanması, alanında uzman olan kişilerin probleme ilişkin yaklaşımlarını, bakış açılarını ortaya çıkarmaya, incelemeye ve bir uzlaşma sağlamaya yönelik bir dizi aşamadan oluşur. Bu çalışmada da tanımlanan aşamalar ve içerikler aşağıdaki gibidir:

i- Katılımcıların seçimi:

Delphi tekniğinde katılımcıların uzman görüşlerini yansıtacak düzeyde olması gerekliliği dikkate alınarak, çalışma evreni, Türkiye'de Tıp fakültelerinde biyoistatistik eğitimi veren ve biyoistatistik yüksek lisans veya doktora derecesine sahip tüm öğretim elemanları olarak tanımlanmıştır. Katılımcıların e-postası Biyoistatistik Derneği tarafından sağlanmış, 153 dernek üyesi öğretim elemanına, davet e-postası gönderilmiştir. Bu e-postadaki linki tıklayıp kayıt formunu dolduranlar, katılımcı olarak tanımlanmıştır. Kayıt formunu toplam 41 öğretim elemanı doldurmuş, bu öğretim elemanları içerisinden çalışmada yer alma kriterlerini sağlayan toplam 33 öğretim elemanına "1. Tur" soruları iletilmiştir. 1.turda yer alan maddeleri yanıtlayan toplam 23 öğretim elemanına "2. Tur" soruları gönderilmiştir. 2.tura da yanıt veren toplam 16 öğretim elemanından "3. Tur" sorularının da incelenmesi istenmiştir. Çalışma nihai sonuçları 16 öğretim elemanından elde edilen veriler ile sağlanmıştır.

ii- Turlardaki anketlerin hazırlanması:

Türkiye'de Tıp Fakültesinde 20 seneyi aşkın Biyoistatistik eğitimi veren anabilim dallarının lisans programları incelenerek 24 ders başlığına ilişkin ders içerikleri oluşturulmuştur. Ders içeriklerinin yanı sıra, ders anlatma yöntemi (*teorik, uygulama (istatistik paket programı kullanarak), uygulama (elle hesaplama-Excel ile), uygulama (programlama dili kullanarak), flipp class, ödev, diğer*), ders tipi (*zorunlu, seçmeli*), teorik ve uygulama ders süresi, dersin verilmesi gereken sınıf düzeyi (*1.sınıf, 2.sınıf, 3.sınıf, 4.sınıf, 5.sınıf, 6.sınıf*) konularını da sorgulayan sorular ankete eklenmiştir. Anket formu, katılımcılardan 1-9 arası likert ölçek tipinde, birden çok seçenekli ve açık uçlu yanıt alınacak şekilde hazırlanmıştır.

Yukarıda da değinildiği gibi, 1.tur anketi, 24 ders başlığı ve 236 ders içeriğinin sorgulanmasına ilişkin yapılandırılmış anket formundan oluşmuştur. Anket sorularına ilişkin %70 fikir birliği oranı yakalandığında, o soruya ilişkin fikir birliğine varıldığı kararının verileceği hipotezi altında çalışma yürütülmüştür. Fikir birliği oranı, yanıt kategorisi 1-9 aralığındaki likert ölçekte olan sorular için 1-2-3 veya 7-8-9 yanıtını verenlerin yüzdesi kullanılarak hesaplanmıştır. Birden çok seçeneğin seçilebildiği sorularda ise en fazla seçilen yanıtta fikir birliği sağlandığına karar verilmiştir. İlk turda sunulan 24 ders başlığından 19'unda fikir birliği oranı %70'in altından olduğu için 2.turda bu konu başlıkları katılımcılara tekrar sorulmuştur. Yine 1.turda sorulan 236 ders içeriğinin 78'i, %70 fikir birliği oranının altında değerlendirildiği için katılımcılara 2.turda tekrar sorulmuştur ve 16 kişiden yanıt alınmıştır. 3.turda 16 ders başlığı ve 61 ders içeriği tekrar sorgulanmış ve nihai sonuçlar 16 kişi üzerinden elde edilmiştir.

BULGULAR

Her bir derse ilişkin elde edilen sonuçlar aşağıda tablolar ile verilmiştir. Fikir birliğine varılamayan ders içerikleri kırmızı ile belirtilmiştir.

Dersin adı ve içeriği	**Fikir birliği sağlanan tur**	**Fikir birliği yüzdesi**
Biyoistatistiğe Giriş	Tur 1	93,5 (müfredatta olmalı)
İstatistik ve Biyoistatistik biliminin tanımı, kullanım alanları	Tur 1	92,9 (içerikte olmalı)
Biyoistatistiğin uygulama alanları	Tur 1	96,4 (içerikte olmalı)
Evren ve örneklem tanımı	Tur 1	100,0 (içerikte olmalı)
Tanımlayıcı istatistik tanımı (veri, denek, değişken, kitle, örneklem, parametre, istatistik);Veri türleri (sıralanabilir, sınıflanabilir, kesikli, sürekli)	Tur 1	100,0 (içerikte olmalı)
Çıkarımsal istatistik tanımı	Tur 1	89,3 (içerikte olmalı)
Örneklem seçiminde temel kavramlar (örneklem çerçevesi, örneklem tasarımı, örneklem birimi, örneklem kesiri)	Tur 1	89,3 (içerikte olmalı)
Örneklem seçiminin önemi	Tur 1	92,9 (içerikte olmalı)
Örneklem hataları (sistematik hata, örneklem hatası, temsil edebilirlik, bilgi yanlılığı, rasgele hata)	Tur 1	82,1 (içerikte olmalı)
Olasılıksız örnekleme yöntemleri (uygun örnekleme, kartopu örnekleme, amaçlı örnekleme)	Tur 3	80,0 (içerikte olmamalı)
Olasılıklı örnekleme yöntemleri (basit rasgele örnekleme, sistematik örnekleme, tabakalı örnekleme, çok aşamalı örnekleme, küme örneklemesi)	Tur 1	74,1 (içerikte olmalı)
Uygun örneklem büyüklüğü seçmenin önemi	Tur 1	92,9 (içerikte olmalı)

Ders tipi: Zorunlu (%100,0)
Anlatım yöntemi ve saati: Teorik (%89,7) - 4,5 saat, Uygulama (%65,5) - 3 saat
Uygulama: İstatistik paket program ile (%94,7)
Sınıf düzeyi: 1. sınıf/dönem (%55,2)
Öneri doğrultusunda Tur 2'de yöneltilen 3 başlık içerisinden "Biyoistatistiğe giriş" başlığı seçildi (%81,3).

Dersin adı ve içeriği	Fikir birliği sağlanan tur	Fikir birliği yüzdesi
Araştırma yöntemleri	Tur 1	92,6 (müfredatta olmalı)
Araştırmanın tanımı	Tur 1	96,3 (içerikte olmalı)
Bir araştırma sorusunun karakteristikleri (etik, anlamlı, açık, uygulanabilir)	Tur 1	92,6 (içerikte olmalı)
Araştırma hipotezi tanımı	Tur 1	96,3 (içerikte olmalı)
Araştırma sürecinde biyoistatistik biliminin yeri ve önemi	Tur 1	96,3 (içerikte olmalı)
Yokluk ve alternatif hipotez tanımı	Tur 1	100,0 (içerikte olmalı)
Tanımlayıcı araştırmalar	Tur 1	96,3 (içerikte olmalı)
Kesitsel araştırmalar	Tur 1	96,3 (içerikte olmalı)
Olgu-Kontrol araştırmaları	Tur 1	96,2 (içerikte olmalı)
Kohort araştırmaları	Tur 1	92,6 (içerikte olmalı)
Tekrarlı ölçüm çalışmaları	Tur 1	88,9 (içerikte olmalı)
Deneysel çalışmalar	Tur 1	96,3 (içerikte olmalı)
Tanı doğruluğu çalışmaları	Tur 1	81,5 (içerikte olmalı)
Geçerlilik ve güvenirlik çalışmaları	Tur 3	70,0 (içerikte olmalı)
Uyum çalışmaları	Tur 3	59,8 (içerikte olmamalı-fikir birliği sağlanamadı)
Sistematik derleme	Tur 3	60,0 (içerikte olmamalı-fikir birliği sağlanamadı)
Meta analizi	Tur 3	26,7 (içerikte olma(ma)lı-fikir birliği sağlanamadı)
Nitel araştırmalar	Tur 3	63,3 (içerikte olmamalı-fikir birliği sağlanamadı)
Araştırmalarda hata kaynakları	Tur 1	85,2 (içerikte olmalı)

Ders tipi: Zorunlu (%90,9)
Anlatım yöntemi ve saati: Teorik (%84,0) - 4 saat, Uygulama (%32,0) - 3,5 saat
Uygulama: İstatistik paket program ile (%100,0)
Sınıf düzeyi: 3. sınıf/dönem (%61,5)
Öneri doğrultusunda Tur 2'de yöneltilen;
Saha uygulaması olmalı mı sorusuna %43,8 olmalı yanıtını verdi.
Üç ders başlığı içerisinden "Araştırma yöntemleri" başlığı seçildi (%75,0).

Dersin adı ve içeriği	**Fikir birliği sağlanan tur**	**Fikir birliği yüzdesi**
Veriye ilişkin temel tanımlar ve değişken tipleri	Tur 1	92,3 (müfredatta olmalı)
Denek, değişken, ölçüm ve veri kavramları	Tur 1	96,2 (içerikte olmalı)
Parametre ve istatistik tanımları	Tur 1	92,3 (içerikte olmalı)
Sıralanabilir kategorik değişken	Tur 1	92,3 (içerikte olmalı)
Sınıflandırılabilir kategorik değişken	Tur 1	92,3 (içerikte olmalı)
Kesikli sayısal değişken	Tur 1	92,3 (içerikte olmalı)
Sürekli sayısal değişken	Tur 1	92,3 (içerikte olmalı)

Ders tipi: Zorunlu (%91,7)
Anlatım yöntemi ve saati: Teorik (%100,0) - 2 saat, Uygulama (%25,0) - 1 saat
Uygulama: İstatistik paket program ile (%100,0)
Sınıf düzeyi: 1. sınıf/dönem (%50,0)
Öneri doğrultusunda Tur 2'de yöneltilen;
Bu ders "Biyoistatistiğe giriş" dersi içerisinde verilmeli midir sorusuna %68,8 "Biyoistatistiğe Giriş ders konusu içerisinde verilmelidir." yanıtını verdi.

Dersin adı ve içeriği	Fikir birliği sağlanan tur	Fikir birliği yüzdesi
Tanımlayıcı istatistikler	Tur 1	96,0 (müfredatta olmalı)
Aritmetik ortalama	Tur 1	100,0 (içerikte olmalı)
Ortanca	Tur 1	100,0 (içerikte olmalı)
Tepe değeri	Tur 1	84,0 (içerikte olmalı)
Oran	Tur 1	96,0 (içerikte olmalı)
Geometrik ortalama	Tur 3	56,8 (içerikte olmamalı-fikir birliği sağlanamadı)
Harmonik ortalama	Tur 3	70,2 (içerikte olmamalı)
Çeyrekler, yüzdelikler	Tur 1	92,0 (içerikte olmalı)
Dağılım aralığı	Tur 1	96,0 (içerikte olmalı)
Standart sapma	Tur 1	100,0 (içerikte olmalı)
Varyans	Tur 1	96,0 (içerikte olmalı)
Çeyreklikler arası genişlik	Tur 1	96,0 (içerikte olmalı)
Çeyrek sapma	Tur 3	33,3 (içerikte olmamalı- fikir birliği sağlanamadı)
Değişim katsayısı	Tur 1	80,0 (içerikte olmalı)

Ders tipi: Zorunlu (%95,5)
Anlatım yöntemi ve saati: Teorik (%87,5) - 2 saat, Uygulama (%87,5) - 1,5 saat
Uygulama: İstatistik paket program ile (%95,2)
Sınıf düzeyi: 1. sınıf/dönem (%54,2)
Öneri doğrultusunda Tur 2'de yöneltilen;
İki ders başlığı içerisinden "Tanımlayıcı istatistikler" başlığı seçildi (%87,5).
Bu ders "Biyoistatistiğe giriş" dersi içerisinde verilmeli midir sorusuna %73,3 "Ayrı bir ders konusu olarak tanımlanmalıdır." yanıtını verdi.
Makale incelemesi olmalı mı sorusuna %68,8 olmalı yanıtını verdi.

Dersin adı ve içeriği	Fikir birliği sağlanan tur	Fikir birliği yüzdesi
Normal dağılım ve özellikleri	Tur 1	68,0 (müfredatta olmalı- fikir birliği sağlanamadı)
Binom dağılımı	Tur 3	66,8 (içerikte olmalı-fikir birliği sağlanamadı)
Poisson dağılımı	Tur 3	63,2 (içerikte olmalı-fikir birliği sağlanamadı)
Üstel dağılım	Tur 3	63,3(içerikte olmamalı-fikir birliği sağlanamadı)
Normal dağılım	Tur 1	100,0 (içerikte olmalı)
Standart normal dağılım	Tur 1	95,7 (içerikte olmalı)
Normal dağılımın özellikleri	Tur 1	100,0 (içerikte olmalı)
Normal dağılımın istatistik biliminde yeri ve önemi	Tur 1	95,7 (içerikte olmalı)
Sayısal bir değişkenin dağılımının grafiksel yöntemler ile incelenmesi	Tur 1	95,7 (içerikte olmalı)
Sayısal bir değişkenin dağılımının tanımlayıcı istatistikler ile incelenmesi	Tur 1	95,7 (içerikte olmalı)

Ders tipi: Zorunlu (%100.0)
Anlatım yöntemi ve saati: Teorik (%94,1) - 2 saat, Uygulama (%70,6) - 1,5 saat
Uygulama: İstatistik paket program ile (%83,3)
Sınıf düzeyi: 1. sınıf/dönem (%70,6)
Öneri doğrultusunda Tur 2'de yöneltilen;

İki ders başlığı içerisinden "Normal dağılım ve özellikleri" başlığı seçildi (%68,8). (Kuramsal dağılımlardı)

Bu ders "Biyoistatistiğe giriş" dersi içerisinde verilmeli midir sorusuna %68,8 "Ayrı bir ders konusu olarak tanımlanmalıdır." yanıtını verdi.

Dersin adı ve içeriği	Fikir birliği sağlanan tur	Fikir birliği yüzdesi
Hipotez testleri-1	Tur 1	91,7 (müfredatta olmalı)
Örneklem dağılımı	Tur 1	83,3 (içerikte olmalı)
Merkezi limit teoremi	Tur 3	66,7 (içerikte olmalı-fikir birliği sağlanamadı)
Kestirim kavramı	Tur 2	87,5 (içerikte olmalı)
Nokta ve aralık kestirimi	Tur 1	75,0 (içerikte olmalı)
Güven aralığı	Tur 1	83,3 (içerikte olmalı)

Ders tipi: Zorunlu (%85,7)
Anlatım yöntemi ve saati: Teorik (%95,5) - 2 saat, Uygulama (%40,9) - 1,5 saat
Uygulama: İstatistik paket program ile (%88,9)
Sınıf düzeyi: 1. sınıf/dönem (%47,6)
Öneri doğrultusunda Tur 2 de yöneltilen;
İki ders başlığı içerisinden "Hipotez testleri-1" başlığı seçildi (%87,5).
Bu ders "Biyoistatistiğe giriş" dersi içerisinde verilmeli midir sorusuna %75,0 "Ayrı bir ders konusu olarak tanımlanmalıdır." yanıtını verdi.

Dersin adı ve içeriği	Fikir birliği sağlanan tur	Fikir birliği yüzdesi
Hipotez testleri-2	Tur 1	100,0 (müfredatta olmalı)
Yokluk ve alternatif hipotez tanımları	Tur 1	100,0 (içerikte olmalı)
Tip-I hata	Tur 1	100,0 (içerikte olmalı)
Tip-II hata	Tur 1	95,8 (içerikte olmalı)
p değeri ve istatistiksel karar	Tur 1	100,0 (içerikte olmalı)
Parametrik ve parametrik olmayan test kavramı ve varsayımları	Tur 1	100,0 (içerikte olmalı)
Bir sürekli değişkenin normal dağılıma uygunluğunun test edilmesi	Tur 1	100,0 (içerikte olmalı)
İki veya ikiden fazla grubun varyanslarının homojenliğinin test edilmesi	Tur 1	91,7 (içerikte olmalı)

Ders tipi: Zorunlu (%95,7)
Anlatım yöntemi ve saati: Teorik (%95,8) - 2 saat, Uygulama (%87,5) - 1 saat
Uygulama: İstatistik paket program ile (%85,7)
Sınıf düzeyi: 3. sınıf/dönem (%45,8)

Dersin adı ve içeriği	Fikir birliği sağlanan tur	Fikir birliği yüzdesi
Veriyi uygun olan tanımlayıcı istatistikler, tablo ve grafiklerle özetleme	Tur 1	95,8 (müfredatta olmalı)
Normal dağılan veri setlerinin uygun tanımlayıcı istatistikler ile özetlenmesi	Tur 1	100,0 (içerikte olmalı)
Normal dağılmayan veri setlerinin uygun tanımlayıcı istatistikler ile özetlenmesi	Tur 1	100,0 (içerikte olmalı)
Sıklık tabloları	Tur 1	91,7 (içerikte olmalı)
Çapraz tablolar	Tur 1	100,0 (içerikte olmalı)
İçiçe tablolar	Tur 3	50,0 (içerikte olmamalı-fikir birliği sağlanamadı)
Çubuk grafik	Tur 1	87,0 (içerikte olmalı)
Daire dilim grafiği	Tur 1	87,5 (içerikte olmalı)
Histogram	Tur 1	82,6 (içerikte olmalı)
Dağılım poligonu	Tur 3	53,2 (içerikte olmamalı-fikir birliği sağlanamadı)
Kutu ve çizgi grafiği	Tur 1	87,5 (içerikte olmalı)
Dal ve yaprak grafiği	Tur 3	70,2 (içerikte olmamalı)
Ortalama-standart sapma grafiği	Tur 1	87,5 (içerikte olmalı)
Saçılım grafiği	Tur 1	91,7 (içerikte olmalı)
Populasyon piramidi	Tur 3	73,3 (içerikte olmamalı)
Forest plot	Tur 3	83,2 (içerikte olmamalı)
Funnel plot	Tur 3	83,2 (içerikte olmamalı)

Ders tipi: Zorunlu (%90,9)
Anlatım yöntemi ve saati: Teorik (%82,6) - 2 saat, Uygulama (%95,7) - 2 saat
Uygulama: İstatistik paket program ile (%90,9)
Sınıf düzeyi: 1. sınıf/dönem (%60,9)

Dersin adı ve içeriği	Fikir birliği sağlanan tur	Fikir birliği yüzdesi
Tek örneklem testleri	Tur 2	87,5 (müfredatta olmalı)
Evren ortalamasının anlamlılık testi	Tur 2	81,3 (içerikte olmalı)
İşaret testi	Tur 2	75,0 (içerikte olmalı)
Evren oranının anlamlılık testi	Tur 2	75,0 (içerikte olmalı)
Tek boyutlu ki-kare	Tur 3	63,2 (içerikte olmalı - fikir birliği sağlanamadı)

Ders tipi: Zorunlu (%100,0)
Anlatım yöntemi ve saati: Teorik (%83,3) - 2 saat, Uygulama (%94,4) - 1 saat
Uygulama: İstatistik paket program ile (%88,2)
Sınıf düzeyi: 3. sınıf/dönem (%38,9)
Öneri doğrultusunda Tur 3 te yöneltilen;
Bu ders "Hipotez testleri-2" dersi içerisinde verilmeli midir sorusuna %69,2 "Hipotez testleri-2 ders konusu içerisinde verilmelidir." yanıtını verdi.

Dersin adı ve içeriği	Fikir birliği sağlanan tur	Fikir birliği yüzdesi
Bağımsız iki örneklem için hipotez testleri	Tur 1	95,8 (müfredatta olmalı)
Bağımlı ve bağımsız grup kavramları	Tur 1	100,0 (içerikte olmalı)
Student's t testi	Tur 1	100,0 (içerikte olmalı)
Mann-Whitney U testi	Tur 1	100,0 (içerikte olmalı)

Ders tipi: Zorunlu (%95,2)
Anlatım yöntemi ve saati: Teorik (%82,6) - 2 saat, Uygulama (%81,3) - 1 saat
Uygulama: İstatistik paket program ile (%100)
Sınıf düzeyi: 3. sınıf/dönem (%43,5)

Dersin adı ve içeriği	Fikir birliği sağlanan tur	Fikir birliği yüzdesi
Bağımsız ikiden çok örneklem için hipotez testleri	Tur 1	100,0 (müfredatta olmalı)
Tek yönlü varyans analizi (ANOVA)	Tur 1	100,0 (içerikte olmalı)
ANOVA'da çoklu karşılaştırmalar için Fisher's LSD testi	Tur 3	49,9 (içerikte olmamalı- fikir birliği sağlanamadı)
ANOVA'da çoklu karşılaştırmalar için Tukey testi	Tur 1	75,0 (içerikte olmalı)
ANOVA'da çoklu karşılaştırmalar için Bonferroni testi	Tur 1	83,3 (içerikte olmalı)
ANOVA'da çoklu karşılaştırmalar için Sidak testi	Tur 3	57,4 (içerikte olmamalı- fikir birliği sağlanamadı)
ANOVA'da çoklu karşılaştırmalar için Tukey's-b testi	Tur 3	53,4 (içerikte olmalı- fikir birliği sağlanamadı)
ANOVA'da çoklu karşılaştırmalar için Dunnet testi	Tur 3	50,2 (içerikte olmamalı- fikir birliği sağlanamadı)
ANOVA'da çoklu karşılaştırmalar için Duncan testi	Tur 3	53,4 (içerikte olmamalı- fikir birliği sağlanamadı)
Kruskal Wallis Varyans analizi	Tur 1	100,0 (içerikte olmalı)
Kruskal Wallis Varyans analizinde çoklu karşılaştırmalar için Dunn-Bonferroni testi	Tur 1	79,2 (içerikte olmalı)
Kruskal Wallis Varyans analizinde çoklu karşılaştırmalar için Bonferroni düzeltmeli Mann Whitney-U testi	Tur 1	75,0 (içerikte olmalı)

Ders tipi: Zorunlu (%91,3)
Anlatım yöntemi ve saati: Teorik (%79,2) - 2 saat, Uygulama (%95,8) - 2 saat
Uygulama: İstatistik paket program ile (%91,3)
Sınıf düzeyi: 3. sınıf/dönem (%45,8)
Öneri doğrultusunda Tur 2 de yöneltilen;
Makale incelemesi olmalı mı sorusuna %75,0 olmalı yanıtını verdi.

Dersin adı ve içeriği	Fikir birliği sağlanan tur	Fikir birliği yüzdesi
Bağımlı iki örneklem için hipotez testleri	Tur 1	100,0 (müfredatta olmalı)
Bağımlı ölçümlerden elde edilen veri yapısının tanımlanması	Tur 1	100,0 (içerikte olmalı)
İki eş arasındaki farkın anlamlılık testi	Tur 1	100,0 (içerikte olmalı)
Wilcoxon testi	Tur 1	100,0 (içerikte olmalı)

Ders tipi: Zorunlu (%95,5)
Anlatım yöntemi ve saati: Teorik (%82,6) - 2 saat, Uygulama (%100,0) - 1 saat
Uygulama: İstatistik paket program ile (%95,7)
Sınıf düzeyi: 3. sınıf/dönem (%47,8)
Öneri doğrultusunda Tur 2 de yöneltilen;

İki ders başlığı içerisinden "Bağımlı iki örneklem için hipotez testleri" başlığı seçildi (%68,8). (önceki İki bağımlı ölçüm arasındaki fark için hipotez testleri idi)

Bu ders "Bağımsız iki örneklem için hipotez testleri" dersi içerisinde verilmeli midir sorusuna %75,0 "Bağımsız iki örneklem için hipotez testleri dersi dışında ayrı bir derste anlatılmalıdır." yanıtını verdi.

Dersin adı ve içeriği	Fikir birliği sağlanan tur	Fikir birliği yüzdesi
İkiden çok bağımlı ölçüm arasındaki farkın testi için hipotez testleri	Tur 3	75,0 (müfredatta olmalı)
Tekrarlı ölçümlerde tek yönlü varyans analizi	Tur 2	81,3 (içerikte olmalı)
Friedman testi	Tur 2	75,0 (içerikte olmalı)

Ders tipi: Zorunlu (%75,0)
Anlatım yöntemi ve saati: Teorik (%81,3) - 1,5 saat, Uygulama (%100,0) - 1 saat
Uygulama: İstatistik paket program ile (%93,8)
Sınıf düzeyi: 3. sınıf/dönem (%56,7)
Öneri doğrultusunda Tur 2 de yöneltilen;

Bu ders "Bağımsız ikiden çok örneklem için hipotez testleri" dersi içerisinde verilmeli midir sorusuna %80,0 "Bağımsız ikiden çok örneklem için hipotez testleri dersi dışında ayrı bir derste anlatılmalıdır." yanıtını verdi.

Dersin adı ve içeriği	Fikir birliği sağlanan tur	Fikir birliği yüzdesi
Ki-kare bağımsızlık testleri	Tur 1	95,7 (müfredatta olmalı)
Pearson ki-kare testi	Tur 1	100,0 (içerikte olmalı)
2x2 tablolarda Yates düzeltmeli ki-kare	Tur 1	87,0 (içerikte olmalı)
2x2 tablolarda Fisher kesin ki-kare testi	Tur 1	100,0 (içerikte olmalı)
Olabilirlik oran testi	Tur 3	53,5 (içerikte olmamalı-fikir birliği sağlanamadı)
Risk ölçütleri	Tur 1	87,0 (içerikte olmalı)
Bağımsız iki oran arasındaki farkın anlamlılık testi	Tur 2	76,2 (içerikte olmalı)
Bağımlı iki oran arasındaki farkın anlamlılık testi	Tur 3	63,5 (içerikte olmalı-fikir birliği sağlanamadı)
McNemar testi	Tur 1	87,0 (içerikte olmalı)
rxc tablolarda bağımsızlık testleri	Tur 1	100,0 (içerikte olmalı)
Cochran Q testi (öneri sonucunda Tur 2 de eklendi)	Tur 3	73,3 (içerikte olmalı)

Ders tipi: Zorunlu (%95,5)
Anlatım yöntemi ve saati: Teorik (%81,8) - 2 saat, Uygulama (%100,0) - 2 saat
Uygulama: İstatistik paket program ile (%100,0)
Sınıf düzeyi: 3. sınıf/dönem (%42,9)
Öneri doğrultusunda Tur 2 de yöneltilen;
Bu ders "Hipotez testleri-2" dersi içerisinde verilmeli midir sorusuna %53,3 "Hipotez testleri-2 ders konusu içerisinde verilmelidir." yanıtını verdi.
Makale incelemesi olmalı mı sorusuna %81,3 olmalı yanıtını verdi.

Dersin adı ve içeriği	Fikir birliği sağlanan tur	Fikir birliği yüzdesi
Örnekleme	Tur 1	95,7 (müfredatta olmalı)
Örneklem seçiminin önemi	Tur 1	95,7 (içerikte olmalı)
Veri toplama yöntemleri ve veri toplarken dikkat edilmesi gereken noktalar	Tur 1	95,7 (içerikte olmalı)
Örneklemde temel kavramlar(örneklem çerçevesi, örneklem tasarımı, örneklem birimi, örneklem kesiri)	Tur 1	91,3 (içerikte olmalı)
Örneklem hataları (sistematik hata, örneklem hatası, temsil edebilirlik, bilgi yanlılığı, rasgele hata)	Tur 1	91,3 (içerikte olmalı)
Randomizasyon tanımı	Tur 1	95,7 (içerikte olmalı)
Randomizasyon yöntemleri	Tur 1	87,0 (içerikte olmalı)
Amaçlı örnekleme yöntemi	Tur 3	56,8 (içerikte olmamalı-fikir birliği sağlanamadı)
Kota örnekleme yöntemi	Tur 3	59,8 (içerikte olmamalı-fikir birliği sağlanamadı)
Saha örnekleme yöntemi	Tur 3	53,2 (içerikte olmamalı-fikir birliği sağlanamadı)
Kartopu örnekleme yöntemi	Tur 3	59,8 (içerikte olmamalı-fikir birliği sağlanamadı)
Monografi	Tur 3	80,0 (içerikte olmamalı)
Yakalama-Yeniden yakalama örnekleme yöntemi	Tur 3	80,0 (içerikte olmamalı)
Basit rasgele örnekleme yöntemi	Tur 1	87,0 (içerikte olmalı)
Sistematik örnekleme yöntemi	Tur 1	87,0 (içerikte olmalı)
Küme örnekleme yöntemi	Tur 1	82,6 (içerikte olmalı)
Tabakalı örnekleme yöntemi	Tur 1	82,6 (içerikte olmalı)
Çok aşamalı örnekleme yöntemi	Tur 3	56,7 (içerikte olmalı-fikir birliği sağlanamadı)
Ders tipi: Zorunlu (%90,9) Anlatım yöntemi ve saati: Teorik (%95,5) - 2 saat Sınıf düzeyi: 1. sınıf/dönem (%45,0)		

Dersin adı ve içeriği	Fikir birliği sağlanan tur	Fikir birliği yüzdesi
Örneklem büyüklüğü hesaplama	Tur 1	81,3 (müfredatta olmalı)
Etki genişliği tanımı	Tur 1	85,7 (içerikte olmalı)
Örneklem büyüklüğünü etkileyen faktörler	Tur 1	90,5 (içerikte olmalı)
Gücü etkileyen faktörler	Tur 1	90,5 (içerikte olmalı)
Tek örneklem ortalaması için örneklem büyüklüğü ve güç hesaplama	Tur 2	75,0 (içerikte olmalı)
Tek örneklem oranı için örneklem büyüklüğü ve güç hesaplama	Tur 2	81,3 (içerikte olmalı)
İki bağımsız örnekleme ilişkin parametrik testler için örnekle büyüklüğü ve güç hesaplama	Tur 1	76,2 (içerikte olmalı)
İki bağımsız örnekleme ilişkin parametrik olmayan testler için örneklem büyüklüğü ve güç hesaplama	Tur 3	60,0 (içerikte olmalı-fikir birliği sağlanamadı)
İkiden fazla bağımsız örnekleme ilişkin parametrik testler için örneklem büyüklüğü ve güç hesaplama	Tur 1	71,4 (içerikte olmalı)
İkiden fazla bağımsız örnekleme ilişkin parametrik olmayan testler için örneklem büyüklüğü ve güç hesaplama	Tur 3	53,2 (içerikte olmalı-fikir birliği sağlanamadı)
Parametrik bağımlı örneklem testleri için örneklem büyüklüğü ve güç hesaplama	Tur 3	66,3 (içerikte olmalı-fikir birliği sağlanamadı)
Parametrik olmayan bağımlı örneklem testleri için örneklem büyüklüğü ve güç hesaplama	Tur 3	53,5 (içerikte olmalı-fikir birliği sağlanamadı)
Korelasyon analizi için örneklem büyüklüğü ve güç hesaplama	Tur 2	68,8 (içerikte olmalı-fikir birliği sağlanamadı)
Basit doğrusal regresyon analizi için örneklem büyüklüğü ve güç hesaplama	Tur 3	56,7 (içerikte olmalı-fikir birliği sağlanamadı)

Ders tipi: Zorunlu (%81,3)
Anlatım yöntemi ve saati: Teorik (%94,1) - 2 saat, Uygulama (%88,2) - 2 saat
Uygulama: İstatistik paket program ile (%100,0)
Sınıf düzeyi: 3. term and 5. term (54.5%)

Dersin adı ve içeriği	Fikir birliği sağlanan tur	Fikir birliği yüzde
Korelasyon analizi	Tur 1	100,0 (müfredatta olmalı)
İlişki kavramı (Doğrusal/eğrisel ilişki, neden-sonuç ilişkisi, doğrudan/dolaylı ilişki)	Tur 1	100,0 (içerikte olmalı)
Pearson korelasyon katsayısı	Tur 1	100,0 (içerikte olmalı)
Spearman korelasyon katsayısı	Tur 1	100,0 (içerikte olmalı)
Kendall's tau-b korelasyon katsayısı	Tur 1	100,0 (içerikte olmalı)
Polyserial korelasyon katsayısı	Tur 1	77,3 (içerikte olmamalı)
Polikorik korelasyon katsayısı	Tur 1	77,3 (içerikte olmamalı)
Çift serili korelasyon katsayısı	Tur 1	72,7 (içerikte olmamalı)
Nokta çift serili korelasyon katsayısı	Tur 1	72,7 (içerikte olmamalı
Kısmi korelasyon	Tur 1	70,0 (içerikte olmalı)
Ders tipi: Zorunlu (%95,2) Anlatım yöntemi ve saati: Teorik (%87,0) – 2 saat, Uygulama (%91,3) - 1,5 saat Uygulama: İstatistik paket program ile (%100,0) Sınıf düzeyi: 3. sınıf/dönem (%43,5)		

Dersin adı ve içeriği	Fikir birliği sağlanan tur	Fikir birliği yüzdesi
Basit doğrusal regresyon analizi	Tur 1	100,0 (müfredatta olmalı)
Bağımlı ve bağımsız değişken kavramı	Tur 1	100,0 (içerikte olmalı)
Basit doğrusal regresyon modelinin matematiksel eşitliği ve regresyon katsayılarının tanımı	Tur 1	100,0 (içerikte olmalı)
Basit doğrusal regresyon analizi varsayımları	Tur 1	100,0 (içerikte olmalı)
Basit doğrusal regresyon modelinin yorumlanması (regresyon katsayılarının ve açıklayıcılık katsayısının yorumu)	Tur 1	100,0 (içerikte olmalı)
Basit doğrusal regresyon analizi içeren bir makalenin değerlendirilmesi	Tur 1	87,0 (içerikte olmalı)
Ders tipi: Zorunlu (%87,0) Anlatım yöntemi ve saati: Teorik (%91,3) - 2 saat, Uygulama (%87,0) - 1,5 saat Uygulama: İstatistik paket program ile (%100,0) Sınıf düzeyi: 3. sınıf/dönem (%39,1)		

Dersin adı ve içeriği	**Fikir birliği sağlanan tur**	**Fikir birliği yüzdesi**
Sağkalım analizi	Tur 2	75,0 (müfredatta olmalı)
Süre, olgu ve sansürlü gözlem tanımı	Tur 2	81,3 (içerikte olmalı)
Yaşam tablosu yöntemi	Tur 2	75,0 (içerikte olmalı)
Kaplan Meier yöntemi	Tur 2	81,3 (içerikte olmalı)
Mantel-Haenszel yöntemi ile sağkalım eğrilerinin karşılaştırılması	Tur 2	74,8 (içerikte olmalı)
Gehan yöntemi ile sağkalım eğrilerinin karşılaştırılması	Tur 3	53,2 (içerikte olmamalı-fikir birliği sağlanamadı)
Tarone ve Ware yöntemi ile sağkalım eğrilerinin karşılaştırılması	Tur 3	50,2 (içerikte olmamalı-fikir birliği sağlanamadı)
Prentice yöntemi ile sağkalım eğrilerinin karşılaştırılması	Tur 3	56,8 (içerikte olmamalı-fikir birliği sağlanamadı)

Ders tipi: Zorunlu (%50,0) Seçmeli (%50,0)
Anlatım yöntemi ve saati: Teorik (%93,3) - 2 saat, Uygulama (%100,0) - 2 saat
Uygulama: İstatistik paket program ile (%100,0)
Sınıf düzeyi: 3. sınıf/dönem (%60,0), 5. sınıf/dönem (%60,0)
Öneri doğrultusunda Tur 2 de yöneltilen;
Bu ders içeriğinde makale incelemesi olmalı mı sorusuna %81,3 oranında olmalı yanıtını verildi.

Dersin adı ve içeriği	Fikir birliği sağlanan tur	Fikir birliği yüzdesi
Tanı testlerinin değerlendirilmesinde kullanılan istatistiksel yöntemler	Tur 1	91,3 (müfredatta olmalı)
Altın standart, tanı testi ve tarama testi tanımları	Tur 1	95,5 (içerikte olmalı)
Tanı doğruluğunun tanımı	Tur 1	95,5 (içerikte olmalı)
Tanı testlerinin doğruluklarına yönelik çalışma tasarımı	Tur 1	90,9 (içerikte olmalı)
Tanı testlerinin doğruluklarına yönelik çalışmalar için gerekli standartlar	Tur 1	90,9 (içerikte olmalı)
İki sonuçlu tanı testleri için doğruluk ölçütlerinin hesaplanması (duyarlılık, seçicilik, yanlış negatif oran, yanlış pozitif oran)	Tur 1	95,5 (içerikte olmalı)
Pozitif ve negatif kestirim değerlerinin tanımları ve hesaplanması	Tur 1	95,5 (içerikte olmalı)
Pozitif ve negatif olabilirlik oranlarının tanımları ve hesaplanması	Tur 1	77,3 (içerikte olmalı)
Sürekli sonuçlu tanı testleri için doğruluk ölçütleri (İşlem karakteristiği eğrisinin tanımı ve matematiksel özellikleri)	Tur 1	95,5 (içerikte olmalı)
İşlem karakteristiği eğrisinin altında kalan alan ve yorumu	Tur 1	95,5 (içerikte olmalı)
Odds oranının tanımı ve hesaplanması	Tur 1	86,4 (içerikte olmalı)
Youden indeks tanımı ve hesaplanması	Tur 3	73,3 (içerikte olmalı)
Ders tipi: Seçmeli (%57,1) Anlatım yöntemi ve saati: Teorik (%95,2) - 2 saat, Uygulama (%85,2) - 2 saat Uygulama: İstatistik paket program ile (%100) Sınıf düzeyi: 5. sınıf/dönem (%50,0)		

Dersin adı ve içeriği	Fikir birliği sağlanan tur	Fikir birliği yüzdesi
Tek değişkenli ikili lojistik regresyon analizi	Tur 2	75,0 (müfredatta olmalı)
Bağımlı ve bağımsız değişken kavramı	Tur 1	85,0 (içerikte olmalı)
Tek değişkenli ikili lojistik regresyon modelinin matematiksel eşitliği ve regresyon katsayılarının tanımı	Tur 1	75,0 (içerikte olmalı)
Tek değişkenli ikili lojistik regresyon analizi varsayımları	Tur 1	80,0 (içerikte olmalı)
Tek değişkenli ikili lojistik regresyon modelinden elde edilen katsayıların yorumlanması	Tur 1	80,0 (içerikte olmalı)
Odds oranı ve güven aralığının yorumlanması	Tur 1	85,0 (içerikte olmalı)
Tek değişkenli ikili lojistik regresyon analizi içeren bir makalenin değerlendirilmesi	Tur 1	77,8 (içerikte olmalı)
Ders tipi: Seçmeli (%58,3) Anlatım yöntemi ve saati: Teorik (%92,3) - 2 saat, Uygulama (%100,0) – 2 saat Uygulama: İstatistik paket program ile (%100,0) Sınıf düzeyi: 5. sınıf/dönem (%50,0)		

Dersin adı ve içeriği	Fikir birliği sağlanan tur	Fikir birliği yüzdesi
Klinik denemeler	Tur 3	71,4 (müfredatta olmalı)
Klinik çalışma tanımı	Tur 2	73,3 (içerikte olmalı)
Klinik denemelerde faz tanımı	Tur 2	73,3 (içerikte olmalı)
Klinik çalışma raporlarının yapısı ve içeriği	Tur 2	80,0 (içerikte olmalı)
Bilimsel etik kavramı	Tur 2	75,0 (içerikte olmalı)
İlaç ve Biyolojik Ürünlerin Klinik Araştırmaları Hakkında Yönetmelik	Tur 3	74,8 (içerikte olmalı)
Etik Kurulların Yapısı	Tur 3	71,4 (içerikte olmalı)
Bir araştırma planının içeriğinde olması gereken konular (dahil edilme-hariç tutulma kriterleri, evren tanımı, örneklem seçimi, istatistiksel analiz planı, örneklem genişliği hesaplanması, kayıp veriyle başa çıkma...vb)	Tur 2	80,0 (içerikte olmalı)
Randomize klinik deneme (avantajları-dezavantajları)	Tur 2	80,0 (içerikte olmalı)
Paralel deneme düzenleri	Tur 2	73,3 (içerikte olmalı)
Çapraz deneme düzenleri	Tur 2	78,6 (içerikte olmalı)
Çok etkenli denemeler	Tur 3	71,8 (içerikte olmalı)
Çok merkezli denemeler	Tur 3	72,0 (içerikte olmalı)
Ardışık denemeler	Tur 3	65,5 (içerikte olmalı- fikir birliği sağlanamadı)
Paralel deneme düzenleri için örneklem büyüklüğü hesaplaması	Tur 3	49,9 (içerikte olmalı- fikir birliği sağlanamadı)
Çapraz deneme düzenleri için örneklem büyüklüğü hesaplaması	Tur 3	49,9 (içerikte olmalı- fikir birliği sağlanamadı)
Çok etkenli denemeler için örneklem büyüklüğü hesaplaması	Tur 3	53,9 (içerikte olmamalı- fikir birliği sağlanamadı)
Çok merkezli denemeler için örneklem büyüklüğü hesaplaması	Tur 3	53,9 (içerikte olmamalı- fikir birliği sağlanamadı)
Ardışık denemeler için örneklem büyüklüğü hesaplaması	Tur 3	53,9 (içerikte olmamalı- fikir birliği sağlanamadı)
Ders tipi: Zorunlu (%50,0) Seçmeli (%50,0) Anlatım yöntemi ve saati: Teorik (%93,3) - 3 saat, Uygulama (%40,0) - 1,5 saat Uygulama: İstatistik paket program ile (%100,0) Sınıf düzeyi: 5. sınıf/dönem (%70,0)		

Dersin adı ve içeriği	**Fikir birliği sağlanan tur**	**Fikir birliği yüzdesi**
Çoklu doğrusal regresyon analizi	Tur 2	75,0 (müfredatta olmalı)
Bağımlı ve bağımsız değişkenler kavramı	Tur 2	83,3 (içerikte olmalı)
Çoklu doğrusal regresyon modelinin matematiksel eşitliği ve regresyon katsayılarının tanımı	Tur 2	83,3 (içerikte olmalı)
Çoklu doğrusal regresyon analizinin temel varsayımları	Tur 2	83,3 (içerikte olmalı)
Çoklu doğrusal regresyon modelinin yorumlanması (regresyon katsayılarının test edilmesi, regresyon katsayıları için güven aralıkları, çoklu açıklayıcılık katsayısı)	Tur 2	83,3 (içerikte olmalı)
Gözlem uzaklıklarının değerlendirilmesi	Tur 2	83,3 (içerikte olmalı)
Aykırı gözlem tanımı	Tur 2	73,3 (içerikte olmalı)
Etkili gözlem tanımı	Tur 2	80,0 (içerikte olmalı)
Değişen varyanslılık sorununun tanımlanması	Tur 2	80,0 (içerikte olmalı)
Hataların normal dağılmama sorununun tanımlanması	Tur 2	80,0 (içerikte olmalı)
Çoklubağlantı tanımı	Tur 2	86,7 (içerikte olmalı)
Hataların ilişkili olma sorununun tanımlanması	Tur 2	80,0 (içerikte olmalı)
Dummy değişken kullanımı (indikatör kodlama yöntemi ile)	Tur 3	73,5 (içerikte olmalı)
Çoklu doğrusal regresyonda değişken seçimi yöntemleri	Tur 2	80,0 (içerikte olmalı)
Çoklu doğrusal regresyon analizi içeren bir makalenin değerlendirilmesi	Tur 2	86,7 (içerikte olmalı)
Ders tipi: Seçmeli (%72,7) Anlatım yöntemi ve saati: Teorik (%73,9) - 3 saat, Uygulama (%73,9) - 2 saat Uygulama: İstatistik paket program ile (%82,4) Sınıf düzeyi: 5. sınıf/dönem (%63,6)		

Dersin adı ve içeriği	Fikir birliği sağlanan tur	Fikir birliği yüzdesi
Çoklu ikili lojistik regresyon analizi	Tur 1	54,5 (müfredatta olmamalı-fikir birliği sağlanamadı)
Bağımlı ve bağımsız değişken kavramı	Tur 1	70,0 (içerikte olmalı)
Çoklu ikili lojistik regresyon modelinin matematiksel eşitliği ve regresyon katsayılarının tanımı	Tur 1	70,0 (içerikte olmalı)
Odds oranının tanımı ve matematiksel hesaplaması	Tur 1	71,0 (içerikte olmalı)
Çoklu ikili lojistik regresyon analizi varsayımları	Tur 1	71,0 (içerikte olmalı)
Etki karışımı, etkileşim ve sıfır frekans sorununun tanımlanması	Tur 1	70,0 (içerikte olmalı)
Çoklu ikili lojistik regresyonda model seçim stratejileri	Tur 1	71,0 (içerikte olmalı)
Çoklu ikili lojistik regresyonda model uyum iyiliği testleri	Tur 1	71,0 (içerikte olmalı)
Çoklu ikili lojistik regresyonda model yeterliliğinin değerlendirilmesi	Tur 1	68,8 (içerikte olmalı-fikir birliği sağlanamadı)
Çoklu ikili lojistik regresyon modelinin yorumlanması (regresyon katsayılarının test edilmesi, regresyon katsayıları için güven aralıkları, açıklayıcılık katsayıları)	Tur 1	73,0 (içerikte olmalı)
Çoklu ikili lojistik regresyon analizi içeren bir makalenin değerlendirilmesi	Tur 1	64,0 (içerikte olmalı-fikir birliği sağlanamadı)
Ders tipi: Seçmeli (%85,7) Anlatım yöntemi ve saati: Teorik (%92,7) - 2 saat, Uygulama (%92,9) - 2 saat Uygulama: İstatistik paket program ile (%100,0) Sınıf düzeyi: 5. sınıf/dönem (%63,6)		

SONUÇ

16 Biyoistatistik uzmanından alınan sonuçlarla tamamlanan bu çalışmada, katılımcılardan 4'ü profesör, 3'ü doçent, 5'i doktor öğretim üyesi, 4'ü araştırma görevlisi idi. Dolayısıyla her unvan grubundan benzer sayıda katılımcı sağlamak, çalışmanın geçerliliği açısından önemlidir. Delphi yönteminde katılımcıların görüşlerini yansıtacak nitelikte olması, yöntemi uygularken dikkat edilmesi gereken bir noktadır. Bu çalışmada da katılımcıların tamamının doktora alanının Biyoistatistik olması, yöntemden elde edilen sonuçların güvenirliği açısından değerli idi. Literatürde, Delphi yöntemi uygulanan çalışmalar için katılımcı sayısının 9-1000 arasında değişebileceğine değinilmiştir. Dolayısıyla 16 uzman tarafından tamamlanan bu çalışmanın sonuçlarının literatüre katkı sağlayacağı düşünülmektedir.

Türkiye'de Tıp Fakültelerinde verilen Biyoistatistik çekirdek müfredatını irdeleyen çalışma yoktur. Bu çalışmadan elde edilen sonuçların, Biyoistatistik müfredatı geliştirmede, konunun uzmanları için değerli olacağını düşünmekteyiz.

Etik kurul onayı:

Çalışmanın etik kurul onayı, Ankara Yıldırım Beyazıt Üniversitesi Sosyal ve Beşerî Bilimler Etik Kurulu'ndan alınmıştır.

Kaynaklar

1- Dalkey, N. C. "Studies In The Quality Of Life: Delphi and Decision Making". Lexington, MA: Lexington Books. (1972).
2-PA Harris, R Taylor, R Thielke, J Payne, N Gonzalez, JG. Conde, Research electronic data capture (REDCap) – A metadata-driven methodology and workflow process for providing translational research informatics support, J Biomed Inform. 2009 Apr;42(2):377-81.
3-PA Harris, R Taylor, BL Minor, V Elliott, M Fernandez, L O'Neal, L McLeod, G Delacqua, F Delacqua, J Kirby, SN Duda, REDCap Consortium, The REDCap consortium: Building an international community of software partners, J Biomed Inform. 2019 May 9 [doi: 10.1016/j.jbi.2019.103208]
4- Saekman, H. "Delphi Critique: Expert Opinion", Lexington, MA: Lexington Books. (l975).
5- Turoff, M. ve Hiltz, S. R. "Computer Based Delphi Processes" London: Kingsley. (2001).

T12

THE STANDARDIZATION OF THE CONTENT OF THE BIOSTATISTICS EDUCATION LECTURED AT THE UNDERGRADUATE TERM OF THE MEDICAL FACULTIES IN TURKEY: A DELPHI STUDY

YUKSEL S[1], ALKAN A[1], DEMIR P[1], SANISOGLU S.Y[1], ERCAN I[2], TURE M[3], KARAHAN S[4], ATES C[5], ELMALI F[6], ERDOGAN S[7], KAYA M.O[8], SAHIN B[9], YAVUZ Z[10], YILDIRIM D[7]

1. Ankara Yildirim Beyazit University, Medical Faculty, Department of Biostatistics, ANKARA
2. Bursa Uludag University, Medical Faculty, Department of Biostatistics, BURSA
3. Adnan Menderes University, Medical Faculty, Department of Biostatistics, AYDIN
4. Hacettepe University, Medical Faculty, Department of Biostatistics, ANKARA
5. Yuzuncu Yil University, Medical Faculty, Department of Biostatistics, VAN
6. Izmir Katip Celebi University, Medical Faculty, Department of Biostatistics, IZMIR
7. Mersin University, Medical Faculty, Department of Biostatistics, MERSIN
8. Firat University, Medical Faculty, Department of Biostatistics and Medical Informatics, ELAZIG
9. Pamukkale University, Medical Faculty, Department of Biostatistics, DENIZLI
10. Ankara University, Medical Faculty, Department of Biostatistics, ANKARA

Abstract

Aim: The purpose of this study was to provide the standardization of the content of Biostatistics curriculum for the undergraduate term of the medical faculties in Turkey by taking advantage of expert ideas about the name, syllabus, instruction method, type, duration of the course and the class level that the course should be given.

Method: An online questionnaire study was conducted using the Delphi methodology with three rounds. The course names and course contents in the questionnaire have been prepared by considering the course names and contents currently lectured at the different medical faculties. An invitation e-mail was sent to 153 instructors registered in the Biostatistics Association to participate in the study. The questionnaire of the first round (Round 1), a structured questionnaire including 24-course titles and 236-course contents, was sent to 33 instructors who satisfied the participation criteria and had agreed to participate to the study. The 19-course titles and 78-course contents of the Round 1 were re-questioned in Round 2 by 23 instructors who completed the Round 1 because there was no consensus amongst the participants in Round 1 for these titles and contents. The 16 participants answered the questionnaire in Round 2. In the third round, the 16-course titles and 61-course contents were asked again and the final results were obtained from 16 instructors.

Results: The consensus was reached in the 22 out of the 24-content strands. “The theoretical distributions” title was changed as “The normal distribution and its properties” and it was decided to lecture the one-sample tests under the hypothesis tests-2 course title. There was no consensus that the multiple logistic regression analysis should be in the curriculum. The 40 (16.9%) out of 236-course contents could not be reached the consensus. Course type and the level of class were

determined for other all courses except two courses and one course, respectively. The duration of theoretical and application were defined.

Conclusion: There is a lack in the literature for the researches about the content of the Biostatistics curriculum lectured in the Medical Faculties in Turkey. Clear content of the Biostatistics curriculum has resulted in this study. It is thought that the results of this study will contribute to improving the Biostatistics core curriculum. Since there is no consensus on some contents, there are issues that need to be examined in detail in Biostatistics education.

Keywords: Biostatistics curriculum, Medical education, Delphi method

INTRODUCTION

The studies about the development of the core curriculum for the Medical Schools have been increased recently. This is probably due to prevent the students from the overload of knowledge, because of the developing and increasing literature in medicine and health. Besides, by the specialization of the fields of expertise, standardization at the national and the international level has led to the need to update the medical education curriculum frequently. Core curriculum content has grown in importance, which has a great impact on the determination of the pathway a medical student will take during the period of graduate education.

The Biostatistics lectures in the medical schools of our country included theoretical statistical information about 20 years ago. In these periods when the use of computers was limited and there were no laboratories, teaching the students the basic statistical calculations and analysis manually, did not allow the students to apply theoretical statistical knowledge gains in the real data set. It was revealed that the effectiveness of the biostatistics education given during the undergraduate period should be questioned, since the students have forgotten the theoretical knowledge, which is not supported by the application, in their graduate education. Nowadays, with the help of developing technology and facilities, most departments support the theoretical education by the statistical package programs, while some departments use the previous teaching method. When considering the syllabus of the medical schools, each department follows different content list. Thus, there is a need for a national consensus in the biostatistics education of undergraduate medical schools. This study was designed and conducted to assist in the preparation of the biostatistics core curriculum.

METHOD

The Delphi method with three rounds was applied in this study. The Delphi method, which is considered as an instrument to provide consensus, aims that the individuals or groups who face a problem from a different point of view compromise without a meeting. This method can be used to get valid and reliable results in the case the decision could be influenced by strong individuals or groups (Turoff, 2001). The basic characteristics of the Delphi method are; secrecy in participation, sequential administration of the structured or semi-structured questionnaire, qualitative and quantitative analyses of group's response, providing feedback about the analyses results to the participants, giving participants the chance to reshape their ideas and make decisions at each stage, keeping consecutive applications until consensus occurs (Dalkey,1972; Sackman, 1975).

The data was collected and organized via REDCap, which was provided by Ankara Yildirim Beyazit University Statistics Consulting Application and Research Center, a partner of REDCap Consortium. REDCap (Research Electronic Data Capture) is a secure, web-based software platform designed to support data capture for research studies, providing 1) an intuitive interface for validated data capture; 2) audit trails for tracking data manipulation and export procedures; 3) automated export procedures for seamless data downloads to common statistical packages; and 4) procedures for data integration and interoperability with external sources. (Harris,2009; Harris 2019).

The Delphi method consists of a series of stages to reveal and examine the approaches and perspectives of the experts about the problem, and to come to an agreement. These stages of our study are as follows:

i-Selection of the participants:

The population has been defined as all instructors with an MSc or Ph.D. in Biostatistics in medical schools of Turkey to ensure the reliability of the Delphi method. The e-mail addresses of the participants were provided by the Biostatistics Association and an invitation was sent to 153 members of the association. The members who registered the form in the invitation were defined as the participants of the study. There were 41 instructors registered for the study, 33 out of them satisfying the inclusion criterion. These instructors were subjected to the Round 1 survey. The Round 2 survey was sent to 23 instructors who answered the Round 1 survey. 16 instructors were remaining in Round 3. The results were obtained by the answers of the 16 instructors.

ii-Preparation of the surveys:

The contents of 24 courses were composed by examining the curriculum of the medical schools where Biostatistics departments lecture at least two decades. The teaching method (*theoretical, application (via statistical package program), application (manual calculation-via Excel), flipped class, homework, other*), course type (*compulsory, elective*), duration of theoretical and application course, class level (*1.term, 2.term, 3.term, 4.term, 5.term, 6.term*) were questioned in addition. The surveys consisted of 1-9 Likert-type, multiple-answer and open-ended questions.

The survey of Round 1 included 24 courses and 236 contents. The minimum 70% of the agreement was defined as consensus. The consensus proportion was calculated as the percentage

of participants giving a response between 1-3 or 7-9 for 1-9 Likert-type questions. The most selected answer was defined as the consensus for multiple-answer questions. The agreement proportion was lower than 70% in 19 courses and 78 contents, which were questioned in Round 2. There were 16 courses and 61 contents requestioned in Round 3.

RESULTS:

The findings of each course are given in the following tables. The syllabus for which no consensus was reached is remarked by red.

Course title and syllabus	**Reached a consensus at**	**Consensus Proportion**
Introduction to Biostatistics	Round 1	93.5 (should be in the curriculum)
Definition of Statistics and Biostatistics Sciences, and their application fields	Round 1	92.9 (covered by this course)
Application fields of Biostatistics	Round 1	96.4 (covered by this course)
The definition of population and sample	Round 1	100.0 (covered by this course)
The definition of the descriptive statistics (data, subject, variable, population, sample, parameter, statistics); Types of data (ordinal, nominal, discrete, continuous)	Round 1	100.0 (covered by this course)
Definition of the inferential statistics	Round 1	89.3 (covered by this course)
Basic concepts of sampling (sampling frame, sampling design, sampling unit, sampling fraction)	Round 1	89.3 (covered by this course)
The importance of sampling	Round 1	92.9 (covered by this course)
Sampling errors (systematic error, sampling error, representability, information bias, random error)	Round 1	82.1 (covered by this course)
Non-Probability Sampling Methods (convenience sampling, snowball sampling, purposive sampling)	Round 3	80.0 (not covered by this course)
Probability Sampling Methods (simple random sampling, systematic sampling, stratified sampling, multiple stages sampling, clustered sampling)	Round 1	74.1 (covered by this course)
The importance of an appropriate sample size	Round 1	92.9 (covered by this course)

Course type: Compulsory (100.0%)
Teaching method and duration: Theoretical (89.7%) – 4.5 hours, Application (65.5%) – 3 hours
Application: via statistical package program (94.7%)
Class level: 1. term (55.2%)
According to the suggestions in Round 1;
"Introduction to Biostatistics" was selected among three titles in Round 2 (81.3%).

Course title and syllabus	Reached a consensus at	Consensus Proportion
Research Methods	Round 1	92.6 (should be in the curriculum)
Definition of a research	Round 1	96.3 (covered by this course)
Characteristics of a research question (ethical, significant, clear, feasible)	Round 1	92.6 (covered by this course)
Definition of a research hypothesis	Round 1	96.3 (covered by this course)
The place and importance of Biostatistics in research	Round 1	96.3 (covered by this course)
The null and alternative hypothesis	Round 1	100.0 (covered by this course)
Descriptive researches	Round 1	96.3 (covered by this course)
Cross-sectional researches	Round 1	96.3 (covered by this course)
Case-control researches	Round 1	96.2 (covered by this course)
Cohort researches	Round 1	92.6 (covered by this course)
Repeated measurements researches	Round 1	88.9 (covered by this course)
Experimental researches	Round 1	96.3 (covered by this course)
Diagnostic accuracy researches	Round 1	81.5 (covered by this course)
Validity and reliability researches	Round 3	70.0 (covered by this course)
Agreement studies	Round 3	59.8 (not covered by this course-no consensus)
Systematic review	Round 3	60.0 (not covered by this course-no consensus)
Meta-analysis	Round 3	26.7 (covered or not covered by this course-no consensus)
Qualitative researches	Round 3	63.3 (not covered by this course-no consensus)
Error sources in researches	Round 1	85.2 (covered by this course)

Course type: Compulsory (90.9%)
Teaching method and duration: Theoretical (84.0%) – 4 hours, Application (32.0%) – 3.5 hours
Application: via statistical package program (100.0%)
Class level: 3. term (61.5%)
According to the suggestions in Round 1;
Field application was selected as a teaching method in Round 2 (43.8%)
"Research Methods" was selected among three titles in Round 2 (75.0%).

Course title and syllabus	**Reached a consensus at**	**Consensus Proportion**
Basic definition of data and data types	Round 1	92.3 (should be in the curriculum)
Concepts of subject, variable, measurement and data	Round 1	96.2 (covered by this course)
Definition of parameter and statistics	Round 1	92.3 (covered by this course)
Ordinal categorical variable	Round 1	92.3 (covered by this course)
Nominal categorical variable	Round 1	92.3 (covered by this course)
Discrete numeric variable	Round 1	92.3 (covered by this course)
Continuous numeric variable	Round 1	92.3 (covered by this course)
Course type: Compulsory (91.7%) Teaching method and duration: Theoretical (100.0%) – 2 hours, Application (25.0%) – 1 hours Application: via statistical package program (100.0%) Class level: 1. term (50.0%) According to the suggestions in Round 1; This course was selected as it should be covered in “Introduction to Biostatistics” in Round 2 (68.8%)		

Course title and syllabus	Reached a consensus at	Consensus Proportion
Descriptive statistics	Round 1	96.0 (should be in the curriculum)
Mean	Round 1	100.0 (covered by this course)
Median	Round 1	100.0 (covered by this course)
Mode	Round 1	84.0 (covered by this course)
Proportion	Round 1	96.0 (covered by this course)
Geometric mean	Round 3	56.8 (not covered by this course-no consensus)
Harmonic mean	Round 3	70.2 (not covered by this course)
Quartiles, percentages	Round 1	92.0 (covered by this course)
Range	Round 1	96.0 (covered by this course)
Standard deviation	Round 1	100.0 (covered by this course)
Variance	Round 1	96.0 (covered by this course)
Interquartile range	Round 1	96.0 (covered by this course)
Quartile deviation	Round 3	33.3 (not covered by this course-no consensus)
Coefficient of variation	Round 1	80.0 (covered by this course)

Course type: Compulsory (91.3%)
Teaching method and duration: Theoretical (87.5%) – 2 hours, Application (87.5%) – 1.5 hours
Application: via statistical package program (95.2%)
Class level: 1. term (54.2%)
According to the suggestions in Round 1;
"Descriptive Statistics" was selected among two titles in Round 2 (87.5%).
This course was selected to be in the curriculum as a separate course rather than covered by "Introduction to Biostatistics" in Round 2 (73.3%).
68.8% of participants decided there should be article review in the lecture.

Course title and syllabus	Reached a consensus at	Consensus Proportion
Normal distribution and its characteristics	Round 1	73.9 (should be in the curriculum)
Binomial distribution	Round 3	66.8 (covered by this course-no consensus)
Poisson distribution	Round 3	63.2 (covered by this course-no consensus)
Exponential distribution	Round 3	63.3 (not covered by this course-no consensus)
Normal distribution	Round 1	100.0 (covered by this course)
Standard normal distribution	Round 1	95.7 (covered by this course)
Characteristics of normal distribution	Round 1	100.0 (covered by this course)
The importance of normal distribution in Statistics	Round 1	95.7 (covered by this course)
Examining the distribution of a numeric variable by plots	Round 1	95.7 (covered by this course)
Examining the distribution of a numeric variable by descriptive statistics	Round 1	95.7 (covered by this course)

Course type: Compulsory (100.0%)
Teaching method and duration: Theoretical (94.1%) – 2 hours, Application (70.6%) – 1.5 hours
Application: via statistical package program (83.3%)
Class level: 1. term (70.6%)
According to the suggestions in Round 1;
"Normal distribution and its characteristics" was selected among two titles in Round 2 (68.8%). (The course title was Theoretical Distributions)
This course was selected to be in the curriculum as a separate course, rather than covered by "Introduction to Biostatistics" in Round 2 (68.8%).

Course title and syllabus	**Reached a consensus at**	**Consensus Proportion**
Hypothesis testing-1	Round 1	91.7 (should be in the curriculum)
Sampling distribution	Round 1	83.3 (covered by this course)
Central limit theorem	Round 3	66.7 (covered by this course-no consensus)
The concept of estimation	Round 2	87.5 (covered by this course)
Point and interval estimation	Round 1	75.0 (covered by this course)
Confidence interval	Round 1	83.3 (covered by this course)

Course type: Compulsory (85.7%)
Teaching method and duration: Theoretical (95.5%) – 2 hours, Application (40.9%) – 1.5 hours
Application: via statistical package program (88.9%)
Class level: 1. term (47.6%)
According to the suggestions in Round 1;
"Hypothesis testing-1" was selected among two titles in Round 2 (87.5%).
This course was selected to be in the curriculum as a separate course, rather than covered by "Introduction to Biostatistics" in Round 2 (75.0%).

Course title and syllabus	Reached a consensus at	Consensus Proportion
Hypothesis testing-2	Round 1	100.0 (should be in the curriculum)
The definition of the null and alternative hypotheses	Round 1	100.0 (covered by this course)
Type-I error	Round 1	100.0 (covered by this course)
Type-II error	Round 1	95.8 (covered by this course)
The p-value and statistical decision	Round 1	100.0 (covered by this course)
The concept of parametric and non-parametric tests and their assumptions	Round 1	100.0 (covered by this course)
Testing normality of a continuous variable	Round 1	100.0 (covered by this course)
Testing homogeneity of variances for at least two groups	Round 1	91.7 (covered by this course)
Course type: Compulsory (95.7%) Teaching method and duration: Theoretical (95.8%) - 2 hours, Application (87.5%) – 1 hour Application: via statistical package program (85.7%) Class level: 3. term (45.8%)		

Course title and syllabus	Reached a consensus at	Consensus Proportion
Summarizing data by appropriate descriptive statistics, tables and plots	Round 1	95.8 (should be in the curriculum)
Summarizing normally distributed data sets by appropriate descriptive statistics	Round 1	100.0 (covered by this course)
Summarizing non-normally distributed data sets by appropriate descriptive statistics	Round 1	100.0 (covered by this course)
Frequency tables	Round 1	91.7 (covered by this course)
Cross tables	Round 1	100.0 (covered by this course)
Nested tables	Round 3	50.0 (not covered by this course-no consensus)
Bar graph	Round 1	87.0 (covered by this course)
Pie graph	Round 1	87.5 (covered by this course)
Histogram	Round 1	82.6 (covered by this course)
Distribution polygon	Round 3	53.2 (not covered by this course-no consensus)
Boxplot	Round 1	87.5 (covered by this course)
Steam and leaf graph	Round 3	70.2 (not covered by this course)
Error plot	Round 1	87.5 (covered by this course)
Scatter plot	Round 1	91.7 (covered by this course)
Population pyramid	Round 3	73.3 (not covered by this course)
Forest plot	Round 3	83.2 (not covered by this course)
Funnel plot	Round 3	83.2 (not covered by this course)
Course type: Compulsory (90.9%) Teaching method and duration: Theoretical (82.6%) – 2 hours, Application (95.7%) – 2 hours Application: via statistical package program (90.9%) Class level: 1. term (60.9%)		

Course title and syllabus	Reached a consensus at	Consensus Proportion
One sample tests	Round 2	87.5 (should be in the curriculum)
Significance test for a population mean	Round 2	81.3 (covered by this course)
Sign test	Round 2	75.0 (covered by this course)
Significance test for a proportion	Round 2	75.0 (covered by this course)
One-way chi-square test	Round 3	63.2 (covered by this course-no consensus)

Course type: Compulsory (100.0%)
Teaching method and duration: Theoretical (83.3%) – 2 hours, Application (94.4%) – 1 hours
Application: via statistical package program (88.2%)
Class level: 3. term (38.9%)
According to the suggestions in Round 2;
This course was selected to be in the curriculum covered by "Hypothesis testing-2" in Round 3 (69.2%).

Course title and syllabus	Reached a consensus at	Consensus Proportion
Hypothesis tests for two independent samples	Round 1	95.8 (should be in the curriculum)
The concepts of dependent and independent groups	Round 1	100.0 (covered by this course)
Student's t-test	Round 1	100.0 (covered by this course)
Mann-Whitney U test	Round 1	100.0 (covered by this course)

Course type: Compulsory (95.2%)
Teaching method and duration: Theoretical (82.6%) – 2 hours. Application (91.3%) – 1 hours
Application: via statistical package program (100.0%)
Class level: 3. term (43.5%)

Course title and syllabus	Reached a consensus at	Consensus Proportion
Hypothesis tests for more than two independent samples	Round 1	100.0 (should be in the curriculum)
One-way analysis of variance (ANOVA)	Round 1	100.0 (covered by this course)
Fisher's LSD test for multiple comparisons	Round 3	49.9 (not covered by this course-no consensus)
Tukey test for multiple comparisons	Round 1	75.0 (covered by this course)
Bonferroni test for multiple comparisons	Round 1	83.3 (covered by this course)
Sidak test for multiple comparisons	Round 3	57.4 (not covered by this course-no consensus)
Tukey's-b test for multiple comparisons	Round 3	53.4 (covered by this course-no consensus)
Dunnet test for multiple comparisons	Round 3	50.2 (not covered by this course-no consensus)
Duncan test for multiple comparisons	Round 3	53.4 (not covered by this course-no consensus)
Kruskal Wallis Variance Analysis	Round 1	100.0 (covered by this course)
Dunn-Bonferroni test for multiple comparisons in Kruskal Wallis Variance Analysis	Round 1	79.2 (covered by this course)
Mann Whitney-U test with Bonferroni correction for multiple comparisons in Kruskal Wallis Variance Analysis	Round 1	75.0 (covered by this course)

Course type: Compulsory (91.3%)
Teaching method and duration: Theoretical (79.2%) – 2 hours, Application (95.8%) – 2 hours
Application: via statistical package program (91.3%)
Class level: 3. term (45.8%)
According to the suggestions in Round 1;
75.0% of participants decided to should be article review in the lecture.

Course title and syllabus	Reached a consensus at	Consensus Proportion
Hypothesis tests for two dependent samples	Round 1	100.0 (should be in the curriculum)
Definition of data design of dependent measurements	Round 1	100.0 (covered by this course)
Paired sample t test	Round 1	100.0 (covered by this course)
Wilcoxon test	Round 1	100.0 (covered by this course)

Course type: Compulsory (95.5%)
Teaching method and duration: Theoretical (82.6%) – 2 hours, Application (100.0%) – 1 hours
Application: via statistical package program (95.7%)
Class level: 3. term (47.8%)
According to the suggestions in Round 1;

"Hypothesis tests for two dependent samples" was selected among two titles in Round 2 (68.8%). (The course title was "Hypothesis test for the difference between two dependent measurements")

This course was selected to be in the curriculum as a separate course, rather than covered by "Hypothesis tests for two independent samples" in Round 2 (75.0%).

Course title and syllabus	Reached a consensus at	Consensus Proportion
Hypothesis tests for the difference between more than two dependent samples	Round 3	75.0 (should be in the curriculum)
One-way repeated measures analysis of variance	Round 2	81.3 (covered by this course)
Friedman test	Round 2	75.0 (covered by this course)

Course type: Compulsory (75.0%)
Teaching method and duration: Theoretical (81.3%) – 1.5 hours, Application (100.0%) – 1 hours
Application: via statistical package program (93.8%)
Class level: 3. term (66.7%)
According to the suggestions in Round 1;

This course was selected to be in the curriculum as a separate course, rather than covered by "Hypothesis tests for more than two independent samples" in Round 2 (80.0%).

Course title and syllabus	Reached a consensus at	Consensus Proportion
Chi-square tests for independence	Round 1	95.7 (should be in the curriculum)
Pearson chi-square test	Round 1	100.0 (covered by this course)
Yates corrected chi-square test for 2x2 tables	Round 1	87.0 (covered by this course)
Fisher exact test for 2x2 tables	Round 1	100.0 (covered by this course)
Likelihood ratio test	Round 3	53.5 (not covered by this course-no consensus)
Risk measures	Round 1	87.0 (covered by this course)
Significance test for the difference between two independent proportion	Round 2	76.2 (covered by this course)
Significance test for the difference between two dependent proportion	Round 3	63.5 (covered by this course-no consensus)
Mc-Nemar test	Round 1	87.0 (covered by this course)
Independence tests for rxc tables	Round 1	100.0 (covered by this course)
Cochran Q test (added in Round 2 w.r.t a suggestion in Round 1)	Round 3	73.3 (covered by this course)

Course type: Compulsory (95.5%)
Teaching method and duration: Theoretical (81.8%) – 2 hours, Application (100.0%) – 2 hours
Application: via statistical package program (100.0%)
Class level: 3. term (42.9%)
According to the suggestions in Round 1;
This course was selected to be in the curriculum covered by “Hypothesis testing-2” in Round 2 (53.3%).
81.3% of participants decided to should be article review in the lecture.

Course title and syllabus	Reached a consensus at	Consensus Proportion
Sampling	Round 1	95.7 (should be in the curriculum)
The importance of sample selection	Round 1	95.7 (covered by this course)
Data collection methods and points to take into account	Round 1	95.7 (covered by this course)
Basic Concept in sampling (Sampling frame, sample design, sampling scheme, sampling unit, sampling fraction)	Round 1	91.3 (covered by this course)
Sampling errors (systematic error, sampling error, representativeness, information bias, precision, random error)	Round 1	91.3 (covered by this course)
The definition of randomization	Round 1	95.7 (covered by this course)
Randomization methods	Round 1	87.0 (covered by this course)
Purposive sampling method	Round 3	56.8 (not covered by this course-no consensus)
Quota sampling method	Round 3	59.8 (not covered by this course-no consensus)
Field sampling method	Round 3	53.2 (not covered by this course-no consensus)
Snowball sampling method	Round 3	59.8 (not covered by this course-no consensus)
Monography	Round 3	80.0 (not covered by this course)
Capture- recapture sampling method	Round 3	80.0 (not covered by this course)
Simple random sampling method	Round 1	87.0 (covered by this course)
Systematic sampling method	Round 1	87.0 (covered by this course)
Clustered sampling method	Round 1	82.6 (covered by this course)
Stratified sampling method	Round 1	82.6 (covered by this course)
Multiphase sampling method	Round 3	56.7 (covered by this course-no consensus)
Course type: Compulsory (90.9%) Teaching method and duration: Theoretical (95.5%) – 2 hours Class level: 1. term (45.0%)		

Course title and syllabus	Reached a consensus at	Consensus Proportion
Sample size calculation	Round 1	81.3 (should be in the curriculum)
Definition of the effect size	Round 1	85.7 (covered by this course)
Factors effecting sample size	Round 1	90.5 (covered by this course)
Factors effecting power	Round 1	90.5 (covered by this course)
Sample size and power calculation for one sample mean	Round 2	75.0 (covered by this course)
Sample size and power calculation for one sample proportion	Round 2	81.3 (covered by this course)
Sample size and power calculation for parametric tests with two independent samples	Round 1	76.2 (covered by this course)
Sample size and power calculation for non-parametric tests with two independent samples	Round 3	60.0 (covered by this course-no consensus)
Sample size and power calculation for parametric tests with more than two independent samples	Round 1	71.4 (covered by this course)
Sample size and power calculation for non-parametric tests with more than two independent samples	Round 3	53.2 (covered by this course-no consensus)
Sample size and power calculation for parametric dependent sample tests	Round 3	66.3 (covered by this course-no consensus)
Sample size and power calculation for non-parametric dependent sample tests	Round 3	53.5 (covered by this course-no consensus)
Sample size and power calculation for correlation analysis	Round 2	68.8 (covered by this course-no consensus)
Sample size and power calculation for simple linear regression analysis	Round 3	56.7 (covered by this course-no consensus)

Course type: Compulsory (81.3%)
Teaching method and duration: Theoretical (94.1%) – 2 hours, Application (88.2%) – 2 hours
Application: via statistical package program (100.0%)
Class level: 3. term and 5. term (54.5%)

Course title and syllabus	Reached a consensus at	Consensus Proportion
Correlation analysis	Round 1	100.0 (should be in the curriculum)
The concept of relation (linear/curvature relation, cause and effect relation, direct/indirect relation)	Round 1	100.0 (covered by this course)
Pearson correlation coefficient	Round 1	100.0 (covered by this course)
Spearman correlation coefficient	Round 1	100.0 (covered by this course)
Kendall's tau-b correlation coefficient	Round 1	100.0 (covered by this course)
Polyserial correlation coefficient	Round 1	77.3 (not covered by this course)
Polychoric correlation coefficient	Round 1	77.3 (not covered by this course)
Biserial correlation coefficient	Round 1	72.7 (not covered by this course)
Point biserial correlation coefficient	Round 1	72.7 (not covered by this course)
Partial correlation	Round 1	70.0 (covered by this course)
Course type: Compulsory (95.2%) Teaching method and duration: Theoretical (87.0%) – 2 hours, Application (91.3%) – 1.5 hours Application: via statistical package program (100.0%) Class level: 3. term (43.5%)		

Course title and syllabus	Reached a consensus at	Consensus Proportion
Simple linear regression analysis	Round 1	100.0 (should be in the curriculum)
Dependent and independent variables	Round 1	100.0 (covered by this course)
The simple linear regression equation and definition of the regression coefficients	Round 1	100.0 (covered by this course)
Assumptions of simple linear regression analysis	Round 1	100.0 (covered by this course)
Interpretation of simple linear regression model (interpretation of regression coefficients and the coefficient of determination)	Round 1	100.0 (covered by this course)
Review of an article containing the simple linear regression analysis	Round 1	87.0 (covered by this course)
Course type: Compulsory (87.0%) Teaching method and duration: Theoretical (91.3%) – 2 hours, Application (87.0%) – 1.5 hours Application: via statistical package program (100.0%) Class level: 3. term (39.1%)		

Course title and syllabus	Reached a consensus at	Consensus Proportion
Survival analysis	Round 2	75.0 (should be in the curriculum)
Definition of time, case and censored data	Round 2	81.3 (covered by this course)
Life table method	Round 2	75.0 (covered by this course)
Kaplan Meier method	Round 2	81.3 (covered by this course)
Mantel-Haenszel method for the comparison of survival curves	Round 2	74.8 (covered by this course)
Gehan method for the comparison of survival curves	Round 3	53.2 (not covered by this course-no consensus)
Tarone and Ware method for the comparison of survival curves	Round 3	50.2 (not covered by this course-no consensus)
Prentice method for the comparison of survival curves	Round 3	56.8 (not covered by this course-no consensus)

Course type: Compulsory (50.0%) Elective (50.0%)
Teaching method and duration: Theoretical (93.3%) – 2 hours, Application (100.0%) – 2 hours
Application: via statistical package program (100.0%)
Class level: 3. term (60.0%), 5. term (60.0%)
According to the suggestions in Round 1;
81.3% of participants decided to should be article review in the lecture.

Course title and syllabus	Reached a consensus at	Consensus Proportion
Statistical methods in evaluation of diagnostic tests	Round 1	91.3 (should be in the curriculum)
Definition of gold standard test, diagnostic test and screening test	Round 1	95.5 (covered by this course)
Definition of diagnostic accuracy	Round 1	95.5 (covered by this course)
Study design for assessing the accuracy of diagnostic tests	Round 1	90.9 (covered by this course)
Standards for reporting of diagnostic accuracy studies	Round 1	90.9 (covered by this course)
Calculating of the accuracy criteria for the diagnostic tests with binary response (Sensitivity, Specificity, false negative rate and false positive rate)	Round 1	95.5 (covered by this course)
The definition and calculation of positive predictive value and negative predictive value	Round 1	95.5 (covered by this course)
The definition and calculation of the positive and negative likelihood ratios	Round 1	77.3 (covered by this course)
The accuracy criteria for the diagnostic tests with the continuous response (definition of the item characteristic curve and its mathematical properties)	Round 1	95.5 (covered by this course)
Area under curve and its interpretation	Round 1	95.5 (covered by this course)
Definition and calculation of Odds ratio	Round 1	86.4 (covered by this course)
Definition and calculation of Youden index	Round 3	73.3 (covered by this course)

Course type: Elective (57.1%)
Teaching method and duration: Theoretical (95.2%) – 2 hours, Application (95.2%) – 2 hours
Application: via statistical package program (100.0%)
Class level: 5. term (50.0%)

Course title and syllabus	Reached a consensus at	Consensus Proportion
Univariate binary logistic regression analysis	Round 2	75.0 (should be in the curriculum)
Dependent and independent variables	Round 1	85.0 (covered by this course)
The univariate binary logistic regression equation and definition of the regression coefficients	Round 1	75.0 (covered by this course)
Assumptions of univariate binary logistic regression analysis	Round 1	80.0 (covered by this course)
Interpretation of coefficients obtained from univariate binary logistic regression model	Round 1	80.0 (covered by this course)
Interpretation of the Odds ratio and confidence interval	Round 1	85.0 (covered by this course)
Review of an article containing the univariate binary logistic regression analysis	Round 1	77.8 (covered by this course)
Course type: Elective (58.3%) Teaching method and duration: Theoretical (92.3%) – 2 hours, Application (100.0%) – 2 hours Application: via statistical package program (100.0%) Class level: 5. term (50.0%)		

Course title and syllabus	Reached a consensus at	Consensus Proportion
Clinical Trials	Round 3	71.4 (should be in the curriculum)
Definition of clinical research	Round 2	73.3 (covered by this course)
Definition of clinical trial phases	Round 2	73.3 (covered by this course)
Structure and content of clinical study reports	Round 2	73.3 (covered by this course)
Concept of scientific ethics	Round 2	80.0 (covered by this course)
Regulation on Clinical Research of Pharmaceutical and Biological Products	Round 3	74.8 (covered by this course)
Structure of ethics committee	Round 3	71.4 (covered by this course)
The contents of a clinical trial protocol (inclusion and exclusion criteria, definition of population, selecting the sample, the plan of statistical analysis, sample size calculation, dealing with missing data…etc.)	Round 2	80.0 (covered by this course)
Randomized clinical trials (the advantages and disadvantages)	Round 2	80.0 (covered by this course)
Parallel group design	Round 2	73.3 (covered by this course)
Cross-over design	Round 2	78.6 (covered by this course)
Factorial design	Round 3	71.8 (covered by this course)
Multicenter clinical trial	Round 3	72.0 (covered by this course)
Sequential trials	Round 3	65.5 (covered by this course-no consensus)
Sample size calculation for parallel group design	Round 3	49.9 (covered by this course-no consensus)
Sample size calculation for cross-over design	Round 3	49.9 (covered by this course-no consensus)
Sample size calculation for factorial design	Round 3	53.9 (not covered by this course-no consensus)
Sample size calculation for multicenter clinical trial	Round 3	53.9 (not covered by this course-no consensus)
Sample size calculation for sequential trials	Round 3	53.9 (not covered by this course-no consensus)
Course type: Compulsory (50.0%) Elective (50.0%) Teaching method and duration: Theoretical (93.3%) – 3 hours, Application (40.0%) – 1.5 hours Application: via statistical package program (100.0%) Class level: 5. term (70.0%)		

Course title and syllabus	Reached a consensus at	Consensus Proportion
Multiple Linear Regression Analysis	Round 2	75.0 (should be in the curriculum)
Definition of the dependent and independent variables	Round 2	83.3 (covered by this course)
The multiple linear regression equation and definition of the regression coefficients	Round 2	83.3 (covered by this course)
Assumptions of multiple linear regression analysis	Round 2	83.3 (covered by this course)
Interpretation of multiple linear regression model (testing of regression coefficients, confidence intervals for regression coefficients, the coefficient of multiple determination)	Round 2	83.3 (covered by this course)
Evaluation of observation distance	Round 2	73.3 (covered by this course)
Definition of outlier observation	Round 2	80.0 (covered by this course)
Definition of influential observation	Round 2	80.0 (covered by this course)
Definition of heteroscedasticity	Round 2	80.0 (covered by this course)
Checking *Normality* of *Residuals*	Round 2	80.0 (covered by this course)
Definition of multicollinearity	Round 2	86.7 (covered by this course)
Testing of autocorrelation	Round 2	80.0 (covered by this course)
Use of dummy variable (Indicator coding)	Round 3	73.5 (covered by this course)
Variable selection methods for multiple linear regression model	Round 2	80.0 (covered by this course)
Review of an article containing the multiple linear regression analysis	Round 2	86.7 (covered by this course)
Course type: Elective (72.7%) Teaching method and duration: Theoretical (92.9%) – 3 hours, Application (92.9%) – 2 hours Application: via statistical package program (100.0%) Class level: 5. term (63.6%)		

Course title and syllabus	Reached a consensus at	Consensus Proportion
Multiple binary logistic regression analysis	Round 1	54.5 (shouldn't be in the curriculum-no consensus)
Definition of the dependent and independent variables	Round 1	70.0 (covered by this course)
The multiple binary logistic regression equation and definition of the regression coefficients	Round 1	70.0 (covered by this course)
Definition of Odds ratio and its calculation	Round 1	71.0 (covered by this course)
Assumptions of multiple binary logistic regression analysis	Round 1	71.0 (covered by this course)
Definition of confounder, interaction and missing cell problem	Round 1	70.0 (covered by this course)
Model building strategies for the multiple binary logistic regression analysis	Round 1	71.0 (covered by this course)
Tests for the goodness of fit in the multiple binary logistic regression analysis	Round 1	71.0 (covered by this course)
Assessing the fit of the model for multiple binary logistic regression analysis	Round 1	68.8 (covered by this course-no consensus)
Interpretation of multiple binary logistic regression model (testing of regression coefficients, confidence intervals for regression coefficients, the coefficient of determination)	Round 1	73.0 (covered by this course)
Review of an article containing the multiple binary logistic regression analysis	Round 1	64.0 (covered by this course-no consensus)
Course type: Elective (85.7%) Teaching method and duration: Theoretical (100.0%) – 2 hours, Application (87.5%) – 2 hours Application: via statistical package program (100.0%) Class level: 5. term (62.5%)		

CONCLUSION

In this study completed with the results of 16-Biostatisticians, there were 4-Professors, 3-Associate professors, 5-Assistant professors, and 4-research assistants. Consequently, providing a similar number of experts from each title is important for the validity of the study. It is crucial to reflect the opinions of the participants in the Delphi method. All participants had/were proceeding a Ph.D. degree in Biostatistics. This was substantial in terms of the reliability of the results obtained from the Delphi method. In the literature, it is addressed that the number of participants can vary between 9 and 1000 for the studies applied to the Delphi technique.

Therefore, the results of this study which was completed by 16 experts are thought to contribute to the literature.

There is no study conducted on the content of the Biostatistics curriculum lectured in the Medical Faculties in Turkey. We think that the results of this Delphi study will be valuable for the experts to develop the Biostatistics curriculum.

Ethics committee approval:

Ethics committee approval for this study was received from Ankara Yildirim Beyazit University, Social and Human Sciences Ethics Committee.

References:

1- Dalkey, N. C. "Studies in The Quality of Life: Delphi and Decision Making". Lexington, MA: Lexington Books. (1972).
2- PA Harris, R Taylor, R Thielke, J Payne, N Gonzalez, JG. Conde, Research electronic data capture (REDCap) – A metadata-driven methodology and workflow process for providing translational research informatics support, J Biomed Inform. 2009 Apr;42(2):377-81.
3- PA Harris, R Taylor, BL Minor, V Elliott, M Fernandez, L O'Neal, L McLeod, G Delacqua, F Delacqua, J Kirby, SN Duda, REDCap Consortium, The REDCap consortium: Building an international community of software partners, J Biomed Inform. 2019 May 9 [doi: 10.1016/j.jbi.2019.103208]
4- Saekman, H. "Delphi Critique: Expert Opinion", Lexington, MA: Lexington Books. (l975).
5- Turoff, M. and Hiltz, S. R. "Computer Based Delphi Processes" London: Kingsley. (2001).

T13

Deep Learning Based Imbalanced Data Classification for Drug Discovery

Selçuk KORKMAZ[1]

[1]*Trakya University Faculty of Medicine Department of Biostatistics and Medical Informatics, Edirne, Turkey*

selcukorkmaz@gmail.com

1. INTRODUCTION

In the early phase of drug discovery studies, there are huge amount of small chemical molecules to be screened in order to detect compounds that show some kind of activity. High-throughput screening (HTS) is a popular method to be used to test large number of small compounds for their activity against specific biological target, such as receptor or enzyme [1]. The datasets that are obtained by using HTS method are uploaded to PubChem database as biological assays (bioassays). The PubChem (https://pubchem.ncbi.nlm.nih.gov/) is a public repository for chemical compounds and their biological activities [2] and is hosted by National Center for Biotechnology Information (NCBI). As of July 2019, the database contains approximately 96 million unique compounds and 1.3 million bioassays and the number of compounds grows exponentially every year. The huge number of experimental data in the PubChem repository provides an important resource for drug discovery studies and it can be used to retrieve useful information for the early phase of drug discovery and development studies.

In order to extract important information from bioassay datasets, a wide range of machine learning (ML) algorithms have been used for classification or activity prediction of compounds. This technique, also called virtual screening, has been extensively studied in the literature. Neural networks (NN) [3], support vector machines (SVM) [4,5], naïve Bayes (NB) [6], k-nearest neighbor (k-NN) [7], random forest (RF) [8] are the most common machine learning algorithms that are used for classification purposes in drug discovery studies. Recently, deep learning (DL), also known as deep neural networks (DNN), has been emerged as a new subfield of machine learning and achieved great performances in various areas, including speech recognition, image classification, language translation, and among many others. DL has been also used in drug discovery studies and achieved great performances. Ma et al. (2015) used a DL model for prediction of quantitative structure-activity relationships and they found that the DNN outperformed RF model [9]. Mayr et al. (2016) used a multi-task deep learning architecture to predict toxicity of compounds and won the Tox21 challenge [10]. Ramsundar et al. (2015) also applied a multi-task DNN on various publicly available molecular compound datasets [11]. Koutsoukas et al. (2017) investigated optimization of hyper-parameters of a DNN model and compared the performance of the DL model with SVM, RF, NB and kNN algorithms [12]. Lenselink et al. (2017) compared performance of a DL model with NB, RF, SVM and logistic regression using ChEMBL bioactivity dataset [13].

Although PubChem is a great resource for drug discovery datasets, the bioassay datasets obtained by the HTS are usually imbalanced, which means that the number of inactive compounds is considerably larger than the number of active compounds. This makes the learning problem harder for ML/DL algorithms, since there is usually little or not enough data for the minority class in these datasets. This problem is not well addressed in the literature for the PubChem repository.

One alternative is to create balanced datasets using over-sampling and/or under-sampling methods. In this study, we aimed to explore the effects of data balancing on DNN using imbalanced PubChem bioassays. We used 8 different types of data balancing methods: 3 under-sampling methods, 3 over-sampling methods and 2 hybrid-sampling methods. First, we retrieved a bioassay data from PubChem database. Then, to create the dataset, we generated molecular descriptors for each compound in the bioassay. Next, we applied data balancing methods to the dataset. Finally, we performed DNN models using the dataset for each data balancing method and compared their performances using various performance measure metrics.

The structure of this paper is as follows. First, we will introduce the dataset that we obtained from PubChem database. Then, we will give information regarding data balancing methods. Next, we will introduce DNN model, our model building strategy and the performance metrics that we used to compare performances of data balancing methods. Finally, we will present our results, discuss our findings and draw some conclusions.

2. METHODS

2.1. Datasets

We used a confirmatory quantitative high-throughput screening (qHTS) assay from the PubChem repository: AID485314 is a confirmatory qHTS assay for inhibitors of DNA polymerase beta. DNA polymerase beta is a key enzyme for repair system of human cells which makes it a promising target for therapeutic modulation of the response to radiation treatment and DNA-damaging drugs.

This assay previously used by Zakharov et al. (2014), where the authors used different data balancing strategies and investigated their performances on QSAR (quantitative structure−activity relationships) models [14]. We downloaded this bioassay in SDF (structure data file) format. Then using the SDF file, we calculated molecular descriptors using PADEL software [15] for each compound in the bioassay. We calculated 2,750 molecular descriptors (1D-2D-3D descriptors and PubChem fingerprints) for each compound using this tool. A more detailed information about molecular descriptors can be found in Yap et al. (2010). After creating the dataset, we applied a pre-processing step. We performed a near zero variance procedure [16] to identify variables that have one or very few unique values. Then, we removed these variables from the datasets and reduced the number of variables to 1,361. The dataset is highly imbalanced, in which it contains 312,009 inactive and 4485 active compounds.

2.2. Data Balancing Methods

2.2.1. Over Sampling

To create a balanced dataset, over-sampling methods try to increase the number of samples in the minority class. The one of the easiest and naïve ways of balancing a dataset is to generate new samples for minority class by randomly sampling with replacement the already available samples. This method, called random over-sampling (ROS), will increase the minority class samples up to size of majority class samples and will create a balanced dataset. Although this method has a straightforward implementation, it is, however, criticized because newly created dataset does not contain more information about the characteristics of the minority class than the old dataset and, thus, causes overfitting [17,18]. Therefore more sophisticated methods have been developed in order to solve the naïve over-sampling problem. The SMOTE (Synthetic Minority Over-sampling

Technique) is one of the over-sampling methods, which creates synthetic samples for the minority class rather than by over-sampling with replacement [19]. This method first finds the k nearest neighbors for each minority class sample. Next, the new samples are generated in the direction of those nearest neighbors. Depending upon the number of over-sampling samples required, a certain number of neighbors from the k nearest neighbors are randomly selected. Another over-sampling method offered by He et al. (2008), called ADASYN (adaptive synthetic), also creates artificial samples for the minority class [20]. The main idea behind the ADASYN is to generate minority class samples that are harder to learn compared to those minority class samples that are easier to learn. In this way, this method can reduce the learning bias and, also, can adaptively shift the decision boundary to focus on those hard to learn samples.

2.2.2. Under Sampling

Unlike over-sampling, under-sampling methods reduce the number of samples in the majority class in order to find a balance between minority and majority classes. Random under-sampling (RUS), like ROS, is a fast and easy way to balance the minority and majority classes by randomly removing samples from the majority class. This method will reduce the number of majority class samples up to size of minority class samples in order to create a balanced dataset. However, this method may potentially remove important samples and, thus, may lead to increase in variance. Another alternative in under-sampling is to apply Tomek's link (Tomek) method [21]. This method detects the Tomek's links between two samples of different classes. A Tomek's link exists if the two samples are nearest neighbors of each other. After detecting Tomek's links, majority class links are removed until all minimally distanced nearest neighbor pairs are of the same class. Removing overlapping samples from the majority class leads to constitute well-defined clusters in the training set and improves the classification performance of machine learning algorithms. Another under-sampling method called edited nearest neighbors (ENN) applies a nearest neighbor algorithm and remove samples which do not agree with their neighborhood [22]. The nearest neighbors are calculated for each sample in the majority class and if the selection criterion is not satisfied then the sample is removed.

2.2.3. Hybrid Sampling

In order to increase the classification performance, the hybrid methods, which combine under-sampling and over-sampling methods have been developed. To achieve this, Tomek and ENN methods have been used in combination with SMOTE, namely SMOTETomek [23] and SOMOTEENN [24]. The main idea behind the hybrid methods is to avoid noisy samples that can be generated by the SMOTE and to obtain a cleaner space.

2.3. Deep Neural Networks

Deep learning, also known as deep neural networks (DNN), is a subfield of machine learning. A DNN uses multiple successive layers of increasingly meaningful representation of data [25] and it includes multiple non-linear hidden layers between an input layer and an output layer [26]. The DNN can learn very complex relationships between input and output [27], since it includes large number of layers and parameters [28]. The layers, which stacked one after the other, are called as neural networks. The input data are stored in these layers as weights. The DNN model tries to find a set of values for the weights of all layers in the network and, finally, tries to correctly map input data to true classes using these weights [25]. A loss function is used to find the optimal values for the weights. The loss function compares predictions of the DNN model and the true classes, and

computes a distance score. This score shows how well the DNN model performed. In order to optimize the loss score, the score produced by the loss function is used as a feedback. This procedure is handled by the optimizer, which implements the backpropagation method [29]. The backpropagation works as follows: (i) the value of the weights are assigned randomly, which makes the loss score very high, (ii) then, the weights are adjusted slightly in the right direction to decrease the loss score, (iii) finally, this process is repeated iteratively for a number of times (i.e. number of epoch) to obtain the weights that minimize the loss function, (iv) finally, a DNN with a minimum loss score yields a trained network which produces predictions that are as close as they can be to the true classes [25]. A general architecture of a DNN and relationship between the DNN layers, loss function, and optimizer is illustrated in Figure 1.

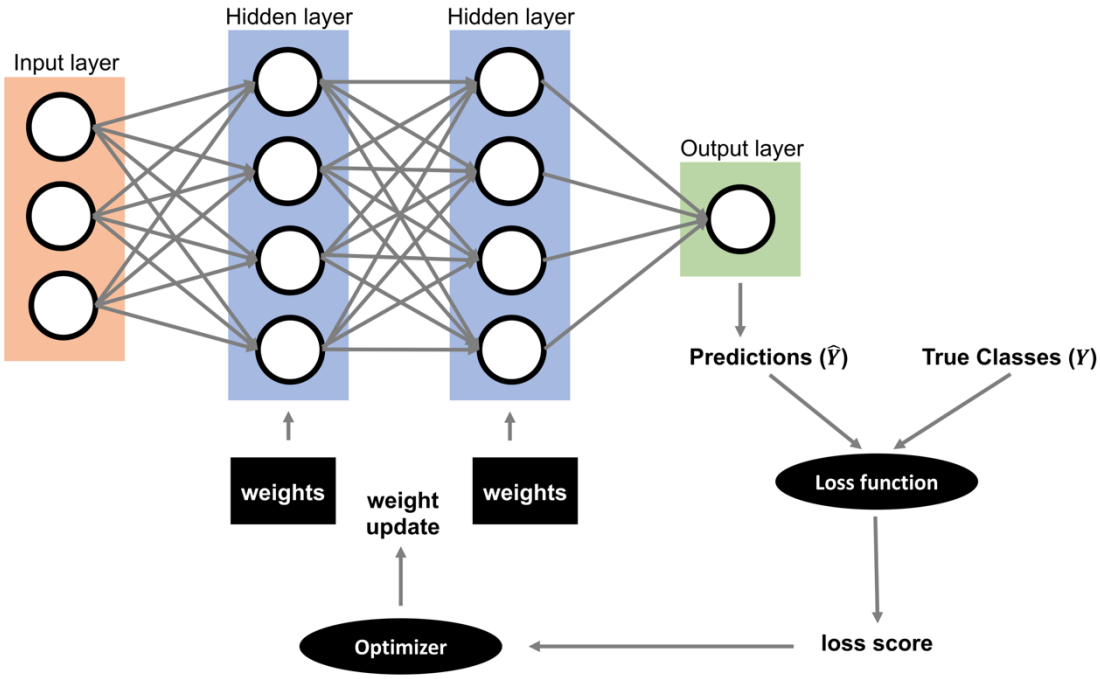


Figure 1. A general architecture of a DNN and relationship between the DNN layers, loss function, and optimizer (adapted from [25]).

2.4. Model Building

First, we applied each data balancing method to each dataset. Then, we centered and scaled each dataset set using z-score transformation. Next, we split each dataset into two parts as 80% training set and 20% test set. We used a fully connected four-hidden-layer DNN architecture. The ReLU (rectified linear unit) function [30] is used as activation function for input and hidden layers and a sigmoid activation function is used for the output layer. A batch normalization is applied in each layer to improve the performance and the stability of the DNN model. We applied 20% dropout rate at each layer in order to avoid the overfitting as suggested by Srivastava et al. (2014). We used binary loss function and Adam method for stochastic optimization. The mini batch size is set to 32 and the number of epoch is defined as 1000.

2.5. Performance Metrics

To compare the performances of data balancing methods, we calculated several performance measures, including balanced accuracy, precision, recall, F1-score and Matthews correlation coefficient (MCC).

Overall acuraccy is not a proper metric when the dataset is highly imbalanced. In this case, balanced accuracy can be used to evaluate general performance of the algorithm. The balanced accuracy is computed as the average between sensitivity and specificity.

$$Balanced\ Accuracy = \left[\frac{TP}{TP + FN} + \frac{TN}{TN + FP}\right]/2$$

Precision, also called as positive predictive value, is the proportion of true positives among samples classified as positive and it varies between 0 and 1.

$$Precision = \frac{TP}{TP + FP}$$

Recall, also called as sensitivity, is the proportion of true positives among all positives and it varies between 0 and 1.

$$Recall = \frac{TP}{TP + FN}$$

F1 score is the weighted average value of the precision and recall. It takes both false negatives and false positives into account and it varies between 0 and 1.

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

MCC another useful performance metric when the dataset is imbalanced. The MCC is a correlation coefficient between observed and predicted classes. It varies between -1 and +1. When the MCC equals -1, it shows a total disagreement between observed and predicted classes, a value of 0 shows no better than random prediction, and a coefficient of +1 shows a perfect prediction.

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$$

AUC (area under the ROC curve) is one of the most widely used metrics which shows the overall performance of a classification algorithm. It varies between 0.5 and 1. A model with AUC equals 1 means there is perfect separation, and when AUC equals 0.5 it means there is no class separation.

$$AUC = \frac{1}{n_0 n_1} \sum_{i=1}^{n_0} \sum_{j=1}^{n_1} \left(I\left[Y_{1i} > Y_{0j}\right] + \frac{1}{2} I\left[Y_{1i} = Y_{0j}\right] \right)$$

where, TP = true positives, TN = true negatives, FP = false positives and FN = false negatives, i runs over all n_0 data points with true label 1, j runs over all n_1 data points with true label 0, Y_{1i} is the i^{th} value for true label 1, Y_{0j} is the j^{th} value for true label 0 and I is the indicator function.

3. RESULTS

We investigated performance of DNN for each data balancing method by calculating performance metrics for the dataset AID485324. The results are summarized in Table 1. When we did not apply any data balancing method the network shows poor performance in terms of balanced accuracy (0.639), precision (0.513), recall (0.282), F1 score (0.364) and MCC (0.374). On the other hand,

under-sampling methods slightly improves the performance of the DNN compared to original dataset. RUS was the best performed data balancing method amongst the under-sampling methods (balanced accuracy = 0.761, precision = 0.766, recall = 0.738, F1 score = 0.752 and MCC = 0.522), whereas Tomek was the worst performed under-sampling method (balanced accuracy = 0.642, precision = 0.638, recall = 0.287, F1 score = 0.396 and MCC = 0.423). Although under-sampling methods slightly improves the performance over the original dataset, the improvement was not adequate for the classification of a small chemical molecule dataset. Over-sampling methods, on the other hand, outperformed under-sampling methods.

Table 1. Performance metric results for each data balancing method

	Balanced Accuracy	**Precision**	**Recall**	**F1**	**MCC**	**AUC**
Original	0.639	0.513	0.282	0.364	0.374	0.831
RUS	0.761	0.766	0.738	0.752	0.522	0.819
Tomek	0.642	0.638	0.287	0.396	0.423	0.812
ENN	0.679	0.733	0.360	0.483	0.509	0.861
ROS	0.996	0.993	0.998	0.996	0.991	0.998
SMOTE	0.997	0.995	0.999	0.997	0.994	0.999
ADASYN	0.995	0.993	0.997	0.995	0.990	0.998
SMOTETomek	0.997	0.994	0.999	0.997	0.993	0.999
SMOTEENN	0.998	0.997	1.000	0.999	0.997	1.000

All of the over*sampling methods performed better than all under-sampling methods in terms of all performance metrics. ROS yielded 0.996 balanced accuracy, 0.993 precision, 0.998 recall, 0.996 F1 score and 0.991 MCC. Similarly, balanced accuracy, precision, recall, F1 score and MCC is founded as 0.997, 0.995, 0.999, 0.997 and 0.994, respectively, for SMOTE method. Likewise, ADASYN performed well in terms of balanced accuracy (0.995), precision (0.993), recall (0.997), F1 score (0.995) and MCC (0.994). Similar to the over-sampling methods, hybrid methods show well performances in terms of all performance measures and outperformed under-sampling methods. SMOTETomek resulted in 0.997 balanced accuracy, 0.994 precision, 0.999 recall, 0.997 F1 score, and 0.993 MCC. Finally, SMOTEENN achieved the best performance among all data balancing methods and resulted in 0.998 balanced accuracy, 0.997 precision, 1.000 recall, 0.999 F1 score and 0.997 MCC.

Moreover, we have performed ROC curve analysis and calculated AUC value for each data balancing method. The ROC curves are illustrated in Figure 2. There were no significant differences between original data set (AUC = 0.831) and under-sampling methods (AUC for RUS is 0.819, AUC for Tomek is 0.812 and AUC for ENN is 0.861, p valueas are 0.565, 0.485, 0.276,

respectively). On the other hand both over-sampling methods (AUC for ROS is 0.998, AUC for SMOTE is 0.999 and AUC for ADASYN is 0.998, all p values are <0.001) and hybrid methods (AUC for SMOTETOMEK is 0.999 and AUC for SMOTEENN is 1.000, all p values are <0.001) significantly improved the performance in terms of AUC.

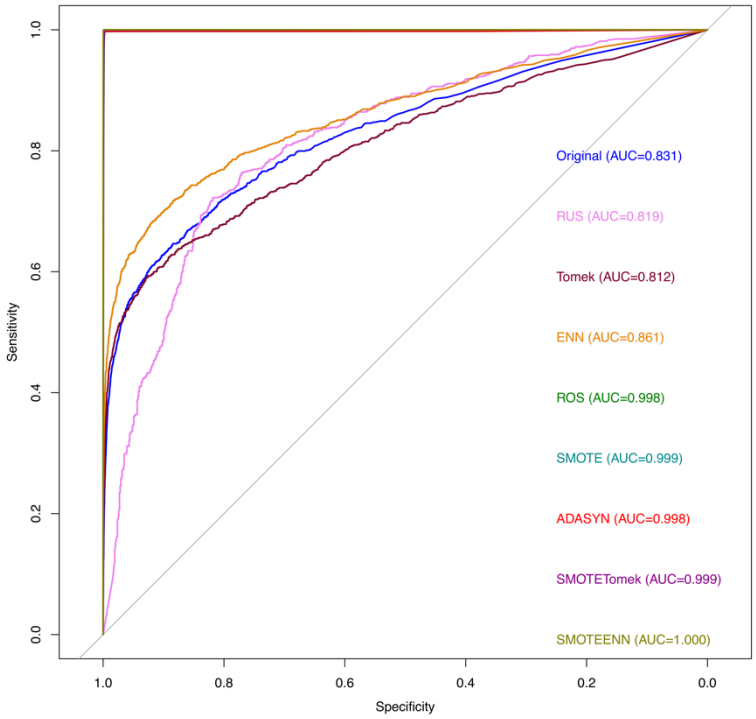


Figure 2. ROC curves for each data balancing method

4. DISCUSSION AND CONCLUSION

Drug discovery process is hard and extremely time consuming. A great deal of effort has been made to decrease both cost and time consumption in the last two decades. It is critical to determine promising molecule compounds that can be drugs in the future in the early phase of drug discovery studies. Experimental methods, such as HTS, have vital importance to detect compounds that show activity towards a specific enzyme or receptor. Datasets obtained by HTS are called bioassays and these datasets uploaded to public databases, such as PubChem. PubChem is a public repository for small chemical compounds and it hosts 1.3 million bioassays as of July 2019. This unique feature of PubChem repository makes it a great resource for machine learning approaches. The classification and activity prediction for drug compounds have been extensively studied throughout the literature. NN algorithm made a breakthrough in the late 1990s in classification of active and inactive compounds. Later, SVM outperformed NN and become the state-of-the-art method almost a decade [4, 5]. Other machine learning algorithms, including NB [6], kNN [7] and RF [8], also used in classification of drug-like and nondrug-like compounds. Recently, DL achieved great performances in almost every application, including text classification, image recognition, compound classification and among others. It is reported that DL outperformed SVM and RF in various studies and become the new state-of-the-art method [9, 10]. Bioassay datasets obtained from PubChem repository includes hundreds of thousands molecules,

which make these datasets perfect for DL algorithm. However, the datasets retrieved from PubChem database are highly imbalanced, which is a well-known disadvantage for machine learning algorithms. To solve this problem, a number of approaches has been suggested. One of the easiest way to deal with such problem is to randomly add (ROS) or subtract (RUS) data points from original dataset. In recent years, more sophisticated methods have been developed. SMOTE and ADASYN are of the most popular over-sampling methods, whereas Tomek and ENN are amongst the most popular under-sampling methods. Moreover, various hybrid methods, such as SMOTETomek and SMOTEENN, which combine under and over-sampling methods, have been proposed to further improve the performance. Here, we have compared performance of different data balancing methods on a DNN model in terms of various performance metrics. Firstly, we observed that raw dataset shows a poor performance, whereas under-sampling methods slightly improved the performance of the DNN. Over-sampling methods, on the other hand, outperformed under-sampling methods in terms of all performance metrics. While the performance of the under-sampling methods was not sufficient, combining them with the over-sampling methods further improved the performance of the DNN. PubChem is a great resource for experimentally derived small chemical molecules and the DNN is a well-performed machine learning method to decrease the cost and the time in drug discovery studies. Data balancing methods, especially over-sampling and hybrid-sampling methods, can further improve the performance of the deep learning models to detect promising active compounds, which can be drug candidates in the future.

REFERENCES

1. Broach, J. R., Thorner, J. (1996). High-throughput screening for drug discovery. Nature, 384(6604), 14-16.

2. Kim, S., Thiessen, P. A., Bolton, E. E., Chen, J., Fu, G., Gindulyte, A., et al.. (2015). PubChem substance and compound databases. Nucleic acids research, 44(D1), D1202-D1213.

3. Sadowski, J., Kubinyi, H. (1998). A scoring scheme for discriminating between drugs and nondrugs. Journal of medicinal chemistry, 41(18), 3325-3329.

4. Byvatov, E., Fechner, U., Sadowski, J., Schneider, G. (2003). Comparison of support vector machine and artificial neural network systems for drug/nondrug classification. Journal of chemical information and computer sciences, 43(6), 1882-1889.

5. Korkmaz, S., Zararsiz, G., Goksuluk, D. (2014). Drug/nondrug classification using support vector machines with various feature selection strategies. Computer methods and programs in biomedicine, 117(2), 51-60.

6. Sun, H. (2005). A naive Bayes classifier for prediction of multidrug resistance reversal activity on the basis of atom typing. Journal of medicinal chemistry, 48(12), 4031-4039.

7. Miller, D. W. (2001). Results of a new classification algorithm combining K nearest neighbors and recursive partitioning. Journal of chemical information and computer sciences, 41(1), 168-175.

8. Ehrman, T. M., Barlow, D. J., & Hylands, P. J. (2007). Virtual screening of Chinese herbs with random forest. Journal of chemical information and modeling, 47(2), 264-278.

9. Ma, J., Sheridan, R. P., Liaw, A., Dahl, G. E., Svetnik, V. (2015). Deep neural nets as a method for quantitative structure–activity relationships. Journal of chemical information and modeling, 55(2), 263-274.

10. Mayr, A., Klambauer, G., Unterthiner, T., Hochreiter, S. (2016). DeepTox: toxicity prediction using deep learning. Frontiers in Environmental Science, 3, 80.

11. Ramsundar, B., Kearnes, S., Riley, P., Webster, D., Konerding, D., & Pande, V. (2015). Massively multitask networks for drug discovery. arXiv preprint arXiv:1502.02072.

12. Koutsoukas, A., Monaghan, K. J., Li, X., & Huan, J. (2017). Deep-learning: investigating deep neural networks hyper-parameters and comparison of performance to shallow methods for modeling bioactivity data. Journal of cheminformatics, 9(1), 42.

13. Lenselink, E. B., Ten Dijke, N., Bongers, B., Papadatos, G., Van Vlijmen, H. W., Kowalczyk, W. et al. (2017). Beyond the hype: deep neural networks outperform established methods using a ChEMBL bioactivity benchmark set. Journal of cheminformatics, 9(1), 45.

14. Zakharov, A. V., Peach, M. L., Sitzmann, M., Nicklaus, M. C. (2014). QSAR modeling of imbalanced high-throughput screening data in PubChem. Journal of chemical information and modeling, 54(3), 705-712.

15. Yap, C. W. (2011). PaDEL-descriptor: An open source software to calculate molecular descriptors and fingerprints. Journal of computational chemistry, 32(7), 1466-1474.

16. Kuhn, M. (2008). Building predictive models in R using the caret package. Journal of statistical software, 28(5), 1-26.

17. Ling, C. X., Li, C. (1998, August). Data mining for direct marketing: Problems and solutions. In Kdd (Vol. 98, pp. 73-79).

18. Japkowicz, N. (2000). The class imbalance problem: Significance and strategies. In Proc. of the Int'l Conf. on Artificial Intelligence.

19. Chawla, N. V., Bowyer, K. W., Hall, L. O., Kegelmeyer, W. P. (2002). SMOTE: synthetic minority over-sampling technique. Journal of artificial intelligence research, 16, 321-357.

20. He, H., Bai, Y., Garcia, E. A., Li, S. (2008, June). ADASYN: Adaptive synthetic sampling approach for imbalanced learning. In 2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence) (pp. 1322-1328). IEEE.

21. Tomek, I. (1976). Two modifications of CNN. IEEE Trans. Systems, Man and Cybernetics, 6, 769-772.

22. Wilson, D. L. (1972). Asymptotic properties of nearest neighbor rules using edited data. IEEE Transactions on Systems, Man, and Cybernetics, (3), 408-421.

23. Batista, G. E., Prati, R. C., & Monard, M. C. (2004). A study of the behavior of several methods for balancing machine learning training data. ACM SIGKDD explorations newsletter, 6(1), 20-29.

24. Batista, G. E., Bazzan, A. L., Monard, M. C. (2003). Balancing Training Data for Automated Annotation of Keywords: a Case Study. In WOB (pp. 10-18).

25. Chollet F., Allaire J. J. (2018). Deep Learning with R. Manning Publications Co.

26. Larochelle, H., Bengio, Y., Louradour, J., Lamblin, P. (2009). Exploring strategies for training deep neural networks. Journal of machine learning research, 10(Jan), 1-40.

27. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R. (2014). Dropout: a simple way to prevent neural networks from overfitting. The journal of machine learning research, 15(1), 1929-1958.

28. Patterson J., Gibson A. (2017). Deep Learning: A Practitioner's Approach. 1st ed. Beijing: O'Reilly.

29. Rumelhart, D. E., Hinton, G. E., Williams, R. J. (1988). Learning representations by back-propagating errors. Cognitive modeling, 5(3), 1.

30. Dahl, G. E., Sainath, T. N., Hinton, G. E. (2013). Improving deep neural networks for LVCSR using rectified linear units and dropout. In 2013 IEEE international conference on acoustics, speech and signal processing (pp. 8609-8613). IEEE.