



**XIX. ULUSAL
ve
II. ULUSLARARASI
BİYOİSTATİSTİK KONGRESİ
BİLDİRİ ÖZETLERİ KİTABI**

Editör

Doç. Dr. Erdem Karabulut

Yardımcı Editörler

Prof. Dr. Atilla Halil Elhan

Prof. Dr. Ersin Ögüş

Doç. Dr. Serdal Kenan Köse

Altın Sponsor



25-28 Ekim 2017
Belek / ANTALYA



XIX. NATIONAL and II. INTERNATIONAL BIostatISTICS CONGRESS BOOK of ABSTRACTS

Editor in Chief

Assoc. Prof. Erdem Karabulut

Associate Editors

Prof. Atilla Halil Elhan

Prof. Ersin Öğüş

Assoc. Prof. Serdal Kenan Köse

Altın Sponsor



October 25-28, 2017
Belek / ANTALYA

Onursal Kurul / Honorary Board

Kongre Onursal Başkanı / Honorary President of the Congress

Prof. Dr. Reha ALPAR Biyoistatistik Derneği Başkanı

Onur Kurulu / Honorary Board

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

Prof. Dr. Kazım ÖZDAMAR O.Ü. Tıp Fak. Biyoistatistik ABD E.
Öğretim Üyesi

Prof. Dr. Ersöz TÜCCAR Ankara Üniversitesi Tıp Fakültesi
Biyoistatistik Anabilim Dalı Emekli
Öğretim Üyesi

Prof. Dr. Fikret İKİZ Ege Üniversitesi Tıp Fakültesi
Biyoistatistik Anabilim Dalı Emekli
Öğretim Üyesi

Prof. Dr. İsmet KAN Uludağ Üniversitesi Tıp Fakültesi
Biyoistatistik Anabilim Dalı Emekli
Öğretim Üyesi

Prof. Dr. Yıldır ATAKURT Ankara Üniversitesi Tıp Fakültesi
Biyoistatistik Anabilim Dalı Emekli
Öğretim Üyesi

Düzenleme Kurulu / Organizing Committee

Prof. Dr. C. Reha ALPAR Hacettepe Üniversitesi Başkan

Prof. Dr. Atilla Halil ELHAN Ankara Üniversitesi Üye

Prof. Dr. Ersin ÖĞÜŞ Başkent Üniversitesi Üye

Doç. Dr. S. Kenan KÖSE Ankara Üniversitesi Üye

Prof. Dr. Erdem KARABULUT Hacettepe Üniversitesi Üye

Bilimsel Danışma Kurulu / Scientific Advisory Board

Prof. Dr. A. Canan YAZICI GÜVERCİN	İstanbul Aydın Üniversitesi
Prof. Dr. A. Ergun KARAAĞAOĞLU	Hacettepe Üniversitesi
Prof. Dr. Ahmet DİRİCAN	İstanbul Üniversitesi
Prof. Dr. Alan TENNANT	University of Lucerne, Switzerland
Prof. Dr. Atilla Halil ELHAN	Ankara Üniversitesi
Prof. Dr. Bahar TAŞDELEN	Mersin Üniversitesi
Prof. Dr. Beyza AKDAĞ ŞAHİN	Pamukkale Üniversitesi
Prof. Dr. Celal Reha ALPAR	Hacettepe Üniversitesi
Prof. Dr. Cemil ÇOLAK	İnönü Üniversitesi
Prof. Dr. Ersin ÖĞÜŞ	Başkent Üniversitesi
Prof. Dr. Erdem KARABULUT	Hacettepe Üniversitesi
Prof. Dr. Fezan MUTLU	Eskişehir Osmangazi Üniversitesi
Prof. Dr. Gülşah SEYDAOĞLU	Çukurova Üniversitesi
Prof. Dr. Hafize SEZER	İstanbul Aydın Üniversitesi
Prof. Dr. Handan ANKARALI	Medeniyet Üniversitesi
Prof. Dr. İlker ERCAN	Uludağ Üniversitesi
Prof. Dr. İmran KURT ÖMÜRLÜ	Adnan Menderes Üniversitesi
Prof. Dr. İsmet DOĞAN	Afyon Kocatepe Üniversitesi
Prof. Dr. Kazım ÖZDAMAR	Eskişehir Osmangazi Üniversitesi
Prof. Dr. Mehmet ORMAN	Ege Üniversitesi
Prof. Dr. Mehtap AKÇİL OK	Başkent Üniversitesi
Prof. Dr. Meriç ÇOLAK	Başkent Üniversitesi
Prof. Dr. Mevlut TÜRE	Adnan Menderes Üniversitesi
Prof. Dr. Mustafa ŞENOCAK	İstanbul Üniversitesi
Prof. Dr. Necdet SÜT	Trakya Üniversitesi
Prof. Dr. Nural BEKİROĞLU	Marmara Üniversitesi
Prof. Dr. Ömer SATICI	Dicle Üniversitesi
Prof. Peter W. ROSE	University of California, USA

XIX. Ulusal ve II. Uluslararası Biyoistatistik Kongresi
XIX. National and II. International Biostatistics Congress
25-28 Ekim 2017 Belek / Antalya

Prof. Rama SHANKER	Eritrea Inst. of Tech., Eritrea
Prof. Dr. Saim YOLOĞLU	İnönü Üniversitesi
Prof. Dr. Sanjib BASU	University of Illinois, USA
Prof. Dr. Setenay ÖNER	Eskişehir Osmangazi Üniversitesi
Prof. Dr. Sıddık KESKİN	Van Yüzüncü Yıl Üniversitesi
Prof. Dr. Vildan SÜMBÜLOĞLU	Sanko Üniversitesi
Prof. Dr. Yasemin YAVUZ	Ankara Üniversitesi
Prof. Dr. Yavuz SANİSOĞLU	Yıldırım Beyazıt Üniversitesi
Prof. Dr. Yusuf ÇELİK	Biruni Üniversitesi
Prof. Dr. Z. Nazan ALPARSLAN	Çukurova Üniversitesi
Prof. Dr. Zeki AKKUŞ	Dicle Üniversitesi
Doç. Dr. Ahmet ÖZTÜRK	Erciyes Üniversitesi
Assoc. Prof. Alex R. de LEON	University of Calgary, Canada
Doç. Dr. Beyza DOĞANAY ERDOĞAN	Ankara Üniversitesi
Doç. Dr. Bülent ÇELİK	Gazi Üniversitesi
Doç. Dr. Canan BAYDEMİR	Kocaeli Üniversitesi
Doç. Dr. Deniz SİĞİRLİ	Uludağ Üniversitesi
Doç. Dr. Derya ÖZTUNA	Ankara Üniversitesi
Assist. Prof. Duchwan RYU	Northen Illinois University, USA
Assoc. Prof. Hakan DEMİRTAŞ	University of Illinois at Chicago, USA
Doç. Dr. Ertuğrul ÇOLAK	Eskişehir Osmangazi Üniversitesi
Doç. Dr. Ferhan ELMALI	İzmir Katip Çelebi Üniversitesi
Doç. Dr. Gökhan OCAKOĞLU	Uludağ Üniversitesi
Doç. Dr. Gülhan OREKİCİ TEMEL	Mersin Üniversitesi
Doç. Dr. Güven ÖZKAYA	Uludağ Üniversitesi
Doç. Dr. İlker ETİKAN	Yakın Doğu Üniversitesi
Doç. Dr. Jale KARAKAYA	Hacettepe Üniversitesi
Assoc. Prof. Lynn SLEEPER	Harvard Medical School, USA
Assoc. Prof. Melody GHAHRAMANI	University of Winnipeg, USA

XIX. Ulusal ve II. Uluslararası Biyoistatistik Kongresi
XIX. National and II. International Biostatistics Congress
25-28 Ekim 2017 Belek / Antalya

Doç. Dr. Nurhan DOĞAN	Afyon Kocatepe Üniversitesi
Doç. Dr. Pınar ÖZDEMİR	Hacettepe Üniversitesi
Assoc. Prof. Recai YÜCEL	University at Albany-SUNY, USA
Doç. Dr. S. Kenan KÖSE	Ankara Üniversitesi
Doç. Dr. Seval KUL	Gaziantep Üniversitesi
Doç. Dr. Şengül CANGÜR	Düzce Üniversitesi
Doç. Dr. Ünal ERKORKMAZ	Sakarya Üniversitesi
Doç. Dr. Timur KÖSE	Ege Üniversitesi
Assist. Prof. Umut ÖZBEK	Icahn School of Medicine, Mount Sinai, USA
Doç. Dr. Yaşar SERTDEMİR	Çukurova Üniversitesi
Dr. Atanu BHATTACHARJEE	Assam, India
Dr. Bernd KLAUS	European Mol. Bio. Lab.,Germany

ÖNSÖZ

Sayın Hocalarım ve Sevgili Meslektaşlarım,

Tıp ve diğer sağlık bilimleri alanlarında çalışan araştırmacıların, yapmış oldukları çalışmaları etkin ve uygun biçimde planlayabilmesi, yaptığı araştırmaları bilimsel temele oturtabilmesi, başarılı bir biçimde yürütebilmesi, güvenilirliği yüksek kararlar verebilmesi için güvenilir verilere, güvenilir istatistiksel yöntemlere ve alanında uzman güvenilir biyoistatistikçilere gereksinimi vardır. Genelde tıp ve diğer sağlık bilimleri gibi önemli çalışma alanlarında kendisine yer bulmuş olan biyoistatistiğin önemi 1920'lerden bu yana her geçen gün artmıştır. Günümüzde biyoistatistik, tıp ve diğer sağlık bilimleri ile ilgili bilimsel çalışmaların vazgeçilmez bir ögesi olmuş ve bu alanda çalışan bilim insanlarının görmesine, duymasına, sezmesine ve karar vermesine önemli derecede katkıda bulunan çok önemli bir bilim alanı konumuna ulaşmıştır.

Günümüzde, tıp ve diğer sağlık bilimleri alanındaki gözlemsel ve deneysel (klinik) araştırmaların tasarım aşamasından, çözümleme ve karar verme aşamasına kadarki süreçte donanımlı biyoistatistik bilgisine sahip biyoistatistik uzmanına duyulan gereksinim eskisinden çok daha fazladır. Bu çerçevede, tıp fakültelerinin temel tıp bilimleri bölümü içerisinde "temel tıp alanı olarak" 1980'li Yılların başından günümüze dek çalışmalarını sürdüren Biyoistatistik Anabilim Dallarının sorumluluğu kat kat artmıştır.

25-28 Ekim 2017 tarihleri arasında Antalya'da yapılacak olan ve bu yıl Biyoistatistik Derneği tarafından düzenlenen "XIX. Ulusal ve II. Uluslararası Biyoistatistik Kongresi" de bu sorumluluğun bir parçasıdır (kongrenin ilk günü kurslara ayrılmıştır). Bu kongreden beklentimiz; ulusal katılımlı bir kongre olmanın yanı sıra uluslararası katılımlı bir kongre konumuna gelerek, Biyoistatistik alanındaki ulusal ve uluslararası bilimsel gelişmeleri görmek, mesleki birikimlerimizi paylaşmak, tartışmak ve bilimsel yayın kalitemizi yükseltmektir.

Üyesi olmaktan her zaman büyük bir mutluluk ve gurur duyduğum Biyoistatistik topluluğunu ve araştırmacıları bu kongrede bir arada görmek en büyük dileğimdir. Dolayısıyla, Biyoistatistiğe gönül vermiş gerek ülke içinden gerekse ülke dışından tüm bilim insanlarını kongremize davet ederken, düzenleme komitesi adına hepinize en içten sevgi ve saygılarımı sunarım.

Kongre Düzenleme Kurulu Adına,

Prof. Dr. Reha ALPAR

Kongre Başkanı

PREFACE

Dear Colleagues,

The researchers working in the field of medicine & health sciences need reliable biostatistical solutions and biostatisticians in order to carry on their studies. Biostatistics which has been usually featured in the fields of medicine and health sciences increased its popularity since the 1920's and it has become a very important field. This importance has been very valuable for the scientists to hear, feel and understand this field better in every aspect. In this regard, responsibility of Biostatistics and biostatisticians has been increased rapidly, especially since the beginning of the 1980s.

XIX. Biostatistics Congress which will be held in Antalya from October 25 to October 28 is also a part of responsibility. Our expectation from this congress is to be recognized internationally by our attendants from all around the world. Our purpose is to gather all Biostatisticians to discuss and analyze the new trends, and to exchange our experience and scientific knowledge.

As a member and temporary chairman of Association of Biostatistics of Turkey, our only wish is to see all the respected biostatisticians and researchers from Turkey and all around the world together in this important congress. We are inviting you to join us in Antalya with all the love and respect.

On behalf of organization board,

Prof. Reha ALPAR

Congress Chair

BİLİMSEL PROGRAM / SCIENTIFIC PROGRAM

25.10.2016 Çarşamba	
9.00 – 17:00	Kongre Kayıt / Registration
14:00-17:30	Kurs 1. R İle Kestirime Yönelik İstatistiksel Modellerin Oluşturulması G. ZARARSIZ, S. KORKMAZ, D. GÖKSÜLÜK Kurs 2. Yapısal Eşitlik Modellemesi, Lisrel ve Amos Uygulamaları K. ÖZDAMAR Kurs 3. Küme Randomize Denemeler ve Çok Düzeyli Modelleme S. KUL
26.10.2016 Perşembe	
9:15 – 9:35	Açılış Saygı Duruşu İstiklal Marşı Açılış Konuşmaları: Prof. Dr. Reha ALPAR, Biyoistatistik Derneği Başkanı Prof. Dr. Kadir SÜMBÜLOĞLU, Biyoistatistik Derneği Onursal Başkanı
9:35 – 10:10	Konferans I: Kanonik Korelasyon ve ilişkili Modeller Prof. Dr. Sıddık KESKİN, Van Yüzüncü Yıl Üniversitesi
10:10 - 10:40	Microsoft Azure Üzerinde Bilimsel Araştırma Senaryoları Semih ESKİN, Bulut Çözümleri Mimarı – Veri ve Yapay Zeka, Microsoft Türkiye Dr. Erdal COŞGUN, Veri Bilimci, Microsoft Research Genetik Ekibi
10:40-10:50	Çay-Kahve Arası
Salon I (Sunum Dili: Türkçe)	
10:50-12:20	Sözlü Bildiriler –I. Oturum Oturum Başkanları: Prof. Dr. Z. Nazan ALPARSLAN, Doç. Dr. Ertuğrul Çolak
10:50-11:05	Ortak Kontrol Grubunun Bulunduğu Faktöriyel Denemelerin İstatistiksel Analizleri: Monte Carlo Simulasyon Çalışması Soner YİĞİT, Mehmet MENDEŞ
11:05-11:20	Bootstrap Güven Aralığı Yöntemlerinin Karşılaştırılması Merve BAŞOL, Dinçer GÖKSÜLÜK, A. Ergun KARAAĞAOĞLU
11:20-11:35	Tedavi-Biyomarker Etkileşiminin BATE Eğrisi ile İncelenmesi Büşra EMİR, Ertuğrul ÇOLAK

XIX. Ulusal ve II. Uluslararası Biyoistatistik Kongresi
XIX. National and II. International Biostatistics Congress
25-28 Ekim 2017 Belek / Antalya

11:35-11:50	GAMLSS Modellerinin Model Yeterliliğinin Belirlenmesinde Worm Eğrileri <u>Meryem ÇAVUŞOĞLU</u> , Gökmen ZARARSIZ, Gözde Ertürk ZARARSIZ, Selçuk KORKMAZ, Dinçer GÖKSÜLÜK, Mümtaz MAZICIOĞLU, Ahmet ÖZTÜRK
11:50-12:05	Veri Füzyon Yöntemleri ile Çoklu Sensör Görüntülerinden Bilgi Çıkarımı <u>Adem DOĞANER</u> , Cemil ÇOLAK
12:05-12:20	Kohonen Öz Örgütlemeli Haritalama Yöntemi İle Psikotik Hastalıkların Kümelenmesi <u>Ash BOZER</u> , Osman DAĞ, Sevilay KARAHAN
12.20-13.45	Öğle Yemeği
13:45-15:15	Sözlü Bildiriler –II. Oturum Oturum Başkanları: Prof. Dr. Vildan SÜMBÜLOĞLU, Doç. Dr. Ferhan ELMALI, Yrd. Doç. Dr. Pınar GÜREL KARADENİZ
13:45-14:00	Yaşam Çözümlemesinde Eğilim Skoru <u>Durdu KARASOY</u> , Hatice IŞIK, Nihal ATA TUTKUN
14:00-14:15	Tekrarlı Gözlem ve Sağkalım Verilerinin Birleşik Model Performans Değerlendirmesi: Bir Simülasyon Çalışması <u>Ebru TURGAL</u> , Beyza DOĞANAY ERDOĞAN, S. Kenan KÖSE
14:15-14:30	Orantılı Hazard Durumunun İhlalinde Log-Rank Testinin Gücünün İncelenmesi <u>Hülya ÖZEN</u> , Cengiz BAL
14:30-14:45	Yaşam Süresi Verilerinin Çok Değişkenli Uyumlu Regresyon Analizinde Kullanımı <u>Merve TÜRKEGÜN</u> , Gülhan OREKİCİ TEMEL, İrem ERSÖZ KAYA
14:45-15:00	Zamana Bağlı Bir Belirteç İçin Optimal Kesim Noktası Belirleme <u>İlker ÜNAL</u> , Yaşar SERTDEMİR, Ceren EFE
15:00-15:15	Zamana Bağlı ROC Analizi Yöntemlerinin Alan Tahminlerinin Karşılaştırılması <u>Ceren EFE</u> , İlker ÜNAL
15:15-15:30	Çay-Kahve Arası
15:30-16:45	Kısa Sözlü Bildiriler – III. Oturum Oturum Başkanları: Prof. Dr. Bahar TAŞDELEN, Yrd. Doç. Dr. Cengiz BAL, Yrd. Doç. Dr. Alev BAKIR
15:30-15:35	Okuduğumuz Meta-analizin Gücünü Biliyor muyuz? <u>Zeliha Nazan ALPARSLAN</u> , Nazlı TOTİK
15:35-15:40	Türkiye’de Solunum Hastalıkları Ölümünün Zaman Serisi Analizi <u>Erkut BOLAT</u> , Sedat ALTIN
15:40-15:45	Zaman Serileri Analiz Yöntemlerinin Sağlık Alanında Kullanılması <u>Vedia Bennu GİLAN</u> , Semra ERDOĞAN, Seyran BOZKURT BABUŞ
15:45-15:50	Güvenilirliği Anlamak ve Cronbach’ın Alfa Katsayısını Yorumlamak Yunus Emre KUYUCU, <u>Osman DEMİR</u>

XIX. Ulusal ve II. Uluslararası Biyoistatistik Kongresi
XIX. National and II. International Biostatistics Congress
25-28 Ekim 2017 Belek / Antalya

15:50-15:55	Yaşam Analizlerinde Esnek Parametrik Bir Yaklaşım: Royston-Parmar Modelleri <u>Doğukan ÖZEN</u> , Hülya ÖZEN, Pınar AMBARCIOĞLU, Ufuk KAYA, Erman GÜLENDAG, İ. Safa GÜRCAN
15:55-16:00	Kırılgnlık İndeksi Kullanılarak Randomize Kontrollü Çalışma Sonuçlarının Güvenilirliğinin Değerlendirilmesi <u>Didem DERİCİ YILDIRIM</u> , Bahar TAŞDELEN
16:00-16:05	Lineer Regresyonda Değişen Varyansın Belirlenmesi ve Giderilmesi <u>Betül DAĞOĞLU</u> , Sevinç Püren YÜCEL, Zeliha Nazan ALPARSLAN
16:05-16:10	Kalıtım için İstatistiksel Ölçüler <u>İsmet DOĞAN</u> , Nurhan DOĞAN
16:10-16:15	Fiber Optik Yansıma Spektroskopi Sistemi ile Tümör Teşhisi Bir Meta Analiz Çalışması <u>Pınar GÜNEL KARADENİZ</u> , Tuba DENKÇEKEN
16:15-16:20	Kronik Obstruktif Akciğer Hastalığının Çeşitli Makine Öğrenmesi Yöntemleri Kullanılarak Sınıflandırılması <u>Ahmet Kadir ARSLAN</u> , Songül ÇUĞLAN, Evren KÖSE, Cemil ÇOLAK
16:20-16:25	İstatistiksel Modellemeden Görsele Geçiş: Nonogram Zeynel Abidin ÖZTÜRK, <u>Seval KUL</u> , İbrahim TÜRKBEYLER, Azer ABİYEY
16:25-16:30	Genelleştirilmiş Additive Poisson Regresyon Modeli Yardımıyla Çevre kirliliği ve Meteorolojik Değişkenlerin Hastane Başvuruları ile İlişkilendirilmesi: Astım Nedenli Acil Başvurularının Modellenmesi <u>Seval KUL</u> , Mehmet KARADAĞ, Saim YOLOĞLU, Mustafa BOĞAN
16:30-16:35	Network Meta Analizi ve Sağlık Bilimleri Alanında Kullanımı <u>Pınar AMBARCIOĞLU</u> , Mehtap AKÇİL OK, Ufuk KAYA, Doğukan ÖZEN, Erman GÜLENDAG, İ. Safa GÜRCAN
16:35-16:40	Radyoloji Dergilerinde Yayınlanan Makalelerin İstatistiksel Hatalar Bakımından İncelenmesi <u>Pınar GÜNEL KARADENİZ</u> , Ender UZABACI, Sema ATIŞ, Fisun Kaşkıır KESİN, Fatma Ezgi CAN, İlker ERCAN
16:40-16:45	K-Ortalama Kümeleme Yönteminde Başlangıç Merkezlerinin Kümeleme Sonuçlarına Etkisi <u>Meryem GÖRAL YILDIZLI</u> , Zeliha Nazan ALPARSLAN
16:50-17:15	<u>MICROSOFT WORKSHOP</u> Microsoft Azure Üzerinde Bilimsel Araştırma Senaryoları
17:15-17:35	Bulut Tabanlı İstatistik Analiz Yazılımı Moderatör: Doç. Dr. Ahmet ÖZTÜRK Yrd. Doç. Dr. Gökmen ZARARSIZ, Yrd. Doç. Dr. Selçuk KORKMAZ, Araş. Gör. Dinçer GÖKSÜLÜK

XIX. Ulusal ve II. Uluslararası Biyoistatistik Kongresi
XIX. National and II. International Biostatistics Congress
25-28 Ekim 2017 Belek / Antalya

Salon II (Sunum Dili: Türkçe/İngilizce)	
10:50-12:20	Sözlü Bildiriler – I. Oturum
	Oturum Başkanları Prof. Dr. Yavuz SANİSOĞLU, Doç. Dr. Derya GÖKMEN, Yrd. Doç. Dr. Selçuk KORKMAZ
10:50-11:05	Novel Method for Epidemiological Information of Dental Clinics by Spatial Regression <u>Adnan KARAİBRAHİMOĞLU</u> , Arslan TERLEMEZ
11:05-11:20	Covariance Patterns for Linear Mixed Models <u>Hatice Tül Kübra AKDUR</u> , Deniz ÖZONUR, Hülya BAYRAK
11:20-11:35	Dealing with the Unknown: Handling Missing Data in Longitudinal Studies <u>Oya KALAYCIOĞLU</u>
11:35-11:50	Assessment and Comparison of Two Source Capture Recapture Methods' Performance <u>Musa Bashir ALBISHIR</u> , Ilker ERCAN
11:50-12:05	Exploring Mutation Profiles of Colon Adenocarcinoma to Infer Potential Ancestral History with Bayesian Networks <u>Uğur TOPRAK</u> , Kemal TURHAN
12:05-12:20	Chronological Age Classifier Enables Evaluation of Pro-Aging Factors at Single Cell Resolution <u>Sumeet Pal SINGH</u> , Gökmen ZARARSIZ, Sharan Kaur JANJUHA, Nikolay NINOV
12.20-13.45	Öğle Yemeği
13:45-15:15	Sözlü Bildiriler – II. Oturum
	Oturum Başkanları: Prof. Dr. Nural BEKİROĞLU, Yrd. Doç. Dr. Leman TOMAK
13:45-14:00	A Glimse at Long-read Sequencing and Megabase-sized Scaffolding Approaches <u>Vahap ELDEM</u> , Gökmen ZARARSIZ, Selçuk KORKMAZ, Dinçer GÖKSÜLÜK, Gözde ERTÜRK ZARARSIZ, Halil BİLGİN, Ahmet ÖZTÜRK
14:00-14:15	MLSeq 2.0: Machine Learning Interface for RNA Sequencing Data <u>Gökmen ZARARSIZ</u> , Dinçer GÖKSÜLÜK, Selçuk KORKMAZ, Vahap ELDEM, Bernd KLAUS, Ahmet ÖZTÜRK
14:15-14:30	Machine Learning Based Virtual Screening in Drug Discovery <u>Selçuk KORKMAZ</u> , Gökmen ZARARSIZ, Dinçer GÖKSÜLÜK, Necdet SÜT
14:30-14:45	Improved Estimation of Body Fat Percentage via Machine-Learning Approaches <u>Halil BİLGİN</u> , Gökmen ZARARSIZ, Volkan ÖZKAYA, Betül ÇİÇEK, Gözde ERTÜRK ZARARSIZ, Selçuk KORKMAZ, Dinçer GÖKSÜLÜK, Ahmet ÖZTÜRK
14:45-15:00	One-Way Tests in Independent Groups Designs: the onewaytests Web Interface <u>Osman DAĞ</u> , Anıl DOLGUN, N. Meric KONAR
15:00-15:15	EVALUATING TIME-DEPENDENT CUT-OFF VALUES FOR LONGITUDINAL DATA <u>N. Meric KONAR</u> , Eda KARAİSMAİLOĞLU, A. Ergun KARAAGAOĞLU

XIX. Ulusal ve II. Uluslararası Biyoistatistik Kongresi
XIX. National and II. International Biostatistics Congress
25-28 Ekim 2017 Belek / Antalya

15:15-15:30	Çay-Kahve Arası
15:30-16:50	Kısa Sözlü Bildiriler – III. Oturum
	Oturum Başkanları: Prof. Dr. C. Reha ALPAR, Doç. Dr. Gülhan ÖZEKİCİ TEMEL, Yrd. Doç. Dr. Adem DOĞANER
15:30-15:35	2x2 Tablolar için Kapsamlı Bir Web Yazılımı <u>Ahu DURMUŞÇELEBİ</u> , Gökmen ZARARSIZ, Selçuk KORKMAZ, Halil BİLGİN, Dinçer GÖKSÜLÜK, Gözde ERTÜRK ZARARSIZ, Ferhan ELMALI, Ahmet ÖZTÜRK
15:35-15:40	Boylamsal Veri Setinde Çok Düzeyli Analiz Yöntem Seçimi ve Bazı Özel Durumlar, Kısıtlılıklar <u>Alev BAKIR</u> , Mustafa Ş. ŞENOCAK
15:40-15:45	Sınıflama ve Regresyon Modellerinin Performanslarının Değerlendirilmesi. Bir Veri Seti Üzerine Uygulama <u>Hülya BİNOKAY</u> , Yaşar SERTDEMİR
15:50-15:55	İşgücüne Dâhil Olma Durumunu Etkileyen Faktörlerin Kategorik Regresyon Analizi ile İncelenmesi Sarp SAKLIYAN, Mehmet KIVRAK, <u>Cemil ÇOLAK</u>
15:55-16:00	Tıp Fakültesi Öğrencilerine Verilen Biyoistatistik Eğitiminde Yeni Yaklaşımlar <u>Ferhan ELMALI</u>
16:00-16:05	Adölesan Dönemde Alkol Tüketimi Risk Faktörlerinin Belirlenmesi ve Risk Tanı Sistemi Modeli <u>Zeliha AYDIN KASAP</u> , Burçin KURT, Kemal TURHAN
16:05-16:10	Bulanık Çıkarsama Sistemleri ile Lojistik Regresyon Sınıflandırıcılarının Performanslarının Apandisit Verisinde Karşılaştırılması <u>S. Kenan KÖSE</u> , İrem KAR, Batuhan BAKIRARAR
16:10-16:15	Bağımlı Değişken İle Ortak Değişken Arasındaki Korelasyona Göre Kovaryans Analizinin Gücü <u>Emre DEMİR</u> , Yasemin YAVUZ, Serdal Kenan KÖSE
16:15-16:20	Gözlemsel Çalışmalarda Ölçülemeyen Karıştırıcı Değişkenlerin Etkisi; Propensity Skor ve Duyarlılık Analizleri <u>Emre DEMİR</u> , Serhat HAYME, Derya GÖKMEN, Atilla Halil ELHAN, Serdal Kenan KÖSE
16:20-16:25	Ayrırlılığın güç olduğu bir klinik durumda farklı karar algoritmalarının performansı <u>Ahmet KESKİNOĞLU</u> , Suriye ÖZGÜR, Burak ORDİN, Önder YAVAŞCAN, Caner ALPARSLAN, Timur KÖSE, Pembe KESKİNOĞLU
16:25-16:30	Tıp Fakültelerinde İlk Biyoistatistik Dersi: Biyoistatistiğin Önemini Açıklama <u>Fatma Ezgi CAN</u> , İlker ERCAN, Deniz SIGIRLI, Guven OZKAYA, Zuleyha ALPER
16:30-16:35	Çoklu Uyum Analizi (MCA): Üniversite Öğrencilerinde Yeme Alışkanlıkları <u>Emrah G. ÖZGÜR</u> , Nural BEKİROĞLU
16:35-16:40	Bland Altman İstatistiksel Yöntemi ve Prostat Kanseri Bir Uygulama <u>Şeyma YAŞAR</u> , Bulut DURAL, Leyla KARACA, Zeynep TUNÇ, Harika G. BAĞ, Saim YOLOĞLU
16:40-16:45	Tıp Eğitiminde Sınavların Değerlendirilmesinde İç Tutarlılık Katsayıları Kullanılabilir mi? <u>Tanyeli KAZAZ</u> , Seval KUL
16:45-16:50	Toplam Prostat Spesifik Antijen Değerlerinin Yapay Zekâ Modellemesi ile Tahmini <u>Şeyma YAŞAR</u> , Ali BEYTUR, Cemil ÇOLAK

XIX. Ulusal ve II. Uluslararası Biyoistatistik Kongresi
XIX. National and II. International Biostatistics Congress
25-28 Ekim 2017 Belek / Antalya

27.10.2017 Cuma

Salon I (Sunum Dili: Türkçe)	
9:00-10:30	Sözlü Bildiriler – IV. Oturum
	Oturum Başkanları: Prof. Dr. Gülşah ŞEYDAOĞLU-Doç. Dr. Nurhan DOĞAN
9:00-9:40	Konferans II: Biyoistatistik Literatürünün Gelişimi ve Trend Konular; Bir Bibliyometrik Analiz Doç. Dr. Serdal Kenan KÖSE
9:45-10:00	Sınıf Dengeleyici: İki Sınıflı Verilerde Sınıf Dengesizliği Problemini Gidermek İçin Web Tabanlı Bir Yazılım <u>Ahmet Kadir ARSLAN, Cemil ÇOLAK</u>
10:00-10:15	Güvenirlilik Çalışmalarında Örnek Büyüklüğü Hesaplama Yaklaşımları <u>Sevinç Püren YÜCEL, Zeliha Nazan ALPARSLAN</u>
10:15-10:30	Çoklu Regresyon Modellerinde Örneklem Büyüklüğü ve Güç Hesabı: Bir Benzetim Uygulaması. <u>Ayhan PARMAKSIZ, Reha ALPAR</u>
10:30-10:45	Çay-Kahve Arası
10:45-12:15	Sözlü Bildiriler – V. Oturum
	Oturum Başkanları: Prof. Dr. İsmet DOĞAN-Doç. Dr. Timur KÖSE
10:45-11:00	Karşılaştırılacak Grup Sayısının Çok Fazla Olduğu Durumlarda Çoklu Karşılaştırma Problemi İçin Alternatif Bir Çözüm Yolu: ANOM Tekniği <u>Mehmet MENDEŞ, Soner YİĞİT</u>
11:00-11:15	Nokta İki Serili Korelasyon (Poin Biserial Correlation) Katsayısı ve İki Örnek t-testinin Performansları Bakımından Karşılaştırılması <u>Soner YİĞİT, Mehmet MENDEŞ</u>
11:15-11:30	Bağımlı Değişkenin Sürekli ve Tekrarlı Olduğu Denemelerde Özet İstatistikler Kullanılarak Örneklem Büyüklüğünün Hesaplanması <u>Duygu SİDDİKOĞLU, Funda Seher ÖZALP ATEŞ, Emre DEMİR, Yasemin YAVUZ</u>
11:30-11:45	Çay-Kahve Arası
11:45-12:00	Yanık Görüntülerinin Kümelenmesinde Bulanık Kümeleme Yaklaşımlarının Başarımlarının Ölçülmesi <u>Yeşim AKBAS, Tolga BERBER</u>
12:00-12:15	Kategorik Kümelenmiş Veri Analizi ve Bir Uygulama <u>Kübra Elif AKBAS, Harika Gözde GÖZÜKARA BAĞ</u>
12:15-12:30	Karar Modelleri için Olasılıksal Duyarlılık Analizi ve Uygulaması <u>Can ATEŞ, Yasemin YAVUZ</u>
12:30-13:00	Kısa Sözlü Bildiriler – VI. Oturum
	Oturum Başkanları: Prof. Dr. Fezan MUTLU, Yrd. Doç. Dr. Mesut AKYOL, Yrd. Doç. Dr. İsmail YILDIZ
12:30-12:35	Geometrik Morfometri Analizi ile Uyku Apneli Olgularda Kraniofasial Şeklinin 3 Boyutlu Analizi <u>Fatma Ezgi CAN, İlker ERCAN, Gokhan OCAKOĞLU, Senem TURAN OZDEMİR, Ezgi DEMİRDÖĞEN ÇETİNOĞLU, Ahmet URSAVAŞ</u>

XIX. Ulusal ve II. Uluslararası Biyoistatistik Kongresi
XIX. National and II. International Biostatistics Congress
25-28 Ekim 2017 Belek / Antalya

12:35-12:40	Akut Koroner Sendromlular için Web Tabanlı Yazılım Geliştirilmesi ve Tıbbi Veri Madenciliği Uygulaması <u>Emek GÜLDOĞAN</u> , Julide YAĞMUR, Cemil ÇOLAK
12:40-12:45	Genetik İlişki Çalışmalarında Veri Kalite Kontrolünün Yanlılık Üzerindeki Etkileri <u>Bahar TAŞDELEN</u>
12:45-12:50	Ridge Regresyon için Ayar Parametresi Seçimi Betül DAĞOĞLU, <u>Sevinç Püren YÜCEL</u> , Zeliha Nazan ALPARSLAN
12:50-12:55	İki Farklı Zaman Dğerlendirici iki farklı Zamanda Stlyoid Processlerinin üç boyutlu bilgisayarlı tomografi ölçümlerinin uyumunun test edilmesinde alternatif uyum yaklaşımları <u>Semra ERDOĞAN</u> , Kaan ESEN, Yüksel BALCI, Sermin TOK, Gülhan TEMEL, Anıl ÖZGÜR
12:55-13:00	LMS Yönteminin Uyum İyiliğinin Ölçülmesinde Kullanılan Yöntemlerin Karşılaştırılması <u>Batuhan BAKIRARAR</u> , İrem KAR, Atilla Halil ELHAN
13:00- 13:05	Suda Çözünür N-heterosiklik Karben Komplekslerinin Kansere Hücrelerine Karşı İlaç Potansiyellerinin Eğri Uydurma Yöntemlerinden İnhibe Eden Konsantrasyon (IC50) Değeriyle İncelenmesi <u>Şeyma YAŞAR</u> , Mitat AKKOÇ, Tuğba KUL KÖPRÜLÜ, Şaban TEKİN, Saim YOLOĞLU, Sedat YAŞAR
13:05-14:00	Öğle Yemeği
Salon II (Sunum Dili: Türkçe/İngilizce)	
9:00-10:30	Sözlü Bildiriler – IV. Oturum
	Oturum Başkanları: Prof. Dr. Atilla Halil ELHAN-Doç. Dr. Seval KUL
9:45-10:00	Measuring Degree of Distinguishability and Agreement in Ordinal Rating Scales <u>Ayfer Ezgi YILMAZ</u> , Serpil AKTAS
10:00-10:15	Çok Değişkenli Varyans Analizi Çözümlemesi için Normal Olmayan ve Dengesiz Örneklemelerin Test İstatistiklerinin I. Tip Hata Oranları Bakımından Karşılaştırılması <u>Mustafa Agah TEKİNDAL</u> , Özlem KAYMAZ, H. Emre KALE, Can ATEŞ
10:15-10:30	Veri Madenciliği Sınıflama Yöntemlerinin Acil Servis Verilerinde Uygulaması <u>Özlem ARIK</u> , Erdem KARABULUT, Emine KADIOĞLU
10:30-10:45	Çay-Kahve Arası
10:45-12:15	Sözlü Bildiriler – V. Oturum
	Oturum Başkanları: Prof. Dr. Erdem KARABULUT, Yrd. Doç. Dr. Semra ERDOĞAN
10:45-11:00	RServe ile Veri Madenciliği Algoritmalarının Web Uygulaması <u>Batuhan BAKIRARAR</u> , İrem KAR, Atilla Halil ELHAN
11:00-11:15	Çoklu Uyum Analizi ve Bir Uygulaması <u>Nihat Burak ZİHNİ</u> , Özlem UZUN, Gülay YEGİNOĞLU, Şahi Nur KALKIŞIM, Canan ERTEMOĞLU ÖKSÜZ
11:30-11:45	Çay-Kahve Arası
11:30-11:45	Kayıp Landmark Tahmininde Yaklaşım Önerilmesi ve Diğer Yöntemlerle Performansının Karşılaştırılması <u>Fatma Ezgi CAN</u> , İlker ERCAN

XIX. Ulusal ve II. Uluslararası Biyoistatistik Kongresi
XIX. National and II. International Biostatistics Congress
25-28 Ekim 2017 Belek / Antalya

11:45-12:00	Kategori Sayısı İkidenden Fazla Olan Bağımsız Değişkenlerde Johnson-Neyman Prosedürü <u>Tuğçe ŞENÇELİKEL</u> , H. Yağmur ZENGİN, Ersin ÖĞÜŞ, K. Setenay ÖNER
12:00-12:15	[0,1] Kapalı Aralığında Tanımlı Sürekli Bir Bağımlı Değişkenin Modellenmesinde Kullanılan Farklı Modelleme Stratejilerinin İncelenmesi <u>Hatice Yağmur ZENGİN</u> , Ersin ÖĞÜŞ, Erdem KARABULUT
12:30-13:00	Kısa Sözlü Bildiriler - VI. Oturum
	Oturum Başkanları: Prof. Dr. Mehtap AKÇİL OK, Doç. Dr. Bülent ÇELİK, Yrd. Doç. Dr. İlker ÜNAL
12:30-12:35	Sürücü Davranışları Ölçeğinin Geçerlik Güvenirlik Analizi ve Faktör Yapısı <u>Erkut BOLAT</u>
12:35-12:40	Çokmerkezli Araştırmalarda Tasarım Etkisi <u>İsmet DOĞAN</u> , Nurhan DOĞAN
12:40-12:45	Afyon Kocatepe Üniversitesi Tıp Fakültesi Öğrencilerinin Toplumsal Cinsiyet Rollerini <u>Nurhan DOĞAN</u> , Bilal YILMAZ, Abdullah Oğuz TAŞBAŞ, Berat Niyazi ÖLMEZ, Rümeyza KILINÇ, Saliha Nur SEZGIN, Tarık Alp ŞAHINOĞLU, Ömer Faruk BALLI
12:45-12:50	Akciğer Kanseri İnsidans ve Mortalitesinin Küresel Dağılımı <u>Nurhan DOĞAN</u> , İsmet DOĞAN
12:50-12:55	Panel Veri Analizi ve Beyin Göçü Verilerinde Uygulama <u>Asena Ayça ÖZDEMİR</u> , Gülhan OREKİCİ TEMEL
12:55-13:00	Biyoeşdeğerlik Çalışmalarında Birey İçi Değişkenlik Katsayısının Büyük Olduğu veya Bilinmediği Durumlarda Örneklem Genişlikleri Hesaplama Adımları <u>Asena Ayça ÖZDEMİR</u> , <u>Elif ERTAŞ</u> , Emine Arzu KANIK
13:00- 13:05	Farklı Sayıda Gözlem İçeren Veri Setleri İçin En Küçük Kareler Ve Temel Bileşenler Regresyonu Sonuçlarının Karşılaştırılması <u>Zeynep TUNÇ</u> , Şeyma YAŞAR, Harika Gözde GÖZÜKARA BAĞ, Saim YOLOĞLU
13:05-14:00	Öğle Yemeği

28.10.2017 Cumartesi

Salon I (Sunum Dili: Türkçe)	
9:00-10:30	Sözlü Bildiriler - VII. Oturum
	Oturum Başkanları: Doç. Dr. S. Kenan KÖSE, Doç. Dr. Jale KARAKAYA
9:00-9:15	Deep Learning ile Yeni Nesil Sekanslama Verilerinde Okuma Kalitelerinin Tahminlenmesi <u>Erdal COŞGUN</u>
9:15-9:30	Akademik İstatistiksel Danışmanlık Merkezleri Üzerine Bir İnceleme ve Model Önerisi <u>Bülent ÇELİK</u> , M. Akif BAKIR, M. Yavuz ATA
9:30-9:45	İnternet Tabanlı Sağkalım Olasılığı Hesaplayıcıların Güvenilirliği <u>Sinan UZUN</u> , Duygu SAÇAR, M. Akif ÖZTÜRK, P.Fulden YUMUK, Nural BEKİROĞLU
9:45-10:00	İkili Sonuçlu Bağımsız Kappa İstatistiklerinin Homojenlik Testleri İçin Bir R Paketi <u>Muammer ALBAYRAK</u> , Kemal TURHAN, Yasemin YAVUZ, Zeliha AYDIN KASAP

XIX. Ulusal ve II. Uluslararası Biyoistatistik Kongresi
XIX. National and II. International Biostatistics Congress
25-28 Ekim 2017 Belek / Antalya

10:00-10:15	Bulanık Çıkarsama Sistemleri ile Veri Madenciliği Yöntemlerinin Sınıflama Performansının Benzetim Çalışması ile Karşılaştırılması <u>İrem KAR</u> , Batuhan BAKIRARAR, S. Kenan KÖSE
10:15-10:30	Ölçeklerde Yapı Geçerliliğinin Değerlendirilmesi: Faktör Analizi ve Rasch Analizi <u>Funda Seher ÖZALP ATEŞ</u> , Derya GÖKMEN, Atilla Halil ELHAN, Şehim KUTLAY
10:30-10:45	Çay-Kahve Arası
10:45-12:00	Sözlü Bildiriler - VIII. Oturum
	Oturum Başkanları: Prof. Dr. Sıddık KESKİN, Doç. Dr. Ünal ERKORKMAZ, Yrd. Doç. Dr. Sevilay KARAHAN
10:45-11:00	İkiden Fazla Bağımsız Oran Karşılaştırılması <u>Soner YİĞİT</u> , Mehmet MENDEŞ
11:00-11:15	Meta-Analizde Güç Değerlendirmesi <u>Nazlı TOTİK</u> , Zeliha Nazan ALPARSLAN
11:15-11:30	Multipl Sklerozlu Hastaların El Yazılarının İstatistiksel Şekil Analizi. <u>Fatma Ezgi CAN</u> , İlker ERCAN, Güven ÖZKAYA, Özlem TAŞKAPILIOĞLU
11:30-11:45	Meta Analizinde Aykırı Değerlerin Tespiti ve Farklı Varyans Tahmin Yöntemleri ile Değerlendirilmesi <u>Mutlu UMAROĞLU</u> , Pınar ÖZDEMİR
11:45-12:00	Konvolüsyonel Sinir Ağları ile 3 Boyutlu Ortamda Oluşturulan Örüntülerin Metin Olarak Tanınması <u>Adem DOĞANER</u> , Cemil ÇOLAK
12:00-12:30	Kapanış
Salon II (Sunum Dili: Türkçe)	
9:00-10:30	Sözlü Bildiriler - VII. Oturum
	Oturum Başkanları: Yrd. Doç. Dr. Gökmen ZARARSIZ
9:00-9:15	Lojistik Regresyon ve Karar Ağacı Tabanlı Yöntemlerinin Sınıflama Performanslarının Karşılaştırılması. <u>Yaşar SERTDEMİR</u> , İlker ÜNAL
9:15-9:30	Dayanıklı regresyon yöntemi (M): Sağlık verisinde bir uygulama <u>Suriye ÖZGÜR</u> , Özge ÇAPKAN, Ş. Büşra MERT, Mert İŞERİ, Timur KÖSE, Pembe KESKİNOĞLU, Mehmet ORMAN
9:30-9:45	Logaritmik Doğrusal Modeller ve Uyum Analizinin Birlikte Kullanımı: Lise Öğrencilerinin Sigara İçme Alışkanlıklarına Uygulanması <u>Fisun KAŞKIR KESİN</u> , Veysel YILMAZ
9:45-10:00	Yarı En Küçük Kareler Regresyonu <u>Erdoğan ASAR</u> , Erdem KARABULUT
10:00-10:15	Sürekli Yapıdaki Çoklu Tanı Testleri İçin Birleştirme Yöntemlerinin ROC Eğrisi Analizi Kullanarak Karşılaştırılması <u>Muzaffer BİLGİN</u> , Ertuğrul ÇOLAK
10:15-10:30	Bayesian Yöntemi Kullanılarak Tekrarlanan Ölçümlerde Kovaryans Yapısının Modellenmesi <u>Mehmet Ziya FIRAT</u>
10:30-10:45	Çay-Kahve Arası

XIX. Ulusal ve II. Uluslararası Biyoistatistik Kongresi
XIX. National and II. International Biostatistics Congress
25-28 Ekim 2017 Belek / Antalya

10:45-11:45	Sözlü Bildiriler – VIII. Oturum
	Oturum Başkanları: Prof. Dr. İlker ERCAN, Doç. Dr. Yaşar SERTDEMİR, Yrd. Doç. Dr. Osman DEMİR
10:45-11:00	Diyabetik Retinopati Hastalığında, Eksuda Lezyonlarının Tespiti ve Sınıflandırılması <u>Uğur ŞEVİK</u>
11:00-11:15	Kısmi En Küçük Kareler Yöntemi ile İki Aşamalı En Küçük Kareler Yöntemlerinin İncelenmesi: Sağlık Alanında Bir Uygulama <u>Sıddık KESKİN</u>
11:15-11:30	Metabolomik Verilerin Sınıflandırılmasında Kısmi En Küçük Kareler Ayırma Analizi Yaklaşımı <u>Meltem ÜNLÜSAVURAN</u> , Gökmen ZARARSIZ, Funda İPEKTEN, Gözde ERTÜRK ZARARSIZ, Selçuk KORKMAZ, Dinçer GÖKSÜLÜK, Halef Okan DOĞAN, Vahap ELDEM, Ahmet ÖZTÜRK
11:30-11:45	Metabolomik biyobelirteçlerinin tespitinde ANOVA-PCA yaklaşımı <u>Funda İPEKTEN</u> , Gökmen ZARARSIZ, Meltem ÜNLÜSAVURAN, Gözde ERTÜRK ZARARSIZ, Selçuk KORKMAZ, Dinçer GÖKSÜLÜK, Halef Okan DOĞAN, Vahap ELDEM, Ahmet ÖZTÜRK
	Meta-Analizinde Odds Oranlarının Aşırı Yayılım Sorununa Beta-Binomial Model yaklaşımı Mehtap AKÇİL OK
12:00-12:30	Kapanış

SÖZLÜ BİLDİRİLER / ORAL PRESENTATIONS

	<u>Sayfa</u>
<u>Soner YİĞİT</u> , Mehmet MENDEŞ. Ortak Kontrol Grubunun Bulunduğu Faktöriyel Denemelerin İstatistiksel Analizleri: Monte Carlo Simulasyon Çalışması	1
<u>Merve BAŞOL</u> , Dinçer GÖKSÜLÜK, A. Ergun KARAAĞAOĞLU. Bootstrap Güven Aralığı Yöntemlerinin Karşılaştırılması	2
<u>Büşra EMİR</u> , Ertuğrul ÇOLAK. Tedavi-Biyomarker Etkileşiminin BATE Eğrisi ile İncelenmesi	4
<u>Meryem ÇAVUŞOĞLU</u> , Gökmen ZARARSIZ, Gözde Ertürk ZARARSIZ, Selçuk KORKMAZ, Dinçer GÖKSÜLÜK, Mümtaz MAZICIOĞLU, Ahmet ÖZTÜRK. GAMLSS Modellerinin Model Yeterliliğinin Belirlenmesinde Worm Eğrileri	5
<u>Adem DOĞANER</u> , Cemil ÇOLAK. Veri Füzyon Yöntemleri ile Çoklu Sensör Görüntülerinden Bilgi Çıkarımı	6
<u>Aslı BOZER</u> , Osman DAĞ, Sevilay KARAHAN. Kohonen Öz Örgütlemeli Haritalama Yöntemi İle Psikotik Hastalıkların Kümelenmesi	7
<u>Durdu KARASOY</u> , Hatice IŞIK, Nihal ATA TUTKUN. Yaşam Çözümlemesinde Eğilim Skoru	9
<u>Ebru TURGAL</u> , Beyza DOĞANAY ERDOĞAN, S. Kenan KÖSE. Tekrarlı Gözlem ve Sağkalım Verilerinin Birleşik Model Performans Değerlendirmesi: Bir Simülasyon Çalışması	10
<u>Hülya ÖZEN</u> , Cengiz BAL. Orantılı Hazard Durumunun İhlalinde Log-Rank Testinin Gücünün İncelenmesi	12
<u>Merve TÜRKEGÜN</u> , Gülhan OREKİCİ TEMEL, İrem ERSÖZ KAYA. Yaşam Süresi Verilerinin Çok Değişkenli Uyumlu Regresyon Analizinde Kullanımı	13
<u>İlker ÜNAL</u> , Yaşar SERTDEMİR, Ceren EFE. Zamana Bağlı Bir Belirteç İçin Optimal Kesim Noktası Belirleme	14
<u>Ceren EFE</u> , İlker ÜNAL. Zamana Bağlı ROC Analizi Yöntemlerinin Alan Tahminlerinin Karşılaştırılması	15
<u>Ahmet Kadir ARSLAN</u> , Cemil ÇOLAK. Sınıf Dengeleyici: İki Sınıflı Verilerde Sınıf Dengesizliği Problemini Gidermek İçin Web Tabanlı Bir Yazılım	16
<u>Sevinç Püren YÜCEL</u> , Zeliha Nazan ALPARSLAN. Güvenirlilik Çalışmalarında Örnek Büyüklüğü Hesaplama Yaklaşımları	18
<u>Ayhan PARMAKSIZ</u> , Reha ALPAR. Çoklu Regresyon Modellerinde Örneklem Büyüklüğü ve Güç Hesabı: Bir Benzetim Uygulaması.	19
Mehmet MENDEŞ, <u>Soner YİĞİT</u> . Karşılaştırılacak Grup Sayısının Çok Fazla Olduğu Durumlarda Çoklu Karşılaştırma Problemi İçin Alternatif Bir Çözüm Yolu: ANOM Tekniği	21
<u>Soner YİĞİT</u> , Mehmet MENDEŞ. Nokta İki Serili Korelasyon (Poin Biserial Correlation) Katsayısı ve İki Örnek t-testinin Performansları Bakımından Karşılaştırılması	22
<u>Duygu SİDDİKOĞLU</u> , Funda Seher ÖZALP ATEŞ, Emre DEMİR, Yasemin YAVUZ. Bağımlı Değişkenin Sürekli ve Tekrarlı Olduğu Denemelerde Özet İstatistikler Kullanılarak Örneklem Büyüklüğünün Hesaplanması	23
Yeşim AKBAŞ, <u>Tolga BERBER</u> . Yanık Görüntülerinin Kümelenmesinde Bulanık Kümeleme Yaklaşımlarının Başarımlarının Ölçülmesi	24
<u>Kübra Elif AKBAŞ</u> , Harika Gözde GÖZÜKARA BAĞ. Kategorik Kümelenmiş Veri Analizi ve Bir Uygulama	25

XIX. Ulusal ve II. Uluslararası Biyoistatistik Kongresi
XIX. National and II. International Biostatistics Congress
25-28 Ekim 2017 Belek / Antalya

	<u>Sayfa</u>
<u>Can ATEŞ</u> , Yasemin YAVUZ. Karar Modelleri için Olasılıksal Duyarlılık Analizi ve Uygulaması	26
<u>Erdal COŞGUN</u> . Deep Learning ile Yeni Nesil Sekanslama Verilerinde Okuma Kalitelerinin Tahminlenmesi	27
<u>Bülent ÇELİK</u> , M. Akif BAKIR, M. Yavuz ATA. Akademik İstatistiksel Danışmanlık Merkezleri Üzerine Bir İnceleme ve Model Önerisi	29
<u>Sinan UZUN</u> , Duygu SAÇAR, M. Akif ÖZTÜRK, P. Fulden YUMUK, Nural BEKİROĞLU. İnternet Tabanlı Sağlıkım Olasılığı Hesaplayıcıların Güvenilirliği	30
<u>Muammer ALBAYRAK</u> , Kemal TURHAN, Yasemin YAVUZ, Zeliha AYDIN KASAP. İkili Sonuçlu Bağımsız Kappa İstatistiklerinin Homojenlik Testleri İçin Bir R Paketi	32
<u>İrem KAR</u> , Batuhan BAKIRARAR, S. Kenan KÖSE. Bulanık Çıkarsama Sistemleri ile Veri Madenciliği Yöntemlerinin Sınıflama Performansının Benzetim Çalışması ile Karşılaştırılması	33
<u>Funda Seher ÖZALP ATEŞ</u> , Derya GÖKMEN, Atilla Halil ELHAN, Şehim KUTLAY. Ölçeklerde Yapı Geçerliliğinin Değerlendirilmesi: Faktör Analizi ve Rasch Analizi	34
<u>Soner YİĞİT</u> , Mehmet MENDEŞ. İkiden Fazla Bağımsız Oran Karşılaştırılması	35
<u>Nazlı TOTİK</u> , Zeliha Nazan ALPARSLAN. Meta-Analizde Güç Değerlendirmesi	36
<u>Fatma Ezgi CAN</u> , İlker ERCAN, Güven ÖZKAYA, Özlem TAŞKAPILIOĞLU. Multipl Sklerozlu Hastaların El Yazılarının İstatistiksel Şekil Analizi.	37
<u>Mutlu UMAROĞLU</u> , Pınar ÖZDEMİR. Meta Analizinde Aykırı Değerlerin Tespiti ve Farklı Varyans Tahmin Yöntemleri İle Değerlendirilmesi	38
<u>Adem DOĞANER</u> , Cemil ÇOLAK. Konvolüsyonel Sinir Ağları ile 3 Boyutlu Ortamda Oluşturulan Örüntülerin Metin Olarak Tanınması	39
<u>Adnan KARAIBRAHİMOĞLU</u> , Arslan TERLEMEZ. Novel Method for Epidemiological Information of Dental Clinics by Spatial Regression	40
<u>Hatice Tül Kübra AKDUR</u> , Deniz ÖZONUR, Hülya BAYRAK. Covariance Patterns for Linear Mixed Models	41
<u>Oya KALAYCIOĞLU</u> . Dealing with the Unknown: Handling Missing Data in Longitudinal Studies	42
<u>Musa Bashir ALBISHIR</u> , İlker ERCAN. Assessment and Comparison of Two Source Capture Recapture Methods' Performance	43
<u>Uğur TOPRAK</u> , Kemal TURHAN. Exploring Mutation Profiles of Colon Adenocarcinoma to Infer Potential Ancestral History with Bayesian Networks	44
<u>Sumeet Pal SINGH</u> , Gökmen ZARARSIZ, Sharan Kaur JANJUHA, Nikolay NINOV. Chronological Age Classifier Enables Evaluation of Pro-Aging Factors at Single Cell Resolution	45
<u>Vahap ELDEM</u> , Gökmen ZARARSIZ, Selçuk KORKMAZ, Dinçer GÖKSÜLÜK, Gözde ERTÜRK ZARARSIZ, Halil BİLGİN, Ahmet ÖZTÜRK. A Glimse at Long-read Sequencing and Megabase-sized Scaffolding Approaches	46
<u>Gökmen ZARARSIZ</u> , Dinçer GÖKSÜLÜK, Selçuk KORKMAZ, Vahap ELDEM, Bernd KLAUS, Ahmet ÖZTÜRK. MLSeq 2.0: Machine Learning Interface for RNA Sequencing Data	48
<u>Selçuk KORKMAZ</u> , Gökmen ZARARSIZ, Dinçer GÖKSÜLÜK, Necdet SÜT. Machine Learning Based Virtual Screening in Drug Discovery	49
<u>Halil BİLGİN</u> , Gökmen ZARARSIZ, Volkan ÖZKAYA, Betül ÇİÇEK, Gözde ERTÜRK ZARARSIZ, Selçuk KORKMAZ, Dinçer GÖKSÜLÜK, Ahmet ÖZTÜRK. Improved Estimation of Body Fat Percentage via Machine-Learning Approaches	50

XIX. Ulusal ve II. Uluslararası Biyoistatistik Kongresi
XIX. National and II. International Biostatistics Congress
25-28 Ekim 2017 Belek / Antalya

	<u>Sayfa</u>
<u>Osman DAĞ</u> , Anıl DOLGUN, N. Meric KONAR. One-Way Tests in Independent Groups Designs: the onewaytests Web Interface	52
<u>N. Meric KONAR</u> , Eda KARAİSMAİLOĞLU, A. Ergun KARAAĞAOĞLU. Evaluating Time-Dependent Cut-Off Values For Longitudinal Data	54
<u>Ayfer Ezgi YILMAZ</u> , Serpil AKTAS. Measuring Degree of Distinguishability and Agreement in Ordinal Rating Scales	56
<u>Mustafa Agah TEKİNDAL</u> , Özlem KAYMAZ, H. Emre KALE, Can ATEŞ. Çok Değişkenli Varyans Analizi Çözümlemesi için Normal Olmayan ve Dengesiz Örneklemelerin Test İstatistiklerinin I. Tip Hata Oranları Bakımından Karşılaştırılması	57
<u>Özlem ARIK</u> , Erdem KARABULUT, Emine KADIOĞLU. Veri Madenciliği Sınıflama Yöntemlerinin Acil Servis Verilerinde Uygulaması	58
<u>Batuhan BAKIRARAR</u> , İrem KAR, Atilla Halil ELHAN. RServe ile Veri Madenciliği Algoritmalarının Web Uygulaması	60
<u>Nihat Burak ZİHNİ</u> , Özlem UZUN, Gülay YEGİNOĞLU, Şahi Nur KALKIŞIM, Canan ERTEMOĞLU ÖKSÜZ. Çoklu Uyum Analizi ve Bir Uygulaması	61
<u>Fatma Ezgi CAN</u> , İlker ERCAN. Kayıp Landmark Tahmininde Yaklaşım Önerilmesi ve Diğer Yöntemlerle Performansının Karşılaştırılması	63
<u>Tuğçe ŞENÇELİKEL</u> , H. Yağmur ZENGİN, Ersin ÖĞÜŞ, K. Setenay ÖNER. Kategori Sayısı İkidenden Fazla Olan Bağımsız Değişkenlerde Johnson-Neyman Prosedürü	64
<u>Hatice Yağmur ZENGİN</u> , Ersin ÖĞÜŞ, Erdem KARABULUT. [0,1] Kapalı Aralığında Tanımlı Sürekli Bir Bağımlı Değişkenin Modellenmesinde Kullanılan Farklı Modelleme Stratejilerinin İncelenmesi	65
<u>Yaşar SERTDEMİR</u> , İlker ÜNAL. Lojistik Regresyon ve Karar Ağacı Tabanlı Yöntemlerinin Sınıflama Performanslarının Karşılaştırılması.	67
<u>Suriye ÖZGÜR</u> , Özge ÇAPKAN, Ş. Büşra MERT, Mert İŞERİ, Timur KÖSE, Pembe KESKİNOĞLU, Mehmet ORMAN. Dayanıklı regresyon yöntemi (M): Sağlık verisinde bir uygulama	68
<u>Fisun KAŞKIR KESİN</u> , Veysel YILMAZ. Logaritmik Doğrusal Modeller ve Uyum Analizinin Birlikte Kullanımı: Lise Öğrencilerinin Sigara İçme Alışkanlıklarına Uygulanması	69
<u>Erdoğan ASAR</u> , Erdem KARABULUT. Yarı En Küçük Kareler Regresyonu	70
<u>Muzaffer BİLGİN</u> , Ertuğrul ÇOLAK. Sürekli Yapıdaki Çoklu Tanı Testleri İçin Birleştirme Yöntemlerinin ROC Eğrisi Analizi Kullanarak Karşılaştırılması	72
<u>Mehmet Ziya FIRAT</u> . Bayesian Yöntemi Kullanılarak Tekrarlanan Ölçümlerde Kovaryans Yapısının Modellenmesi	73
<u>Uğur ŞEVİK</u> . Diyabetik Retinopati Hastalığında, Eksuda Lezyonlarının Tespiti ve Sınıflandırılması	74
<u>Siddik KESKİN</u> . Kısmi En Küçük Kareler Yöntemi ile İki Aşamalı En Küçük Kareler Yöntemlerinin İncelenmesi: Sağlık Alanında Bir Uygulama	76
<u>Meltem ÜNLÜSAVURAN</u> , Gökmen ZARARSIZ, Funda İPEKTEN, Gözde ERTÜRK ZARARSIZ, Selçuk KORKMAZ, Dinçer GÖKSÜLÜK, Halef Okan DOĞAN, Vahap ELDEM, Ahmet ÖZTÜRK. Metabolomik Verilerin Sınıflandırılmasında Kısmi En Küçük Kareler Ayırma Analizi Yaklaşımı	77
<u>Funda İPEKTEN</u> , Gökmen ZARARSIZ, Meltem ÜNLÜSAVURAN, Gözde ERTÜRK ZARARSIZ, Selçuk KORKMAZ, Dinçer GÖKSÜLÜK, Halef Okan DOĞAN, Vahap ELDEM, Ahmet ÖZTÜRK. Metabolomik biyobelirteçlerinin tespitinde ANOVA-PCA yaklaşımı	78
<u>Mehtap AKÇİL OK</u> . Meta-Analizinde Odds Oranlarının Aşırı Yayılım Sorununa Beta-Binomiyal Model yaklaşımı	79

POSTER BİLDİRİLERİ / POSTER PRESENTATIONS

	<u>Sayfa</u>
<u>Zeliha Nazan ALPARSLAN</u> , Nazlı TOTİK. Okuduğumuz Meta-analizin Gücünü Biliyor muyuz?	82
<u>Erkut BOLAT</u> , Sedat ALTIN. Türkiye’de Solunum Hastalıkları Ölümlerinin Zaman Serisi Analizi	83
Vedia Bennu GİLAN, <u>Semra ERDOĞAN</u> , Seyran BOZKURT BABUŞ. Zaman Serileri Analiz Yöntemlerinin Sağlık Alanında Kullanılması	84
Yunus Emre KUYUCU, <u>Osman DEMİR</u> . Güvenilirliği Anlamak ve Cronbach’ın Alfa Katsayısını Yorumlamak	85
<u>Doğukan ÖZEN</u> , Hülya ÖZEN, Pınar AMBARCIOĞLU, Ufuk KAYA, Erman GÜLENDAG, İ. Safa GÜRCAN. Yaşam Analizlerinde Esnek Parametrik Bir Yaklaşım: Royston-Parmar Modelleri	87
<u>Didem DERİCİ YILDIRIM</u> , Bahar TAŞDELEN. Kırılganlık İndeksi Kullanılarak Randomize Kontrollü Çalışma Sonuçlarının Güvenilirliğinin Değerlendirilmesi	88
Betül DAĞOĞLU, <u>Sevinç Püren YÜCEL</u> , Zeliha Nazan ALPARSLAN. Lineer Regresyonda Değişen Varyansın Belirlenmesi ve Giderilmesi	89
<u>İsmet DOĞAN</u> , Nurhan DOĞAN. Kalıtım için İstatistiksel Ölçüler	90
<u>Pınar GÜNEL KARADENİZ</u> , Tuba DENKÇEKEN. Fiber Optik Yansıma Spektroskopisi Sistemi ile Tümör Teşhisi Bir Meta Analiz Çalışması	92
<u>Ahmet Kadir ARSLAN</u> , Songül ÇUĞLAN, Evren KÖSE, Cemil ÇOLAK. Kronik Obstrüktif Akciğer Hastalığının Çeşitli Makine Öğrenmesi Yöntemleri Kullanılarak Sınıflandırılması	93
Zeynel Abidin ÖZTÜRK, <u>Seval KUL</u> , İbrahim TÜRKBEYLER, Azer ABİYEYEV. İstatistiksel Modellemeye Görsel Geçiş: Nonogram	94
<u>Seval KUL</u> , Mehmet KARADAĞ, Saim YOLOĞLU, Mustafa BOĞAN. Genelleştirilmiş Additive Poisson Regresyon Modeli Yardımıyla Çevre kirliliği ve Meteorolojik Değişkenlerin Hastane Başvuruları ile İlişkilendirilmesi: Astım Nedenli Acil Başvurularının Modellenmesi	95
<u>Pınar AMBARCIOĞLU</u> , Mehtap AKÇİL OK, Ufuk KAYA, Doğukan ÖZEN, Erman GÜLENDAG, İ. Safa GÜRCAN. Network Meta Analizi ve Sağlık Bilimleri Alanında Kullanımı	96
<u>Pınar Günel KARADENİZ</u> , Ender UZABACI, Sema ATIŞ, Fisun Kaşır KESİN, Fatma Ezgi CAN, İlker ERCAN. Radyoloji Dergilerinde Yayımlanan Makalelerin İstatistiksel Hatalar Bakımından İncelenmesi	97
<u>Meryem GÖRAL YILDIZLI</u> , Zeliha Nazan ALPARSLAN. K-Ortalama Kümeleme Yönteminde Başlangıç Merkezlerinin Kümeleme Sonuçlarına Etkisi	98
<u>Fatma Ezgi CAN</u> , İlker ERCAN, Gökhan OCAKOĞLU, Senem TURAN OZDEMİR, Ezgi DEMİRDÖĞEN ÇETİNOĞLU, Ahmet URSAVAŞ. Geometrik Morfometri Analizi ile Uyku Apneli Olgularda Kraniofasial Şeklinin 3 Boyutlu Analizi	99
<u>Emek GÜLDOĞAN</u> , Julide YAĞMUR, Cemil ÇOLAK. Akut Koroner Sendromlular için Web Tabanlı Yazılım Geliştirilmesi ve Tıbbi Veri Madenciliği Uygulaması	100
<u>Bahar TAŞDELEN</u> . Genetik İlişki Çalışmalarında Veri Kalite Kontrolünün Yanlılık Üzerindeki Etkileri	101

XIX. Ulusal ve II. Uluslararası Biyoistatistik Kongresi
XIX. National and II. International Biostatistics Congress
25-28 Ekim 2017 Belek / Antalya

	<u>Sayfa</u>
<u>Sevinç Püren YÜCEL</u> , Betül DAĞOĞLU, Zeliha Nazan ALPARSLAN. Ridge Regresyon için Ayar Parametresi Seçimi	102
<u>Semra ERDOĞAN</u> , Kaan ESEN, Yüksel BALCI, Sermin TOK, Gülhan TEMEL, Anıl ÖZGÜR. İki Farklı Zaman Değerlendirici iki farklı Zamanda Stlyoid Processlerinin üç boyutlu bilgisayarlı tomografi ölçümlerinin uyumunun test edilmesinde alternatif uyum yaklaşımları	103
<u>Batuhan BAKIRARAR</u> , İrem KAR, Atilla Halil ELHAN. LMS Yönteminin Uyum İyiliğinin Ölçülmesinde Kullanılan Yöntemlerin Karşılaştırılması	104
<u>Seyma YAŞAR</u> , Mitat AKKOÇ, Tuğba KUL KÖPRÜLÜ, Şaban TEKİN, Saim YOLOĞLU, Sedat YAŞAR. Suda Çözünür N-heterosiklik Karben Komplekslerinin Kansere Hücrelerine Karşı İlaç Potansiyellerinin Eğri Uydurma Yöntemlerinden İnhibe Eden Konsantrasyon (IC50) Değeriyle İncelenmesi	105
<u>Ahu DURMUŞCELEBİ</u> , Gökmen ZARARSIZ, Selçuk KORKMAZ, Halil BİLGİN, Dinçer GÖKSÜLÜK, Gözde ERTÜRK ZARARSIZ, Ferhan ELMALI, Ahmet ÖZTÜRK. 2x2 Tablolar için Kapsamlı Bir Web Yazılımı	106
<u>Alev BAKIR</u> , Mustafa Ş. ŞENOCAR. Boylamsal Veri Setinde Çok Düzeyli Analiz Yöntem Seçimi ve Bazı Özel Durumlar, Kısıtlılıklar	107
<u>Hülya BİNOKAY</u> , Yaşar SERTDEMİR. Sınıflama ve Regresyon Modellerinin Performanslarının Değerlendirilmesi. Bir Veri Seti Üzerine Uygulama	108
Sarp SAKLIYAN, Mehmet KIVRAK, <u>Cemil ÇOLAK</u> . İşgücüne Dâhil Olma Durumunu Etkileyen Faktörlerin Kategorik Regresyon Analizi ile İncelenmesi	109
<u>Ferhan ELMALI</u> . Tıp Fakültesi Öğrencilerine Verilen Biyoistatistik Eğitiminde Yeni Yaklaşımlar	110
<u>Zeliha AYDIN KASAP</u> , Burçin KURT, Kemal TURHAN. Adölesan Dönemde Alkol Tüketimi Risk Faktörlerinin Belirlenmesi ve Risk Tanı Sistemi Modeli	112
<u>S. Kenan KÖSE</u> , İrem KAR, Batuhan BAKIRARAR. Bulanık Çıkarsama Sistemleri ile Lojistik Regresyon Sınıflandırıcılarının Performanslarının Apandisit Verisinde Karşılaştırılması	113
<u>Emre DEMİR</u> , Yasemin YAVUZ, Serdal Kenan KÖSE. Bağımlı Değişken İle Ortak Değişken Arasındaki Korelasyona Göre Kovaryans Analizinin Gücü	114
<u>Emre DEMİR</u> , Serhat HAYME, Derya GÖKMEN, Atilla Halil ELHAN, Serdal Kenan KÖSE. Gözlemsel Çalışmalarda Ölçülemeyen Karıştırıcı Değişkenlerin Etkisi; Propensity Skor ve Duyarlılık Analizleri	115
Ahmet KESKİNOĞLU, <u>Suriye ÖZGÜR</u> , Burak ORDİN, Önder YAVAŞCAN, Caner ALPARSLAN, Timur KÖSE, Pembe KESKİNOĞLU. Ayırıcılığın güç olduğu bir klinik durumda farklı karar algoritmalarının performansı	116
<u>Fatma Ezgi CAN</u> , İlker ERCAN, Deniz SIGIRLI, Güven OZKAYA, Zuleyha ALPER. Tıp Fakültelerinde İlk Biyoistatistik Dersi: Biyoistatistiğin Önemini Açıklama	117
Emrah G. ÖZGÜR, <u>Nural BEKİROĞLU</u> . Çoklu Uyum Analizi (MCA): Üniversite Öğrencilerinde Yeme Alışkanlıkları	118
<u>Seyma YAŞAR</u> , Bulut DURAL, Leyla KARACA, Zeynep TUNÇ, Harika G. BAÇ, Saim YOLOĞLU. Bland Altman İstatistiksel Yöntemi ve Prostat Kanseri Bir Uygulama	119
Tanyeli KAZAZ, <u>Seval KUL</u> . Tıp Eğitiminde Sınavların Değerlendirilmesinde İç Tutarlılık Katsayıları Kullanılabilir mi?	121

XIX. Ulusal ve II. Uluslararası Biyoistatistik Kongresi
XIX. National and II. International Biostatistics Congress
25-28 Ekim 2017 Belek / Antalya

	<u>Sayfa</u>
Şeyma YAŞAR, Ali BEYTUR, Cemil ÇOLAK. Toplam Prostat Spesifik Antijen Değerlerinin Yapay Zekâ Modellemesi ile Tahmini	122
<u>Erkut BOLAT</u> . Sürücü Davranışları Ölçeğinin Geçerlik Güvenirlik Analizi ve Faktör Yapısı	123
<u>İsmet DOĞAN</u> , Nurhan DOĞAN. Çokmerkezli Araştırmalarda Tasarım Etkisi	124
<u>Nurhan DOĞAN</u> , Bilal YILMAZ, Abdullah Oğuz TAŞBAŞ, Berat Niyazi ÖLMEZ, Rümeysa KILINÇ, Saliha Nur SEZGIN, Tarık Alp ŞAHINOĞLU, Ömer Faruk BALLI. Afyon Kocatepe Üniversitesi Tıp Fakültesi Öğrencilerinin Toplumsal Cinsiyet Roller	126
<u>Nurhan DOĞAN</u> , İsmet DOĞAN. Akciğer Kanseri İnsidans ve Mortalitesinin Küresel Dağılımı	127
<u>Asena Ayça ÖZDEMİR</u> , Gülhan OREKİCİ TEMEL. Panel Veri Analizi ve Beyin Göçü Verilerinde Uygulama	128
Asena Ayça ÖZDEMİR, <u>Elif ERTAŞ</u> , Emine Arzu KANIK. Biyoesdeğerlik Çalışmalarında Birey İçi Değişkenlik Katsayısının Büyük Olduğu veya Bilinmediği Durumlarda Örneklem Genişlikleri Hesaplama Adımları	129
<u>Zeynep TUNÇ</u> , Şeyma YAŞAR, Harika Gözde GÖZÜKARA BAĞ, Saim YOLOĞLU. Farklı Sayıda Gözlem İçeren Veri Setleri İçin En Küçük Kareler Ve Temel Bileşenler Regresyonu Sonuçlarının Karşılaştırılması	130

KONFERANS - I

Kanonik korelasyon analizi ve ilişkili modeller

Sıddık KESKİN

Van Yüzüncü Yıl Üniversitesi, Tıp Fakültesi, Biyoistatistik AD, Tuşba-Van

Amaç: Bu çalışmada, Kanonik korelasyon analizi ve bununla ilişkili modellerden olan Redundancy analizi ile Kanonik uyum analizi incelenerek sağlık alanında bir uygulama yapılmıştır.

Yöntem: Kanonik korelasyon analizi: Kanonik korelasyon analizi, **X** ve **Y** değişken kümeleri arasındaki korelasyonu maksimum yapacak şekilde, orijinal değişkenlerin doğrusal kombinasyonlarından oluşan yeni değişkenler elde etmeyi amaçlar. Orijinal değişkenlerin doğrusal kombinasyonlarından oluşan yeni değişkenlere kanonik değişken adı verilir. Bu kanonik değişken çiftleri arasındaki hesaplanan korelasyonlara ise kanonik korelasyon denir. Kanonik korelasyon analizi, boyut indirgeme amaçlı kullanılabilir.

Redundancy analizi: Redundancy analizi, **X** ve **Y** gibi iki değişken kümesini içeren tablolar veya veri setleri arasındaki ilişkiyi analiz etmeye yarayan yöntemlerden birisidir. Kanonik korelasyon analizi simetrik bir yöntem, Redundancy analizi ise simetrik olmayan bir yöntemdir. Kanonik korelasyon analizinde bileşenler yada yeni değişkenler, veri setleri arasındaki korelasyonu maksimum yapacak şekilde oluşturulurken, Redundancy analizinde, **Y** değişken seti ile mümkün olduğunca yüksek korelasyonlu olacak şekilde, **X** değişken setinden bileşenler veya yeni değişkenler oluşturulur. Sonraki adımda, **Y** değişken setinden bileşenler elde edilir. Böylece, bunlar daha önce **X** değişken setinden elde edilmiş olan bileşenler ile olabildiğince yüksek korelasyonlu olur.

Kanonik uyum analizi: Kanonik uyum analizi, bir yörede bulunan topluluklar (**Y** veri seti) ile çevresel değişkenler (**Z** veri seti) arası ilişkiyi incelemek üzere geliştirilmiştir. Kanonik uyum analizi, uyum analizinin açıklayıcı veya tahmin edici değişkenler ile genişletilmiş hali olup, uyum analizi ile regresyon analizinin birleşimi olarak düşünülebilir.

Çalışmada, yöntemler ile ilgili uygulama yapmak üzere; sağlık alanında, 72 bireye ait, 7 değişkenden oluşan veri seti kullanılmıştır. Değişkenlerden BKİ, Yaş, ALT ve AST **X** değişken seti, TPSA, SPSA ve GLEASON skalası **Y** değişken seti olarak alınmıştır.

Bulgular: Kanonik korelasyon analizinde birinci kanonik değişken çifti arasındaki kanonik korelasyon % 61.1olarak bulunmuştur. Redundancy analizinde, constrained ve unconstrained varyans oranları sırası ile % 40.246 ve %59.74 olarak bulunmuştur. Kanonik uyum analizinde ise birinci ve ikinci boyuta ait varyans açıklama oranları sırası ile % 64.02 ve %15.57 olarak bulunmuştur.

Sonuç: Sonuç olarak, yöntemlerin birbirine benzer olduğu; Redundancy analizinin, Temel bileşenler analizinin kanonik versiyonu ve constrained formu olduğu, Kanonik uyum analizinin ise uyum analizinin kanonik versiyonu olduğu belirtilerek, üç yöntemin de temelde ağırlıklandırılmış tekil değer ayrışımı olarak, matrislerin satır ve sütunları için yeni koordinatların bulunmasını amaçladığı vurgulanmıştır.

Anahtar Sözcükler: Kanonik değişken, uyum, redundancy, constrained

KONFERANS - II

Biyoistatistik Literatürünün Gelişimi ve Trend Konular; Bir Bibliyometrik Analiz

Doç.Dr. Serdal Kenan KÖSE

Ankara Üniversitesi, Tıp Fakültesi, Biyoistatistik Anabilim Dalı, Ankara

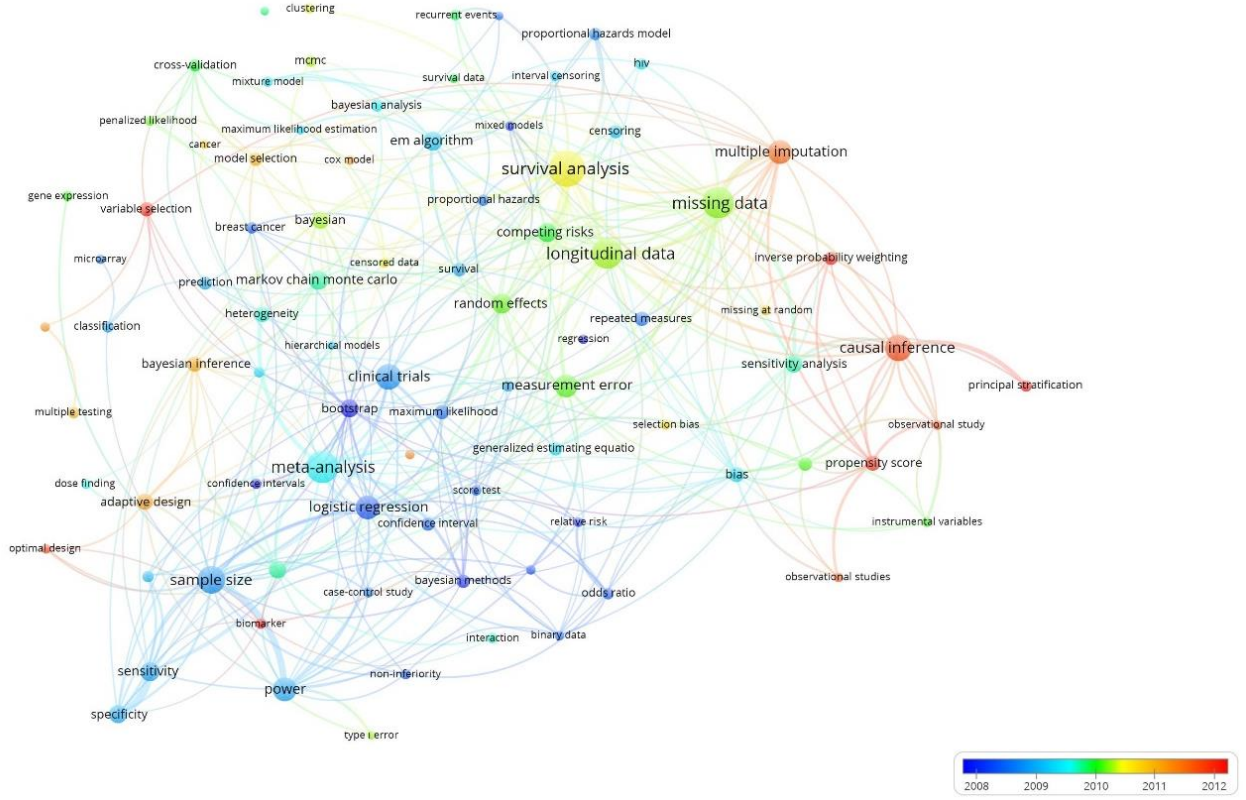
Özet

Amaç: Her geçen gün artan Biyoistatistik araştırmaları üzerine yayınlanan yayınlarda birçok yeni yöntem önerilmektedir. Araştırmacıların birçok dergide yayınlanan yeni yöntemleri takip etmeleri zorlaşmaktadır. Bu çalışmada Dünyadaki Biyoistatistik üzerine yapılan çalışmaların geniş bir literatür taramasını yapmak, trend konuları belirlemek, Ülkeler arası yayın ve atıf ilişkilerini ortaya koymak ve Ülkemizdeki Biyoistatistik araştırmacılarına yön gösterecek ve literatürü özetleyecek bilgiler sunmak amaçlanmıştır.

Yöntem: Bibliyometri, akademik literatür içerisindeki makale veya kitap gibi yazılı yayınların istatistiksel analizleridir. Birçok araştırma alanı, kendi alanlarının etkisini, bir yazarın etkisini veya belirli bir bilimsel çalışmanın uluslararası etkisini keşfetmek için bibliyometrik yöntemler kullanır. Atıf analizi ise bilimsel dokümanlar, yazarlar veya makaleler arasındaki ilişkileri gösteren bir bibliyometrik yöntemdir. Araştırmamızda 1984-2016 yıllarında etki faktörü yüksek Biyoistatistik dergilerinde yayımlanan tüm yayınlar Thomson Reuters Web of Science (WoS; Thomson Reuters, New York, NY, USA) veri tabanından indirilerek analiz edilmiştir. WoS veritabanında arama yapmak için Anahtar Sözcükler olarak belirlenen dergi isimleri ("statistics in medicine", "international journal of biostatistics" vb.) kullanılmıştır. 1984 ve 2016 yılları arasında yayınlanan belgeler çalışmaya dâhil edilirken, 2017 yılında yayınlananlar hariç tutulmuştur. Bibliyometrik Ağ Görselleştirmeleri VOSviewer (Versiyon 1.6.5) kullanılarak oluşturulmuştur. Bibliyometrik analizler kullanılarak biyoistatistik yayınlarında aktif olan yazarlar, ülkeler, bilimsel dergiler ve üniversiteler belirlenmiş, ortak yazarlık, atıf, ortak-atıf analizleri ve en çok kullanılan anahtar kelimler arasındaki ilişkiler araştırılmıştır.

Bulgular: Biyoistatistik Literatürüne en çok katkı yapan ilk üç yazarın Zhou XH (65; %0,64), White IR (61; %0,60) ve Senn S (58; %0,57), en çok katkı sağlayan ilk üç ülkenin USA (5766; %57,06), İngiltere (1505; %14,89) ve Kanada (631; %6,24), en çok katkı sağlayan üniversitenin Harvard Üniversitesi (658; %6,51) olduğu belirlenmiştir. En çok yayın 2016 yılında (683; %6,76) yapılmıştır. Toplam 9351 ve yıllık ortalama 584 atıf ile en çok atıf alan çalışmanın "statistics in medicine" dergisinde Higgins, JPT ve Thompson, SG (2002) tarafından yayınlanan "Quantifying heterogeneity in a meta-analysis" başlıklı makale olduğu belirlenmiştir. En popüler Anahtar Sözcükler Şekil 1'de gösterilmiştir. Grafiğin sağ alt kısmındaki renkler maviden kırmızıya doğru son yıllarda daha yaygın kullanılan Anahtar Sözcükleri göstermektedir.

XIX. Ulusal ve II. Uluslararası Biyoistatistik Kongresi
XIX. National and II. International Biostatistics Congress
25-28 Ekim 2017 Belek / Antalya



Şekil 1. Yıllara göre Biyoistatistikte kullanılan popüler Anahtar Sözcükler

Sonuç: Bu çalışmada 1984 ve 2016 yılları arasında biyoistatistik araştırma alanında WoS veri tabanında indekslenen yayınlar üzerine bibliyometrik analizler yapılarak, Ülkemizdeki Biyoistatistik araştırmacılarına yön gösterecek önemli bilgiler ortaya konulmuştur.

Anahtar Sözcükler: Biyoistatistik, bibliyometrik analiz, trend konular, ağ görselleştirme, atıf analizi

SÖZLÜ SUNUMLAR
ORAL PRESENTATIONS

Ortak Kontrol Grubunun Bulunduğu Faktöriyel Denemelerin İstatistiksel Analizleri: Monte Carlo Simulasyon Çalışması

Mehmet MENDEŞ, **Soner YİĞİT**

*Çanakkale Onsekiz Mart Üniversitesi Ziraat Fakültesi, Zootekni Bölümü, Biyometri ve Genetik ABD-
Çanakkale*

Amaç: Araştırmacılar denemelerini kurarken birçok durumda ‘uygulanan muamelelerin (deneme gruplarının) ne kadar etkili olup olmadıklarını ya da mevcut duruma göre üzerinde durulan özelliği önemli düzeylerde etkileyip etkilemediklerini belirleyebilmek amacıyla’ denemelerine bir kontrol grubu da koyarlar. Çünkü böylece etkisi araştırılan muamelelerin mevcut duruma göre etkilerini daha güvenilir bir şekilde araştırma ya da ortaya koyma imkânına kavuşabilirler. Kontrol grubunun bulunduğu denemelerden (özellikle faktöriyel denemelerde) elde edilen verilerin istatistik analizlerinde dikkatli olması gerekir. Özellikle uygulamada kurulan denemelerin oldukça önemli bir kısmının faktöriyel düzende olduğu gerçeği dikkate alındığında, bu tür denemelerde bir tek kontrol grubunun (ortak kontrol grubu) bulunduğu durumlarda elde edilen verilerin istatistiksel analizlerinde bilinen klasik faktöriyel denemelere ilişkin hesaplamaların uygun olmayacağı ortadadır. Ancak, uygulamada birçok durumda araştırmacıların bu tür denemelerden elde ettikleri veri setlerinin istatistiksel analizlerinde hala klasik faktöriyel denemelerdeki hesaplamaları yaptıkları dikkati çekmektedir. Doğal olarak böyle bir yola başvurulması sonucunda da elde edilecek sonuçlar güvenilir olmayacaktır. Çünkü böyle bir çözüm yoluna gidilmesi, deneme başında kararlaştırılan 1.Tip hata olasılığının deneme başında kararlaştırılan düzeyde korunamamasına ve testin gücünde olumsuz değişmelerin meydana gelmesine sebep olacaktır. Bunun sonucunda da ‘gerçekte istatistiksel olarak önemli olan farklılıklar önemli bulunamayacak ya da istatistiksel olarak önemli olmayan farklılıklar önemli’ bulunabilecektir. Böyle durumlarda; Winer ve ark. (1991) farklı bir çözüm yolu önermişlerdir. Bu noktadan hareketle yürütülen bu simülasyon çalışmasında faktöriyel denemelerden elde edilen veri setlerinin hem bilinen klasik faktöriyel denemelerin analizi gibi hem de Winer ve ark. (1991) tarafından önerilen çözüm yolu ile istatistiksel analizleri yapılarak bu iki yaklaşımın birçok deneme koşulundaki performansları ortaya konulmuştur. Söz konusu yaklaşımların birçok deneme koşulundaki performanslarının ortaya konulmuş olması ise; elde edilen sonuçların genelleştirilebilmesi bakımından oldukça önemlidir.

Yöntem: Söz konusu yöntemlerin performanslarının karşılaştırılması amacıyla Monte Carlo Simulasyon Tekniğinden yararlanılarak farklı deneme koşullarında üretilen tesadüf sayıları üretilmiştir.

Bulgular: Yapılan Monte Carlo Simulasyon çalışması sonucunda; dikkate alınan deneme koşulları ne olursa olsun Winer ve ark. (1991) tarafından önerilen çözüm yolunun oldukça güvenilir sonuçlar verdiği görülmüştür.

Sonuç: Ortak bir kontrol grubunun bulunduğu faktöriyel denemelerde dikkate alınan deneme koşulları ne olursa olsun istatistiksel analizlerin Winer ve ark. (1991) tarafından önerilen çözüm yoluna göre yapılması gerektiği bildirilmiştir.

Anahtar Sözcükler: Kontrol grubu, faktöriyel düzen, 1.tip hata, Testin gücü

Bootstrap Güven Aralığı Yöntemlerinin Karşılaştırılması

Merve BAŞOL¹, Dinçer GÖKSÜLÜK¹, A. Ergun KARAAĞAOĞLU¹

¹ Hacettepe Üniversitesi, Tıp Fakültesi, Biyoistatistik, 06100, Ankara, TÜRKİYE

merve.basol@hacettepe.edu.tr, dincer.goksuluk@hacettepe.edu.tr, ekaraaga@hacettepe.edu.tr

Amaç: Bootstrap yöntemi, parametre tahminlerinin yanlışlığının ölçülmesi, standart hata kestirimi ve benzeri amaçlar ile kullanılan bilgisayar tabanlı bir yeniden örnekleme yöntemidir. Ortalama gibi bazı istatistiklerin standart hata değeri bir formül yardımı ile kolaylıkla tahmin edilebilirken daha karmaşık istatistikler için bu durum olası olmayabilir [1]. Bootstrap yöntemi, bu karmaşık istatistiklerin standart hatasını tahmin etmede günümüzde de sıklıkla kullanılan bir yöntemdir. Bu yöntem, aynı zamanda güven aralığı ve yanlışlık miktarını tahmin etmede de sıklıkla tercih edilmektedir. Bu çalışmada bootstrap örnekleme yöntemi ile güven aralığı hesabı üzerinde durulacaktır. Bootstrap yöntemi ile güven aralığı tahmin etmek için geliştirilmiş birçok yöntem vardır. Bu yöntemler birbirlerinin eksikliklerinden yola çıkılarak geliştirilmiş yöntemlerdir. Bu çalışma kapsamında bootstrap güven aralığı yöntemlerinden standart normal bootstrap, yüzdelik (percentile), BCa (bias corrected and acceleration), Bootstrap - t ve ABC (approximation bias corrected) yöntemleri [1-5] kullanılacaktır. Bu çalışmanın amacı, bootstrap güven aralığı yöntemlerini tanıtmak, yöntemlerin hangi eksikliklerden yola çıkılarak geliştirildiğini, olumlu ve olumsuz yanlarını göstermek ve yöntemleri karşılaştırmaktır.

Yöntem: Bootstrap güven aralığı yöntemlerinin performansını değerlendirmek ve birbirleri ile karşılaştırmak için bir benzetim çalışması yapılmıştır. Benzetim çalışması ile yöntemler farklı dağılımlarda (normal dağılım, ki-kare dağılımı vb.), farklı istatistiklerde (ortalama, korelasyon vb.) ve farklı örneklem büyüklüklerinde (20,100) incelenmiştir. Ayrıca, yöntemler harcadıkları süre bakımından da incelenmiştir. Elde edilen sonuçlara göre yöntemlerin kapsama yüzdeleri ve güven aralıklarının genişlikleri hesaplanmıştır. Benzetim çalışmasına ek olarak yöntemlerin sonuçları gerçek bir veri seti kullanılarak da incelenmiştir. Analizler, R 3.3.2 programında “bootstrap” ve “boot” paketleri kullanılarak yapılmıştır.

Bulgular: Benzetim çalışması ile ortalama, varyans, korelasyon gibi istatistiklerin güven aralıkları her bir yöntem ile tahmin edilmiştir. Elde edilen sonuçlara bakıldığında yöntemlerin kullanılan istatistik türüne, dağılım türüne, örneklem büyüklüğüne göre farklılık gösterdiği görülmüştür. Dağılım çarpık olduğu durumda, standart ve yüzdelik güven aralığı yöntemlerinin kapsama yüzdeleri diğer yöntemlere göre daha düşüktür. En yüksek kapsama yüzdesi ise bootstrap-t yönteminde elde edilmiştir. Ancak, güven aralığı genişlikleri incelendiğinde bootstrap-t yönteminin oldukça geniş aralıklar verdiği görülmüştür. Dağılım normal olduğu durumda ise, yöntemler hem kapsama yüzdesi hem de güven aralığı genişliği açısından benzer sonuçlar vermiştir. Ayrıca, örneklem büyüklüğünün artması da yöntemlerin benzer sonuçlar vermesini sağlamıştır. Bu sonuçlar, literatür taramasında elde edilen bilgiye de uymaktadır. Hesaplama süreleri incelendiğinde ise bootstrap -t yöntemi diğer yöntemlere göre daha uzun sürede analizleri bitirmektedir. Bunun nedeni, bu yöntemin iç içe üç bootstrap kullanmasından kaynaklanmaktadır.

Sonuç ve değerlendirme: Bootstrap güven aralığı yöntemleri birbirlerinin eksikliklerinden yola çıkılarak geliştirilmiş yöntemlerdir. Bu nedenle, aslında birbirini güncelleyen yöntemlerdir. Yine

de farklı özelliklere ve varsayımlara sahiptirler. Bu nedenle, bir çalışmada hangi bootstrap güven aralığının kullanılması gerektiği, hangi varsayımların sağlanıp sağlanmadığı iyi belirlenmelidir. Uygun yöntemle karar vermede yalnızca kapsama yüzdelerinin dikkate alınması yanlış sonuçlar elde edilmesine neden olabilir. Geniş güven aralıklarının kapsama yüzdelerini yanıltıcı bir şekilde artıracak görülmektedir. Bu nedenle, kapsama yüzdeleri ile birlikte güven aralıklarının genişliklerinin de dikkate alınması önerilmektedir.

Anahtar Sözcükler: bootstrap, güven aralığı, BCa, ABC

KAYNAKLAR

1. Efron, B., Tibshirami, R.J. (1993). An Introduction to the Bootstrap. *Springer –Science*.
2. Chernick, M.R., LaBudde, R.A. (2011). An Introduction to Bootstrap Methods with Applications to R. *A John Wiley & Sons, inc, publication*.
3. Davison, A.C., Hinkley, D.V. (1997). Bootstrap Methods and their Application. *Cambridge University Press*.
4. Chernick, M.R. (2008). Bootstrap Methods A Guide for Practitioners and Researchers. *A John Wiley & Sons, inc, publication, second edition*.
5. Carpenter, J., Bithell, J. (2000). Bootstrap Confidence Intervals: When, Which, What? A Practical Guide For Medical Statisticians. *Statistics in Medicine*. 19:1141 – 1164.

Tedavi-Biyomarker Etkileşiminin BATE Eğrisi ile İncelenmesi

Büşra EMİR¹, Ertuğrul ÇOLAK¹

¹ Eskişehir Osmangazi Üniversitesi, Tıp Fakültesi, Temel Tıp Bilimleri Bölümü, Biyoistatistik AD, Eskişehir

Amaç: Klinik araştırmalarda, farklı biyomarker değerlerine sahip hastalar, yeni tedavilere veya ilaçlara farklı tepkiler verebilir. Kişiselleştirilmiş tıpta, tedaviden fayda görebilecek hastaları açık bir şekilde tanımlamada tedavi ile biyomarker değerleri arasındaki etkileşimi incelemek önemlidir. Bu çalışmanın amacı, biyomarker değerinin bir fonksiyonu olarak tedavi etkisinin grafiksel gösterimi olan Biyomarker'a göre düzeltilmiş Tedavi Etkisi (Biomarker Adjusted Treatment Effect; BATE) eğrisinin kullanımını örnek bir veri seti kullanarak tanıtmaktır.

Yöntem: BATE eğrisi, yaşam sürelerini içeren bir veri setinde biyomarker değerinin klinik faydasını değerlendirmede ve belirli bir hasta için optimal tedaviyi seçmede kullanılması önerilen yeni bir yöntemdir. Biyomarker'a göre bir değer aralığı belirleyerek aynı amaca yönelik kullanılan farklı tedavi yöntemlerinin seçiminde karar vermeyi sağlar. Belirlenen bu değer aralığına göre hangi tedavi yönteminin daha etkili olabileceğini önerir. Biyomarker'a göre değer aralığının belirlenebilmesi için tedavi-biyomarker etkileşiminin önemli olması gerekmektedir. Bu çalışmada etkileşimin önemliliği, Fan J, Lin H, Zhou Y. tarafından önerilen yerel kısmi olabilirlik tahmin (local partial likelihood estimation) yöntemi kullanılarak belirlendi. Bu yöntemde gaussian kernel düzeltilmesi kullanıldı. Çalışmada kullanılan uygulama veri seti, 1974 ve 1984 yılları arasında Mayo Clinic tarafından yürütülen klinik bir çalışmadan elde edilmiştir. Karaciğer primer biliyer siroz (PBC) hastalığı tanısı almış 424 hastadan 312'si randomize olarak 2 tedavi grubuna ayrılmış. Birinci gruba D-penisilamin ilaç tedavisi uygulanmış diğer gruba ise plasebo verilmiştir. Bu hastalara ait değişkenlerden yaşam süreleri, sansür durumları, tedavi grupları, aspartat aminotransferaz (AST) ve alkalın fosfataz (ALP) değerleri BATE eğrisi ile incelendi. Bu çalışmada, ALP değeri PBC hastalığının belirlenmesinde biyomarker olarak kullanıldı. AST, ortak değişken olarak modele dahil edildi.

Bulgular: Yerel kısmi olabilirlik tahmin yöntemi kullanılarak tedavi grupları ile biyomarker olarak belirlenen ALP değerleri arasında önemli bir etkileşim belirlendi ($Q_1=34.27558$, $p=0.002$). Sonraki aşamada oluşturulan BATE eğrisine göre ALP için belirlenen değer aralığı yaklaşık olarak 900-2100 (Ü/litre) hesaplandı.

Sonuçlar: ALP değeri 900-2100 (Ü/litre) arasında olan PBC hastaları için D-penisilamin ilaç tedavisi uygulanması önerilebilir. ALP'si bu aralık dışında olan hastalar D-penisilamin ilaç tedavisine duyarlı olmadığından farklı tedavi yöntemlerinin uygulanması gerekebilir. BATE eğrisi, tedavi etkinliğine göre biyomarker aralığı belirlemesi sayesinde kişiselleştirilmiş tıp alanında kullanılabilecek önemli bir yöntemdir.

Anahtar Sözcükler: BATE eğrisi, Tedavi-Biyomarker Etkileşimi, Yerel kısmi olabilirlik tahmin yöntemi

GAMLSS Modellerinin Model Yeterliliğinin Belirlenmesinde Worm Eğrileri

Meryem ÇAVUŞOĞLU^{1,2}, Gökmen ZARARSIZ^{1,2}, Gözde ERTÜRK ZARARSIZ¹, Selçuk KORKMAZ^{2,3},
Dinçer GÖKSÜLÜK^{2,4}, Mümtaz MAZICIOĞLU⁵, **Ahmet ÖZTÜRK**^{1,2}

¹Erciyes Üniversitesi, Tıp Fakültesi, Biyoistatistik Anabilim Dalı, Kayseri

²Turcosa Analitik Çözümler Ltd. Şti., Erciyes Teknopark, Kayseri

³Trakya Üniversitesi, Tıp Fakültesi, Biyoistatistik Anabilim Dalı, Edirne

⁴Hacettepe Üniversitesi, Tıp Fakültesi, Biyoistatistik Anabilim Dalı, Ankara

⁵Erciyes Üniversitesi, Tıp Fakültesi, Aile Hekimliği Anabilim Dalı, Kayseri
meryemhokelek@hotmail.com

Özet

Büyüme eğrileri, boy, kilo gibi antropometrik ölçümlerin yaşa bağlı değişimini izlemede sıklıkla kullanılan toplum standartlarıdır. Bu eğriler ile toplumların genel sağlık durumları hakkında bilgi elde edilebilir, bireylerin farklı zaman noktalarındaki ölçümleri incelenerek büyümenin sağlıklı olarak gerçekleşip gerçekleşmediği izlenebilir, büyüme ile ilişkili sağlık problemleri açısından alt ve üst risk limitleri belirlenebilir ve sağlıklı toplum verileri ile kıyaslama yapılarak olağan dışı ölçümlere sahip bireyler tespit edilebilir. Büyüme eğrilerini oluşturma da çok sayıda istatistiksel yöntem geliştirilmiştir. GAMLSS modelleri bu yöntemlerin birçoğunu kapsayan, çok sayıda olasılık dağılımını ve düzleştirme yöntemlerini içeren bir regresyon yaklaşımıdır. Örneğin, büyüme eğrileri oluştururken LMS yöntemi Box-Cox normal dağılımını, LMSP yöntemi ise Box-Cox power üstel dağılımlarından faydalanmaktadır. Her iki yöntemde de düzleştirme yöntemi olarak kübik eğriler tercih edilmektedir. Bu yöntemler ile büyüme eğrileri oluştururken önemli bir problem referans dağılımın yaşa göre uygun biçimde düzgünleştirilmesidir. Yöntemlerin içerisinde yer alan farklı dağılım parametrelerinin kullanımı, sonuçların farklılaşmasına, farklı toplum referansları elde edilmesine yol açmaktadır. Bu nedenle oluşturulan modellerin model yeterliliklerinin incelenmesi ve elde edilen sonuçlar üzerinde dikkatli bir denetim yapılması gerekmektedir. Worm eğrileri bu amaçla geliştirilmiş basit ve esnek bir grafiksel yaklaşımdır. Bu eğriler, referans dağılımın gözlenen ve beklenen farklılıklarını yaşa bağlı olarak görselleştirmektedir. Bu sayede kullanıcılar, grafikteki eğrilerin biçimine göre oluşturdukları modellerin yeterli olup olmadığına karar verebilmekte, yeterli değilse ne gibi değişiklikler yapmaları gerektiğine ilişkin yorumlar elde edebilmektedir. Bu çalışmada, worm eğrilerinin gerçek bir antropometrik büyüme eğrisi verisinde kullanımı gösterilecektir. Ayrıca, kullanıcılara bu eğrileri basit olarak oluşturabilmelerine yönelik kullanıcı-dostu bir web uygulaması sunulacaktır.

Anahtar Sözcükler: Büyüme eğrileri, GAMLSS, LMS, LMSP, worm eğrileri

Veri Füzyon Yöntemleri ile Çoklu Sensör Görüntülerinden Bilgi Çıkarımı

Adem DOĞANER¹, Cemil ÇOLAK²

¹Kahramanmaraş Sütçü İmam Üniv., Tıp Fakültesi, Biyoistatistik ve Tıbbi Bilişim AD, K.Maraş

²İnönü Üniversitesi, Tıp Fakültesi, Biyoistatistik ve Tıp Bilişimi AD, Malatya

Amaç: Verilerden bilgi çıkarımı gerçekleştirebilecek çok sayıda model ve yöntem geliştirilmiştir. Niceliksel veri yığınlarından makine öğrenme algoritmaları veya farklı istatistiksel teknikler ile bilgi çıkarımı gerçekleştirmek mümkündür. Niceliksel veri yığınlarında başarılı sonuçlar üretebilen bu yöntemler, görüntü gibi veri kaynaklarında başarılı sonuçlar üretememektedir. Mevcut yöntemlerin görüntülerden bilgi çıkarımında kısıtlı çözüm üretmesi ve insan gözünün sadece belirli bir dalga boyu aralığındaki görüntüleri görebilmesi görüntülerden bilgi çıkarımını zorlaştırmaktadır. Bu çalışmada 3 farklı dalga boyutunda çalışan kameralardan elde edilen görüntülerin veri füzyon yöntemleri ile birleştirilerek çoklu gözlem dizisine sahip Saklı Markov Modeli ile bilgi çıkarımı yapılması amaçlanmaktadır.

Yöntem: Çalışmanın uygulama aşamasında, görüntülerden bilgi çıkarımını gerçekleştirecek modelin geliştirilmesi için farklı dalga boylarında görüntüler alabilen sensörlere sahip kameralardan yararlanarak görüntüler alınmıştır. İnsan gözü sadece görünür ışık dalga boyu aralığındaki görüntüleri görebilmektedir. Özel sensörlere sahip kameralar ışığın farklı dalga boylarında insan gözünün doğrudan göremediği görüntüler elde edilebilmektedir. Uygulamada kızılötesi, ultraviyole ve görünür ışık dalga boyunda görüntüler alabilen kameralar ile t zamanında aynı açıdan ve aynı cilt dokusundan görüntüler alınmıştır. Bu görüntüler veri füzyon yöntemleri ile birleştirilmiştir. Saklı Markov Modeli (SMM) durumlar arası geçişlerin gözlemler dizisi aracılığı ile tahmin edildiği iki stokastik süreçten oluşan modellerdir. Uygulamada ise 3 farklı kameradan elde edilen gözlem dizileri ve bu gözlemlerden tahmin edilmek istenen bir durum dizisi bulunduğu için klasik SMM ile modellemek yerine Saklı Markov Modeli tabanlı yeni bir model üzerine çalışılmıştır. Modelin uygulanması sonucu elde edilen görüntülere göre cildin risk değerlendirmesi gerçekleştirilmiştir.

Bulgular: 3 farklı kameradan elde edilen görüntüler veri füzyon yöntemleri ile birleştirilmiştir. Birleştirme sonucunda görüntülerde ve cilt dokuları üzerinde farklı renk tonları ortaya çıkmıştır. Görüntülerde piksel kümeleri elde edilmiştir. Bu görüntüler ile cilde ilişkin cilt kalınlığı, ısı geçirgenliği, melanin yoğunluğu, pigmentasyon problemleri ve güneşe hassasiyet konusunda bilgi çıkarımı gerçekleştirilmiştir.

Sonuç: Farklı kameralardan elde edilen görüntülerin veri füzyonu ile birleştirilmesi ve yeni geliştirilen SMM tabanlı model başarılı bir şekilde uygulanmıştır. Cilt üzerinde risk teşkil eden bölgelerde farklı renkte ve yoğunlukta piksel kümelerinin olduğu gözlemlenmiştir. Görüntüdeki bu kümelerle ilgili bilgi çıkarımı gerçekleştirilmiştir.

Anahtar Sözcükler: Markov Modelleri, Örüntü Birleştirme, Bilgi Çıkarımı

Kohonen Öz Örgütlemeli Haritalama Yöntemi İle Psikotik Hastalıkların Kümelenmesi

Aslı BOZER, Osman DAĞ, Sevilay KARAHAN

Hacettepe Üniversitesi Tıp Fakültesi Biyoistatistik Anabilim Dalı

Özet: Veri madenciliği; büyük miktarda veri içerisinde, gizli kalmış, değerli, kullanılabilir bilgileri açığa çıkaran ve stratejik karar destek sağlamak amacıyla kullanılan; verilerin analizini temel alarak karar verme modelleri yaratan bir yöntemdir (1). Veri madenciliği, öngörülen modelleri uygulayarak gelecekteki olaylara daha iyi kararlar verilmesini, eylemde bulunulmasını sağlamaktadır (2). Veri madenciliği, veri tabanında beklenmeyen ilişkilerin bulunmasına yardım etmesi açısından geleneksel istatistiksel analizlerden farklılık göstermektedir. Yüksek boyutluluk ve büyük hacimli verilerden ötürü, veri madenciliğinde geleneksel istatistiksel yöntemlerin kullanımı sınırlıdır. Dolayısıyla veri madenciliğinde, yapay zeka ve veri görselleştirme alanındaki çalışmalar gerekli hale gelmektedir. Yapay sinir ağları; insan beyninin özelliklerinden olan öğrenme yolu ile yeni bilgiler türetebilme, yeni bilgiler oluşturabilme ve keşfedebilme gibi yetenekleri gerçekleştiren bilgisayar sistemleridir. Yapay sinir ağları arasında; Hebb Net, Perceptron, Kohonen Map (Kohonen Öz Örgütlemeli Haritalama Yöntemi) Backpropagation Net, Hopfield Yapay Sinir Modeli yer almaktadır (1). Bu çalışmada Tuevo Kohonen tarafından geliştirilmiş Kohonen Öz Örgütlemeli Haritalama yöntemi kullanılacaktır. Kohonen Öz Örgütlemeli Haritalama Yöntemi; n-boyutlu uzaydaki girdi verilerini, orjinal topolojik ilişkilerini koruyarak daha düşük boyutlu çizimle (genellikle 2 boyutlu) haritalamaktadır. Verideki benzer örüntüleri ortaya çıkaran bir yöntemdir. Bu yöntem, denetimsiz öğrenme algoritmalarını kullanan, yani öğrenmeyi kendi kendine gerçekleştiren, ileri beslemeli bir yapay sinir ağıdır (3, 4).

Çalışmanın amacı; veri madenciliğinde, kümeleme amacı ile kullanılan Kohonen Öz Örgütlemeli Haritalama yöntemini kuramsal olarak tanıtmak, bu yöntemin özelliklerini incelenmek, sağlık alanında sınıflama gereksinimi olduğu durumlarda bu yöntemin kullanılabilirliğini göstermek, bu yöntem sonucunda elde edilen sınıflamanın performansının (doğruluk oranının vb.) diğer klasik sınıflama yöntemlerine göre daha yüksek olacağını göstermektir.

Yöntem: Çalışmada Kohonen öz örgütlemeli haritalama yöntemi kullanılacaktır. Bu çalışma; psikiyatrik hastalıklardan biri olan şizofreni hastalarının verileri kullanılarak uygulanacaktır. Şizofreni, hezeyanlar ve hallüsinasyonlar olarak bilinen psikotik semptomlardan, fonksiyonel bozulmalara kadar geniş bir semptom spektrumunda, değişik şekillerde görülen, kronik ve sıklıkla relapslarla seyreden bir akıl hastalığıdır. Şizofreni tüm psikiyatrik rahatsızlıklar içinde, tanımlanması, betimlenmesi en zor olanlardan biridir (5). Şizofreninin tanınması ile ilgili olarak çeşitli sınıflandırma sistemleri geliştirilmiştir. Bu sınıflandırma sistemleri ile şizofreni tanınması için çeşitli alt tipler oluşturulmuştur. Alt tiplerin sınıflandırılmasında kullanılan belirtilerin ve semptomların kohonen map yöntemi kümelenmesiyle, sınıflandırmanın ve teşhisin istatistiksel olarak analiz edilmesi hedeflenmektedir. Çalışmanın verileri toplanmış olup, analiz sonrasında Kohonen map öz örgütlemeli haritalama yöntemi ile klasik kümeleme yöntemlerinin sonuçları karşılaştırılacaktır.

Anahtar Sözcükler: Veri Madenciliği, Kohonen Öz Örgütlemeli Haritalama Yöntemi, Şizofreni

Kaynaklar

- 1-Kohonen T, Kaski S, Somervuo P, Lagus K, Oja M. Paatero. (2003).Self-organizing Map.İki Yılda Bir Değerlendirme Raporu. <http://cis.legacy.ics.tkk.fi/research/reports/biennial02-03/cis-biennial-report-20022003-8.pdf>
- 2-Nisbet, R., Elder, J., Miner, G. (2009). Handbook of Statistical Analysis and Data Mining Applications. İngiltere:Academic Press.
- 3-Kohonen, T. Self-Organizing maps. 3th ed. Newyork: Springer; 2001.
- 4-Valova I, Georgiev G, Gueorguievac N, Olsona J. Initialization issues in self-organizing maps. Procedia Computer Science. 2013; 20: 52-57.
- 5-Ozturk MO:(1997)Ruh Sağlığı ve Bozuklukları.7. Basım. Hek. Birl. Yay. Ankara.

Yaşam Çözümlemesinde Eğilim Skoru

Durdu KARASOY¹, Hatice IŞIK¹, Nihal ATA TUTKUN¹

¹Hacettepe Üniversitesi Fen Fakültesi İstatistik Bölümü, Ankara.

Özet

Amaç: Eğilim skoru yöntemleri gözlemsel çalışmaların çözümlemesi için popüler bir araçtır. Bu yöntemlerin kullanımı, iki tedavi grubu arasındaki ölçülen temel karakteristiklerdeki farklılıklar nedeniyle meydana gelen bozulmaların etkisini azaltmayı sağlar. Eğilim skoru yöntemleri, gözlemsel çalışmalarda tedavi etkilerini tahmin etmek için kullanılır. Tıbbi ve epidemiyolojik çalışmalarda çıktılar, çoğunlukla olay süresi olmaktadır. Eğilim skoru yöntemleri ile ortalama tedavi etkileri için, tedavi üzerine ortalama tedavi etkileri için ve potansiyel çıktı ortalamaları için değişik tahmin ediciler elde edilir. Tedavi etkileri, regresyon düzeltmesi, ters olasılık ağırlıklılıkları, ters olasılık ağırlıklandırılmış regresyon düzeltmesi ve ağırlıklandırılmış regresyon düzeltmesi kullanılarak tahmin edilebilir. Bu çalışmada, mide kanseri için kemoterapi etkisi bu yöntemler kullanılarak incelenmiştir ve yorumlanmıştır.

Yöntem: Bu çalışmada, yaşam çözümlemesinde tedavi etkilerini tahmin etmede kullanılan regresyon düzeltmesi, ters olasılık ağırlıklılıkları, ters olasılık ağırlıklandırılmış regresyon düzeltmesi ve ağırlıklandırılmış regresyon düzeltmesi yöntemleri kullanılmıştır.

Bulgular: Kullanılan yöntemler sonucunda, tüm hastalar kemoterapi aldığı ortalama yaşam süresi, kemoterapi almayan hastalardan yaklaşık 30 ay daha uzun olarak tahmin edilmiştir. Hastalar kemoterapi almadığında tahmin edilen ortalama yaşam süresi ise yaklaşık 35 ay olarak bulunmuştur. Tüm hastalar kemoterapi aldığı ortalama yaşam süresi, hastaların hiç biri kemoterapi almadığındaki duruma göre yaklaşık %83 artmaktadır.

Sonuç: Eğilim skorları, grup farklılıklarını düzeltmek için değerli fakat kullanılmayan bir yaklaşımdır. İki grubun karşılaştırılabilir hale getirilmesi için kullanılır. Bu yöntemle, tedavi etkilerini ortaya çıkarmak ve bireysel faktörleri kontrol etmek mümkündür. Bu çalışmada, çıktısı olay süresi olan yaşam çözümlemesinde eğilim skoru yöntemlerinin kullanılması ele alınmış ve gerçek bir veri seti olan mide kanseri verileri için kemoterapinin etkisi incelenmiş ve elde edilen sonuçlar yorumlanmıştır.

Anahtar Sözcükler: Yaşam çözümlemesi, Eğilim skoru, tedavi etkisi

Tekrarlı Gözlem ve Sağkalım Verilerinin Birleşik Model Performans Değerlendirmesi: Bir Simülasyon Çalışması

Ebru TURGAL¹, S. Kenan KÖSE²

¹ Hitit Üniversitesi, Tıp Fakültesi, Biyoistatistik Anabilim Dalı, Çorum.

² Ankara Üniversitesi, Tıp Fakültesi, Biyoistatistik Anabilim Dalı, Ankara.

ebruturgal@hitit.edu.tr

ÖZET

Amaç: Birleşik modelleme sağkalım verileri ile farklı zaman noktalarında ölçülmüş tekrarlı gözlemleri analiz eden bir teknik olarak yaygınlaşmaya başlamıştır. Ortak rastgele etkiler ile boylamsal süreç ve sağkalım sürecinin birbirine bağlandığı birleşik modelleme, bu iki alt model için eş zamanlı kestirim yapma olanağı sağlar. Ancak çoğu zaman birleşik modellemenin matematiksel hesaplama zorluğu ön plana çıkmakta ve olabilirlik kestiriminde bazı paketler integral çözümlemede sıkça başarısız olmaktadır. Bu çalışmanın amacı, simülasyon tekniği ile zamana bağlı değişkenlerin çeşitli örneklem büyüklükleri ve varyans-kovaryans yapısı için performanslarının değerlendirilmesidir.

Yöntem: Çalışma R’da joiner, JM paketleri kullanılarak gerçekleştirilmiştir. JM paketine ait birleşik model için uyarlayıcı ve yalancı uyarlayıcı (pseudo-adaptive) Guass-Hermite kare yapma yöntemi kullanılmış, sonuçlar performans bakımından kıyaslanırken yanlışlık esas alınmıştır.

Bulgular: Ibrahim, J. G., Chu, H., Chen, L. M. (2010) Rastgele Kesim Noktası ve Rastgele Eğim ile tedavi etkisine sahip doğrusal karışık etkili modelden üretilmiş tekrarlı gözlem verisinin birleşik modellenmesine ait sağkalım üzerindeki zamana bağlı tahmin edici etkisinin daha küçük standart hataya sahip olduğunu simülasyon çalışmalarında göstermiştir. Paralel olarak, hazard oranındaki tedavi etkisi, oransal hazard modeli ile birleşik modelde birbirine çok yakın ve neredeyse yansız bulunmakla birlikte tahmine ilişkin standart hataların da hemen hemen aynı olduğunu söyleyebiliriz. Bununla beraber, sağkalım üzerindeki tedavi (grup) etkisi sıfır alınmadığında simülasyon süresi oldukça uzamakta ve elde edilen benzetim sonuçları performans açısından kabaca bu çalışma ile benzer bulunmuştur.

Sonuç: Sabit ve rasgele etkiler ile hata terimleri üzerine çeşitli senaryolar üretilerek simülasyon yapılmıştır. Örneklem sayısı (n) 150, 200 alındığında sağkalım üzerindeki zamana bağlı tahmin edici etkisi genişletilmiş Cox regresyon için simülasyon çalışmalarının çoğunda güven aralıklarının dışına çıkmıştır. JM, joiner paketine göre daha iyi sonuç göstermiştir. $h_0(t)$ temel risk fonksiyonlarından parçalı sabit ve yiv(spline) fonksiyonlar sağkalım üzerindeki zamana bağlı tahmin edici etkisini kapsama açısından birbirlerine yakın performans göstermişlerdir. Simülasyon çalışmaları devam etmektedir. Sunuma kadar olan süre zarfında çalışmalar tamamlanacaktır.

Anahtar Sözcükler: Birleşik Modelleme, Boylamsal Veri Analizi, Sağkalım Analizi, Simülasyon

Kaynaklar:

Cox D. (1972). Regression models and life-tables (with discussion). *J Roy Stat Soc B*; **34**: 187–220.

Ibrahim, J. G., Chu, H., Chen, L. M. (2010). Basic concepts and methods for joint models of longitudinal and survival data. *Journal of Clinical Oncology*, 28(16), 2796-2801.

Rizopoulos D. (2012). Joint Models for Longitudinal and Time-to-Event Data: with Applications in R. Boca Raton: Chapman and Hall/CRC.

Tsiatis A, Davidian M. (2004). An overview of joint modeling of longitudinal and time-to-event data, *Statistica Sinica*, **14**, 809–834.

Yuen, H. P., Mackinnon, A. (2016). Performance of joint modelling of time-to-event data with time-dependent predictors: an assessment based on transition to psychosis data. *PeerJ*, **4**, e2582.

Orantılı Hazard Durumunun İhlalinde Log-Rank Testinin Gücünün İncelenmesi

Hülya ÖZEN¹, Cengiz BAL¹

¹*Eskişehir Osmangazi Üniversitesi Tıp Fakültesi Biyoistatistik AD., 26480 Eskişehir*

Amaç: Yaşam analizinin temel amaçlarından biri farklı gruplara ait yaşam fonksiyonlarının homojen olup olmadığını değerlendirmektir. Bu amaçla literatürde kullanılan yöntemler genellikle log-rank ve ağırlıklandırılmış log-rank testleridir. Log-rank testi, orantılı hazard durumunda gücü oldukça yüksek karşılaştırma sonuçları vermektedir. Bu çalışmada orantılı hazard durumu ihlal edildiğinde log-rank ve ağırlıklandırılmış log-rank testlerinin gücünün araştırılması amaçlanmaktadır.

Yöntem: Monte Carlo benzetim yöntemi kullanılarak orantılı hazard durumu ve bu durumun ihlaline ait çeşitli senaryolar oluşturuldu. Orantılı hazard varsayımı ihlaline ait senaryolar kesişen iki yaşam fonksiyonunu karşılaştırma üzerine kuruldu. Karşılaştırmalar log-rank ve ağırlıklandırılmış log-rank (Wilcoxon, Tarone-Ware, Peto, Modified Peto, Fleming-Harrington) testleri kullanılarak gerçekleştirildi. Yaşam sürelerinin türetilmesinde farklı parametrelili üstel dağılımlardan ve sağdan sansürlü mekanizma oluşturmak için ise tekdüze dağılımdan faydalandı. Benzetim çalışmalarında tekrar sayısı 1000 kabul edildi. Grupların örnek hacimleri birbirine eşit ve 20, 50 ve 100 olarak belirlendi. Sansür oranları ise her iki grupta birbirine eşit ve %0, %10, %20 ve %30 olarak seçildi. Anlamlılık düzeyi ise $\alpha=0.05$ kabul edildi. Testlerin gücü; anlamlı fark bulunan benzetim sonuçlarının tekrar sayısına oranı ile elde edildi.

Bulgular: Orantılı hazard yaklaşımı altında tüm testlerin gücü örnek hacmi arttıkça %100'e yakınsamaktadır. Özellikle örnek hacmi 50 ve 100 olduğunda, güç değerleri sansür oranlarındaki artıştan oldukça az etkilenmektedir.

Kesişen yaşam fonksiyonlarının yer aldığı orantılı hazard durumunun ihlalinde testlerin gücü oldukça düşmektedir. Ağırlıklandırılmış log-rank testlerinden sadece Fleming-Harrington'ın diğer testlere göre daha güçlü sonuçlar vermektedir. Sansür oranındaki artış tüm testlerin gücünü olumsuz yönde etkilerken, örnek hacmindeki artış gücün artmasını sağlamaktadır.

Sonuç: Orantılı hazard durumunda güçlerinin yüksek olmasından dolayı log-rank testlerinin kullanılması uygundur. Ancak, orantılı hazard durumu ihlalinde özellikle küçük örnek hacmi ve yüksek sansür oranında log-rank ve ağırlıklandırılmış log-rank testlerinin gücü azalmaktadır. Bu nedenle bu testlere alternatif önerilen diğer testlerin kullanılması önerilmektedir.

Anahtar Sözcükler: kesişen yaşam fonksiyonları, log-rank, orantılı hazard durumu

Yaşam Süresi Verilerinin Çok Değişkenli Uyumlu Regresyon Analizinde Kullanımı

Merve TÜRKEGÜN¹, Gülhan OREKİCİ TEMEL¹, İrem ERSÖZ KAYA²

¹Mersin Üniversitesi Tıp Fakültesi, Biyoistatistik ve Tıbbi Bilişim Anabilim Dalı, Mersin

²Mersin Üniversitesi Tarsus Teknoloji Fakültesi, Yazılım Mühendisliği Bölümü, Mersin

Amaç: Bu çalışmanın amacı; Survival MARS tekniğinin teorik çerçevelerini anlatma ve yaygın kullanılan artık türlerinin modeldeki performansını değerlendirmektir.

Yöntem: Bir canlının yaşam süresinin tahmin edilmesi için geliştirilen yöntemler veri yapısının sansürlü olmasından kaynaklı olarak klasik regresyon yöntemlerinden farklıdır. Sansürlü verilerin bulunduğu regresyon yöntemlerinde Cox regresyon analizi kullanılmaktadır. Ayrıca regresyon yöntemlerine alternatif olarak geliştirilen Çok Değişkenli Uyumlu Regresyon Analizi (MARS) mevcuttur. MARS yönteminin bağımsız değişkenler arasındaki çoklu bağlantı sorunundan ve uç değerlerden etkilenmemesi bakımından klasik regresyon modellerine göre üstünlük göstermektedir. Fakat amaç hem yaşam süresinin tahmin edilmesi hem de bu üstünlükleri bir arada kullanabilmek olduğunda her iki yöntem de tek başına kullanılamamaktadır. Son yıllarda MARS yöntemi ile birlikte Cox regresyon yöntemini bir arada analiz eden Survival MARS yöntemi kullanılmaktadır. Bu analizin ilk aşamasında Cox regresyon analizi uygulanmakta, ikinci aşamasında ise MARS analizi yapılmaktadır. Cox regresyon modelinde yaygın olarak Cox-Snell, Martingale ve Sapma (Deviance) artıkları kullanılmaktadır. En yaygın kullanılan artık türü ise Cox-Snell artıklarıdır. Cox-Snell ve Martingale artıkları; açıklayıcı değişkenlerin Cox orantısal hazard modeli üzerindeki en iyi etkisinin açıklanmasında kullanılmakta ve Martingale artıkları üzerinden dönüştürme uygulanarak Sapma artıkları elde edilmektedir. Ancak bu artıkların, ağırlıklı olarak gözlenen yaşam sürelerine bağlı olması ve kümülatif hazard fonksiyonunun tahminine ihtiyaç duyması birer dezavantajdır. Bu çalışmada; Cox-Snell, Martingale ve Sapma artıklar kullanılarak Survival MARS yaklaşımı sonuçları elde edilmiştir. Verilerin analizinde Matlab 6.0 paket programı kullanılmıştır. Survival MARS analiz performansları Genelleştirilmiş Çapraz Geçerlilik (GCV) ve Açıklayıcılık Katsayı (R^2) değerlerine bakılarak karşılaştırılmıştır. Ayrıca Cox Regresyon ve Survival MARS analizinden elde edilen model katsayılarına ait standart hata değerleri karşılaştırılmıştır.

Bulgular: Cox regresyon modeli istatistiksel olarak anlamlı bulunmuştur ($p=0,005$ Ki-Kare=12.57). Survival MARS sonuçlara göre; Cox-Snell, Sapma ve Martingale artıklı modellerde GCV ve R^2 değerleri sırasıyla ($GCV_{C-S}=0.622$ $R^2_{C-S}=0.48$), ($GCV_S=0.990$, $R^2_S=0.20$), ($GCV_M=0,658$, $R^2_M=0.05$) olarak elde edilmiştir.

Sonuç: Yaşam verilerinin analizinde sonuç değişkeni ile açıklayıcı değişkenler arasındaki lineer ve lineer olmayan ilişki yapıları ve açıklayıcı değişkenler arasındaki interaksiyonların yorumlanmasında Cox regstreyon analizi yetersiz kalmaktadır. Cox regresyon modelinden elde edilen artıkları kullanan Survival MARS yaklaşımı; bu ilişkilerin kolaylıkla yorumlanabilmesini sağlamaktadır. Survival MARS modeli; Cox regresyon modeline karşı güçlü ve uygulanabilir alternatif bir yöntemdir ve Cox-Snell artıklı Survival MARS modeli diğer artıklı modellere göre en uygun yaklaşımdır.

Anahtar Sözcükler: Çok Değişkenli Uyumlu Regresyon, Cox Regresyon, Cox-Snell, Martingale, Sapma (Deviance)

Zamana Bağlı Bir Belirteç İçin Optimal Kesim Noktası Belirleme

İlker ÜNAL, Yaşar SERTDEMİR, Ceren EFE

Çukurova Üniversitesi Tıp Fakültesi Biyoistatistik Anabilim Dalı Adana

ilkerun@cu.edu.tr

Sayısal bir belirtecin sınıflama başarısı ROC analizi ile ölçülür Bu analizde belirtecin tüm bireyler için aynı olan tek bir zamanda ölçüldüğü varsayılır. Belirtecin zamana bağlı olması halinde zaman etkisini dikkate alan zamana bağlı ROC analizine başvurulur. Bu durumda da bireylerin ölçümlerinin farklı zamanlarda olması nedeniyle tüm bireyler için geçerli tek bir kesim noktası vermek doğru olmaz. Zamana bağlı belirtecin ölçümü zamanı dikkate alan bir kesim noktasına göre değerlendirilmesi gerekir. Bu çalışmada literatürde sık kullanılan optimal kesim noktası belirleme yöntemlerini zamana bağlı ROC analizine adaptasyonu yapılarak uygun kesim noktası belirlenmeye çalışılacaktır. Çalışmada literatürde verilen gerçek veri setleri üzerinde uygulamalar yapılarak yöntemlerin sonuçları tartışılacaktır.

Anahtar Sözcükler: Zamana bağlı belirteç, ROC analizi, optimal kesim noktası

Zamana Bağlı ROC Analizi Yöntemlerinin Alan Tahminlerinin Karşılaştırılması

Ceren EFE¹, İlker ÜNAL¹

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

cefe@cu.edu.tr

ROC analizi, hastalık tanısında kullanılan niceliksel tanı testlerinin veya biyolojik belirteçlerin sınıflama başarısını ölçmeye yarayan popüler bir yöntemdir. Ancak sağkalım verilerinde veya boylamsal verilerde olduğu gibi, ilgilenilen olay ile ölçüm yapılan belirteç arasında bir zamana bağlılık söz konusu ise, klasik ROC analizi belirtecin gerçek performansını tahmin edemeyebilir. Bu nedenle, sonuçları zamana bağlı olarak değişkenlik gösteren bir biyolojik belirtecin tanı performansını değerlendirirken zaman kavramını da hesaba katarak işlem yapan “zamana bağlı ROC analizi” yöntemi geliştirilmiştir. Takip gerektiren çalışmalarda her hastanın ölçüm bilgisi düzenli olarak kaydedilemeyebilir. Böyle bir durum söz konusu olduğunda, herhangi bir zaman noktasına ait ölçümü eksik olan bireylerin belirteç bilgisi, aynı zaman noktasına ait ölçümleri tam olan bireylerin verisinden faydalanılarak modellenenabilir. Belirtecin ilgilenilen zamandaki sınıflama başarısı bu modellenmiş veriden elde edilen zamana bağlı ROC eğrisi altında kalan alan yoluyla belirlenir. Bu çalışmada, üretilen geniş boylamsal veriden her birey için rastgele zaman noktaları seçilerek zamana bağlı ölçüm içeren sağkalım verisi oluşturulacaktır. Elde edilen yeni veride belirtecin sınıflama başarısı zamana bağlı ROC yöntemleri ile elde edilecek ve sonuç, orijinal veriden elde edilen sınıflama başarısı ile karşılaştırılarak yöntemlerin tahminlerinin ne kadar iyi olduğu araştırılacaktır.

Anahtar Sözcükler: Zamana Bağlı ROC, Biyolojik Belirteçler, En Yakın Komşu Kestirimi, Kaplan-Meier Kestirimi

Sınıf Dengeleyici: İki Sınıflı Verilerde Sınıf Dengesizliği Problemini Gidermek İçin Web Tabanlı Bir Yazılım

Ahmet Kadir ARSLAN¹, Cemil ÇOLAK¹

¹İnönü Üniversitesi Tıp Fakültesi Biyoistatistik ve Tıp Bilişimi Anabilim Dalı, Malatya

Amaç: Sınıf dengesizliği problemi, makine öğrenmesi alanındaki önemli konulardan biridir. Üzerinde çalışılan veri setinde, örneklerin oluşturduğu sınıflardan birinin diğer örneklerin oluşturduğu sınıf veya sınıflara göre sayıca yüksek derecede fazla olması dengesiz sınıf problemini ortaya çıkarır. Örneğin, tıbbi veri setlerinde hasta sınıfına ilişkin örnek sayısı, hasta olmayan (kontrol) sınıfına göre azdır (özellikle nadir görülen hastalıklarda). Bu durumda hasta ve hasta olmayan bireylerin sınıflandırılmasında yanlılık (bias) sorunu ortaya çıkmaktadır. Çünkü sınıflandırma problemlerinde kullanılan standart makine öğrenmesi modelleri büyük sınıf(lar)ın etkisi altında kalmaktadır ve küçük sınıfların varlığı ortadan kaybolmaktadır. Literatürde bu problemin çözümüne yönelik birçok yaklaşım önerilmiştir. Genel olarak bu yaklaşımlar içsel ve dışsal yaklaşımlar olmak üzere ikiye ayrılabilir. Dışsal yaklaşımlara SMOTE, Borderline SMOTE, LN-SMOTE, SMOTE-RSB, Safe-Level-SMOTE ve SMOTE-IPF vb. algoritmaları, içsel yaklaşımlara ise SMOTEBoost, SMOTEBagging, RAMOBoost, UnderBagging, Roughly Balanced Bagging ve RUSBoost vb. algoritmaları örnek verilebilir. Fakat bu algoritmalar çeşitli programlama dillerinde kod olarak çalıştırıldığı ve kullanıcı arayüzüne sahip olmadıkları için kullanıcı dostu değildir. Bu sebeple dengesiz sınıf probleminin giderilmesi için dili Türkçe olan kullanıcı dostu bir yazılım geliştirilmiştir.

Yöntem: Bu yazılımda, sınıf dengesizliği probleminin giderilmesi için dışsal yaklaşımlar kullanılacaktır. Dışsal yaklaşımlarda, azınlık sınıf verilerinden yeniden örnekleme yöntemleri yardımıyla sentetik veriler türetilerek çoğunluk sınıfın sayısına eşit veya hemen hemen eşit duruma getirilir. Azınlık sınıf kullanılarak yapılan bu yeniden örnekleme yöntemine üst örnekleme (oversampling) denir. Diğer yandan çoğunluk sınıf verilerinin sayısı alt örnekleme (undersampling) yöntemi kullanılarak azınlık sınıfın sayısına çekilebilmekte ve veri seti dengeli hale getirilebilmektedir.

Yazılımın arka planını R programlama dilinde yazılmış paketler oluşturmaktadır. Kullanıcı arayüzü ise Shiny paketi ile oluşturulmuştur. Shiny, R dilinde oluşturulan uygulamaların birer web uygulamasına dönüşmesini sağlayan bir RStudio projesidir. Yazılımda, veri yükleme, yüklenen veri setinin değişken tipinin belirlenmesi, aşırı/aykırı değer ve kayıp değer analizleri, tanımlayıcı istatistikler, dağılım tanılama, alt-üst örnekleme ve modelleme aşamaları bulunmaktadır.

Bulgular: Yazılımda; yüklenen veri setinin değişken tipleri belirlendikten sonra, aşırı/aykırı değer ve kayıp değer analizleri yapılarak çeşitli veri alt-üst örnekleme analizine hazır hale getirilir. Tanımlayıcı istatistikler aşamasında, değişkenlerin grup bazında bazı temel istatistikleri hesaplanır ve tablo olarak sunulur. Dağılım tanılama aşamasında, sayısal değişkenlerin bazı teorik olasılık dağılımlarına uygunluklarına bakılır ve değişken bazında olasılık yoğunluk grafikleri verilir. Alt-üst örnekleme aşamasında veri çeşitli algoritmalar kullanılarak dengeli hale getirilir. Modelleme aşamasında ise sınıf dengesizliği giderilen verinin, çeşitli makine öğrenmesi modelleri kullanılarak sınıflandırma performansı incelenir.

Sonuç: Geliştirilen bu yazılım ile araştırmacılar, veri madenciliği ve makine öğrenmesi disiplinlerindeki önemli problemlerden biri olan sınıf dengesizliği sorununu gidermek için yaygınlıkla kullanılan yöntemleri bir arada bulabilecekler ve bu yöntemlerin teorik arka planını öğrenmeye gerek kalmaksızın alt-üst örnekleme analizlerini uygulayabileceklerdir.

Anahtar Sözcükler: Alt-Üst Örnekleme, Shiny, Web Tabanlı Yazılım

Güvenirlilik Çalışmalarında Örnek Büyüklüğü Hesaplama Yaklaşımları

Sevinç Püren YÜCEL, Zeliha Nazan ALPARSLAN

Çukurova Üniversitesi Tıp Fakültesi Biyoistatistik Anabilim Dalı, Adana

sevincpurenyucel@gmail.com

Giriş: Güvenirlilik çalışmalarında, ölçeklerin güvenirliliğini belirlemek adına birçok yöntem vardır. Burada önemli olan nokta, bu güvenirlilik çalışmalarının planlama aşamasında, “bir ölçeğin güvenirlilik katsayısını doğru olarak belirleyebilmek için ne kadar örnek büyüklüğü almamız gerektiği” sorusuna cevap bulmaktır. Cronbach alfa güvenirlilik katsayısı, sağlık bilimlerinde en çok kullanılan güvenirlilik ölçütlerinden biridir. Cronbach alfa güvenirlilik katsayısına bağlı olarak istenilen doğruluk (desired precision) ve istenilen güç (desired power) için olmak üzere iki farklı örnek büyüklüğü hesaplama yaklaşımı geliştirilmiştir.

Amaç: Güvenirlilik çalışmalarının planlanması aşamasında, Cronbach’ın alfasına bağlı örnek büyüklüğü yaklaşımlarının; beklenen farklı Cronbach alfa katsayılarına, ölçeklerin içerdiği madde sayısına, farklı güven aralığı genişliklerine ve farklı güç seviyelerine bağlı olarak değişimlerinin incelenmesi amaçlanmıştır.

Yöntem: İstenilen doğruluk için örnek büyüklüğü elde etme üzerine kurulu simülasyon çalışmasında, paralel ölçüm varsayımı göz önüne alınarak, beklenen farklı Cronbach alfa değer aralıkları (0,60-0,69 ve 0,70-0,79 “orta güvenirlilik”, 0,80-0,89 ve 0,90-0,99 “yüksek güvenirlilik”), ölçekte bulunabilecek farklı madde sayıları (5, 10, 20, 30, 40, 50, 75, 100) ve istenen farklı güven aralığı genişlikleri (0.05, 0.1 ve 0.15) alınarak 1000 tekrarlı simülasyonla örnek büyüklüğü yaklaşımı değerlendirilmiştir.

İstenilen güç için örnek büyüklüğü elde etme üzerine kurulu simülasyon çalışmasında ise, planlanan farklı Cronbach alfa değer aralıkları ve bu aralıklara bağlı olarak $H_0: \alpha = h$ hipotezini test etmek adına farklı h değerleri dikkate alınmıştır. Ölçekte bulunan madde sayıları istenilen doğruluk için yapılan çalışma ile aynı iken güç değerleri 0.80, 0.85 ve 0.90 olarak alınmıştır.

Bulgular ve Sonuç: Simülasyon çalışmalarından elde edilen genel bulgular:

- Güven aralığı genişliği arttıkça örnek büyüklüğünün azalması,
- Beklenen Cronbach alfa düzeyi arttıkça örnek büyüklüğünün azalması,
- İstenilen güç düzeyi arttıkça örnek büyüklüğünün artması,
- Madde sayısı arttıkça örnek büyüklüğünün azalmasıdır.

Sonuç olarak, herhangi bir bilimsel araştırmada olduğu gibi güvenirlilik çalışması yaparken de, çalışmanın amacına uygun minimum örnek büyüklüğünün hesaplanması önemlidir. Sonuçlar kullanımı kolaylaştırmak için mümkün olduğunca genellenerek verilecektir.

Anahtar Sözcükler: Cronbach alfa, güvenirlilik, güç, örnek büyüklüğü, güven aralığı genişliği.

Çoklu Regresyon Modellerinde Örneklem Büyüklüğü ve Güç Hesabı: Bir Benzetim Uygulaması.

Ayhan PARMAKSIZ¹, Reha ALPAR²

¹ Hacettepe Üniversitesi Biyoistatistik AD Doktora Öğrencisi, ANKARA

² Hacettepe Üniversitesi Tıp Fakültesi Biyoistatistik AD, ANKARA
aparmaksiz@gmail.com

Amaç: Bu çalışmanın amacı, çoklu regresyon modellerinde örneklem büyüklüğünün belirlenmesine ilişkin kural ve yöntemleri tartışmak, bağımlı ve bağımsız değişkenler arası korelasyon yapılarının örneklem büyüklüğü üzerine etkisini (Maxwell,2001) bir benzetim çalışması ile değişik örneklem büyüklüklerinde R^2 ve katsayı kestirimlerine ilişkin güç değerlerini hesaplamak suretiyle göstermektir.

Yöntem: Çoklu regresyon modellerinde, örneklem genişliği genellikle pratik bir yaklaşımla parmak hesabı (rules of thumb) dediğimiz bağımsız değişken sayısı ile ilişkili olarak önerilen kurallarla kabaca, güç hesabı dikkate alınmadan belirlenmektedir. Güç hesabı prosedürlerini dikkate alarak örneklem büyüklüğünün belirlenmesine yönelik çalışmalar da yapılmıştır (Green,1991). Maxwell'in (2001), Green'in çalışmasını bir adım öteye taşıyarak bağımlı ve bağımsız değişkenler arasındaki korelasyon yapılarını da dikkate alan bir hesaplama yöntemi geliştirmiştir. Bu çalışmada 1 bağımlı, 5 bağımsız değişkenli, bağımsız değişkenler arası korelasyon %40, bağımlı değişken ile bağımsız değişkenler arası korelasyon ise %30 ve R^2 %17,5 olacak şekilde üretilen 1000000 gözlemlik bir popülasyondan, 100, 250, 500 ve 700 büyüklüklerinde rastgele örneklemeler çektik. Her bir örneklem büyüklüğü için 10000 tekrar yapılmıştır. Her örneklem için regresyon modeli kurulmuş, R^2 ve katsayı kestirimleri yapılmıştır. Örneklem dağılışının istatistikleri (Ortalama, SS(SH), Minimum, Maksimum, Coverage Probability, Average Width, Güç, RMSE) hesaplanmıştır.

Bulgular: Örneklem dağılışları parametre etrafında normal dağılım göstermesine rağmen standart hata değerleri küçükten büyüğe örneklem büyüklükleri için sırasıyla 11,7,5 ve 4 şeklinde bulunmuştur. Kestirimlere ilişkin % 95 güven aralıkları ve güç hesapları yapılmıştır. Yine sırasıyla güç değerleri yaklaşık olarak %18, %40, %65 ve %80 olarak bulunmuştur. Sonuçlar Maxwell'in bulgularıyla uyumlu çıkmıştır.

Sonuç: Regresyon modellerinde elde edilen katsayıların yokluk hipotezi yanlış iken yani evren parametresinin 0 'dan farklı olduğu durumlarda elde edilen kestirim için yapılan hatanın düşük (yani gücün yüksek) olması kestirimin güvenilirliği için gereklidir. Benzetim çalışmamız göstermiştir ki değişkenler arası korelasyon ilişkisi göz önünde bulundurularak yeterli örneklem büyüklüğü hesaplanmalıdır. Bunun için ise ya önceki çalışmalardan yararlanarak ya da bir pilot çalışma ile değişkenler arası korelasyon bilgilerine ulaşarak örneklem büyüklüklerinin belirlenmesi gerekmektedir. Günümüze kadar yapılan regresyon çözümlemelerine ilişkin çalışmaların birçoğunun gücünün düşük olduğu rahatlıkla söylenebilir. Bu çalışmalardan elde edilen katsayı kestirimlerine güvenilemeyeceğinden ancak bir meta analiz çalışması ile daha anlamlı hale gelebilir. Ayrıca şunu da belirtmek gerekir ki aynı korelasyon yapısında gerekli örneklem büyüklüğü G Power paket programıyla hesaplandığında 250 civarı bir örneklem büyüklüğü hesap etmektedir ki bu sayının katsayı kestirimleri için yeterli gücü sağlamayacağı açıktır.

Anahtar Sözcükler: Çoklu Regresyon Modelleri, Güç hesabı, Örneklem Büyüklüğü, Benzetim Uygulaması

Karşılaştırılacak Grup Sayısının Çok Fazla Olduğu Durumlarda Çoklu Karşılaştırma Problemi İçin Alternatif Bir Çözüm Yolu: ANOM Tekniği

Mehmet MENDEŞ, **Soner YİĞİT**

Çanakkale Onsekiz Mart Üniversitesi Ziraat Fakültesi, Zootekni Bölümü, Biyometri ve Genetik ABD-17100-Çanakkale

Amaç: Bu çalışmanın amacı; grup ortalamalarının, varyansların, oranların ve diğer yer ve değişim ölçülerinin karşılaştırılmasında etkin bir şekilde kullanılabilen Ortalamaların Analizi Yönteminden (Analysis of Means (ANOM)) özellikle karşılaştırılacak grup sayısının çok fazla olduğu durumlarda ($k > 30$) bir çoklu karşılaştırma yöntemi olarak kullanılabileceğini göstermektir. Çünkü, ANOM grafiksel bir yöntem olduğu için sonuçların yorumlanması çok kolay olmakta, gözlenen farkın istatistiksel önemliliğinin yanında Pratik ya da klinik önemliliğinin değerlendirilmesine de imkan verebilmekte ve karşılaştırılan muamele gruplarından hangilerinin daha üstün hangilerinin de daha kötü olduğu oldukça kolay bir şekilde belirlenebilmektedir.

Yöntem: Bu çalışmada Monte Carlo Simulasyon Tekniğinden yararlanılarak ANOM yönteminin birçok deneme koşulundaki performansı ortaya konulmuştur. Aynı zamanda karşılaştırılacak grup sayısının oldukça fazla olduğu durumlarda ($k=80$) ANOM yönteminden bir çoklu karşılaştırma testi olarak nasıl etkin bir şekilde yararlanılabileceğini göstermek amacıyla gerçek bir very seti üzerinden ANOM yöntemi ile Tukey çoklu karşılaştırma testi karşılaştırılmıştır.

Bulgular: Simulasyon sonuçları, ANOM yönteminin özellikle karşılaştırılacak grup sayısının çok fazla olduğu durumlarda oldukça iyi bir performans sergilediğini ve dikkate alınan deneme koşulları ne olursa olsun toplam doğruluk seviyesinin %60'ın altına düşmediğini göstermiştir. Aynı zamanda ANOM yönteminin performansının etkileyen major faktörün örnek hacmi ve etki büyüklüğü olduğu belirlenmiştir

Sonuç: Simulasyon sonuçları ve gerçek veri seti üzerinden yapılan değerlendirmeler sonucunda ANOM yönteminin özellikle karşılaştırılacak muamele grup sayısının fazla olduğu durumlarda klasik çoklu karşılaştırma yöntemlerine göre oldukça avantajlı olduğu ve bu yöntemden karşılaştırılacak grup sayısının çok fazla olduğu durumlarda grup ortalamaları arasındaki farkların, varyansların, oranların karşılaştırılması yanında bir çoklu karşılaştırma yöntemi olarak ta etkin bir şekilde yararlanılabileceği sonucuna varılmıştır.

Anahtar Sözcükler: ANOM, Çoklu karşılaştırma yöntemleri, Duyarlılık, Toplam doğruluk

Nokta İki Serili Korelasyon (Poin Biserial Correlation) Katsayısı ve İki Örnek t-testinin Performansları Bakımından Karşılaştırılması

Soner YİĞİT, Mehmet MENDEŞ

Çanakkale Onsekiz Mart Üniversitesi Ziraat Fakültesi, Zootekni Bölümü, Biyometri ve Genetik ABD-17100-Çanakkale

Amaç: Bu çalışmanın amacı, değişik deneme koşullarında Nokta İki Serili korelasyon katsayısı ve İki Örnek t-testinin Monte Carlo Simülasyon tekniği kullanılarak değişik deneme koşulları altında 1. Tip Hata ve Testin Gücü bakımından karşılaştırılmasıdır.

Yöntem: Bu çalışmanın materyalini Fortran Power Station Developer Studio'nun IMSL kütüphanesinin RNNOA, RNBET ve RNSTT fonsiyonlarından üretilmiş tesadüf sayıları oluşturmaktadır. Çalışmada Monte Carlo Simulasyon Tekniği kullanılarak farklı etki büyüklüğü (0.0, 0.75 ve 1.50), istenen olayın popülasyondaki oluş ihtimali ($p=0.20, 0.40, 0.50, \text{ve } 0.80$) ve farklı örnek genişliği kombinasyonları ($n=10, 20, 30, 50, 100 \text{ ve } 200$) dikkate alınarak $N(0,1)$ $\beta(10,10)$ ve $t(10)$ dağılımlarından tesadüf sayıları üretilmiştir. Söz konusu testlerin her bir deneme koşulundaki performansları 100.000 simulasyon denemesi sonucunda belirlenmiştir.

Bulgular: Varyanslar homojen olduğu sürece deneme koşulları ne olursa olsun Nokta İki Serili korelasyon katsayısının 1.tip hata olasılıkları neredeyse %5.00 seviyesine eşittir. Diğer taraftan İki Örnek t-testinin 1. tip hata olasılıkları da %5.00 seviyesine oldukça yakın olmasına rağmen $n=10$ olduğunda istenen olayın oluş ihtimali ve dağılımın şekline etkilenmiştir. Varyansların heterojen olması durumunda ise her iki test için gerçekleşen 1. tip hata olasılıkları %5.00 seviyesinden ciddi sapmalar göstermiştir. Bu durum özellikle $n \leq 30$ olduğunda belirgin olarak görülmektedir. Testin gücü değerleri incelendiğinde her iki testin de benzer sonuçlar verdiği görülmüştür. Varyansların heterojen olması durumunda her iki testinde varyansların homojen olduğu durumlara göre testin gücü değerlerinin belirgin şekilde düştüğü görülmüştür. Tüm sonuçlar incelendiğinde ise örnek hacmi, etki büyüklüğü ve istenen olayın oluş ihtimalinin testin gücü değerlerini etkileyen en önemli faktörler olduğu görülmüştür.

Sonuç: Nokta İki Serili korelasyon katsayısı ve İki Örnek t-testi neredeyse tüm deneme koşullarında (özellikle varyanslar eşit olduğunda) benzer sonuçlar vermişlerdir. Ancak t-testi sürekli bir değişken ile ikili (binary) bir değişken arasındaki ilişkinin derecesini vermez. Sadece ikili (binary) değişkenin seviyelerinin arasında sürekli değişken bakımından fark olup olmadığını belirler. Buna karşın Nokta İki Serili korelasyon katsayısı ise hem ikili değişkenin seviyeleri arasında fark olup olmadığını hem de sürekli ve ikili (binary) değişken arasındaki ilişkinin derecesini vermektedir. Bu nedenle bağımsız iki grup ortalamasının karşılaştırılmasında Nokta İki Serili korelasyon katsayısının tercih edilmesi araştırmacılara daha detaylı bilgi sağlayacaktır.

Anahtar Sözcükler: Point Biserial Korelasyon, t-testi, 1. Tip hata, Testin gücü

Bağımlı Değişkenin Sürekli ve Tekrarlı Olduğu Denemelerde Özet İstatistikler Kullanılarak Örneklem Büyüklüğünün Hesaplanması

Duygu SİDDİKOĞLU¹, Funda Seher ÖZALP ATEŞ¹, Emre DEMİR², Yasemin YAVUZ¹

¹Ankara Üniversitesi Tıp Fakültesi, Biyoistatistik Anabilim Dalı, Ankara

²Hitit Üniversitesi Tıp Fakültesi, Biyoistatistik Anabilim Dalı, Çorum

duygu.sddk@gmail.com

Amaç: Bu çalışmada iki gruplu sürekli yanıt değişkenine sahip tekrarlı veri tasarımlarında özet istatistikler kullanılarak örneklem büyüklüğü belirleme ilkeleri sunulacak, özet istatistik seçiminden ve kovaryans yapısından örneklem büyüklüğünün nasıl etkilendiğini gösteren simülasyon sonuçları doğrusal karma modeller yaklaşımıyla bulunan örneklem büyüklükleri ile karşılaştırılacaktır.

Yöntem: Analizin birincil hipotezi genellikle izlem boyunca gruplar arasında değişim oranlarının aynı olup olmadığının veya zaman*grup etkileşimin test edilmesidir. Birim profillerin paterni doğrusal veya monoton olduğunda en küçük kareler yöntemi ile elde edilen regresyon doğrusunun eğimi, her birimin yanıt ve zaman değişkeni arasındaki ilişkiyi özetlemek üzere uygulanmıştır. Anlamlı bir etkileşim etkisinin olmadığı hipotez doğru varsayıldığında, birimlerdeki eğimler önemli ölçüde farklı olmamalıdır. Bu amaçla, her birim için eğimler elde edildiğinde, gruplardaki eğimlerin benzerliği tek yönlü ANOVA testi ile değerlendirilmiştir. Monoton ve doğrusal birim profillerinde çalışmanın son zaman noktasındaki ortalamalar kullanılarak bağımsız gruplarda t testi yöntemi ile, başlangıç noktasında meydana çıkabilecek profil farklılıkları olduğunda ANCOVA yöntemi ile örneklem büyüklüğü hesaplanmıştır. Bu yaklaşımlardan yararlanarak hesaplanan örnek genişlikleri simule verilerle karşılaştırmalı olarak incelenmiştir.

Sonuç: Her iki yaklaşımın da literatürde kullanılan yaklaşımlar olduğu göz önünde bulundurularak, grup içi regresyon doğrusunun eğimi, son ölçümlerin ortalaması gibi özet istatistikler orijinal ölçümler yerine kullanılarak veriler kolay, etkili ve güvenilir bir şekilde özetlenebilir ve örneklem büyüklüğünün kullanılmasında özet istatistikler kullanılabilir. Ancak örneklem büyüklüğünün özet istatistikler yardımıyla belirlenmesi sırasında trendin lineer olması, kovaryans yapısının ve birim-içi korelasyonun doğru belirlenmesi gibi varsayımların önemi göz önünde bulundurulmalıdır.

Anahtar Sözcükler: tekrarlı ölçüm, örneklem büyüklüğü, özet istatistikler

Yanık Görüntülerinin Kümelenmesinde Bulanık Kümeleme Yaklaşımlarının Başarımlarının Ölçülmesi

Yeşim AKBAŞ¹, Tolga BERBER²

¹ Karadeniz Teknik Üniversitesi, Fen Bilimleri Enstitüsü, İstatistik ve Bilgisayar Bilimleri ABD

² Karadeniz Teknik Üniversitesi, Fen Fakültesi, Bilgisayar Bilimleri Anabilim Dalı, Trabzon

Amaç: : Günümüzde birçok tıp alanında olduğu gibi yanık alanında da hastalık teşhisinde karar destek sistemleri kullanılmaktadır. Dünya Sağlık Örgütü her yıl yanık nedeni ile ölüm sayısını 265.000 olarak belirlemiştir ve bu sayı yanık tedavisinin ne denli önemli olduğunu gözler önüne sermektedir.

Tedavinin gerçekleşmesinde göz önüne alınması gereken birçok parametre vardır. Yanık yüzdesi bu parametrelerin en önemlilerindendir. Bu çalışmada söz konusu parametrenin belirlenebilmesi adına yanık/normal deri bölgeleri belirlemek amacıyla kümeleme yöntemleri denenmiştir.

Yöntem: Çalışmada Karadeniz Teknik Üniversitesi Tıp Fakültesi Farabi Hastanesi yanık ünitesine başvuran hastaların hiçbir standart gözetmeksizin çekilen görüntülerini içeren veri kümesinden 10 adet örnek görüntü seçilmiştir. Seçilen örnek görüntüler aracılığıyla en uygun küme sayısının belirlenmesi ve en uygun bulanık kümeleme yaklaşımının seçilmesi hedeflenmektedir. Bu amaçla, RGB uzayında FCM (Fuzzy C Means), LAB uzayında FCM ve PCM (Possibilistic C Means) yaklaşımları denenerek hangi yöntemin daha iyi sonuçlar verdiği araştırılmıştır. En uygun küme sayısının belirlenebilmesi için, yöntemler farklı küme sayıları ile tekrar tekrar denenerek her deneme için “Ayırt Edicilik” ve “Bağlılık” değerleri hesaplanmıştır.

Bulgular: Her bir yanık görüntüsü RGB uzayında FCM, LAB uzayında FCM ve PCM olmak üzere 3 farklı yöntem kullanılarak öncelikle 2 kümeye ayrılarak “Ayırt Edicilik” ve “Bağlılık” değerleri hesaplanmıştır. Bu işlem küme sayısı 20 oluncaya kadar tekrarlanmıştır. Tüm görüntüler için elde edilen ayırt edicilik ve bağlılık değerlerinin ortalamaları alınarak, “Ortalama Bağlılık” ve “Ortalama Ayırt Edicilik” değerleri elde edilmiştir.

	Ortalama Bağlılık Değerleri			Ortalama Ayırt Edicilik Değerleri		
Görüntüler	FCM-RGB	FCM-LAB	PCM	FCM-RGB	FCM-LAB	PCM
Görüntü 1	60.465,90	7.747,84	441.413,00	453.654,00	25.022,70	32.503,70
Görüntü 2	41.336,10	3.731,27	2.809.680,00	490.962,00	32.663,90	5,00
Görüntü 3	22.761,60	2.628,82	408.399,00	261.636,00	16.510,60	710,43
Görüntü 4	29.904,60	4.761,40	788.916,00	334.462,00	24.117,50	145,88
Görüntü 5	33.350,60	3.610,85	358.089,00	229.613,00	17.867,30	8.616,99
Görüntü 6	457,16	1.063,97	86.969,60	3.821,30	4.910,16	1,31
Görüntü 7	667,18	1.222,22	152.443,00	5.569,70	9.160,16	26.975,80
Görüntü 8	36.344,70	4.472,17	528.977,00	321.068,00	27.224,60	42.183,90
Görüntü 9	55.987,30	8.153,59	861.332,00	512.347,00	52.753,40	87.663,00
Görüntü 10	58.136,80	8.305,80	339.421,00	373.296,00	52.710,50	81.648,60

Tablo 1. Ortalama bağlılık ve ortalama ayırt edicilik değerleri (10⁸)

Sonuç: Gerçekleştirilen deneyler sonucunda, seçilen örneklem için en uygun yöntemin LAB uzayında çalıştırılmış olan FCM olduğu gözlemlenmiş ve seçilen örneklem için en uygun küme sayısının da 13 olduğu tespit edilmiştir. Ayrıca, görüntülerin farklı renk uzayları ile denenmesiyle kullanılan yöntemin başarımlarının artacağı öngörülmektedir.

Anahtar Sözcükler: Yanık, Kümeleme, FCM, PCM

Kategorik Kümelenmiş Veri Analizi ve Bir Uygulama

Kübra Elif AKBAŞ¹, Harika Gözde GÖZÜKARA BAĞ¹

¹*İnönü Üniversitesi, Tıp Fakültesi, Biyoistatistik ve Tıp Bilişimi Anabilim Dalı, Malatya*

Amaç: Sağlık alanında yapılan çalışmalarda sıklıkla kümelenmiş veri yapısı karşımıza çıkabilmektedir. Kümelenmiş verileri diğer verilerden ayıran en önemli özelliği aynı kümeden elde edilen sonuçların birbiriyle ilişkili olmasıdır. Bu korelasyon yapısının göz ardı edilmesi ise istatistiksel çıkarımda yanlılığa neden olabilir. Bu çalışmanın amacı, kümelenmiş kategorik veriler için korelasyon yapısını dikkate alan alternatif yöntemleri tanıtmak ve bir veri üzerinde uygulamasını göstermektir. Böylece kümelenmiş kategorik verilere uygulanan standart istatistiksel analiz yöntemlerinin yanlı sonuç verebileceği ve doğru analiz yöntemi ile geçerli bulgular elde edilebileceği örneklenecektir.

Yöntem: Standart ki-kare testi ile önerilen düzeltilmiş testlerin karşılaştırılması ve örneklenmesi amacıyla hipotetik bir sağlık verisi kullanılmıştır. Bu verinin analiz edilmesinde ise standart Pearson ki-kare testi ve kümelenmiş kategorik veriler için önerilen Rosner (1982) düzeltilmiş ki-kare, Dallal (1988) düzeltilmiş ki-kare, Donner (1989) düzeltilmiş ki-kare ve Rao ve Scott (1992) düzeltilmiş ki-kare testleri kullanılmıştır.

Bulgular: Her bir hastanın iki gözü üzerinde yapılan ölçüm sonucunda katarakt varlığı değerlendirilmiştir. Yaş gruplarına göre pozitiflik oranları karşılaştırılmak istenmektedir. Rosner düzeltilmiş ki-kare test istatistiği $T=9.04$; $p=0.010$, Dallal düzeltilmiş ki-kare test istatistiği $D=8.27$; $p=0.040$, Donner düzeltilmiş ki-kare istatistiği $X_A^2=9.12$; $p=0.010$ ve Rao ve Scott düzeltilmiş ki-kare test istatistiği $\tilde{X}^2=8.9$; $p=0.011$ olarak elde edilmiştir. Ayrıca standart ki-kare test istatistiği $X^2=9.2$; $p=0.055$ olarak bulunmuştur.

Sonuç: Kümelenmiş veri yapısını dikkate alan düzeltilmiş test istatistikleri ile H_0 hipotezi reddedilirken, standart ki-kare testi ile H_0 hipotezi kabul edilmiştir. Veri yapısına uygun olmayan istatistiksel yöntemlerin kullanılması yanlış sonuçlara ulaşılmasına neden olabilir.

Anahtar Sözcükler: Kümelenmiş veri, küme-içi korelasyon, yanlılık, düzeltilmiş ki-kare, iki durumlu veri.

Karar Modelleri için Olasılıksal Duyarlılık Analizi ve Uygulaması

Can ATEŞ¹, Yasemin YAVUZ²

¹ Van Yüzüncü Yıl Üniversitesi Tıp Fakültesi Biyoistatistik AD.

² Ankara Üniversitesi Tıp Fakültesi Biyoistatistik AD.

Amaç: Günümüzde, bilim ve teknoloji alanındaki hızlı gelişime paralel olarak “kolay, kullanışlı, hızlı, daha az maliyetli, girişimsel olmayan, ucuz” gibi avantajlarıyla karşımıza çıkan yeni yöntemlerin eskilerinin yerini alıp alamayacakları konusunda daha objektif ve net değerlendirmelere olan ihtiyaç artmaktadır.

Klinik karar problemlerinin çözümü için kullanılan karar analizi modellerinde, bazı değişkenlerin verilen karar üzerinde etkili olduğu bilinmektedir. Bu değişkenlerin değişen değerleri için, model sonucu da değişiklik gösterebilmektedir. Bu nedenle, ilgili değişkenlerin model sonucunda verilen karar üzerinde ne düzeyde etkili olduğu, diğer bir deyişle modelin değişkenlere ne düzeyde duyarlı olduğu incelenmelidir.

Bu çalışmanın amacı, klinik karar modellerinde karşılaşılabilecek belirsizliklerin, varılacak karara olan etkilerini belirlemek amacıyla kullanılan Olasılıksal Duyarlılık Analizi yöntemini tanıtmak ve klinikte karşılaşılan bir karar probleminin incelendiği analizler ile varılan kararın hangi faktörlerin ne seviyedeki etkileri ile farklılık gösterebileceğini ortaya koymaktır.

Yöntem: Bu amaç doğrultusunda, Erişkin Omurga Bozukluğu tanısı almış hastalar için etkin tedavi yönteminin seçilmesi amacıyla klinik bir karar analizi modeli oluşturulmuş ve bu modelle varılacak karar etkili olduğu düşünülen yaş ve vücut kitle indeksi (VKİ) parametreleri, Olasılıksal Duyarlılık analizleri ile incelenmiştir.

Bulgular: Temel karar modeli olarak tanımlanan başlangıç modelimizde cerrahi strateji daha iyi bulunurken, modele duyarlı bulunan yaş ve VKİ değişkenlerinin etkilerini içeren final modelimizde konservatif tedaviye ilişkin beklenen fayda değerinin cerrahi tedaviye göre daha yüksek olduğu gözlenmiştir (sırasıyla beklenen faydalar; 0,84 ve 0,82).

Sonuç: Bu anlamda zor ve karmaşık yapıdaki klinik karar analizleri için Duyarlılık Analizleri gerçekleştirilmeden verilmiş bir kararın güvenilirliğinin çok düşük olduğu ve bu yanlış karara ilişkin geri dönüşü olmayan klinik sıkıntılar yaşanabileceği göz ardı edilmemelidir.

Anahtar Sözcükler: Duyarlılık Analizi, Karar Analizi, Karar Ağacı, Olasılıksal Duyarlılık Analizi.

Deep Learning ile Yeni Nesil Sekanslama Verilerinde Okuma Kalitelerinin Tahminlenmesi

Erdal COŞGUN

Data Scientist, Genomics Team, Microsoft Research, Redmond, WA, US

ercosgun@microsoft.com

Amaç: Genetik araştırmalar 3 aşama ile ifade edilebilir: 1. Biyolojik örneklerin elde edilmesi ve sekanslanması 2. Variant Calling (VC-referans genom ile örnek genom eşlendikten sonra varyasyonların bulunması), 3. Biyoinformatik Analizler (Tertiary Analysis). Bu aşamalardan birincisi ilk olarak 1953'te Watson ve Crick tarafından DNA'nın 3 boyutlu yapısının çözülmesi ile başlamıştır. Ancak bu çalışmalar hazır ve görelili olarak daha basit RNA yapısına sahip türlerde tRNA veya mikrobiyal ribozomal'a odaklanmıştır. İkinci nesil sekanslama daha fazla sayıda paralel sekanslama yapan (daha sonradan Roche tarafından satın alınan) 454 tarafından başlatıldığı kabul edilir. Bu sistem çok sayıda okuma yapabilen önemli bir teknoloji getirmişti. Yeni Nesil Sekanslama (NGS)'nin önceki yöntemlere göre en önemli farklılığı ise: Tek Molekül Sekanslama, anlık sekanslama olarak açıklanabilir. Biyoinformatik analizler ise Biyoistatistik bilimini bilgisayar bilimi ile birleştiren önemli bir aşamadır. Bu çalışmada temel amaç: NGS deney düzenlerinin ikincil analizleri sonrası elde edilen variant calling format (VCF) verisinden kalite skorlarını deep learning ile tahminlemektir. Kalite skorları özellikle "kişiselleştirilmiş tıp" çalışmaları için çok önemli olmakla birlikte sekanslama cihazının ve VC çalışmalarının etkinliğini belirlemede kullanılan bir belirteçtir.

Yöntem: Variant Calling için en sık kullanılan ardışık düzen olan (pipeline) Broad's Enstitüsünün BWA+ GATK tekniği Microsoft Genomics Servis aracılığıyla kullanılmıştır. Tüm genom sekanslama için 'Genome In the Bottle' veri setlerinden **NA12877, NA12878, NA12879** hastalarına ait veriler kullanılmıştır. Her hastanın verisi için **VC işlemi 20 kez** tekrarlanmıştır. Büyük boyuttaki verinin analizi için Microsoft Cognitive Toolkit (CNTK) Neural Network ve kontrol olarak Random Forest, klasik Neural Network kullanılmıştır. VCF veri setindeki DP:Toplam Okuma derinliği, AF: Allel frekansı, Read Depth: Okuma derinliği, HQ: Haplotip Kalitesi değişkenliği ile QS: Kalite Skoru tahmini (Phred-scaled quality score <10) yapılmıştır.

Bulgular: Sekanslama sonucu beklenen kalite düzeyini tahminlemek için 3 farklı model kurulmuştur. (Tablo 1)

Tablo 1. Tahminleme modellerinin karşılaştırmalı sonuçları

%F20'ün Ortalaması)	NA-12877			NA-12878			NA-12879		
	CNTK	RF	Nnet	CNTK	RF	Nnet	CNTK	RF	Nnet
Doğruluk	78.75	72.1	79.45	57.15	51.48	53.15	68.59	62.71	59.14
Recall/Duyarlılık	74.86	71.6	77.37	56.24	49.57	54.9	69.45	62.54	57.19
Precision	75.36	72.7	80.54	59.14	53.49	51.58	67.25	63.14	60.17

Sonuç: Bu çalışma özellikle giderek ucuzlayan NGS çalışmalarının analizi için deep learning tekniklerinin uygulanabilir olduğunu göstermek için önemlidir. Her bir örnek için (30x coverage, tüm genom) büyüklüğü 100 GB olan NGS verilerinden elde edilen VC verileri (~ 2 GB) deep learning teknikleri ile GPU bazında yaklaşık **45 dakikada** analiz edilmiştir. Random Forest ve Neural Network ile analiz süresi **440 dakika** olarak kaydedilmiştir. Hem hız hem de tutarlılık

bakımından deep learning araştırmacılara avantaj sağlamaktadır. NA-12877 kodlu hasta dışında CNTK daha yüksek doğruluk ile tahminleme yapmıştır. Bu hastadaki kötü performansın temel nedeni hastada ortaya çıkan varyantların 'nadir' görülen hastalıklarla ilişkili olması olarak öngörülmüştür. Bu sonuç da henüz deep learning'in bu tip hastalarda çalışmak için yeterli olmadığını göstermektedir (olası neden:multi-threading). Ancak yine de bu hastada dahi Random Forest gibi çok tercih edilen bir yöntemle göre daha güvenilir sonuçlar vermiştir. Sonuç olarak yakın gelecekte bulut sistemlerinde Graphics Processing Unit (GPU) tabanlı analizlerin biyoistatistik uzmanları tarafından kullanımının bir zorunluluk haline geleceği açıktır.

Anahtar Sözcükler: *Deep learning, Yeni Nesil Sekanslama, Variant Calling Format, Microsoft Cognitive Toolkit*

Kaynaklar:

1. Heather, J. M., & Chain, B. (2016). The sequence of sequencers: The history of sequencing DNA. Genomics, 107(1), 1–8. <http://doi.org/10.1016/j.ygeno.2015.11.003>
2. Yu, J. et al. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nat. Genet. 38, 203–8 (2006)
3. Lippert, C. et al. FaST linear mixed models for genome-wide association studies. Nat. Methods 8, 833–5 (2011).
4. <https://www.microsoft.com/en-us/cognitive-toolkit/> - Erişim tarihi 3 Temmuz 2017

Akademik İstatistiksel Danışmanlık Merkezleri Üzerine Bir İnceleme ve Model Önerisi

Bülent ÇELİK, M. Akif BAKIR, M. Yavuz ATA

Gazi Üniversitesi Fen Fakültesi İstatistik Bölümü, Ankara

Özet

Amaç: Bu çalışmada, Amerika Birleşik Devletleri'nin bazı üniversitelerindeki Akademik İstatistik Danışma Merkezleri hakkında yapılan inceleme sonuçları verilerek yeni bir model önerisi geliştirilmiştir.

Yöntem: Birinci aşamada akademik çalışmaları istatistiksel yeterlilik açısından değerlendiren yayınlar incelenmiştir. İkinci aşamada istatistiksel danışmanlık ve akademik istatistiksel danışmanlık merkezleri hakkında yazılan bilimsel makaleler incelenmiştir. Üçüncü aşamada üniversitemizde faaliyet gösteren istatistiksel danışmanlık merkezleri incelenmiştir. Son aşamada ise araştırmacının bizzat görev yaptığı veya görüşme yaptığı dünyadaki ve ülkemizdeki istatistiksel danışma merkezlerinin çalışmaları incelenmiştir.

Bulgular: Bu çalışmada, ülkemizde ve diğer ülkelerde yapılan akademik çalışmaların istatistiksel açıdan değerlendirmesi ile ilgili çeşitli yayınlar incelenmiştir. Bu incelemeler neticesinde diğer disiplinlerdeki bilimsel çalışmaların istatistiksel analizlerinde yöntem ve sonuç değerlendirmeleri bakımından dikkate değer yoğunlukta yanlışlıkların yapıldığı görülmüştür. Ayrıca üniversitemizde yapılan tezlerin istatistiksel açıdan incelenmesi neticesinde benzer sonuca ulaşılmıştır. Yapılan bu çalışmalar akademisyenler için istatistiksel danışmanlık hizmetinin gerekliliğini ortaya çıkarmıştır. Bu nedenle üniversitemizde istatistiksel danışmanlık merkezi kurma fikri ortaya çıkmıştır. Bu merkezin faaliyetinin en iyi şekilde yürütülebilmesi için düşüncelerimizi uygulamaya geçirmeden önce yerinde incelemeler yapmak amacıyla dünyadaki örnekler incelenerek, Michigan State University, Center for Statistical Training and Consulting'de yerinde üç aylık araştırma yapılmış ve bu merkezin çalışmaları da örnek alınarak yeni bir model geliştirilmiştir.

Sonuç: Sayısal olarak gittikçe artan akademik yayınların kalitesinin artırılmasında ve bu yayınlarda yer alan istatistiksel süreçlerin doğru ve yeterli kullanılmasında İstatistiksel Danışmanlık faaliyeti önemli bir yer tutmaktadır. Araştırmacılar üniversitelerin istatistik, biyoistatistik ve bazen diğer bölümlerinden istatistiksel danışmanlık hizmeti almakta ancak başvuru bu bölümlerin artan talebi karşılama yetersiz kaldıkları dikkat çekmektedir. İstatistik ile ilgili disiplinlerde öğrenim gören lisans ve lisansüstü öğrencilerin gerçek uygulamalardan uzak oldukları, yapılan bilimsel çalışmalarda istatistik yöntemlerin kullanımında hatalar olduğu da gözlenen önemli hususlardır. Üniversitelerde iyi bir yapıyla ve destekle kurulan veya kurulacak İstatistiksel Danışmanlık Merkezleri talepleri karşılama ve akademik yayın kalitesinin artırılmasında bir çözüm olabilir. Ayrıca kurulan akademik istatistiksel danışma merkezlerinin kalite denetim prosesine sahip olmadıkları da tespit edilmiştir. Görülen aksaklıklardan hareketle araştırmacılar tarafından geri beslemeli kalite denetim sistemi prosesine sahip bir model önerilmiştir.

Anahtar Sözcükler: İstatistiksel danışmanlık, İstatistiksel danışma merkezleri, Akademik yayınlar

İnternet Tabanlı Sağkalım Olasılığı Hesaplayıcıların Güvenilirliği

Uzun S.¹, Sacar D.², Öztürk MA.², Yumuk PF.², Bekiroğlu N.¹

¹Marmara Üniversitesi, Tıp Fakültesi, Biyoistatistik Anabilim Dalı, İstanbul.

²Marmara Üniversitesi, Tıp Fakültesi, İç Hastalıkları Anabilim Dalı, İstanbul.

Amaç: Sağkalım çözümlemesinde; son günlerde oldukça popüler olan internet tabanlı kişisel sağkalım olasılığı hesaplayıcılarının, biyoistatistik alanındaki Kaplan- Meier yöntemi ile sonuçlarının karşılaştırılması yapılmış ve bu karşılaştırma sonucunda Kaplan- Meier yönteminin, kişisel sağkalım hesaplayıcılarına göre, daha yüksek sağkalım olasılıkları verdiği görülmüştür.

Yöntemler: Veriler, Marmara Üniversitesi Hastanesi'nin Medikal Onkoloji servisinde, 1994 yılı ile 2009 yılları arasında meme kanseri tanısı konulan hastalardan oluşmaktadır. Veri temizleme işleminden sonra, 698 adet hasta çalışmaya katılmıştır.

Bu araştırmada internet tabanlı sağkalım olasılığı hesaplayıcılarından 'PREDICT' programından yararlanılmış olup, meme kanseri tanısı konulan hastaların herbiri için 5 yıllık ve 10 yıllık olmak üzere, sağkalım olasılıkları elde edilmiştir. Ancak PREDICT programında sağkalım olasılıklarını hesaplamak için, yaş, tanı tespit yöntemi, MetLN, tümör boyutu, tümör derecesi, ER, Her2, Ki67 ve tedavi türü gibi değişken bilgilerinin girilmesi gerekmektedir. 698 kişiye ait sağkalım olasılıklarının ortalaması hem 5 yıllık hemde 10 yıllık için hesaplanmıştır.

Aynı hasta grubuna Kaplan- Meier yöntemi uygulanarak, 5 ve 10 yıllık sağkalım olasılıkları elde edilmiştir. 7 değişken için, tekdeğişkenli analiz yapılmış ve kategorik verilerde Logrank testi, sürekli değişkenlerde ise Cox Regresyon analizi yapılarak, anlamlı değişkenler elde edilmiştir. Sonrasında bu anlamlı değişkenlere çokdeğişkenli Cox regresyon analizi yapılmıştır.

Sonuçlar: Predict programı ile 698 hasta için 7 değişken kullanarak elde edilen 5 ve 10 yıllık ortalama sağkalım olasılıkları sırasıyla %84 ve %72 olarak hesaplanmıştır.

Aynı hasta grubu için Kaplan- Meier yöntemi ile sağkalım olasılıkları, 5 yıllık için % 93,7 bulunurken, 10 yıl için ise %80,2 olarak hesaplanmıştır.

Predict programında sağkalım olasılıkları hesaplanırken kullanılan değişkenlerin Kaplan-Meier yöntemi ile anlamlı değişkenler olup olmadığını araştırmak üzere yapılan tekdeğişkenli analizdeki Logrank testi ve Cox regresyon sonuçlarına göre söz konusu 7 değişkenden; yaş, MetLN, tümör boyutu ve tümör derecesi anlamlı çıkmıştır. Yapılan çokdeğişkenli Cox regresyon çözümlemesinde ise hastanın sağkalım süresine etki eden değişkenler öncelik sırasına göre; MetLN ve yaş olarak sıptanmıştır.

Tartışma ve Sonuç: Son zamanlarda kullanım ve ulaşım kolaylığı açısından internet ortamında çok hızla yayılan PREDICT gibi sağkalım olasılıklarını hesaplayan yazılımlar oldukça kullanılmaktadır.

Oysa bu tür sağkalım olasılıkları hesaplayan yazılımlarda, anlamlı olup olmadığı test edilmeden bazı değişkenler dikkate alınarak hesaplanan sağkalım olasılıkları, Kaplan -Meier yöntemiyle hesaplanan sağkalım olasılıklarından çoğu zaman beklendiği gibi daha düşük çıkmaktadır. Ancak bu değişkenlerin istatistiksel olarak anlamlılıkları tartışma konusu olabilir ve sağkalım

olasılıkları hesaplanırken kullanılan değişkenler uygulamamızda olduğu gibi anlamlı değişkenler çıkmayabilir.

Sonuç olarak, bu tür hazır paketlerden yararlanan medikal onkoloji doktorlarının hastalarını bilgilendirme ve tedavilerine karar verme aşamasında bu tip hesaplayıcılara ne ölçüde güvenebilecekleri, hala tartışma konusu olmaktadır.

Anahtar Sözcükler: İnternet Tabanlı Sağkalım Olasılığı Hesaplayıcı, Kaplan Meier Yöntemi, Cox Regresyon, PREDICT.

İkili Sonuçlu Bağımsız Kappa İstatistiklerinin Homojenlik Testleri İçin Bir R Paketi

Muammer ALBAYRAK¹, Kemal TURHAN¹, Yasemin YAVUZ², **Zeliha AYDIN KASAP¹**

¹*Karadeniz Teknik Üniversitesi Tıp Fakültesi, Biyoistatistik ve Tıp Bilişimi Anabilim Dalı, Trabzon*

²*Ankara Üniversitesi Tıp Fakültesi, Biyoistatistik Anabilim Dalı, Ankara*

Amaç: Bu çalışma, ikili sonuçlu bağımsız birden fazla çalışmada ya da tabakalara ayrılmış tek bir çalışmada gözlemciler arası uyumun değerlendirilmesi amacıyla elde edilen kappa istatistiklerinin homojen olup olmadığının test edilmesinde kullanılması önerilen beş farklı homojenlik testinin kolaylıkla gerçekleştirilmesini için bir R uygulaması geliştirilmesini amaçlamıştır.

Yöntem: Bu çalışmada, R yazılımı ile geliştirilen algoritma kappa uyum tablosunu temel almaktadır. Her bir test için hazırlanan fonksiyonların tümü üç parametre ile çalışmaktadır. Bunlar sırasıyla, J tane çalışma ya da tabaka için pozitif uyumlu değerlendirmelerin sayısını içeren vektör, uyumsuz değerlendirmelerin sayısını içeren vektör ve negatif uyumlu değerlendirmelerin sayısını içeren vektördür. Her bir homojenlik testi fonksiyonu, aldığı bu parametreleri kullanarak bir ki-kare istatistiği hesaplamakta ve hipotez testi ile farklı çalışma veya tabakalara ait kappa istatistikleri arasındaki farkın istatistiksel olarak anlamlı olup olmadığına karar vermektedir. İlgili R paketi <https://cran.r-project.org/web/packages/kaphom/> adresinde yayınlanmıştır.

Bulgular: Kappa katsayısı, bir durumun varlığı ya da yokluğu üzerine yapılmış iki değerlendirme arasındaki uyumun değerlendirilmesinde kullanılması yaygın olarak kabul görmüş bir ölçüdür. Cohen'in kappa katsayısı ve sınıf içi kappa katsayısı olmak üzere iki tür kappa katsayısı vardır. Bu çalışmanın kapsamı kappa katsayısının sınıf içi türü ile ilgilidir. Çok merkezli çalışmalarda ya da tabakalanmış tek bir çalışmada, elde edilen kappa istatistikleri karşılaştırılmak ve eldeki tüm bilgi kullanılarak ortak ya da özet bir kappa uyumu verilmek istenir. Eğer elde edilen bu J tane kappa istatistiğinin homojenlik testi reddedilmezse, bütün çalışmaları özetleyen tek bir kappa istatistiği üzerinden çıkarımlar yapılabilir. Literatürde iki sonuçlu bağımsız kappa istatistiklerinin homojenlik testi için önerilen beş yöntem bulunmaktadır. Bunlar; Fleiss, Donner Uyum İyiliği, Olabilirlik Skor, Modifiye Edilmiş Skor ve Pearson homojenlik testleridir.

Sonuç: Geliştirilen uygulama, her biri farklı şartlar altında daha iyi performans gösteren beş farklı homojenlik testini de içerdiğinden kullanıcıya istediği testi kullanma seçeneğinin yanında tüm testleri uygulayarak karşılaştırma yapma olanağı vermesi bakımından oldukça kullanışlıdır. Ayrıca bu paket, çalışma ya da tabaka sayısı bakımından herhangi bir sınır olmadığından teoride istenen sayıda kappa istatistiğinin homojenlik testini gerçekleştirmek amacıyla kullanılabilir. Ancak, çalışma ya da tabaka sayısının artmasının pratikte performans ve doğruluğu nasıl etkileyeceğinin ileri simülasyon ve testlerle incelenmesi bir gerekliliktir.

Anahtar Sözcükler: Kappa istatistiği; Homojenlik; Sınıf-içi kappa

Bulanık Çıkarsama Sistemleri ile Veri Madenciliği Yöntemlerinin Sınıflama Performansının Benzetim Çalışması ile Karşılaştırılması

İrem KAR, Batuhan BAKIRARAR, S. Kenan KÖSE,

Ankara Üniversitesi Tıp Fakültesi, Biyoistatistik Anabilim Dalı, Ankara

Özet

Sınıflandırma, veri setinde her gözlem için hedef sınıfı doğru tahmin etmeyi amaçlayan veri madenciliğinin en önemli araştırma konularından birisidir. Sınıflandırma işleminde sistemin eğitim verisi üzerinden öğrenmesi ve benzer bir veri seti ile karşılaşması halinde o verinin ait olacağı sınıfı tahmin etmesi sağlanır. Bilgisayar sistemlerinde öğrenme iki şekilde yapılabilir. Bunlardan birisi makine öğrenmesi bilim dalında da temel kabul edilen veri ile öğrenme; diğeri ise uzman görüşüne dayalı öğrenmedir. Herhangi bir problemin çözümünde elde veri olması halinde makine öğrenmesi teknikleri ile gerekli çıkarsama işlemlerini yapmak mümkün iken, mevcutta veri olmadığı durumlarda, problemin çözümüne ulaşmada uzman görüşünden yararlanan öğrenme teknikleri kullanılmaktadır. Bu çalışmada, uzmandan öğrenen bulanık çıkarsama sistemleri ile veriden öğrenen veri madenciliği yöntemlerinin farklı örnek büyüklüklerinde (100, 250, 500, 1000) sınıflandırma performansları karşılaştırılmıştır. Karşılaştırmada kullanılan veri madenciliği algoritmaları literatürde sıklıkla karşılaşılan sınıflandırıcılar olan, Yapay Sinir Ağları, Random Forest ve Destek Vektör Makinesi'dir. Sınıflandırma performans ölçütleri olarak doğruluk, kesinlik, duyarlılık ve F-ölçütü kullanılmıştır.

Bu çalışmada, University of California Irvine (UCI) makine öğrenmesi veritabanında yer alan kalp hastalığı verisinin özellikleri korunarak değişken önemine göre belirlenen 6 bağımsız değişken (cinsiyet, göğüs ağrı tipi, maksimum kalp atım hızı, egzersizin neden olduğu anjin, oldpeak ve ana damar sayısı) için türetilen veriler kullanılmıştır. Çalışma sonucunda, veri madenciliği yöntemlerinden biri olan Yapay Sinir Ağları ve Bulanık Çıkarsama Sistemleri benzer sonuçlar vermiştir ve Random Forest ile Destek Vektör Makinesi yöntemlerine göre daha iyi sınıflama performansı göstermişlerdir. Verilerin analizi R 3.3.2 programlama dili kullanılarak yapılmıştır.

Anahtar Sözcükler: Bulanık Çıkarsama Sistemleri, Sınıflama, Veri Madenciliği

Ölçeklerde Yapı Geçerliliğinin Değerlendirilmesi: Faktör Analizi ve Rasch Analizi

Funda Seher ÖZALP ATEŞ¹, Derya GÖKMEN¹, Atilla Halil ELHAN¹, Şehim KUTLAY²

¹Ankara Üniversitesi Tıp Fakültesi Biyoistatistik Anabilim Dalı, ANKARA

²Ankara Üniversitesi Tıp Fakültesi Fiziksel Tıp ve Rehabilitasyon Anabilim Dalı, ANKARA

Amaç: Sağlık alanında daha duyarlı gözlemler yaparak objektif ve standart ölçümler elde edilebilmesi için ölçme aracının geçerli olması gereklidir. Geçerlilik, ölçme aracının ölçülecek özelliği, tam ve doğru olarak, başka özelliklerle karıştırmadan ölçebilmesidir. Bir ölçek geliştirme sürecinde analitik olarak değerlendirilen ve belki de en önemli geçerlilik türü, (içsel) yapı geçerliliğidir. Ölçeklerin içsel yapı geçerliliklerinin değerlendirilmesinde, faktör analizi yaklaşımlarından “Açıklayıcı Faktör Analizi (AFA)” ve “Doğrulamalı Faktör Analizi (DFA)”, modern test teorisi yaklaşımlarından “Rasch Analizi” en yaygın kullanılan yaklaşımlar arasındadır. Çalışmanın amacı, romatoid artritli hastalar için “kendine bakım-hareket-ev işleri” alanlarında özürüllük düzeylerinin değerlendirilmesi için geliştirilen ölçme araçlarının yapı geçerliliğinin değerlendirilmesinde AFA, DFA ve Rasch Analizi uygulanması ve sonuçlarının incelenmesidir.

Yöntem: Toplamda 32 madde ile gerçekleştirilen AFA sonuçlarına göre, maddelerin iki boyutta toplandığı belirlenmiş, faktör yükleri ve klinik sınıflamaları da göz önünde bulundurulduğunda ilk boyut “Alt ekstremité ile ilişkili kendine bakım-hareket-ev işleri”, ikinci boyut “Üst ekstremité ile ilişkili kendine bakım-hareket-ev işleri” olarak adlandırılmıştır. Rasch Analizi uygulamasında, yukarıda bahsedilen iki boyut için ayrı ayrı ve klinik yapı göz ardı edilip her ikisinin birlikte ele alındığı durumlarda, ölçme araçları Rasch analizi aşamaları (modele uyum göstermeyen, yaş grubu, cinsiyet ve/veya hastalık süresi bakımından madde işlev farklılığı (MIF) gösteren ve tek boyutluluk varsayımını bozan maddelerin çıkarılması) dikkate alınarak değerlendirilmiştir.

Bulgular: AFA’dan elde edilen sonuçlar, klinisyen görüşüyle karşılaştırıldığında sadece bir maddede uyumsuzluk belirlenmiş; bu maddenin klinik yapının uygun gördüğü yerde değerlendirilmesine karar verilmiştir. Klinik olarak belirlenen yapı DFA ile test edildiğinde, uyum iyiliği istatistiklerine göre, elde edilen sonuçların klinik yapı ile uyumlu olduğu sonucuna ulaşılmıştır. Rasch analizi’nden elde edilen sonuçlarda, “alt ekstremité ile ilişkili kendine bakım-hareket-ev işleri” boyutu için 18 madde içerisinde modele uyumsuzluk gösteren maddeler ölçme aracından çıkarıldıktan sonra model uyumunun yeterli olduğu belirlenmiştir. “Üst ekstremité ile ilişkili kendine bakım-hareket-ev işleri” boyutu için 14 maddenin tümü modele uyum göstermiş; tek boyutluluk varsayımını bozan maddeler analizden çıkarılmış; yaş grubuna göre bir maddenin tekdüze MIF gösterdiği belirlenmiştir. Klinik yapı gözardı edilerek yapılan değerlendirmede toplam 32 madde analize dahil edilmiştir. Tek boyutluluk yapısının sağlanmadığı görülmüştür. Elde edilen iki boyut, klinik yapı ile benzerdir.

Sonuç: Ölçeklerde yapı geçerliliğinin incelenmesinde, ölçeğin sadece boyut yapısını inceleyen faktör analizi yöntemlerine kıyasla, içsel yapı geçerliliğinin madde uyumu, birey uyumu, madde yanlılığı gibi farklı özellikler bakımından da değerlendirildiği Rasch Analizi’nin beraber kullanılması önerilmektedir.

Anahtar Sözcükler: Açıklayıcı Faktör Analizi, Doğrulamalı Faktör Analizi, Ölçek, Rasch Analizi, Yapı Geçerliliği

İkiden Fazla Bağımsız Oran Karşılaştırılması

Soner YİĞİT, Mehmet MENDEŞ

*Çanakkale Onsekiz Mart Üniversitesi Ziraat Fakültesi, Zootekni Bölümü, Biyometri ve Genetik ABD-
17100-Çanakkale*

Özet

Amaç: Uygulamada araştırmacıların ilgilendikleri konulardan bir tanesi de binomiyal dağılım gösteren populasyonlardan elde edilmiş oranların karşılaştırılmasıdır. Bu amaçla Z-testinden oldukça sık yararlanılmaktadır. Ancak Z-testi sadece iki bağımsız oranın karşılaştırılmasında kullanılabilir. İkiden fazla oranın karşılaştırılmasında ANOM Tekniği ve Ki-Kare temelli oran testlerinden yararlanılabilir. Ancak söz konusu testlerin performanslarına ilişkin (1. Tip hata ve Testin gücü) herhangi bir çalışmaya rastlanılmamış olması bir eksiklik olarak değerlendirilmektedir. Bu noktadan hareketle yürütülen bu çalışmada k-tane bağımsız oranın karşılaştırılmasında kullanılan ANOM ve Ki-Kare temelli testlerin değişik deneme koşullarındaki performanslarının ortaya konulması amaçlanmıştır.

Yöntem: Bu amaçla Monte Carlo simulasyon tekniğinden yararlanılarak farklı grup sayısı ($k=3, 4$ ve 5), istenen olayın populasyondaki oluş ihtimali ($p=0.10, 0.25, 0.50$ ve 0.75), etki büyüklüğü ($0.20, 0.40$ ve 0.60) ve örnek genişliği ($n=10, 20, 50, 100$ ve 300) kombinasyonları dikkate alınarak Bernoulli dağılımı gösteren populasyonlardan tesadüf sayıları üretilmiştir.

Bulgular: Yapılan 100000 simulasyon denemesi sonucunda dikkate alınan deneme koşulları ne olursa olsun en güvenilir sonuçların ANOM testinin kullanılması durumunda elde edildiği görülmüştür. Genel olarak ANOM testi $n>10$ olduğunda %5 civarında 1. Tip hata olasılığı gerçekleştirirken, Ki-Kare temelli testin oldukça sapmalı sonuçlar verdiği görülmüştür. Benzer durum testin güç değerleri bakımından da geçerlidir.

Sonuçlar: ANOM Tekniği grafiksel bir yöntem olması nedeniyle sonuçların yorumlanması oldukça kolay olmaktadır. Gerek simulasyon sonuçları değerlendirildiğinde gerekse de sonuçların yorumlanmasında sağladığı kolaylık nedeniyle ikiden fazla bağımsız oranın karşılaştırılmasında ANOM Tekniğinin kullanılmasının uygun olduğu görülmüştür.

Anahtar Sözcükler: ANOM, Ki-Kare, Oran, 1. Tip hata, Testin gücü

Meta-Analizde Güç Değerlendirmesi

Nazlı TOTİK, Zeliha Nazan ALPARSLAN

Çukurova Üniversitesi Tıp Fakültesi Biyoistatistik Anabilim Dalı, Adana

nazlitotik.biostat@gmail.com

Giriş: Meta-analiz çalışmaları, birbirinden bağımsız çalışmalardan elde edilmesi mümkün farklı sonuçları bir araya getirmek, bu sonuçlardaki çeşitliliği açıklamak, daha güvenilir ve daha doğru sonuçlar elde etmek amacıyla kullanılırlar. Yapılan çalışmaların genellenebilirliği için kullanılan güç analizi, sıfır hipotezinin reddedildiği durumlarda kabul edilmemesi olasılığı olarak tanımlanır. İstatistiksel gücün hesaplanması, birçok araştırma tasarımında olduğu gibi meta-analiz çalışmalarının planlanmasında ve yorumlanmasında da önemlidir. Ancak literatürde meta-analiz çalışmaları için güç hesaplamalarına yönelik çalışma sayısı azdır. Bir meta-analiz çalışması için güç analizi yapıldığında, genel etkinin beklenen büyüklüğü, dahil edilen çalışma sayısı ve çalışmaların örnek büyüklüğü göz önüne alınarak araştırmacıya, istatistiksel açıdan anlamlı bir sonuç bulma ihtimali sağlanabilmekte aynı zamanda da bulunduğu sonucun gücünü değerlendirme imkânı verilebilmektedir.

Amaç: Meta-analizlerde sonuçların birleştirilmesi aşamasında kullanılan sabit etki ve rastgele etki istatistiksel modelleri altında, belirli bir etki büyüklüğünde kabul edilebilir (%80 ve üstü) güç için makalelerdeki minimum örnek büyüklüğünün (MMÖB) ve çalışma sayılarının ne olacağını belirleyerek gücün farklı durumlardaki değişimini analiz etmektir.

Yöntem: İlgilenilen meta-analizin çalışma sayılarının (k) 2, 5, 7, 10, 15, 18 ve etki büyüklüklerinin (d) 0.1-0.2, 0.20001-0.3, 0.30001-0.4, 0.40001-0.5 ve 0.50001-0.7 değerleri arasında ve hasta kontrol oranlarının 1:1, 1:2 ve 1:3 olması durumunda, 1000 tekrarlı simülasyon çalışması ile güç değerleri elde edilmiştir. Simülasyon senaryosuna heterojenliğin üç seviyede (0.33, 0.67 ve 1) eklenmesiyle güç değerlendirmeleri farklı heterojenlik durumlarında, farklı etki büyüklükleri, farklı çalışma sayıları ve farklı hasta kontrol oranları ile de yapılmıştır.

Bulgular ve Sonuç: Meta-analiz çalışmalarında, sabit etki ve rastgele etki modeli olmak üzere kullanılan iki yaklaşım altında yapılmış olan simülasyon çalışmaları ile farklı etki büyüklüğü, çalışma sayısı ve MMÖB ile güç hesaplamaları yapılmıştır. Bu sonuçların yorumlanması ile homojen ve heterojen yapı kıyaslanmasının yanı sıra incelenen senaryolarda en uygun güç çıkarsamaları elde edilmiş, sonuçlar uygulanabilirlik açısından da incelenmiştir. Meta analizlerin kanıt piramidindeki yeri göz önüne alındığında, güç değerlendirmelerinin de bilinçli olarak yapılması gereği açıktır.

Anahtar Sözcükler: Sabit etki modeli, rastgele etki modeli, meta-analizde güç.

Multipl Sklerozlu Hastaların El Yazılarının İstatistiksel Şekil Analizi

Fatma Ezgi CAN^{1,2}, İlker ERCAN³, Güven ÖZKAYA³, Özlem TAŞKAPILIOĞLU⁴

¹*Biyoistatistik ve Tıp Bilişimi Anabilim Dalı, Tıp Fakültesi, İzmir Katip Çelebi Üniversitesi, İzmir*

²*Biyoistatistik Anabilim Dalı, Sağlık Bilimleri Enstitüsü, Uludağ Üniversitesi, Bursa*

³*Biyoistatistik Anabilim Dalı, Tıp Fakültesi, Uludağ Üniversitesi, Bursa*

⁴*Nöroloji Anabilim Dalı, Tıp Fakültesi, Uludağ Üniversitesi, Bursa*

Özet

Amaç: İstatistiksel şekil analizi ile 5 yıldır takip edilen multipl skleroz hastalarının el yazısı karakteristiklerinin değişimi incelenmek istenmektedir.

Yöntem: Multipl skleroz hastalarının el yazı örnekleri çalışma başlangıcında, 6 ve 12. aylarda ve 5 yıl sonra olmak üzere dört farklı zaman diliminde alınmıştır. El yazı karakteristiklerine istatistiksel şekil analizi uygulanmıştır. İstatistiksel şekil analizi procrustes analizi yapılarak uygulanmış ve hastaların grafik üzerinde dağılımını değerlendirmek için de temel koordinat analizi uygulanmıştır.

Bulgular: Hasta gruplarına ait dağılım grafiklerinde her bir hastanın ilk ölçüm, 6. ay, 12. ay ve 5. yıldaki Expanded Disability Status Scale skorları dikkate alınarak, hastaların farklı yazı örnekleri incelendiğinde temel koordinat analizi sonucunda elde edilen grafiklerinde hastaların sistematik bir konumlanma göstermediği görülmüştür.

Sonuç: Çalışmamızda 5 yıl sonunda MS hastalarının ve kontrollerin farklı yazı formatlarında, örnek alınan zamanlara göre bir farklılık göstermediği görülmüştür.

Anahtar Sözcükler: El yazısı; İstatistiksel şekil analizi; Temel koordinat analizi

Meta Analizinde Aykırı Değerlerin Tespiti ve Farklı Varyans Tahmin Yöntemleri ile Değerlendirilmesi

Mutlu UMAROĞLU¹, Pınar ÖZDEMİR¹

¹Hacettepe Üniversitesi Tıp Fakültesi Biyoistatistik Anabilim Dalı

Özet

Meta analizi aynı konuda yapılmış benzer çalışmaların birleştirilip ortak bir sonuca varılması yöntemidir. Meta analizinde çalışmalar birleştirilirken genellikle çalışmaların etki büyüklükleri birleştirilir. Bu nedenle çalışmalar birleştirilmeden önce çalışmaların etki büyüklüklerinin hesaplanması ve etki büyüklüğünün dağılım yapısının incelenmesi gerekir. Etki büyüklüklerinin dağılım yapısındaki benzerlik çalışmaların homojen olduğunu, farklılık ise çalışmaların heterojen olduğunu göstermektedir. Bazı durumlarda ise bir yada birkaç çalışma diğer çalışmaların dağılım yapısından tamamen farklı olabilmektedir. Eğer bir çalışmaya ait etki büyüklüğü diğer çalışmalardan farklı olacak şekilde bir değer alıyorsa bu çalışma meta analizindeki aykırı değer olarak adlandırılmaktadır.

Bir meta analizinde aykırı değer olup olmadığı hakkında forest plot, radial plot, labbe plot gibi grafiksel yöntemlerle fikir sahibi olunabilir. Ancak aykırı değerlerin tespit edilmesi için artıkların incelenmesi gerekmektedir.

Bir meta analizinde aykırı değer olması etki büyüklüğünün dağılım yapısını olduğundan daha heterojen gösterme eğilimindedir ve heterojen yapıya sahip meta analizi çalışmalarında rasgele etki modeli kullanılması önerilmiştir. Rasgele etki modelleri içerisinde DerSimonian-Laird, en çok olabilirlik yöntemi, kısıtlanmış en çok olabilirlik yöntemi, Sidik-Jonkman yöntemi, deneysel Bayes yöntemi gibi farklı varyans tahmin yöntemleri bulunmaktadır.

Bu çalışmada aykırı değer bulunan meta analizleri bu beş yöntem açısından değerlendirilmiştir. Birleştirilmiş etki büyüklüğü DerSimonian-Laird ve kısıtlanmış en çok olabilirlik yönteminde aykırı değerden en az etkilenmiş deneysel Bayes yöntemi ise en çok etkilenmiştir. Güven aralığı deneysel Bayes yönteminde en dar çıkarken kısıtlanmış en çok olabilirlik yönteminde en geniş çıkmıştır. Çalışmalar arası varyans tahminleri (τ^2) karşılaştırıldığında DerSimonian-Laird yöntemi en düşük varyansa sahip çıkmıştır.

Anahtar Sözcükler: Meta-analizi, aykırı değer, heterojenlik, rasgele etki modeli

Kaynaklar:

1. W. Viechtbauer, M.W.L. Cheung, 2010, Wiley, Outlier and influence diagnostics for meta-analysis
2. F.N. Gumedze, D. Jackson, BMC, 2011, A random effects variance shift model for detecting and accommodating outliers in meta-analysis
3. L. Lin, H. Chu, J.S. Hodges, 2016, Alternative Measures of Between-Study Heterogeneity in Meta-Analysis: Reducing the Impact of Outlying Studies, Biometrics
4. J. Higgins, S.G. Thompson, J. Deeks, DG. Altman, 2003, Measuring inconsistency in meta-analyses, BMJ
5. Hedges LV and Olkin I (1985). Statistical Method for Meta-Analysis. Academic Press, Orlando, FL.
6. K. Sidik, K.N. Jonkman, Simple heterogeneity variance estimation for meta-analysis, Journal of the royal statistical society, 2005

Konvolüsyonel Sinir Ağları ile 3 Boyutlu Ortamda Oluşturulan Örüntülerin Metin Olarak Tanınması

Adem DOĞANER¹, Cemil ÇOLAK²

¹*Kahramanmaraş Sütçü İmam Üniv. Tıp Fakültesi, Biyoistatistik ve Tıbbi Bilişim AD, K.Maraş*

²*İnönü Üniversitesi, Tıp Fakültesi, Biyoistatistik ve Tıp Bilişimi AD, Malatya*

Amaç: Sinir ağları uzun yıllar boyunca sınıflandırma ve tahmin işlemlerinde başarılı sonuçlar üretse de, modelin öğrenmesini sağlayacak yeterli veri bulunamaması ve bilgisayar işlemcilerin performansının düşük olması nedeniyle büyük boyuttaki verilerde tahmin ve sınıflandırma başarısı belirli bir düzeyi aşamamıştır. Büyük veri yığınlarının hızla artması ve bilgisayar işlemcilerinin çok güçlü hale gelmesi ile derin öğrenme ve makine öğrenme sistemleri, tanıma ve sınıflandırma işlemlerinde büyük başarılar sağlamıştır. Derin öğrenme mimarilerinde örüntü tanıma işlevini Konvolüsyonel sinir ağları (CNN) yerine getirmektedir. İleri beslemeli bir ağ modeline sahip olan Konvolüsyonel sinir ağları görüntü tanımda oldukça başarılıdır. Bu çalışmada el ve parmak ucu hareketleri ile üç boyutlu ortamda yazılan el yazılarının sensörler ile takip edilerek bilgisayar ortamına aktarılması ve bu el yazılarının konvolüsyonel sinir ağları ile tanınması amaçlanmaktadır.

Yöntem: Çalışmada ilk olarak parmak ucu ile el yazıları yazılarak 3 boyutlu ortamda el ve parmak hareketlerini takip eden sensörler aracılığı ile el yazılarının bilgisayara işlenmesi sağlanmıştır. İkinci aşamada el yazılarının Konvolüsyonel sinir ağları ile tanınması gerçekleştirilmektedir. Konvolüsyonel sinir ağları görüntü boyutunu her katmanda daha az sayıda piksel içeren boyutlara dönüştürerek öznetelik sağlamaktadır. Konvolüsyon katmanları ve pooling katmanları ile görüntü içerisindeki daha küçük boyutlu görüntülere filtreler uygulanarak yeni matrisler elde edilmektedir. Bu şekilde her katmanda yeni öznetelikler elde edilerek öğrenme sağlanmaktadır. Öğrenmenin sağlanması aşamasında CNN için çok sayıda örneğe ihtiyaç duyulmaktadır. Çalışmada el yazısı ile oluşturulan harflerin tanınması için 28*28 boyutunda çok sayıda örneğin model tarafından öğrenilmesi sağlanmıştır. Bu işlemler için Python dili ve konvolüsyonel sinir ağlarını model alan derin öğrenme kütüphanelerinden yararlanılmıştır.

Bulgular: Uygulamada el ve parmak ucu hareketleri ile 3 boyutlu ortamdan elde edilen sinyaller bilgisayar ortamına aktararak 2 boyutlu el yazısı formlarına dönüştürülmüştür. Oluşturulan el yazıları daha önce eğitilen konvolüsyonel sinir ağları modeline tanıtılmış ve Python dili ve kütüphanelerinde işlenmiştir. Modelin ilk aşamalarında tanıma işlemlerinde başarılı sonuçlar elde edilmiştir.

Sonuç: El ve parmak ucu hareketleri ile oluşturulan örüntülerin 2 boyutlu ortamda el yazısı olarak tanımlanması ve konvolüsyonel sinir ağları ile el yazılarının tanımlanmasında model başarılı şekilde uygulanmış ve sonuçlar alınabilmektedir. Bu modelin ileriki aşamada işitme engelli bireylerin diğer insanlar ile daha rahat iletişim kurabilmesine olanak sağlaması öngörülmektedir.

Anahtar Sözcükler: Örüntü tanıma, Konvolüsyonel Sinir Ağları, Python Dili

Novel Method for Epidemiological Information of Dental Clinics by Spatial Regression

Adnan KARAIBRAHIMOGLU¹, Arslan TERLEMEZ²

¹Necmettin Erbakan University, Meram Medical Faculty, Medical Education and Informatics
Department, Biostatistics Unit, KONYA

²Necmettin Erbakan University, Faculty of Dentistry, Department of Endodontics, KONYA
adnankaraibrahim@gmail.com

Aim: Oral health, separated generally into two parts: self-care and clinical care, is very important and affects many aspects of our life. In recent years, the number of patients in dental clinics is increasing dramatically. The aim of this study is to find the effects of some demographic characteristics on dental clinics by spatial regression taking into account the districts in central Konya.

Methods: There were totally 113,411 electronic records and since we want to analyze the data in municipality level, data consisted of 16,054 records of patients. The variables were dental clinics, age, gender, year and month of the application and district names of residence. The noisy terms in dataset were cleaned at first due to some misediting data entry in the application procedure. Next, 180 district names were matched with longitude and latitude coordinate values of districts by related shapefile in a GIS media. A spatial multinomial logistic regression model was conducted considering the dental clinics as DV and other attributes as IV. There were eight clinics in the Faculty: oral and maxillofacial surgery, oral diagnosis and imaging, restorative dentistry, endodontics, orthodontics, prostodontics, pedodontics and periodontics & preventive dentistry. Since the number of patients for this clinic doesn't indicate the real spatial dependency, it was considered that the oral diagnosis & imaging clinic was taken as a reference category in the model. The code was run and density maps for each dental clinic were plotted.

Results: According to the results, $R^2=0.456$, $AIC=27606.7$ and Pearson GOF test was significant. Likelihood ratio test results were significant for age, gender and the month of application, but not for the year of application. Advanced age, male gender and frequently March and November months were significant for majority of clinics. Regions with higher income were prominently correlated with orthodontics and prostodontics. The districts having easy urban transport facility to the Faculty were associated with all clinics.

Conclusion: This study showed that the spatial analysis is useful if there is a spatial dependency and gives more detailed information. Categorical spatial data can be analyzed by semi- or non-parametric methods as well as continuous data is analyzed easily by spatial lag or spatial error models. It is obvious that epidemiological results of this study should be examined carefully by dentistry clinics to make more accurate decisions for oral health.

Key words: Epidemiology, dentistry, geographical information system (GIS), spatial multinomial logistics regression

Covariance Patterns for Linear Mixed Models

Hatice Tül Kübra AKDUR^{1,*}, Deniz ÖZONUR² and Hülya BAYRAK³

¹*Gazi University, Faculty of Sciences, Department of Statistics, Ankara, Turkey*

²*Gazi University, Faculty of Sciences, Department of Statistics, Ankara, Turkey*

³*Gazi University, Faculty of Sciences, Department of Statistics, Ankara, Turkey*

¹*hatice_senol@wsu.edu*, ²*denizozonur@gazi.edu.tr*, ³*hbayrak@gazi.edu.tr*

Objective: Linear mixed model is very useful tool for analyzing extensive variety of data structures which statisticians usually encounter. Despite this extensive applicability, the results of such analyses are not very frequent in the applied research literature. Even though commercial and open source soft wares are limited in the past, now many statistical soft wares such as R, SPSS, and SAS are allowed to analyze linear mixed models. **Material and Methods:** Longitudinal and repeated measures datasets require elaborative attention especially on the step of choosing the right covariance pattern. This types of datasets include correlated data because subjects contribute multiple observations to the dataset. To illustrate analyzing longitudinal data by linear mixed model, lung function growth data (Dockery et al., 1983) is taken into consideration. The dataset consists of children from six cities in US and is analyzed by lme4 R package. **Discussion:** Main aim of the study is to provide sufficient information for selecting accurate covariance pattern to practitioners of statistics before the analyzing the dataset. Furthermore, we attempted to demonstrate important features of linear mixed model and its analysis to applied statisticians. Some of the hybrid covariance patterns are introduced and the use of importance is discussed. We also give a detailed case study with longitudinal follow-up.

Keywords: Linear mixed models; longitudinal data, hybrid covariance patterns

References

Dockery, D. W., Berkey, C. S., Ware, J. H., Speizer, F. E., & Ferris Jr, B. G. (1983). Distribution of Forced Vital Capacity and Forced Expiratory Volume in One Second in Children 6 to 11 Years of Age 1–3. American Review of Respiratory Disease, 128(3), 405-412.

Dealing with the Unknown: Handling Missing Data in Longitudinal Studies

Oya KALAYCIOĞLU

Abant İzzet Baysal University, Department of Econometrics, Bolu

Abstract

Aim: Missing data is a common problem in medical research even in well-controlled situations, especially in longitudinal studies. A recent review of randomised controlled trials with longitudinal measurements found that the majority of the publications failed to make any attempt to explain the reasons for their choice of missing data handling method. Ad-hoc techniques such as complete case analysis, mean imputation, last observation carried forward, etc. have fallen out of favour in methodological literature in recent years, as they are subject to serious bias under most of the missing data settings. Alternatively, multiple imputation (MI) method which is based on the statistical theory is increasingly being used as a sophisticated tool to alleviate the problem of bias and efficiency of the parameter estimates, however using MI is not straightforward with longitudinal data especially when non-normally distributed variables are incomplete. The aim of this talk is to discuss various missing data handling methods for longitudinal studies and make some suggestions on how to choose the optimal missing data handling technique.

Methods: Different MI strategies that can be used with longitudinal data are compared qualitatively in terms of the imputation model that they use and the incomplete variable types that they can handle. In order to make matters concrete, a longitudinal case study where, data on both outcome and explanatory variables are collected repeatedly at several time points is considered which concerns the continuity of care among cancer patients. The regression coefficient estimates obtained with simple methods and MI techniques are compared. In addition, a simulation study has been carried out to compare the methods regarding the bias and efficiency of the regression coefficient estimates.

Results: Various missing data methods produced substantially different estimates of the regression coefficients and their standard errors. When the results of the simulations are considered, MI technique which uses multivariate normal imputation model produces the least bias in most situations. The only exception is the regression coefficient estimates of the incomplete categorical variables with extreme prevalences.

Conclusion: When missing data occurs in repeatedly measured variables, the choice of appropriate missing data handling methods depends on various factors such as the distribution of the incomplete variables, missingness pattern and the number of incomplete variables. The simulation studies showed that, these factors should be considered with care before applying any imputation method in order to improve estimation.

Keywords: Missing data, multiple imputation, longitudinal data, repeated measurements

Assessment and Comparison of Two Source Capture Recapture Methods' Performance

^{1,2}**Musa Bashir ALBISHIR**, ³Ilker ERCAN

¹*Department of Biostatistics, Uludag University Institute of Health Sciences, Bursa, Turkey*

²*Faculty of Science, Department of Statistics, Federal University Birninkebbi, Birnin Kebbi - Kebbi State, Nigeria*

³*Department of Biostatistics, Uludag University Faculty of Medicine, Bursa, Turkey*

Abstract

The aims of this study were to assess the performance of some Capture Recapture methods with simulation studies on data obtained from survey to estimate the population size of families with at least one disabled individual in Turkey with small, medium and large sample I and II, and to compare the performance of these estimators with different sample sizes by Mean Squared Error (MSE) measure.

From the comparison part of the mentioned capture recapture methods, the Lincoln - Petersen method provides more reliable results than Chapman and Bailey methods in small sample size I and II, but undefined when there were no element recaptured. In respect to big sample I and small sample II Chapman and Bailey methods provides better estimates than Lincoln - Petersen method. When working with big sample size I and II are all the three estimates provides better estimates with least error.

Keywords: Capture Recapture, Assessment and Comparison, Lincoln - Petersen method, Chapman method, Bailey method.

Exploring Mutation Profiles of Colon Adenocarcinoma to Infer Potential Ancestral History with Bayesian Networks

Uğur TOPRAK¹, Kemal TURHAN¹

¹Karadeniz Technical University, Medical Faculty, Department of Biostatistics and Medical Informatics

Introduction: Recent studies in genome sequencing have shown that the somatic mutations that drive cancer development are distributed across a large number of genes. This mutational heterogeneity complicates efforts to distinguish interactions between driver genes. It is known that mutations are the reason of cancer but it is still not clear that which mutation(s) are acting together and how they trigger each other.

Aim: This study aims to reveal some hidden mutation dependencies and order during the early stages of colon adenocarcinoma progression.

Methods: In this paper, we compared several methods to infer mutation relations and order underlying somatic mutations by utilizing evolutionary information. Genetic data gathered from GDC (Genomic Data Commons) data portal and important genes selected from OMIM (Online Mendelian Inheritance in Man) database. Ancestral history had been interpreted as a structure learning problem in the presence or absence of data. Unnecessary and redundant information was pruned from data with the association rules. Remaining data sorted into pairs in terms of to their prevalence. These pairs are scored according to Akaike Information Criterion, Bayesian Information Criterion, Bayesian Dirichlet equivalent, and K2 scoring methods. Bayesian network was built by using mutation data analyzed from colon adenocarcinoma patient samples which could be found in GDC. Associated mutations were evaluated and Hill Climbing and Tabu structure learning algorithms were applied on data set. Finally, a simulated test dataset generated from negative binomial distribution also known as a Poisson-Gamma mixture according literature to manually create ancestral trees. Produced dataset is used to build, compare and validate inferred network structure.

Results: Bayesian networks are formed by different local search and learning algorithms gave similar results according to the scoring methods. DAGs (Directed Acyclic Graphs) formed with AIC and BIC scores gave more linear results and generally one mutation does not effected from other mutations. However, BDe and K2 methods gave us more complex DAG structures and not only gene to gene but also genes to gene interactions were observed.

Conclusion: We anticipate that our method demonstrates that interactivity and evolutionary histories of mutations are crucial information for treatment of this disease. These techniques can be applied to different cancer types or genetic data to understand better about genetic material changing during life span of the living.

Keywords: Bayesian Networks, Mutation Dependency, Structural Learning

Chronological Age Classifier Enables Evaluation of Pro-Aging Factors at Single Cell Resolution

Sumeet Pal SINGH¹, Gökmen ZARARSIZ², Sharan Kaur JANJUHA^{1,3}, Nikolay NINOV^{1,3,4}

¹*DFG-Center for Regenerative Therapies Dresden, Cluster of Excellence, Technische Universität Dresden, Dresden, Germany*

²*Erciyes University, Faculty of Medicine, Dept. of Biostatistics, Melikgazi/Kayseri*

³*Paul Langerhans Institute Dresden of the Helmholtz Zentrum München at the University Hospital and Faculty of Medicine Carl Gustav Carus of Technische Universität Dresden, Dresden, Germany*

⁴*German Center for Diabetes Research (DZD e.V.), Neuherberg, Neuherberg, Germany.*

Abstract

Aim: Everyone grows old, but at different rates. Why does the rate of aging differ between individuals? The answer lies in identifying factors that trigger premature aging. We propose identifying pro-aging factors via an unbiased, data-driven approach that models aging of single beta-cells. The model, called GERAS* (Genetic Reference for Age of Single-cell), uses artificial intelligence (AI) to 'learn' transcriptional changes that cells naturally undergo with age. With this training, GERAS utilizes age-specific biomarkers for detecting premature cellular aging: young cells attaining the genetic signature of old cells.

* Geras was the Greek God of old age.

Method: Using zebrafish beta-cells, we developed a chronological age classifier that categorizes individual beta-cells solely from their genetic profile. To obtain the temporal dynamics of beta-cells, we profiled 460 zebrafish beta-cells from five ages: juvenile (1 month), young (3 and 6 months) and old (10 and 14 months). 70% of these cells were then randomly chosen to train an AI algorithm (three-layer neural network) to predict the cell's chronological age. The remaining 30% cells were used to test the accuracy of the model. An accuracy of 97.5% was obtained for our dataset, demonstrating that our AI model, GERAS, successfully categorizes individual beta-cells into the five ages.

Results: We utilized GERAS to evaluate the impact of pro-aging factors. High-calorie diet, a condition known to accelerate aging, was used as a model of accelerated aging. We separated 3 months-old zebrafish siblings into two groups. One group was put on low-calorie diet (single feeding on alternative days), while the other group was put on high-calorie diet (three times feeding per day). After one month of treatment, the beta-cells were isolated and the age of individual beta-cells evaluated using GERAS. The analysis showed a striking difference in age between the two set of beta-cells obtained from young zebrafish. While 90% of the beta-cells from low-calorie fed animals were classified as young, only 20% of the beta-cells from high-calorie fed animals were similarly classified; the rest 80% were categorized as old. This suggests that high-calorie diet shifts the transcriptional profile of young beta-cells towards an older population, causing premature aging.

Conclusion: Our study provides the first-ever age classifier for individual beta-cells. Additionally, it provides proof-of-principle of the utility of age classifier by evaluating the impact of high-calorie diet on the age of beta-cells. The project allows rapid evaluation of beta-cell aging, a key phenomenon underlying development of diabetes.

Keywords: Single-cell Sequencing, Aging, Neural Network

A Glimse at Long-read Sequencing and Megabase-sized Scaffolding Approaches

Vahap ELDEM¹, Gökmen ZARARSIZ^{2,3}, Selçuk KORKMAZ^{3,4}, Dinçer GÖKSÜLÜK^{3,5},
Gözde ERTÜRK ZARARSIZ², Halil BİLGİN^{3,6}, Ahmet ÖZTÜRK^{2,3}

¹ *İstanbul Üniversitesi, Fen Fakültesi, Biyoloji Bölümü, Zooloji Anabilim Dalı, İstanbul*

² *Erciyes Üniversitesi, Tıp Fakültesi, Biyoistatistik ve Tıbbi Bilişim Anabilim Dalı, Kayseri*

³ *Turcosa Analitik Çözümler Ltd Şti, Erciyes Teknopark, Kayseri*

⁴ *Trakya Üniversitesi, Tıp Fakültesi, Biyoistatistik ve Tıbbi Bilişim Anabilim Dalı, Edirne*

⁵ *Hacettepe Üniversitesi, Tıp Fakültesi, Biyoistatistik Anabilim Dalı, Ankara*

⁶ *Abdullah Gül Üniversitesi, Mühendislik Fakültesi, Bilgisayar Mühendisliği, Kayseri*

Background: Generating high-quality finished genomes is considered as an essential prerequisite for understanding genome function, organization and evolution. Despite a significant progress in genome researches achieved over the last decades using a variety of second-generation platforms, they have often failed in resolving the repeat structure of highly-complex repeat genome, structural variation detection, complex genomic rearrangements, haplotype phasing and genome assembly contiguity due to mainly their short read length (<300 bp). With advent of long-read Sequencing or “third-generation DNA sequencing technologies”, above-mentioned issues in genome analysis have been partly ameliorated by increasing average read lengths up to 10,000-100,000 bp. Moreover, third-generation mapping technologies such as; optical mapping, “Hi-C” chromatin interaction and linked-reads systems are recently introduced and seems to pave the way for chromosome-scale assembly. In this study, we aim to compare and assess the effectiveness of long-read sequencing and scaffolding approaches on diploid and complex genome analysis.

Method: We systematically investigated the influence of single-molecule real-time sequencing (PacBio SMRT) or nanopore-based (Oxford Nanopore, MinION, GridION) sequencing approaches on previously sequenced genome assemblies using 3C method (Correctness, Continuity and Completeness) with the help of basic genome assembly metrics and the lineage-specific sets of Benchmarking Universal Single-Copy Orthologs analysis. We also surveyed three mapping technologies; optical mapping system (BioNano Genomics), “Hi-C” chromatin interaction system (Dovetail Genomics) and linked-reads system (10X Genomics) to see whether they contribute significantly on *de novo* genome assembly by measuring sequence based basic assembly metrics such as N50, NG50, L50, contig and scaffold numbers, largest scaffold lengths as well as sequence alignment success ratio.

Results: It is clear that the long-read informations at least 80-100X genome depth can considerably improve the continuity of draft genome assemblies at least 100 fold by spanning tandem repeats. In addition, long reads enable to discover numerous misassemblies, missing sequences in gaps and repeat structure errors in draft genome assemblies. The main contributions of state-of-art mapping methods (10X Genomics, Dovetail and Bionano Genomics) scaffolding approaches regardless of technology used are (i) the correct orientation of scaffolds and enhancing continuity, (ii) reducing the number of contigs in initial assemblies by closing gaps, enabling full chromosome-length scaffolding.

Conclusion: We prospected that long-read sequencing technologies and true full chromosome-length scaffolding procedure will replace over NGS-based approaches alone and be accepted as standard methods for *de novo* genome studies.

Keywords: Long read sequencing, state-of-art scaffolding approaches, *de novo*, genome assembly.

MLSeq 2.0: Machine Learning Interface for RNA Sequencing Data

Gökmen ZARARSIZ^{1,2}, Dinçer GÖKSÜLÜK^{2,3}, Selçuk KORKMAZ^{2,4}, Vahap ELDEM⁵,
Bernd KLAUS⁶, Ahmet ÖZTÜRK^{1,2}

¹ *Erciyes University, Faculty of Medicine, Department of Biostatistics, Kayseri, Turkey*

² *Turcosa Analytics Solutions Ltd. Co., Erciyes Teknopark, Kayseri, Turkey*

³ *Hacettepe University, Faculty of Medicine, Department of Biostatistics, Ankara, Turkey*

⁴ *Trakya University, Faculty of Medicine, Department of Biostatistics, Edirne, Turkey*

⁵ *İstanbul University, Faculty of Science, Department of Zoology, İstanbul, Turkey*

⁶ *European Molecular Biology Laboratory (EMBL), Heidelberg, Germany*

Abstract

During the last half decade, MLSeq R/BIOCONDUCTOR package has become the de facto tool for application of machine-learning algorithms in classification of next-generation RNA-Sequencing (RNA-Seq) data. Researchers have appealed to MLSeq for various purposes, which include prediction of disease outcomes, identification of best subset of features (genes, transcripts, other isoforms), and sorting the features based on their predictive importances. Along with the innovations in the RNA-Seq based machine-learning literature, we now release MLSeq 2.0, the second version of the software. Previous version of the software included five machine-learning algorithms as well as several preprocessing approaches. This version of this software has come with significant updates. Using this package, researchers can upload their raw RNA-seq count data, preprocess their data and perform a wide range of machine-learning algorithms. Preprocessing approaches include deseq median ratio and trimmed mean of M means (TMM) normalization methods, as well as the logarithm of counts per million reads (log-cpm), variance stabilizing transformation (vst), regularized logarithmic transformation (rlog) and variance modeling at observational level (voom) transformation approaches. Normalization approaches can be used to correct systematic variations. Transformation approaches can be used to bring discrete RNA-seq data hierarchically closer to microarrays and conduct microarray-based classification algorithms. Currently, MLSeq package contains 90 microarray-based classifiers including the recently developed voom-based discriminant analysis classifiers. Besides these classifiers, MLSeq package also includes discrete-based classifiers, such as Poisson linear discriminant analysis (PLDA) and negative binomial linear discriminant analysis (NBLDA). Over the preprocessed data, researchers can build classification models, apply parameter optimization on these models, evaluate the model performances and compare the performances of different classification models. Moreover, the class labels of test samples can be predicted with the built models. MLSeq is a user friendly, simple and currently the most comprehensive package developed in the literature for RNA-Seq classification. Researchers can access this tool through <https://github.com/gokmenzararsiz/MLSeq/>.

Keywords: Classification, genomics, machine-learning, next-generation sequencing, RNA-Sequencing, voom

Machine Learning Based Virtual Screening in Drug Discovery

Selçuk KORKMAZ¹, Gökmen ZARARSIZ², Dinçer GÖKSÜLÜK³, Necdet SÜT¹

¹Trakya University, Faculty of Medicine, Department of Biostatistics and Medical Informatics, Edirne

¹Erciyes University, Faculty of Medicine, Department of Biostatistics, Kayseri

¹Hacettepe University, Faculty of Medicine, Department of Biostatistics, Ankara

Abstract

Drug discovery and development studies are increasingly becoming time consuming, costly and hard. This process can take more than 10 years and cost over one billion dollars. Virtual screening is an important step in early-phase of drug discovery process. Since there are thousands of compounds, this step should be both fast and effective in order to distinguish drug-like and nondrug-like molecules. Statistical machine learning methods are widely used in drug discovery studies for classification purpose. We aimed to develop a web-based tool, which can classify molecules as drug-like and nondrug-like based on various machine learning methods, including discriminant, tree-based, kernel-based, ensemble and other algorithms. To construct this tool, first, performances of twenty-three different machine learning algorithms are compared by ten different measures, then, ten best performing algorithms have been selected based on principal component and hierarchical cluster analysis results. Besides classification, this application has also ability to create heat map and dendrogram for visual inspection of the molecules through hierarchical cluster analysis. Moreover, users can connect the PubChem database to download molecular information and to create two-dimensional structures of compounds. This application is freely available through www.biosoft.hacettepe.edu.tr/MLViS/. In recent years, deep learning algorithms show successful performance in many applications, therefore they have started to be used in the classification of drug molecules. As a further research, a deep neural network algorithm with high predictive performance will be developed with a high dimensional training data obtained from the PubChem database. The performance of deep learning algorithm will be compared with Random Forests and Support Vector Machines. In addition, a web-based application which contain a deep learning algorithm will be developed and offered to the researchers in this field.

Keywords: Machine Learning, Deep Learning, Virtual Screening, Drug Discovery

Improved Estimation of Body Fat Percentage via Machine-Learning Approaches

Halil BİLGİN^{1,2}, Gökmen ZARARSIZ^{2,3}, Volkan ÖZKAYA⁴, Betül ÇİÇEK⁴, Gözde ERTÜRK ZARARSIZ³, Selçuk KORKMAZ^{2,5}, Dinçer GÖKSÜLÜK^{2,6}, Ahmet ÖZTÜRK^{2,3}

¹ *Abdullah Gul University, Engineering Faculty, Department of Computer Engineering, Kayseri, Turkey*

² *Turcosa Analytics Solutions Ltd. Co., Erciyes Teknopark, Kayseri, Turkey*

³ *Erciyes University, Faculty of Medicine, Department of Biostatistics, Kayseri, Turkey*

⁴ *Erciyes University, Faculty of Health Sciences, Department of Nutrition and Dietetics, Kayseri, Turkey*

⁵ *Trakya University, Faculty of Medicine, Department of Biostatistics, Edirne, Turkey*

⁶ *Hacettepe University, Faculty of Medicine, Department of Biostatistics, Ankara, Turkey*

bilginhalil@gmail.com

Abstract

Aim: Obesity is among the most common risk factors for cancer, hypertension, cardiovascular disease and diabetes. Body fat percentage (BF%) is a commonly used measure to diagnose obesity. Various methods are present to predict the BFP, which includes dual energy X-ray absorptiometry (DEXA), bioelectric impedance analysis (BIA), anthropometric measurements such as skinfold thickness and circumferences. DEXA is an accurate and precise method for predicting BF% and reported as the gold standard method. However, its cost, portability, radiation exposure and convenience limits the use of this method. BIA is less expensive than DEXA, but its cost is still high and this method is not convenient for field studies. Anthropometric measures are simple, however combining them in an accurate way is another challenging problem. Most of the combined methods are based on the ordinary least squares regression method. To provide a simple and accurate prediction of BF%, we designed a comprehensive study and applied numerous machine-learning methods in order to find the optimal prediction model. We compared our results with BIA and equations computed from the anthropometric measurements.

Methods: We collected the data of 131 healthy subjects. Each subject has the data of DEXA and BIA measurements, as well as the anthropometric measurements such as waist and neck circumference, height, weight and skinfold thicknesses obtained with calipers at various skinfold sites on the body. The dataset is split into training (70%) and test (30%) sets. All model building process is applied in training set, while the performances of each model are assessed in test sets. At first, for each gender, we performed forward, backward, genetic algorithm and simulated annealing algorithms for feature selection. Next, 40 different machine-learning models are built. These models include least-square based models, additive models, regression trees, shrinkage methods, Bayesian methods, kernel-based methods and ensemble methods. We applied 5 fold cross-validation method with 10 repeats to identify the optimal parameters of each model. All model building process is repeated 50 times to generalize the results. Root mean squared error, mean absolute deviation, mean absolute percentage error, the coefficient of determination and Spearman's correlation coefficient are used as evaluation metrics. Hierarchical clustering analysis is applied to identify the most accurate models.

Results: The best-performed algorithms for both male and female groups were support vector machines with the linear kernel, Bayesian regularized neural networks and Gaussian process regression with the linear kernel. Obtained successful machine learning models were found to be more accurate than BIA and other equations according to the performance metrics used.

Conclusion: The generalized performance of 40 different ML algorithms, BIA and equations based on anthropometric measures in the literature for BFP prediction were compared. It is found that ML algorithms were more accurate than the current methods for BFP prediction. A web application for successful ML models was created and it is publicly available from <http://www.opensoft.turcosa.com.tr/BodyFatML>

Keywords: Bioelectric impedance analysis, body fat percentage, dual-energy X-ray absorptiometry, machine-learning, obesity, regression,

One-Way Tests in Independent Groups Designs: the onewaytests Web Interface

Osman DAG¹, Anıl DOLGUN² and N. Meric KONAR³

¹ *Department of Biostatistics, Hacettepe University, Ankara, Turkey;*

² *School of Science, RMIT University, Melbourne, Australia.*

Abstract: One-way tests in independent groups designs are the most commonly utilized statistical methods with applications on the experiments in medical sciences, pharmaceutical research, agriculture, biology, chemistry, engineering, and social sciences. In this study, especially for non-R users, the onewaytests web interface is introduced to investigate treatment effects on the dependent variable. This application is available at <http://www.biosoft.hacettepe.edu.tr/onewaytests>.

Objective: In this study, our primary focus was to develop a shiny-based web interface where any user can apply the aforementioned tests to their own datasets without R.

Methodology: The onewaytests web interface implements Welch's heteroscedastic F test with trimmed means and Winsorized variances, Alexander-Govern test and James second order test, in addition to the ANOVA, Welch's heteroscedastic F test, Kruskal-Wallis test, and Brown-Forsythe test. Besides the omnibus tests, the tool provides pairwise comparisons to investigate which groups create the difference providing that a statistically significant difference is obtained by the omnibus test. Moreover, the tool offers several graphic types such as grouped box-and-whisker plot and error bar graph. Also, it provides statistical tests and plots (i.e. Q-Q plot and histogram) to assess variance homogeneity and normality.

Implementation of the onewaytests web interface: This onewaytests web interface provides the users with one-way tests in independent groups designs, pairwise comparisons, graphical approaches, and assess variance homogeneity and normality of data in each group via tests and plots. At times, it is difficult for new R users or applied researchers to deal with R codes. Thus, we have developed a web interface of onewaytests package (Dag et al., 2017) by using shiny (Chang et al., 2017).

Users can upload their data to the tool via Data upload tab (Figure 1a). There exist two demo datasets on this tab for the users to test the tool. Basic descriptive statistics can be obtained through Describe data tab (Figure 1b). Moreover, assumptions of variance homogeneity and normality can be checked via this tab to decide on which one-way test is appropriate to test the statistical difference. Variance homogeneity is checked by variance homogeneity tests (Levene's test, Bartlett's test, Fligner-Killeen test). The normality of data in each group is assessed by normality tests (Shapiro-Wilk, Cramer-von Mises, Lilliefors (Kolmogorov-Smirnov), Shapiro-Francia, Anderson-Darling, Pearson Chi-Square tests) and plots (Q-Q plot and Histogram with normal curve). After describing the data, users can test statistical difference between groups through One-way tests tab (Figure 1c). In this tab, there exist one-way tests (ANOVA, Welch's heteroscedastic F, Welch's heteroscedastic F test with trimmed means and Winsorized variances, Brown-Forsythe, Alexander-Govern, James second order test, Kruskal-Wallis tests) and pairwise comparison methods (Bonferroni, Holm, Hochberg, Hommel, Benjamini-Hochberg, Benjamini-Yekutieli, no corrections). Moreover, users can obtain the graphics of given groups through Graphics tab (Figure 1d).

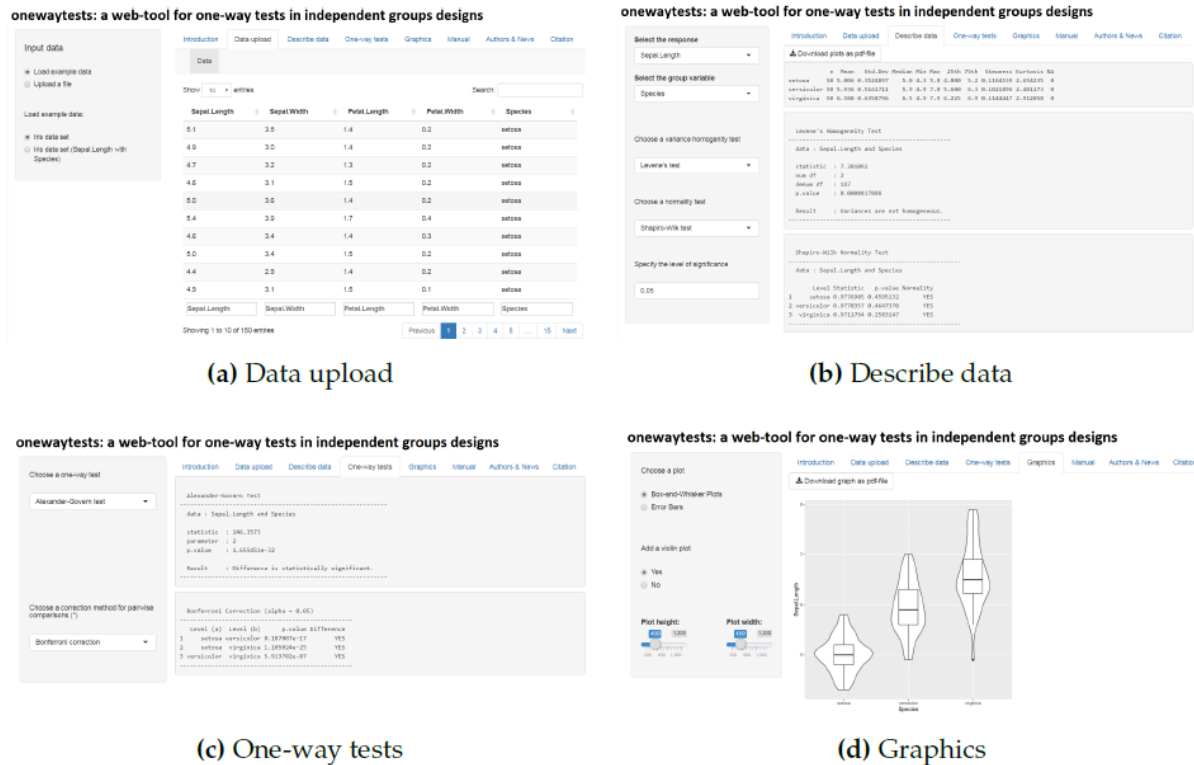


Figure 1: The onewaytests web interface

Conclusion: In this study, a shiny-based web interface was developed for any non-R users to apply the aforementioned tests to their own datasets without R. The detailed information and the R codes running in the background of the tool can be found in Dag et al. (2017).

Keywords: Web Tool, ANOVA, Alexander-Govern Test.

References

- Chang, W., Cheng, J., Allaire, J., Xie, Y., McPherson, J. (2017). shiny: Web Application Framework for R. R package version 1.0.1.
- Dag, O., Dolgun, A., Konar, N.M. (2017). onewaytests: One-Way Tests in Independent Groups Designs. R package version 1.4.

Evaluating Time-Dependent Cut-Off Values For Longitudinal Data

Naime Meric KONAR¹, Eda KARAİSMAİLOGLU², A. Ergun KARAAGAOGLU¹

¹:Hacettepe University, Faculty of Medicine, Department of Biostatistics, Ankara

²: Kastamonu University, Faculty of Medicine, Department of Biostatistics, Kastamonu

nmeric.konar@hacettepe.edu.tr

Aim: C-Reactive Protein (CRP) and Procalcitonin (PCT) are most commonly used biomarkers in intensive care units. Both of the biomarkers are used in diagnosis and treatment of several types of diseases including infectious disorders, cardiovascular diseases and so on. The aim of this study is to determine and evaluate the time-dependent cut-off values of serial CRP and PCT measurements in discriminating event and event-free individuals for 1, 2 and 3 days after the follow-up period.

Method: Joint modeling approach is used to model serial biomarker measurements (CRP and PCT) and survival times simultaneously. Data were collected from Hacettepe University Hospitals Adult Intensive Care Units retrospectively. Serial CRP measurements of 457 individuals and serial PCT measurements of 534 individuals were recorded along with demographic characteristics such as age and gender. The follow – up period is taken as the first 30 days in any intensive care unit and the event is defined as death in Intensive Care Units (ICU mortality). Different datasets were created for women and men for both CRP and PCT. For each dataset, sub-groups were formed for each internal care unit. For each sub - group, time – dependent ROC Curves are used for evaluating the diagnostic performance and time-dependent cut-off values are utilized to distinguish the event and event-free individuals for 1, 2 and 3 days after the the follow-up period.

Results: Results revealed that time-dependent diagnostic performance is better in men than women for both CRP and PCT biomarkers. On the other hand time-dependent cut-off values

Table1. Time-dependent AUC values and time-dependent cut-off values for women for CRP and for men for PCT

	Intensive Care Unit	Time - Dependent AUC			Intensive Care Unit	Time - Dependent Cut - Off Values		
		$\Delta t = 1$	$\Delta t = 2$	$\Delta t = 3$		$\Delta t = 1$	$\Delta t = 2$	$\Delta t = 3$
CRP for Women (n=223)	Emergency	0,564	0,562	0,561	Emergency	3,396	3,396	3,396
	Brain Surgery	0,56	0,558	0,557	Brain Surgery	3,643	3,643	3,581
	General Surgery	0,567	0,566	0,565	General Surgery	4,814	4,752	4,752
	Internal Diseases	0,549	0,548	0,546	Internal Diseases	2,842	2,842	2,842
	Cardiology	0,605	0,602	0,599	Cardiology	3,704	3,704	3,643
	Neurology	0,558	0,556	0,556	Neurology	2,472	2,472	2,472
	Intensive Care Unit	Time - Dependent AUC			Intensive Care Unit	Time - Dependent Cut - Off Values		
		$\Delta t = 1$	$\Delta t = 2$	$\Delta t = 3$		$\Delta t = 1$	$\Delta t = 2$	$\Delta t = 3$
PCT for Men (n=282)	Emergency	0,718	0,713	0,709	Emergency	0,64	0,586	0,586
	Brain Surgery	0,724	0,718	0,712	Brain Surgery	0,316	0,262	0,208
	General Surgery	0,719	0,714	0,709	General Surgery	1,665	1,611	1,557
	Internal Diseases	0,637	0,635	0,634	Internal Diseases	-0,115	-0,115	-0,115
	Cardiology	0,718	0,712	0,707	Cardiology	0,694	0,64	0,586
	Neurology	0,626	0,625	0,624	Neurology	-1,032	-1,087	-1,087

seems to decrease almost each unit in the 3rd day after the last measurement for both biomarkers. Brief results are given as an example for CRP biomarker for women and PCT biomarker for men:

Conclusions: Changes in diagnostic performance show that PCT is better biomarker than CRP for each intensive care unit in terms of predicting ICU mortality. On the other hand it's revealed that cut-off values either decrease or do not change through the 3rd day after the follow-up for CRP and PCT. It shows that the risk of death in ICU seems to decrease after the end of follow-up period for any intensive care unit.

Keywords: time-dependent ROC Curve, cut-off values, longitudinal data

References

1. Rizopoulos, D. (2011). "Dynamic Predictions and Prospective Accuracy in Joint Models for Longitudinal and Time-to-Event Data." *Biometrics* 67, 819–829.
2. Rizopoulos, D. (2010). "JM: An R Package for the Joint Modelling of Longitudinal and Time-to-Event Data." *Journal of Statistical Software*, 35:9.

Measuring Degree of Distinguishability and Agreement in Ordinal Rating Scales

Ayfer Ezgi YILMAZ, Serpil AKTAS

Department of Statistics, Hacettepe University Faculty of Science, Ankara-Turkey

Aim: When the categories of a square contingency table is ordinal, weighted kappa or Gwet's AC2 coefficients are used to summarize the degree of agreement between two raters. In addition to investigate the agreement between raters, the term of category distinguishability should be considered. The aim of this study is to assess the agreement coefficients and category distinguishability in ordinal rating scales together.

Methods: Degree of distinguishability, average degree of distinguishability, weighted kappa, and AC2 coefficients are applied to carcinoma in situ of the uterine cervix data. The results are discussed over the pathologist pairs.

Results: AC2 coefficient indicates that the agreement is higher than weighted kappa coefficient. Incorrect classification affects the level of the agreement in this respect. In that case, it will be beneficial to use the degree of distinguishability to detect if the categories are distinguishable or not.

Conclusion: It is proposed to use agreement coefficients and degree of distinguishability simultaneously.

Keywords: ordinal rating scales; agreement; degree of distinguishability; weighted kappa; Gwet's AC2 coefficient.

Çok Değişkenli Varyans Analizi Çözümlemesi için Normal Olmayan ve Dengesiz Örneklemelerin Test İstatistiklerinin I. Tip Hata Oranları Bakımından Karşılaştırılması

Mustafa Agah TEKİNDAL¹, Özlem KAYMAZ², H. Emre KALE³, Can ATEŞ⁴

² Ankara Üniversitesi, İstatistik Bölümü Uygulamalı İstatistik ABD, Ankara

³ Ankara Üniversitesi, Beyin Araştırmaları Merkezi, Disiplinlerarası Sinirbilimi ABD, Ankara

⁴ Van Yüzüncü Yıl Üniversitesi, Tıp Fakültesi Biyoistatistik ABD, Van

Amaç: Varyansın çok değişkenli analizi, aynı zamanda MANOVA (Multivariate Analysis of Variance) olarak da bilinen iki veya daha fazla bağımlı değişkenin bulunduğu verileri analiz etmek için kullanılan bir yöntemdir. Tüm parametrik yöntemlerde olduğu gibi MANOVA yönteminde hataların normal dağılması, varyansların homojen olması ve hata terimlerinin birbirinden bağımsız olması varsayımları vardır. Ancak uygulamada normal olmayan dağılımlar, normal dağılıma göre daha yaygındır. Bu çalışmada MANOVA yönteminin normallik ve homojen varyans varsayımlarının sağlanmadığı durumlarda Wilks Lamda istatistiği (Wilks' Lambda), Pillai iz istatistiği (Pillai's Trace), Hotelling iz istatistiği (Hotelling's Trace), Roy'un en büyük özdeğer (Roy's Largest Root) test istatistiklerinin nasıl etkileneceği araştırılmıştır. Başka bir deyişle bu durumlarda testlerin güvenilir olup olmadığı (dayanıklılığı) incelenmiştir.

Yöntem: Bu amaçla, farklı senaryolarda simülasyon çalışması yapılmıştır. Farklı grup, değişken sayıları ve farklı örneklem hacimlerinde, grup varyanslarının sabit ($\sigma_1^2 = \sigma_2^2 = \dots = \sigma_g^2$) ve artan ($\sigma_1^2 < \sigma_2^2 < \dots < \sigma_g^2$) olduğu durumları göz önünde bulundurularak Gamma(4-4-4; 0.5), Gamma(4-9-36;0.5), Student's t(2) ve Normal(0;1) dağılımlardan rasgele sayılar üretilmiştir. Ayrıca gruptaki gözlem sayılarının dengeli ve dengesiz olduğu durumlarda dikkate alınmıştır.. Simülasyon çalışmalarında 10000 tekrar yapılarak $\alpha = 0.05$ iken her bir test için birinci tip hata değerleri hesaplanmıştır. Simülasyon çalışması RStudio program dili (Mass (Modern Applied Statistics with S'-2017.04.21) ve lestat (A package for LEarning STATistics-20.02.2015) paketi kullanılarak yapılmıştır.

Bulgular: Simülasyon çalışması sonuçlarına göre, Gamma Dağılımında sabit varyans ve artan varyans durumunda 2 değişkenli dengeli ve dengesiz örneklemelerde Pillai iz test istatistiğinin I. Tip hata oranlarının $\alpha = 0.05$ değerine en yakın sonuçlar verdiği görülmüştür. 3 değişkenli olduğu durumda ise en yakın sonuçları dengeli örneklemelerde Roy'un en büyük özdeğer test istatistiği, dengesiz örneklemelerde Pillai iz test istatistiği göstermiştir. Student's t dağılımında ise belirlenen senaryolarda test istatistiklerinin I. Tip hata oranlarının $\alpha = 0.05$ değerine oldukça uzak sonuçlar verdiği görülmüştür. Normal dağılımın sabit varyans durumunda, 2 değişken için dengeli ve dengesiz örneklemelerde en yakın sonuçlar Roy'un en büyük özdeğer test istatistiği, 3 değişkenli dengeli ve dengesiz örneklemelerde ise en yakın sonuçlar Wilks Lamda test istatistiği göstermiştir. Normal dağılımın artan varyans durumunda ise 2 ve 3 değişkenli dengeli ve dengesiz örneklemelerde en yakın sonuçları Roy'un en büyük özdeğer test istatistiği göstermiştir.

Sonuç: Çalışmada belirlenen senaryolar sonucunda Gamma Dağılımının Student's t dağılımına göre daha tutarlı sonuçlar verdiği görülmüştür. Normal dağılım ile karşılaştırıldığında ise yine Gamma dağılımı Student's t dağılımına göre daha benzer sonuçlar vermiştir. Bu çalışma, farklı dağılım ve parametreler ile simülasyon çalışması yapılarak genişletilebilir.

Anahtar Sözcükler: MANOVA, Wilks' Lambda, Pillai's Trace, Hotelling's Trace, I.tip hata, Robustness

Veri Madenciliği Sınıflama Yöntemlerinin Acil Servis Verilerinde Uygulaması

Özlem ARIK¹, Erdem KARABULUT², Emine KADIOĞLU³

¹Dumlupınar Üniversitesi, Tıp Fakültesi, Biyoistatistik Anabilim Dalı, KÜTAHYA

²Hacettepe Üniversitesi, Tıp Fakültesi, Biyoistatistik Anabilim Dalı, ANKARA

³Dumlupınar Üniversitesi, Tıp Fakültesi, Acil Tıp Anabilim Dalı, KÜTAHYA

Amaç: Veri madenciliği; çok büyük miktardaki karmaşık ve düzensiz veriler içinden karar verme ve belli bir amacı gerçekleştirmek için uygun modellerin ortaya çıkarılmasıdır. Sınıflama, mevcut verilerin belirlenen özelliklere göre sınıflandırılması ve yeni bir veri eklendiğinde bu verinin sınıfının tahmin edilmesi işlemidir. Bu çalışmada temel amaç, Dumlupınar Üniversitesi Kütahya Evliya Çelebi Eğitim ve Araştırma Hastanesinin acil servis birimine başvuran 444 hastaya ait bazı bilgiler kullanılarak hastanın gerçekten acil durum gerektirip gerektirmediği yönünde bir sınıflama yapmak ve sınıflama yöntemlerinin performansını ölçmektir.

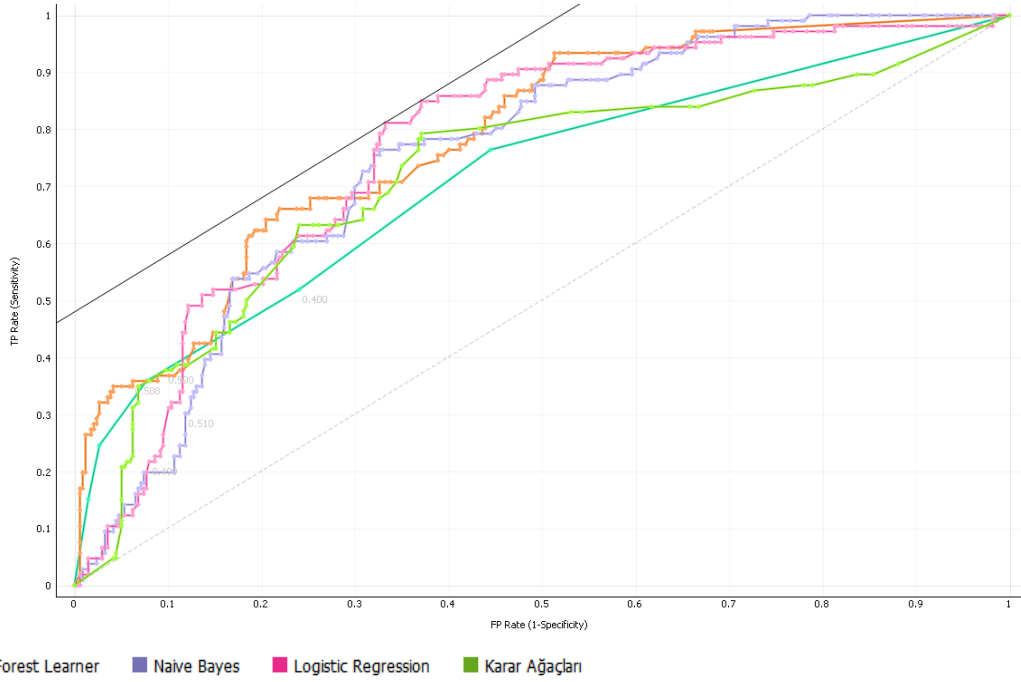
Yöntem: Bu çalışmada, acil servise gelen 444 hastaya ait özgeçmiş, fizik muayene, ekg, wbc, nötrofil, hemoglobin, glukoz, üre, ast, potasyum, troponin, inr, tit lökosit, akciğer grafisi, ek görüntüleme, konsültasyon, tedavi, kalış süresi ve triaj sınıfı bilgileri kullanıldı. Veri ile ilgili ön işlemeden sonra veri madenciliği sınıflama yöntemlerinden karar ağaçları, random forest, örnek tabanlı yöntemler: k en-yakın komşu, bayes sınıflandırıcı: naive-bayes ve lojistik regresyon kullanıldı ve doğru sınıflama oranı, duyarlılık, seçicilik, F1 ölçüsü, pozitif kestirim değeri, Mathews korelasyon katsayısı (MKK) ve ROC eğrisi altında kalan alan ölçüleri aracılığıyla yöntemlerin performansları belirlendi.

Bulgular: Acil servis verisi için kırmızı ve diğerlerine (sarı, yeşil) göre hedef sınıf kırmızı olacak şekilde sınıflama modellerine ait elde edilen bulgular;

Tablo 1: Performans Ölçüleri

Sınıflama Modeli	Doğru Sınıflama Oranı	Duyarlılık	Seçicilik	F1	Pozitif Kestirim Değeri	MKK
kNN	0.79	0.36	0.92	0.45	0.60	0.34
Random Forest	0.80	0.36	0.93	0.46	0.62	0.36
Naive Bayes	0.74	0.30	0.88	0.40	0.44	0.21
Lojistik Regresyon	0.75	0.20	0.92	0.27	0.45	0.17
Karar Ağaçları	0.78	0.38	0.90	0.45	0.55	0.32

Sınıflama yöntemlerine ait ROC eğrileri de Şekil 1`de verilmiştir.



Şekil 1: ROC Eğrileri

Sonuç: Orange 3.4 programı aracılığıyla yapılan çalışmada, Tablo 1’de verilen değerlere göre göre en iyi sınıflama yöntemi Random Forest’tır. kNN ve Karar Ağaçları yöntemleri de Naive-Bayes ve Lojistik Regresyon yöntemlerine göre daha iyi performansa sahiptir. Şekil 1 incelendiğinde de hastaları acil durumlarına göre sınıflara ayırmak için kullanılan yöntemler içerisinde Random Forest modelinin diğerlerine göre daha iyi olduğu sonucuna varılmıştır.

Anahtar sözcükler: Veri Madenciliği, Sınıflama, Acil Servis

Kaynaklar

Hacettepe Üniversitesi, BİS786-Veri Madenciliği Dersi Notları

<http://yunus.hacettepe.edu.tr/~hcingi/ist376a/6Bolum.doc>

Köktürk, F., Ankaralı, H. ve Sümbüloğlu, V., (2009), “Veri Madenciliği Yöntemlerine Genel Bakış”, Türkiye Klinikleri J Biostat,1(1):20-5.

Coşkun, E., Karabulut, E. ve Karaağaoğlu, E., (2009), “Random Forest ve Destek Vektör Makinası Yöntemleri ile Gen Seçimi ve Sınıflama”, 6. İstatistik Kongresi, Antalya, 29 Nisan-3 Mayıs 2009.

RServe ile Veri Madenciliği Algoritmalarının Web Uygulaması

Batuhan BAKIRARAR, İrem KAR, Atilla Halil ELHAN

Ankara Üniversitesi Tıp Fakültesi, Biyoistatistik Anabilim Dalı, Ankara

Özet

Rserve, diğer programların R programlama dilini başlatmasına veya R kütüphanelerine bağlantı yapmasına gerek olmadan R'nin sahip olduğu tüm imkanları kullanmasını sağlayan bir TCP/IP sunucusudur. Her bağlantı ayrı bir çalışma alanı ve çalışma dizinine sahiptir. C/C++, PHP ve Java gibi popüler diller için istemci uygulamaları mevcuttur. Rserve uzaktan bağlantıyı, kimlik doğrulama ve dosya transferini destekler. Genellikle diğer uygulamalardaki istatistiksel model kurmada, grafik çiziminde, vs. R sunucusunu entegre etmek için kullanılır. Java ise açık kaynak kodlu, nesneye yönelik, platformdan bağımsız, yüksek performanslı, çok işlevli, yüksek seviye, adım adım işletilen (interpreted) bir programlama dilidir.

Çalışmadaki amacımız, Java platformunda geliştirilen arayüz ile web sitesi üzerinde veri madenciliğinin hiçbir program/programlama dili gereksizdirin yapılmasını sağlamaktır. Bu websitesi arka planda R programlama dilinde geliştirilen tüm veri madenciliği kütüphanelerini ve algoritmalarını kullanmaktadır. Sürekli gelişen ve zenginleşen R kütüphaneleri sayesinde websitesinin de geliştirilmesi ve güncel kalması amaçlanmaktadır.

Anahtar Sözcükler: Java, R, RServe

Çoklu Uyum Analizi ve Bir Uygulaması

Nihat Burak ZİHNİ¹, Özlem UZUN¹, Gülay YEGİNOĞLU², Şahi Nur KALKIŞIM¹, Canan ERTEMOĞLU ÖKSÜZ¹

¹ Sağlık Hizmetleri Meslek Yüksekokulu, Karadeniz Teknik Üniversitesi, Trabzon, Türkiye

² Tıp Fakültesi Temel Tıp Bölümü, Karadeniz Teknik Üniversitesi, Trabzon, Türkiye
ygulay@ktu.edu.tr

nbzihni@ktu.edu.tr, ozlemuzun@ktu.edu.tr, skalkisim@ktu.edu.tr, certemoglu@ktu.edu.tr

Giriş: Verileri negatif değer almayan, varsayımlardan bağımsız olan ve parametrik olmayan bir teknik olarak ifade edilen çoklu uyum analizi, kategorik değişkenlerin yorumlanmasını kolaylaştıran, olumsuzluk tablolarında (çapraz tablo) satır ve sütun değişkenleri arasındaki benzerlikleri, farklılıkları, ilişkileri ve bu değişkenlerin birlikte değişimlerini daha az boyutlu bir uzayda grafiksel olarak gösteren bir yöntemdir. Bu yöntemi diğer geleneksel istatistik yöntemlerinden ayıran en önemli özelliği, doğrulayıcı değil verinin içeriğini ortaya çıkarmaya çalışan açıklayıcı bir analiz olmasıdır. Bu nedenle çoklu uyum analizi çalışmalarında anlamlılık testi eksikliği bulunmaktadır. Çoklu uyum analizindeki anlamlılık testi eksikliği logaritmik doğrusal modeller ile tamamlanmakta ve bu modeller kullanılarak ikili ya da daha yüksek dereceli ilişkiler ve etkileşimler, uygun modelin anlamlılığının sınanması suretiyle belirlenebilmektedir.

Amaç: Bu çalışmada Karadeniz Teknik Üniversitesi Sağlık Hizmetleri Meslek Yüksekokulu 1. ve 2. sınıfındaki öğrencilerin okudukları program tercihleriyle fiziksel özellikleri (boy, kilo, cinsiyet) arasındaki ilişkileri “Logaritmik doğrusal Modeller ve Çoklu Uyum Analizi” kullanılarak ortaya koymak amaçlanmıştır.

Yöntem: Çalışmanın amacına ulaşabilmek için öncelikle incelenen değişkenler arasındaki ilişki logaritmik doğrusal analiz (LDA) ile bir model olarak ortaya konulmaya çalışılmış ardından da incelenen değişkenler ile değişken düzeyleri arasındaki ilişkiler grafiksel bir biçimde sunularak uyum analizi uygulanmıştır. Bu şekilde LDA ve uygunluk analizi birbirlerinin tamamlayıcısı olarak da kullanılmıştır.

Bulgular: Çalışma kapsamında yer alan öğrencilerden 83’ü tıbbi laboratuvar, 89’u tıbbi dokümantasyon ve sekreterlik, 120’si ilk ve acil yardım, 108’i ise tıbbi görüntüleme programına kayıtlıdır. Öğrencilerin (n=400) okudukları program tercihleriyle, kategorik hale çevrilmiş fiziksel özellik (boy, kilo) değişkenlerine LDA uygulandığında değişkenler arasındaki ana etkilerin, ikili etkileşimlerin ve üçlü etkileşimin L.R.(Olabilirlik oranı) Ki-kare değerine göre anlamlı olduğu görülmüştür. ($p < \alpha = 0,05$) LDA sonucunda elde edilen bir başka bulgu ise kısmi ilişkilerdir. Burada ise tüm ana etkilerin ve ikili etkileşimlerden ise sadece tercih edilen program ile boy değişkeni arasındaki etkileşimin anlamlı olduğu ifade edilebilir. ($p < \alpha = 0,05$) Söz konusu değişkenler için çoklu uyum analizi bulguları ise şu şekildedir. Birinci ve ikinci temel hareketsizlik değerleri sırasıyla 0,432 ve 0,408’dir. Birinci boyut toplam hareketsizliğin %19’unu açıklamaktayken, ikinci boyut %18’ini açıklamaktadır. Toplam hareketsizliğin açıklanabilmesi için gerekli boyut sayısı ise 7’dir.

Sonuç: Çalışma sonucunda “Öğrencilerin okudukları program tercihi, boy ve kilo” değişkenlerine çoklu uyum analizi uygulanmış ve analiz sonucunda uzun boylu ve fazla kilolu öğrencilerin tıbbi görüntüleme programını, orta boylu ve zayıf öğrencilerin ilk ve acil yardım programını, kısa boylu

ve normal kilolu öğrencilerin ise tıbbi laboratuvar ile tıbbi dokümantasyon ve sekreterlik programlarını tercih ettikleri görülmüştür.

Anahtar Sözcükler: olumsuzluk tabloları, çoklu uyum analizi, logaritmik doğrusal analiz

Kayıp Landmark Tahmininde Kullanılabilecek Yöntemlerin Performansının Karşılaştırılması

Fatma Ezgi CAN^{1,2}, İlker ERCAN³

¹*Biyoistatistik ve Tıp Bilişimi Anabilim Dalı, Tıp Fakültesi, İzmir Katip Çelebi Üniversitesi, İzmir*

²*Biyoistatistik Anabilim Dalı, Sağlık Bilimleri Enstitüsü, Uludağ Üniversitesi, Bursa*

³*Biyoistatistik Anabilim Dalı, Tıp Fakültesi, Uludağ Üniversitesi, Bursa*

Özet

Amaç: İstatistiksel şekil analizi, benzer şekillerin ya da benzer şekillerden oluşan farklı grupların özelliklerini tanımlamak amacıyla ölçülen istatistiklere ait şekiller kümesinin geometrik analizidir. Şekil analizinde kullanılan landmarklar, anakütle içinde veya anaküteller arasında eşleşen her nesneye karşılık gelen noktalardır. Şekil analizi çalışmalarında landmarklarda kayıp veri olması bilgi kaybına yol açmaktadır. Bu çalışmada 2 boyutlu uzayda landmarklar ele alınmıştır. İki boyutlu uzayda landmarklarda kayıp verileri tahmin etmek için F istatistiğine dayalı bir yaklaşım bu çalışmada önerilmiştir.

Yöntem: Normal dağılımdan $n=30$ birey için 3 landmarklı veri türetilmiştir. 30 birimlik 3 landmarklı veri setinde, kayıp landmark oluşturulmuştur. Bu çalışmada önerilen F istatistiğine dayalı yaklaşımın performansını görmek amacıyla Çoklu Regresyon Atama yöntemi ile karşılaştırılmıştır. Simülasyon çalışmasında kafatası üzerinde tanımlanan 3 landmarklı şekle yönelik 500 tekrar yapılarak yöntemler karşılaştırılmıştır.

Bulgular: Simülasyon çalışmasında yapılan 500 tekrar sonucunda F istatistiğine dayalı yaklaşımdan elde edilen root mean square error (RMSE) sonuçları x koordinatı için 155.22, y koordinatı için 6.40 ve x ile y birlikte 138.53 olarak bulunmuşken; çoklu regresyon atama yönteminde ise x koordinatı için 349.54, y koordinatı için 13.29 ve x ile y birlikte 312.74 olarak bulunmuştur. Relatif bias (RB) sonuçları ise F istatistiğine dayalı yaklaşımda x için 19.10, y için 0.55 ve x ile y birlikte 16.87 olarak elde edilmişken; çoklu regresyon atama yönteminde ise x için 349.54, y için 13.29 ve x ile y birlikte 35.63 olarak bulunmuştur.

Sonuç: Bu çalışmada önerilen F istatistiğine dayalı yöntemin çoklu regresyon atama yöntemine göre x, y ve x ile y birlikte değerlendirildiği durumlarda RMSE ve RB bakımından çok daha iyi sonuçlar verdiği görülmüştür.

Anahtar Sözcükler: Kayıp veri, İstatistiksel şekil analizi, Landmark

Kategori Sayısı İki'den Fazla Olan Bağımsız Değişkenlerde Johnson-Neyman Prosedürü

Tuğçe ŞENÇELİKEL¹, H. Yağmur ZENGİN¹, Ersin ÖĞÜŞ¹, K. Setenay ÖNER²

¹Başkent Üniversitesi Tıp Fakültesi, Biyoistatistik Anabilim Dalı, Ankara, Türkiye

²Eskişehir Osmangazi Üniversitesi Tıp Fakültesi, Biyoistatistik Anabilim Dalı, Eskişehir, Türkiye

Sağlık alanında iki değişken arasındaki ilişki incelenirken ilişkiyi etkileyen farklı değişkenler vardır. Değişkenler arasındaki ilişkide bu ilişkinin yapısında veya yönünde değişikliğe neden olabilecek değişken Moderatör (Düzenleyici) değişken olarak adlandırılır. Moderatör, bağımlı ve bağımsız değişken arasındaki ilişkinin gücüne ve yönüne etki ederken moderatörün etkisi istatistiksel olarak etkileşim terimi ile karakterize edilir. Bu durumu test eden yöntemler sadece etkileşimin anlamlı olup olmadığını göstermektedir. Test sonucunda, ilişki üzerinde moderatörün varlığı tespit edildiğinde; ilişkinin daha iyi analiz edilebilmesi ve yorumlanabilmesi için etkileşim hakkında bilgi sahibi olmak gerekir. Bu amaçla kullanılan en popüler yöntemlerden biri Johnson-Neyman yöntemidir. Bu yöntem moderatör değişkenin hangi noktalarında bağımlı ve bağımsız değişken arasındaki ilişkinin anlamlı ya da anlamsız olduğunu belirleme amacına yönelik bir yöntemdir. Bağımsız değişken, sürekli ya da 2 kategorili ($k=2$) olduğunda yöntemin uygulanmasında sorunla karşılaşılmaz iken, çok kategorili olduğu durumda literatürde yöntemin nasıl uygulanacağına dair detaylı bilgi bulunmamaktadır. Bu nedenle Amanda Kay Montoya, bölgesel anlamlılık için Johnson-Neyman prosedürüne dayalı analitik hesaplamalar yapmıştır. Ancak bu prosedür, 2^{k-1} dereceden polinomiyal denklemlerin köklerinin bulunmasına dayandığından $k \geq 5$ olduğunda denklemlerin kökleri bulunamamıştır. Bunun üzerine Montoya, Omnibus Groups Region Significance (OGRS) adlı aracı geliştirmiş ve kategorik bağımsız değişken ile sürekli bir moderatör değişken arasındaki etkileşimi incelemiştir.

Amaç: Bu çalışmada, OGRS yöntemi ile çok kategorili bağımsız değişken, sürekli moderatör ve bağımlı değişkenin olduğu bir veri seti kullanılarak SAS paket programında uygulanması amaçlanmıştır.

Yöntem: Veri seti Başkent Üniversitesi Adli Tıp Anabilim Dalı'ndan alınan Trafik psikolojisi verilerinin bir bölümünden oluşmaktadır. Bağımlı değişken riskli sürüş davranışı, bağımsız değişken anksiyete (Hafif-Orta-Şiddetli), moderatör (M) değişken ise tutumdur. Bireylerin anksiyete durumları ile riskli sürüş davranışları arasındaki ilişkinin tutuma bağlı olup olmadığı OGRS yöntemi ile test edilecektir. Şiddetli anksiyete D1 ve orta anksiyete D2 olarak kodlanmış ve referans kategori olarak hafif anksiyete alınmıştır.

Bulgular: Regresyon modelinde, D1 ve D2 katsayıları önemli bulunduğundan ($p < 0.001$), tutum skoru sıfır iken sırasıyla hafif ve şiddetli anksiyete arasında; hafif ve orta anksiyete arasında önemli düzeyde ilişki görülmüştür. D1 ve M etkileşimi önemli bulunduğundan ($p < 0.001$) hafif ve şiddetli anksiyete grupları arasındaki; D2 ve M etkileşimi önemli bulunduğundan ($p < 0.05$) hafif ve orta anksiyete grupları arasındaki riskli sürüş davranışı açısından farklılığın tutum değişikçe değişmesi beklenmektedir.

Sonuç: OGRS yöntemi ile bağımlı değişken ve çok kategorili bağımsız değişken arasında sürekli bir moderatör değişkenden etkilenen ilişkinin detaylı olarak analiz edilebileceği ve dolayısı ile araştırmacıların ilişki konusunda daha iyi ve anlaşılır yorum yapabilecekleri belirlenmiştir.

Anahtar Sözcükler: Moderatör, Johnson-Neyman, OGRS

[0,1] Kapalı Aralığında Tanımlı Sürekli Bir Bağımlı Değişkenin Modellenmesinde Kullanılan Farklı Modelleme Stratejilerinin İncelenmesi

Hatice Yağmur ZENGİN¹, Ersin ÖĞÜŞ¹, Erdem KARABULUT²

¹ *Başkent Üniversitesi Tıp Fakültesi, Biyoistatistik Anabilim Dalı, Ankara*

² *Hacettepe Üniversitesi Tıp Fakültesi, Biyoistatistik Anabilim Dalı, Ankara*

Sürekli ve sınırlı bir bağımlı değişken ile bir ya da birden çok bağımsız değişken arasındaki ilişkinin doğasının ortaya çıkartılması pek çok disiplinde sıklıkla karşılaşılan yaygın bir problemdir. Bu tür değişkenler genellikle cevap olarak oran, olasılık veya hız üreten süreçler sonucunda doğal olarak ortaya çıkmaktadır. Literatürde bu türden verilerin modellenmesinde kullanılan farklı stratejiler söz konusudur. Bu bağlamda, 2000’li yıllardan önce sıklıkla doğrusal regresyon analizi uygulanmış olsa da bağımlı değişkenin tanım aralığı dışında tahminler üretmesi sonucunda Lojistik regresyon analizi gibi farklı modelleme seçenekleri değerlendirilmeye başlanmıştır. Örneğin Lojistik regresyonun yanı sıra Tobit model başlarda tercih edilmiş ancak izlem yapılamamasından dolayı değil, bağımlı değişkenin sınırlı olmasından dolayı bazı değerlerin gözlenememesi bu yaklaşımın da verinin modellenmesi için uygun olmadığı görüşünü hâkim kılmıştır. Bu nedenle son yıllarda araştırmacılar bu tür verilerin ortaya çıkmasına neden olan süreçleri daha iyi temsil edebilecek farklı model arayışlarına devam etmektedir. Özellikle, Beta ve Simpleks regresyon modeli gibi modeller son zamanlarda tercih edilmektedir. Ancak bu modellerin altında yatan olasılık yoğunluk fonksiyonlarının tanım bölgelerinin 0 ve 1 sınır değerlerini kapsamaması sebebiyle araştırmacılar sıklıkla kapalı aralıkta tanımlı oran, olasılık veya hız verilerinin modellenmesinde Beta ya da Simpleks regresyon analizi öncesi dönüşüm yöntemlerinin veya doğrudan yığılımlı, genişletilmiş veya karma modellerin kullanımına yönelmişlerdir.

Amaç: Sıfır-ve-Bir-Yığılımlı Beta regresyon, Lojistik Kuantil regresyon ile farklı dönüşüm yöntemleri sonrası uygulanan Lojistik, Beta ve Simpleks regresyon analizlerinin [0,1] aralığında tanımlı sürekli bağımlı değişken tahmin performanslarının R’da gerçek veri seti kullanılarak karşılaştırılması amaçlanmıştır.

Yöntem: Çalışma kapsamında orijinal bir mikrobiyolojik veri seti [2] üzerinde doğrusal (polinomial), Sıfır-ve-Bir-Yığılımlı Beta, Lojistik Kuantil, farklı dönüşüm yöntemleri sonrası Lojistik, Beta, Simpleks regresyon yaklaşımları uygulanarak bağımlı değişken tahmin performansları ortalama karesel hatanın karekökü (Root Mean Square Error –RMSE) kullanılarak karşılaştırılmıştır.

Bulgular: Farklı dönüşüm yöntemleri sonrası Beta regresyon analizi, alternatiflerine göre daha iyi bir bağımlı değişken tahmin performansı sergilemiştir. Sıfır-ve-Bir-Yığılımlı Beta regresyon modeli umulan performansı küçük örnek genişliğinde verememiştir. En düşük tahmin performansı ise Lojistik Kuantil regresyon ile elde edilmiştir.

Sonuç: Araştırmacılar bu tür verilerin modellenmesi sırasında birden fazla yöntem uygulayıp elde edilen sonuçları karşılaştırarak ilgili veriye en uygun modeli tercih edebilir.

Anahtar Sözcükler: Sınırlı ve Sürekli Bağımlı Değişken, Dönüşüm, Beta Regresyon, Simpleks Regresyon, Lojistik Kuantil Regresyon

Referanslar

1. R Core Team (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
2. Zhao, L., T. J. Montville, and D. W. Schaffner. 2000. Inoculum Size of *Clostridium botulinum* 56A Spores Influences: Time-to-Detection and Percent Growth-Positive Samples. *Journal of Food Science* 65(8):1369-1375.

Lojistik Regresyon ve Karar Ağacı Tabanlı Yöntemlerinin Sınıflama Performanslarının Karşılaştırılması.

Yaşar SERTDEMİR, İlker ÜNAL

Çukurova Üniversitesi Tıp Fakültesi Biyoistatistik Anabilim Dalı /ADANA

Giriş: Vaka kontrol çalışmalarında genel amaç, risk faktörlerinin belirlenmesi veya bir tahmin modelinin elde edilmesidir. Bu amaç için lojistik regresyon (LR), K en yakın komşu, diskriminant analizi ve karar ağacı tabanlı yöntemler (karar ağacı (KA) ve topluluk yöntemleri (random forest RF)) kullanılabilir. LR yöntemi uzun yıllardır kullanımda olması nedeniyle daha iyi bilinmekte ve analizde elde edilen katsayıların (risk artışı) kolay yorumlanması nedeniyle daha çok tercih edilmektedir. Son yıllarda sağlık alanında sınıflama ve tahmin amaçlı karar ağacı tabanlı yöntemlerin popüleritesi artmaktadır. Karar ağaçları, karar verecek kişiye, aldığı kararın nasıl ve neye dayanarak alındığını anlamada ve yorumlamada kolaylık sağlaması nedeniyle kullanışlı bir yöntemdir. Topluluk tabanlı karar ağaçlarının avantajı ise, tek bir karar ağacı yerine birden çok karar ağacının topluca değerlendirilmesidir. Pubmed de yapılan taramada son 10 yılda sağlık alanında 50200 vaka kontrol çalışmasında lojistik regresyon analizinin uygulanmasına karşın, sadece 2180 karar ağacı ve 250 topluluk tabanlı karar ağacı uygulaması gözlenmektedir. Bu 250 yayının da 193 tanesi son 5 yılda yapılmıştır.

Amaç: Lojistik regresyon ve karar ağacı tabanlı yöntemlerinin, literatürde verilen veri setleri kullanarak elde edilecek performans ölçütlerinin karşılaştırılması amaçlanmaktadır.

Material Yöntem: Literatürde verilen veri setleri kullanacak ve her veri seti önce %70 lik eğitim ve %30'luk test veri seti olmak üzere bölünecektir. Eğitim veri setlerine LR, KA ve RF yöntemleri uygulandıktan sonra elde edilen modeller test veri setlerine uygulanarak duyarlılık, özgüllük, pozitif prediktif değer(PPD), negatif prediktif değer(NPD), genel doğruluk(GD) ve Eğri altında kalan alan(EAKA) performans ölçütleri hesaplanacaktır. Bu üç yöntem için performans ölçütlerinin sıraları elde edilecektir ve bu sıra istatistikleri karşılaştırılacaktır.

Anahtar kelime: Lojistik regresyon, Sınıflama, Karar ağacı, Performans.

Dayanıklı Regresyon Yöntemi (M): Sağlık Verisinde Bir Uygulama

Suriye ÖZGÜR¹, Özge ÇAPKAN², Ş. Büşra MERT², Mert İŞERİ², Timur KÖSE¹, Pembe KESKİNOĞLU³, Mehmet ORMAN¹

¹Ege Üniversitesi Tıp Fakültesi Biyoistatistik ve Tıbbi Bilişim AD

²Dokuz Eylül Üniversitesi Fen Fakültesi, İstatistik Bölümü

³ Dokuz Eylül Üniversitesi Tıp Fakültesi Biyoistatistik ve Tıbbi Bilişim AD

Özet

Giriş: Çoklu doğrusal regresyon analizindeki en büyük problemlerden biri, regresyon parametreleri için en düşük kareler yöntemi ile güvenilmez tahminlere yol açan aykırı değerlerdir. Ayrıca, artıkların normal dağılıma uymaması ve çoklu doğrusal bağlantı durumunda, sağlam tahmin ediciler üzerine dayalı yöntemlerin kullanılması önerilmektedir.

Amaç: Verilerdeki genel karakteristik yapıya uymayan gözlemlerin varlığı, EKK (En Küçük Kareler) tahmin sonuçlarını, bu tahminlere ait standart hataları ve güven aralıklarını olumsuz şekilde etkilemektedir. Bu çalışmanın amacı, EKK'ye alternatif olarak dayanıklı yöntemlerin kullanılması ile aykırı değerlerin verideki etkinliği azaltılarak, tahminciler üzerine etkilerini araştırmaktır.

Materyal ve Metot: Bu çalışmada kullanılan veri seti, 65 ve üzeri yaşta 247 hastanın demans tanısı için bilişsel işlevlerinin belirlendiği bir çalışmanın gerçek verilerinden alınmıştır. Bilişsel işlevleri temsil eden Y bağımlı değişkeni, standartlaştırılmış mini mental test skoru kullanılarak tanımlanmıştır. Modellerde bağımsız değişkenler olarak (X_i) yaş (yıl), cinsiyet (kadın-erkek), eğitim (okur-yazar olma durumu) değişkenleri kullanılmıştır. Demans ve kognitif durum bozukluğu için modelde yer alan açıklayıcı değişkenler literatürdeki öngörülen şekildedir. Araştırmanın analizi, R yazılımı ve NCSS 11 Veri Deneme sürümü kullanılarak gerçekleştirilmiştir. Değerlendirmeler, artıkların normalliği ve çoklu doğrusal bağlantı kontrolü yapıldıktan sonra çoklu doğrusal regresyon ve Huber tarafından ortaya atılan M tahmincisi ile değerlendirilmiştir.

Bulgular: Regresyon modelinde artıkların normal dağılmadığı görülmüştür ($p < 0,001$). Veri seti aykırı gözlemleri de içermektedir. Çoklu lineer regresyon modelinin açıklayıcılık katsayısı 0,30 olarak hesaplanırken, dayanıklı yöntemin açıklayıcılığı 0,46 olarak bulunmuştur. Bu modelde mini mental test skoru, yaş, cinsiyet ve eğitim durumu ile açıklanmaya çalışılmıştır.

Sonuçlar: Aykırı ve uç değerlerin olduğu veri setlerinde EKK yöntemi ile elde edilen tahmin sonuçları, standart hatalar ve güven aralıkları olumsuz şekilde etkilenmiştir. Diğer yandan, dayanıklı tahmincilerle aykırı değerler belirlenerek, güvenilir tahminler elde edilebilir.

Anahtar Sözcükler: Çoklu lineer regresyon, Dayanıklı regresyon, Aykırı değerler, Mini mental test skoru

Logaritmik Doğrusal Modeller ve Uyum Analizinin Birlikte Kullanımı: Lise Öğrencilerinin Sigara İçme Alışkanlıklarına Uygulanması

Fisun KAŞKIR KESİN^{a,b}, Veysel YILMAZ^c,

^a *Düzce Üniversitesi, Sosyal Bilimler Meslek Yüksekokulu, Sosyal Güvenlik Bölümü, Düzce, Türkiye*

^b *Uludağ Üniversitesi, Sağlık Bilimleri Enstitüsü, Biyoistatistik AD, Bursa, Türkiye*

^c *Eskişehir Osmangazi Üniversitesi, Fen-Edebiyat Fakültesi, İstatistik Bölümü, Eskişehir, Türkiye*

Amaç: Bu çalışmada amaç, çok değişkenli Kontenjans Tablosu ile özetlenen sigara içme alışkanlığına sahip lise öğrencisinin anne-baba eğitim durumu, okul türü, günlük içilen sigara miktarı, okuldaki başarı durumu gibi kategorik değişkenlerine, geriye doğru aşamalı Logaritmik Doğrusal Modelleri uygulamak, anlamlı bulunan etkileşimleri belirlemek, ilgili değişkenlerin alt kategorileri arasındaki ilişkinin net bir şekilde görülebilmesi için en uygun modelin bileşenlerine Uyum Analizi uygulamak ve bileşenlerin alt kategorilerinin birbirleri ile olan etkileşimlerini grafiksel olarak yorumlamaktır. Logaritmik Doğrusal Modeller ile Uyum Analizinin birbirlerini bütünleyen analiz yöntemleri olduğunu sergilemektir.

Yöntem: Bu kapsamda, 2010 yılında Eskişehir’de il merkezindeki liselerde okuyan 662 öğrenciye ait, sigara içme alışkanlıklarına ilişkin veri derlenmiştir. Değişkenlere önce geriye doğru aşamalı Logaritmik Doğrusal Model, elde edilen modelin bileşenlerine de Uyum Analizi uygulanmıştır.

Bulgular: Geriye doğru aşamalı Logaritmik Doğrusal Modeller sonucunda en fazla üçlü etkileşimler istatistiksel olarak anlamlı bulunmuştur. Geriye doğru aşamalı Logaritmik Doğrusal analiz sonucunda gelecek 5 yıl içinde sigara içme durumu*okuldaki başarı durumu*sigaranın sağlığa zararlı olduğu düşüncesi ve günlük içilen sigara miktarı*sigara içildiğinden ailenin haberdar olup olmaması*Okul türü üçlü etkileşimleri anlamlı bulunmuştur. Uyum Analizi ile gelecek 5 yıl içinde sigara kullanımını arttırma niyetinde olan öğrencilerin sigaranın sağlığa zararsız olduğu düşüncesinde oldukları ve okuldaki başarı durumlarının çok başarısız olduğu; gelecek 5 yıl içinde sigarayı bırakma niyetinde olan öğrencilerde ise sigaranın sağlığa zararlı düşüncesi etkilidir ve bu öğrencilerin başarı durumlarının başarılı ve başarısız olduğu görülmektedir. Ayrıca, sigara içtiğinden ailesinin haberi olan düz lise ve meslek lisesi öğrencilerinin günlük sigara içme miktarının 21 ve üzeri, 11–20 ve 6–10 gibi çok sayıda olduğu; sigara içtiğinden ailesinin haberi olmayan Anadolu ve fen lisesi öğrencilerinin ise günlük 1–5 adet sigara içtiği söylenebilir.

Sonuç: Logaritmik Doğrusal Modeller ve Uyum Analizi, kategorik değişkenler arasındaki ilişkilerin yorumunu kolaylaştıran istatistiksel birer tekniktir. Kontenjans Tablosu analizinde ki-kare analizi özellikle satır ve sütun değişkenlerine ait kategori sayısının çok olması durumunda kullanılamaz hale gelebilmektedir, etkileşimler ve birlikte değişimler analiz edilememektedir. Bu nedenle değişkenler arasındaki etkileşimi ortaya koymak için Uyum Analizi ve Logaritmik Doğrusal Modeller gibi tekniklere ihtiyaç vardır. Uyum Analizinde anlamlılık testi desteği olmadığından, Logaritmik Doğrusal Modeller gibi bir analiz ile kontrol edilmelidir.

Anahtar Sözcükler: Logaritmik Doğrusal Model; Uyum Analizi; Çoklu Uyum Analizi; Kontenjans Tabloları

Yarı En Küçük Kareler Regresyonu

Erdoğan ASAR¹, Erdem KARABULUT²

¹*Hacettepe Üniversitesi Sağlık Bilimleri Enstitüsü Biyoistatistik Anabilim Dalı Doktora Öğrencisi, Ankara*

²*Hacettepe Üniversitesi Tıp Fakültesi Biyoistatistik Anabilim Dalı Öğretim Üyesi, Ankara*

Özet

Amaç: GEE yaklaşımı çerçevesinde geliştirilen bir ilişkili veri analiz yöntemi olan QLS tanıtılacak, elde edilecek gerçek verilerle GEE ve QLS sonuçları değerlendirilecektir.

Genelleştirilmiş Kestirim Eşitlikleri (The Generalized Estimating Equation - GEE) veri analizinde sağladığı yararlarından dolayı özellikle biyoistatistik alanında, boylamsal (panel) verilerin modellenmesinde yaygın kullanılan yöntemlerden birisidir.

GEE'nin çıkış noktası, GEE'nin Genelleştirilmiş Doğrusal Modelleri (The Generalized Linear Models - GLM) ilişkili verilere uygun bir biçimde uyarlayarak genişletilmesidir.

Yarı En Küçük Kareler (Quasi-Least Squares – QLS) ise Liang ve Zeger (1986) tarafından tanıtılan GEE yaklaşımı çerçevesinde, Shults ve Chaganty (1996-1999) tarafından geliştirilen ilişkili bir veri analiz yöntemidir.

Kesitsel veriler; gözlemlerin bir zaman dilimindeki ölçümleri içerir ve sadece gözlemler arası değişimlerin analizinin yapılmasına olanak sağlar. Boylamsal veriler ise aynı birimlere ait özelliklerin zaman içinde tekrarlı olarak ölçülmesi ile elde edilen verilerdir ve gözlemlerin zaman içerisindeki değişimleri incelenebilir. Verilerin kümelenmiş (bir gözlemin farklı zamanlardaki ölçümleri) olmasından dolayı küme içindeki korelasyonun da göz önüne alınarak modellerin kurulması ve analizlerin yapılması gerekmektedir. Bu tür durumlarda klasik regresyon modelleri ile analizler yapılamadığından özel regresyon modelleri geliştirilmiştir.

GLM, geleneksel regresyon modelinin bir uzantısıdır ve bu tür modellerde hataların dağılımının normal olması ve bağımlı değişkenin varyansının sabit olması varsayımları sağlanmalıdır. Fakat boylamsal verilerin analizinde bu varsayımlar her zaman sağlanamamaktadır. GEE ise GLM yaklaşımı çerçevesinde geliştirilmiş olup, boylamsal verilerin GLM'de istenen varsayımların sağlanamadığı durumlarda da analiz edilmesine olanak sağlamaktadır. Ayrıca GEE, dengeli olmayan (gözlemlere ait eksik ölçümler) verilerde de kullanılabilir. Aynı gözleme ait ölçümler arasındaki bağımlılığı açıklamak için "çalışan korelasyon matrisi (working correlation matrix)" belirlenir.

QLS, GEE çerçevesinde korelasyon parametrelerinin tahmini için aşamalı bir algoritması olan bilgisayar hesaplamalarına dayalı sayısal bir yaklaşımdır. α korelasyon parametresi ve β regresyon parametresi olmak üzere bu kestirimleri sırayla hesaplar:

- β değerinin $\alpha=0$ varsayarak bir başlangıç değeri elde edilir. Sonra β 'nın GEE kestirim eşitliği için $\alpha=0$ değerlendirilerek bir çözüm elde edilir. Bulunan değerler için hesaplamalar tekrarlanır.
- β kestiriminde yakınsama (convergence) oluncaya kadar belirli adımlar sıra ile yapılır.

Aynı gözleme ait ölçümler arasındaki bağımlılığı açıklamak için sıklıkla kullanılan çalışan korelasyon çeşitlerinden bazıları aşağıda verilmiştir.

- **Değiştirilebilir Korelasyon** (Exchangable, Equicorrelated): Aynı gözlemin ölçümleri arasındaki korelasyonun aynı olduğunu varsaymaktadır. Tekrarlı ölçümlerde zamana bağıllık olmadığı durumlarda yaygın olarak kullanılan korelasyon yapılarındandır.
- **Otoregresif Korelasyon-AR-1** (The First-Order Autoregressive-AR-1): Sağlık alanında zamana bağlı tekrarlı ölçümler olduğunda sıklıkla kullanılmaktadır. Özellikle yakın zaman aralıklarıyla elde edilen verilerde en uygun korelasyon yapısıdır.
- **Markov Korelasyon Yapısı** (The Markov Correlation Structure): AR-1'in eşit zaman aralıklı olmayan ölçümler için genelleştirilmiş korelasyon yapısıdır.
- **Yapılandırılmamış Korelasyon** (Unstructured): Bu korelasyon yapısında, birim içi korelasyon için varsayılan bir örüntü yoktur. Bu yapı, tüm gözlemler için ölçüm zamanlarının ortak olduğu çalışmalar için uygundur. Buradaki çekince, korelasyon parametresinin boyutunun gözlem sayılarının makul olduğunda bile büyük olmasıdır.

QLS, eşit zaman aralıklı olmayan boylamsal verinin analizine uygun olan Markov korelasyon yapısının kullanımına olanak tanımaktadır.

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

Sürekli Yapıdaki Çoklu Tanı Testleri İçin Birleştirme Yöntemlerinin ROC Eğrisi Analizi Kullanarak Karşılaştırılması

Muzaffer BİLGİN¹, Ertuğrul ÇOLAK¹

*¹ Eskişehir Osmangazi Üniversitesi, Tıp Fakültesi, Temel Tıp Bilimleri Bölümü, Biyoistatistik AD,
Eskişehir*

Amaç: Bir hastalığa ait tanının belirlenmesinde birden fazla medikal tanı testinden yararlanılabilir. Bu testlerin birleştirilmesi ile daha hassas hasta sağlam sınıflaması yapmak mümkündür. Bu çalışmanın amacı, medikal tanı testlerini doğrusal olarak birleştiren çoklu Reciever Operator Characteristics (ROC) modelinde kullanılan en iyi doğrusal birleştirme yöntemi ile (Best Linear Combination Method) konvansiyonel yöntemlerden olan ve medikal tanı testlerinin birleştirilmesinde yaygın olarak kullanılan lojistik regresyon ve ayırma analizi yöntemlerinin performanslarını karşılaştırmaktır.

Yöntem: En iyi doğrusal birleştirme yöntemi ile konvansiyonel yöntemlerin karşılaştırılması simülasyon çalışması ile yapıldı. Simülasyon çalışması, karaciğer yetmezliği tanısının belirlenmesinde kullanılan 6 farklı medikal test baz alınarak gerçekleştirilmiştir. Hasta ve sağlam bireyler için bu testlere ortalama vektörleri ve varyans-kovaryans matrisleri için parametre değerleri literatürde yer alan çalışmalardan elde edilmiştir.

Belirlenen parametre değerleri kullanılarak medikal test değerleri hem hasta hem de sağlam grup için çok değişkenli normal dağılımdan farklı örnek büyüklüklerinde türetili. Türetilen bu veri setleri ile en iyi doğrusal birleştirme ve konvansiyonel yöntemler kullanılarak birleştirilmiş ROC eğrisi analizi ile yöntemlerin performansları karşılaştırıldı. Karşılaştırma kriterleri olarak eğri altında kalan alan (AUC), optimum kesim noktası ve bu kesim noktasına göre belirlenen duyarlılık, özgüllük değerleri kullanıldı. Simülasyonlarda 10000 tekrar kullanıldı. Simülasyon çalışmalarına ek olarak, yöntemlerin karşılaştırılması gerçek uygulama veri seti kullanılarak yapılmıştır. Bu veri seti over kanserli 60 hasta ile 70 sağlıklı kadın bireylere ait CA125, CA19.9, CA15.3 ve CEA medikal tanı test değerlerini içeren veri seti, İzmir'de bulunan Sağlık Bilimleri Üniversitesi Tepecik Eğitim ve Araştırma Hastanesi'nden elde edilmiştir.

Bulgular: Simülasyon çalışmaları sonucunda değerlendirme kriterlerine göre varyans-kovaryans matrislerinin homojenliğinden etkilenmeyip, performansı en yüksek yöntem, en iyi doğrusal birleştirme yöntemi olarak bulunmuştur. Gerçek uygulama veri setine göre karşılaştırmalar yapıldığında, en yüksek AUC değerini veren birleştirme yöntemi en iyi doğrusal birleştirme yöntemi olarak belirlenmiştir.

Sonuçlar: En iyi doğrusal birleştirme yöntemi konvansiyonel yöntemlerden olan lojistik regresyon analizi ve ayırma analizi yöntemlerine göre daha başarılı birleştirme yöntemi olarak belirlenmiştir. Ayrıca en iyi doğrusal birleştirme yöntemi kullanılarak çoklu medikal tanı testlerinden tek bir optimum kesim noktası hesaplamak mümkündür. Bu yönü ile diğer konvansiyonel yöntemlere göre avantajlıdır.

Anahtar Sözcükler: Medikal Tanı Testi, En İyi Doğrusal Birleştirme Yöntemi, Lojistik Regresyon Yöntemi, Ayırma Analizi Yöntemi

Modelling Covariance Structure for Repeated Measures Data Using Bayesian Method

Mehmet Ziya FIRAT

*Akdeniz University, Faculty of Agriculture, Department of Animal Science, Biometry and Genetics
Unit, Antalya*

Repeated measures data come from experiments in which observations are made on subjects at regular, planned times. These experiments have two or more treatments and are set up using familiar designs. Repeated measures designs are widely used throughout the life sciences. For example, in most clinical trials, repeated measurements are taken on the same experimental subject over time. It is usually plausible to assume that observations on the same unit are correlated. Ignoring covariance structure may result in erroneous inference, and avoiding it may result in inefficient inference. Therefore, to estimate the treatment effect, it is important to adequately model the covariance structure of the repeated measurements. Recently available mixed model methodology permits the covariance structure to be incorporated into the statistical model. The General Linear Model (GLM) methodology was usually used to apply methods based on the mixed model with special parametric structure on the covariance matrices. A mixed model is a generalization of the standard model used in the GLM procedure, the generalization being that the data are permitted to exhibit correlation and variability. However, recent advances in statistical software made possible by the rapid development of computer technology in the past decade have made many new methods available to data analysts, such as Bayesian method. This method, which formally accounts for the uncertainty in the parameter values for a mixed model which allows quadratic changes over time, circumvents many of the problems associated with conventional procedures. The objective of this paper is to provide a background understanding of mixed model methodology for modelling covariance structure of repeated measures data and to use a real data set from a rheumatologic trial comparing disease groups in terms of the joint pain scores to illustrate the use of Bayesian Method.

Keywords : Repeated measures, Bayesian, Covariance structure, Mixed model

Diyabetik Retinopati Hastalığında, Eksuda Lezyonlarının Tespiti ve Sınıflandırılması

Uğur SEVİK

Karadeniz Teknik Üniversitesi, Fen Fakültesi, Bilgisayar Bilimleri Anabilim Dalı, Trabzon

Amaç: Diyabetik Retinopati (DR), diyabet sonucu oluşan ve erken teşhis edilmediği takdirde körlüğe neden olan bir göz hastalığıdır. Bu göz hastalığının görme kaybına yol açmaması için diyabet hastalarının düzenli olarak oftalmologlar tarafından renkli retina görüntüleri alınıp erken teşhis ve tedavi yöntemlerinin uygulanması gerekmektedir. Bu amaçla, DR hastalığı sonucu oluşan parlak lezyonlardan eksuda lezyonu, istatistiksel görüntü analiz yöntemleri ile tespit edilip, hekimler için bir karar destek sistemi (KDS) oluşturulmuştur. Eksuda, sert ve yumuşak olmak üzere iki farklı formda oluşabilmektedir. Dolayısıyla bu iki lezyonunda ayrı ayrı sınıflandırılması yapılmıştır.

Yöntem: Yapılan çalışmada, KTÜ Tıp Fakültesi, Göz Hastalıkları AD retina polikliniğinden Fundus Kameralar ile alınan DR hastalığına sahip 125 retina görüntüsü kullanılmıştır. Ayrıca bu görüntüler DRIMDB ismiyle açık erişimli olarak literatüre kazandırılmıştır. DR lezyonları koyu ve parlak olmak üzere iki ana grupta toplanabilir. Çalışmada, parlak lezyonların tespiti için arka plan kestirim yöntemleri uygulanmış ve koyu lezyonlar elimine edilmiştir. Geriye kalan parlak lezyonlar eksuda lezyonları olup, KDS ile birlikte sert ve yumuşak olarak iki sınıfa ayrılmıştır. KDS için retina görüntü veri tabanının % 60'ı eğitim %40'ı test verisi olarak kullanılmıştır. Çalışmada, eksuda lezyonlarının parlaklık değerleri kullanılarak medyan, standart sapma ve ortalama öznitelikleri çıkartılmıştır. Bununla birlikte eksuda lezyonundan alan, ağırlık merkezi, sınırlayıcı çerçeve, büyük eksen boyutu, dış merkezlilik, yönelme, dışbükey alan, dolu alan, euler sayısı, eşdeğer çap, sağlamlık, kapsam, çevre ve yuvarlaklık gibi şekilsel öznitelikler de çıkartılmıştır. Tüm bu öznitelikler k-En Yakın Komşu (k-EYK), Naive Bayes (NB), Yapay Sinir Ağları (YSA), Karar Ağaçları (KA), Rastgele Orman (RO) ve Adaboost sınıflandırıcılarına eğitim verisi olarak verilmiş ve her yöntem ayrı ayrı sınıflandırma performansları ölçülmüştür. Sınıflandırıcıların eğitim aşamasında ezberleme gibi yan etkilerinden kurtulmak için veri seti 10-kere Çapraz Doğrulama yöntemi kullanılmıştır

Bulgular: Bu çalışmada, 527 adet sert eksuda lezyonu ile 9 adet yumuşak eksuda lezyonu sınıflandırıcılar tarafından eğitilip test edilmiştir.

Sınıflandırıcı	Sınıf	TP	TN	FP	FN	Hassasiyet	Duyarlılık	F1
DVM	SE	526	0	9	1	98,32%	99,81%	99,06%
	YE	0	526	1	9	0,00%	0,00%	0,00%
AdaBoost	SE	524	2	7	3	98,68%	99,43%	99,05%
	YE	2	524	3	7	40,00%	22,22%	28,57%
YSA	SE	523	1	8	4	98,49%	99,24%	98,87%
	YE	1	523	4	8	20,00%	11,11%	14,29%
KA	SE	522	1	8	5	98,49%	99,05%	98,77%
	YE	1	522	5	8	16,67%	11,11%	13,33%
RO	SE	525	0	9	2	98,31%	99,62%	98,96%
	YE	0	525	2	9	0,00%	0,00%	0,00%
NB	SE	511	6	3	16	99,42%	96,96%	98,17%
	YE	6	511	16	3	27,27%	66,67%	38,71%
1NN	SE	521	1	8	6	98,49%	98,86%	98,67%
	YE	1	521	6	8	14,29%	11,11%	12,50%

Tüm öznitelikler kullanıldığında parlak lezyonların sınıflandırma performansları (SE: Sert Eksuda, YE: Yumuşak Eksuda) aşağıdaki tabloda verilmiştir. Tabloda performans ölçütleri olarak hassasiyet, duyarlılık ve F1 ölçütü kullanılmıştır.

Sonuç: Yapılan çalışmalar sonucunda arka plan kestirimi sonucu ortaya çıkan görüntüler üzerinde sert ve yumuşak eksuda lezyonlarının sınıflandırılmasında en başarılı sınıflandırıcı Naive Bayes yöntemi olduğu görülmüştür. Yumuşak lezyonların başarısının düşük olması nadir görülen bir lezyon olduğundan yeterince eğitim verisinin olmamasıdır.

Anahtar Sözcükler: Diyabetik Retinopati, İstatistiksel Karar Destek Sistemi, Medikal Görüntü Analizi

Kısmi En Küçük Kareler Yöntemi ile İki Aşamalı En Küçük Kareler Yöntemlerinin İncelenmesi: Sağlık Alanında Bir Uygulama

Sıddık KESKİN

Van Yüzüncü Yıl Üniversitesi, Tıp Fakültesi, Biyoistatistik AD, Tuşba-Van

Amaç: Bu çalışmada, standart regresyon analizi yöntemine benzerlik gösteren ancak varsayımlardan bazılarının sağlanamaması durumunda kullanılabilecek analiz yöntemlerinden olan Kısmi en küçük kareler regresyonu ile İki aşamalı en küçük kareler regresyonu incelenerek sağlık alanında bir uygulama yapılmıştır.

Yöntem: Kısmi En Küçük Kareler Regresyonu (PLS): Açıklayıcı değişkenler arasında yüksek korelasyonun (multicollinearity) olması veya açıklayıcı değişken sayısının gözlem sayısından fazla olması durumunda, standart regresyon analizi yerine tercih edilebilen bir yöntemdir. PLS, orijinal değişkenlerin doğrusal kombinasyonu ile gizli (latent) değişken, faktör veya bileşen olarak adlandırılan yeni değişkenler elde ederek, bu yeni değişkenleri standart regresyon analizi modeline dahil eder. PLS birden fazla cevap değişkenini tek modelde özetleyebilir. Temel bileşenler analizi ve çoklu regresyon analizi yöntemlerini kombine eden bir yöntemdir. Analizde temel amaç, **X** ve **Y** arasındaki ilişkinin, t adet ortogonal gizli (latent) değişken olarak bilinen bileşenler yardımıyla bulunmasıdır. Birinci aşamada açıklayıcı değişkenler arasındaki kovaryansın maksimizasyonu yapılır. İkinci aşamada ise t adet gizli değişken ya da bileşen kullanılarak standart regresyon analizi uygulanır.

İki Aşamalı En Küçük Kareler Yöntemi (2SLS): Standart regresyon analizinde, “**X** değişkenleri hatasız ölçülmüştür” varsayımı sağlanmadığında İki Aşamalı En Küçük Kareler (2SLS) yöntemi kullanılabilir. İki aşamalı en küçük kareler yöntemi, aracı veya yardımcı (instrumental variable) değişkenler kullanılarak, açıklayıcı içsel (endogenous) değişkenler ile yapılan özel bir regresyon analizi yöntemidir. İki aşamalı en küçük kareler regresyonu yaklaşımında genel ilke, model parametrelerini tahmin etmek için hata terimi ile ilişkisiz olan yardımcı ya da aracı değişkenlerin kullanılmasıdır. Bu aracı değişkenler, içsel değişkenler ile korelasyonludur. Ancak hata terimi ile korelasyonlu değildir. Analizde iki aşama vardır ve bu iki aşama aracı değişkenler kullanılarak eş zamanlı olarak yapılabilir.

Çalışmada, yöntemler ile ilgili uygulama yapmak üzere; sağlık alanında, 64 bireye ait, 7 değişkenden (Yaş, SKB, BKİ, PLT, AST, ALT, LDH) oluşan veri seti kullanılmıştır. Her iki yöntemi için; LDH bağımlı değişken olarak alınmıştır. İki aşamalı en küçük kareler yönteminde; Yaş, SKB ve BKİ aracı değişken olarak; PLT, AST ve ALT ise içsel değişken olarak alınmıştır.

Bulgular: Kısmi en küçük kareler yöntemi için ALT dışında diğer değişkenlerin modele katkıları istatistik olarak önemli bulunmuştur. Modellere e ait Belirleme katsayıları (R^2) ise PLS’de %54.2, 2SLS de ise %15.2 olarak bulunmuştur.

Sonuç: Sonuç olarak, standart regresyon analizinin bazı varsayımları sağlanmadığında, her iki yöntemin de kullanılabileceği, ayrıca PLS’ nin boyut indirgeme amaçlı olarak da kullanılabileceği vurgulanmıştır.

Anahtar Sözcükler: Çoklu bağlantı, gizli değişken, aracı değişken, temel bileşen

Metabolomik Verilerin Sınıflandırılmasında Kısmi En Küçük Kareler Ayırma Analizi Yaklaşımı

Meltem ÜNLÜSAVURAN^{1,2}, Gökmen ZARARSIZ^{1,2}, Funda İPEKTEN^{1,2}, Gözde ERTÜRK ZARARSIZ², Selçuk KORKMAZ^{2,3}, Dinçer GÖKSÜLÜK^{2,4}, Halef Okan DOĞAN⁵, Vahap ELDEM⁶, Ahmet ÖZTÜRK^{1,2}

¹Erciyes Üniversitesi, Tıp Fakültesi, Biyoistatistik Anabilim Dalı, Kayseri

²Turcosa Analitik Çözümlemeler Ltd. Şti., Erciyes Teknopark, Kayseri

³Trakya Üniversitesi, Tıp Fakültesi, Biyoistatistik Anabilim Dalı, Edirne

⁴Hacettepe Üniversitesi, Tıp Fakültesi, Biyoistatistik Anabilim Dalı, Ankara

⁵Cumhuriyet Üniversitesi, Tıp Fakültesi, Biyokimya Anabilim Dalı, Sivas

⁶İstanbul Üniversitesi, Fen Fakültesi, Biyoloji Bölümü, İstanbul

meltemunlusavuran@gmail.com

Özet

Metabolomik, belirli bir zaman diliminde dokularda, hücrelerde ve fizyolojik sıvılarda lipid, karbonhidratlar, vitaminler, hormonlar ve diğer hücre bileşenlerinden ortaya çıkan küçük molekül ağırlıklı metabolitlerin yüksek kapasiteli teknolojiler kullanılarak saptanması, miktarının belirlenmesi ve tanımlanmasıdır. Genomik ve proteomik gibi diğer -omik bilimlerinde ne olabileceğinin bilgisine ulaşılabilirken, metabolomik biliminde gerçekte ne olduğunun bilgisine ulaşılabilir. Bu nedenle, tüm metabolitlerin ayrıntılı ve kantitatif ölçümünde, hastalıkların teşhisinde veya toksik ajanların fenotip üzerindeki etkilerini araştırmada en ideal yöntemler arasında yer almaktadır. Metabolomik verileri, diğer -omik verileri gibi, yüksek boyutludur ve değişken sayısı (p) gözlem sayısından (n) oldukça fazladır. Bu tür verilerde geleneksel istatistiksel yöntemler kullanılamamaktadır. Bu tür veriler ile baş etmede bir çözüm veri boyutunu indirgeyerek, $p > n$ problemini aşmaktır. Sıklıkla kullanılan boyut indirgeme yöntemlerinden biri temel bileşenler analizidir. Bu yöntem, sıklıkla kullanılan etkin bir istatistiksel yaklaşım olmasına rağmen, sınıf değişkenini kullanmamasından dolayı gruplar arası değil örnekler arasındaki farka odaklanmaktadır. Sınıf değişkenini kullanarak çözümleme yapabilen bir yöntem kısmi en küçük kareler ayırma analizi yöntemidir. Bu yöntem, gruplar arasındaki farkı daha iyi vurgulayabilmekte, ayrıca gürültülü ve ilişkili yapıda olan metabolomik verileri ile baş edebilmektedir. Ayrıca, değişken önemini tanımlayan çeşitli istatistikler sunabilmektedir. Bu çalışmada kısmi en küçük kareler ayırma analizi yönteminin gerçek metabolomik verileri üzerinde kullanımı gösterilecektir.

Anahtar Sözcükler: Boyut indirgeme, kısmi en küçük kareler ayırma analizi, metabolomik, sınıflandırma

Metabolomik biyobelirteçlerinin tespitinde ANOVA-PCA yaklaşımı

Funda İPEKTEN^{1,2}, Gökmen ZARARSIZ^{1,2}, Meltem ÜNLÜSAVURAN^{1,2}, Gözde ERTÜRK ZARARSIZ²,
Selçuk KORKMAZ^{2,3}, Dinçer GÖKSÜLÜK^{2,4}, Halef Okan DOĞAN⁵, Vahap ELDEM⁶, Ahmet
ÖZTÜRK^{1,2}

¹Erciyes Üniversitesi, Tıp Fakültesi, Biyoistatistik Anabilim Dalı, Kayseri

²Turcosa Analitik Çözümlemeler Ltd. Şti., Erciyes Teknopark, Kayseri

³Trakya Üniversitesi, Tıp Fakültesi, Biyoistatistik Anabilim Dalı, Edirne

⁴Hacettepe Üniversitesi, Tıp Fakültesi, Biyoistatistik Anabilim Dalı, Ankara

⁵Cumhuriyet Üniversitesi, Tıp Fakültesi, Biyokimya Anabilim Dalı, Sivas

⁶İstanbul Üniversitesi, Fen Fakültesi, Biyoloji Bölümü, İstanbul

fundaipekten@gmail.com

Özet

Metabolomik, canlı bir organizmada yer alan küçük moleküllerin (metabolitlerin) tanımlanması ve ölçülmesi ile ilgilenen bir bilim dalıdır. Özellikle, hastalıklar ile ilgili biyobelirteçlerin tespitine yönelik etkin teknik ve yaklaşımları kapsamaktadır. Bireylerden alınan örnekler üzerinden vücuttaki çok sayıda metabolit düzeyi ölçülebilmekte ve hastalıklar ile ilişkilendirilebilmektedir. Tek yönlü varyans analizi yaklaşımını kullanarak ilgilenilen bir durum açısından istatistiksel anlamlı metabolitler tayin edilebilse de, bu yaklaşım metabolitler arasındaki etkileşimi dikkate almamaktadır. Çok değişkenli tek yönlü varyans analizi (MANOVA), bu problemin çözümünde kullanılabilen çok değişkenli bir yaklaşım olsa da, metabolomik verileri gibi değişken sayısının (p), gözlem sayısından (n) fazla olduğu durumlarda kullanımını yitirmektedir. Ayrıca çok değişkenli normal dağılım ve varyans-kovaryans matrislerinin homojenliği gibi varsayımlar metabolomik verilerinde bu yöntemin kullanımını kısıtlamaktadır. ANOVA-PCA bu problemlerin üstesinden gelebilen bir yaklaşımdır. İstatistikte sıklıkla kullanılan tek yönlü varyans analizi ve temel bileşenler analizi yöntemlerini bir araya getirmekte, deney tasarımındaki her bir faktörün düzeyine bağlı olarak, verideki değişkenliği artıklardan elde edilen hata miktarı ile karşılaştırmaktadır. Bu çalışmada ANOVA-PCA yönteminin gerçek metabolomik verileri üzerinde kullanımı gösterilecektir.

Anahtar Sözcükler: Biyobelirteç tespiti, metabolomik, tek yönlü varyans analizi, temel bileşenler analizi

Meta-Analizinde Odds Oranlarının Aşırı Yayılım Sorununa Beta-Binomiyal Model Yaklaşımı

Mehtap AKÇİL OK

Başkent Üniversitesi, Sağlık Bilimleri Fakültesi, Beslenme ve Diyetetik Bölümü

Amaç: Meta-Analizi (MA), aynı amaçla yapılmış pek çok çalışmanın sonuçlarını birleştirerek daha kesin ve güvenilir tahminler yapılmasını amaçlayan istatistiksel bir yöntemdir. MA'lerinin çoğu etki ölçütü olarak odds oranlarını (OR) yada logaritmik OR'larını kullanmaktadır. Çalışmaların sonuçları homojen dağıldığında sonuçların birleştirilmesinde herhangi bir problem yoktur ve standart sabit etkiler modelinden (SEM) yararlanılmaktadır. Ancak heterojenlik söz konusu olduğunda genellikle toplamsal (additive) rasgele etki modeli (REM) kullanılarak sonuçlar birleştirilir. Aşırı yayılımın olduğu durumlarda alternatif olarak geliştirilen çarpımsal (multiplicative) REM ile tahmin yapmak daha güvenilir sonuçlar vermektedir. Bu modeldeki çarpımsal faktör, daha önce bildiğimiz sınıf içi korelasyon katsayısı (Intra Class Correlation Coefficient-ICCC) gibi yorumlanabilmektedir. Bu model genellikle klinik denemelerde karşılaştırma kollarının birinde ya da birden fazlasında ilgilenilen olay olasılığının beta dağılımlı olmasından kaynaklanmaktadır. Bu çalışmanın amacı, OR'ların aşırı yayılım göstermesi durumunda ve ICCC'nin farklı kestiricilerinin kullanıldığı BB modellerini yapılmış uygulamalarla tanıtmaktır.

Yöntem: SEM'de her bir çalışmanın tamamen aynı etkiyi kestirdiği varsayımı yapılmakta bireysel (individual) çalışma sonucunun varyansının tersiyle elde edilen ağırlıklar kullanılarak en küçük varyanslı ağırlıklı ortalama kesitimi yapılmaktadır. Varsayım sağlanamadığında ise toplamsal REM kullanılır ve bu modelde hem çalışma içi hem de çalışmalar arası varyansın tersiyle ağırlıklandırılmış sonuçlar elde edilir. Ancak varyans tersiyle elde edilmiş kestirimlerin örneklem sayısı küçük ve seyrek veri (sparse) olduğu durumlarda yanlış sonuçlar verdiği araştırmalarla saptanmıştır. OR'ların aşırı yayılım durumunda Beta-Binomiyal (BB) yaklaşımla birleştirilmesinde yeni geliştirilen farklı kestiricilerden yararlanılması literatürde çok yeni çalışılmaktadır. BB dağılım, ilgilenilen bir olayın başarı olasılığının (p) Beta dağılımı ile binomiyal dağılımın Binom (n,p) karmasından oluşmaktadır. $Y \sim \text{Binom}(n,p)$ ve $p \sim \text{Beta}(\alpha, \beta)$ dağılımı gösterdiğinde, Y koşulsuz olarak α , β ve n parametreleriyle BB dağılım gösterir. Dolayısıyla aşırı yayılım gösteren bireysel çalışmaların bu modelle birleştirilmesinde iki aşamalı model yaklaşımı söz konusudur. α ve β parametrelerinin kestirmek için momentler ya da en çok olabilirlik yöntemlerinden yararlanılmaktadır. Kestirilen parametreler aracılığı ile ağırlıklar hesaplanarak sonuçlar birleştirilir. Elle hesaplanması mümkün olmayıp açık yazılım kaynaklı olan R'da eklenmiş kodlardan yararlanılarak ya da çeşitli araştırmacıların yazdığı macrolar ile SAS NLMIXED kullanılarak kestirimler yapılmaktadır.

Bulgular: BB modeli ile OR'ların birleştirilmesinde farklı araştırmacıların yaptıkları uygulama sonuçları incelenmiştir. Bakbernuly ve Kulinskaya'nın Ocak 2017 yayınlanan çalışmalarında, Pre-eklemsi'de diüretiklerin etkisini inceleyen 9 çalışmanın (toplam 6942 hasta) sonuçları SEM, toplamsal REM ve farklı ICC kestirim yöntemleri kullanılarak BB modellerini incelemişlerdir. ICC arttıkça en doğru kestirimlerin BB modele dayandığı sonucuna varmışlardır. Aynı şekilde, Young-Xu ve Chan (2017) çalışmalarında anti-mantar ajanlarına yönelik hem gözlemsel çalışmaların hem de klinik denemelerin %0.82'den %13.9'a değişen oranları üzerinde BB modelleriyle kestirim

sonuçları elde etmişlerdir. Standart meta-analizi yöntemlerine göre daha kesin kestirim sonuçları ve daha dar güven aralıkları bulmuşlardır.

Sonuç: Farklı konularda uygulanmış meta-analizleri incelediğinde, aşırı yayımlı binomiyal verilerin birleştirilmesinde BB modellerini kullanmanın daha doğru özet kestirimler vereceği sonucuna varılmıştır.

Anahtar Sözcükler: Meta-Analizi; Beta-Binomiyal Model; Aşırı Yayımlı Veri; Odds Oranı; Sınıf İçi Korelasyon

KISA SÖZLÜ SUNUMLAR
SHORT ORAL PRESENTATIONS

Okuduğumuz Meta-analizin Gücünü Biliyor muyuz?

Zeliha Nazan ALPARSLAN, Nazlı TOTİK

Çukurova Üniversitesi Tıp Fakültesi Biyoistatistik Anabilim Dalı, Adana

nazan@cu.edu.tr

Giriş: Meta-analizler, özellikle de randomize kontrollü deneme çalışmalarından yapılanlar, kanıt değeri nedeni ile, son yıllarda giderek popüleritesi artan çalışmalardır. Ancak yapımı aşamasında bütün gerekli şartlar sağlansa bile, meta-analizin gücü göz önüne alınmadan varılacak her sonuç anlamlı olmayabilir. Zira, güç hesaplanması birçok araştırmada olduğu gibi meta-analiz çalışmalarının planlanmasında ve yorumlanmasında da önemlidir. Uzun süredir meta-analiz ile sonuç çıkarımı mevcut olsa da meta-analizde istatistiksel testlerin gücünü hesaplama konusunda çok az çalışma vardır. Bir meta-analiz için yapılan güç analizleri, genel etkinin beklenen büyüklüğü, dahil edilen makale sayısı ve makalelerdeki minimum örnek büyüklüğü (MMÖB) göz önüne alındığında, araştırmacıya istatistiksel açıdan anlamlı bir sonuç bulma imkânı verir. Meta-analizdeki güç hesabı, sabit ve rastgele etki modelleri altında incelenebilir.

Amaç: Meta-analizlerin gücü kavramını kuramsal olarak gözden geçirmenin yanında, güç konusunu sıklıkla oluşan bazı gerçek durumlarda tartışmaya açmaktır.

Yöntem: Literatürde meta-analizler genellikle konu alt başlıklarının değerlendirilmesi ile verilirler. Çalışmanın alt çalışmalarında birincil çıktı değişkenin aynı olması (örneğin VAS ve azalması) yorumu kolaylaştırır, ancak literatürde alt başlıklarının farklı birincil değişkenlerle oluşturulduğu meta-analizler de vardır. Yapılan örnek bir çalışmanın sonuçları incelenecektir.

Sonuç: Günümüzde dergiler gücü %80'den küçük olan çalışmaları özendirmezler. Ama meta-analiz çalışmalarında güç konusu pek tartışılan bir konu değildir. Meta-analizlerin raporlanması için önerilen QUOROM (Quality of Reporting of Meta-Analyses) da bu konuda hiçbir atıf ya da sorgulama yapmamaktadır. İçerdiği çalışmaların münferit güçleri ne olursa olsun, meta-analizlerin gücü hesaplanabilir, hesaplanmalıdır.

Örneklerle de anlatılan birleşik tip meta-analizlerde alt grupların farklı güçlerinin genel olarak güce nasıl katkıda bulunacağı konusu tartışılmaya açıktır.

Anahtar Sözcükler: Güç hesaplaması, sabit etki modeli, rasgele etki modeli, meta-analizin gücü.

Türkiye’de Solunum Hastalıkları Ölümünün Zaman Serisi Analizi

Erkut BOLAT¹, Sedat ALTIN²

¹*İstanbul Üniversitesi Cerrahpaşa Tıp Fakültesi Biyoistatistik Anabilim dalı, İstanbul*

²*Yedikule Göğüs Hastalıkları ve Göğüs Cerrahisi E.A. Hastanesi, İstanbul*

Amaç: Bilindiği üzere solunum sistemi hastalıkları, hem hava kirliliğinin artışı hem de sigarayla ilişkili hastalıkları olması nedeniyle ülkemizde de giderek artmaktadır. Bu artışın doğal yansıması ölüm oranlarındaki artıştır. Biz de bu çalışmada Türkiye’de solunum hastalıklarındaki ölüm sayılarının yıllara göre değişimini araştırmayı amaçladık.

Yöntem: 2010-2015 yılları arasındaki TÜİK ölüm verileri kullanılarak solunum sistemi ölümleri cinsiyete göre karşılaştırılarak analiz edildi. Ölüm sayılarının çalışılan zaman diliminde hangi yıllar arasında istatistiksel olarak anlamlı değişim gösterdiğini saptamak için Joinpoint Regresyon Analizi kullanıldı. Ölüm sayılarındaki değişimin (azalma veya artma) olduğu en uygun nokta belirlendi. Yıllık yüzde değişim (APC: Annual Percent Change) her bir trend için hesaplanmıştır.

Bulgular: İncelenen çalışma süresince toplam solunum hastalıkları ölüm sayısı 205.801 dir. Bunların 123.022’si erkek(%60), 82.779’u kadınlardan(%40) oluşmuştur. Ölümünün 40.644(%19.7)’ü pnömoni, 124.960(%60.7)’i kronik obstrüktif akciğer hastalığı ve bronşiektazi, 10.569(%5.2)’u astım ve 29.628(%14.4)’ü diğer solunum hastalıklarıdır.

2010 ve 2015 yılları arasında toplam solunum hastalıkları ölüm sayısında ortalama yıllık %11.4 (%95 Güven Aralığı: 7.3 ile 15.5) oranında anlamlı bir artış görülmektedir (p<0.001). Erkeklerde %10.1 (6.07-14.4)’lik önemli bir artış (p<0.001), kadınlarda ise %13.3(9.3 -17.4) oranında erkeklere göre biraz daha fazla artış değişimi görülmektedir (p<0.001). Toplam pnömöni ölümlerinde %34.1 (24.9-43.9), erkek pnömöni ölümlerinde %33.3(26.6-42.6) ve kadın pnömöni ölümlerinde %33.7(24.5-43.7) ile oldukça yüksek yüzde ise artış olmuştur (p<0.001). Toplam KOAH ve bronşiektazi ölümlerinde %8(3.2-13.1), erkeklerde %7.1(2.4-12.1), kadınlarda ise %9.1(4.0-14.5) artış görülmüştür(p<0.001). Astım ölümlerinde toplamda yıllık %8.9(2.0-16.3) ve kadınlarda %11.4(5.7-17.4) artış olmuştur(p<0.001) ancak erkeklerde anlamlı bir değişim görülmemiştir. Diğer solunum hastalıkları ölümlerinde de anlamlı bir değişim olmamıştır.

Sonuç: Solunum sistemi hastalıklarında pnömöni, KOAH ve astıma bağlı ölümlerin son 5 yıl içerisinde anlamlı derecede artış gösterdiği tesbit edilmiştir. Özellikle pnömöni ve KOAH ta iki cinste de artış anlamı olarak bulunmuştur.

Anahtar Sözcükler: Zaman trendi, Joinpoint regresyon, Solunum hastalıkları ölümü

Zaman Serileri Analiz Yöntemlerinin Sağlık Alanında Kullanılması

Vedia Bennu GİLAN¹, **Semra ERDOĞAN¹**, Seyran BOZKURT BABUŞ²

¹Mersin Üniversitesi Tıp Fakültesi, Biyoistatistik ve Tıbbi Bilişim Anabilim Dalı, Mersin

²Mersin Üniversitesi Tıp Fakültesi, Acil Tıp Anabilim Dalı, Mersin

ÖZET

Amaç: Zaman serileri analiz yöntemleri pek çok alanda yaygın olarak kullanılmakta olup sağlık alanında kullanımı yeni yeni ortaya çıkmaktadır. Oysaki diğer alanlarda olduğu gibi sağlık alanında da karşımıza çıkan pek çok veri için zamansal değişimlerin gözlenmesi ve ileriye yönelik tahminlere ihtiyaç duyulması durumlarında zaman serileri analiz yöntemlerinden yararlanılabilir. Çalışmamızda ilk olarak, zaman serileri yöntemleri içerisinde en başarılı yöntem belirlenmeye çalışılmıştır. Ardından, Mersin Üniversitesi Tıp Fakültesi Sağlık Araştırma ve Uygulama Merkezi Hastanesi Acil Servisine 2010-2016 yılları arasında darp, keskin cisimle yaralanma, ateşli silah ile yaralanma vb durumlarla gelen 18 yaş üstü kadınların verileri elde edilmiştir. Elde edilen bu verilere üstel düzleştirme yöntemi uygulanarak 2017 yılında acil polikliniğine aynı şikayetle gelecek olan kadınların sayısı tahmin edilmeye çalışılmıştır.

Yöntem: Çalışmamızda, hastaneden alınan orijinal verinin yapısına benzer olacak şekilde Minitab 16.0 paket programı kullanılarak veriler üretilmiştir. Elde edilen verilere SPSS 17.0 paket programı kullanılarak, zaman serisi analiz yöntemlerinden yaygın olarak kullanılan Üstel düzleştirme ve Box-Jenkins (ARIMA) yöntemleri uygulanmış ve sonucunda üstel düzleştirme yönteminin daha başarılı sonuçlar verdiği gözlenmiştir. Ardından, 2010-2016 yılları arasında darp, keskin cisimle yaralanma, ateşli silah ile yaralanma vb durumlarla acil servise gelen 18 yaş üstü kadınların verilerine üstel düzleştirme yöntemi uygulanarak 2017 yılına ait tahminler yapılmıştır.

Bulgular: Üstel düzleştirme ve Box-Jenkins (ARIMA) yöntemlerinden elde edilen tahminler incelendiğinde, üstel düzleştirme yönteminin daha yüksek açıklama oranı ($0.792 > 0.638$), daha düşük Normalized BIC (Bayes bilgi kriteri) ($2.694 < 3.344$) değerine sahip olduğu ve elde edilen tahmin değerleri ile simülasyon verileri arasında istatistiksel olarak anlamlı bir fark olmadığı ($p=0.922$) gözlenmiştir. Hastaneden alınan orijinal veriye üstel düzleştirme yöntemi uygulandığında, 2017 yılında hastanenin acil servisine gelecek kadın sayısı ise 597 olarak tahmin edilmiştir.

Sonuç: Üstel düzleştirme yöntemi hastaneden alınan orijinal veriye uygulandığında 2010 yılından 2016 yılına kadar artarak devam eden şiddet vakalarının 2017 yılında da artarak devam edeceği görülmüştür. Ayrıca diğer yıllarda gözlendiği gibi 2017 yılında da kadına uygulanan şiddetin en sık yaz mevsiminde yaşanacağı sonucu ortaya çıkmaktadır. Çalışmamız sadece hastanenin acil servisine gelen şiddet görmüş kadınlardan oluştuğu için gerçek rakamları tam olarak yansıtmamaktadır. Şiddet görmüş ancak acil servise gelmemiş pek çok kadının daha var olduğu da bilinmektedir.

Anahtar Sözcükler: Zaman serisi analizi, Kadına şiddet, Üstel düzleştirme, Box-Jenkins

Güvenilirliği Anlamak ve Cronbach'ın Alfa Katsayısını Yorumlamak

Yunus Emre KUYUCU, Osman DEMİR

Gaziosmanpaşa Üniversitesi, Tıp Fakültesi, Biyoistatistik A.D. ,Tokat

Amaç: Güvenirlilik, genellikle yanlış anlaşılmış bir kavram olmuştur. Güvenirlilik, kullanılan aracın kendisine değil, bir ölçme aracıyla elde edilmiş ölçümlere (veya sonuçlara) tanımlamada bulunur. Güvenirlilik testin kendisinin değil, eldeki veriler veya ölçümlerin bir özelliğidir. Böylelikle, testin güvenirliliği veya aracın güvenirliliği ya da test güvenilirliktir diye ifade etmek doğru değildir. (Bademci, 2008) Zira ölçümler güvenilir veya güvenilir değildir. Böylelikle, “test güvenirliliği” yerine, “ölçüm güvenirliliği” kavramının kullanılması çok daha uygundur. Geçerlik iddiaları ise, test ölçümlerinin belirli kullanımları ve yorumlarına ilişkin yapılmalıdır. Cronbach'ın (1951) alfa katsayısı yöntemi, ölçüm güvenirliliği kestirimi için kullanılmaktadır. Alfa katsayısı değişik faktörler tarafından etkilenmektedir. Örneğin, toplam test ölçüm varyansı, alfa katsayısı üzerinde en büyük etkiye sahiptir. Böylece, toplam test ölçümlerinin varyansı, madde varyansları toplamından daha küçükse, sonrasında Cronbach'ın alfa katsayısı negatif olabilecektir. Matematiksel olarak, Cronbach alfa, negatif küçük olabilir (Bademci, 2006). Çalışmada güvenirlilik ile ilgili kavramlar incelenmiş ve Cronbach alfanın negatif olma durumu örnek veri setleri üzerinde incelenmiştir.

Yöntem: Test ölçümlerinin güvenirliliğini kestirmenin yöntemlerinden birisi de, Cronbach'ın (1951) alfa katsayısı yöntemidir (Linn ve Miller, 2005; Mehrens ve Lehmann, 1991). Alfa katsayısı, aşağıdaki formül yoluyla hesaplanmaktadır (Reinhardt, 1996; Worthen, White, Fan ve Sudweeks, 1999).

$$[Alfa] \alpha = k / (k-1) * [1 - (\sum \sigma^2_i / \sigma^2_T)]$$

k = test üzerindeki madde sayısı

σ^2_i = i madde ölçüm varyansı [ya da bir madde üzerindeki bir grup bireyden elde edilen ölçümlerin varyansı]

$\sum \sigma^2_i$ = i madde ölçüm varyanslarının toplamı

σ^2_T = toplam test ölçümlerinin varyansı

Bulgular: 30 kişiden oluşan homojen ve heterojen özellik gösteren iki ayrı örneklem grubuna ilişkin denencel veriler oluşturulmuştur. Buna göre bu örneklemelere 10 maddelik bir ölçek uygulanmış ve her madde (0) Katılmıyorum; (1) Kısmen Katılıyorum; (2) Katılıyorum şeklinde ölçümlenmiştir. Yapılan hesaplama sonucunda α katsayısı homojen örneklem grubu için negatif ($\alpha = -0,275$); heterojen örneklem grubu için ise pozitif ($\alpha = 0,877$) bulunmuştur.

Sonuç: Eğer bir test, homojen bir gruba verilirse, toplam test ölçümü içindeki değişkenlik azalacak, dolayısıyla alfa katsayısı küçülecek, aynı test daha heterojen bir gruba verilirse toplam test ölçümü içindeki değişkenlik artacak, dolayısıyla alfa katsayısı da büyüyecektir (Arnold, 1996; Helms, 1999). İncelenen denencel örnekten de görüleceği üzere, güvenirlilik (alfa katsayısı) örneklem özelliklerinden güçlü şekilde etkilenmektedir. Bu durumda aynı ölçme aracı [veya test], daha ayrışık ya da daha bağdaşık öğrencilerden oluşan gruplara ya da örneklemelere uygulandığında, birbirlerini onamayan ölçümler güvenirliliği ortaya çıkacaktır (Thompson, 1994). Tüm bu bulgular, güvenirliliğin, örnekleme, dolayısıyla örneklemden elde edilen ölçümlere bağlı olduğunu göstermektedir (Guthrie, 2000). Buradan, güvenirliliğin testlerin değil, eldeki verilerin veya ölçümlerin bir özelliği olduğu (Thompson, 1999) ifade edilebilir. Buradan hareketle güvenirliliği ölçümlerin değil de, testlerin bir özelliği olarak kabul etmek ve “test güvenilirliktir”/“testin güvenirliliği” ifadelerini kullanmak doğru olmayacaktır.

Anahtar Sözcükler: Güvenirlilik; Cronbach'ın Alfa Katsayısını; Ölçüm.

Kaynaklar

- Arnold, M. E. (1996). Influences on and limitations of classical test theory reliability estimates. (ERIC Document Reproduction Service No. ED 395 950).
- Bademci V.(2008). Araştırmalarda Ölçme İle İlgili Bazı Büyük Hataları Düzeltmek Ve Eğitimde Yeniden Yapılanmayı Sürdürmek: Güvenirlilik, Testlerin Bir Özelliği Değildir. Gazi Üniversitesi Endüstriyel Sanatlar Eğitim Fakültesi Dergisi Sayı: 22, 2008.
- Bademci V.(2011).Türk Eğitim Ve Biliminde Bilimsel Devrim: Testler Ya Da Ölçme Araçları Güvenilir Ve Geçerli Değildir. Dicle Üniversitesi Ziya Gökalp Eğitim Fakültesi Dergisi, 16 (2011) 116-132.
- Helms, L. S. (1999). Basic concepts in classical test theory: Tests aren't reliable, the nature of alpha, and reliability generalization as meta-analytic method. (ERIC Document Reproduction Service No. ED 427 083).
- Guthrie, A. C. (2000). A review of coefficient alpha and some basic tenets of classical measurement theory. (ERIC Document Reproduction Service No. ED 438 307).
- Linn, R. L. ve Miller, M.D. (2005). Measurement and assessment in teaching. (Ninth Edition). Upper Saddle River, New Jersey: Pearson.
- Reinhardt, B. (1996). Factors affecting coefficient alpha: A mini monte carlo study. Advances in Social Science Methodology, Volume 4. Ed. B. Thompson. Greenwich, Connecticut: JAI.
- Thompson, B. (1999). Five methodology errors in educational research: A pantheon of statistical significance and other Faux Pas. Advances in Social Science Methodology, Volume 5. Ed. B. Thompson. Stamford, Connecticut: JAI.
- Worthen, B. R., White, K. R., Fan, X ve Sudweeks, R. R. (1999). Measurement and assessment in schools. (Second Edition). New York: Addison Wesley Longman.

Yaşam Analizlerinde Esnek Parametrik Bir Yaklaşım: Royston-Parmar Modelleri
Doğukan ÖZEN¹, Hülya ÖZEN², Pınar AMBARCIOĞLU¹, Ufuk KAYA¹, Erman GÜLENDAG¹, İ. Safa GÜRCAN¹

¹ Ankara Üniversitesi, Veteriner Fakültesi, Biyoistatistik Anabilim Dalı, 06110, Dışkapı, Ankara

² Eskişehir Osmangazi Üniversitesi, Tıp Fakültesi, Biyoistatistik Anabilim Dalı, 26480, Eskişehir

Sansürlü veriler içeren sağkalım verilerinin modellenmesi için sıklıkla Cox orantılı hazard regresyonundan yararlanılmaktadır. Bununla birlikte, bu tür veriler için parametrik modellerin kullanılmasının da bazı avantajları bulunmaktadır. Örneğin, Cox modeli kurulurken orantılı olmayan hazard durumunun görülmesi ile yaşanabilecek potansiyel bir zorluk, parametrik modeller yardımıyla daha kolay bir şekilde ele alınabilir. Ayrıca hazard fonksiyonunun görselleştirilmesi de daha kolay olacaktır. Log kümülatif hazard ve log kümülatif odds temel başarısızlık fonksiyonlarının kübik eğriler kullanılarak düzgünleştirilmesinde Weibull ve log-lojistik modellerin uzantıları önerilmektedir. Bu noktada, Weibull, log-lojistik ve log-normal modellerin bir uzantısı niteliği taşıyan ve birçok uygulama alanına sahip olan Royston-Parmar model ailesinin kullanılması başarılı sonuçlar gözlemlenmede önemli bir tercih olabilir. Modelde kovariyetlerin etkisinin uygun ölçekte (hazard oranı, başarısızlık olasılığı veya başarısızlık olasılığının probiti gibi) orantılı olduğu varsayılmaktadır. Bunun yanı sıra genellikle uygulamada sınırlayıcılık yaratabilen $\ln t$ formunda yaşam süresi dönüşümlerine ihtiyaç duyulmamaktadır. Model, doğrusallık üzerindeki sınırlamaları kaldırıp doğrusal olmayan fonksiyonların kullanımına da izin vermektedir. Bu çalışmada modellenen yaşam fonksiyonlarının dağılımlarında esneklik sağlayan ve yaşam analizleri için parametrik bir yaklaşım olan Royston-Parmar modellerinin bir uygulama ile tanıtılması amaçlanmaktadır. Bu doğrultuda, Royston-Parmar modellerinin pratik kullanışlılığı ve uygulamaya yönelik kolaylıkları örnek bir veri seti de kullanılarak ele alınmıştır.

Anahtar Sözcükler: Hazard fonksiyonu, Royston Parmar modelleri, sağkalım analizi

Kırılganlık İndeksi Kullanılarak Randomize Kontrollü Çalışma Sonuçlarının Güvenilirliğinin Değerlendirilmesi

Didem DERİCİ YILDIRIM¹, **Bahar TAŞDELEN¹**

¹ Mersin Üniversitesi, Biyoistatistik ve Tıbbi Bilişim Anabilim Dalı, MERSİN

Giriş: Randomize kontrollü çalışmalarda tedavinin etkinliğini belirleyen birçok faktör vardır. Belli bir kesim noktasına sahip p değerine bakılarak tedavinin etkili olduğunu söyleyebilmek yıllardır tartışılmakta olan bir konudur. Çünkü p değeri çalışmanın örnek genişliğine ve kaç kişide ilgilenilen durum gözlemlendiğine bakmaksızın benzer değeri vermektedir. Bu durumu örneklendirecek olursak, bir çalışmada A tedavisi alan grupta 100, plasebo grubunda 100 birey olmak üzere, gruplarda sırasıyla 1 ve 9 bireyde miyokard infarktüsü gözlenmiştir. Bu bilgiler doğrultusunda elde edilen çapraz tablo için Fisher exact testi sonucunda p değeri 0,02 olarak bulunmuştur. Başka bir çalışmada ise B tedavisi alan 4000 birey, plasebo grubunda 4000 birey olmak üzere, tedavi grubunda 200 bireyde, plasebo grubunda 250 bireyde miyokard infarktüsü görülmüştür. Elde edilen çapraz tabloda p değeri 0,02 olarak hesaplanmıştır. Her iki çalışma içinde aynı p değerleri elde edilmiştir. Ancak ilk çalışmada tedavi grubunda bir kişide daha miyokard infarktüsü görüldüğü durumda p değeri 0,06 olmaktadır.

Böyle bir durumda, istatistiksel olarak anlamlılık gösteren bir çalışmada ($p < 0,05$), tedavi grubundaki hasta olmayanların en az kaç daha hasta olsaydı istatistik anlamlılık ortadan kalkardı sorusunun yanıtı kırılganlık indeksidir. Bu indeks yardımıyla bir hasta ile değişen p değerlerine güvenerek tedavi etkin demek doğru değildir.

Amaç: Bu çalışmada amaç, klasik bir yaklaşım ve metrik bir ölçüm olan p değerinden yararlanarak değerlendirilen çapraz tablolar için daha doğru bir yaklaşım olan kırılganlık indeksinin tanıtılmasıdır.

Yöntem: Pubmed veri tabanında son beş yıl içerisinde, "The New England journal of medicine", "JAMA", "JAMA internal medicine", "British medical journal", "Lancet (London, England)", "Annals of internal medicine", "Chest", "American journal of respiratory and critical care medicine" ve "The Lancet. Respiratory medicine" gibi alanında en iyi puana sahip dergilerde yayınlanmış, sonuç değişkeni mortalite olan, pediatrik olmayan, tam metni ücretsiz olan, dili İngilizce ve istatistik analiz olarak ki-kare veya Fisher exact test içeren randomize kontrollü çalışmaların sistematik olarak değerlendirilmesi planlanmıştır.

Bulgular: Çalışmaya dahil edilen randomize kontrollü klinik çalışma örnek genişliği, sonuç değişkeni, ilgilenilen durumun gözlemlendiği birey sayısı ve rapor edilen p değerleri bakımından incelenmiştir. Sonuç değişkeni ile ilişkisi araştırılan değişkenler için p değerleri yanı sıra kırılganlık indeksi elde edilmiştir. P değeri anlamlı olmasına karşın çok düşük çıkan kırılganlık indeksine sahip çalışmaların oranı tespit edilmeye çalışılmıştır.

Sonuç: Randomize kontrollü çalışmalarda çapraz tablolar için p değeri ile birlikte kırılganlık indeksinin de verilmesi gerektiği sonucuna varılmıştır.

Anahtar Sözcükler: Randomize kontrollü çalışma, kırılganlık indeksi, çapraz tablo.

Lineer Regresyonda Değişen Varyansın Belirlenmesi ve Giderilmesi

Betül DAĞOĞLU, **Sevinç Püren YÜCEL**, Zeliha Nazan ALPARSLAN
Çukurova Üniversitesi Tıp Fakültesi Biyoistatistik Anabilim Dalı, Adana
sevincpurenyucel@gmail.com

Giriş: Hata terimlerinin sıfır ortalamalı, sabit varyanslı ve ilişkisiz olması lineer regresyon varsayımlarından biridir. Ancak bağımlı değişkeninin ortalama ve varyansının birbiri ile ilişkili olması durumunda sabit varyans varsayımı gözardı edilmiş olur. Bu ilişkililiğe bağlı olarak ortaya çıkan değişen varyans problemi, En Küçük Kareler(EKK) tahmin edicisinin yansızlık varsayımını etkilemez ancak tahmin parametrelerinin varyansında şişmeye neden olur. Bu problem, grafik metotlar ve istatistiksel testler kullanılarak saptanabilir. Değişen varyans varlığında ise Box-Cox Metodu kullanılarak model varsayımları sağlanabilir.

Amaç: Lineer regresyonda, değişen varyansın belirlenmesi için kullanılan yöntemlerin ve bu problemin giderilmesi için uygulanabilecek Box-Cox metodunun değerlendirilmesi amaçlanmıştır.

Yöntem: Çok değişkenli normal dağılım varsayımı altında beş açıklayıcı değişken kendi aralarında ilişkisiz ve hata terimlerinin varyansları sabit olmayacak şekilde üretilmiştir. Box-Cox metodu uygulanarak elde edilen tanımlayıcılık katsayısı (R^2) ve hata kareler ortalaması (MS_{Res}) EKK ile elde edilen değerlerle kıyaslanmaktadır.

Bulgular ve Sonuç: Lineer regresyon varsayımlarını bozduğundan değişen varyansın giderilmesi önemlidir. Değişen varyansı gidermede en sık kullanılan Box-Cox yönteminin uygulanması sonucunda elde edilen regresyon parametrelerinin EKK ile elde edilen regresyon parametrelerinden daha düşük varyansa sahip olduğu ve modelin tahmin performansının daha yüksek olduğu görülmüştür.

Anahtar Sözcükler: Lineer regresyon, değişen varyans, Box-Cox yöntemi, En Küçük Kareler Yöntemi.

Statistical Measures For Heritability

İsmet DOĞAN, Nurhan DOĞAN

*Afyon Kocatepe Üniversitesi Tıp Fakültesi Biyoistatistik ve Tıbbi Bilişim Anabilim Dalı,
 AFYONKARAHİSAR*

Heritability is an important concept in genetics and provides a quantitative measure for the proportion of trait variation that is attributed to genetic variation. It is used in a variety of fields including psychology, behavioral genetics, and breeding. How much phenotypic variation is genetic; and how much of the genetic variation is additive and can be explained by fitting all genetic variants simultaneously in one model? In order to give the answers of the questions were given above, the degree of heritability must be estimated. The heritability coefficient is the numeric index of heritability ranging from 0 (no genetic contribution) to 1 (complete heritability). The estimation of the heritability from the regression of offspring on parents is comparatively straightforward. The data are obtained in the form of measurements of parents – one or the mean of both-and the mean of their offspring. When the mean values of a trait, derived from both parents, is used the heritability is equivalent to the regression coefficient (b),

$$h^2 = b_{(offspring-midparent)} \quad (1)$$

but if the offspring are compared with just one parent the heritability,

$$h^2 = 2 * b_{(daughter-dam \text{ or } full \text{ sibs})} \quad (2)$$

Half-sibs share 25% of their genes. Therefore,

$$h^2 = 4 * b_{(half \text{ sibs})} \quad (3)$$

Among various proposed methods to estimate heritability, variance component methods are advantageous when estimating heritability using markers. Broad sense heritability (h^2) of a trait is the ratio of genetic variability σ_g^2 to phenotypic variability ($\sigma_g^2 + \sigma_e^2$) and is given by

$$h^2 = \sigma_g^2 / (\sigma_g^2 + \sigma_e^2) \quad (4)$$

The analysis is based on calculating variances, and the goal is to find out how much of the variance in phenotype can be attributed to genetic variance. Generally, estimation of variance components is based on analysis-of-variance estimates obtained as outlined in Table 1.

Table 1. Analysis of variance for half-sib data

Source	Degrees of Freedom	Mean Square	Expected Mean Square
Between Bulls	$b - 1$	M_g	$\sigma_e^2 + b\sigma_g^2$
Within Bulls	$N - b$	M_e	σ_e^2
Total	$N - 1$		

Estimates of σ_e^2 and σ_g^2 are $\sigma_e^2 = M_e$ and $\sigma_g^2 = (M_g - M_e)/b$

Interestingly, heritabilities are not constant. Heritability can change over time because the variance in genetic values can change, the variation due to environmental factors can change, or the correlation between genes and environment can change. Genetic variance can change if allele frequencies change (e.g., due to selection or inbreeding), if new variants come into the population (e.g., by migration or mutation), or if existing variants only contribute to genetic variance following a change in genetic background or the environment.

Keywords: Population genetics, breeding, genetic variation

References

Falconer DS. Introduction to Quantitative Genetics. 2nd ed. Essex, England: The Pitman Press; 1981. p. 148-169.

Kim Y, Lee Y, Lee S, Kim NH, Lim J, Kim YJ, Oh JH, Min H, Lee M, Seo HJ, Lee SH, Sung J, Cho NH, Kim BJ, Han BG, Elston RC, Won S, Lee J. On the estimation of heritability with family-based and population-based samples. Biomed Research International 2015. doi: 10.1155/2015/671349
Rudra P, Shi WJ, Vestal B, Russell PH, Odell A, Dowell RD, Radcliffe RA, Saba LM, Kechris K. Model based heritability scores for high-throughput sequencing data. BMC Bioinformatics 2017; 18: 143-158.

Fiber Optik Yansıma Spektroskopi Sistemi ile Tümör Teşhisi: Bir Meta Analiz Çalışması

Pınar GÜNEL KARADENİZ¹, Tuba DENKÇEKEN²

¹SANKO Üniversitesi, Tıp Fakültesi, Biyoistatistik AD, Gaziantep, Türkiye

²SANKO Üniversitesi, Tıp Fakültesi, Biyofizik AD, Gaziantep, Türkiye

Amaç: Bu çalışmanın amacı, fiber optik yansıma spektroskopi sisteminin tümör dokusunu teşhis etmedeki duyarlılık ve seçiciliğini meta analiz ile değerlendirmektir.

Yöntem: Bu çalışmada literatür taraması Ağustos 2017 son tarih olmak üzere “Pubmed” veri tabanı kullanılarak yapıldı. “Kanser veya displazi veya hiperplazi”, “doku veya hasta”, “duyarlılık”, “reflektans veya ışık saçılımı veya fiber optik prob” Anahtar Sözcükleri kullanılarak yapılan tarama sonucunda 541 makaleye ulaşıldı. Bunlardan; vaka raporu, yorum, derleme, mektup türünde olan, anahtar kelime ve özetinde nano, floresans, raman, nanoparçacık, plazmon, ilaç, lazer kelimelerini içeren, insan deneyleri olmayan ve gerekli istatistikleri içermeyen makaleler çalışma dışı bırakılarak toplam 29 adet makale meta analize dahil edildi. R programında “mada” paketi ile teşhis yönteminin doğruluğunu değerlendirmek üzere meta analiz uygulandı. Uygulama için doğru pozitif, yanlış negatif, yanlış pozitif ve doğru negatif sayıları kullanılarak duyarlılık ve seçiciliğe ait “forest” grafikleri oluşturuldu. Tanılayıcı odds oranı hesaplandı ve tanı testleri doğruluğu meta analizi için literatürde önerildiği şekilde rasgele etkiler modeli kullanılarak genel duyarlılık ve seçicilik değerleri elde edildi.

Bulgular: 29 çalışma için rasgele etkiler modeli kullanılarak yapılan meta analiz sonucunda genel duyarlılık ve seçicilik değerleri sırasıyla 0,83 ve 0,84 olarak bulundu. Tanılayıcı odds oranı ise 27,293 (%95 Güven Aralığı:16,011-46,525) olarak elde edildi.

Sonuç: Bu çalışma ile fiber optik yansıma spektroskopi sisteminin tümör dokusunu tespit etme potansiyelinin yüksek olduğu sonucuna ulaşıldı. Bu çalışma sayesinde, literatürdeki çalışmaların sonuçlarından faydalanılarak özet istatistikler elde etmeyi sağlayan meta analizin, tanı testi doğruluğu konusunda da uygulanabileceği gösterilmiştir.

Anahtar Sözcükler: Tanı testi doğruluğu için meta analiz, fiber optik, yansıma, tümör.

Kronik Obstruktif Akciğer Hastalığının Çeşitli Makine Öğrenmesi Yöntemleri Kullanılarak Sınıflandırılması

Ahmet Kadir ARSLAN¹, Songül ÇUĞLAN², Evren KÖSE², Cemil ÇOLAK¹

¹: İnönü Üniversitesi Tıp Fakültesi Biyoistatistik ve Tıp Bilişimi Anabilim Dalı, Malatya

²: İnönü Üniversitesi Tıp Fakültesi Anatomi Anabilim Dalı, Malatya

Özet

Amaç: Bu çalışmada ileri beslemeli tek katmanlı yapay sinir ağları, topluluk öğrenme yöntemlerinden rastgele orman ve yeni geliştirilmekte olan derin öğrenme makinesi modelleri, Kronik Obstruktif Akciğer Hastalığını (KOA) sınıflandırma performansları çeşitli ölçütler kullanılarak karşılaştırılmıştır.

Yöntem: Çalışmada kullanılan veri seti, 40 yaşın üzerinde 107 sağlıklı birey ile KOAH tanısı almış, atak döneminde olan 107 hasta olmak üzere toplam 214 kişiyi kapsamaktadır. Ayrıca hasta ve kontrol gruplarının sınıflandırılmasında grup değişkeni hariç toplam 39 (36'sı nicel (sayısal), 3'ü nitel (kategorik)) klinik değişken kullanılmıştır. Bu değişkenler, deneklere ilişkin demografik değişkenler ile milimetre ve santimetre cinsinden baş ve boyun antropometrik ölçümlerinden oluşmaktadır. Sınıflandırma modelleri olarak rastgele orman, yapay sinir ağları ve derin öğrenme makinesi kullanılmıştır. Bu algoritmaların yanlı ve şansa bağlı sonuç verme olasılıklarını en aza indirmek için 5-katlı çapraz geçerlilik yöntemi kullanılmıştır. Ayrıca ilgili modellerin optimizasyon parametreleri izgara arama (grid search) tekniği kullanılarak optimize edilmiştir. En uygun parametrenin hangisi olduğunun belirlenmesinde doğruluk ölçütü kullanılmıştır. Çalışmada kullanılan modellerin sınıflandırma performanslarının değerlendirilmesinde doğruluk, duyarlılık, seçicilik ve işlem karakteristik eğrisi altında kalan alan (AUROC) ölçütleri kullanılmıştır. Analizler, RapidMiner Studio 7.5.001 Free Edition yazılımı kullanılarak gerçekleştirilmiştir.

Bulgular: Kullanılan sınıflandırma modellerinin performans ölçütleri bazında sonuçları Tablo 1'de verilmiştir.

Tablo 1: Sınıflandırma modellerinin performans metrikleri bazında sonuçları

Performans ölçütleri	Modeller		
	İleri Beslemeli Tek Katmanlı Yapay Sinir Ağları	Rastgele Orman	Derin Öğrenme Makinesi
Doğruluk	0.9534	0.9953	0.9581
Duyarlılık	0.9628	0.9905	0.9446
Seçicilik	0.9442	0.9999	0.9723
AUROC	0.9730	0.9999	0.990

Sonuç: Bulgular incelendiğinde, topluluk öğrenme yöntemlerinden biri olan rastgele orman modelinin dört farklı performans ölçütünde de diğer modellere nazaran daha iyi sınıflandırma performansı gösterdiği tespit edilmiştir. Toplu öğrenme yöntemi olması ve çalışmada kullanılan diğer modellere nazaran çok daha kısa zamanda sonuç vermesi nedeniyle sınıflandırma ve regresyon problemlerinde rastgele orman modelinin kullanılması önerilmektedir.

Anahtar Sözcükler: Kronik Obstruktif Akciğer Hastalığı, Yapay Sinir Ağları, Rastgele Orman, Derin Öğrenme

İstatistiksel Modellemeden Görsele Geçiş: Nomogram

Zeynel Abidin ÖZTÜRK¹, **Seval KUL²**, İbrahim TÜRKBEYLER¹, Azer ABİYEYEV¹

¹ *Gaziantep Üniversitesi, Tıp Fakültesi, Geriatri AD, Gaziantep*

² *Gaziantep Üniversitesi, Tıp Fakültesi, Biyoistatistik AD, Gaziantep*

Amaç: İstatistiksel modelleme tıpta hem gözlemsel hem de deneysel çalışmalarda çok yaygın olarak kullanılmaktadır. Modelleme hastalıkların risk faktörlerinin belirlenmesinden prognozunun tahminine kadar çok geniş bir kullanım alanına sahiptir. Buna karşın modelleme sonuçlarının anlaşılması için ileri bir istatistik bilgisine sahip olmayı gerektirmektedir. Nomogram, modelleme sonucunda anlamlı bulunan değişkenleri görsel bir ölçeğe çevirmektedir, böylelikle karmaşık modellerin anlaşılabilir bir grafiğe dönmesi mümkün olmaktadır. Bu çalışmada bir uygulama ile nomogram kullanımı tartışılmıştır.

Yöntem: Çalışmaya Gaziantep Üniversitesi hastanesi Geriatri polikliniğine Ocak 2016-2017 tarihleri arasında başvuran 193 obez hasta dahil edilmiştir. Obez hastalarda sarkopeni gelişmesi için risk faktörü olabilecek 12 değişken modele dahil ederek backward logistic regresyon analizi uygulanmıştır. Geriye doğru yapılan elemelerden sonra istatistiksel olarak anlamlılığı devam eden değişkenler yardımıyla nomogram grafiği çizilmiştir.

Bulgular: Analizlerde 45 sarkopenik obez, 148 obez ile karşılaştırıldığında vücut yağ oranındaki (OR=1.16 95%CI=1.03-1.31), bel çevresindeki (OR=1.62 %95 GA=1.24-2.11) ve baldır genişliğindeki artışın (OR=1.31 %95 GA=1.05-1.64), ayrıca MMSE skorunun 24 den düşük olmasının (OR=27.30 %95 GA=1.65-45) obez vakalarda sarkopeni gelişmesinde etken olduğu saptandı. Bu değişkenlere ek olarak yüksek ADL skoruna (OR=1.16 95% CI=1.04-1.29) ve 5 ve daha yüksek GDL skoruna sahip olmakta (OR=11.01 95%CI=1.18-86.25) sarkopeni için risk faktörü olarak belirlendi. Bu modelin sensitivite ve spesifite değerleri sırasıyla %87.9, ve %94 tü.

Sonuç: Çok sayıda anlamlı değişken bir araya getirilerek nomogram üzerinde modelin etkin gösterimi yapılmış ve klinik bir soruya cevap bulunmuştur.

Anahtar Sözcükler: nomogram, modelleme, lojistik regresyon

Genelleştirilmiş Additive Poisson Regresyon Modeli Yardımıyla Çevre kirliliği ve Meteorolojik Değişkenlerin Hastane Başvuruları ile İlişkilendirilmesi: Astım Nedenli Acil Başvurularının Modellenmesi

Seval KUL¹, Mehmet KARADAĞ², Saim YOLOĞLU³, Mustafa BOĞAN⁴

¹ *Gaziantep Üniversitesi, Tıp Fakültesi, Biyoistatistik AD, Gaziantep*

² *İnönü Üniversitesi, Sağlık Bilimleri Enstitüsü, Biyoistatistik AD, Malatya*

³ *İnönü Üniversitesi, Tıp Fakültesi, Biyoistatistik AD, Malatya*

⁴ *Şehitkamil Devlet Hastanesi, Gaziantep*

Amaç: Çok sayıda çalışma tarafından hava kirliliği değişkenleri, hava sıcaklığı, nem, basınç ve toz nedeni ile hastane başvuruları arasında ilişki olduğu gösterilmiştir. Bir başka bilinen ise bu değişkenlerin etkisinin gecikmeli olarak (lag effect) özellikle solunumsal hastalıkların alevlenmesi üzerine etkisi olduğudur. Gecikmeli etkinin modele dahil edilmesi multicollinearity sorununu ortaya çıkarması nedeni ile modellemede özel istatistiksel yöntemlere ihtiyaç duyulmaktadır.

Yöntem: Bu çalışmada modelleme için bir öneri olarak Genelleştirilmiş Additive Poisson Regresyon modeli kullanılmıştır. Uygulamada Eylül 2009-Ocak 2014 tarihleri arasında Gaziantep il merkezinde hastanelerin acillerine astım nedeni ile yapılan başvurular ve meteorolojik değişkenler kullanılmıştır.

Bulgular: Modele dahil edilen değişkenlerin standart hatalarının oldukça düşük olduğu gözlenmiştir. Ayrıca Modelleme sonucu bulunan anlamlı ilişkiler literatürle uyumludur. Buda modelin geçerliğini gösteren diğer bir kriter olarak değerlendirilmiştir. Modele başvuru gününün hafta içi ya da hafta sonu olma durumu eklendiğinde modelin log-likelihood değeri belirgin şekilde düşmüştür.

Sonuç: Genelleştirilmiş Additive Poisson Regresyon Modeli yardımıyla çevresel olaylar ile hastalıklar alevlenmesi arasında ilişki kurulabileceği gösterilmiştir.

Network Meta Analizi ve Sağlık Bilimleri Alanında Kullanımı

Pınar AMBARCIOĞLU¹, Mehtap AKÇİL OK², Ufuk KAYA¹, Doğukan ÖZEN¹, Erman GÜLENDAG¹, İ. Safa GÜRCAN¹

¹ *Ankara Üniversitesi, Veteriner Fakültesi, Biyoistatistik Anabilim Dalı, 06110, Dışkapı, Ankara*

² *Başkent Üniversitesi, Sağlık Bilimleri Fakültesi, Beslenme ve Diyetetik Bölümü, 06790, Ankara*

Meta analizi, aynı konu üzerinde birbirinden bağımsız olarak yapılmış gözlemsel ve deneysel çalışmalardan elde edilen sonuçları bir araya getirmek, bu sonuçlardaki değişimi açıklamak, daha güvenilir ve doğru sonuçlar elde etmek için istatistiksel yöntemlerin kullanılması süreci olarak tanımlanmaktadır. Meta analizinde uygulanan niceliksel sentezleme işlemi, bilimsel çalışmalarda uygulanan istatistiksel yöntemlere göre hesaplanan etki büyüklüklerinin tahminine dayanmaktadır. Meta analizi, çalışma sonuçlarını birleştirerek genel bir sonuca ulaşmayı sağlasa da bazı eksik yönleri de bulunmaktadır. Bunlardan bir tanesi meta analizi ile aynı anda yalnızca iki müdahale/tedavi/ilaç/deney grubunun karşılaştırılabilmesidir. Network meta analizi, standart meta analizine ek bir yöntem olarak geliştirilmiş olup, iki veya daha çok müdahale/tedavi/ilaç/deney grubunun karşılaştırılmasına olanak sağlamaktadır. Network meta analizinin temel özelliği; örneğin A ve B tedavi yöntemlerinin karşılaştırılmak istendiği bir meta analizine hem A ve B tedavi yöntemlerinin karşılaştırıldığı çalışmaların dahil edilebilmesi (doğrudan etki), hem de A ile C ve B ile C tedavi yöntemlerinin karşılaştırıldığı çalışmaların (dolaylı etki) dahil etme olanağını sağlamasıdır. Böylece ikili karşılaştırmalar yapan meta-analizler yapmak yerine doğrudan ve dolaylı bulguları kullanarak uygulanan network meta analizi sayesinde tek bir analiz ile tüm yöntemleri aynı anda karşılaştırmak mümkün hale gelmektedir. Bu çalışmanın amacı, Türkiye’de yapılan bilimsel araştırmalarda kullanımı henüz yaygınlaşmamış olan network meta analizini tanıtmaktır. Çalışmada, network meta analizinin teorik alt yapısı, sağlık bilimleri alanında yapılan uygulama örnekleriyle desteklenerek özetlenmiştir.

Anahtar Sözcükler: Etki büyüklüğü, Network meta analizi, Meta analizi

Radyoloji Dergilerinde Yayınlanan Makalelerin İstatistiksel Hatalar Bakımından İncelenmesi

Pınar GÜNEL KARADENİZ¹, Ender UZABACI^{2,3}, Sema ATIŞ^{3,4}, Fisun KAŞKIR KESİN^{3,5}, Fatma Ezgi CAN^{3,6}, İlker ERCAN⁷

¹SANKO Üniversitesi, Tıp Fakültesi, Biyoistatistik AD, Gaziantep, Türkiye

²Uludağ Üniversitesi, Veteriner Fakültesi, Biyoistatistik AD, Bursa, Türkiye

³Uludağ Üniversitesi, Sağlık Bilimleri Enstitüsü, Biyoistatistik AD, Bursa, Türkiye

⁴Haliç Üniversitesi, Tıp Fakültesi, Biyoistatistik ve Tıp Bilişimi AD, İstanbul, Türkiye

⁵Düzce Üniversitesi, Sosyal Bilimler Meslek Yüksekokulu, Sosyal Güvenlik Programı, Düzce, Türkiye

⁶İzmir Katip Çelebi Üniversitesi, Tıp Fakültesi, Biyoistatistik Anabilim Dalı, İzmir, Türkiye

⁷Uludağ Üniversitesi, Tıp Fakültesi, Biyoistatistik AD, Bursa, Türkiye

Amaç: Bu çalışmanın amacı, radyoloji alanında Science Citation Index (SCI) ya da Science Citation Index-Expanded (SCI-E) indekslerinde yer alan dergilerde yayınlanan makaleleri, yapılan istatistiksel yanlışlar bakımından değerlendirmektir. Bu sayede, radyoloji alanında bilimsel yayın yapan araştırmacıların çalışmalarını yaparken istatistik konusunda duyarlı ve dikkatli olmalarını sağlayarak kaliteli bilimsel yayınlar üretilmesine katkıda bulunmaktadır.

Yöntem: SCI ya da SCI-E kapsamındaki dergilerde yayınlanan 79 makale çalışmaya dahil edilmiştir. Makaleler, Thomson Reuters Clarivate Analytics veri tabanında yer alan SCI veya SCI-E kapsamındaki radyoloji dergilerinde 2016-2017 yıllarında yayınlanan makaleler arasından rasgele seçilmiştir. Makaleler; p değerine ilişkin yanlışlar, istatistiksel testlerle ilgili yanlışlar, terminoloji ve yorumlama ile ilgili yanlışlar olmak üzere 3 ana başlık altında toplam 16 hata bakımından incelenmiştir.

Bulgular: 79 makalenin yalnızca 4 (%5) tanesinde istatistiksel bakımdan hiç hataya rastlanmazken; 75 tanesinde en az bir istatistiksel hata mevcuttur. En çok karşılaşılan hata, %65 (n=51) oranı ile “verilerin özetlenmesinde yapılan hatalar” olmuştur. Bunu %37 (n=29) oranı ile “p değerinin yanlış gösterimi” izlemiştir. En az karşılaşılan hata ise %4 (n=3) oranı ile “istatistiksel sembol hataları” olmuştur.

Sonuç: Tıbbın hemen hemen her alanında olduğu gibi radyoloji alanında yapılan bilimsel yayınlarda da istatistiksel hatalar ortaya çıkmaktadır. Bu hataları önlemek amacıyla hem araştırmacıların hem de araştırmaların yayınlandığı dergi editörlerinin üzerine düşen sorumluluklar vardır. Araştırmacılar, temel istatistik konusunda bilgiye sahip olmalıdır ve çalışmanın planlama, analiz, yorumlama ve raporlama gibi tüm aşamalarında bir biyoistatistik uzmanından destek almalıdır. Editörler ise sorumlu oldukları dergilere gönderilen araştırmaları, değerlendirme aşamasında bir biyoistatistik uzmanına göndermelidir.

Anahtar Sözcükler: İstatistiksel hatalar, İstatistiksel inceleme, Radyoloji alanındaki makaleler

K-Ortalama Kümeleme Yönteminde Başlangıç Merkezlerinin Kümeleme Sonuçlarına Etkisi

Meryem GÖRAL YILDIZLI, Zeliha Nazan ALPARSLAN

Çukurova Üniversitesi Tıp Fakültesi, Biyoistatistik Anabilim Dalı, Adana

Giriş: Bilgi sistemleri ve iletişim teknolojilerindeki gelişmelerine bağlı olarak sağlık alanında toplanan manuel veya tıbbi cihazlardan elde edilen farklı formattaki tıbbi veriler dijital ortamda ve kolaylıkla erişilebilecek veri tabanlarında saklanabilmektedir.

Sağlık kurumlarında bu amaçla oluşturulan büyük veri niteliğindeki verileri içeren Hastane Bilgi Yönetim Sistemleri (HBYS) tıp alanındaki araştırmalar için veri madenciliği yöntemlerinin kullanılabileceği en önemli kaynaklardır. Verilerin ulaşılabilir olması kadar veri-bilgi dönüşüm sürecinin iyi yönetilebilmesi de önemlidir. Tıp alanındaki bilimsel çalışmalarda veri madenciliği yöntemlerinden K-ortalama algoritması yaygın olarak kullanılmaktadır.

HBYS verileri üzerinde gerek yönetsel gerekse tıbbi analizlerin elde edilmesinde kullanılan K-ortalama algoritması ile başlangıç parametre değerlerine göre çok farklılık gösteren kümeleme sonuçları bulunabilmektedir. Algoritmanın parametre değerlerinden küme sayısının(k) değişmesi ile oluşturulan kümeleme sonuçlarının farklılık göstermesi beklenen durumdur ancak k sabit tutulduğunda başlangıç merkezlerine bağlı olarak kümeleme sonuçlarındaki farklılığın incelenmesi gerekir.

Amaç: Bu çalışmada HBYS'den alınan bir veri setinde farklı başlangıç merkez değerleri ile analiz edilerek elde edilen kümelerin, niteliksel ve kliniksel sonuçları karşılaştırılarak optimal kümelerin belirlenmesinde başlangıç merkez seçiminin öneminin vurgulanması amaçlanmaktadır.

Yöntem: VitaminB12 değerlerine göre sınıflandırılan verilerin diğer değişkenlerle ilişkisini belirleyen araştırmada HBYS verileri kullanılmıştır. Veri setinde hastalara ait; anabilim dalı, cinsiyet, yaş, vitaminB12 değerleri içeren veriler sistemden oluşturulup, K-ortalama algoritması Weka veri madenciliği programı ile analiz edilmiştir.

Bulgular: Küme sayısını sabit tutarak farklı başlangıç merkezleriyle oluşturulan kümeleme sonuçları incelendiğinde; kümeleme sonuçlarının bazılarında yalnız cinsiyet faktörü etkiliyken, bazılarının da hem cinsiyet hem de yaş faktörünün etkili olduğu gözlenmiştir. Ayrıca VitaminB12 değerleri sınıflandırılma yüzdeleri kümeleme sonuçlarında değişkenlere göre farklılıklar göstermiştir. Kümeleme sonuçlarını değerlendirme kriterlerine göre küme içi hata kareler toplamında (SSE) da farklılıklar bulunmuştur.

Sonuç: Analiz sonuçlarındaki tutarsızlıklar bilimsel çalışmaları geçersiz kılan durumlardır. Çalışmanın sonuçları değerlendirildiğinde, kümeleme sonuçlarına göre birbirinden farklı kararların alınabileceği görülmüştür. Bu kararların gerek yönetsel gerekse kliniksel alanlarda kullanılması, sağlık personelinin tıbbi hatalar, sağlık yöneticilerin yönetsel hatalar yapmasında etken faktörler olabilir. Optimal kümeleme sonuçları için başlangıç merkezlerinin dikkatli seçilmesi gerekir.

K-ortalama kullanım kolaylığı olan bir algoritmadır, ancak genelde başlangıç parametrelerin, özde de başlangıç merkezlerinin seçiminin profesyonelce yapılması algoritmanın kullanılabilişliğini ve kümeleme sonuçlarına güvenilirliğini artırmak açısından önemlidir..

Anahtar Sözcükler : K-ortalama, HBYS, başlangıç parametreleri, başlangıç merkezi.

Three-Dimensional Analysis Of Craniofacial Shape In Obstructive Sleep Apnea Syndrome Using Geometric Morphometrics

Fatma Ezgi CAN^{1,2}, Ilker ERCAN³, Gokhan OCAKOGLU³, Senem TURAN OZDEMIR⁴, Ezgi DEMIRDOGEN ÇETINOGLU⁵, Ahmet URSAVAS⁵

¹*Department of Biostatistics and Medical Informatics, Faculty of Medicine, Izmir Katip Celebi University, Izmir, Turkey*

²*Department of Biostatistics, Institute of Health Sciences, Uludag University, Bursa, Turkey*

³*Department of Biostatistics, Faculty of Medicine, Uludag University, Bursa, Turkey*

⁴*Department of Anatomy, Faculty of Medicine, Uludag University, Bursa / TURKEY*

⁵*Department of Pulmonary Diseases, Faculty of Medicine, Uludag University, Bursa / TURKEY*

Abstract

Aims: Craniofacial disharmony is an important risk factor for obstructive sleep apnea syndrome (OSAS). The general aim of this study was to investigate the craniofacial shape in adults with and without OSAS using geometric morphometric methods. To localize and quantify geometric morphometric differences in craniofacial soft tissue profile among obstructive sleep apnea patients and control subjects were studied.

Methods: Selected 106 adult Turkish subjects, consisting of 50 patients with OSAS diagnosed by polysomnography (PSG) and 56 non-OSAS controls, were studied. Three dimensional craniofacial scanning process were performed on patients subjects on the same day as the PSG. Twelve standart craniofacial landmarks were collected from each subject's 3D craniofacial scan. Geometric morphometric analysis was used to compare the craniofacial shape differences between the OSAS and non-OSAS control groups.

Results: No statistically significant difference in terms of sex was found between the groups ($p=0.318$). However, statistically significant difference was found according to age and BMI ($p<0.001$; $p<0.001$). No statistically significant difference in terms of general shape in face shapes was found between the OSAS group and control group. However local significance differences were found ($p < 0.05$).

Conclusion: In this study, the differences in 3D craniofacial shape in adults Turkish with and without OSAS were investigated using landmark based geometric morphometric methods. There are many studies in the literature investigating craniofacial morphologic features in the OSAS patient group. In these studies, attempts were made to identify craniofacial phenotypic features for the OSAS patient group. We investigated the craniofacial region in this study by considering the mean shape and localize changes. While there was no significant difference between the OSAS subjects and controls in terms of general craniofacial features, some significant local differences were seen.

Key Words: Obstructive sleep apnea, craniofacial, geometric morphometry, three dimensional

Development of Web-Based Software for Acute Coronary Syndrome and a Medical Data Mining Application

Emek GULDOĞAN¹, Julide YAGMUR², Cemil COLAK¹

*¹Inonu University Faculty of Medicine, Department of Biostatistics and Medical Informatics,
Malatya, Turkey*

²Inonu University Faculty of Medicine, Department of Cardiology, Malatya, Turkey

Abstract

Aim: Medical data mining is based on data mining methods and related intelligent methods (e.g., granular computing, neural networks and soft computing) used in medicine. In this research, it was aimed to develop a web-based software and to implement medical data mining on the records of the patients with acute coronary syndrome.

Materials and Methods: The data in this study included retrospective observations recorded in the database from the web-based software developed for Cardiology Department, Turgut Özal Medical Center, Inonu University. PHP (Personal Home Page) programming language and MySQL Database Management System were employed for the development of the web-based software system. Laplace Support Vector Machines (LSVM) was constructed to predict absence or presence of diabetes mellitus in patients with acute coronary syndrome.

Results: A web based software performing data entry, query, delete, update, etc. was developed. As a result of medical data mining application, the accuracy and area under ROC curve with 95% CI were obtained as; 0.9804 (0.9716 - 0.987) and 0.9332 (0.9096 - 0.9567), respectively.

Conclusion: The developed web-based software created a very important infrastructure for implementing medical data mining applications. It was determined that the LSVM model produced very good predictive results to estimate absence or presence of diabetes mellitus in patients with acute coronary syndrome.

Keywords: Diabetes Mellitus; Laplace Support Vector Machine; Medical Data Mining.

Genetik İlişki Çalışmalarında Veri Kalite Kontrolünün Yanlılık Üzerindeki Etkileri

Bahar TAŞDELEN

Mersin Üniversitesi Tıp Fakültesi Biyoistatistik ve Tıbbi Bilişim AD, Mersin

Amaç: Yanlılık, üzerinde çalışılan konu, araştırmanın tipi veya deneme materyali ne olursa olsun, Biyoistatistik biliminde sürekli irdelenen bir alandır. Yanlılıkla baş etme yolları her araştırma tipi için farklıdır. Genetik araştırmalar, her geçen gün hızla artarken, bu tip çalışmalarda gerçek genetik ilişkileri tanımlama gücü, öncelikle verinin kalite kontrolüne bağlıdır. Bu çalışmada, sözü edilen kalite kontrol adımlarının istatistik detaylarının incelenmesi amaçlanmıştır.

Yöntem: Kalite kontrol adımlarının yapılmazsa doğurabileceği sonuçlar, uygulamadaki istatistiksel yaklaşımlar ve kullanılan yazılımlar detaylı olarak incelenmiştir. Popülasyon tabakalanması, örneklerin ilişkili olma durumu, genotipleme oranı, Hardy-Weinberg dengesi ve minör allel frekansının sonuçlara etkileri istatistik detayları ile incelenerek, çözüm önerileri üzerinde durulmuştur.

Bulgular: Genom boyu ilişki (genome-wide association, GWAS) çalışmaları insanlarda hastalığın genetik kökenini belirlemek amacıyla binlerce, tek nükleotid polimorfizmi (single nucleotid polymorphism, SNP) üzerinde, vaka-kontrol, popülasyon veya klinik tabanlı prospektif ve kesitsel çalışma tiplerinde sıklıkla yapılmaktadır. Çok büyük boyutta marker lokus test edildiğinden, çok küçük bir hata bile kötü sonuçlar doğurabilir, bu da yanlış pozitif veya yanlış negatif ilişkilerin ortaya çıkmasına sebep olabilir. Bu nedenle cinsiyet tutarsızlığı, örneklerin ilişkili olma durumu, popülasyon tabakalanması, genotipleme oranı, Hardy-Weinberg dengesi, minör allel frekansı gibi analiz öncesinde üstesinden gelinmesi gereken kalite kontrol adımları vardır. Ayrıca, genetik verinin kalite kontrolüne özel yöntemler ve bu yöntemlerin uygulaması amacıyla kullanılan pek çok yazılım mevcuttur.

Sonuç: Genetik ilişki çalışmalarında sonuçların yanlılığını en aza indirmek ve hastalıkla ilişkili genlerin en yüksek doğrulukla tespit edilebilmesi için öncelikle verinin kalite kontrolünün yapılması, marker sayısı ve örnek genişliği de dikkate alınarak verinin arındırılması şarttır. Tabi ki bu durum genom boyu ilişki çalışmaları ve aday (candidate) gen çalışmaları için farklı boyutlarda incelenmelidir.

Anahtar Sözcükler: Genom boyu ilişki çalışmaları, yanlılık, karıştırıcı etki, kalite kontrol.

Ridge Regresyon için Ayar Parametresi Seçimi

Sevinç Püren YÜCEL, Betül DAĞOĞLU, Zeliha Nazan ALPARSLAN

Çukurova Üniversitesi Tıp Fakültesi Biyoistatistik Anabilim Dalı, Adana

sevincpurenyucel@gmail.com

Giriş: Çoklu iç ilişki, lineer regresyon modelinde açıklayıcı değişkenler arasındaki lineer bağımlılık olarak tanımlanır. Çoklu iç ilişki varlığında En Küçük Kareler (EKK) Yöntemi, regresyon parametreleri için kötü koşullu (ill-conditioned) tahmin ediciler elde edilmesine neden olmaktadır. Kötü koşulluluğun temeli, çoklu iç ilişki durumunda EKK kestiricisinin Gauss-Markov teoremi gereği yansız kestirici olmasına, ancak yansız kestiriciler arasındaki minimum varyansa sahip kestirici olmamasına dayanmaktadır. Bu problem, yanlış tahminler elde etmek için pek çok yöntem mevcuttur. Bu yöntemlerden en sık kullanılanı parametre tahminlerinin daha durağan bir kümesinin elde edilebildiği, Hoerl ve Kennard(1970)'in önerisi olan Ridge Regresyon'dur. Ancak, Ridge tahmin edicilerinin hesaplanmasında uygun ayar parametresi (k) seçimi önem kazanır; zira, k arttıkça hata kareler toplamı (SS_{Res}) artar ve tanımlayıcılık katsayısı (R^2) azalır. Optimal k değerinin belirlenmesinde ise birçok yöntem vardır.

Amaç: Ridge regresyonda kullanılan ayar parametresinin belirlenmesi için önerilen yöntemlerin istatistiksel olarak kıyaslanması amaçlanmıştır.

Yöntem: Çok değişkenli normal dağılım varsayımı altında dört açıklayıcı değişken farklı korelasyon senaryosu ile 1000 tekrarlı üretilmiştir. Hata terimi, sıfır ortalamalı sabit varyanslı olarak normal dağılıma bağlıdır ve yanıt değişkeni, açıklayıcı değişken ve hata terimine göre ilişkilendirilerek modellenmiştir. Çoklu iç ilişkinin parametre tahmini üzerindeki etkisini azaltmak adına kullanılan Ridge regresyonuna bağlı ayar parametresini belirlemeye yönelik yöntemler uygulanıp, ortalama hata kareler büyüklüğüne bağlı olarak kıyaslamalar yapılmaktadır.

Bulgular ve Sonuç: Durağan bir parametre kümesi elde etmede en iyi yöntem, hata kareler ortalaması en düşük Ridge ayar parametresini veren yöntem olarak belirlenmiştir. Genel olarak ulaşılabilecek sonuç, çoklu iç ilişki varlığında EKK tahmin edicilerinden ziyade yanlış tahmin ancak minimum varyanslı önerilen Ridge ayar parametresinin kullanılmasıdır.

Anahtar Sözcükler: Lineer regresyon, çoklu iç ilişki, Ridge regresyon, ayar parametresi.

**Alternative Agreement Approaches in Testing the Agreement of 3 Dimensional
Computerized Tomography Measurements of Styloid Processes of Two Different
Observers and Two Different Time**

Semra ERDOĞAN¹, Kaan ESEN², Yüksel BALCI², Sermin TOK², Gülhan TEMEL¹,
Anıl ÖZGÜR²

¹ *Department of Biostatistics and Bioinformatics, Faculty of Medicine, Mersin University, 33343,
Mersin, Turkey*

² *Department of Radiology, Faculty of Medicine, Mersin University, 33343, Mersin, Turkey*

Abstract

Introduction: Elongated styloid process or calcified stylohyoid ligament causes symptoms described by Eagle known as Eagle's syndrome. However, the attachment point of the SP to the temporal bone may not always be observed clearly and measurements of the SP may vary between radiologists. We aimed to evaluate the intraobserver and interobserver agreement of SP measurements on three-dimensional (3D) CT images between two experienced radiologists by using Concordance Correlation Coefficient (CCC) and Coefficient of Individual Agreement (CIA) agreement statistical methods.

Material and Methods: In this study, contrast-enhanced carotid CT angiography images of 134 patients that have been performed between June, 2015 and December, 2015 were retrospectively evaluated. The length between the attachment point of the SP to the temporal bone and the distal end was measured on 3D CT images which were performed with a 64-slice CT scanner. There was a 3-week period between the replicated measurement of the same radiologist.

Result: The agreement between two radiologists and the replicated measurements of the radiologists themselves were calculated with the help of CIA and CCC agreement indexes. When CIA is used for agreement statistics, it can be said that the two radiologists disagree in the right and left measurements of the individual, (respectively 0.195; 0.281) and when the CCC agreement statistics are used, it can be said that there is a perfect agreement between the measurements of these two radiologists (respectively 0.991; 0.992).

Conclusion: The variance values of between-subject and within-subject should be taken into consideration for each observer in cases of two replicated measurements. If these values show huge differences from each other, CCC_{total} values may be calculated larger than the real value and that the results may mislead the researchers. However, CIA statistics is more stable and when such a case is encountered, researchers are advised that between-subject and within-subject variances are calculated over the differences between the two replicated measurements of each observer. Therefore, σ_B^2 / σ_W^2 rate should absolutely be taken into consideration.

Keywords: inter-observer agreement; disagreement; replicated measurement; styloid processes

LMS Yönteminin Uyum İyiliğinin Ölçülmesinde Kullanılan Yöntemlerin Karşılaştırılması

Batuhan BAKIRARAR, İrem KAR, Atilla Halil ELHAN

Ankara Üniversitesi Tıp Fakültesi, Biyoistatistik Anabilim Dalı, Ankara

Özet

Uyum iyiliğinin ölçülmesi için bugüne kadar birçok farklı yöntem önerilmiştir. Önerilen yöntemlerin hepsinin altında yatan mantık, smoothing eğri tarafından sağlanan uyumu, veriyi ayrı gruplara ayırarak elde edilen uyumla karşılaştırmaktan geçmektedir. Bunun için kullanılan yöntemlerden biri de Cole ve Green tarafından önerilen LMS yöntemidir. LMS yönteminde; M eğrisi ortalama ya da ortanca değeri, S eğrisi değişim katsayısını ve L eğrisi Box-Cox kuvvet dönüşümünü göstermektedir.

Literatürde, uyum iyiliğini ölçmek için çeşitli yöntemler önerilmiştir. Bunlardan en sık kullanılanlar, LR (Likelihood Ratio) testi, Royston'ın Q testleri ve van Buuren'ın worm plot testidir. LR yönteminde Cole ve Green, uyum iyiliğinin Ki-Kare dağılımında verilen serbestlik derecesindeki belli bir artışa karşılık gelen sapmadaki azalmaya göre hesaplanmasını önermiştir. Royston ve Wright tarafından tanımlanan Q testleri, tahmin edilen Z skorlarındaki grup varyasyonları arasındaki rastgele olmamayı test eder. Q testleri konum, ölçek, çarpıklık, basıklık ve normallik testlerini içerir. Van Buuren's Q-Q plot testi ise sıklıkla verilerin belli bir dağılımı takip edip etmediğini belirlemek için kullanılır. Worm plot bir dizi ardışık yaş gurubuna yönelik Q-Q çizimlerinin bir kare dizisinden oluşur.

Bu çalışmada, LMS kullanılarak LR, Q test ve worm plot yöntemlerinin aynı veri seti üzerinde uygulanması ve sonuçlarının karşılaştırılması amaçlanmıştır. Çalışmada kullanılan veri seti 9 yaş grubuna ait toplam 4500 kişiyi (2250 kız, 2250 erkek) içermekte olup, R 3.4.0 programlama dili kullanılarak benzetim ile üretilmiştir. Sonuçların karşılaştırılmasında yöntemlere ait grafikler ve LMS değerleri kullanılacaktır.

Anahtar Sözcükler: LMS, LR, Q plot, Q-Q plot,

Suda Çözünür N-heterosiklik Karben Komplekslerinin Kansere Hücrelerine Karşı İlaç Potansiyellerinin Eğri Uydurma Yöntemlerinden İnhibe Eden Konsantrasyon (IC₅₀) Değeriyle İncelenmesi*

Seyma YAŞAR¹, Mitat AKKOÇ², Tuğba KUL KÖPRÜLÜ³, Şaban TEKİN³, Saim YOLOĞLU¹, Sedat YAŞAR²

¹İnönü Üniversitesi Tıp Fakültesi Biyoistatistik ve Tıp Bilişimi Anabilim Dalı, Malatya

²İnönü Üniversitesi Fen Edebiyat Fakültesi Kimya Bölümü, Malatya

³Gaziosmanpaşa Üniversitesi Fen Edebiyat Fakültesi Biyoloji Bölümü, Tokat

Amaç: Rutenyum kompleksleri, antikanser aktivite çalışmalarında platin'den sonra ön plana çıkan en önemli bileşikler arasında yer almaktadır. Rutenyum komplekslerinin DNA, RNA, proteinler ve kolesterol içeren nükleolipidler gibi farklı biyomoleküllere bağlanabilme (lipofilik ve hidrofilik) özellikleri sebebiyle özellikle antikanser ilaç olarak kullanılabilme potansiyeline sahip önemli moleküller olduğu bilinmektedir (1,2). Bu maddelerin anti kanser aktivitelerinin test edilmesindeki en önemli kriter inhibe edici değer (IC₅₀)dir. IC₅₀, inhibe edici özelliği olduğu bilinen maddenin tam inhibisyonunu sağlayacak konsantrasyonunun yarı değeridir (3).

Gereç ve Yöntemler: Bu çalışmada komplekslerin *in vitro* antikanser aktiviteleri, HeLa (İnsan serviks kanseri) ve C6 (Fare Glioma) hücre hatları üzerinde hücre proliferasyon testi ve BrdU hücre proliferasyon ELISA kiti (Roche) kullanılarak gerçekleştirilmiştir. Graphad Prism 7 deneme sürümü ile M-941, M-936, M-929, M-925 ve cisplatin maddelerinin IC₅₀ değerleri istatistiksel olarak hesaplandı. M-941, M-936, M-929, M-925 ve cisplatin maddelerinin sonuçları IC₅₀ değeri ± Standart hata olarak verildi.

Bulgular: HeLa hücresi için M-941, M-936, M-929, M-925 ve cisplatin kodlu maddelerin IC₅₀ ± standart hata değerleri sırasıyla 51,62 ± 11,2, 57,91 ± 13,74, 11,15 ± 3,92, 22,8 ± 2,78 ve 93,98 ± 12,28 olarak hesaplanmıştır. C6 hücresi için M-941, M-936, M-929, M-925 ve cisplatin kodlu maddelerin IC₅₀ ± standart hata değerleri sırasıyla 96,75 ± 14,27, 139,1 ± 10,49, 11,58 ± 3,79, 24,27 ± 3,61 ve 127,6 ± 13,92 olarak hesaplanmıştır. Son olarak M-941, M-936, M-929, M-925 ve cisplatin maddeleri için HeLa hücre hattı ve C6 hücre hattında IC₅₀ değerlerinin hesaplanmasında uydu edilen eğriler için R² değerleri M-941, M-936, M-929, M-925 ve cisplatin maddeleri için sırasıyla 0,18, 0,07, 0,25, 0,91, 0,65 ve 0,56, 0,93, 0,35, 0,82 ve 0,81 dir.

Sonuçlar: IC₅₀ ve R² değerleri incelendiğinde hem HeLa hücre hattında hem de C6 hücre hattında sentezlenen M-925 kodlu maddenin halen kanser tedavisinde kullanılan cisplatin'den daha iyi antiproliferatif etkiye sahip olduğu bulunmuştur. Bu bağlamda sentezlenen M-925 kodlu maddenin ilaç etkin maddesi olma potansiyelinin olduğu düşünülmektedir

Anahtar Sözcükler: inhibe edici konsantrasyon (IC₅₀), eğri uydurma yöntemleri, cisplatin

Kaynaklar

- 1) Doug A M, Khadijah M H, Matthew J P, Andrew J D, Yang H Y, Wiley J Y. Anticancer activity of Ag (I) N-heterocyclic carbene complexes derived from 4, 5-dichloro-1H-imidazole. Metal-based drugs. 2008;2008.
- 2) Iqbal MA, Haque RA, Nasri SF, Majid AMSA, Ahamed MBK, Farsi E, et al. Potential of silver against human colon cancer: (synthesis, characterization and crystal structures of xlyl (Ortho, meta, & Para) linked bis-benzimidazolium salts and Ag(I)-NHC complexes: In vitro anticancer studies). Chemistry Central Journal. 2013;7:27-.
- 3) Sebaugh J. Guidelines for accurate EC50/IC50 estimation. Pharmaceutical statistics. 2011;10(2):128-34.

(*): Bu çalışma COST 2515- CM1105 aksiyonu kapsamında TÜBİTAK tarafından 114Z036 nolu proje ile desteklenmiştir.

2x2 Tablolar için Kapsamlı Bir Web Yazılımı

Ahu DURMUŞCELEBİ¹, Gökmen ZARARSIZ^{1,2}, Selçuk KORKMAZ^{2,3}, Halil BİLGİN²,
Dinçer GÖKSÜLÜK^{2,4}, Gözde ERTÜRK ZARARSIZ¹, Ferhan ELMALI⁵, Ahmet ÖZTÜRK^{1,2}

¹Erciyes Üniversitesi, Tıp Fakültesi, Biyoistatistik Anabilim Dalı, Kayseri

²Turcosa Analitik Çözümler Ltd. Şti., Erciyes Teknopark, Kayseri

³Trakya Üniversitesi, Tıp Fakültesi, Biyoistatistik Anabilim Dalı, Edirne

⁴Hacettepe Üniversitesi, Tıp Fakültesi, Biyoistatistik Anabilim Dalı, Ankara

⁵Katip Çelebi Üniversitesi, Tıp Fakültesi, Biyoistatistik Anabilim Dalı, İzmir

Özet

Sağlık alanındaki araştırmaların birçoğunda kategorik veri analizi gerçekleştirmek üzere 2x2 çapraz tablolara ihtiyaç duyulmaktadır. Bir hastalığın varlığını ya da yokluğunu belirlemede tanı testlerinin performansını araştırmak, hastalıklar ile ilgisi olduğu düşünülen değişkenler açısından risk kestirimleri yapmak, klinik iki kategorili iki değişken arasındaki neden-sonuç ilişkisini araştırmak, iki değerlendirici arasındaki uyumu araştırmak, sınıflandırma yöntemlerinin performansını değerlendirmek, ya da uzaklık-benzerlik ölçülerini hesaplamak gibi çok sayıda problemin çözümüne yönelik bu tablolar üzerinden istatistiksel hesaplamalar yapılabilmektedir. Literatürde 2x2 tablolar ile hesaplamalar yapmak için çok sayıda yazılım bulunmasına rağmen, bu yazılımlar belli problemlerin çözümüne yönelik geliştirilmiş olup sınırlı sayıda hesaplama yapmaktadır. Bu çalışmada, çok sayıda problemin çözümüne yönelik gerekli istatistikleri hesaplayan ve yorumlayan, kolay erişilebilen ve kullanılabilen, kapsamlı bir web yazılımı geliştirilmesi amaçlanmıştır. Web uygulamasını geliştirmek üzere R programlama dilinin **shiny**, **shinydashboard**, **dplyr** kütüphanelerinden; Javascript dilinin ise **HighCharts.js** kütüphanesinden faydalanılmıştır. Geliştirilen sistem Turcosa Analitik bünyesinde yer alan iş istasyonuna aktarılmış ve kullanıma açılmıştır. Geliştirilen 2x2 yazılımı ile araştırmacılar: (i) 38 adet tanı ölçüsü, (ii) 21 adet ilişki ölçüsü, (iii) 13 adet uyum ölçüsü, (iv) 9 adet hipotez testi, (v) 73 adet uzaklık-benzerlik ölçüsü, (vi) 21 adet risk ölçüsü olmak üzere 6 ana başlık halinde toplam 175 ölçüyü hesaplayabilmektedir. Kullanıcılar veri yüklemeye ihtiyaç duymadan 2x2 web yazılımı üzerinden çapraz tablo gözelerine ilişkin değerleri girerek hızlı ve kolay bir şekilde sonuçlarını yorumları ile birlikte elde edebilmektedir. Kullanıcılar elde ettikleri sonuçları *.csv, *.pdf, *.xls gibi çeşitli formatlarda indirebilir, ayrıca interaktif grafikler ile birlikte raporlayabilirler. 2x2 yazılımına <http://www.opensoft.turcosa.com.tr/2x2/> adresi üzerinden erişebilirler.

Anahtar Sözcükler: 2x2, çapraz tablo, kategorik veri analizi, risk ölçüleri, tanı ölçüleri, uzaklık ölçüleri

Boylamsal Veri Setinde Çok Düzeyli Analiz Yöntem Seçimi ve Bazı Özel Durumlar, Kısıtlılıklar

Alev BAKIR¹, Mustafa Ş. ŞENOC AK²

¹Haliç Üniversitesi, Tıp Fakültesi, Biyoistatistik ve Tıp Bilişimi ABD, İstanbul

² İstanbul Üniversitesi, Cerrahpaşa Tıp Fakültesi, Biyoistatistik ABD, İstanbul
alefbakir@gmail.com, mssenocak@gmail.com

Amaç: Klinikte durumun, prognozunu, sonucunu anlaşılabilirliği, öngörülebilirliği için yapılan çalışmalarda veri toplama işlemi tek bir zamandaki veya önce-sonra düzeyindeki gözlem değerleri ile gerçekleştirilebileceği gibi, aynı veri setine ait farklı zaman dilimlerindeki değerlerden faydalanılarak da incelemelerin detaylandırılması söz konusu olabilmektedir.

Yöntem: Zamana bağlı olarak hem bireyler arası farklılıkları hem de birey içi farklılıkları göz önüne alıp toplanmış veri setleri Boylamsal Data (Longitudinal Data) olarak bilinmekte ve Panel Veri Çözümlemesi ile sabit etkinin, rassal etkinin bulunabildiği modeller ile parametre tahminleri yapılabilmektedir. Bu tip veri setleri Genelleştirilmiş Doğrusal Modeller (Generalized Linear Models-GLM) gibi klasik yöntemler ile çözümlenememekle birlikte istatistikte çok düzeyli (multilevel) analiz yöntemleri ile incelenebilmekte ve çalışma uzun sürdüğü takdirde kayıp data ile daha fazla karşılaşma ihtimaline karşı bazı özel durumlar da söz konusu olabilmektedir.

Bulgular: Kesitsel veri seti ile çalışıldığında ve varsayımlar uygun olduğunda Lineer Regresyon veya Lojistik Regresyon yöntemler, tahmin edilecek değişken sayım yöntemi ile elde edildiğinde Poisson Model, aşırı yayılım söz konusu olduğunda da Negatif Binomial Model kullanılmaktadır. Veri seti Boylamsal Data olduğunda Genelleştirilmiş Lineer Modeller zaman boyutunun eklenmesi ile Ağırlıklandırılmış Genelleştirilmiş Eşitlik Kestirim (Weighted Generalized Estimating Equations-WGEE) (1) ve Genelleştirilmiş Lineer Karma Model (Generalized Linear Mixed-Effects Model-GLMM) (2) yöntemleri kullanılmalıdır. Bu iki model birbirinden, çalışmaya “random effect ($\mathbf{z}_{it}^T \mathbf{b}_i$)” eklenmesi ve devamında model parametrelerinin farklı yorumlanması ile ayırt edilebilir.

$$y_{it}|x_{it} \sim f(\mu_{it}),$$

$$g(\mu_{it}) = \mathbf{x}_{it}^T \boldsymbol{\beta}_i, \quad 1 \leq i \leq n, 1 \leq t \leq T \quad (1)$$

$$y_{it}|x_{it}, z_{it}, \mathbf{b}_i \sim f(\tilde{\mu}_{it}),$$

$$g(\tilde{\mu}_{it}) = \mathbf{x}_{it}^T \tilde{\boldsymbol{\beta}} + \mathbf{z}_{it}^T \mathbf{b}_i, \quad \mathbf{b}_i \sim N(0, \Sigma_b), \quad 1 \leq i \leq n, 1 \leq t \leq T \quad (2)$$

Sonuç: Boylamsal data ile çalışıldığında çok düzeyli analiz yöntemi seçiminde GLMM ‘nin GLM ‘nin parametrik, WGEE ‘nin semi-parametrik bir uzantısı olduğu doğru sonuç yorumlama açısından unutulmamalıdır. Ayrıca bu modeller klinikte boylamsal datadaki amaç değişkeninin farklı değişken dağılımlarında da (Bernoulli, Poisson), karşılaştırılması muhtemel olan kayıp datanın çok olması durumunda da çalışılabilmesine imkan sağlamaktadır. Bu iki tip yaklaşım birbirinden oldukça farklı sonuçlar doğurabileceği için model seçimi, değişken sayısı, randomness gibi durumlarda daha da bir önem kazanmaktadır.

Anahtar Sözcükler: Boylamsal Data, Çok Düzeyli Analiz Yöntemleri, Genelleştirilmiş Lineer Karma Model, Ağırlıklandırılmış Genelleştirilmiş Eşitlik Kestirimi

**Sınıflama ve Regresyon Modellerinin Performanslarının Değerlendirilmesi. Bir
Veri Seti Üzerine Uygulama**

Hülya BİNOKAY, Yaşar SERTDEMİR

Çukurova Üniversitesi Tıp Fakültesi Biyoistatistik Anabilim Dalı, Adana
hbinokay@hotmail.com

Özet: Sağlık bilimlerinde sınıflama amaçlı yapılan çalışmalar, daha çok belirli bir hastalığın tanısının konulmasında etkili olan risk faktörlerinin, istatistiksel olarak modellenerek belirlenmesi şeklinde yapılmaktadır. Bu modeller tahmin edici ve tanımlayıcı modeller olarak ikiye ayrılmaktadır.

Tahmin edici modellerin amacı sonuçları bilinen verilerden bir model geliştirmek ve bu model yardımıyla sonuçları bilinmeyen yeni veri kümelerinin sonuç değerlerini tahmin etmektir. Tahmin edilecek değişken eğer sürekli ise tahmin problemi regresyon, kategorik ise sınıflama problemi olarak adlandırılmaktadır. Sınıflama ve regresyon modelleri, tahmin edici modeller içinde en yaygın kullanıma sahiptir. Bu modeller içinde en çok kullanılan yöntemler Lojistik Regresyon, Karar Ağaçları, Topluluk Yöntemleri (bagging, boosting, random forest), Destek Vektör Makinesi, K en yakın komşu ve Naive Bayes dir. Bu yöntemlerde, modeller belirlenirken farklı algoritmalar kullanılmaktadır. Bu algoritmaların başarı sıralaması veri setine göre farklılık gösterebilmektedir.

Amaç: Lojistik regresyon, Karar ağacı ve Random Forest yöntemleri ile analiz edilen yeni doğan sepsis verilerinde üç yöntemin tanı ile ilgili performanslarının değerlendirilmesidir.

Yöntem: Veriler %70 eğitim ve %30 test veri setleri olarak ikiye ayrılacak. Eğitim veri setine Lojistik regresyon, Karar Ağacı ve Random Forest yöntemleri uygulanarak, test seti için doğruluk, eğri altında kalan alan, duyarlılık, seçicilik değerleri 1000 kez tekrar edilecektir. Analizler R 3.3.3 ile gerçekleştirilmiştir.

Bulgular: Sepsis verisi üzerinde yapılan çalışmalarda doğruluğun(Accuracy) 3 yöntem içinde birbirine yakın olduğu, Lojistik regresyon yönteminde ise etkileşim terimlerinin de bulunduğu modelde duyarlılığının diğer yöntemlere göre daha iyi olduğu gözlemlenmiştir. Random Forest ve Lojistik Regresyon yöntemlerinde seçiciliğin daha iyi olduğu, eğri altında kalan alanlarda ise Lojistik regresyon ve Random Forest yöntemlerinin daha iyi sonuç verdiği gözlemlenmiştir.

Sonuç: Bu veri seti için Lojistik Regresyon ve Random Forest yöntemleri Karar Ağaçlarına göre daha iyi performans gösterdiği saptanmıştır.

Anahtar Sözcükler: Karar Ağacı, Random Forest, Lojistik Regresyon

İşgücüne Dâhil Olma Durumunu Etkileyen Faktörlerin Kategorik Regresyon Analizi ile İncelenmesi

Sarp SAKLIYAN¹, Mehmet KIVRAK², Cemil ÇOLAK¹

¹ İnönü Üniversitesi Tıp Fakültesi Biyoistatistik ve Tıp Bilişimi AD, Malatya, Türkiye.

Özet

Amaç: Kategorik Regresyon analizi, bağımsız değişkenin kategorik olduğu durumlarda kullanılabilen ve optimal ölçeklendirme ile değişkenler arasında hem doğrusal hem de doğrusal olmayan ilişkileri incelemeyi sağlayan bir yöntemdir. Ekonomik ve sosyal göstergeler (nüfus ve göç hareketliliği, enflasyon, hane halkı büyüklüğü, ücret politikası vb.), istihdamı etkileyen göstergelerdir ve kategorik ölçeklendirmeye dayalı istatistiksel modellemeler ile analizleri gerçekleştirilebilir. Bu araştırma, optimal ölçeklendirmeye dayalı kategorik regresyon analizi ile işgücüne dahil olma durumunu etkileyen faktörlerin belirlenmesini amaçlamaktadır.

Yöntem: Türkiye İstatistik Kurumu (TÜİK) hane halkı İşgücü verilerinin 2016 yılına ait son altı aylık kayıtları üzerinden 2463 hane verisine kategorik regresyon analizi uygulanmıştır. Bu çalışmada, işgücüne dâhil olma durumu bağımlı değişken, cinsiyet, göç, eğitim düzeyi, gelir düzeyi, hanehalkı büyüklüğü, çalışılan sektör, enflasyon ve işsizlik beklenti endeksi ise bağımsız değişkenler olarak modelde dikkate alınmıştır.

Bulgular: Öncelikle kurulan modelin anlamlılığına ve bağımsız değişkenlerin bağımlı değişkendeki değişimi (R^2) ne ölçüde açıkladığına bakılmıştır. Model anlamlılığı sağlanmış; fakat bağımlı değişkendeki değişimi açıklama oranı ($R^2=0.115$) düşük çıkmıştır. Standartlaştırılmış katsayılar incelendiğinde; yaş değişkeni (0.236) modele en fazla pozitif yönde katkı sağlarken, enflasyon (-0.190) ve işsizlik beklenti endeksi (-0.289) modele negatif yönde en fazla katkı sağlamışlardır. Modele etki eden değişkenler üzerine yoğunlaşarak, uç değerlerin tespiti ve verilerde kategorileşme sağlanarak veri yapısına uygun optimum ölçeklendirme yapılmıştır. Yaş verisi üzerinde kısıtlamaya gidildiğinde, ölçeklendirme sonrası model anlamlılığı sağlanarak, bağımsız değişkenlerin bağımlı değişkendeki değişimi açıklama oranında ($R^2=0.474$) maksimum artış sağlandığı görülmüştür. Optimum ölçeklendirme sonrasında standartlaştırılmış katsayılar incelendiğinde; çalışılan sektör değişkeni (0.698) modele en fazla pozitif yönde katkı sağlarken, enflasyon (-0.189) ve işsizlik beklenti endeksi (-0.293) modele negatif yönde en fazla katkı sağlamışlardır. Etki katsayısı hesaplanarak, bağımsız değişken kategorilerinin bağımlı değişkenle ne yönlü bir ilişki içerisinde olduğunu tahmin edilmeye çalışılmıştır. Hane halkı büyüklüğünün 3.2 (0.310) [değişken kategorisi (etki katsayısı)] ve 3.3 (0.295) olan hanelerin işgücüne olumlu etkileri olduğunu, 3.4 (-0.147) ve 3.5 (-0.196) olan hanelerin ise işgücüne ters etkileri olduğu ve hane halkı büyüklüğü arttıkça işgücü durumu negatif yönde etkilenmiştir.

Sonuçlar: Önerilen yaklaşım, optimal ölçeklendirme sayesinde veri yapısına uygun seçimler türeterek, en uygun modellemeyi elde etme imkanı sağlamış ve kategorik değişkenlerin olduğu veri setlerinde oldukça başarılı tahmin sonuçları verdiği gözlemlenmiştir.

Anahtar Sözcükler: İşgücüne dâhil olma durumu, kategorik regresyon analizi, optimal ölçeklendirme.

Teşekkür: Bu çalışma, İnönü Üniversitesi Bilimsel Araştırma Projeleri Koordinasyon Birimi tarafından 2016/145 nolu proje ile desteklenmiştir.

Tıp Fakültesi Öğrencilerine Verilen Biyoistatistik Eğitiminde Yeni Yaklaşımlar

Ferhan ELMALI

*İzmir Kâtip Çelebi Üniversitesi, Tıp Fakültesi, Biyoistatistik Anabilim Dalı, İzmir
ferhan.elmali@ikc.edu.tr*

Özet

Biyoistatistik dersi ülkemiz ve dünya tıp fakültelerinin çoğunda birinci sınıfta verilmektedir. Öğrencinin ilk yılında kliniklerle ve araştırmalarla tanışmamış olması, Biyoistatistik kavram ve yöntemlerini hekimlikle bağdaştıramaması dersin gerekliliğinin ve öneminin anlaşılmamasına yol açmaktadır. Dersin kalabalık sınıflarda formüllere dayalı olarak verilmesi, uygulamalarda kendileri tarafından toplanmayan gerçek verilerin ya da eğitici tarafından oluşturulan sanal verilerin kullanılması öğrencinin derse ilgisini azaltmaktadır. Bunların sonucu olarak öğrenci, Biyoistatistik eğitiminin sadece akademisyen ve uzmanlık yapan asistanlar için gerekli olduğuna, pratisyen hekimler için bu eğitimin gerekli olmadığına inanmaktadır.

Tıp fakültelerinde Biyoistatistik eğitiminin gerekliliğinin öğrenciler tarafından anlaşılabilmesi için bilim insanları tarafından sorunlar tespit edilmeye çalışılmış ve farklı çözüm önerileri sunulmuştur. Amerikan Medikal Sağlık İstatistikçileri Eğitim Kurulu (1953) yayınladığı raporda istatistiksel yöntemlerin tüm tıp öğrencilerine verilmesi gerektiğini ancak bu yöntemlerin bir dizi hazır teknik olarak değil sorunları aydınlatan bir süreç olarak öğretilmesini bildirmektedirler. Hopkins (1958), Amerika Birleşik Devletleri ve Kanada'daki 94 tıp fakültesinde yapmış olduğu ankette Biyoistatistiğin detaylı istatistiksel tekniklerin verildiği bir ders olmasından ziyade bilimsel yöntemin genel bir aracı olması gerektiği sonucuna ulaşmıştır. Colton (1975) öğrencilerin "Biyoistatistik eğitiminin temel amacının tıp literatürünün eleştirel okunmasını sağlamak" olduğunu düşündüklerini bildirmiştir. Hogg (1991) öğrencilerden gerçek olmasına rağmen 'eski' veriler üzerinde çalışmalarını istemek yerine, kendi verilerini bulmalarını veya üretmelerini sağlamanın daha iyi olduğunu belirtmiştir. Smith (1998) büyük sınıflarda verilen eğitimin verimini arttırmak için sınıfların takımlara ayrılmasını ve her takıma ödev verilmesini önermiştir. Ercan ve ark. (2008) öğrenciler tıp eğitimi süresince literatürü takip etmedikleri için Biyoistatistik dersinde edindikleri bilgileri hemen unuttuklarını ifade etmiştir. Sami (2010) Biyoistatistik derslerinde formüller, elle hesaplamalar ve resmi sunumların dersin sıkıcı hale gelmesine neden olduğunu ifade etmiştir. Miles ve ark.(2010) Biyoistatistiğin pratisyen hekimlikte de gerekli olduğunu kavratmak için örneklerin günlük hayattan verilmesini önermiştir. Shakeri (2016) tıp fakültelerindeki Biyoistatistik müfredatlarının bilgisayar ve yazılım uygulamalarına dayalı olarak tasarlanmasını önermiştir. Evans ve ark.(2016) güncel bir eğitim modeli olan "flipped classroom" modelini Standford Üniversitesi Tıp Fakültesi'nde uygulamışlar, uyguladıkları bu eğitim modelinin geleneksel modele göre üstün olduğunu göstermişlerdir. Schwartz ve ark. (2016) "flipped classroom" modelini Biyoistatistik eğitime uyarlayarak esnek bir model olan "Flipped Biostatistics Classroom" modelini geliştirmişlerdir. Fong ve ark. (2017) wiki olarak isimlendirilen ve sosyal medya üzerinden toplantılar düzenleyerek verilen Biyoistatistik dersinin klasik yönetime göre öğrenci notlarında daha iyi sonuç verdiğini göstermişlerdir.

Gelişen teknoloji, geliştirilen yeni öğrenme modelleri, değişen öğrenci profilleri ve toplum ihtiyacı göz önüne alınarak ülkemiz tıp fakültelerindeki Biyoistatistik eğitiminin uluslararası düzeyde verilmesinin öğrencinin derse olan yaklaşımını olumlu yönde etkileyeceğini düşünmekteyiz.

Anahtar Sözcükler: Biyoistatistik eğitimi, tıp fakültesi, aktif öğrenme, flipped classroom.

Kaynaklar:

- 1- Fertig, J.W., Badger, G.F., Hopkin, C.E., Merrell, M., Puffer, R.R., Shackelford, M.F., Yerushalmy, J., "Training in Medical and Public Health Statistics", *American Journal of Public Health Year Book*, 1953; 43(2):129-134.
- 2- Hopkins, C.E., "Biostatistics Instruction in Medical Schools", *Journal of Medical Education*, 1958; 33(4): 370-372.
- 3- Colton, T., "An Inventory of Biostatistics Teaching in American and Canadian Medical Schools.", *Journal of Medical Education*, 1975; 50: 594-604.
- 4- Hogg, R. V., "Statistical Education: Improvements are Badly Needed" *The American Statistician*, 1991; 45: 342-343.
- 5- Smith, G., "Learning Statistics By Doing Statistics", *Journal of Statistics Education*, 1998; 6(3): <https://www2.amstat.org/publications/jse/v6n3/smith.html#hogg>.
- 6- Ercan, I., Ozkaya, G., Ocakoglu, G., Yazici, B., Sezer, A., Ediz, B., Kan, I. "Determining Biostatistics Knowledge of Students and Physicians in Medical School." *InterStat*, 2008; 3: 1-17.
- 7- Sami, W., "Biostatistics Education for Undergraduate Medical Students.", *Biomedica*, 2010; 26:80-84.
- 8- Miles, S., Price, G.M., Swift, L., Shepstone, L., Leinster, S.J., "Statistics Teaching in Medical School: Opinions of Practising Doctors.", *BMC Medical Education*, 2010; 10 (1): 75.
- 9- Shakeri, N. "Effectiveness of Some Educational Methods and Tools on Improving the Level of Understanding of Biostatistics among Medical Students and Paramedical Postgraduate Students.", *Global journal of health science*, 2016; 11(8): 302.
- 10- Evans K.H., Thompson A.C., Colin O'Brien, C., Bryant, M., Basaviah, P., Prober, C., Popat, R.A., "An Innovative Blended Preclinical Curriculum in Clinical Epidemiology and Biostatistics: Impact on Student Satisfaction and Performance.", *Academic Medicine*, 2016; 91(5) :696-700.
- 11- Schwartz, T.A., Andridge, R.R., Sainani, K.L., Stangle, D.K., Neely, M.L., "Diverse Perspectives on a Flipped Biostatistics Classroom.", *Journal of Statistics Education*, 2016; 24(2) :74-84.
- 12- Fong, S.S.M., Chu, S.K.W., Lau, W.W.F., Doherty, I., Hew, K. F., (2017). "Incorporating wiki technology in a traditional biostatistics course: Effects on university students' collaborative learning, approaches to learning and course performance." *Journal of Information Technology Education: Research*, 2017;16: 285-299.

Adölesan Dönemde Alkol Tüketimi Risk Faktörlerinin Belirlenmesi ve Risk Tanı Sistemi Modeli

Zeliha AYDIN KASAP¹, Burçin KURT¹, Kemal TURHAN¹

1- *Karadeniz Teknik Üniversitesi Tıp Fakültesi, Biyoistatistik ve Tıp Bilişimi Anabilim Dalı, Trabzon*

Özet

Amaç: Adölesanlarda alkol tüketimi dünyada önemli bir sorun haline gelmiştir. Çalışmanın amacı, adölesanların alkole yönelme sebeplerini belirlemek ve bu konuda eğitim kurumları ile ebeveynleri destekleyen koruyucu yaklaşımlar geliştirmektir.

Yöntem: Yapılan çalışmada, uluslararası bir veri tabanındaki 395 adölesana ait 33 parametre istatistiksel metotlarla analiz edilmiştir. Veri setindeki 395 kayıttan 353 ü alkol tüketimi risk grubunda, 42 tanesi ise risk grubunda değil şeklinde belirtilmiştir. Özellik seçimi sonrası elde edilen anlamlı parametrelere, karar ağacı ve Naive Bayes yöntemleri uygulanarak, adölesanların alkol tüketimine ilişkin risk grubunda olup olmadığına karar veren otomatik bir koruyucu tanı sistemi geliştirilmiştir. Çalışma R Studio programı ortamında gerçekleştirilmiştir.

Bulgular: Yapılan çalışmada cinsiyet, yaş, yerleşim yeri türü, devamsızlık süresi, dışarı çıkma sıklığı, gezme süresi, ders çalışma süresi, serbest vakit, sınıf tekrarı sayısı olmak üzere 9 parametre, adölesanlarda, alkol tüketiminde anlamlı bulunmuştur ($p < 0,05$). Bu parametrelerle geliştirilen Naive Bayes algoritması modelinin duyarlılığı %100, özgüllüğü %91.41, doğruluk değeri %91.8, Kappa istatistiği 0.48 ve AUC (ROC eğrisinin altında kalan alan) değeri ise yaklaşık 0.96 iken karar ağacı yöntemi ile bu değerler sırasıyla %100, %99.67, %99.68, %96 ve %99.83 olarak elde edilmiştir.

Sonuç: Çalışmada elde edilen sonuçlara göre karar ağacı yöntemi ile geliştirilen modelin, Naive Bayes algoritmasına göre geliştirilen modelden daha başarılı olduğu görülmüştür. Böylece adölesanların alkol tüketimi risk grubunda olup olmadığına ilişkin koruyucu amaçlı bir tanı sistemi geliştirilmiştir. Buna göre, adölesanlardan belli zamanlarda anket yardımıyla önerilen parametre bilgileri alınarak, bir adölesan risk grubuna girdiği anda ebeveyn ve eğitim kurumlarını uyarabilen koruyucu bir sistem olarak kullanılabilir. Bu şekilde risk grubundaki adölesanın alkol tüketimine ilişkin takibi daha iyi yapılabilir ve erkenden önleyici bir yaklaşım izlenebilir.

Anahtar Sözcükler – adölesan, alkol tüketimi, koruyucu tanı sistemi, karar ağacı, naive bayes

**Bulanık Çıkarsama Sistemleri ile Lojistik Regresyon Sınıflandırıcılarının
Performanslarının Apandisit Verisinde Karşılaştırılması**

S. Kenan KÖSE, İrem KAR, Batuhan BAKIRARAR

Ankara Üniversitesi Tıp Fakültesi, Biyoistatistik Anabilim Dalı, Ankara

Özet

Sınıflandırma, veri sınıflarını tanımlayan ve ayıran bir model (veya fonksiyon) bulma işlemidir. Model, eğitim verilerinin (yani sınıf etiketlerinin bilindiği veri nesnelerinin) analizine dayanarak türetilir ve test verilerinin (sınıf etiketinin bilinmediği nesnelerin) sınıf etiketini tahmin etmek için kullanılır. Bu çalışmada, bulanık küme teorisine dayalı uzman sistemler olarak tanımlanan bulanık çıkarsama sistemleri ile lojistik regresyon sınıflandırıcılarının performansları karşılaştırılacaktır. Değerlendirmelerde, sınıflandırma performans ölçütleri olarak doğruluk, kesinlik, F-ölçütü ve duyarlılık ele alınacaktır.

Çalışmada, Knowledge Extraction based on Evalutionary Learning (KEEL) veritabanında yer alan yedi bağımsız sürekli değişkene sahip 106 kişilik örneklemde oluşan apandisit verisi kullanılmıştır. Çalışma kapsamında yapılan analizler sonucunda, doğruluk değerleri sırasıyla Bulanık Çıkarsama Sistemleri ve Lojistik Regresyon için 0.813 ve 0.781 olarak bulunmuştur. Verilerin analizi R 3.3.2 programlama dili kullanılarak yapılmıştır.

Anahtar Sözcükler: Bulanık Çıkarsama Sistemleri, Lojistik Regresyon, Tahmin

Bağımlı Değişken İle Ortak Değişken Arasındaki Korelasyona Göre Kovaryans Analizinin Gücü

Emre DEMİR¹, Yasemin Yavuz², Serdal Kenan KÖSE²

¹Hitit Üniversitesi, Tıp Fakültesi, Biyoistatistik Anabilim Dalı, Çorum

²Ankara Üniversitesi, Tıp Fakültesi, Biyoistatistik Anabilim Dalı, Ankara

Özet

Amaç: İstatistiksel araştırmalarda nedensellik bağıntısı araştırılırken deney sonunda ulaşılan bulguların araştırılan etkenin dışında başka bir faktöre bağlı olup olmadığı sonucuna sonuç değişkenli ile ilişkili diğer değişkenlerin etkilerinin arıtılması ile ulaşılabilir. Bu işlem istatistiksel analiz öncesinde (randomizasyon, eşleştirme, kısıtlama) veya analiz evresinde (tabakalandırma, kovaryans analizi) yapılır. Kovaryans analizi bir çalışmada etkisi incelenen faktör haricindeki bağımlı değişken ile korelasyon bulunan sürekli bir veya birden fazla değişkenin (covariate-ortak değişken) etkisinin arıtılarak istatistiksel olarak kontrol edilmesini sağlayan bir yöntemdir. Kovaryans analizinde bağımlı değişkenler ile ortak değişkenler arasındaki ilişkiden dolayı, ortak değişkenlerin bağımlı değişken üzerindeki etkisi arındırıldıktan sonra bulgular değerlendirilir. Kovaryans Analizi grup içi hata varyansını küçülterek açıklanamayan hata varyansının bir kısmını deneye dâhil edilen diğer ortak değişkenler yoluyla açıklayarak hata varyansını azaltır ve bağımsız değişkenin etkisinin daha kesin olarak kestirilmesini sağlar. Ortak değişkenden kaynaklanan değişim ölçülebilir ve hata varyansından ayırt edilebilir dolayısıyla hata varyansının azaltılması ile testin gücü artar. Kovaryans analizinin önemli varsayımlarından bir tanesi ortak değişken ile bağımlı değişken arasında doğrusal bir korelasyon olmasıdır. Bu çalışmada bu korelasyonun düzeyine göre kovaryans analizinin gücünün nasıl değiştiği bir simülasyon çalışması ile araştırılmıştır.

Yöntem: Bağımlı değişken ile ortak değişken arasında 0 dan 1 e kadar 0,1 birimlik artış ile korelasyon bulunacak şekilde $n=20, 30, 50, 100$ ve 500 örneklem büyüklüğünde 3 farklı veri seti üretilerek 1000 simülasyon yapılmıştır. Uygulamada R paket programında açık kaynak kodlu pwr (Version 1.2-1; Champely, 2017) paketi kullanılmıştır.

Bulgular: 1000 simülasyon sonucunda farklı korelasyon değerlerine göre örneklem sayısı $n=20, 30, 50, 100$ ve 500 için Varyans analizi (ANOVA) ve Kovaryans analizi (ANCOVA) güç analizi sonuçları raporlanmıştır. Elde edilen simülasyon sonuçlarına göre sonuç değişkeni ile ortak değişken arasındaki korelasyon 0,2 ve üzeri olduğunda Varyans analizi yerine Kovaryans analizi kullanımının araştırma gücünü arttırdığı belirlenmiştir.

Sonuç: Gruplar arası kareler toplamı/Grup içi kareler toplamı olan F değeri kritik bir değerden büyükse, gruplar arasında anlamlı bir farklılığın olduğu sonucuna varılır. Açıklanamayan varyans, bireysel farklılıklar gibi hata varyansını ve diğer faktörlerin etkisini içerir. Bu nedenle, ortak değişkenlerin bağımlı değişken üzerindeki etkisi kontrol edildiğinde, grup içi kareler toplamından çıkan bu etki F değerini büyütecek ve varsa anlamlı bir farklılık bulmak için istatistiksel gücü artırmış olacaktır. Dolayısıyla deneysel hatayı ve bulgulardaki yanlılığı azaltmak için sonuç değişkeni ile ortak değişken arasında 0,2 üzerinde korelasyon bulunduğunda Varyans analizi yerine Kovaryans analizi kullanımı araştırmamızın gücünü artırarak yanlılığı azaltacaktır.

Anahtar Sözcükler: Kovaryans analizi, ortak değişken düzeltmesi, korelasyon, güç analizi, simülasyon

**Gözlemsel Çalışmalarda Ölçülemeyen Karıştırıcı Değişkenlerin Etkisi;
Propensity Skor ve Duyarlılık Analizleri**

Emre DEMİR¹, Serhat Hayme², Derya GÖKMEN², Atilla Halil Elhan², Serdal Kenan KÖSE²

¹Hitit Üniversitesi, Tıp Fakültesi, Biyoistatistik Anabilim Dalı, Çorum

²Ankara Üniversitesi, Tıp Fakültesi, Biyoistatistik Anabilim Dalı, Ankara

Özet

Amaç: Randomizasyonun yapılamadığı gözlemsel araştırmalarda, tedavi ve kontrol grupları arasında araştırma öncesinde birçok ortak değişken veya faktöre göre farklılıklar bulunabilir. Gruplar araştırılan etken dışında diğer özellikler açısından homojen olmadığında saf tedavi etkisi kestirilemeyecek ve yanlış tahminlere ulaşılabacaktır. Gözlemsel araştırmalarda tedavi etkisinin yansız tahminlerini elde etmek ve karıştırıcı (confounder) değişkenlerin etkisini gidermek için literatürde Propensity skor tahminine dayalı eşleştirme analizi kullanılmaktadır. Ancak eşleştirme analizi sonucu her zaman yansız tahminler elde edilemeyebilir. Bazı durumlarda eşleştirmeye alınacak ortak değişken veya faktörlerin dışında ölçülemeyen değişkenler bulunabilir ve tahminlere gizli yanlılık (hidden bias) karışabilir. Bu durumda Rosenbaum (2002, 2005) eşleştirme ile giderilemeyen veya araştırmacı tarafından belirlenemeyen değişkenlerden kaynaklı yanlılığı tanımlamak için eşleştirme sonrası duyarlılık analizlerinin yapılmasını önermiştir. Bu araştırma saf tedavi etkisinin kestirimi amacıyla yapılan propensity skora dayalı eşleştirme analizi sonucunda gözlenemeyen değişkenlerden kaynaklı gizli yanlılığı belirlemek için kullanılabilecek duyarlılık analizlerini bir uygulama üzerinde sunmayı amaçlamaktadır.

Yöntem: Sürekli bir sonuç değişkeni için Duyarlılık analizleri için Wilcoxon işaret testi ve Hodges-Lehmann nokta tahminine dayalı Rosenbaum sınırları kullanılmıştır. Farklı senaryolarda üretilen veri setleri üzerinden propensity skora dayalı eşleştirme analizi yapıldıktan sonra duyarlılık analizleri yapılmış ve senaryolara göre gizli yanlılığın değişimleri araştırılmıştır. Uygulamada R paket programında açık kaynak kodlu Matching (Sekhon, 2015) ve rbounds (Keele, 2015) paketleri kullanılmıştır.

Bulgular: Bir duyarlılık analizinde, yanlılık taşımayan bir çalışmadan ayrılma derecesinin ölçüsü olarak Gamma (Γ) kullanılır. Eşleştirme sonucunda yanlılığın giderildiği, diğer bir deyişle geriye dönük bir çalışma için randomizasyonun sağlandığı varsayıldığı için Gamma (Γ) 1 değerini aldığı Wilcoxon işaret test sonucundaki olasılık değeri ile eşleştirme analizi sonrasındaki olasılık değeri birbirine benzer olacaktır. Gözlenemeyen değişkenlerden dolayı tedaviye atanan gizli yanlılığın oddsu olan Gamma değerlerinin 1'in üzerindeki artışına göre hesaplanan alt ve üst sınır olasılık p değerlerinin değişimi istatistiki olarak anlamlılıktan anlamsızlığa dönüştüğünde gamma değerlerindeki farklılık gizli yanlılık duyarlılığını göstermektedir. Yani sonuç değişkenimiz üzerinde yaptığımız çıkarımı değiştirecek en küçük Gamma değerine göre ölçülemeyen ortak değişken varlığından kaynaklı gizli yanlılık olduğu belirlenmektedir.

Sonuç: Gözlemsel araştırmalarda propensity skora dayalı olarak yapılan eşleştirme analizi ile araştırılan tedavi etkisi dışındaki ortak değişken veya faktörlerin gruplarda benzer olması sağlanarak neden-sonuç ilişkisi doğru olarak saptanabilmektedir. Ancak araştırmalarda ölçülemeyen veya araştırmacı tarafından propensity skor hesabında kullanılmayan bazı ortak değişken veya faktörlerden kaynaklı gizli yanlılık oluşabilir. Bu nedenle eşleştirme sonrası gizli yanlılığın azaltılması ve tedavinin gerçek etkisinin belirlenmesi için duyarlılık analizlerinin araştırmalarda dikkate alınması önerilmektedir.

Anahtar Sözcükler: Gözlemsel araştırmalar, ölçülemeyen karıştırıcı etkisi, propensity skor, duyarlılık analizi

Ayırıcılığın Güç Olduğu Bir Klinik Durumda Farklı Karar Algoritmalarının Performansı

Ahmet KESKİNOĞLU¹, **Suriye ÖZGÜR²**, Burak ORDİN³, Önder YAVAŞCAN⁴, Caner ALPARSLAN⁴,
Timur KÖSE², Pembe KESKİNOĞLU⁵

¹Ege Üniversitesi Tıp Fakültesi Çocuk Sağlığı ve Hastalıkları AD

²Ege Üniversitesi Tıp Fakültesi Biyoistatistik ve Tıbbi Bilişim AD

³Ege Üniversitesi Fen Fakültesi Matematik Bölümü

⁴Tepecik Eğitim ve Araştırma Hastanesi, Çocuk Sağlığı ve Hastalıkları

⁵Dokuz Eylül Üniversitesi Tıp Fakültesi Biyoistatistik ve Tıbbi Bilişim AD

ÖZET

Giriş: Sağlık alanında edilen verilerle genellikle olasılık dahilinde bir karar verilir. Son dönemlerde sağlık alanındaki araştırmalarda tıbbi karar için veri madenciliğinden yararlanılarak veri madenciliğinin tüm teknikleri uygulanmaktadır. Sınıflandırma, kümeleme ve özellik seçimi sağlık alanında da uygulanan veri madenciliğinin temel alanlarıdır.

Amaç: Bu çalışmada klinik olarak birbirinden ayırıcı tanısı oldukça güç yapılan iki hastalık için (Vezikoureteral Reflü/ Tekrarlayan idrar yolu enfeksiyonu-VUR/İYE), veri madenciliği temel alanlarında yer alan analizler gerçekleştirilmiştir. Üç farklı karar algoritmasının performansını farklı veri yapısı ile birlikte değerlendirmek amaçlanmıştır.

Yöntem: Araştırmanın verileri hastane kayıtlarından retrospektif olarak elde edilmiştir. 425 VUR tanılı ve 186 İYE tanılı çocuğun kaydı incelenmiştir. Bu iki hastalık için dosyalardaki kayıtlarda aynı değerlendirme ve inceleme süreçleri mevcuttur. Klinik bulgu, laboratuvar sonuçlar ve radyolojik inceleme sonuçlarından 39 özellik ile sonuç (VUR/İYE) tahminlenmiştir. Kural temelli karar tablosu (decision table), fonksiyon temelli MLP (Multilayer perceptron-çok katmanlı algılayıcı) ve Naive-Bayes sınıflandırma algoritmalarının performansı veri özelliklerine göre değerlendirilmiştir. Analiz için WEKA 3.6.2 programından yararlanılmıştır.

Bulgular: Farklı üç yöntemin uygulanması ile farklı başarımlar elde edilmiştir.

Decision table: %97,38

MLP: %84,34

Naive-bayes: %48,77

Karar tablosu sınıflama performansı, analize alınan açıklayıcı değişken sayısı farklılaştırıldığında da en yüksek ve Naive-Bayes yöntemi en düşük performansta saptanmıştır. Açıklayıcı değişkenler ilk modelde klinik ve laboratuvar değişkenleri, ikinci modelde bu değişkenlere görüntüleme değişkenleri eklenerek incelenmiştir. Değişken sayı ve özellikleri değiştirildiğinde de yukarıdaki performanslara benzer sonuçlar elde edilmiştir.

Tartışma ve Sonuç: Naive-Bayes, istatistiksel yöntemler kullanarak sınıflandırma yapan ve sınıflardaki niteliklerin birbirinden bağımsız olduğunu varsayan basit bir tekniktir. Oysa bu veri özelliğinde bağımlı bir yapı mevcuttur. Bu nedenle bu algoritmanın performansının kötü olduğu düşünülmektedir.

Farklı özellikteki verileri bir arada analize olanak sağlayan veya doğrusal olmayan ilişkilerin analiz edilebildiği çok katmanlı algılayıcılar ve karar tablosu ile iyi başarımlar elde edilmiştir. Araştırmanın veri yapısında bu sayılan özellikler (nominal, ordinal ve sürekli veri varlığı) olduğu için başarımlar Naive-Bayes yönteminden daha iyi performans sağlamıştır.

Bu çalışma TÜBİTAK tarafından desteklenen 114S011 no'lu projenin verilerinden üretilmiştir.

Anahtar Sözcükler: Karar Tablosu, Naive-Bayes, Çok katmanlı algılayıcı, Ayırıcı klinik tanı

First Course in Biostatistics at Medical Faculties: Explaining the Importance of Biostatistics

Fatma Ezgi CAN^{1,2}, Ilker ERCAN³, Deniz SIGIRLI³, Guven OZKAYA³, Zuleyha ALPER^{4,5}

¹*Department of Biostatistics and Medical Informatics, Faculty of Medicine, Izmir Katip Celebi University, Izmir, Turkey*

²*Department of Biostatistics, Institute of Health Sciences, Uludag University, Bursa, Turkey*

³*Department of Biostatistics, Faculty of Medicine, Uludag University, Bursa, Turkey*

⁴*Department of Medical Education, Faculty of Medicine, Department of Medical Education, Uludag University, Bursa, Turkey*

⁵*Department of Family Medicine, Faculty of Medicine, Department of Family Medicine, Uludag University, Bursa, Turkey*

Abstract

Aims: Statistics is a mandatory course in health sciences and in the fields of science and social sciences in bachelor's degree curricula. The aim of our study is: (i) to evaluate the effect of explaining the importance of biostatistics to medical students with the support of visual materials, sample events and, detailed information from clinicians about the place they put biostatistics in their professional lives, to arouse the interest and curiosity of students about biostatistics, (ii) to evaluate students first predictions about Biostatistics, and (iii) to ensure to give the importance it deserves by the implementation of new compelling educational techniques.

Methods: Three methods were performed to explain the importance of Biostatistics to three different student groups in the first class in medicine faculty, to see the impact of visual material, detailed information and sample events. One week later, three questions with a five-point likert scale were administered to the students of each group.

Results: The reliability level of the applied scale may be expressed as intermediate level.

Conclusion : Educators must put 'experiences and practising' to the centre of learning in lessons and should improve their techniques to show students the importance of the subject.

Keywords: Biostatistics education; Medical students; Medical education; Medical curriculum.

Multiple Correspondence Analysis (MCA): Eating habits among university students

Gökay EÖ.¹, **Bekiroğlu N.**²

¹ *Ankara University, Medical School, Biostatistics Department, Ankara, Turkey*

² *Marmara University, Medical School, Biostatistics Department, İstanbul, Turkey*

Background: Categorical variables are frequently used in studies related to health sciences. Methods that used in the analysis of studies having such type of data, in some cases, may be insufficient and it is also difficult to interpret the relationship between such variables. One of the methods used to eliminate this problem, is the **multiple correspondence analysis(MCA)**.

Aim: In this study, we aimed to determine Marmara University students' eating habits by means of the model "four-leaf-clover", and to investigate the relationship between the factors affecting students' eating habits.

Materials and Methods: Questionnaire consisted of 17 questions concerning to their eating habits among university students. "Four-leaf-clover" model is divided into four groups; such as milk and milk (dairy) products, meat-egg- and-legume group, bread-cereal group, and fruit-vegetable group. This questionnaire was distributed to Marmara University students at Marmara Universty. Data of the survey have been obtained by face-to-face interviewing with students and the results of MCA are discussed.

Results: 204 students were accepted to participate in survey. 54.4% (n=111) of these students were female and 45.6% (n=93) were male. Age of the students was between 19-25. Among the students, 29.1% of the students (n=59) were living with their family, 29.1% (n=59) were living in their own student house, 22.6% (n=46) were living in a public dormitory, 19.2% (n=39) were living in a private dormitory.

47.0% (n=95) of these students preferred to eat 2 meal courses a day while 44,6% (n=90) were eating 3 meal courses a day. More than half of the students 52,9% (n=108) declared that their knowledge level about nutrition as "I know a lot".

When it comes to the main factors such as gender, the level of nutrition knowledge and the place where students live during their university education according to the MCA coordinate plot; females who live with their family have high level of knowledge about nutrition. The percentage of inertia (variance in the cross-table) explained by the first dimension is %21,29 and by the second dimension is %16,25. The percentage of inertia explained by the sum of these two dimensions is 37.54%.

Conclusion: According to the model of "four-leaf clover", female students rather than male students, and students who live with their parents have more regular and healthy nutrition habits and they consume more fruits and vegetables.

Keywords: multiple correspondence analysis, students' eating habits, categorical data analysis

Bland Altman İstatistiksel Yöntemi ve Prostat Kanserinde Bir Uygulama
Seyma YAŞAR¹, Bulut DURAL², Leyla KARACA³, Zeynep TUNÇ¹, Harika G. G. BAĞ¹, Saim YOLOĞLU¹

¹İnönü Üniversitesi Tıp Fakültesi Biyoistatistik ve Tıp Bilişimi Anabilim Dalı, Malatya

²İnönü Üniversitesi Turgut Özal Tıp Merkezi Üroloji Anabilim Dalı, Malatya

³İnönü Üniversitesi Turgut Özal Tıp Merkezi Radyoloji Anabilim Dalı, Malatya

Amaç: Yöntem karşılaştırma çalışmalarında aynı denek üzerinden farklı yöntemlerle elde edilen ölçümler arasındaki uyumun düzeyi merak edilir. Bunun için kullanılan yöntemlerden birisi de Bland Altman istatistiksel yöntemidir (1). Prostat kanseri erkeklerde cilt kanserinden sonra en sık tanı konulan ve akciğer kanserinden sonra ikinci sıklıkta ölüme neden olan kanserdir. Prostat kanserini yakalamak için dijital rektal muayene (DRM) ve prostat spesifik antijen (PSA) kullanılmakta ve tanı ultrason eşliğinde transrektal biyopsi ile konulmaktadır. Klinik olarak önemli prostat kanserinin tespitini sağlayacak, gereksiz biyopsilerin önüne geçecek, hastaya uygulanacak en uygun tedavinin seçilmesini ve gerektiğinde aktif takibin yapılmasını sağlayacak görüntüleme yöntemlerine olan ihtiyaç belirgin hale gelmiştir (2). Prostat kanserinin belirlenmesinde kullanılan görüntüleme yöntemlerinden ikisi Bilgisayarlı Tomografi (BT) ve Manyetik Rezonans (MR) dır. Bu çalışmada bu iki görüntüleme yöntemi arasındaki uyum Bland Altman uyum analizi ile incelenmiştir.

Gereç ve Yöntemler: Bu çalışmada İnönü Üniversitesi Tıp Fakültesi Turgut Özal Tıp Merkezi Üroloji Anabilim Dalına başvuruda bulunan, prostat kanseri teşhisi konulan 43 hastanın Bilgisayarlı Tomografi(BT) ve Manyetik Rezonans (MR) sonucu Bland-Altman uyum analizi ile analiz edilmiştir. Çalışmada yapılan analizler için MedCalc version 10.1.6.0 ve IBM SPSS Statistics 24.0 programları kullanılmıştır.

Bulgular: Bu çalışmada kullanılan 43 hastadan alınan BT ve MR sonuçları için ortalama ve standart sapma sırasıyla $22,78 \pm 9,3$ ve $19,53 \pm 9,25$ dür. Yapılan istatistiksel analizler sonucunda BT ve MR ölçümleri arasındaki farklar normal dağılım göstermektedir ($p=0,34$). Bu değerlere ait fark ve ortalamalar arasında ise istatistiksel olarak anlamlı bir ilişki bulunmamıştır ($p=0,953$). Bland Altman grafiğindeki BT ve MR ölçüm değerlerinin farkları ve ortalamalarına ait noktaların önemli bir miktarı uyum sınırları içinde olduğundan, bu değerlere ait farklar ve ortalamalar arasında ilişki olmadığı bulunmuştur. 43 hasta üzerinden iki metot ile elde edilen ölçüm değerlerinin farklarına ait ortalama 3,24, standart sapma ise 4,72 dir. Farkların ortalamasına ait %95 güven aralığı ise 1,79 ile 4,7 dir. Farkların ortalaması ve standart sapması kullanılarak elde edilen uyum sınırları %95 güven düzeyinde -6,01 ve 12,51 olarak bulunmuştur. Uyum sınırlarının alt limiti olan -6,01 değeri için %95 güven aralığı -8,51 ile -3,5 bulunurken, üst limit değeri olan 12,51 için %95 güven aralığı 10 ile 15,01 bulunmuştur.

Sonuçlar: Yapılan analizler sonucunda MR görüntüleme tekniğinden elde edilen sonuçlar, BT görüntüleme tekniğinden elde edilen tümör çapları $6,01 \text{ cm}^2$ büyük, $12,51 \text{ cm}^2$ küçük bulunabilir. Dolayısıyla bu sonuçlar gözönünde bulundurularak hastaneye yeni başvuruda bulunan ve kanser şüphesi olan yada prostat kanseri hastası olup takip gereken hastalar için sadece bir görüntüleme tekniğinin kullanılması yeterli olacaktır.

Anahtar Sözcükler: Bland Altman, Metot Karşılaştırma, Prostat Kanseri

Kaynaklar

1. Yasemin G, Sertkaya D, Demirtaş S. Klinik araştırmalarda iki ölçüm tekniğinin uyumunu incelemede kullanılan istatistiksel yöntemler. Ankara Üniversitesi Tıp Fakültesi Mecmuası. 2003;56(01).
2. EROL C. Prostat Kanserinde Görüntüleme Teknikleri. Türkiye Klinikleri Journal of Urology Special Topics. 2017;10(2):121-32.

Tıp Eğitiminde Sınavların Değerlendirilmesinde İç Tutarlılık Katsayıları Kullanılabilir mi?

Tanyeli KAZAZ, **Seval KUL¹**

¹Gaziantep Üniversitesi, Tıp Fakültesi, Biyoistatistik AD, Gaziantep

Amaç: Tıpta ölçme ve değerlendirme eğitimin en önemli bileşenlerinden birleşenlerinden biridir. Çoğunlukla sınavlar çoktan seçmeli şekilde yapılır ve toplam puan hesaplanır. Madde analizi yapılarak soruların ayırıcılık ve zorluk düzeyleri değerlendirilmektedir. Ayrıca basıklık ve çarpıklık katsayıları yardımıyla sınavın genelinin zorluk düzeyinin de belirlenmesi mümkündür. Buna karşın soruların bir bütün ya da alt bölümlerinin puanları arasındaki tutarlılığı ye yönelik olup olmadığı değerlendirilmemektedir. Bu çalışmada Cronbach alfa, KR20, Spearman Brown katsayılarını kullanarak sınav geçerliğinin test edilip edilemeyeceğinin ortaya konulması amaçlanmıştır.

Yöntem: Binomiyal dağılım yardımıyla doğru ve yanlış cevapları ifade eden binary yapıda 100 soru içeren 20,30,40,50 ve 60 kişinin sınava katıldığı varsayılan veri setleri üretilmiş, bu veriler için Cronbach alfa, KR20, Spearman Brown değerleri hesaplanmıştır. Öğrencilerin toplam puanları ile geçerlik istatistikleri arasındaki korelasyonlar incelenmiştir. Simulasyonlar R programında yapılmıştır.

Bulgular: Yöntemler arasındaki en yüksek ICC değerleri KR20 ile Cronbach alfa arasında ($ICC=0,996$ (%96 $GA=0,994-0,997$) olarak bulunmuştur ($P=0,001$). Bu değerlerin sınav ortalaması ile zayıf bir negatif anlamlı korelasyonu vardır (sırasıyla $r=-0,287$, $p=0,004$; $r=-0,295$, $p=0,003$). Spearman Brown katsayısı ile sınav puanı arasında anlamlı korelasyon saptanmamıştır ($r=-0,154$, $P=0,127$).

Sonuç: Sınav puanları ile yöntemlerden elde edilen değerler arasında sınav sorularının geçerliği test edebilecek düzeyde anlamlı bir ilişki olmadığına dolayısı ile yöntemlerin uygulamada kullanılamayacağına karar verilmiştir.

Toplam Prostat Spesifik Antijen Değerlerinin Yapay Zekâ Modellemesi ile Tahmini

Seyma YAŞAR¹, Ali BEYTUR², Cemil ÇOLAK¹

¹ İnönü Üniversitesi Tıp Fakültesi Biyoistatistik ve Tıp Bilişimi AD, Malatya, Türkiye.

² İnönü Üniversitesi Tıp Fakültesi Üroloji AD, Malatya, Türkiye.

Özet

Giriş: Prostat kanserinin tanısı için kullanılan serum belirteçlerinden birisi olan toplam prostat spesifik antijen (PSA) taramasının klinik olarak yararlı olduğu gösterilmiştir. Bu çalışmada, toplam PSA değerinin çok katmanlı algılayıcı (Multilayer Perceptron, MLP) yapay sinir ağı (YSA) modeli yardımıyla tahmin edilmesi amaçlanmıştır.

Gereç ve Yöntem: Bu çalışmadaki toplam PSA değerlerine yönelik veriler (n=1422), İnönü Üniversitesi Tıp Fakültesi Üroloji Anabilim Dalına ait hasta kayıtlarının saklandığı veritabanından rastgele olarak yapısal sorgulama dili (SQL) kullanılarak seçilmiştir. Toplam PSA değerleri hedef/bağımlı değişken, yaş (yıl), kan grubu (A/B/0/AB), ex durumu (sağ/ölü), lenfosit (LY) (%), hemoglobin (HGB) (g/dL), nötrofil (NE) (%), albumin (g/dL), kalsiyum (mg/dL), ortalama eritrosit hemoglobini (MCH) (pg), lökosit sayısı (WBC) ($10^3/\text{ml}$), trombosit (PLT) ($10^3/\text{ml}$) açıklayıcı/tahminleyici değişkenler olarak analizlerde değerlendirilmiştir. Aykırı/aşırı gözlemler analiz edilmiş, nicel değişkenler Z-skor ya da Box-Cox dönüşümü yeniden ölçeklendirilmiş ve değişken seçim yöntemi kullanıldıktan sonra, MLP YSA modeli toplam PSA değerini tahmin etmek için oluşturulmuştur. Modelin tahmin başarımı, ortalama mutlak hata, standart sapma ve korelasyon katsayısı değerleri ile incelenmiştir.

Bulgular: MLP YSA modeli, 993'ü eğitimde ve 429'u ise teste olmak üzere toplam 1422 veri seti kullanılarak oluşturulmuştur. Ortalama mutlak hata, standart sapma ve korelasyon katsayısı değerleri eğitim veri seti için; 0.744, 0.895 ve 0.452; test veri seti için ise; 0.773, 0.935 ve 0.355 olarak hesaplanmıştır. Oluşturulan modelin tahmini doğruluk oranı; %20.3 olarak kestirilmiştir. MLP YSA modelinde değişkenlerin önemlilik değerleri ise; HGB için 0.33, NE için 0.22, Kalsiyum için 0.16, PLT için 0.13, yaş için 0.10 ve ex durumu için 0.06 olarak elde edilmiştir.

Sonuçlar: MLP YSA modeli, seçilen değişkenlere dayalı olarak toplam PSA değerinin tahmininde oluşturulmuş ve ilişkili faktörlerin önemlilik düzeylerini hesaplamıştır. Toplam PSA değerlerinin tahmininde daha iyi tahmin sonuçları, farklı ilave değişkenlerin, alternatif yeniden örnekleme yöntemlerinin ve değişik modellerin kullanımı ile elde edilebilir.

Anahtar Sözcükler: Risk faktörleri, toplam prostat spesifik antijen (PSA), yapay zekâ modellemesi.

Sürücü Davranışları Ölçeğinin Geçerlik Güvenirlik Analizi ve Faktör Yapısı

Erkut BOLAT¹

¹İstanbul Üniversitesi Cerrahpaşa Tıp Fakültesi Biyoistatistik Anabilim dalı

Amaç: Sürücü davranışları ölçeği (SDÖ), ihlaller (10 madde), ihmaller (8 madde) ve hatalar (8 madde) olmak üzere toplam 26 maddeden oluşmaktadır. Ölçek, sürücülerin araç kullanırken nasıl tavrı sergilediklerini ve trafikte diğer sürücülere karşı tutumlarının nasıl olduğunu öğrenmek için kullanılmaktadır. Bu çalışmadaki amacımız ülkemizde bu ölçeğin geçerlilik ve güvenilirlik değerlerini öğrenmek ve faktör bileşenlerinin birbiriyle karşılaştırıp bir değerlendirme yapmak olacaktır.

Yöntem: Bu çalışmada, 2015 yılında 891 kişilik örneklem grubuna uygulanan sürücü davranışları ölçeği kullanılmıştır. Çalışma İstanbul'da Haziran 2015'ten Aralık 2015'e kadar olan süreçte kesitsel bir araştırma olarak yürütülmüştür. 236'sı kadın 655'i erkek olmak üzere 891 sürücü çalışmada yer almıştır. Katılımcıların ne sıklıkta 26 davranışı işlediğini öğrenmek için altı noktalı ölçek (0=Hiç bir za man, 1=nadiren, 2= Bazen, 3= Oldukça sık , 4= Sık sık, 5= Her zaman) kullanılmıştır. Faktör bileşenlerini öğrenmek için Principal component analysis; Varimax rotation with Kaiser Normalization ve Eigen değer >1 olarak yapılmıştır. Geçerlilik ve güvenilirlik katsayılarını öğrenmek için Cronbach Alfa değeri incelenmiştir.

Bulgular: Faktör analizi, Principal component analysis; Varimax rotation with Kaiser Normalization ve Eigen değer >1 olarak yapıldığında 6- faktörlü çözüm önermesine karşın 3-faktörlü çözüm tercih edilmiştir. Birinci bileşen 10 madde içerir. Birinci bileşenin maddeleri trafikteki ihlalleri kapsadığından birinci bileşen "ihlaller"den oluşur. İkinci bileşen, 8 maddeden oluşan yoldaki hatalardan bahseder. Üçüncü bileşen de 8 madde olan ihmallerden oluşur. Faktör yükleri ihlallerde en yüksek "Otobanda hız limitlerini dikkate almamak" (0.71), hatalarda "Kaygan bir yolda ani fren veya patinaj yapmak" (0.65) ve ihmallerde "Aracınızı park alanında nereye bıraktığınızı unutmak" (0.75) maddeleri bulunmuştur.. Cronbach Alfa değeri ise ihlaller için 0.76, hatalar için 0.80, ihmaller 0.70 için ve tüm maddeler için ise 0.87 olup yeterli derecede yüksek bulunmuştur.

Sonuç: Sonuç olarak, sürücü davranışları ölçeği faktör yapısı ve geçerlik güvenirlik bakımından elde edilen değerler yüksek olup ölçeğin sürücü davranışlarını ölçmede geçerli ve güvenilir bir ölçme aracı olduğu söylenebilir.

Anahtar Sözcükler: Sürücü davranışları ölçeği, Faktör Analizi, Geçerlik, Güvenirlik.

Design Effect in Multicenter Surveys

İsmet DOĞAN, Nurhan DOĞAN

*Afyon Kocatepe Üniversitesi Tıp Fakültesi Biyoistatistik ve Tıbbi Bilişim Anabilim Dalı,
AFYONKARAHİSAR*

Multicenter (= cluster based) studies involve correlation in data because subjects from the same center are more similar than are those from different centers. Such a correlation potentially affects the power of standard statistical tests, and conclusions made under the assumption that data are independent can be invalidated. In cluster based studies, the between clusters variation causes inflation of error variance. This needs to be accounted for at the time of designing a study, or it will reduce the power to give desired results. If the clustering is ignored, the standard error will (generally) be overoptimistic and the *P*-values will be artificially small. Such an analysis will not give a biased estimate of the effect, but will overestimate the precision of that estimate. To have adequate power, the sample size has to be increased by a design effect (*Deff*) or variance inflation factor. A usual measure of the clustering effect on an estimator is the *Deff*. *Deff* is defined as the ratio of two variances: the variance of the estimator when the center effect is taken into account over the variance of the estimator under the hypothesis of a simple random sample.

$$Deff = \frac{\text{Complex Survey Design Variance}}{\text{Simple Random Sample Variance}}$$

The *Deff* represents the amount by which the sample size needs to be multiplied to account for the design of the study. Ignoring clustering can lead to over or underpowered studies. A *Deff* greater than 1.0 connotes that the sampling design decreases precision of estimate compared to simple random sample, and a *Deff* less than 1.0 confirms that the sampling design increases precision of estimate compared to simple random sample. The standard measure of the extent to which observations are correlated by cluster is the intraclass correlation coefficient (ICC). A better known measure related to the ICC is the design effect due to clustering, defined as 'the loss of effectiveness (resulting from) use of cluster sampling, instead of simple random sampling'. Although most of the ICCs are small, for a large cluster size they can increase the design effect substantially, which in turn affects sample size. Hence even the small ICC should not be ignored while calculating the sample size. The *Deff* is a function of average cluster size ($\bar{n}$) and intraclass correlation coefficient (*ICC*, ρ):

$$Deff = 1 + (\bar{n} - 1)\rho \quad (1)$$

The ICC quantifies the variation between the clusters and can be defined as proportion of total variation that is attributed to differences between the clusters:

$$ICC = \rho = \frac{\sigma_b^2}{\sigma_b^2 + \sigma_w^2} \quad (2)$$

σ_b^2 : between cluster component of variance,

σ_w^2 : within cluster component of variance,

Deff indicates how sampling design influences the computation of the statistics under study and accommodates for the miscalculation of sampling error. *Deff* is used to adjust for inaccurate variances. Also, *Deff* are used to express the effects of sample design beyond the elemental variability, removing both the units of measurement and sample size as 'nuisance parameters'.

The ICC does not depend on the cluster size or on the numbers of clusters. The design effect, however, is affected by the sample design and is strongly dependent on cluster size.

Keywords: Cluster based survey, error variance, experimental design

References

Kish L. Methods for design effects. Journal of Official Statistics 1995; 11(1): 55-77.

Masood M, Reidpath DD. Intraclass correlation and design effect in BMI, physical activity and diet: a cross-sectional study of 56 countries. BMJ Open 2016; 6: e008173. doi:10.1136/bmjopen-2015-008173

Vierron E, Giraudeau B. Design effect in multicenter studies: gain or loss of power? BMC Medical Research Methodology 2009; 9(39).

Afyon Kocatepe Üniversitesi Tıp Fakültesi Öğrencilerinin Toplumsal Cinsiyet Rollerini

Nurhan DOĞAN¹, Bilal YILMAZ², Abdullah Oğuz TAŞBAŞ², Berat Niyazi ÖLMEZ², Rümeyza KILINÇ², Saliha Nur SEZGİN², Tarık Alp ŞAHİNOĞLU², Ömer Faruk BALLI²

¹ Afyon Kocatepe Üniversitesi, Biyoistatistik ve Tıbbi Bilişim Anabilim Dalı, Afyonkarahisar

² Afyon Kocatepe Üniversitesi 3.Sınıf Tıp Fakültesi Öğrencisi

Özet

Amaç: AKÜ Tıp Fakültesi öğrencilerinin toplumsal cinsiyet rollerini belirlemek ve toplumsal cinsiyet rollerinin çeşitli sosyo-demografik özelliklere göre farklılıklarını araştırmaktır.

Yöntem: Araştırmaya, Afyon Kocatepe Üniversitesi Tıp Fakülte'sinde eğitim görmekte olan toplam 431 öğrenci katıldı. Öğrencilerin toplumsal cinsiyet rollerini belirlemek amacıyla Cinsiyet Rolü Değerlendirme Envanteri (Bem, 1974) kullanıldı. Envanter toplam 60 maddeden oluşmaktadır. Bu 60 maddenin 20 tanesi "kadınsılık" (Feminine), 20 tanesi "erkeksilik" (Masculine) ve 20 tanesi ise "toplumsal isteklilik" (Social Desirability) olmak üzere üç alt ölçek içermektedir. Söz konusu ifadeler 7'li likert ölçeği ile ölçeklendirilmiştir. Veriler yüz yüze görüşme yöntemiyle elde edildi. Ölçeğin güvenilirliğinin hesaplanmasında Cronbach alfa katsayısı kullanıldı. Verilerin değerlendirilmesinde ki-kare testi ve anlamlı olan ki-kare testleri için düzeltilmiş artıklar (Adjusted Residual) kullanılarak hücre bazında değerlendirmeler yapıldı. İstatistiksel anlamlılık için $p < 0.05$ değeri kabul edildi. Verilerin değerlendirilmesinde SPSS 20.0 paket programından yararlanıldı.

Bulgular: Ölçeğin güvenilirlik katsayısı toplam da %87 (alt ölçeklerde kadınsılık (%84) ve erkeksilik (%85))'dir. Kadınsılık ortanca değeri 5,15 (2,5;6,90) erkeksilik ortanca değeri ise 4,90 (2,05;6,85) olarak elde edildi. Erkeklerin %34'ü androjen, %31'i erkeksi, %29'u farklılaşmamış (undifferentiated) ve %6'sı ise kadınsı cinsiyet rolüne sahipken, kadınların %36'sı kadınsı, %28'i androjen, %24'ü farklılaşmamış ve %12'si ise erkeksi cinsiyet rolüne sahiptir. Otoriter aile ortamında yetişen öğrenciler erkeksi, demokrat aile ortamında yetişenler öğrenciler androjen ve koruyucu aile ortamında yetişen öğrenciler ise daha kadınsı cinsiyet rolüne sahiptirler ($p=0.01$). Cinsiyet rolü olarak kadınsı cinsiyet profiline sahip olan öğrenciler mezun olduklarında Dahili Tıp Bilimleri bölümünü seçmeyi düşünürken androjen profiline sahip olan öğrenciler ise Cerrahi Tıp Bilimleri bölümünü seçmeyi düşünmektedirler ($p < 0.001$).

Sonuç: Katılımcılar, yüksek düzeyde kadınsı ve erkeksi özellikler geliştirerek, bulunduğu ortamın gerektirdiği kadınsı veya erkeksi özellikleri gösterebilecek esnek bir kişiliğe sahip oldukları görülmüştür. Yani hem erkeksi hem de kadınsı özellikleri desteklediği görülmektedir. Farklılaşmamış cinsiyet rolüne sahip olan katılımcıların bir toplumsal cinsiyet rolü belirleme kapsamında karmaşa yaşadıkları ifade edilebilir. Elde edilen sonuçlar toplumdaki cinsiyet rolünün değiştiğini düşündürmektedir. Bu yönde daha kapsamlı araştırmalara ihtiyaç duyulmaktadır.

Anahtar Sözcükler: Bem Cinsiyet Rolü Envanteri, Üniversite Öğrencisi, Toplumsal Cinsiyet Rollerini

Global Patterns of Incidence and Mortality in Lung Cancer

Nurhan DOĞAN¹, İsmet DOĞAN¹

¹*Afyon Kocatepe University, Faculty of Medicine Department of Biostatistics and Medical Informatics, Afyonkarahisar.*

Abstract

Aim: This study aims to compare the distribution of the six WHO regions and the more developed and less developed regions of all cases incidence and mortality in lung cancer (for males and females, respectively) with the distribution of the all age groups by Configural Frequency Analysis (CFA).

Methods: National data (Globocan-2012) were included in the analysis. These include six World regions and the more developed and less developed regions. Data were analysed with CFA which is used for the analysis of multiway contingency tables.

Results: In 2012 over 1.6 million people died due to lung cancer (69.1% in male and 30.9% in female). In Africa, East Mediterranean, Europe and South-East Asia, lung cancer incident cases were observed to be higher than expected. Under the years <70, in the Africa, East Mediterranean, Europe and South-East Asia lung cancer mortality are significantly underrepresented and therefore antitypes. Here, lung cancer mortality was observed lower than expected. In the more developed regions, lung cancer incident cases and mortality were significantly overrepresented in almost all age groups both male (years <55) and female, but in the less developed regions it was significantly underrepresented in almost all age groups.

Conclusion: The majority of the global lung cancer burden occurred in the more developed countries. If the current situation remains unchanged, incidence and mortality may continue to increase.

Keywords: Configural frequency analysis, Lung cancer, Mortality, Incidence, WHO regions

Panel Veri Analizi ve Beyin Göçü Verilerinde Uygulama

Asena Ayça ÖZDEMİR, Gülhan OREKİCİ TEMEL

Mersin Üniversitesi Tıp Fakültesi, Biyoistatistik Anabilim Dalı, Mersin

Özet

Amaç: Bu çalışmanın amacı panel veri analizini kullanmak ve 192 ülkenin beyin göç verisi üzerinde uygulamaktır.

Yöntem: Sağlık alanına uygun verilerde iki farklı veri türü karşımıza çıkmaktadır. Bu veriler ya zaman serisi verileri ya da kesitsel veri setleridir. Ve her veri türü kendine özgü analizlerle değerlendirilmektedir. Günümüzde birçok veri tabanlarında zaman serileri ve yatay kesit verileri oluşturulmaya başlanmış bir vakaya ya da vaka grubuna ait gözlemler her biri ayrı zamanlarda gözlenerek kaydedilmiştir. Bu tipteki zaman ve kesit verilerinin birleştirilmesi ile oluşan verilere panel veri denilmektedir. Bu tipteki verilerin kullanılması ile oluşan ilişkilerin tahmin edilmesi yöntemine de Panel Veri Analizi denilmektedir.

Panel veri analizinin en önemli avantajı gözlem sayısının zaman ve kesit verilerine göre daha fazla olmasıdır. Ayrıca zaman serisi verilerinde çoklu bağlantı problemi olmasına rağmen panel veri analizinde çoklu bağlantı sorunu yoktur. Örneğin kesitsel yatay verilerde sadece vakalar arasındaki farklılıklar incelenebilirken, panel veri analizi kullanılarak hem vakalar arasındaki farklılıklar hem de bu vakanın zaman içerisindeki farklılıkları birlikte incelenmektedir. Panel veri analizinin dezavantajları ise, tekrarlı verilerin elde edilmesi, verinin popülasyonu temsili için henüz güç analizinin olmaması, panelin sonlarına doğru kayıp veri sorunu ve zamana bağlı yanlılıkların olmasıdır.

Panel veri analizi ile oluşturulan modeller katsayıların sabit ya da tesadüfi varsayıldığına bağlı olarak iki grupta incelenmektedir. Bütün katsayıların vakalara ve/veya zamana göre değişmediği ve vakaların aynı popülasyondan geldiği vakaların homojen olduğu varsayımın olduğu modeller katsayıları sabit modellerdir. Vakalara veya zamana göre meydana gelen farklılığın modele dahil edildiği sabit etkili modeller ve tesadüfi etkili modeller olmak üzere modellerde mevcuttur. Birim ya da birim ve zaman farklılıklarını temsil eden katsayılara ait modelin sabit ve tesadüfi etki modeli olup olmadığını ki kare dağılımı gösteren Hausman testi ile test etmek mümkündür.

Bulgular: Bu çalışmada ülkelerin beyin göçlerine ait veriler panel verisi olarak kullanılmıştır. Veri seti 192 ülkenin, 1991-2004 yılları arasındaki beyin göçü, ülke popülasyonu verilerini içermektedir. Verilerin analizi E-Views 10 paket programında yapılmıştır.

Sonuç: Hausman testine göre model random etkili modeldir. Kurulan model istatistik açıdan anlamlı bulunmuştur ($p < 0.001$). Hem zamanın hem de birimlerin etkisini ortaya koymak açısından panel veri analizi oldukça önemli bir yöntemdir.

Anahtar Kelime: Panel Veri, Boylamsal Veri, Kesitsel Veri

Biyoeşdeğerlik Çalışmalarında Birey İçi Değişkenlik Katsayısının Büyük Olduğu veya Bilinmediği Durumlarda Örneklem Genişlikleri Hesaplama Adımları

Asena Ayça ÖZDEMİR, Elif ERTAŞ, Emine Arzu KANIK

Mersin Üniversitesi Tıp Fakültesi, Biyoistatistik Anabilim Dalı, Mersin

Özet

Amaç: Bu çalışmanın amacı biyoeşdeğerlik çalışmalarında birey-içi değişkenlik katsayısının büyük olduğu veya hakkında yeterli bilgiye sahip olunmadığı durumlarda örneklem genişliği hesaplama adımlarının belirlenmesidir. **Gerekçe:** Biyoeşdeğerlik çalışmaları ilaç geliştirme faz çalışmalarının 1. aşamasına denk gelen genellikle sağlıklı yetişkin gönüllerde ve özel durumlarda ise hastalarda, çocuklarda ve kadınlarda da yürütülmesine izin verilen önemli araştırmalardır. Bu araştırmalarda post hoc güç hesaplanması klinik anlamlı olmayıp, araştırma başında en az %80 güç koşullarını sağlayan örneklem hacmi ile başlamak hem yanlışlıkla reddedilecek hipotezlerde toplumun riskini kontrol etmek hem de yanlışlıkla kabul edilecek hipotezlerde üretici firmanın riskini kontrol altına almak açısından hem toplum sağlığı hem de ekonomik dengeler bakımından önemlidir. Birey-içi değişkenlik (CV) çapraz geçişli denemelerde ANOVA hata kareler ortalamasından yararlanılarak hesaplanan bir katsayı olup 30 ve daha fazla olması aynı etken maddeye maruz kalan kişilerde farmakokinetik parametrelerin yüksek değişkenlik gösterdiğini gösterir. Eğer yeterli örneklem ile çalışılır ve tekrarlı veya sıralı gibi uygun bir deney tasarımı kullanılırsa ilacın klinik denemesi sorunda %80-%125 olan güven sınırlarında firmalar tarafından esnetme talep edebilmektedir. Bu nedenle birey-içi CV bu alanda önemli bir konudur.

Yöntem: Bu çalışmada literatürler incelenmiş ve biyoeşdeğerlik çalışmalarında yüksek veya bilinmeyen birey içi CV durumunda tasarım ve örneklem genişliği konusunda öneriler içeren makalelerden yararlanarak bu durumlarda kullanılacak adımlar bu çalışma sonunda elde edilen bir akış şeması ile özetlenmiştir. Paket programlarda CV ile örneklem hesaplama durumları karşılaştırmalı olarak incelenmiştir.

Tartışma : Biyoeşdeğerlik çalışmalarında kullanılan tasarımlar; maddelerin uzun yarılanma ömürleri biliniyorsa paralel tasarım, bilinmiyorsa çapraz geçişli tasarım yöntemleri kullanılır. Bu tasarımlarda ikiden fazla araştırma kol yer alıyorsa çok kollu paralel tasarım ya da yüksek dereceli çapraz geçişli tasarım kullanırken, iki kol yer alması durumunda birey-içi değişkenlik güvenilirliği sorgulanır. Eğer birey-içi değişkenlik hakkında bir bilgiye ulaşılabiliyorsa sabit örneklem tasarımı kullanılması gerekmektedir. Rehberlerde birey-içi değişkenliğin 20'den küçük olduğu durumlar için gruplarda 12 bireyin yeterli olduğu söylenmektedir. Bir parametre için birey-içi değişkenliği %30'dan yüksek olması ise yüksek düzeyde değişkenliği göstermektedir. Sabit örneklem tasarımında birey-içi değişkenlik<30 ise tekrarlı tasarımlar kullanılırken, birey-içi değişkenlik>30 olması durumunda 2X2 çapraz geçişli tekrarlı tasarım yöntemi uygulanmaktadır. Birey-içi değişkenlik hakkında kesin bir bilgiye sahip olunmadığı durumlarda pilot çalışmaların sonuçlarını da tartışmalıdır. Bu durumda iki aşamalı sıralı tasarım (two-stage sequential design) yöntemi kullanılması önerilmektedir.

Sonuç: Bu çalışmanın sonucunda birey-içi değişkenliğin yüksek olduğu ve bilinmediği durumlar için bir akış şeması çıkarılmıştır. Farklı değişkenlik katsayıları için de bir simülasyon çalışması ile grup sıralı tasarımların grup tekrarlı tasarımlarla karşılaştırılmasının yapılmasının literatüre katkı sağlayacağı düşünülmektedir.

Anahtar Kelime: Biyoeşdeğerlik, örnek genişliği, birey-içi değişkenlik katsayısı, sıralı denemeler

Farklı Sayıda Gözlem İçeren Veri Setleri İçin En Küçük Kareler Ve Temel Bileşenler Regresyonu Sonuçlarının Karşılaştırılması

Zeynep TUNÇ¹, Şeyma YAŞAR¹, Harika Gözde GÖZÜKARA BAĞ¹, Saim YOĞLU¹

¹İnönü Üniversitesi Tıp Fakültesi Biyoistatistik ve Tıp Bilişimi Anabilim Dalı, Malatya

Amaç: Regresyon analizinde en küçük kareler yöntemi, gözlem değerleri, değişkenler ve hataların dağılımı hakkında birtakım varsayımların sağlandığı durumlarda geçerlilik kazanır. Bu varsayımlar geçerli olmadıkça elde edilen sonuçlar güvenilir olmaz. Buna bağlı olarak elde edilen regresyon denkleminde yapılacak kestirimlerin hatalı olma olasılığı yüksek olur (1). Çoklu regresyon çözümlemesinin varsayımlarından biri bağımsız değişkenler arasında kuvvetli bir ilişki olmamasıdır. Bu varsayımın bozulması, yani bağımsız değişkenler arasında bir ya da daha fazla doğrusal bağıntının olması çoklubağlantı sorununu gündeme getirir. Uygulamaların çoğunda bağımsız değişkenler arasında ilişki söz konusudur. Hatta bazı durumlarda bağımsız değişkenler arasında çok kuvvetli doğrusal ilişki vardır ve böyle durumlarda regresyon modeli yardımıyla yapılacak çıkarsamalar yanlış yönlendirmelere ve hatalara neden olur. Bağımsız değişkenler arasında çoklubağlantı olduğunda uygulanabilecek regresyon yöntemlerinden biri olan temel bileşenler regresyonu doğrusal regresyon modelinde bağımsız değişkenler arasında güçlü ilişkiler olduğunda en küçük kareler tekniğine alternatif olarak geliştirilmiştir (2). Bu çalışmada farklı gözlem sayısı içeren çoklubağlantı koşulunu sağlayan veri setlerine en küçük kareler ve temel bileşenler regresyonları uygulanmıştır. Çoklubağlantısı olan farklı gözlem sayısı içeren veri setlerinde, çoklubağlantının etkisinin gözlem sayısına göre nasıl değiştiğinin incelenmesi amaçlanmıştır.

Gereç ve Yöntemler: Bu amaçla, farklı gözlem sayısı içeren veri setleri benzetim tekniğiyle türetilmiştir. Veri setleri türetilirken gözlem sayıları 1000'den 10000'e kadar olan 10 veri seti oluşturulmuştur. Veri setlerinde 3 bağımsız (x_1, x_2, x_3) ve bir bağımlı değişken (y) kullanılmıştır. Bağımlı değişken ortalaması 60, standart sapması 5, bağımsız değişkenler ise ortalamaları sırasıyla 40, 50, 90 ve standart sapmaları sırasıyla 4, 8, 12 olacak şekilde türetilmiştir. Bağımsız değişkenler arasındaki korelasyonlar; x_1 ile x_2 arasında 0.8, x_1 ile x_3 arasında 0.9, x_2 ile x_3 arasındaki 0.5, y ile x_1 arasındaki 0.75, y ile x_2 arasındaki 0.65, y ile x_3 arasındaki 0.55 olarak tanımlanarak çoklu bağlantı sorunu elde edilmiştir.

Bulgular: Her iki regresyon tekniği de uygulandığında tüm veri setleri için tüm bağımsız değişkenler ve tümel modeller istatistiksel olarak anlamlı bulunmuştur. Çoklubağlantının sonuçlarından biri olarak en küçük kareler yönteminde iki bağımsız değişkenin regresyon katsayısının işareti, bağımlı değişken ile arasındaki ilişkinin yönüne ters olarak bulunmuştur. Temel bileşenler regresyonunda ise bu sorunun ortadan kalkmıştır. Bununla birlikte, temel bileşenler regresyonundan elde edilen katsayıların standart hataları, en küçük kareler yöntemine göre daha küçük elde edilmiştir. Ayrıca gözlem sayısı baz alındığında da gözlem sayısı arttıkça beklenildiği üzere katsayıların standart hataları daha da azalmıştır.

Sonuçlar: Üretilen 10 veri seti için yapılan en küçük kareler ve temel bileşenler regresyonu analizi sonucunda çıkan regresyon katsayılarının işaretlerinden ve standart hatalarından yararlanılarak temel bileşenler regresyonunun veri sayısı arttıkça regresyon katsayılarına ait standart hatayı en küçük kareler yöntemine göre daha da azalttığı gözlemlenmiştir.

Anahtar Sözcükler: çoklubağlantı, en küçük kareler regresyonu, temel bileşenler regresyonu, örneklem genişliği

Kaynaklar

1. ŞAHİNLER, S.(2000). En Küçük Kareler Yöntemi ile Doğrusal Regresyon Modeli Oluşturmanın Temel Prensipleri. MKÜ Ziraat Fakültesi Dergisi 5 (1-2): 57- 73.
2. ALBAYRAK, A.S.2005. Çoklu Doğrusal Bağlantı Halinde En küçük Kareler Tekniğinin Alternatifi Yanlı Tahmin Teknikleri Ve Bir Uygulama. ZKÜ Sosyal Bilimler Dergisi Cilt 1, Sayı 1

YAZARLAR DİZİNİ / INDEX

A. Ergun KARAAĞAOĞLU.....	2, 54	Dinçer GÖKSÜLÜK.....	2,5,46,48,49,50,77,78,106
Abdullah Oğuz TAŞBAŞ.....	126	Doğukan ÖZEN.....	87, 96
Adnan KARAİBRAHİMOĞLU.....	40	Durdu KARASOY.....	9
Ahmet Kadir ARSLAN.....	16, 93	Duygu SAÇAR.....	30
Ahmet KESKİNOĞLU.....	116	Duygu SİDDİKOĞLU.....	23
Ahmet ÖZTÜRK.....	5,46,48,50,77,78,106	Ebru TURGAL.....	10
Ahmet URSAVAŞ.....	99	Eda KARAİSMAİLOĞLU.....	54
Ahu DURMUŞÇELEBİ.....	106	Elif ERTAŞ.....	129
Alev BAKIR.....	107	Emek GÜLDOĞAN.....	100
Ali BEYTUR.....	122	Emine Arzu KANIK.....	129
Anıl DOLGUN.....	52	Emine KADIOĞLU.....	58
Anıl ÖZGÜR.....	103	Emrah G. ÖZGÜR.....	118
Arslan TERLEMEZ.....	40	Emre DEMİR.....	23, 114, 115
Asena Ayça ÖZDEMİR.....	128, 129	Ender UZABACI.....	97
Aslı BOZER.....	7	Erdal COŞGUN.....	27
Atilla Halil ELHAN.....	34, 60, 104, 115	Erdem KARABULUT.....	58, 65, 70
Ayfer Ezgi YILMAZ.....	56	Erdoğan ASAR.....	70
Ayhan PARMAKSIZ.....	19	Erkut BOLAT.....	83, 123
Azer ABİYEYEV.....	94	Erman GÜLENDAG.....	87, 96
Bahar TAŞDELEN.....	88, 101	Ersin ÖĞÜŞ.....	64, 65
Batuhan BAKIRARAR.....	33, 60, 104, 113	Ertuğrul ÇOLAK.....	4, 72
Berat Niyazi ÖLMEZ.....	126	Evren KÖSE.....	93
Bernd KLAUS.....	48	Ezgi DEMİRDÖĞEN ÇETİ.....	99
Betül ÇİÇEK.....	50	Fatma Ezgi CAN.....	37, 63, 97, 99, 117
Betül DAĞOĞLU.....	89, 102	Ferhan ELMALI.....	106, 110
Beyza DOĞANAY ERDOĞAN.....	10	Fisun KAŞKIR KESİN.....	69, 97
Bilal YILMAZ.....	126	Funda İPEKTEN.....	77, 78
Bulut DURAL.....	119	Funda Seher ÖZALP ATEŞ.....	23, 34
Burak ORDİN.....	116	Gökhan OCAKOĞLU.....	99
Burçin KURT.....	112	Gökmen ZARARSIZ.....	5,45,46,48,49,50,77,78,106
Bülent ÇELİK.....	29	Gözde ERTÜRK ZARARSIZ.....	5,46,50,77,78,106
Büşra EMİR.....	4	Gülay YEGİNOĞLU.....	61
Can ATEŞ.....	26, 57	Gülhan OREKİCİ TEMEL.....	13, 103, 128
Canan ERTEMOĞLU ÖKSÜZ.....	61	Güven ÖZKAYA.....	37, 117
Caner ALPARSLAN.....	116	H. Emre KALE.....	57
Cemil ÇOLAK.....	6, 16, 39, 93, 100, 109, 122	Halef Okan DOĞAN.....	77, 78
Cengiz BAL.....	12	Halil BİLGİN.....	46, 50, 106
Ceren EFE.....	14, 15	Harika Gözde GÖZÜKARA.....	25, 119, 130
Deniz ÖZONUR.....	41	Hatice İŞİK.....	9
Deniz SIGIRLI.....	117	Hatice Tül Kübra AKDUR.....	41
Derya GÖKMEN.....	34, 115	Hatice Yağmur ZENGİN.....	64, 65
Didem DERİCİ YILDIRIM.....	88		

Hülya BAYRAK.....	41	Nihal ATA TUTKUN.....	9
Hülya BİNOKAY.....	108	Nihat Burak ZİHNİ.....	61
Hülya ÖZEN.....	12, 87	Nikolay NINOV.....	45
İ. Safa GÜRCAN.....	87, 9	Nural BEKİROĞLU.....	30, 118
İbrahim TÜRKBEYLER.....	94	Nurhan DOĞAN.....	90, 124, 126, 127
İlker ERCAN.....	37, 43, 63, 97, 99, 117	Osman DAĞ.....	7, 52
İlker ÜNAL.....	14, 15, 67	Osman DEMİR.....	85
İrem ERSÖZ KAYA.....	13	Oya KALAYCIOĞLU.....	42
İrem KAR.....	33, 60, 104, 113	Ömer Faruk BALLI.....	126
İsmet DOĞAN.....	90, 124, 127	Önder YAVAŞCAN.....	116
Julide YAĞMUR.....	100	Özge ÇAPKAN.....	68
K. Setenay ÖNER.....	64	Özlem ARIK.....	58
Kaan ESEN.....	103	Özlem KAYMAZ.....	57
Kemal TURHAN.....	32, 44, 112	Özlem TAŞKAPILIOĞLU.....	37
Kübra Elif AKBAŞ.....	25	Özlem UZUN.....	61
Leyla KARACA.....	119	P. Fulden YUMUK.....	30
M. Akif BAKIR.....	29	Pembe KESKİNOĞLU.....	68, 116
M. Akif ÖZTÜRK.....	30	Pınar AMBARCIOĞLU.....	87, 96
M. Yavuz ATA.....	29	Pınar Günel KARADENİZ.....	92, 97
Mehmet KARADAĞ.....	95	Pınar ÖZDEMİR.....	38
Mehmet KIVRAK.....	109	Reha ALPAR.....	19
Mehmet MENDEŞ.....	1, 21, 22, 35	Rümeysa KILINÇ.....	126
Mehmet ORMAN.....	68	S. Kenan KÖSE.....	II, 10, 33, 113, 114, 115
Mehmet Ziya FIRAT.....	73	Saim YOLOĞLU.....	95, 105, 119, 130
Mehtap AKÇİL OK.....	79, 96	Saliha Nur SEZGIN.....	126
Meltem ÜNLÜSAVURAN.....	77, 78	Sarp SAKLIYAN.....	109
Mert İŞERİ.....	68	Sedat ALTIN.....	83
Merve BAŞOL.....	2	Sedat YAŞAR.....	105
Merve TÜRKEGÜN.....	13	Selçuk KORKMAZ.....	5, 46, 48, 49, 50, 77, 78, 106
Meryem ÇAVUŞOĞLU.....	5	Sema ATIŞ.....	97
Meryem GÖRAL YILDIZLI.....	98	Semra ERDOĞAN.....	84, 103
Mitat AKKOÇ.....	105	Senem TURAN OZDEMİR.....	99
Muammer ALBAYRAK.....	32	Serhat HAYME.....	115
Musa Bashir ALBISHIR.....	43	Sermin TOK.....	103
Mustafa Agah TEKİNDAL.....	57	Serpil AKTAS.....	56
Mustafa BOĞAN.....	95	Seval KUL.....	94, 95, 121
Mustafa Ş. ŞENOC AK.....	107	Sevilay KARAHAN.....	7
Mutlu UMAROĞLU.....	38	Sevinç Püren YÜCEL.....	18, 89, 102
Muzaffer BİLGİN.....	72	Seyran BOZKURT BABUŞ.....	84
Mümtaz MAZICIOĞLU.....	5	Sharan Kaur JANJUHA.....	45
N. Meric KONAR.....	52, 54	Siddık KESKİN.....	I, 76
Nazlı TOTİK.....	36, 82	Sinan UZUN.....	30
Necdet SÜT.....	49	Soner YİĞİT.....	1, 21, 22, 35

Songül ÇUĞLAN.....	93
Sumeet Pal SINGH	45
Suriye ÖZGÜR	68, 116
Ş. Büşra MERT	68
Şaban TEKİN	105
Şahi Nur KALKIŞIM.....	61
Şehim KUTLAY	34
Şeyma YAŞAR.....	105, 119, 122, 130
Tanyeli Kazaz.....	121
Tarık Alp ŞAHINOĞLU	126
Timur KÖSE	68, 116
Tolga BERBER	24
Tuba DENKÇEKEN	92
Tuğba KUL KÖPRÜLÜ	105
Tuççe ŞENÇELİKEL.....	64
Ufuk KAYA.....	87, 96
Uğur ŞEVİK.....	74

Uğur TOPRAK	44
Vahap ELDEM.....	46, 48, 77, 78
Vedia Bennu GİLAN.....	84
Veysel YILMAZ	69
Volkan ÖZKAYA	50
Yasemin YAVUZ	23, 26, 32, 114
Yaşar SERTDEMİR.....	14, 67, 108
Yeşim AKBAŞ	24
Yunus Emre KUYUCU	85
Yüksel BALCI.....	103
Zeliha AYDIN KASAP	32, 112
Zeliha Nazan ALPARSLAN	18,36,82,89,98,102
Zeynel Abidin ÖZTÜRK.....	94
Zeynep TUNÇ	119, 130
Zuleyha ALPER.....	117