# XX$^{th}$ NATIONAL
# and
# III$^{rd}$ INTERNATIONAL
# BIOSTATISTICS CONGRESS
# BOOK of PROCEEDINGS


**Editor in Chief**
Prof. Vildan SÜMBÜLOĞLU

**Associate Editors**
Assoc. Prof. Seval KUL
Assist. Prof. Pınar GÜNEL KARADENİZ
Assist. Prof. İlkay DOĞAN



26-29 October 2018
Gaziantep
TURKEY

**XX$^{th}$ National and III$^{rd}$ International Biostatistics Congress**
**26-29 October 2018 GAZIANTEP**

# ORGANIZING COMMITTEE

| **Honorary Board** | |
|---|---|
| Abdulkadir KONUKOĞLU | SANKO University, Chair of the Board of Trustees |
| Prof. Güner DAĞLI | SANKO University, Rector |
| Prof. Salih Murat AKKIN | SANKO University, Dean of Faculty of Medicine |
| Prof. Kadir Sümbüloğlu | Hacettepe University, Faculty of Medicine Department of Biostatistics Retired Professor, Honorary President of Association of Biostatistics of Turkey |

| **Chairs of the Congress** | |
|---|---|
| Prof. Erdem KARABULUT | President of Association of Biostatistics of Turkey, Hacettepe University Faculty of Medicine Department of Biostatistics |
| Prof. Vildan SÜMBÜLOĞLU | SANKO University Faculty of Medicine Department of Biostatistics |

| **Organizing Committee** | | |
|---|---|---|
| Prof. İsmet DOĞAN | Afyonkarahisar Health Sciences University | Member |
| Prof. Atilla Halil ELHAN | Ankara University | Member |
| Assoc. Prof. S. Kenan KÖSE | Ankara University | Member |
| Assoc. Prof. Seval KUL | Gaziantep University | Member |
| Assist. Prof. Gökmen ZARARSIZ | Erciyes University | Member |
| Instructor Tanyeli KAZAZ | Gaziantep University | Member |

| **Congress Scientific Secretary** | | |
|---|---|---|
| Assist. Prof. Pınar Günel KARADENİZ | SANKO University | Member |
| Assist. Prof. İlkay DOĞAN | Gaziantep University | Member |
| Dr. Mehmet KARADAĞ | İnönü University | Member |

**XX$^{th}$ National and III$^{rd}$ International Biostatistics Congress**
**26-29 October 2018 GAZIANTEP**

## SCIENTIFIC ADVISORY BOARD

| | |
|---|---|
| Prof. A. Ergun KARAAĞAOĞLU | Hacettepe University |
| Prof. Ahmet DİRİCAN | İstanbul University |
| Prof. Ahmet ÖZTÜRK | Erciyes University |
| Prof. Alan TENNANT | University of Lucerne, Switzerland |
| Prof. Atilla Halil ELHAN | Ankara University |
| Prof. A. Canan YAZICI GÜVERCİN | İstanbul Aydın University |
| Prof. Bahar TAŞDELEN | Mersin University |
| Prof. Beyza AKDAĞ ŞAHİN | Pamukkale University |
| Prof. Celal Reha ALPAR | Hacettepe University |
| Prof. Cemil ÇOLAK | İnönü University |
| Prof. Erdem KARABULUT | Hacettepe University |
| Prof. Ersin ÖĞÜŞ | Başkent University |
| Prof. Fezan MUTLU | Eskişehir Osmangazi University |
| Prof. Gülşah SEYDAOĞLU | Çukurova University |
| Prof. Hafize SEZER | İstanbul Aydın University |
| Prof. Handan ANKARALI | Medeniyet University |
| Prof. İlker ERCAN | Uludağ University |
| Prof. İmran KURT ÖMÜRLÜ | Adnan Menderes University |
| Prof. İsmet DOĞAN | Afyonkarahisar Health Sciences University |
| Prof. K. Setenay ÖNER | Eskişehir Osmangazi University |
| Prof. M. Nurullah ORMAN | Ege University |
| Prof. M. Yusuf ÇELİK | Biruni University |
| Prof. Mehtap AKÇİL OK | Başkent University |
| Prof. Meriç ÇOLAK | Başkent University |
| Prof. Mevlüt TÜRE | Adnan Menderes University |
| Prof. Mustafa ŞENOCAK | İstanbul University |
| Prof. Nazan ALPARSLAN | Çukurova University |
| Prof. Necdet SÜT | Trakya University |
| Prof. Nural BEKİROĞLU | Marmara University |
| Prof. Ömer SATICI | Dicle University |
| Prof. Pınar ÖZDEMİR | Hacettepe University |
| Prof. Saim YOLOĞLU | İnönü University |
| Prof. Sanjib BASU | Northern Illinois University, USA |
| Prof. Sıddık KESKİN | Yüzüncü Yıl University |
| Prof. Vildan SÜMBÜLOĞLU | SANKO University |
| Prof. Yasemin YAVUZ | Ankara University |
| Prof. Yavuz SANİSOĞLU | Yıldırım Beyazıt University |
| Prof. Zeki AKKUŞ | Dicle University |
| Prof. Peter W. ROSE | University of California, USA |

| | |
|---|---|
| Prof. Rama SHANKER | Eritrea Inst. of Tech., Eritrea |
| Assoc. Prof. Alex R. de LEON | University of Calgary, Canada |
| Assoc. Prof. Beyza DOĞANAY ERDOĞAN | Ankara University |
| Assoc. Prof. Bülent ÇELİK | Gazi University |
| Assoc. Prof. Canan BAYDEMİR | Kocaeli University |
| Assoc. Prof. Deniz SIĞIRLI | Uludağ University |
| Assoc. Prof. Derya ÖZTUNA | Ankara University |
| Assoc. Prof. Ertuğrul ÇOLAK | Eskişehir Osmangazi University |
| Assoc. Prof. Ferhan ELMALI | İzmir Katip Çelebi University |
| Assoc. Prof. Gökhan OCAKOĞLU | Uludağ University |
| Assoc. Prof. Gülhan ÖREKECİ TEMEL | Mersin University |
| Assoc. Prof. Güven ÖZKAYA | Uludağ University |
| Assoc. Prof. Hakan DEMİRTAŞ | University of Illinois, USA |
| Assoc. Prof. İlker ETİKAN | Yakın Doğu University |
| Assoc. Prof. Jale KARAKAYA KARABULUT | Hacettepe University |
| Assoc. Prof. Lynn SLEEPER | Harvard Medical School, USA |
| Assoc. Prof. Melody GHAHRAMANİ | University of Winnipeg, USA |
| Assoc. Prof. Nurhan DOĞAN | Afyonkarahisar Health Sciences University |
| Assoc. Prof. Recai YÜCEL | University at Albany-SUNY, USA |
| Assoc. Prof. S. Kenan KÖSE | Ankara University |
| Assoc. Prof. Semra ERDOĞAN | Mersin University |
| Assoc. Prof. Seval KUL | Gaziantep University |
| Assoc. Prof. Şengül CANGÜR | Düzce University |
| Assoc. Prof. Ünal ERKORKMAZ | Sakarya University |
| Assoc. Prof. Timur KÖSE | Ege University |
| Assoc. Prof. Yaşar SERTDEMİR | Çukurova University |
| Assist. Prof. Duchwan RYU | Northen Illinois University, USA |
| Assist. Prof. Umut ÖZBEK | Icahn School of Medicine at Mount Sinai, USA |
| Dr. Atanu BHATTACHARJEE | Assam, India |
| Dr. Bernd KLAUS | European Mol. Bio. Lab.,Germany |

# ORAL PRESENTATIONS

Sayfa

O23. Damla Hazal SARAL, Tuğba ARIKOĞLU, Bahar TAŞDELEN. The Determination of Optimal Test Combination by Evaluating The Units of the Different Diagnostic Tests Used in The Estimation of Airway Hyperresponsiveness Using Boole Algebra and Latent Class Analysis

O24. Meral AYDIN, Pınar ÖZDEMİR. Reducing the Bias in the Use of Imperfect Gold Standard Test

O25. Seda KESKİN, Neslihan BAKİ, Kemal TURHAN, Burçin KURT, Melek BULUT KUZU, Deha Denizhan KESKİN, Zeliha AYDIN KASAP. A Forecast Model Proposal for HPV Development Risk in Women

O26. İlker ÜNAL, Yaşar SERTDEMİR, Ceren EFE, Esin ÜNAL. Type II Error Rates for Measuring the Classification Accuracy of a Biomarker in Two Groups

O27. Necdet SÜT, F. Betül ÖRS. Effect of Kernel Density Estimation on Naive Bayes Classification Results

O28. Duygu KORKMAZ, Sıddık KESKİN. Multivariate Regression Tree Method and Its Application

O29. Selcen YÜKSEL, Afra ALKAN, Pervin DEMİR. Evaluation of Receiver Operating Characteristic Analysis to Determine the Minimum Clinically Important Difference on Interpretability of Scales

O30. Harika Gözde GÖZÜKARA BAĞ, Yücel DUMAN. 3-Way ROC Analysis: An Illustration on Ertapenem Sensitivity

O31. Şeyma YAŞAR, Zeynep TUNÇ, İpek BALIKÇI ÇİÇEK, Ayşe Nur AKATLI, Leyla KARACA, Ramazan ALTINTAŞ, Cemil ÇOLAK, Saim YOLOĞLU. The Evaluation of Agreement between Measurement Methods used in Diagnosis and the Comparison by Statistical Methods

O32. Gülden HAKVERDİ, Ömer Faruk DADAŞ, Timur KÖSE, Mehmet N. ORMAN. Utilization of Odds Ratio in Prospective Studies

O33. İpek BALIKÇI ÇİÇEK, Zeynep TUNÇ, Şeyma YAŞAR, Ahmet Kadir ARSLAN, İbrahim ŞAHİN, Cemil ÇOLAK, Saim YOLOĞLU. A Comparison of Multivariate Statistical Methods to Detect Risk Factors For Type 2 Diabetes Mellitus

O34. Yasar SERTDEMİR, İlker ÜNAL, Hülya BİNOKAY. Comparing the Performance of Logistic Regression and Some Classification Methods Using Real Data Sets

O35. Hülya BİNOKAY, Yaşar SERTDEMİR. Comparing the Performance of Logistic Regression and Different Classification Models

O36. Meltem ÜNLÜSAVURAN, Gözde ERTÜRK ZARARSIZ, Gökmen ZARARSIZ, Funda İPEKTEN, Selçuk KORKMAZ, Dinçer GÖKSÜLÜK, Halef Okan DOĞAN, Vahap ELDEM, Ünal ERKORKMAZ, İsmail KOÇYİĞİT, Zafer AYDIN, Ahmet ÖZTÜRK. Investigation of H2O Automated Machine Learning Algorithm Performance on Metabolomic Data

O37. Senem KOÇ, Leman TOMAK, Sezgin GÜNEŞ. Classification of Chronic Peridontitis by Using Different Machine Learning Algorithms

O38. Osman DAĞ, Erdem KARABULUT, C. Reha ALPAR. GMDH-Type Neural Network Algorithms for Binary Classification: GMDH2 R Package and Its Web Interface

O39. Emre DEMİR, Bedreddin Ali AKÇA, Serhat HAYME, Serdal Kenan KÖSE. A New Mobile Application for Sample Size Calculations in Medical Research

O40. Şengül CANGÜR, Handan ANKARALI. Comparison of Fast Regression Methods Used in Model Selection in High Dimensional Data

O41. Hüseyin KUTLU, İlker ETİKAN. Classification of ECG Data by Deep Learning Methods
O42. Pervin DEMİR, Atilla Halil ELHAN, Şehim KUTLAY, Ayşe A. KÜÇÜKDEVECİ. Developing a Common Metric for Assessment of Functioning in Patients with Osteoarthritis
O43. Alev BAKIR, Aydın ERAR. Projection Pursuit Method, Algorithms and Indexes
O44. Ebru OZTURK, Batuhan BAKIRARAR, Sevilay KARAHAN, Ergun KARAAGAOGLU. An Extensive Web Interface for Validity and Reliability with RServe
O45. Batuhan BAKIRARAR, İrem KAR, Atilla Halil ELHAN. Developing Web-based Software Clustering Algorithms with RServe
O46. Batuhan BAKIRARAR, İrem KAR, Beyza DOĞANAY ERDOĞAN, Atilla Halil ELHAN. Assessing Performance of Missing Value Imputation Methods Using Data Mining Classification Algorithm
O47. Uğur TOPRAK, Emrah Gökay ÖZGÜR, Serhat HAYME, Beyza DOĞANAY ERDOĞAN. Comparing Prediction Abilities of Classification Algorithms on Determining Specific Chemical Substances
O48. Çağla ŞAFAK, Zeynep YAVUZ, Beyza DOĞANAY ERDOĞAN. Functional Data Analysis and an Application of the "Functional t-test"
O49. Ferhan ELMALI, Funda İfakat TENGİZ. Innovation in Biostatistics Education: Multidisciplinary Approach in Undergraduate Medical Education
O50. Ayhan PARMAKSIZ, C. Reha ALPAR. Power Analysis in Multiple Linear Regression Models

# SHORT ORAL PRESENTATIONS

Sayfa

SO18. Sıddık KESKİN, Birhan KUNTER, Nurhan KESKİN. Principle Coordinates Analysis and An Application

SO19. Nazlı TOTİK, Sevinç Püren YÜCEL, Zeliha Nazan ALPARSLAN. Investigation of Causal Structure Using Mediation Analysis

SO20. Nurgül BULUT, Hayriye ERTEM VEHİD, Setenay ÖNER. Comparison of Two Independent Group Data with Quantitative Data with Three Different Methods

SO21. Yağmur POLAT, Mesut AKYOL, S. Yavuz SANİSOĞLU. New Experimental Design Model Based on Mutation: the Umbrella Design

SO22. Çağla SARITÜRK, Funda PEPEDİL TANRIKULU, Nazan ŞEN, Nurhilal BÜYÜKKURT, İlknur KOZANOĞLU. A novel approach for analysing differential cell count of Bronchoalveolar lavage samples: Flow-cytometry

SO23. Emre DİRİCAN, Zeki AKKUŞ, Erol KILIÇ, İsmail YILDIZ, Ömer SATICI. Machine Learning Approaches in Ki-67 Scoring

SO24. Ragıp Onur ÖZTORNACI, Bahar TAŞDELEN, Mehmet Tevfik DORAK. Using a Feature Selection Method for Machine Learning Algorithms on GWAS Data

SO25. Su Özgür, Turhan KAHRAMAN, Mete EMİNAĞAOĞLU, Ahmet ÖZKURT, Berril DÖNMEZ ÇOLAKOĞLU, Betül YÜRDEM, Uğur ELİİYİ, Pembe KESKİNOĞLU, Arzu GENÇ. Analysis of EEG and Motion Signal Data for Walking Action in a Neurological Disease

SO26. Aslı BOZ, Gülşah SEYDAOĞLU, Özge VURAL. The Multi-Criteria Decision Methods in Precision Medicine - Individual Treatment

SO27. Afra ALKAN, Pervin DEMİR, Selcen YÜKSEL. Two New Measures to Evaluate the Accuracy of Screening Tests in Diagnostic Studies Involving Paired Organs: Binocular Sensitivity and Binocular Specificity

SO28. Meryem GÖRAL YILDIZLI, Zeliha Nazan ALPARSLAN. Data Envelopment Analysis Approach For Measuring the Efficiency of University Hospital Units

SO29. Serdar DENİZ, Mehmet KIVRAK, Sarp ÜNER. The Frequency of Obesity and Related Factors in Kindergarden, Primary and Middle School Students in Mersin

SO30. Betül DAĞOĞLU HARK, Zeliha Nazan ALPARSLAN. Marginal Modeling of Clustered Data with MAR Mechanism

SO31. Ceren EFE, İlker ÜNAL. Comparison of Variable Selection Criteria: A Simulation Study

# ORAL PRESENTATIONS

# O1

# Novel Statistical Approaches in Clustering RNA-Sequencing Data

**Ahu Durmuşçelebi**[1], Gökmen Zararsız[1,2], Gözde Ertürk Zararsız[1], Selçuk Korkmaz[2,3], Dinçer Göksülük[2,4], Vahap Eldem[5], Ahmet Öztürk[1,2]

[1]*Erciyes University Medical Faculty Department of Biostatistics, Kayseri*
[2]*Turcosa Analytics Solutions Ltd. Co. Erciyes Teknopark, Kayseri*
[3]*Trakya University Medical Faculty Department of Biostatistics, Edirne*
[4]*Hacettepe University Medical Faculty Department of Biostatistics, Ankara*
[5]*İstanbul University Faculty of Science Department of Biology, İstanbul*

ahudurmuscelebi.87@gmail.com

**Objective**: RNA-sequencing (RNA-Seq) is a new transcriptomics tool which is one of the next-generation sequencing technologies and can perform their operations quickly and cheaply based on the principle of high-throughput sequencing. When compared to microarrays, RNA-sequencing offers a number of advantages: (i) having less noisy data, (ii) ability to detect new transcriptome and coding regions, (iii) not requiring the preparation of transcripts of interest. One major task using gene expression datasets is the clustering of samples and detecting the disease subtypes. The aim is to find samples with similar expression levels that could lead to the discovery of new cancer subtypes. Even both microarrays and RNA-Seq are techniques used for the same purpose, microarray based algorithms cannot directly be applied to RNA-Seq data. The reasons for this are the fact that RNA-sequencing data have non-negative integers, and the RNA-sequencing data need to be modeled correctly in order to obtain unbiased results due to the over-dispersion problem. In the first studies, RNA-sequencing data were transformed into microarrays and microarray based algorithms are used to these transformed data. In later studies, it has been suggested that RNA-sequencing data should be analyzed by statistical methods designed specifically for discrete counts. However, since the mathematical theory of discrete distributions is less tractable than the normal distribution, these distributions tend to limit both the performance and usefulness of RNA-sequencing analysis methods. In this study, it is aimed to integrate the existing clustering methods with the voom method, which estimates the mean and variance relationship of log-counts and produces precision weights for each observation to be used in subsequent analyzes, thus benefiting from the power of the normal distribution in count data.

**Method**: The voom transformation outputs, i.e. log-cpm transformed values and the precision weights were integrated into weighted distance matrices. Both hierarchical and k-means clustering algorithms were applied to the weighted distance matrices, which is obtained from raw, normalized and transformed RNA-Seq data. In addition, Poisson clustering method was used as well for comparisons. The clustering results were compared using adjusted Rand index. The analyses were applied in two real RNA-Seq datasets in R programming language (www.r-project.org ).

**Results**: For both clustering methods, while the developed approaches show less performance in some scenario, it has been observed that they give superior results in most of the scenario as compared to other methods.

**Conclusion**: It has been proposed that this new clustering approach is comparable with previous methods and researchers can evaluate them among the available methods.

**Keywords**: Gene expression, clustering, RNA sequencing, voom, next generation sequencing

This study was supported by Scientific Research Projects (BAP) Coordination Unit in ERCIYES UNIVERSITY with TYL-2018-8142 project number.

# O2

# The Effect of SNP Number, Minor Allel Frequency and Sample Size on the Performance of Machine Learning Methods in Genetic Association Studies

**Ragıp Onur Öztornacı**[1], Bahar Taşdelen[1]

[1]*Mersin University Department of Biostatistics and Medical Information, Mersin*

onur.oztornaci@gmail.com

**Objective:** The study aim is investigation to determine performance of machine learning methods in the genetic association studies increasingly nowadays to detect SNPs related to disease. Using simulated data, the effect of sample size, number of SNPs and minor allele frequency on SNP selection is evaluated.

**Method:** First, the SNP data frames are simulated that different size and different structures for using machine learning algorithms. Additive heritage model was used in the data frames. 10-fold cross validation was used all the data frames and thus it was avoided to overfitting problem in the training data frame. Dataset was divided to training (67%) and testing (33%) sets, in this way the validation of the models was tested. R language was used for the all simulation and machine learning algorithms. The simulation steps are,

- Case and control groups to be two groups has been created,
- All the sample sizes are n=50, 100, 200, 500,
- Number of the SNPs are k=10, 30, 50, 100, 250, 500, 1000,
- Minor allele frequencies are p=0.001, 0.01, 0.05,
- Heritage model are AA, AB, BB,

The data was created by following these steps.

**Results:** According to investigated methods, support vector machine has one of the best results. If number of SNPs could be over the 250, it can be reached 99% accuracy rate. Between the methods of C4.5 and C.5 has no big differences, but it can be saying that getting the result of C5 is faster than J48. To use Random Forest method at the small number of SNPs in the small sample size is obtaining higher accuracy rate and it should be consider to high sensitivity rate.

**Conclusions:** As a result, as the number of SNPs increases, there is also an increase in the accuracy. For all scenarios, support vector machine method generally has the best classification performance; logistic regression, C5.0, C4.5 (J48) methods follow support vector machines. As the minor allele frequency drops, the likelihood of capturing SNPs in the genome is also reduced for small sample sizes and the accuracies of methods reduce.

**Keywords:** Classification, single nucleotide polymorphism, simulation study

# O3

# Smoothing Spline ANOVA Models

**Nurgül Bulut**[1], Handan Ankaralı[1]

[1]*İstanbul Medeniyet University Faculty of Medicine Department of Biostatistics and Medical Informatics, İstanbul*

nurgul.bulut@medeniyet.edu.tr

**Objective:** The simple linear prediction model consists of a dependent variable (Y), an independent variable (X) and an error term (ε). Excessive or noisy data in the data set may affect the average of Y for a given X value. Smoothing Spline ANOVA (SS-ANOVA) models; is a family of versatile correction methods used for models with noisy univariate, multivariate and longitudinal data. SS-ANOVA model for proportional hazard model from this family is explained with an example and it is aimed to show that the use of these models in the presence of noisy data is more appropriate.

**Method:** Smoothing spline; it is used to estimate an unknown function *f(x)* giving the relation between variables X and Y properly. Semiparametric or nonparametric models are preferred because of the possibility of impaired normality and the increased frequency of noisy data in multivariate data. For this reason, smoothing methods are used to estimate the functions obtained. By dividing the intervals defined by the smoothing spline into subintervals with the help of observation values, we model the relation between a polynomial and the variables X and Y in each subinterval to obtain a continuous function with the desired grade. In our work; The SS-ANOVA model was used to evaluate the survival of 184 patients who underwent cardiac transplantation at Standford between 1967 and 1980, the survival time after surgery, survival (live/death) and age variables. Parameter estimations were made by penalized likelihood method. The interaction effects of the model were examined using the Kullback-Leibler method.

**Results:** Using the SS-ANOVA model for the Hazard model, we examined whether the age and time variables are additive or interaction to the model and found that the variables will be included as separate factors in the model ($\rho_{additive}<\rho_{interaction}$). After installing the appropriate model, hazard estimation values were obtained using the "hzdrate.sshzd" function. Age variable charts were formed with the predicted values found.

**Conclusion:** One of the most common problems when using parametric models is the existence of extreme deviations in the data set. With SS-ANOVA Models it is possible to obtain more accurate estimation results by eliminating such noisy data. Especially SS-ANOVA models can be recommended when models with time-dependent longitudinal data are established.

**Keywords:** Smoothing spline ANOVA, penalized likelihood method, longitudinal data

# O4

# A Piecewise Exponential Model, Logit Model and Complementary Log-log Model in Ranked Set Sampling

**Nursel Koyuncu**[1], Nihal Ata Tutkun[1]

[1]*Hacettepe University Faculty of Science Department of Statistics, Ankara*

nkoyuncu@hacettepe.edu.tr

**Objective:** Continuous-time models are typically employed when exact timing of event is known whereas discrete time models are generally used when only the interval in which an event occurs is known or the event itself occurs in discrete intervals. Cox regression model (CRM) is the most popular survival model for continuous-time survival data, but it lacks the facility to test hypotheses about the shape of the hazard function. One way to get some of the flexibility of the CRM without losing the hypotheses testing capability is to employ the piecewise exponential model (PEM). In PEM, survival time scale is divided into intervals and PEM is similar to logit and complementary log-log models. PEM uses information on the exact timing of events, while the others are based on interval-censored data. In applications, studying with population can be difficult or costly. For this reason using correct sampling designs gains importance. In this study, our aim is to evaluate the effect of sampling designs to parameter estimation of the mentioned survival models.

**Method:** We conduct a simulation study using ranked set sampling with different sampling sizes and compare our results with classical simple random sampling. The analysis are done in R program using "survival" package.

**Results:** The parameter estimations are made with simulation study by different sample size and sampling designs using Exponential Model, Logic Model and Complementary Log-log Model.

**Conclusion:** As a result of the simulation study, it is seen that the parameter estimates using ranked set sampling have smaller mean square error.

**Keywords:** Ranked set sampling, survival analysis, Monte Carlo simulation

# O5

# The Analysis of Breast Cancer Patients' Data Using Cure Model

**Şirin Çetin**[1], İsa Dede[2]

[1]*Ondokuz Mayıs University, Samsun*
[2]*Mustafa Kemal University Faculty of Medicine Department of Medical Oncology, Hatay*

cetinsirin55@gmail.com

**Objective**: The aim of this study is to review the definition of a cure model, explain when cure models can be used, and use a breast cancer data set as a sample in describing a cure model. Breast cancer is a heterogeneous disease caused by many complicated factors. The causes and factors leading to its onset remains an obscure topic even today. As breast cancer is a heterogeneous disease, it shows differences with other natural diseases. The developing technology and increasing treatment options have proven to have a positive effect on both the survival period and cure rate of breast cancer patients. The analysis of data acquired from breast cancer patients is done using a different model from the routine survival analysis model. This is due to the fact that in the Cox proportional hazards model used in normal survival model an assumption that every patient will experience the event of interest if the follow up time is long enough is made. Yet this is not always applicable. This is because in some cancer conditions, for instance breast cancer the patient can get cured of the disease long after the cancer treatment and since the survival period of the patient being followed up is unknown at the end of the study, the Cox proportional hazards model isn't suitable. Hence the cure model is a better option.

**Method:** In this retrospective cohort study we examined 200 breast cancer patients referred to Medical Oncology Clinic of Antakya State Hospital between 2016 and 2018. The statistical analyses were computed using R version 3.3.2 software and SPSS software (version 21.0; IBM, Armonk, NY). The Cure model was computed using the R version 3.3.2. Survival analyses and Cure Models analysis were done and the result of considered variables on patient's survival was analyzed.

**Result**: The median of survival time of 121 breast cancer patients analyzed was determined as 48 months. Estrogen Receptor (ER) and c-erbB2 were found as important factors influencing the failure ($p<0.005$).

**Conclusion:** The Cure model is different from the Cox regression model in that it allows different interpretations to be made together with the calculated cure rates and this makes it the more important.

**Keywords**: Breast cancer, cure model, survival analysis

# O6

# Comparison of Performances of Different Smoothing Functions on Respiratory Diseases Data in Generalized Additive Models

**Mehmet Karadağ[1],** Seval Kul[2], Saim Yoloğlu[3], Behçet Al[4], Mustafa Boğan[5]

[1]*İnönü University Institute of Health Science Department of Biostatistics and Medical Informatics, Malatya*
[2]*Gaziantep University Faculty of Medicine Department of Biostatistics, Gaziantep*
[3]*İnönü University Faculty of Medicine Department of Biostatistics and Medical Informatics, Malatya*
[4] *Gaziantep University Faculty of Medicine Department of Emergency Medicine, Gaziantep*
[5] *Gaziantep Şehit Kamil Public Hospital Department of Emergency Medicine, Gaziantep*

mkarad@gmail.com

**Objective:** In this study, we aimed to compare the performance of Cyclic Cubic Regression Splines (CC), Thin Plate Regression Splines (TPRS) and Cubic Regression Splines (CRS) smoothing functions, which are frequently used in GAM (Generalized Additive Models) method, on the estimation of respiratory diseases.

**Method:** Between 1 January 2009 and 31 March 2014 a total of 1916 days, the number of hospitalized patients for of asthma, COPD, pneumonia and pulmonary thromboembolism was modeled as dependent variable. The smoothing function of the time was added as an estimator and the performance of different smoothing functions were compared. Performance criteria are AIC, Corrected AIC, BIC, Difference and $R^2$.

**Results:** In the light of previous studies in the models using GAM method, the degree of freedom of smoothing functions has been determined as 12. The models with CC, TPRS and CRS smoothing functions were compared to the performance criteria in the models in which the asthma disease dependent variable was estimated. It was observed that the smoothing function, which performs the best is CRS. The models with CC, TPRS and CRS smoothing functions were compared to the performance criteria in the models in which the KOAH disease dependent variable was estimated. It was observed that the smoothing function, which performs the best is TPRS. The models with CC, TPRS and CRS smoothing functions were compared to the performance criteria in the models in which the pneumonia disease dependent variable was estimated. It was observed that the smoothing function, which performs the best is TPRS. The models with CC, TPRS and CRS smoothing functions were compared to the performance criteria in the models in which the pulmonary thromboembolism disease dependent variable was estimated. It was observed that the smoothing function, which performs the best is TPRS.

**Conclusion:** In the GAM method, the performance of the models with Cyclic Cubic Regression Splines smoothing function in the independent variables was lower than the other smoothing functions**.** It was observed that the performance of Thin Plate Regression Splines and Cubic Regression Splines were close to each other.

**Keywords:** Smoothing functions, GAM, modeling of respiratory tract diseases

# O7

# miRNA Name Standardization in Meta-Analysis Studies

**Elif Pala**[1], Tuba Denkçeken[2]

[1]*SANKO University Faculty of Medicine Department of Medical Biology, Gaziantep*
[2]*SANKO University Faculty of Medicine Department of Biophysics, Gaziantep*

elifonur78@yahoo.com

**Objective**: MicroRNAs (miRNAs); are 22 nucleotides in length, endogenous, noncoding RNAs which are found in a variety of organisms, including plants, viruses and humans. miRNAs play a key role in important biological processes such as cell differentiation, proliferation and apoptosis. In recent years, increase in the number of studies with miRNAs has made meta-analysis favorite, which has led to an increase in the number of publications aimed to detect disease-related biomarkers. MiRBase is the most frequently used main database, considering the number of citations for miRNAs. The current version being used is MiRBase 22 version. Differences in the continuously updated versions of MiRBase cause inconsistencies in miRNA names and lead overlapping the studies. miRNAs can be added to or deleted from the database during development, and also their names may change. In previous version 21, there were 35828 mature miRNAs from 223 organisms, while in the current version of MiRBase version 22 there are 48885 mature miRNAs from 271 organisms. Likewise the mature miRNA hsa-miR-190a-5p was named hsa-miR-190 in MiRBase release 2.0. This name was changed to hsa-miR-190a in release 18.0, and was again renamed to hsa-miR-190a-5p in version 19.0. For this reason, in order to guide researchers correctly in meta-analysis studies, the names of detected miRNAs should be standardized and updated.

**Method:** Correction of such name inconsistencies can lead to difficulties and mistakes when working with large miRNA data sets in meta-analysis. *miRNAmeConverter* application is presented in this study to minimize such mistakes and to standardize the names of miRNAs.

**Results:** *miRNAmeConverter* is the only publicly available, free tool found in the Bioconductor package of the R Studio program (https://bioconductor.org/packages/release/bioc/html/miRNAmeConverter.html) and is used to translate miRNA names into the most up-to-date version, especially when working with large miRNA data sets. The main algorithm implemented enables translation of species-independent mature miRNA names to user selected MiRBase versions automatically and quickly. In addition to the MiRBase version specific miRNA name, the user can also verify the sequence of the corresponding miRNA.

**Conclusion:** In this study, the *miRNAmeConverter* which is developed in 2017, found within Bioconductor package in the R Studio program to eliminate inconsistencies in the mature miRNA names in different MiRBase versions is presented. The not yet widely used web interface is straightforward to use and is therefore accessible to researchers who are less familiar with the R programming language.

**Keywords:** MiRBase, miRNA, miRNAmeConverter, R Studio

# O8

# metaHUN: a web tool for Meta Analysis

**Mutlu Umaroğlu**[1], Pınar Özdemir[1]

[1]*Hacettepe University Faculty of Medicine Department of Biostatistics, Ankara*

mutlu.umaroglu@hacettepe.edu.tr

**Objective:** Meta-analysis is a statistical method that integration of the results from similar separate studies on the same subject. There are two general statistical methods are available for meta-analysis, namely, fixed effect and random effects model. In meta-analysis, combining the studies is usually done by effect size and generic inverse variance method (GIVM) is used. There are several software packages that implement for meta-analysis. However, there is no free, user friendly and comprehensive software. Therefore, it is aimed to develop an R-based, free, user-friendly and comprehensive meta-analysis tool with metaHUN software.

**Method:** metaHUN software allows the calculation of effect size from the data, meta-analysis from the calculated effect sizes, drawing the meta analytic plots, making cumulative meta-analysis and investigating bias and distant observations.

**Conclusion:** As a result, metaHUN software has been developed in addition to existing software. With metaHUN software, it is possible to perform meta-analysis with only user interface without writing code.

**Keywords:** Meta-analysis, R, metafor, web-tools

# O9

# A Meta Analysis of Pregnancy Rates of Dairy Cattles with Ovsynch Protocol in Turkey

Aytaç Akçay[1], Murat Abay[2], **Elif Çelik**[3]

[1]*Erciyes University Faculty of Veterinary Medicine Department of Biometrics, Kayseri*
[2]*Erciyes University Faculty of Veterinary Medicine Department of Obstetrics and Gynecology, Kayseri*
[3]*Erciyes University Graduate School of Health Sciences, Department of Animal Science, Kayseri*

elifcelik149@gmail.com

**Objective:** To improve the reproductive efficiency of the dairy cattle, detect oestrus and inseminate in the appropriate time period, eliminate anoestrus problems observed after birth, oestrus or ovulation synchronizations techniques are used. Ovsynch protocol is the most widely used ovulation synchronization technique with GnRH and PGF2α combinations. Cows with ovulation time can be inseminated in the desired time period and the necessary pregnancy rate for reproductive yield can be provided. In this study, in addition to estimating population ratio by combining pregnancy rates in herds with Ovsynch protocol, subgroup and meta regression analyzes according to years, animals, races and regions were performed to determine the sources of heterogeneity between studies.

**Method:** In the study 33 primitive studies carried out in Turkey between years 1999-2015 and made with Ovsynch protocol were included in the meta-analysis. These studies were selected from a total of 4.910 results from a search with the keyword "Ovsynch Protocol" in the Google Scholar search engine. The inclusion criteria included the use of Ovsynch synchronization in the study and pregnancy rates as effect size. In total, 1855 cows were analyzed. In the sub-group analysis conducted according to the regions, since there were not enough studies in the Southeastern Anatolia Region and the Black Sea Region, these regions were excluded from the analysis and 30 of these studies were used and 1759 cows were included in the analysis. Meta analyzes were performed by CMA (Comprehensive Meta Analysis) software.

**Results:** As a result of the meta-analysis, high heterogeneity was found between the studies ($Q= 56.911$, $df = 32$, $I^2 = 43.772$). Therefore, random effect model was used and pooled prevalence value was 0.400 (0.366-0.435) and it was significant ($p<0.001$). Subgroup analysis and meta regression analysis were performed to find the source of the heterogeneity between studies. As a result of these analyzes, the only source of heterogeneity was the regions variable and there was a significant difference between pregnancy rates of the Mediterranean Region, the Marmara Region, the Central Anatolia Region and the Aegean Region, and pregnancy rates of the Eastern Anatolia Region. According to meta regression model, Ovsynch protocol increases pregnancy rate of the Marmara Region by 1.01 units and decreases the pregnancy rate of the Eastern Anatolia by -1.50 units.

**Conclusion:** In conclusion, it has been concluded that systematic review of a large number of studies by meta-analysis can be an effective tool for increasing the fertility with Ovsynch protocol and to be a resource for future studies.

**Keywords:** Pregnancy rate, meta-analysis, meta-regression, ovsynch protocol, dairy cattle

# O10

## Application of Time Series Analysis to Clinical Data

**Kamber Kaşali**[1], Ahmet Dirican[2]

[1]*Atatürk University Faculty of Medicine Department of Biostatistics and Bioinformatics, Erzurum*
[2]*İstanbul University Faculty of Medicine Department of Biostatistics, Istanbul*

kmbrkasali76@gmail.com

**Objective:** Heart rate (HR), systolic blood pressure (SBP) and diastolic blood pressure (DBP), etc. In the studies done with variables like this, very variable analysis methods are frequently used in evaluating the data about the causality. Although trying to solve the relationship between the results of these approaches and the available time and criteria, it is not enough to correctly interpret the role of time causality. In order to investigate the relationship between these types of time-varying clinical variables and to reach the clinical facts, the "time series analysis" which is the original biostatistical method should be used. In this respect, it is also necessary to demonstrate that it is a more correct approach to analyze the clinical variables that change over time with time series analysis.

**Method:** Data used in the study; The 24-hour rhythm and blood pressure recorded in the patient's complaint of cardiology policlinic complained of blood pressure and heart attack were obtained from the holter results. 450 files from hospital records were obtained from 250 files, 125 male and 125 female according to the scanned criteria. Heart rate rates (HR), systolic blood pressure (SBP) and diastolic blood pressure (DBP) variables were recorded from the files that matched the criteria.

**Results:** According to the results, there is a causal relationship between HR and SBP and DBP for male and female patients. Because the p values are 0.0017 and 0.0084 for males and 0.0056 and 0.0001 for females, respectively. These p values indicate that the hypothesis Ho is rejected, which states that "the series of SBP and DBP are not the cause of granger of HR series". This result shows us that SKB and DBS can be used predicted that HR is granger cause.

In the other case, DBP was not the cause of granger in male patients ($p = 0.553$). In female patients DBP was the granger cause of DBP ($p = 0.0343$). In male and female patients, SBP does not appear to be a granger cause of DBP (male, $p = 0.126$; female, $p = 0.2156$). In male and female patients, both HR and both SBP and DBP are not the cause of granger (male, $p = 0.2649$, $p = 0.7100$; female, $p = 0.3749$, $p = 0.7026$).

**Conclusion:** The results show that the results of the time series analysis reveal more detailed results than the multiple comparison methods. According to the results of the time series analysis, it is shown that the correlations of HR and SBP and DBP variables are immediate, and that they are in balance in the long term.

In our study, it has been shown that applying time series analysis will give more detailed results to the time-varying data in the field of medicine.

**Key words:** Time series, cointegration, Granger causality

# O11

## Negative Binomial Regression and its Application

Sadi Elasan[1], Can Ateş[1], **Sıddık Keskin**[1]

[1]*Van Yüzüncü Yıl University Faculty of Medicine Department of Biostatistics, Van*

skeskin973@gmail.com

**Objective:** This study aims to introduce Negative binomial regression (NBR) which is one of the analysis methods that can be used when the dependent or response variable is count data with an application by giving information about it.

**Method:** When the dependent or response variable is expressed as the number of occurrence of the event for the count data obtained from any area and time interval, Poisson regression method is widely used to examine the relationship between this variable and the explanatory or independent variables. In order to apply this method, the response variable ($Y$) has a Poisson distribution, in other words, mean and variance equal. When the observed variance is higher than the mean of the Poisson data, overdispersion has occurred. In over-dispersed count data, the application of the Poisson regression lead to change the standard errors of the estimated statistics. In this case, it is proper to use Negative binomial regression model which takes into consideration overdispersion. The most basic or standard form of Negative binomial regression is NBR2 and this uses the natural log link function. NBR is similar to the standard Multiple regression analysis. Only difference is that response variable is count data. As designed for the Poisson-gamma mixture model in NBR, the parameters are estimated with Maximum likelihood using the Marquardt algorithm which is a similar form of the Newton-Raphson algorithm. As goodness of fit statistics, goodness of fit and residuals statistics are used. In addition to these; the score test and the Lagrange multiplier test can also be used. In the study, data obtained from free access website for the surgical age, number of axillary lymph nodes and survival times of 306 patients who had breast cancer surgery were used as application material. Axillary lymph node number was considered as the response variable and the relationships with surgical age and survival times were investigated. Poisson and Negative binomial regression models were used for this purpose.

**Results:** While approximately 25% of the patients had less than 5 years of survival, mean and standard deviation for the number of axillary lymph nodes were found 4.03 and 7.19, respectively. The coefficients for age and survival were 0.956 ± 0.0027 (p=0.001) and 2.743±0.157 (p=0.001) in Poisson regression while 0.983±0.0099 (p=0.092) and 2.770±0.636 (p=0.001) in Negative binomial regression. Log likelihood values of Poisson and Negative binomial regression were -1474.440 and -683.961, respectively.

**Conclusion:** As compared to the Poisson regression, a significant change was observed for the significance test of the coefficients to the independent variables in Negative binomial regression. Also, Log likelihood value shown two times decrease. Poisson and Negative binomial regression are very similar. There is no assumption that the mean and variance are equal for Negative binomial regression, therefore the over-dispersion is not a problem for this regression. In addition, as the variance and the mean are closer, both regression methods give similar results.

**Keywords:** Count data, overdispersion, Marquardt algorithm, Poisson-gamma mixed model

# O12

# Regression Approaches for Modelling Zero-Inflated Count Data

**Merve Başol Göksülük**[1], Dinçer Göksülük[1], A. Ergun Karaağaoğlu[1]

[1]*Hacettepe University Faculty of Medicine Department of Biostatistics, Ankara*

merve.basol@hacettepe.edu.tr

**Objective**: Besides continuous variables such as hemoglobin and serum bilirubin levels, count data such as the number of symptoms, number of risk behaviors at a given time point (smoking, alcohol use etc.) or number of seizures of epilepsy can be used as a response variable for regression analysis in clinical studies. When count data is of interest, it is likely to observe excess zeros in the data which is called as zero-inflation. In this case, it is not appropriate to use the regular Poisson and negative binomial regression models for modelling the count response. For this reason, zero-inflated and hurdle models which are based on Poisson and negative binomial distributions are suggested when response variable has excess zeros. These models take the effect of zero-inflation into account by modelling count data in two stages which are zero and non-zero parts. The aim of this study is to compare the performance of regression models proposed for zero-inflated data through a comprehensive simulation study and real datasets.

**Method:** In the simulation study, we investigated the effects of different sample sizes, zero ratios and dispersion parameters on the model performances which are the Poisson regression, negative binomial regression, zero-inflated negative binomial (ZINB), Poisson hurdle and negative binomial hurdle regression models. For each scenario, 1000 different datasets were generated. Model performances were evaluated based on the Akaike's information criterion (AIC), coverage ratio of confidence intervals and interval widths. Furthermore, the model performances were evaluated on two real datasets, which are West Nile virus and Horseshoe Crab datasets. For a zero-inflated dataset, Vuong test is used to examine the significance of the difference between regular model and its zero-inflated extension, e.g Poisson regression vs. zero-inflated Poisson regression. When there is a statistically significant difference between two models, we concluded that the amount of zero-inflation is significant; hence, results of zero-inflated model should be considered.

**Results:** According to simulation results, Poisson and negative binomial models give similar results when the amount of dispersion is up to moderate. However, negative binomial models' performances increase as the amount of dispersion increases. Zero-inflated and hurdle models performed better as expected when zero-inflation ratio is statistically significant. However, it is found that zero-inflated models had better model fit than hurdle models. When models are compared through confidence interval widths and coverage rates, zero-inflated models gave narrower confidence intervals even though the coverage ratios for hurdle and zero-inflated models were similar. This result indicates that zero-inflated models have lower standard error estimates than that of hurdle models.

**Conclusion:** Zero-inflation ratio and dispersion were found to have a significant effect on regression analysis results. Therefore, the use of appropriate probability distribution is important for the reliability of the results**.** When the researchers decide which method to use, they should consider how to model excess zeros. If the aim is to evaluate all zeros together, hurdle models should be used. Otherwise, if the aim is to evaluate the amount of structural zeros, then, zero-inflated models should be preferred.

**Keywords:** Excessive zero, hurdle model, zero-inflated model, poisson regression, negative binomial regression

# O13

# Many-Facet Rasch Model and An Application

**İrem Kar**[1], Batuhan Bakırarar[1], Atilla Halil Elhan[1], Serdal Kenan Köse[1]

[1]*Ankara University Faculty of Medicine Department of Biostatistics, Ankara*

ikar@ankara.edu.tr

**Objective:** In this study, it is aimed to explain many facet Rasch model with examples. In addition, it was examined how the results change when the attitudes of the graders (lenient / severe) are similar or different from each other.

**Method:** In order to analyze the facet21 data set (included in RUMM2030), RUMM 2030 program was used. The variables included in the data set are Graders (D1, D2), Gender (women, men) and 10 items that are answered for health assessment. The responses of 47 people in the data set for 10 items were scored by 2 graders between 1-5. The weighted maximum likelihood method was used as the parameter estimation method. In order to compare the attitudes of the graders in the study, separate scenarios were designed in addition to the facet21 data set and data were generated according to these scenarios.

**Results:** In the original data set and in the data set created by the two graders' attitudes being different from each other, the difference between the graders in many facet Rasch model did not changed from item to item, whereas in partial credit model the difference varied from item to item. The correlation between partial credit model and may facet Rasch model for the two graders was found to be quite high and statistically significant.

**Conclusion:** As a result, similar results are obtained by applying many facet Rasch model and by the average of the scores given by the graders to the same person by using partial credit model. If you want to have information about the attitudes of the graders and to get additional information about the graders, many facet Rasch model can be preferred.

**Keywords:** Many-facet Rasch model, partial credit model, Rasch models

# O14

# The Effect of Different Strategies for Combining Disordered Thresholds on Rasch Model Fit

**Ömer Faruk Dadaş**[1], Timur Köse[1], Derya Gökmen[2], Atilla Halil Elhan[2]

[1]*Ege University Faculty of Medicine Department of Biostatistics and Medical Informatics, Izmir*

[2]*Ankara University Faculty of Medicine Department of Biostatistics, Ankara*

omer.faruk.dadas@ege.edu.tr

**Objective:** In the examination of the internal construct validity of the measurement tools used in the healthcare field, Rasch analysis is one of the most common used approaches in the context of item response theory. In this study, on an example where the measurement tool consists of items with a multi response category, it is aimed to examine the effect of combining with different strategies of the categories of items with disordered thresholds on Rasch model fit.

**Method:** For this purpose, a measurement instrument was used to evaluate the disability levels in the areas of "self-care, getting around, hold activities" of Rheumatoid Arthritis patients who were previously developed within the scope of TUBITAK project. (Data obtained from these patients were used as application data). The internal structure validity of this measurement tool was evaluated by Rasch analysis and it was decided to use the Partial Credit Model as the appropriate Rasch model. The strategies used in the study were have been obtained by combining the categories with disordered thresholds, "to the left, to the right, to middle-left, and to middle-right". In this study, RUMM 2030 student version was used to evaluate the internal structure validity of the measurement tool.

**Results:** In the measurement tools obtained after the category combining, information loss occurred due to the combining of the categories and it has been seen that standard errors of the ability of individuals have risen. However, it has been seen that the model fit and internal consistency of measurement tools has increased after the combining strategies. Furthermore, in the measurement tools obtained after four different strategies, both the items from each scale and the representation of subsection of each scale are important for evidence of the validity of the scope of these measurement tools.

**Conclusion:** The category combining strategies discussed in this study were carried out on only one sample data, and since the single strategy was applied for all items in the combining process, the size of the disordered thresholds was ignored. Therefore, it is necessary to carry out a simulation study in order to generalize the results obtained. In addition, in this study, there has not been performed a combining that has at least 10 individuals in each item category while the categories are combined. In other words, the number of individuals who responded to the categories was not evaluated. It is considered that it would be more appropriate to combine the categories with less than 10 frequencies before the implementation of these strategies and then combine them with the strategies used in the study.

**Keywords:** Disordered thresholds, model fit, Rasch analysis, Rasch models

# O15

# Discrimination Tests for Normal and Cauchy Distributions Under Progressive Censoring

**M. Kazım Körez**[1], Ahmet Pekgör[2], İsmail Kinaci[1], Coşkun Kuş[1]

[1]*Selçuk University Science Faculty Department of Statistics, Konya*
[2]*Necmettin Erbakan University Science Faculty Department of Statistics, Konya*

mkkorez@gmail.com

**Objective:** In this paper, we consider normal and Cauchy distribution and introduce two discrimination tests for selection the true distribution based on progressive censored data.

**Method:** First discrimination method is ratio of the maximized likelihood which is used in choosing between two distributions. The Second one is Kullback-Leibler divergence which is a measure of certainty between two functions. Simulation study is conducted to observe the power of tests.

**Result:** The results indicate that our procedures work quite well and they can be used to discriminate between the normal and Cauchy distribution under progressive censoring.

**Conclusion:** These two proposed tests will allow to researchers to discrimination between Normal and Cauchy distributions for a censored data set.

**Keywords:** Normal distribution, Cauchy distribution, Kullback-Leibler divergence, ratio of the maximized likelihood, progressive censoring.

# O16

# Cox Proportional Hazard Mixture Cure Model and An Application

**Büşra Emir**[1,2], Ertuğrul Çolak[2], Ferhan Elmalı[1], Setenay Öner[2]

[1]*Izmir Katip Celebi University Faculty of Medicine Department of Biostatistics, Izmir*
[2]*Eskisehir Osmangazi University Faculty of Medicine Department of Biostatistics, Eskisehir*

busra.emir@ikc.edu.tr

**Objective:** Survival analysis is a collection of methods developed to assess the time from a specified time point to the occurrence of the event of interest. One of the main variables of survival analysis is survival time and Cox proportional hazards model is the most preferred model established to determine the factors affecting the survival time. Cox (1972) has developed this model taking into account the likelihood that individuals who have survived the beginning of a certain period of time would die in this interval. At any point in time, he assumed that the hazard function in a group would be proportional to the hazard function in another group for the same time point. It is based on the assumption that each individual is predisposed to failure (death, metastasis, or recurrence) according to what is being investigated. However, this assumption may not be true if a large part of the patients survives after an adequate observation period. Particularly in recent years, investigations in some types of cancer, including breast cancer, liver cancer, childhood lymphomas and leukemia have shown that a significant proportion of patients with these diseases have improved after the treatment. The "cure models" developed by Boag (1949) have become an important statistical model used to analyze the survival data of cancer patients. In this study, it was aimed to introduce the Cox proportional hazard mixture cure model using a sample data set.

**Method:** Mixture cure models are examined in two parts, the cured part and the uncured part. This structure allows explanatory variables to differ in the effects on cured and uncured individuals. The practice data set used in the study was obtained from a clinical trial conducted by Mayo Clinic between 1974 and 1984. 312 patients from 424 patients with liver primary biliary cirrhosis (PBC) were randomly assigned to 2 treatment groups. The first group received D-penicillamine drug treatment and the other group received placebo. The first group received D-penicillamine drug treatment and the other group received placebo. Firstly, the proportional hazard assumption for the application data set was examined by the Schoenfeld residual approach. Survival time, censorship status, treatment groups, age of patients, and histological stages of this disease were analyzed by Cox proportional hazard model and Cox proportional hazard mixture cure model. The Logit link function was used to model the effect of the explanatory variables on the cured part. Parameter estimates of Cox proportional hazard mixture cure model were performed by EM algorithm.

**Results:** In the Cox proportional hazard mixture cure model for the liver primary biliary cirrhosis (PBC) disease dataset, no variables were found statistically significant for the cured part. It was found that the histological stages of disease and age were statistically significant variables in the uncured part ($p=0.007$; $p=0.002$).

**Conclusion:** It is recommended to use the Cox proportional hazard mixture cure model if there is a large number of cured individuals who do not occurred event of interest after a long observation period when the study is completed and if the data structure provides a proportional hazard assumption.

**Keywords:** Cox proportional hazard model, mixture cure model, Cox proportional hazard mixture cure model

# O17

# A Two-step Approach for Modeling Longitudinal and Survival Data

**Ebru Turgal**[1], Beyza Doğanay Erdoğan[2]

[1]*Hitit University Faculty of Medicine Department of Biostatistics, Çorum*
[2]*Ankara University Faculty of Medicine Department of Biostatistics, Ankara*

ebruturgal@hitit.edu.tr

**Objective**: The purpose of studies involving longitudinal data is generally to examine how the mean response profiles differ between groups and also the course of responses over time. The answers to these questions can be found with the help of various statistical models. Mixed-effect models include both fixed and random effects. In the literature, two-stage approaches for the evaluation of longitudinal data and survival processes jointly have been proposed. In this approach, the parameters in the longitudinal process and the survival process are estimated separately. In the first step, the longitudinal response value estimated from this model for each observation time using the mixed-effect model is taken as the explanatory variable in the expanded Cox model in the second step. With this study, we propose a new two-stage approach for the evaluation of longitudinal data and survival process together.

**Method:** As the first stage of the two-step approach, error-free longitudinal data estimates from the regression tree analysis for the longitudinal data will be obtained and used as an explanatory variable in the extended Cox regression model in the second stage. In the first stage, the reason of taking the regression tree approach for the analysis of longitudinal data is that when there are many covariates (fixed effects), the number of fixed effects to be added to the model and their interactions with each other increase and the modeling of this complex structure becomes more difficult. Since this method is based on the EM algorithm, the name of the regression tree is also called the random effects expectation maximization.

**Results:** When the -2 log-likelihood values of the models created for $PaO_2$ / $FiO_2$ variable are taken into consideration, the lowest value is obtained by using RE-EM AR (1) tree. The parameter estimations obtained by using RE-EM tree instead of linear mixed effect model were taken as explanatory variable in the extended Cox regression analysis and it was found that the standard errors obtained were lower than the classical two-stage approach.

**Conclusion:** In this study, a two-stage approach previously proposed by Albert, P. S. and Shih (2010) in the modeling of longitudinal data and survival data has been reviewed. In the first stage, a regression tree was used for longitudinal data and a new approach was presented. This result may be supported by further simulation studies. In addition, regression trees can be preferred for modeling the longitudinal process because it offers a more flexible solution to the researcher.

**Keywords:** Longitudinal data analysis, mixed effects, random effects, RE-EM trees, survival analysis

# O18

## Study of the Effect of Karnofsky Performance Scale on the Survival of Various Cancer Cohorts

Nihat Burak Zihni[1], Kemal Turhan[2], **Zeliha Aydın Kasap**[2]

[1]*Karadeniz Technical University Vocational School of Health Science, Trabzon*
[2]*Karadeniz Technical University Institute of Health Sciences, Department of Biostatistics and Medical Informatics, Trabzon*

zelihaaydinkasap@ktu.edu.tr

**Objective:** In survival studies, the survival of patients according to various factors is investigated. However, in recent studies, it is emphasized that quality of life is as important as the life span of the patient. The Karnofsky Performance Scale Index (KPI) is one of the tools used to measure the quality of life of cancer patients. The Cancer Genome Atlas (TCGA) is a publicly available database of genetic and clinical information of 33 types of cancer. TCGAbiolinks is a R package developed for the R language that makes it easy to import data from the TCGA database. In this study, Karnofsky Performance Scale Score (KPI) data is analyzed within the cancer data of the TCGA database. The aim of the study is to examine the effect of KPI score on the survival of cancer patients independent of the effect of other variables.

**Method:** Data of different cancer cohorts obtained with TCGAbiolinks package using R language were analyzed with KTÜ licensed SPSS 22.0 software. The differences between the descriptive statistics of cancer cohorts were analyzed with independent samples t-test and with Mann-Whitney U test in non-normally distributed data. Kaplan-Meier method was used to compare survival rates among cancer cohorts. Cox regression analysis backward LR method was used for multivariate analyzes.

**Results:** GBM cancer (14.81 ± 16.58 months) was found to be more fatal than LGG cancer (24.71±30.58 months; $p<0.001$). Patients with GBM cancer had lower KPI score (77.45±14.55) than those with LGG cancer (87.71±12.10; $p<0.001$). In addition according the data which has KPI score, it was found that patients with GBM cancer had 3.041 times more relative risk than those with LGG cancer ($p<0.001$).

**Conclusion:** There was a moderately positive relationship between KPI score and survival ($r=0.309$; $p<0.001$). Based on this finding, it can be said that those with high KPI score lived longer. It is important to consider that the quality of life is as important as the life span of the patient. In this study, it was concluded that LGG cancer patients had better quality of life than those who had GBM cancer.

**Keywords:** Karnofsky performance scale index, TCGAbiolinks package, Kaplan-Meier, Cox regression

# O19

# Robust Chart for Monitoring the Contaminated Skew Normal Process Under Ranked Set Sampling

**Derya Karagöz**[1], Nursel Koyuncu[1]

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

deryacal@hacettepe.edu.tr

**Objective:** When the quality variable has a skewed distribution, it might be misleading to observe the process by using the Shewhart control charts. The usage of Shewhart control charts in skewed distributions causes an increase of Type I risk when the skewness increases because of the variability in population.

**Methods:** For this reason, the weighted variance and skewness correction methods are proposed as an alternative to the classical Shewhart method. In this study, we propose to construct control limits of robust control chart for monitoring the skewed and contaminated process by using ranked set sampling design. We consider the modified Shewhart (MS) and the modified weighted variance (MWV) methods proposed by Karagöz (2018). To get improvement in control charts, we take into account the simple random sampling and ranked set sampling schemes. The performance of the robust charts based on simple and ranked set sampling designs are compared in terms of the Type I risk probabilities by Monte Carlo simulation.

**Results:** When the skewness of distribution increases, the p values of the RSS mean charts gives desirable results for large sample size.

**Conclusion:** For large sample sizes, MS and MWV mean charts under the RSS design perform better than simple sampling designs. Finally, the modified methods using different sampling designs for the mean chart can be used to monitor the contaminated skew-normal process.

**Keywords:** Ranked set sampling, control charts, process control

# O20

# Optimization of Epicardial Fat Thickness Change in Obese Patients with Weight Loss by Bariatric Surgery Using Central Composite and Box-Behnken Design

**Mustafa Agâh Tekindal**[1], Cihan Altin[2], Can Ateş[3], Varlık Erol[4]

[1]*Selçuk University Faculty of Veterinary Medicine Department of Biostatistics, Konya*
[2]*Başkent University Faculty of Medicine Zübeyde Hanım Research and Application Center Cardiology Department, İzmir*
[3]*Van Yüzüncü Yıl University Faculty of Medicine Department of Biostatistics, Van*
[4]*Başkent University Faculty of Medicine Zübeyde Hanım Research and Application Center, Department of Genetic Surgery, İzmir*

matekindal@gmail.com

**Objective:** The aim of our study was to determine the optimization of the change in epicardial fat thickness in obese patients who underwent bariatric surgery with Central Composite (CCD) and Box-Behnken Experimental design (BBD).

**Method:** In order to meet the needs of the executive in the experimental design methods; knowing the limits of the design being developed, understanding the effects of the variables within these limits and finding the most analytical solution is extremely important for the designer. However, if an analytical relationship cannot be expressed between the variables that define the design and the evaluation criteria to measure the quality of the design, it may be necessary to use other methods to reach the optimum solution. In such cases, Response Surface Methods are used to see the sensitivity of the assessment criterion to changes in design variables, and even to obtain the necessary correlations experimentally. Response surface methods are evaluated in two different ways as CCD and BBD Layouts. In response surface methods, the model is generated by Regression Analysis. Regression coefficients are determined to what extent the main effect or interaction effect of a factor has an important effect on the value of the response variable. The first step in response surfaces methods; the factors that are thought to have an effect on the response variable, and the levels they have. These two criteria usually determine the trial patterns to be established to create the regression model. In this study, $3^3$ experimental designs were designed. This study was approved by Baskent University Institutional Review Board and Ethics Committee and supported by Baskent University Research Funding (Approval number: KA16/281). The study data consisted of 40 obese patients who lost weight by bariatric surgery between February 2015 and December 2016. Body Mass Index (BMI), Age and HOMA values were evaluated in 3 categories and 3 levels, and response variable was the change in Epicardial Fat Thickness (ΔEFT).

**Results:** As a result of CCD analysis, Age=30.52, BMI=45.30, HOMA=34.62, the optimum ΔEFT=2.571. As a result of BBD analysis, Age=38.36, BMI=63.18, HOMA=14.95, the optimum ΔEFT=3.756. Optimum ΔEFT is modeled with Contour and Response surface graphics.

**Conclusion:** According to the results of the analysis, it was found that BBD analysis for optimum ΔEFT was more positive than CCD and optimum age, BMI and HOMA combinations were determined to reach maximum ΔEFT.

**Keywords:** Response surface, central composite trial layout, Box-Behnken trial order

# O21

# An Alternative New Approach to ITT and PP Analysis

**Betül Dağoğlu Hark**[1], Zeliha Nazan Alparslan[2]

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

betuldagoglu@gmail.com

**Objective:** Protocol violations and missing data are among the most common problems in evaluating treatment efficacy in randomized controlled clinical trials. The most preferred solution for these problems is the utilization of ITT (Intention-to-treat) and/or PP (Per Protocol) analyses. ITT analysis includes each subject assigned according to a randomized treatment assignment; in other words, the protocol violations and missing data after randomization are ignored. PP analysis on the other hand, intends to measure the 'pure' treatment effect; that is, the effect in the set of patients who comply with therapy and have no other complicating factors. Treatment efficiencies calculated by ITT and PP analyzes are different.

In this study, it is aimed to propose the Begg and Greenes correction used to correct validation bias in diagnostic tests, as an alternative to ITT and PP analysis. Thus, it may be possible to obtain a more unbiased estimator from the ITT and PP analyses.

**Method:** Different treatment success rates, sample size, missing data mechanisms and missing rates and data sets are used to evaluate the success of the new method. The results are compared with the success rate, ITT and PP results.

**Results and Conclusion:** The prognostic balance based on randomization is maintained with ITT analysis. Therefore, the estimation of treatment effect in the ITT analysis is generally conservative. In PP analysis, the treatment effect can be estimated higher. The new method based on the proposed Begg and Greenes correction seems to give results closer to the actual treatment effect than the ITT and PP analyses.

**Keywords:** Clinical trial, ITT, PP, Begg and Greenes correction

# O22

# The Effect of Quantitative Data Discretization Before Classification Analysis on Classification Performance

**Fatma Betül Örs**[1], Necdet Süt

[1]*Trakya University Medical Faculty Department of Biostatistics and Medical Informatics, Edirne*

f.betul.oers@gmail.com

**Objective:** The aim of this study was to investigate the effects of continuous variables discretization before classification analysis on classification performance.

**Method:** Six real datasets that can be obtained open access via internet were used in the study. Each of these dataset was analyzed by using the C5.0 and naïve Bayes (NB-Kernel Probability-Density-Function) algorithms. Then, the continuous variables in the datasets were discretized by using MDLP (Minimum Description Length Principle), CAIM (Class-Attribute Interdependence Maximization), CACC (Class-Attribute Contingency Coefficient), Ameva, ChiMerge (α=0.05), Chi2, ExtendedChi2 methods. New datasets were classified again by C5.0 and NB classifiers, and then before-after discretization of them were compared. Sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), area under the curve (AUC) and accuracy rates were calculated to evaluate the classification performances of the algorithms. The R statistical program were used for data analysis.

**Results:** The sensitivity, specificity, PPV, NPV, AUC and accuracy values of the C5.0 in the raw-datasets that had less than 300 subjects were found to be 0.78, 0.825, 0.735, 0.855, 0.831, 0.807, respectively. After discretization, these ratios increased to 0.798-(ExtendChi2), 0.893-(CACC), 0.815-(CACC), 0.864-(ExtendedChi2), 0.859-(Ameva) and 0.831-(Ameva). The values for NB in the raw-datasets were found to be 0.786, 0.853, 0.757, 0.879, 0.889, 0.834, respectively. After discretization, these ratios increased to 0.885-(Chi2), 0.896-(CACC), 0.825-(Chi2, ChiM), 0.926-(Chi2), 0.964-(Chi2, ChiM) and 0.893-(Chi2).

The sensitivity, specificity, PPV, NPV, AUC and accuracy rates of the C5.0 in the raw-datasets consisting of more than 1000 observations were found as 0.735, 0.858, 0.817, 0.802, 0.822, 0.819, respectively. After discretization, these ratios increased to 0.798-(Chi2, ChiM), 0.871-(CACC), 0.821-(MDLP, Chi2, ChiM), 0.853-(AMEVA), 0.845-(Chi2, ChiM) and 0.85-(Chi2, ChiM). The values for NB in the raw-data were found as 0.613, 0.907, 0.727, 0.762, 0.796, 0.776, respectively. After discretization, these ratios increased to 0.836-(Chi2, ChiM), 0.913-(Chi2, ChiM), 0.827-(Chi2, ChiM), 0.872-(AMEVA), 0.937-(Chi2, ChiM) and 0.88-(Chi2, ChiM).

**Conclusion:** Independent from observation numbers, discretization of continuous variables increased to the classification performance. In datasets with more than 1000 observations, mean increase rate after discretization (11.4%) was higher than in datasets with less than 300 observations (6.5%), however it was close to each other in C5.0 (3% vs 3.8%). In datasets with 1000 observational, Chi2 and ChiM were achieved higher results in 5 of the 6 criterions, however, in datasets less than 300 observations, NB with discretized Chi2 achieved higher values. On the basis of the highest discretization rates, the NB achieved higher results than the C5.0 algorithm in all criteria.

**Keywords:** Classification, discretization, C5.0, naive bayes

# O23

# The Determination of Optimal Test Combination by Evaluating the Units of the Different Diagnostic Tests Used in the Estimation of Airway Hyperresponsiveness Using Boole Algebra and Latent Class Analysis

**Damla H. Saral**[1], Tuğba Arıkoğlu[2], Bahar Taşdelen[1]

[1]*Mersin University Faculty of Medicine Department of Biostatistics and Medical Information, Mersin*
[2]*Mersin University Faculty of Medicine Department of Pediatric Allergy and Immunology, Mersin*

dmlhzl5@gmail.com

**Objective:** In this study, determination of optimal test combination by using clinically significant cut-off points to assess co-use of Fractional nitric oxide (FeNO), clinical important when diagnosing asthma, IOS and Plethysmography parameters, known to be more sensitive paremeters than spirometry in (airway hyperresponsiveness) AHR estimation, is intended. In determining the optimal test combinations, the diagnostic performance measures for combinations of parameters with low, middle and high distinguishing will be compared using Boole Algebra and LCA.

**Method:** Because most of the diagnostic tests can not distinguish between patients and healthy individuals with high accuracy, simultaneous or sequential administration of multiple tests can increase the diagnostic accuracy. When a diagnostic test is used alone; sensitivity, specifity and other performance measures may be poor but when multiple tests are used together, a test combination with a higher performance than a single test may be obtained. One of the combined test acquiring approaches is the Boole Algebra (and / or rule); another approach is Latent Class Analysis (LCA). Combined test result may be obtained according to Boole Algebra with and / or rule. LCA is a statistical method developed to reduce the bias that is used in the evaluation of alternative tests and is formed in the situation of uncertain reference tests where there is no gold standard test. LCA is used in cases of real disease where all monitored variables are affected by a latent variable. The structure of relations between the other variables is examined to define latent variable.

**Findings:** AHR is defined as one of the prominent clinical features of asthma, and Methacholine challenge test with a spirometer is performed to confirm this. As a result, it is concluded that those with a methacholine concentration of <4 have severe aırway hyperresponsıveness. In this study, clinically significiant cut-off points were first determined using ROC analysis to evaluate the co-use of IOS and Plethysmography parameters known to be more sensitive than spirometer with Fractional nitric oxide (FeNO) value which have clinically importance in AHR estimation. In distinguishing mild AHR from moderate and severe AHR, R5% parameter which has a statistically significant but weak accuracy was accepted. The cut-off value for R5% of the IOS parameters was determined as >104.14 (AUC=0.642, p=0.051) in the separation of those with severe and moderate AHR sensitivities. The cut-off values for Srtotal%, which is the FeNO and Plethysmography parameter, were obtained from a previous study (FeNO>28, AUC=0.723, p<0.001; SRtot>294.9, AUC=0.674, p=0.013). Using these cut-off values, the performances of FeNO, R5% and SRTot combinations were compared with Boole Algebra and GSA.

**Results:** When using Boole Algebra, the discrimination power of individuals with moderate AHR level is increased by combining according to AND rule, and the sensitivity of combination of FeNO, R5% and SRTot parameters are low because the number of positive individuals decreases gradually.

**Conclusion:** Combining according to OR rule, the power of separating the severe AHR individuals increases, but this result is misleading. For this reason, LCA power which considers the inter-parameter relationships in determining the optimal combination is a higher method.

**Keywords:** Diagnostic test, boole algebra, latent class analysis, sensitivity, specifity, gold standart test

# O24

## Reducing the Bias in the Use of Imperfect Gold Standard Test

**Meral Aydın**[1], Pınar Özdemir[1]

[1]*Hacettepe University Faculty of Medicine Department of Biostatistics, Ankara*

meralcubuk@gmail.com

**Objective:** Diagnostic tests used to confirm the presence or absence of a disease, to provide information about the progress of the disease, and in some cases to determine the response to treatment are important in the health field. The aim of this study is to provide information about the methods developed to evaluate the performance of diagnostic tests in case of the use of imperfect reference tests, which is one of the sources of bias that affect the diagnostic accuracy.

**Method:** In order to evaluate the performance of diagnostic tests in cases where there is no definitive reference test, models were created with discrepant resolution, composite reference standard and latent class analysis methods for different test combinations by using data from 90 patients who were diagnosed as Cystic Fibrozist and the results were compared.

**Results:** The results obtained on the basis of the application data revealed that the test performance measures changed due to the absence of the perfect reference test and the methods applied. In this context the specificity value increased from 0.80 to 0.92 in the discrepant solution with exotoxin, elastase and culture tests; as a result of composite reference standard analysis, it increased from 0.80 to 0.88. However, the sensitivity value increased from 0.72 to 0.92 in the discrepant solution and from 0.72 to 0.64 in the composite reference standard analysis. Similar results were obtained with exotoxin, alkaline protease and culture tests. These differences between the test performance measures will also change according to the decision rule determined in the composite reference standard method, index and reference tests. However, the results obtained with the composite reference standard method in this study were found to be closer to the actual values. Based on the results of the latent class analysis with the data obtained, only the use of model 2 consisting of combining the results of alkaline protease, elastase and exotoxin evaluated by blood test was found to be statistically significant.

**Conclusion:** If Pseudomonas aeruginosa (Pa) infection is not diagnosed and treated in the early period, chronic colonization occurs in the lungs in cystic fibrosis patients. In some centers in our country, the difficulty of producing Pa in cultures and the diagnosis of Pa infection in young children who cannot remove the sputum are delayed in diagnosis. In this context, it was statistically significant to use diagnostic tests consisting of combining the results of alkaline protease, elastase and exotoxin results which are easily accessible, cost effective, rapid results and evaluated by blood test compared to culture test. It is also necessary to be aware of the fact that there may be multiple sources of bias and variability in order to obtain more accurate and reliable results. In this respect, the results of the studies conducted in cooperation with clinicians and statisticians are considered to be more efficient both technically and clinically.

**Keywords:** Evaluation of diagnostic tests, imperfect gold standard test, bias

# O25

# A Forecast Model Proposal for HPV Development Risk in Women

Seda Keskin[1], **Neslihan Baki**[2], Kemal Turhan[2], Burçin Kurt[2], Melek Bulut Kuzu[3], Deha Denizhan Keskin[1], Zeliha Aydin Kasap[2]

[1]*Ordu University Faculty of Medicine Department of Obstetrics and Gynecology, Ordu*
[2]*Karadeniz Technical University Faculty of Medicine Department of Biostatistics and Medical Informatics, Trabzon*
[3]*Ministry of Health Batıkent 2nd Family Health Centre, Ankara*

ebirdilektut@gmail.com

**Objective:** In this study, it is aimed to develop a prediction model by examining the existence of pre-determinants of HPV, which is a very important risk factor for cervical cancer. Using the prediction model developed, patients with a high probability of HPV will be identified and referred for screening tests. In this way, HPV infection will be prevented from being transformed into cervical cancer by early detection.

**Method:** In this study, a total of 860 patient data, 356 HPV positive and 504 HPV negative, which were screened for retrospective cervical cancer, were HPV infection risk factors; age, smoking, age at first sexual intercourse, number of pregnancies, number of sexual partners were evaluated. Four models have been developed by using four different core (linear, radial, polynomial, sigmoid) functions with the help of R Programming and Support Vector Machine (SVM) classification algorithm.

**Results:** With SVM linear core function, sensitivity 94%, specificity 92% and AUC 90%, radial sensitivity 94%, specificity 93% and AUC 91.5%, polynomial sensitivity 93%, specificity 94% and AUC 91.7%, sensitivity with sigmoid 91% specificity 92% and AUC 89% were obtained.

**Conclusion:** The model developed with the polynomial core function was the most successful prediction model with a sensitivity of 93%, specificity 94% and AUC 91.7%. Thus, a successful estimation model was developed to identify patients in the HPV risk group at an early stage.

**Keywords*:*** Support vector machine, Human Papillom Virus(HPV), risk factors

# O26

# Type II Error Rates for Measuring the Classification Accuracy of a Biomarker in Two Groups

**İlker Ünal**[1], Yaşar Sertdemir[1], Ceren Efe[1], Esin Ünal[2]

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

ilkerun@cu.edu.tr

**Objective**: Studies on the comparison of two groups' means are frequently conducted. In these studies, when the difference between the means of groups is significant, the researchers also might want to investigate whether the classification accuracy of this measurement (i.e. the area value under the ROC curve, AUC-ROC) is above a certain level. However, for this second hypothesis, which was not foreseen at the planning stage of the study, the sample size calculation for this hypothesis was generally not considered. Therefore, the type II error rate for testing this hypothesis was not controlled at the design stage of the study and it may exceed the acceptable levels.

**Method**: In this study, based on the sample size calculated for a study which is planned to show the difference between the groups' means, for different scenarios, we will show how the power of the second analysis for the classification accuracy changes and what the required sample size should be.

**Results and Conclusion**: As a result of the this study, we will show that conducting the study with the initially calculated sample size could increase the type 2 error rate by two to three times.

**Keywords**: Sample size, AUC-ROC, type II rate

# O27

## Effect of Kernel Density Estimation on Naive Bayes Classification Results

Necdet Süt[1], **Fatma Betül Örs**[1]

[1]*Trakya University Medical Faculty Department of Biostatistics and Medical Informatics, Edirne*

f.betul.oers@gmail.com

**Objective:** The aim of this study was to investigate the effect of kernel density estimation (KDE) on naïve bayes classification results by a simulation study.

**Method:** Ten different variables in two different situations (7 categorical-3 continuous and 3 categorical-7 continuous) with 30, 100, 500 and 1000 observations were derived from 1000 replicates. Outcome variable, a binary variable, was derived from 2 different class distributions (50%-50%) and (30%-70%). Derived data were firstly analyzed with normal distribution probability density function (NDPDF) naive Bayes, then analyzed with KDE. The classification results were compared based on sensitivity, specificity, positive predictive value, negative predictive value, accuracy ratio and area under the curve. The R program was used in the derivation and analysis of the data.

**Results:** When the continuous variable rate was 30% in 4 different combinations, accuracy rates of the simulation with 30 subjects were 88.9%-66%-91%-72.8% for NDPDF, 86.9%-60.5%-88.8%-69.3% for KDE; with 100 subjects those were 91.6%-64.8%-92.6%-76.4% for NDPDF, 90.7%-65.7%-91.8%-75.2% for KDE; with 500 subjects those were 91.6%-62.2%-93.2%-76.4% for NDPDF, 91.5%-71.0%-93.0%-77.9% for KDE; with 1000 subjects those were 92.0%-61.2%-93.2%-76.3% for NDPDF, 91.9%-70.6%-93.1%-78.3% for KDE.

When the continuous variable rate increased to 70%, accuracy rates of the simulation with 30 subjects were 98.9%, 70.5%-99.1%-65.9% for NDPDF, 98.3%-69.5%-98.4%-71% for KDE; with 100 subjects those were 99.6%-69.0%-99.6%-56.5% for NDPDF, 99.4%-78.9%-99.5%-80.1%; with 500 subjects those were 99.7%-65.5%-99.7%-51.0% for NDPDF, 99.7%-87.3%-99.7%-87.8% for KDE; with 1000 subjects those were 99.8%-63.7%-99.8%-48.7% for NDPDF, 99.7%-89.0%-99.7%-89.0% for KDE.

**Conclusion:** When the number of observations is below 100 and the categorical variable ratio in the data set is high, the Kernel method has lower accuracy rate than the NDPDF method in the naive Bayes classification. There is no significant difference between two probability density functions when the categorical variable ratio is high in data with 100 observations, but higher accuracy rates are achieved with KDE when the ratio of non-normal distributed continuous variables is high. Accuracy rates of the two probability density functions are very similar when the number of observations is 500 or higher and the continuous variables have normal distribution. When the rate of continuous variables that have normal distribution in the data set increases, accuracy rate of KDE and NDPDF increases, too. Furthermore, the presence of non-normally distributed variables when the continuous variables ratio increases and the class distributions are different leads to a significant reduction in the accuracy rates of NDPDF. The KDE has higher accuracy rates than NDPDF in non-normally distributed variables.

**Keywords:** Classification, Kernel, probability density function, naive Bayes

# O28

# Multivariate Regression Tree Method and Its Application

**Duygu Korkmaz**[2], Sıddık Keskin[1]

[1]*Yüzüncü Yıl University Faculty of Medicine Department of Biostatistics, Van*
[2]*Çukurova University Faculty of Medicine Department of Biostatistics, Adana*

korkmazduygu157@hotmail.com

**Objective:** Multivariate statistical analysis methods are commonly used to analyze complex data sets. These methods handle multiple independent and / or dependent variables in a single analysis, rather than performing several univariate or bivariate analyzes, where all these variables are related to each other at different levels. One of these techniques is the multivariate regression tree method. As a result of the literature review, it has been noted that there is almost no Turkish literature on this method. The aim of this study designed to eliminite this deficiency is to introduce the method, to give the basic concepts and to examine the theoretical structure together with the application.

**Method:** The analysis was performed by a multivariate regression tree method, which is an extension of the univariate regression tree, which is similar to constrained cluster analysis, where the response variable is continuous and multiple, also explanatory variables can be used either continuously or categorically. This method is based on minimizing the heterogeneity within each node and maximizing the variation between the nodes. In the application phase of the study; The aim of this study is to show the applicability of the method and to make the statistical interpretations of the method rather than the biological interpretation. Therefore, hypothetical data were used instead of actual data. In the creation of hypothetical data, a scenario similar to the study by Suvak et al. (2017) has been considered. For this, it was assumed that 95 patients from 21 districts of Van, Bitlis and Ağrı provinces were included. There are 3 dependent and 8 independent variables related to these patients. In this study, analyzes were carried out in R programming language.

**Results:** Albumin [ALB (g / dL)] variable was found to be root node in tree building according to diseases. Only 3 of the explanatory variables were in the tree. According to the "1SE" rule, which is another method used to determine the tree size, ALB variable was root and single node.

| **Tree Size** | **$R^2$** | **Cross Validation Error** | **Standart Error** |
|---|---|---|---|
| CVRE rule | 0.55 | 2.6 | 0.48 |
| 1SE rule | 0.19 | 2.39 | 0.49 |

**Conclusion:** The method is a non-parametric test method, there are no assumptions about the underlying distribution of the data, it is robust to outliers, it can be interpreted easily (graphically), it has the ability to handle missing data. Therefore, it can be said that this method can be used to determine the relationships between multiple answer variables and multiple explanatory variables, especially to determine species-environment relationships and it can be used easily in the other areas (Questier, Put, Coomans, Walczak and Wander Heyden, 2005).

**Keywords:** Cross Validation, node, relative error

# O29

# Evaluation of Receiver Operating Characteristic Analysis to Determine the Minimum Clinically Important Difference on Interpretability of Scales

**Selcen Yuksel**[1], Afra Alkan[1], Pervin Demir[1]

[1]*Ankara Yildirim Beyazit University Medical School Department of Biostatistics, Ankara*

selcenpehlivan@gmail.com

**Objective:** The aim of this study was to examine the accuracy of cut-off point obtained by widely-used method in the literature -ROC Analysis- to determine the minimum clinically important difference (MCID), which is the estimator of responsiveness for scales, by a simulation study.

**Method:** The baseline person parameters were firstly generated and, by using these values, two gold standard groups were constructed as "improved" and "non-improved" after the treatment. Five point-likert response patterns were obtained for 20 items in each group, representing pre- and post-treatment responses of individuals. The mean change score between post treatment and baseline scores for the improved group was considered as real MCID ($MCID_R$), after baseline and post-treatment total scores were calculated from response patterns. The cut-off for change score specified by ROC analysis, which best discriminates between improved group and not improved group, $MCID_{ROC}$, was compared to $MCID_R$. The scenarios of simulation were consisted of sample size and ability level of persons. The data were generated for each of 36 scenarios with 10 000 repeats by using Markov Chain Monte Carlo, (MCMC) algorithm.

**Results:** It was observed that the $MCID_R$ and $MCID_{ROC}$ were not affected by sample size and the ability level. However, $MCID_{ROC}$ overestimated the $MCID_R$ values in all scenarios.

**Conclusion:** The cut-off points obtained by ROC analysis found to be greater than the real MCID values. Therefore, alternative methods are required to calculate MCID.

**Key words:** Minimum clinically important difference, ROC, simulation

# O30

## 3-Way ROC Analysis: An Illustration on Ertapenem Sensitivity

**Harika Gözde Gözükara Bağ**[1], Yücel Duman[2]

[1]*İnönü University Faculty of Medicine Department of Biostatistics and Medical Informatics, Malatya*
[2]*İnönü University Faculty of Medicine Department of Medical Microbiology, Malatya*

harika.gozukara@inonu.edu.tr

**Objective:** ROC (Receiver Operating Characteristic) analysis is a widely used method for determination of the best cut-off value of a diagnostic test and also for evaluation of the diagnostic ability. The aim of this study is to introduce the 3-way ROC analysis which can be used when the outcome variable has three categories and to make an illustration on a clinical data.

**Method:** In clinical microbiology laboratories, determining the antibiotic sensitivities of isolated microorganisms is one of the most important parameters for the management of the diseases related to these microorganisms. CLSI (Clinical Laboratory Standards Institute) and EUCAST (European Committee on Antimicrobial Susceptibility Testing) guidelines are widely used in the world to interpret sensitivities. In this study, in Enterobacteriaceae strains (n=1002), 3-way ROC analysis was performed to determine the performance of disc diffusion results of the EUCAST and CLSI interpretation guidelines for ertapenem when BMD (Broth MicroDilution) considered as gold standard. Correct classification probabilities (CCP) for 3 outcomes (sensitive, intermediate sensitive and resistant) and volume under the ROC surface (VUS) were calculated for both interpretation guides. 3-way ROC analysis and VUS calculations were performed by DiagTest3grp package in R (3.3.1) software.

**Results:** For CLSI guideline, group correct classification probabilities are found to be $CCP\_1_{CLSI}=0.9134$, $CCP\_2_{CLSI}=0.6087$ and $CCP\_3_{CLSI}=1.0000$ respectively for sensitive, intermediate sensitive and resistant groups. The VUS for CLSI guideline was obtained as $VUS_{CLSI}=0.771$ (95% CI: 0.679-0.8525, $p<0.001$). For EUCAST guideline, CCP corresponding to three classes were obtained, respectively, as $CCP\_1_{EUCAST}=0.8058$, $CCP\_2_{EUCAST}=0.0435$ and $CCP\_3_{EUCAST}=1.0000$. The VUS for EUCAST guideline is found as $VUS_{EUCAST}=0.5114$ (95% CI: 0.488-0.5452, $p< 0.001$).

**Conclusion:** Both EUCAST and CLSI guidelines classifications are found to have VUS values greater than an uninformative three-class diagnostic test value 1/6 and found to be statistically significant.

**Keywords:** 3-way ROC analysis, the volumeunder the ROC surface, ertapenem sensitivity

# O31

# The Evaluation of Agreement Between Measurement Methods Used in Diagnosis and the Comparison by Statistical Methods

**Şeyma Yaşar**[1], Zeynep Tunç[1], İpek Balıkçı Çiçek[1], Ayşe Nur Akatlı[2], Leyla Karaca[3], Ramazan Altıntaş[4], Cemil Çolak[1], Saim Yoloğlu[1]

[1]*İnönü University Faculty of Medicine Department of Biostatistics and Medical Informatics, Malatya*
[2]*İnonu University Turgut Özal Medical Center Department of Pathology, Malatya*
[3]*İnönü University Faculty of Medicine Turgut Ozal Medical Center Department of Radiology, Malatya*
[4]*İstinye University Bahçeşehir Liv Hospital Department of Urology, Istanbul*

seyma.yasar@inonu.edu.tr

**Objective:** Along with the technological developments, the methods used in the diagnosis and treatment of diseases are simultaneously developing. Each new method has to be compared with the existing reference method to determine its accuracy and precision in diagnosing and treating. In method comparison studies, a number of methods have been developed to determine the degree of agreement between measurements obtained by different methods on the same subject. The aim of this study is to evaluate the compatibility of computed tomography (CT) and magnetic resonance imaging (MR) methods, which are the imaging techniques used to describe prostate cancer, with the pathology accepted as the reference method.

**Method:** In this study, 37 prostate cancer patients, the most common type of cancer in men, were taken from Inonu University Turgut Özal Medical Center Urology Department. The agreement between CT and MR results of prostate cancer patients and the pathology results were evaluated Bland-Altman, Class correlation coefficient (CCT), Concordance Correlation Coefficient (CKK), Deming Regression and Passing-Bablok method comparison methods.

**Results:** After the analysis, CCT and CKK values for CT-Pathology results were 0.62 and 0.62, respectively. Similarly, CCT and CKK values for MR-Pathology results were 0.74 and 0.75, respectively. The regression line equation for the deming regression method is y = -6.21 + 1.03 for CT-pathology, whereas for the MR-pathology y = -6.86 + 1.11x. Finally, when the BT-Pathology measurement values are evaluated by Bland-Altman statistical method, the mean values of the measurement differences are 2.42 and the standard deviation is 33.50. The agreement limits at 95% confidence level calculated by means of the mean and standard deviation values for these difference measurement values are -63.24 and 68.09. Similarly, when the agreement between MR-Pathology measurement values is examined by Bland-Altman method, the mean values for the difference between the measurement values are 4.14 and the standard deviation is 26.78. The agreement limits of 95% confidence level calculated using the mean and standard deviation values for the difference values of these measurements are -48.34 and 56.63.

**Conclusion:** When the confidence intervals of Deming regression and Passing-Bablok regression methods were examined, no systematic or proportional error was observed between the results of CT-Pathology and MR-Pathology. Therefore, it may be preferable to use Bland-Altman method to determine the interobserver agreement. When all of these analysis results are taken into consideration, the tumor sizes obtained by the pathology method can be found as 63.24 cc large, 68.09 cc smaller than the measured values obtained from the CT result. Similarly, tumor size obtained by pathology method can be found as 48.34 cc large,

56.63 cc smaller than the measurement values obtained by MR imaging technique. When CCT and CKC values were examined, it was observed that MR was more compatible with the pathology results than CT.

**Keywords:** Bland-Altman, deming regression, passing-bablok regression, intraclass correlation coefficient, concordance correlation coefficient,prostate cancer

# O32

# Utilization of Odds Ratio in Prospective Studies

**Gülden Hakverdi**[1], Ömer Faruk Dadaş[1], Timur Köse[1], Mehmet N. Orman[1]

[1]*Ege University Faculty of Medicine Department of Biostatistics and Medical Informatics, İzmir*

g.hakverdi1706@hotmail.com

**Objective:** Odds ratio is a common measure of size effect in case-control studies, cohort studies and clinical trials. As the odds ratio is not easy to understand, it is often interpreted as relative risk in prospective studies. However, it is not correct to interpret the odds ratio as relative risk when the initial risk (disease incidence) is prevalent in the study population (greater than 10%). As a result of misinterpretation of the odds ratio obtained from logistic regression, the actual treatment effect on the disease or the result is not shown. The aim of this study was to examine the alternative statistical methods used to estimate relative risk in cases with widespread outcome and to compare the results of these methods on hypothetical data.

**Method:** 5 hypothetical data sets were used in the study and each data set consist of 600 subjects. The data sets were formed by taking into consideration the subjects' disease, exposure to factors and confounder variables. In this study, the estimated relative risks and confidence intervals of Mantel-Haenszel, Poisson regression, Log-binomial model and Zhang-Yu methods were calculated when the actual relative risk was 1, 2, 3, 4 and 5. According to the estimated relative risks, the odds ratios obtained from logistic regression were compared with the results of these methods.

**Results:** By examining the findings, the estimated relative risks and confidence intervals of all methods were found to be similar when the real relative risk was 1 and 2; where the actual relative risk is 3, 4 and 5, the methods give different results. In the Zhang-Yu method, the relative risk was smaller than expected and the confidence interval was narrower than other methods. Mantel-Haenszel and Log-binomial methods are closer to true relative risk in relative risk estimation. The confidence interval of the Poisson regression model was found to be wider than the Log-binomial and Mantel-Haenszel methods. The odds ratio obtained from logistic regression was estimated to be greater than the actual relative risk.

**Conclusion:** Using odds ratio as an estimator of relative risk in prospective studies with widespread results causes the risk factor to be acquired biased. This negatively affects the results of treatment and interventions in the field of health. According to the findings obtained, in the estimation of relative risk, the values which are closer to the actual relative risk were achieved by Mantel-Haenszel and Log-binomial methods. Therefore, Mantel-Haenszel and Log-binomial methods can be used to estimate the relative risk in cases with widespread outcome. Since the data in this study are hypothetical, a simulation study is needed to generalize the results.

**Keywords:** Cohort studies, Log-binomial model, Mantel-Haenszel, odds ratio, relative risk

# O33

# A Comparison of Multivariate Statistical Methods to Detect Risk Factors for Type 2 Diabetes Mellitus

**İpek Balıkçı Çiçek**[1], Zeynep Tunç[1], Şeyma Yaşar[1], Ahmet Kadir Arslan[1], İbrahim Şahin[1], Cemil Çolak[2], Saim Yoloğlu[1]

*[1]İnönü University Faculty of Medicine Department of Biostatistics and Medical İnformatics, Malatya*

*[2]İnönü University Faculty of Medicine Deparmetn of Internal Medicine Malatya*

ipek.balikci@inonu.edu.tr

**Objective**: The aim of this study was to compare the classification performance of Logistic Regression Analysis (LRA), Artificial Neural Networks (ANN) and Decision Trees Methods by using data from patients with and without Type 2 Diabetes Mellitus (DM) and to estimate importance the risk factors for Type 2 DM

**Method:** The data analyzed in this study were collected from the patients who came to the İnönü University Medical Faculty Turgut Özal Medical Center Internal Medicine Department Diabetes and Thyroid Polyclinic. The data set consists of a total of 313 patient records with 146 (47%) Type 2 DM and 167 % (53%) without Type 2 DM. In this context, possible risk factors that may be associated with Type 2 DM; gender, family history, long-term drug use, cortisone use, concomitant disease, hypertension, stress factor, heart disease, total cholesterol, smoking, alcohol consumption, exercise status, carbohydrate use, vegetable use, meat use, age, weight, height , starting age, daily bread consumption, HDL, LDL, triglyceride. Artificial Neural Networks (ANN), Logistic Regression Analysis (LRA) and Decision Trees methods were compared using the data of patients with and without type 2 DM. Accuracy, classification error, AUC (area under the ROC curve), sensitivity, specificity, and F- measure were used to compare the classification performance of the methods.

**Results:** Using the data of patients with and without Type 2 DM, the method of deciding the best classification performance from Logistic Regression, Artificial Neural Networks and Decision tree methods is Artificial Neural Networks (ANN) method. The values of classification performance criteria of this method were respectively found to be 98.94, sensitivity 100, selectivity 97.73, precision 98.04, F-measurement 99.01, AUC 0.978 and classification error 1.06. According to the results of Artificial Neural Networks (ANN) method; from the risk factors affecting the disease, sex, family history, long-term drug use, cortisone use, concomitant disease, high blood pressure, stress factor, heart disease, high cholesterol, smoking, alcohol consumption, exercise status, carbohydrate use, vegetable use, meat use, age, weight , height, starting age, daily bread consumption, HDL, LDL, Triglyceride, Total Cholesterol, fasting blood sugar independent variables obtained weight values respectively; 0.017, 0.013, 0.009, 0.008, 0.017, 0.008, 0.016, 0.024, 0.053, 0.006, 0.007, 0.023, 0.040, 0.020, 0.007, 0.046, 0.083, 0.049, 0.024, 0.066, 0.083, 0.084, 0.031, 0.020, 0.244.

**Conclusion**: When the Logistic Regression, Artificial Neural Networks and Decision Trees classification methods are applied, the Artificial Neural Networks (ANN) method has shown the best performance and according to this method, the most important risk factor that may cause diabetes is fasting blood sugar**.** The success of classification of diseases can be increased by using machine learning algorithms in this subject future studies.

**Keywords:** Type 2 diabetes mellitus, risk factors, logistic regression analysis, artificial neural networks, decision trees, multivariate statistical methods

# O34

# Comparing the Performance of Logistic Regression and Some Classification Methods Using Real Data Sets

**Yasar Sertdemir**[1], Ilker Unal[1], Hulya Binokay[1]

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

yasarser@cu.edu.tr

**Objective:** The purpose of case control studies is to obtain a prediction model or to determine risk factors. If the aim is a prediction model; Logistic regression (LR), Decision Tree (DT) Random Forest (RF), Support vector machines (SVM) and Naive Bayes (NB) can be used. The LR method has been in use for many years which may the reason why it is better known than other methods another reason why it is preferred may be the easy interpretation of the coefficients obtained in the analysis. In recent years, the popularity of other methods for classification and prediction in the field of health has been increasing. Decision trees are useful for decision making persons and it is easy to understand how and to what extent the decision is taken. The advantage of RF is that it makes a prediction using multiple decision trees instead of a single decision tree. The relationship between the DVM dependent variable and the explanatory variable/s is similar to LR when it is linear, whereas DVM is more successful than LR when it is not linear. The NB method provides posterior probabilities using prior knowledge (such as prevalence) in small samples. To compare the performance of the LR, DT, RF, SVM and NB methods using the scoring obtained from selected performance criteria from data sets in the literature.

**Method:** The data set will be divided into 70% training and 30% test data set. Training data set: LR, SVM, DT, RF and NB methods were applied. Then the predicted models were applied to the test data set, the geometric mean of the Matthews Correlation Coefficient (MCC), geometric mean of Sensitivity and Specificity (GM), the area under the ROC curve (AUC ROC), the area under the Precision Recall curve (AUC PR) and adjusted Performance criteria such as the F criterion will be calculated. This process was repeated 150 times for each data set for each method based on the averages of performance criteria rank (s) were obtained and the performance of the methods was evaluated according to the scores generated from these ranks.

**Results:** There were no significant differences between LR, DT, RF, SVM and NB scores in the data sets with low prevalence. When medium and high prevalence data sets are evaluated, RF and DVM are more successful than other methods. The mean scores obtained for the prevalence>0.35; the difference between LR (13.71±4.6), DT (13.55±5.1), RF (19.5±5.4), SVM (20.0±5.7) and NB (9.79±2.6) were found to be statistically significant ($p<0.001$). RF and SVM methods seem to be more successful in the $R^2$ group than in other methods when the $R^2$ values are low medium and high.

**Conclusion:** Based on the results of the data sets in our study, the RF and SVM methods were found to be more successful than the remaining methods in our study.

**Keywords:** Logistic regression, decision tree, naive Bayes, support vector machines

# O35

# Comparing the Performance of Logistic Regression and Different Classification Models

**Hülya Binokay**[1], Yaşar Sertdemir[1]

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

hbinokay@cu.edu.tr

**Objective:** Classification models are frequently used in health for prediction purposes and diagnosis. Logistic Regression (LR) is the most common used and best-known among these models**.** Other methods used are Decision Tree (DT), Random Forest (RF), Support Vector Machine (SVM) and Naive Bayes (NB). The aim of this study is to compare the performances of classification models for different sample size (n), prevalence and coefficient of determination ($R^2$).

**Method**: Data were simulated using the LR model with interaction terms. We simulated 1000.000 observations for each scenario (combinations of prevalence=(0.3, 0.5) and $R^2$=(0.3, 0.5, 0.7). Samples of desired size (150, 300 and 500) were sampled from simulated 1000.000 observations. These samples were divided into 70% training and 30% test sets. LR, DT, RF, SVM and NB methods were applied to the training data set. Classification models obtained from training sets were applied to the test set and the Accuracy, F-score, AUC were recorded for analysis.

**Results:** In scenarios where the coefficient of determination and prevalence was low, the DT and SVM methods were not able to do any classification in 70% and 36% respectively. We only included the simulations where all method did any classification. We observed that the NB method was the most successful method in scenarios where sample size, $R^2$ and prevalence were low. For scenarios with higher $R^2$ the RF and NB methods were the most successful methods. LR and NB methods were the most successful methods in low prevalence, high sample size and low $R^2$ scenario. LR and (RF, SVM) methods were the most successful methods for high $R^2$ scenario.

**Conclusion**: We observed that RF method were the most other methods in low sample size and high $R^2$ scenarios. The NB method was the most successful method in the scenarios where $R^2$ and sample size were low.

**Keywords**: Logistic regression, random forest, support vector machine, naive Bayes

# O36

# Investigation of H2O Automated Machine Learning Algorithm Performance on Metabolomic Data

**Meltem Ünlüsavuran**[1,2], Gözde Ertürk Zararsız[1], Gökmen Zararsız[1,2], Funda İpekten[1,2], Selçuk Korkmaz[2,3], Dinçer Göksülük[2,4], Halef Okan Doğan[5], Vahap Eldem[6], Ünal Erkorkmaz[7], İsmail Koçyiğit[8], Zafer Aydın[9], Ahmet Öztürk[1,2]

[1]*Erciyes University Faculty of Medicine Department of Biostatistics, Kayseri*
[2]*Turcosa Analitical Analysis Company Erciyes Technopark, Kayseri*
[3]*Trakya University Faculty of Medicine Department of Biostatistics, Edirne*
[4]*Hacettepe University Faculty of Medicine Department of Biostatistics, Ankara*
[5]*Cumhuriyet University Faculty of Medicine Department of Biochemistry, Sivas*
[6]*İstanbul University Faculty of Science Department of Biology, İstanbul*
[7]*Sakarya University Faculty of Medicine Department of Biostatistics, Sakarya*
[8]*Erciyes University Faculty of Medicine Department of Internal Medicine, Kayseri*
[9]*Abdullah Gül University Faculty of Engineering Department of Computer Engineering, Kayseri*

meltemunlusavuran@gmail.com

**Objective**: The H2O algorithm, which has been introduced recently in the literature, is an automatic machine learning algorithm that can automate the time-consuming and require experience stages such as parameter optimization and the determination of the appropriate method. Although this algorithm has been shown to be effective in different problems in the literature, it use has not been investigated in metabolomic data. Therefore, this study is aimed to determine the application of H2O automatic machine learning algorithm in metabolomic data, to determine which machine learning algorithms will be used automatically and to evaluate the performance of parameter optimization automatically.

**Method**: Three different metabolomic data sets were used in the study. In order to classify these data, random forest (RF) and support vector machines (SVM) methods were applied as well as H2O automatic machine learning algorithm. Performance evaluation was performed with balanced accuracy rate and Matthew correlation coefficient.

**Results**: The balanced accuracy rate, in the isoniazid class of tuberculosis data, 95% in random forest method, 100% in support vector machine and 91.7% in H2O algorithm; in the rifampicin class of tuberculosis data, 88.3% in random forest method, 90% in support vector machine and 94.1% in H2O algorithm; in the streptomycin class of tuberculosis data, 85% in random forest method, 85% in support vector machine and 95.2% in H2O algorithm was find. The balanced accuracy rate in the Cachexia data, 68.7% in random forest method, 75% in support vector machine and 92.9% in H2O algorithm was find. The balanced accuracy rate in the HCC (hepatocellular carcinoma) data, 75% in random forest method, 87.5% in support vector machine and 100% in H2O algorithm was find.

**Conclusion**: As a result, In this field has been shown that H2O method can use which has not been used previously for metabolomics data. In addition, the performance of H2O algorithm with random forest and support vector machines which are frequently used in the literature as strong statistical classification algorithms were compared. The H2O algorithm provide can be used more easily and better performance than the previously used methods of classification for application data.

**Keywords**: Automated machine learning, H2O, metabolomics, random forest, support vector machines

# O37

# Classification of Chronic Peridontitis by Using Different Machine Learning Algorithms

**Senem Koç**[1],Leman Tomak[1], Sezgin Güneş[2]

[1]*Ondokuz Mayıs University Faculty of Medicine Department of Biostatistics, Samsun*
[2]*Ondokuz Mayıs University Faculty of Medicine Department of Medical Biology, Samsun*

kocsenem79@hotmail.com

**Objective:** The aim of this study is to compare classification performances of chronic peridontitis, which is one of the most important causes of tooth loss by using different Machine Learning Algorithms.

**Method:** The dataset used in this study was collected at Ondokuz Mayıs University Medical Faculty Hospital, Faculty of Dentistry, Department of Oral Diagnosis and Radiology and Periodontology between September 2003 and November 2006 from 174 patients who had gingival complaints, 102 of whom were healthy controls and 72 of whom were chronic peridontitis patients. The dataset contained attributes like gender, vitamin D receptor gene variants, haplotype sequences, total number of teeth and caries index. The Preprocessing step included missing values and outliers determination afterwards the data was ready to use for machine learning. The dataset was split into training (60%) and test (40%) sets. Classification algorithms like K nearest Neighborhood (KNN), Naive Bayes, Support Vector Machines (SVM) and Random Forest were applied to the dataset. In the training phase, LOOCV method was applied during the analyses to overcome overfitting. Accuracy, sensitivity and specificity were used to evaluate the classification performance of the models. The analyses were performed using the R software.

**Results:** As a result of KNN algorithm the accuracy, sensitivity and specificity were obtained as; 0.7377, 0.5926 and 0.8529, respectively. Results of Naïve Bayes algorithm for accuracy, sensitivity and specificity were obtained as; 0.7541, 0.6667 and 0.8235, respectively. Results of SVM algorithm for accuracy, sensitivity and specificity were obtained as; 0.8033, duyarlılık 0.7407 and 0.8529, respectively. Finally results of Random Forest algorithm for accuracy, sensitivity and specificity were obtained as; 0.7705, 0.6667 and 0.8529, respectively.

**Conclusion:** It was determined that SVM algorithm produced better predictive results compared to other algorithms. SVM algorithm is recommended since it is effective in classifying data that have complex structure, such as genetic data with its kernel functions based on statistical learning theory.

**Keywords:** Peridontitis, pre-processing, LOOCV, nearest k neighborhood, naive Bayes, support vector machines, random forest

# O38

# GMDH-Type Neural Network Algorithms for Binary Classification: GMDH2 R Package and Its Web Interface

**Osman Dag**[1], Erdem Karabulut[1], C. Reha Alpar[1]

[1]*Hacettepe University Faculty of Medicine Department of Biostatistics, Ankara*

**Objective:** Group Method of Data Handling (GMDH) - type neural network algorithms are the self-organizing algorithms for modeling complex systems. In this study, we develop an R package, GMDH2, for binary classification. The package offers two main algorithms, GMDH and diverse classifiers ensemble based on GMDH (dce-GMDH) algorithms. The package also provides a well-formatted table of descriptive statistics in different format (R, LaTeX, and HTML). Moreover, it produces confusion matrix and related statistics, and scatter plot (2D and 3D) with classification labels of binary classes to assess the prediction performance. All properties of the package are demonstrated on Wisconsin Breast Cancer data. A Monte Carlo simulation study is also conducted to compare GMDH algorithms to the other well-known classifiers under the different conditions. Moreover, a user-friendly web-interface of the package is developed especially for new R users or applied researchers. This web-interface is available at http://www.softmed.hacettepe.edu.tr/GMDH2. In this study, our main objective is to develop an R package and its shiny-based web interface where any user can apply the aforementioned algorithms to their own datasets for binary classification.

**Method:** GMDH algorithm is a system of layers where the neurons exist. GMDH algorithm learns the relations among the variables, selects the important variables and makes the classification. Diverse classifiers ensemble based on GMDH (dce-GMDH) algorithm is the GMDH algorithm which assembles the well-known classifiers - support vector machines, random forest, naive bayes, elastic net logistic regression, artificial neural network - to make classification for binary output.

**Demonstration of GMDH2 Package and Its Web-Interface:** The GMDH2 package includes several functions especially designed for binary response. After installing and loading GMDH2 package, the functions designed for binary response are available to be used. In this part, we include some portion of the code to save space. Due to word limitation, Wisconsin Breast Cancer data set is not included in this part. It is possible to contact the authors to reach its full version if desired.

R> model <– GMDH(x.train, y.train, x.valid, y.valid, alpha = 0.6, maxlayers = 10, maxneurons = 15, exCriterion = "MSE", verbose = TRUE)

R> model <– dceGMDH(x.train, y.train, x.valid, y.valid, alpha = 0.6, maxlayers = 10, maxneurons = 15, exCriterion = "MSE", verbose = TRUE)

A web interface of GMDH2 package (Dag et al., 2018) is developed for non-R users by using shiny.

**Conclusion:** In this study, an R package and its shiny-based web interface were developed for the users to apply the aforementioned algorithms to their own datasets. The detailed information and the R code can be found in Dag et al. (2018). In this study, we compared GMDH and dce-GMDH algorithms to support vector machines, random forest, naive Bayes, elastic net logistic regression, artificial neural network with a Monte Carlo simulation. In the light of this

simulation study, the use of dce-GMDH algorithm seems to be beneficial, since it performs well under almost all scenarios and takes advantage from other classifiers when needed.

**Keywords:** R Package, web tool, data mining, machine learning algorithms, Monte Carlo simulation.

# O39

# A New Mobile Application for Sample Size Calculations in Medical Research

**Emre Demir**[1], Bedreddin Ali Akça[2], Serhat Hayme[3], Serdal Kenan Köse[3]

[1]*Hitit University Faculty of Medicine Department of Biostatistics, Çorum*
[2]*Hitit University Department of Computer Technology Osmancık Ömer Derindere Vocational School, Çorum*
[3]*Ankara University Faculty of Medicine Department of Biostatistics, Ankara*

emredemir82@gmail.com

**Objective:** While most researchers are designing a research, they first answer the question "How many subjects should I take for this study?" Accurate determination of sample size, especially in medical research, is one of the first and most important stages of the research. Correctly calculated sample size will ensure that a study has sufficient power, avoids wastage, and does not expose patients to unnecessary experimental treatment. Findings obtained with an incorrect number of subjects may result in a less powerful research and could require the repetition of the study. A study conducted with a smaller sample size may cause low power, while a study conducted with over-sample necessitates additional time and loss of costs. Although there are only a few simple mobile applications for sample size calculations in the literature, a free and comprehensive application that researchers can access at any time is not yet available. This study aims to develop a new Android mobile application that calculates sample size for medical researchers.

**Methods:** With the Android mobile application, we have developed, the sample size required for one sample t test, independent two sample t test, paired sample t test, one and two sample tests for proportions, analysis of variance (ANOVA), correlation analysis for continuous variables, and chi-square analysis for categorical variables, which is used as a common statistical test in many studies, can be determined.

**Results:** As a result of two different sample size analyses performed with our mobile application, the sample size required to detect an effect size of 0.3 with a significance level (alpha) of 0.05 and 80% power using a 1 degree of freedom (df) chi-square test was reported as a total of 88 subjects. It was reported that a total of 16 subjects should be taken to detect a difference of 3 units an infinite population with a two-tailed significance level of 0.05 and 80% power using a paired t-test. Our mobile application carried out the same results with G*Power and PASS packages.

**Conclusion:** The application we have developed is more comprehensive and faster than all of the existing applications. In addition, the report cannot be obtained via e-mail to print the results in other applications. The final sample size report of this mobile application, which can be downloaded from within the Google Play Store, is expected to be accepted in the ethics committees of our country. All researchers in our country will be able to easily calculate the sample size to use in their studies with this application and they will be able to use the report they receive from the mobile application in the ethics committee applications. Development work is continuing to make our mobile application a popular international application.

**Keywords:** Mobile application, sample size, power analysis, medical research

# O40

# Comparison of Fast Regression Methods Used in Model Selection in High Dimensional Data

**Sengul Cangur**[1], Handan Ankaralı[2]

[1]*Duzce University Medical Faculty Department of Biostatistics and Medical Informatics, Duzce*
[2] *İstanbul Medeniyet University Medical Faculty Department of Biostatistics and Medical Informatics, Istanbul*

sengulcangur@gmail.com

**Objective:** In a study involving a large number of variables, it is very important to get the most accurate predictive statistical model quickly and easily. In particular, algorithms are used to achieve accurate and interpretable models in high dimensional data where there are multicollinearity problems. The purpose of this study is to introduce Dimensional Reduction of Correlation Matrix (DRCM), Nonparametric Dimensional Reduction of Correlation Matrix (DRCM-N), Variance Inflation Factor (VIF) Regression and Robust VIF Regression methods and to compare the performances of methods by a simulation study constructed with different scenarios.

**Method:** In the presence of multicollinearity and outliers and in the case of increasing the number of independent variables considered to be included in the model, the performances of the four fast regression methods are investigated by a simulation study. The simulation work was done in the MATLAB program and by computer with Intel(R) Core(TM) i7-6500U CPU @ 2.50 GHz, 2592 Mhz, 2 core(s), 4 logical processor(s).

**Results:** In all scenarios, all target variables were selected to final model with four methods. The methods that arrive to the final model within the shortest time were DRCM and DRCM-N, respectively. The time to reach the final model of R.VIF regression method was shorter nearly two times than the time obtained by VIF regression. In all scenarios, while the maximum numbers of the total noise variables to be included in the model were obtained by DRCM and DRCM-N methods, they were selected by R.VIF regression at least. As a result, R.VIF regression method showed better performance than the other approaches in both data structures.

**Conclusion:** We propose the use of fast regression methods because of the disadvantages of traditional methods when there are large data sets and outliers in the data set, and the loss of information in the results obtained with these methods. In particular, we recommend using the R.VIF regression method as fast regression estimator in the presence of multicollinearity and outliers in the data set.

**Key Words:** Variance inflation factor, robust, dimensional reduction, high dimensional data

# O41

# Classification of ECG Data by Deep Learning Methods

**Hüseyin Kutlu**[1], İlker Etikan[2]

[1]*Adıyaman University, Adıyaman*
[2]*Yakındoğu University Faculty of Medicine Deparment of Biostatistics, Lefkoşa*

hkutlu@adiyaman.edu.tr

**Objective:** Time series have an important place in biostatistics applications. In the time series, deep learning methods in estimation and classification processes show higher performance than traditional methods. In this study, a deep learning based method for the classification of electrocardiogram (ECG) signals, which is a time series, has been proposed. ECG is widely used by physicians to examine heart health. In the manual analysis of ECG signals, the problem similar to other time series data is the difficulty of detecting and categorizing different waveforms and morphologies in the signal. For physicians, this task is both time consuming and prone to errors. The aim of this study is to propose a method that can accurately classify five different cardiac arrhythmias according to AAMI EC57 standard with LSTM from ECG signals.

**Method:** ECG signals are a time series signal. In this study, long - short term memory (LSTM) has been used in recent years. The time-varying values of the data of ECG classes with the designed LSTM model were learned. ”ECG Heartbeat Categorization Dataset“ data set was used. and ECG signal classes were learned and classified by LSTM.

**Results:** According to the AAMI EC57 standard there are five different cardiac arrhythmias. These are called N, S, V, F, Q. In the study, an average of 96.95% accuracy, 92.22% sensitivity, 97.97% specificity, 92.77% sensitivity and 92,40 f score values were obtained for 5 classes. Class F is classified as 97.29%, class N 94.42, class Q 99.26, class S 96.72% and class V 97.09. F Score, Recall, Precission, Sensivity metrics are given for the classifier evaluation. A comparative table is given in the literature**.**

**Conclusion:** The results of the study were compared with the studies in the literature. LSTM structure of the deep learning methods compared to existing methods have been seen. Deep learning methods were found to be effective in the classification of time series.

**Keywords:** ECG signals, deep learning, LSTM

# O42

## Developing a Common Metric for Assessment of Functioning in Patients with Osteoarthritis

**Pervin Demir**[1], Atilla Halil Elhan[2], Şehim Kutlay[3], Ayşe A. Küçükdeveci[3]

[1]*Ankara Yıldırım Beyazıt University Faculty of Medicine Department of Biostatistics, Ankara*
[2]*Ankara University Faculty of Medicine Department of Biostatistics, Ankara*
[3]*Ankara University Faculty of Medicine Department of Physical Medicine and Rehabilitation, Ankara*

pervin.demr@gmail.com

**Objective:** The functional assessment of patients with osteoarthritis (OA) is very important for evaluation and follow-up of treatment. There are some generic and disease-specific instruments that are valid and reliable, to evaluate functionality of the patients. To compare directly the results obtained from these instruments, the scale units of them need to be equalized. The aim of this study is to develop a common metric for three outcome measures and, examine the comparability of the results from the patients with knee OA.

**Method:** In this study, the data set from 284 patients with knee OA who completed the Health Assessment Questionnaire (HAQ), Physical functions of WOMAC Osteoarthritis Index (WOMAC) and Nottingham Health Profile (NHP) and provided the inclusion criteria was used. The common metric was construct with the simultaneous analysis of scales and the assumptions of Rasch Model (RM) for this metric was tested. The reliability of the scale was examined with the person separation index. Statistical analysis and calculations were conducted with IBM SPSS Statistics 21.0 and RUMM2030. A $p<0.05$ was accepted as statistically significant.

**Results:** The mean age of patients was 61.72±10.45 years. The p=0.331 was for the fit of RM, while the assumptions of unidimensionality and local dependence were met for the common metric. The person separation index was 0.840. There was no differential item functioning for age, gender or duration of disease. The common metric scores were defined in the range 0-100, and the equivalent values on the common metric were determined for the other scale scores.

**Conclusions:** The common metric of HAQ, WOMAC and NHP was created, which is convenient in practice for clinicians, i.e., total scores obtained from HAQ, WOMAC and NHP scales can be compared directly with this common 0-100 metric. For example, the score of 46 on the common metric is equivalent to 6 on the HAQ, 27 on the physical function subscale of WOMAC, and 4 on the physical function subscale of NHP. With the same manner, the total score of each instrument can be compared with the corresponding score in the common metric.

**Keywords:** Rasch model, osteoarthritis, common metric, functionality

# O43

# Projection Pursuit Method, Algorithms and Indexes

**Alev Bakır**[1], Aydın Erar[2]

[1]*Haliç University Faculty of Medicine Department of Biostatistics and Bioinformatics, Istanbul*
[2]*İstanbul Mimar Sinan Fine Arts University Department of Statistics, Istanbul*

alevbakir@yahoo.com

**Objective**: Problems such as the presence of extreme values and multicollinearity in multidimensional data analysis also affect statistical analysis and lead to incorrect estimates. The reduction of the data size plays an important role in eliminating these problems. Projection Pursuit is a technique for the exploratory analysis of multidimensional data sets, the method seeks out linear projection of the multidimensional data onto a line or a plane, that first described by Friedman and Tukey (1974).

**Method**: Projection Pursuit Method is a form of linear graphing in multidimensional data. The method aims at discovering and interpreting the pattern in the data with the distribution of the data by moving the high dimensional original data to a lower dimensional (one or two dimensional) space. This method has two basic steps: 1- Finding a projection pursuit index which is the size of the degree of structure or separation 2- Finding the best plane that will give the greatest value of the index. Different projection search algorithms and various index calculations have been developed.

**Results**: As a result of the practises, it can be said that the projection search method does not give very good results in the low dimensional data sets, and the method is effected even if it is not in the big rate in the presence of extreme values. It has been understood that the projection pursuit method is better suited to implement in data sets that transformed data and do not have multicollinearity. Despite the combination of various index and algorithm, Posse's Chi-Square Index has been shown to be the best end result in any kind of data distribution.

**Conclusion**: The projection pursuit method can be used as the first step because it can lead us to a comment about the data without any information about multidimensional data. However, when detailed data analysis is concerned, it is understood that it may be insufficient and supportive methods are needed.

**Keywords**: Projection pursuit technique, projection pursuit algorithms, projection pursuit index, reducing dimensional, structure removal

# O44

# An Extensive Web Interface for Validity and Reliability with RServe

**Ebru Ozturk**[1], Batuhan Bakirarar[2], Sevilay Karahan[1], Ergun Karaağaoglu[1]

[1]*Hacettepe University Department of Biostatistics, Ankara*
[2]*Ankara University Department of Biostatistics, Ankara*

ebru.ozturk3@hacettepe.edu.tr

**Objective:** The main motivation of this study is to develop a web interface for validity and reliability via using RServe on Java platform that users can conduct their analysis without knowing how to use R.

**Method:** The two main components of measurement are validity and reliability. While validity can be defined as a measurement approach that gives the degree of success in evaluating what it is designed to measure, on the other hand; reliability can be defined as the consistency in obtaining similar results upon repeated measurement applications.

In this web interface, we offer to users content analysis based on Lawesh method with respect to content validity ratio and content validity index. While for concurrent validity, the correlation coefficient with the scatter plot and Bland-Altman graphics are given, for predictive validity the results of linear regression are given based on cross-validation. To investigate convergent and divergent validity, the correlation coefficient and scatter plot again are examined. For factor analysis, explanatory factor analysis results are given. Moreover, independent two samples tests might also have presented to construct validity. For categorical data, the results of ROC analysis are observed with optimum cut off points can be listed in predictive validity.

In the reliability part of this interface, the paired samples t-test and correlation coefficient are represented in order to analyze parallel-form reliability. For interval consistency the Cronbach's alpha coefficient, intraclass correlation coefficient based on one-way ANOVA, two-way ANOVA and two-way mixed methods and Kendall's coefficient of concordance (Kendall's W) are given in the test-retest method. For Split-Half reliability, the Spearman-Brown, Rulon and Guttman formulas are supplied. The Cohen's kappa and weighted kappa coefficients proposed for the reliability analysis of categorical data.

Except for validity and reliability, item analysis is also a substantial part of the measurement. In this web interface, the results of item analysis of scale are also offered to the users.

In order to develop a web interface, we prefer to use RServe on Java platform. Rserve is "a TCP/IP server which allows other programs to use facilities of R from various languages without the need to initialize R or link against R library" (RServe-Binary RServer, 2018). For the popular languages such as C/C++, PHP, and Java, there exist client-side applications. RServe allows the users authentication, remote connection, and file transfer. Usually, RServe is used to integrate R backend for computation of statistical models, graphics in other applications. Java is a programming language that is open-source, object-oriented and concurrent. It is also a computing platform for application development.

**Conclusion:** In this study, we developed a web interface for validity and reliability in measurement based on RServe on Java platform. Our aim is serving to this interface especially for non-R users to conduct their analysis without R.

**Keywords:** Validity, reliability, RServe, web interface

# O45

# Developing Web-based Software Clustering Algorithms with RServe

**Batuhan Bakırarar**[1], İrem Kar[1], Atilla Halil Elhan[1]

[1]*Ankara University Faculty of Medicine Department of Biostatistics, Ankara*

batuhan_bakirarar@hotmail.com

**Objective:** The aim of the study is to enable users to use data mining methods without knowing the R and Java programming languages and without having to use data mining programs.

**Method:** For the analysis of the data, the USArrests data set from the data sets in the R programming language was used as the sample data set. In this study, the results of clustering algorithms were given on this data set.

**Results:** Hierarchical and K-means clustering methods were used as examples and the results of these methods were summarized with dendogram graphs.

**Conclusion:** Rserve allows to use data mining methods by using rich R libraries. At the same time, it is free and has an advantage of continuously improving and adding new functions.

**Keywords:** Rserve, clustering, data mining

# O46

# Assessing Performance of Missing Value Imputation Methods Using Data Mining Classification Algorithm

**Batuhan Bakırarar**[1], İrem Kar[1], Beyza Doğanay[1], Atilla Halil Elhan[1]

[1]*Ankara University Faculty of Medicine Department of Biostatistics Ankara*

batuhan_bakirarar@hotmail.com

**Objective:** The aim of the study is to evaluate the performance of three different imputation methods using a data mining classification algorithm.

**Method:** Mammographic Mass data set in University of California Irvine (UCI) Machine Learning database was used in this study. There are data from 6 variables of 961 patients in the data set. These variables are BIRADS, Age, Mass Shape, Mass Margin, Mass Density and Class. Data set was randomly deleted with having 10%, 20% and 30% of missing data. Three methods used for imputation were using mean or mode, Chained Equations and Multilayer Perceptron.

**Results:** Multilayer Perceptron was the best performing method in all three cases. The Multilayer Perceptron method was followed by Chained Equations and using mean or mode values, respectively.

**Conclusion:** As a result, imputation with prediction methods, using the relations between variables and data structure to fill missing value, were better results than simple imputaion methods and were recommended to be used in the literature.

**Keywords:** Missing value, imputation, prediction methods, chained equations

# O47

# Comparing Prediction Abilities of Classification Algorithms on Determining Specific Chemical Substances

**Uğur Toprak**[1], Emrah Gökay Özgür[1], Serhat Hayme[1], Beyza Doğanay Erdoğan[1]

[1]*Ankara University Faculty of Medicine Department of Biostatistics, Ankara*

toprakugur@ymail.com

**Objective:** Over the past few decades advances in molecular sciences, especially in chemometrics, allowed systematic approaches to understand structure of chemical substances, and has facilitated new discoveries in drug discovery, pharmaceutical chemistry and medicine. Determining chemical materials has vital function in human health and response to environment. Recent studies showed that investigation of chemicals could provide useful information for chemical substance classification at the molecular level due to their ability to measure the unique features and their interactions for thousands of raw chemical materials simultaneously. Although, chemical material classification methods have improved, there is still a need for a fully automated and objective method for determining chemical substances. This study aims to detect best classification model for raw chemical substance dataset among several classification algorithms and to compare their performances.

**Method:** Raw chemical substance dataset, which is selected for this study, consists of four different classes; AcDiSol, DiCaP, Kollidion, and MCC, chemical substance types and some other samples which does not belong to any class. We applied four widely used classification algorithms to determine membership of chemical substances on data by using The UnscramblerX version 10.5.1., Soft Independent Modeling of Class Analogies (SIMCA), Support Vector Machines (SVM) and Linear Discriminant Analysis (LDA) algorithms were used for classification. Cross validation and split train were applied to train the dataset. Classification performances were detected to find the best classifier.

**Results:** Prediction accuracy of the classification methods differs according to how they handle the data. Our results showed that test data gave the classification success of SIMCA 8-11% over SVM and 12-16% over LDA. SIMCA had superior performance (100%) in classification of DiCaP and Kollidion, and LDA was the least successful classifier (76%) in AcDiSol.

**Conclusion:** According to our results, SIMCA individually showed superior to the others methods on this raw chemical substance data set. However, it should be taken into account that statistical, computational and representational limitations of the algorithms were determinant in the classification. Consequently, for both raw chemical substance data set and simulated data set, SIMCA was the best and fastest solution to classify less subjectively. However, results still needs to be checked by experts in case of strong bias or some rare conditions.

**Key Words:** Chemometrics, classification, SIMCA

# O48

# Functional Data Analysis and an Application of the "Functional t-test"

**Çağla Şafak**[1], Zeynep Yavuz[1], Beyza Doğanay Erdoğan[1]

[1]*Ankara University Faculty of Medicine Department of Biostatistics, Ankara*

caglasafak01@gmail.com

**Objective**: The first step of a research plan is to determine the appropriate statistical method for the research data. This study will introduce Functional Data Analysis (FDA), a method for repeated measures which has started to show popularity in recent years, and an application for the two sample case will be discussed. Methods used for repeated measurements gain more importance, especially in the analysis of the data collected for conditions such as the success of treatments in the health field like the healing process of the patients. Functional data analysis allowing the evaluation of repeated measurements collected at different time points as well as data collected at the same time points makes it a preferable method. In this study, discrete time points were transformed into a function to make predictions about the characteristics of an underlying function of the measurements taken from the same individual at different time points on a real medical data set. In order to make statistical comparisons and inferences, functional data analysis methods were applied to the functions obtained and the results are discussed.

**Method:** In the scope of this study female breast cancer patients aged 29 and older who underwent radical mastectomy and had their lymph nodes removed at Ankara Numune Hospital were randomly divided into two groups. The first group (n=27) received treatment at hospital while the second group (n=30) received a home program. Preoperative (day 0), post operative day 5, 30, 90 and 180 flexion (shoulder range of motion) measurements were collected from each patient at the same time points. Each individual had the exact same number of measurements. Flexion measurement functions were created for each patient and Functional t –test was used to compare the patients who were randomly assigned to either hospital treatment or home program groups.

**Results:** In this study, the flexion values of individuals in the hospital treatment and home program groups were compared. The first step of the analysis was to convert the data points taken from the same individual at five different time points into functions and then to smooth them in order to use in the proceeding steps. After these procedures, the individuals in each treatment group had similar mean values at the beginning, however, at the later time points, the mean curve of the hospitalized group was observed to be higher than the average curve of the individuals receiving treatment at home. In addition, individuals in these treatment groups had similar variations at baseline, but it was observed that individuals receiving treatment at home after mastectomy had higher variation afterwards. According to the t-statistic plot obtained from the Functional t-test, patients in both hospital treatment and home program groups showed no statistical difference until the operation however in the post-operative process the statistical differentiation of the flexion values are evident.

**Conclusion:** The t-statistic graph obtained from the Functional t- test allows us to say that the results are consistent with the analysis of the same data with mixed-effect models. As a result, it can be said that functional data analysis approach can be used in the analysis of repeated measurements along with the approaches in the literature.

**Key words:** B-splines, smoothing, functional t test, functional data analysis.

# O49

# Innovation in Biostatistics Education: Multidisciplinary Approach in Undergraduate Medical Education

**Ferhan Elmalı**[1], Funda İfakat Tengiz[2]

[1]*İzmir Kâtip Çelebi University Faculty of Medicine Department of Biostatistics, İzmir*
[2]*İzmir Kâtip Çelebi University Faculty of Medicine Department of Medical Education, İzmir*

ferhan.elmali@ikc.edu.tr

**Objective:** The aim of this study is to provide information to the researchers about the education of Biostatistics which will be given with different teaching methods within the scope of "System Based Integrated Education Model" which is implemented in the undergraduate medical education at the University of Izmir Kâtip Çelebi (IKCU) from 2018-2019 academic year.

**Method**: Developing technology in recent years and increasing number of students in faculties with Flipped Class teaching method is recommended in Biostatistics education. With the Flipped Class teaching method, students can follow the lesson videos prepared by the pre-lesson instructor and practice with the small education groups they are involved in during the lesson. Until 2017-2018 academic year, Biostatistics topics given in the first two boards of the first semester in the Faculty of Medicine at the University of Izmir Kâtip Çelebi. As of the 2018-2019 academic year, Human and Community Health (HCS) in the vertical corridor will be given in the semester 1, 2 and 3 with different teaching methods including the Flipped Class method. In addition to the Flipped Class method, small group trainings, assignments, projects and field studies will be carried out with different disciplines.

**Results**: At the end of the 2020-2021 academic year, students' view to Biostatistics, achievement perception and satisfaction levels will be evaluated by doing research. The effectiveness of the teaching methods will be evaluated.

**Conclusion**: By using all these teaching methods, it is aimed to be accepted by physician candidates how important and necessary of biostatistics in general medicine and scientific researches.

**Keywords:** Biostatistics, integrated education, effective learning, flipped class, team study education.

# O50

# Power Analysis in Multiple Linear Regression Models

**Ayhan Parmaksız**[1], Reha Alpar[2]

[1]*Hacettepe University Health Sciences Institute Department of Biostatistics, Ankara*
[2]*Hacettepe University Faculty of Medicine Department of Biostatistics, Ankara*

aparmaksiz@gmail.com

**Objective:** The aim of this study is to discuss the rules and methods for determination of sample size for 80% power in multiple linear regression analysis and to determine the sample sizes required for different correlation structures between dependent and independent variables.

**Method:** In multiple linear regression models, the required sample size is usually calculated by some rules those are relation with the number of independent variables. These rules are generally proposed by researchers who study on a specific area, ignoring power analysis procedures. The most difficult stage of the procedure for theoretical (a priori) power analysis is to determine the effect sizes. Even for the difference between the averages of the group, similar but different effect size definitions have been made. In fact, different effect sizes have been made by scanning the literature in different fields. As will be understood, the effect size problem is complex even for univariate analyzes, and is an even greater problem for multiple linear regression analysis that requires a multivariate effect size definition. For linear regression analysis, studies have been done to determine the sample size by taking into account the power analysis procedures (Green, 1991). Maxwell (2001), by taking Green's work further into account the correlation structures between dependent and independent variables have made an effect size definition and suggested a formula to determine the sample size. In some studies conducted outside of these studies, $R^2$ was used as the effect size and it was tried to calculate the required sample size for $R^2$. In our study, the efficacy of R2, which is used to determine the sample size, was examined. In addition, Maxwell's (2001) suggested that the average correlation between the variables used in the formula is also discussed in the adequacy of solving the problem.

**Results:** With the data sets produced for different correlation structures, sufficient sample sizes for 80% power were calculated and presented with summary tables and graphs. While applying the simulation study, solutions were taken into consideration for the variance swelling rates for the combinations of interest and relatively high values were observed in very few scenarios where the sample size was less than 20. For each scenario, very different sample sizes were calculated for different correlation scenarios which calculated the $R^2$ value and produced similar $R^2$ values. Different combinations of the same correlation values were studied and different sample sizes were determined.

**Conclusion:** In multiple linear regression analyzes, the $R^2$ value used to determine the sample size for 80% power is not an appropriate effect size. The average correlation value that summarizes the correlation between variables for Maxwell's suggested correlation is not sufficient to determine the correct sample size.

**Keywords:** Multiple linear regression models, power analysis, sample size, simulation study

# SHORT ORAL PRESENTATIONS

# SO1

# Application of the Proportional Hazards Mixed Cure Model on Patients with Breast Cancer

**Özge Pasin**[1], Handan Ankaralı[2]

[1]*Istanbul University Faculty of Medicine Department of Biostatistics, Istanbul*
[2]*Istanbul Medeniyet University Faculty of Medicine Department of Biostatistics, Istanbul*

ozgepasin90@yahoo.com.tr

**Objective:** Cox regression method is the most frequently used method in investigating the risk factors affecting the survival times in survival studies. Proportional hazards assumption must be provided to apply the Cox regression model. The method assumes that there were many individuals experiencing the event and that the individuals were sensitive to experience the event. However, this is not provided in some cases. Currently, there has been a rapid improvement in studies investigating particularly cancer associated mortalities and metastasis. Therefore, there are many individuals who have not experienced the event (death, recurrence, etc.) in cancer research. The use of Cox model will result with misleading results when the number of patients receiving treatment is high because Cox regression analysis is based on the assumption that there some individuals experiencing the event. In such cases cure models are used as an alternative method. The aim of the study was to explain the theoretical features of the cure models, and to apply on an appropriate data set, and compare the results using the Cox regression model.

**Method:** The study was performed on 295 women diagnosed with breast cancer obtained from the Holland Cancer Institute. The death of the patients were evaluated as event, and other patients were identified as censored during the study. The risk factors that might affect the survival time were taken as age (aged below or above 44), receiving chemotherapy, receiving hormone treatment, and histopathological stage(mild-moderate differentiated, well-differentiated). Both Cox regression method, and proportional hazards mixed cure model(using the logit function) were applied to the data set after testing the proportional hazards assumption. The data were analysed using the R package.

**Results:** The data in the study provided the proportional hazards assumption. The outcome of the Cox regression analysis showed that age, receiving chemotherapy, receiving hormone therapy had no significant effect on the hazard ratio ($p>0.05$), however, histopathological stage was found to have a significant effect on the hazard ratio ($p<0.001$). The outcome of the proportional hazards mixed cure model showed that the factors had no significant effect on the cure ratio for the cured section ($p>0.05$). The investigation of the results for the uncured section showed that age, receiving chemotherapy, receiving hormone therapy, but the histopathological stage had statistically significant effects ($p<0.001$, for all).

**Conclusion:** Cure models may be used rather than Cox regression method when the presence of individuals who did not experience the investigated event in the long term, and will give more detailed results**.**

**Keywords:** Survival time, cure model, treatment, Cox regression, proportional hazards

## SO2

# Parametric Accelerated Failure Time (Aft) Model Application in Lung Cancer Patients

Nural Bekiroğlu[1], **Ilgın Cebeci**[2], Hüseyin Aykut[2], Melissa Acar[2]

[1]*Marmara University Faculty of Medicine Department of Biostatistics, Istanbul*
[2]*Marmara University Institute of Health Sciences Department of Biostatistics, Istanbul*

ilgincebeci@gmail.com

**Objective:** Cox Proportional Hazards (CPH) model is the most frequently used multivariate regression model for survival analysis. However in practice proportional hazards(PH) assumption may not hold. In this case, Parametric Accelerated Failure Time (AFT) models are applied. In this study, AFT and CPH models were tested on lung cancer data, and the results were compared.

**Method:** Survival data of 60 lung cancer patients were analyzed. The dependent variable, PFS (progression-free survival time) and 10 independent variables (sex, age, cancer stage, tumor location, tumor type, tumor size and invasion, lymph node involvement, vascular invasion, adjuvant chemotherapy and recurrence) were tested. For univariate analysis; LogRank for categorical variables, and simple Cox regression for continuous random variables were employed. Statistically significant variables were included in multivariate Cox model. Hazard Functions of PFS for log-normal, log-logistic, Weibull and exponential distributions were examined. Model goodness of fit test and Akaike Information Criteria (AIC) values were computed to find the best AFT model. Log-logistic regression was used as AFT model. The data were analyzed in STATA 10 and SPSS 15 statistical package programs. $P < 0.05$ was considered statistically significant.

**Results**: Variables of lymph node involvement, tumor type, cancer stage, vascular invasion, adjuvant chemotherapy and recurrence were found significant as the result of univariate analyses. The variables of recurrence, tumor type, vascular invasion and adjuvant therapy were found statistically significant as the result of the multivariate Cox model. For AFT model, Hazard Functions of log-normal, log-logistic, Weibull and exponential distributions of dependent variable PFS were examined. AIC values were 205, 204, 216 and 242 respectively. According to the AIC value and hazard function distributions, the optimal AFT model was determined as Log-logistic. The result of log-logistic regression showed that adjuvant chemotherapy, recurrence and tumor type variables were statistically significant for the AFT model.

**Conclusion:** CPH model is one of the most commonly used survival model. Models such as the AFT regression models could be more suitable to use where the PH assumption does not hold. Also, it is important to note that the interpretation of the parameters obtained from the AFT and CPH models differ. Comparisons of survival times are made in AFT models while hazards are compared in CPH model. As biostatisticians, we should recommend the survival models, suchs as AFT which would allow for more accurate analysis and interpretation,other than CPH model where applicable.

**Keywords:** Survival analysis, Cox model, AFT model, log-logistics

# SO3

# Cox Proportional Hazard Regression Model and Random Forest Model Application on Patients with Glioblastoma Multiform Brain Tumor

**Ilgın Cebeci**[1], Özkan Alan[2], Hüseyin Aykut[1], Melissa Acar[1], Fulden Yumuk[2], Nural Bekiroğlu[3]

*[1]Marmara University Institute of Health Sciences Department of Biostatistics, İstanbul*
*[2]Marmara University Faculty of Medicine Department of Internal Medicine Division of Medical Oncology, İstanbul*
*[3]Marmara University Faculty of Medicine Department of Biostatistics, İstanbul*

ilgincebeci@gmail.com

**Objective:** In this study, it was aimed to apply and to compare the results of Cox Proportional Hazard Regression (CPHR) model and Random Forest model in order to determine more accurately prognostic factors of glioblastoma multiforme brain tumor patients.

**Method:** This study is carried out with Marmara University Medical Oncology Department; a total of 38 independent variables of 99 glioblastoma multiforme brain tumor patients were analyzed by log rank test. Univariate survival analysis were performed according to the pathology result, with the time to progression up to progression. The stepwise CPHR model was applied to the meaningful variables. In addition, a model comparison was made after applying Random Forest model (Salford Predictive Modeler 8.2).

**Results:** 7 variables such as; the performance status which is called ecoggrup, presence of a mutation which is called artx, the reason for drug withdrawal, proteinuria, hypertension, sex, and presence of deep vein thrombosis (dvt) were found to be significant. Stepwise Cox Proportional Hazard regression model proteinuria and hypertension were found respectively significant. Random Forest analysis the reason for drug withdrawal, proteinuria, sex and hypertension were found significant.

**Conclusion:** Prognestic factors of proteinuria, gender and hypertension has great impact in patients with glioblastoma multiforme brain tumors, especially when the reason for drug withdrawal is due to progression.We think that interpretation of CPHR model results supported by machine learning models in survival analysis will provide more accurate diagnosis and treatment.

**Keywords:** Cox proportional hazard regression model, survival analysis, random forest model, machine learning

# SO4

## Determination of Effective Variables by Cox Regression Method for Appropriate Matching in Kidney Transport

**Pembe Keskinoğlu**[1], Mehmet N. Orman[2], Timur Köse[2], Su Özgür[2], Erhan Tatar[3], Adam Uslu[3]

[1]*Dokuz Eylül University Faculty of Medicine Department of Biostatistics and Medical Informatics, İzmir*
[2]*Ege University Faculty of Medicine Department of Biostatistics and Medical Informatics, İzmir*
[3]*İzmir Bozyaka Education and Research Hospital, İzmir*

pembe.keskinoglu@deu.edu.tr

**Objective:** Researches in the health field focused on information discovery using large databases with the contribution of informatics and studies at the molecular level with new approaches. However, the objectives for biostatistics are unchanged. The main targets to be achieved are to determine the relationships between variables and to do forward-looking predictions. Some of the data in the health field, where predictions are made forward, should be tested with time dependence. Occurance of the result variable does not matter only. The passing time and the presence of variables within this time that affect the result is also important. This data structure is the subject of Cox regression analysis which takes into account time dependence. The prediction of the result variable also determines the realization of the medical procedure to be performed in the previous process. One of the most appropriate issues for this condition is kidney transplantation. In particular, it is predicted that survival of the cadaveric kidney transplanted with a successful decision for donor-recipient matching will be long. In appropriate donor-recipient matching, HLA tissue compatibility and some features of the donor-recipient are effective. In this study, the characteristics affecting the survival of the cadaveric transplanted kidney will be determined and the contribution of decision support systems will be investigated in appropriate matching for transplantation.

**Methods:** The data of 130 patients in Bozyaka Training and Research Hospital who underwent cadaveric kidney transplantation between February 2013 and July 2017 were used. Cox regression method was used to estimate the factors affecting the decision of transplantation in the organ transplantation decision system.

**Results:** The mean age of the patients and donors was; 46.6±10.2, 46.4±15.9, respectively. The mean follow-up was 22.2 ± 15.5 months. Graft rejection was occured in 19.9% of the subjects (n=22). Predictions were made with the models prepared for the recipient (age, cold ischemia time), donor (age, diabetes, hypertension) and tissue compliance (HLA)for transplant decision. Recipient age and tissue adaptation variables were statistically significant for transplant decision ($p<0.05$).

**Conclusion:** The organ transplantation decision system in our country is carried out by a legal regulation based on experience, which is not based on biostatistics mechanisms. The development of relevant decision support systems will not only provide scientific contribution, at the same time will contribute to the development of organ transplant system in Turkey. In this limited study, the variables of tissue adaptation and recipient age were found to be important for the successful decision of cadaveric renal transplantation. When the organ transplantation matching systems in Europe and America are examined, it is seen that many variables belonging to the recipient and donor are effective. The variable of tissue adaptation

in the models was included in our database as the only count value. However, compliance assessment should be calculated according to the prevalence in the community by Hardy Weinberg equality. There are some limitations in the decision to tranplantation. First, five-year follow-up data is insufficient. Second, the independent variables are included in databases categorically. Third, time dependence of the independent variables is not included in models. Correct transport decision estimates should be made using larger samples where the frequency of HLA tissue structure of the population is evaluated and modelled.

**Keywords:** Cox regression, kidney transplantation, decision support system

# SO5

# Determination of Learning Styles of Medical Faculty Students: Preliminary Study for a Comprehensive Structural Equation Modeling

Kemal Turhan[1], Burçin Kurt[1], **Zeliha Aydin Kasap**[1], Serbülent Ünsal[1], Emin Mollahasanoğlu[1], Erşan Kalayci[1], Merve Gülbahar Çakmak[1], Yusuf Baykal[1], Merve Yürüyen[1], Ayşe Pak[1], Amine Bayrakli[1]

[1]*Karadeniz Teknik University Faculty of Medicine Department of Biostatistics and Medical Informatics, Trabzon*

zelihaaydin86@gmail.com

**Objective:** This study is a preliminary preparation of a comprehensive structural equation model study which is still in progress on the students of Karadeniz Technical University Medical Faculty. The aim of this study was to determine the learning styles of medical students and the factors affecting them. So; The first step of a comprehensive structural equation modeling study to be done by integrating the students' success levels and learning environments in the committee tests with the learning styles defined in the same students has been completed.

**Method:** In 2017-2016 academic year, the Grasha and Riechmann Learning Styles Scale (GRLSS) was applied to the students studying at Karadeniz Technical University (KTU) Faculty of Medicine, to determine the learning styles and to examine their relationship with the success levels of the courses. In addition, the students were asked questions about their demographic characteristics. GRLSS; It describes the learning styles in 6 sub-headings, including avoidant, participatory, competitive, collaborative, dependent and independent, and consists of 32 items. The sample of the study consists of 172 students. In the study, by taking into consideration the factor loads of each learning style of the scale, the scores of the students were calculated and the learning style levels were also low, medium and high. Descriptive statistics were calculated, multiple chi-square test was used, data were analyzed at $\alpha = 0.05$ significance level and analyzed in IBM SPSS Statistics 22 program. In this context, the required permits are taken from the related persons and authorities.

**Results:** The scales were applied to 236 people and the analysis was carried out with 172 data by excluding the missing data. 54.1% of the students were female and 45.9% were male. The mean age was 20.48±1.2 years. When the social media usage levels of the students were examined, it was seen that 5.8% never used it, 10.5% used it rarely, 25.6% used it a little, 36.6% used it, 21.5% used it a lot. When the reading levels of the students were investigated, it was found that 6.9% never read, 16.7% read rarely, 39.6% read less, 30.6% read them, 6.3% read much. In addition, students' places of residence were examined and they preferred to stay at the dormitory (40.1%). Each of the learning styles of the students showed moderate accumulation. According to the multiple chi-square test, when learning style levels were compared according to 6 learning styles, a statistically significant difference was found between them ($p<0.001$).

**Conclusions**: This is a preliminary preparation of a comprehensive structural equation modeling study whose details are still ongoing. Here, only the age, gender, accommodation preferences, social media usage levels and parental education levels of the KTU medical faculty students and the results of GRLSS were evaluated. Each of the learning styles of the students showed a moderate level of accumulation. Thus, it was concluded that medical students could perform learning in each learning style.

**Keywords:** Grasha and Riechmann learning styles scale, GRLSS, learning styles, multi chi-square

# SO6

# Application of the Heckman Sample Selection Model on Smoking Addiction

Özge Pasin[1], **Başak Gürtekin**[1], Elif Gülen Onur[2]

[1]*İstanbul University Faculty of Medicine Department of Biostatistics, İstanbul*
[2]*Düzce University Faculty of Medicine Department of Family Practice, Düzce*

bgurtek@istanbul.edu.tr

**Objective:** Although Heckman sample selection models are frequently used in econometry, the use of this models in health area is rare. The aim of the study was to explain the reasons of use, and theoretical features of Heckman sample selection model and to apply the model on an appropriate data set on health area.

**Method:** Selection of the sample is a problem encountered in various research. A non-randomised selection procedure will cause a source of bias in the research. Heckman model was recommended for correcting the selection bias, and has been used for the statistically correction of bias which may develop due to non-randomised sample selection. In addition, these models are used in case of mullicollinearity, and in the presence of heterogenity in the error term. The risk factors affecting the smoking cessation period, and smoking cessation condition included in the selection model were investigated using the data of patients who presented to smoking cessation polyclinic in the Department of Family Medicine in Duzce University between March 2015, and March 2016. The risk factors were taken as the age, sex, the number of pack-year, alcohol use, and Fagerstrom nicotine dependence scores. The presence of sample selection error was tested using the Inverse Mill's Ratio. The STATA 15 program was used in the application of the model.

**Results:** 263 participants out of 473 gave up smoking. The ratio of the individuals included in the selection model was 55.6%. A statistically significant Inverse Mill's Ratio was obtained ($p=0.023$). It showed that neglecting the observations which had no cessation time on the model gave erroneous results. Although age had a statistically significant effect on the smoking-free period on the Heckman model ($p=0.005$), sex, and the use of alcohol had no significant effect ($p>0.05$). Fagerstrom nicotine dependence score was found statistically significant for the selection model ($p=0.001$).

**Conclusion:** Heckman selection model is an alternative method that may be used in case of sample selection bias on the selection model and may eliminate the biased, inconsistent prediction results in linear regression. The use of this method in health area must be extended, and more studies must be performed on this topic.

**Keywords:** Heckman model, sample selection, bias, ınverse Mill's ratio, smoking addiction

# SO7

# New Experimental Design Model Affecting Cure Development Process: The Basket Design

**Mesut Akyol**[1], Yağmur Polat[1], Yavuz Sanisoğlu[1]

[1]*Yıldırım Beyazıt University Faculty of Medicine Department of Biostatistics, Ankara*

mesutakyol@gmail.com

**Objective:** The objective is this study is to promote the Basket Design, which is a clinical trial design that can be designed based on the tumour mutation and requires much smaller sample size compared to other designs.

Clinical trials are studies designed specifically to diagnose different diseases and improve the ongoing process. In early stage clinical cancer studies where cure efficacy is being examined, determining which type of cancer the cure in development process has healing property on and detecting if the cure really has an impact on the relevant disease is crucial for the researcher.

The optimal two staged experiment design Simon has designed for Phase II clinic trials, the optimal three staged design conceived by Ensign et al. for Phase II continuum and the optimal three staged experiment design of Chen et al. for Phase II cancer research are some of the most used experiment designs used in cancer research. In cancer research, Phase II studies are vital since it is the phase where the target cure is first experimented on patients of the related disease. However, it's extremely challenging to test the development of the cure in low prevalence diseases as the development tests require extensive sampling.

For studies that require extensive sampling, to acquire the satisfactory number for extensive sampling and comprehend the efficacy of the cure that can be used as a cure for the disease, a longer time period is required than it is for other studies. Therefore, Phase II trials for diseases with low prevalence cost more time and fund. Rising costs affect the development process of the cure adversely, and causes the researchers and their supporter, pharmaceutical industry to lose their interest. In result of this, the cure is either not developed, or come to existence as Orphan medicinal product. Due the fact that the mandatory clinical trials were not finished, the patients have difficulties accessing the cure.

**Method:** In the study, the Basket Design will be promoted in accordance with information that was obtained from the literature, and will be compared with other designs in the light of the literary information.

**Results:** The recent cancer researches showed that in early stage clinical cancer trials, the experiment designs based on the mutation of the primary tumour that causes the disease are more efficient than the experiment designs based on the anatomical location where cancer is located. In the light of that information, new clinical experiment designs that can be used especially in Phase II and Phase III trials came to existence. The Basket Design is one of those experiment designs. The simulation studies in the literature for the Basket Design shows that when the design plan is made correctly and when the right restrictions are applied, the development process of the cure can be shortened compared to other designs.

**Keywords:** Experimental design, clinical cancer trials, phase II clinical studies, the basket design

# SO8

# The Commonly Used Statistical Methodologies in Genetic Epidemiology for Different Types of Data

**Yusuf Kemal Arslan**[1], Selen Begüm Uzun[2], Derya Gökmen[2]

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

ykarslan@gmail.com

**Objective:** Advances in technology are rapidly generating vast amounts of genotyping data in the human genome. To maximize the use of this information in unravelling the etiology of complex diseases, statistical approaches are needed. The purpose of this study is to present the statistical methodologies for genetic epidemiology that using common like association analysis, data mining, neural and gene networks, model selection and Bayesian methods, multivariate analysis, multistage designs, structural equation modeling (SEM). Especially SEM can be used for solving systems of linear and non-linear equations, to test the overall fit of "causal" disease model that models individual genes as latent variables defined by multiple single-nucleotide polymorphisms (SNPs).

**Method:** SEM comprises two general sub-models: i) a measurement model that develops the relationships between the observed variables and the latent variables ii) a structural model that develops the relationship between the latent variables. For fitting functions, the robust weighted least-squares estimator (WLSMV), which allows for estimation of binary and categorical dependent variables, and a ML estimator robust to non-normality (MLF) could be used. Factor analysis (FA) could be used to determine the proportion of average variance explained (AVE). The chi-square test could be used whether the specified model is significantly different from the alternative model, which assumes the data are from a multivariate normal distribution with an unconstrained covariance matrix. Therefore, when categorical or non-normal dependent variables are modeled, a modified chi-square test or other fit index robust to non-normality is needed.

**Results:** The root mean squared error of approximation (RMSEA) is an absolute fit index which represents dispersal of data to model discrepancy across degrees of freedom; and, a RMSEA value of less than or equal to 0.05 is believed to represent the boundary of acceptable fit. A good fitting model could be checked by Comparative Fit Index (CFI), RMSEA or WRMR fit index standards.

**Conclusion:** In this study model fitting indexes for SEM in genetic epidemiology was explained.

**Keywords**: Genetic epidemiology, statistical methods, SEM

# SO9

# Statistical Methods in Genetic Epidemiology: Investigation of Familial Clustering

**Selen Begüm Uzun**[1], Yusuf Kemal Arslan[2], Yaşar Sertdemir[3], Gülşah Seydaoğlu[3]

[1]*Ankara University Faculty of Medicine Department of Biostatistics, Ankara*
[2]*Erzincan University Faculty of Medicine Department of Biostatistics, Erzincan*
[3]*Cukurova University Faculty of Medicine Department of Biostatistics, Adana*

selenbegumuzun@gmail.com

**Objective:** Genetic epidemiology is the field which study the genetic structure in human populations and the effects of the compound on the disease and the transition of the disease in families. Investigation of the genetic aspects of the disease are done in several stages, one of these stages includes familial clustering studies. Familial clustering is mainly based on the evidence that clustering is in the family. One of the first clues of genetic etiology is that the disease is observed in families more than normal. When analyzing clustering in the family, many different statistical methods are used such as Binary Logistic Regression, Exact Inference for Family Disease Cluster, Regressive logistic regression, Generalized Estimation Equation (GEE) and some others.

Hematology can be defined as a science that deals with the structure and function of blood in normal and diseased conditions. There are three different types of hematological cancer; leukemia, lymphoma, myeloma for which each was shown to have the risk of familial clustering. However questions of the existence of familial clustering on all hematological cancers are truly not investigated.

The objective of this study is to evaluate Exact Inference for Family Disease Cluster and Regressive Logistic Regression to be applied for the identification of the familial cluster in a data on hematological cancers from population based case control study.

**Method:** Information of hematological cancer patients(353) compared with Solid CA(353) and Non-CA(352) control groups. In the first method, family history of disease took into account. Results of regressive logistic regression are extended version of first method. All statistical analyses were performed using SPSS (IBM Corp. Released 2011. IBM SPSS Statistics for Windows, Version 20.0.) and pfc.sim (probability of familial clustering of disease-simulation) code of gap (a genetic analysis package) in R.

**Results:** In regressive logistic regression, informations of all siblings calculated and the highest probability is in solid group and then hematological group and the control group has the lowest probability of cancer. Results of exact inference method, there is a no familial aggregation in solid cancer and hematologicak cancer group.

**Conclusion:** The advantages and disadvantages of the methods are discussed and the appropriate method for the aim is proposed.

**Keywords**: Familial cluster, regressive logistic regression, hematological cancer

# SO10

# Measures of Clustering and Heterogeneity in Multilevel Logistic Regression

**Sevinç Püren Yücel**[1], Meryem Göral Yıldızlı[1], Nazlı Totik[1], Zeliha Nazan Alparslan[1]

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

sevincpurenyucel@gmail.com

**Objective:** Multilevel data structures occur frequently in public health, health care research, behavioral research, and epidemiological research (especially social epidemiology). One suitable method for the analysis of such data is multilevel regression analysis. Multilevel models include cluster-specific random effects that allow one to explain the total variance as cluster-based and individual-based variance. In these models, some measures of variance components and heterogeneity are needed to estimate the general contextual effect, i.e. the effect of the clustering on the response. One of these measures is the variance partition coefficient which represents the total observed variation ratio in the response attributed to the cluster variation. Another is the median odds ratio, which measures the magnitude of the clustering effect when a multilevel logistic regression model is used.

This study aims to investigate the applicability and interpretability of the above mentioned measures in multilevel logistic regression with a data set obtained from the database of Balcalı Hospital, Faculty of Medicine in Çukurova University.

**Method:** The death status of the patients will be modeled by multilevel logistic regression analysis using variables including the hospitalization information of patients (i.e. age, gender, degree of complexity of the operation, department, etc.) operated in the surgical departments of Balcalı Hospital (Çukurova Medical School). Afterwards, measures defined in the related literature are evaluated.

**Results and Conclusions:** Expected results can be summed up as theoretical and practical ones:

Theoretical result: It will be emphasized that researchers should use a combination of the variance partition coefficient (which determines the variance ratio resulting from the difference between the clusters) and the median odds ratio(which measures the magnitude of the overall contextual effect).

Practical result: Clustering and heterogeneity measures in multilevel logistic regression will shed light on the use of Hospital Information Management Systems (HIMS) and will lead to beneficial use of hospital data for information retrieval.

**Keywords:** Variance partition coefficient, median odds ratio, multilevel modelling, logistic regression

# SO11

# Population Attributable Fractions of Tobacco Related Cancers in Turkey and Seven Geographical Regions

Sultan Eser[1], **Su Özgür**[2]

[1]*Balıkesir University Faculty of Medicine Department of Public Health, Balıkesir*
[2]*Ege University Faculty of Medicine Department of Biostatistics and Medical Informatics, İzmir*

suozgur35@gmail.com

**Objective:** Anti-tobacco interventions have been effective in many developed countries where lung and other tobacco-related cancers are declining. In Turkey, the first tobacco legislation put into practice in 1996 and strengthened in 2008. Aim of this work was to quantify the size of the problem in each of the seven geographical regions and in the whole of Turkey.

**Methods:** We followed the methods proposed by Peto et al. and Parkin et al. Incidence rates of lung cancer and other tobacco-related cancers were obtained from Cancer Incidence in Five Continents Volume XI (CI5C) for 8 provinces and projected to the regions. Estimates of the incidence of lung cancer in non-smokers and relative risks for the other cancer sites were obtained from the large CPS II cohort of the American Cancer Society. By combining these parameters in the usual formula due to Cole & McMahon we obtained estimates of the fraction of tobacco-related cancer attributable to smoking (Population Attributable Fraction: PAF).

**Results:** For Turkey, in total, 59.4% of the 351591 TRC cases can be attributed to tobacco smoking; 74.5% in males, 6.8% in female. PAFs by cancer sites are as follow: 89.6%, 38.4% of lung; 86.4%, 5.1% of laryngeal; 70.3%, 4.3% of esophageal; 70.5%, 3.1% of oral cavity & pharyngeal; 54.4%, 0.4 % of kidney; 53.2%, 1.7% of bladder; 38.4%, 1.1% of stomach, 39.1%, 1.1% of liver, 41.3%, 1.5 % of the pancreas cancers and 30.4%, 0.5% of myeloid leukemias in men and women respectively; 1.03 % of cervical cancers in women. The highest PAF for all TRCs is in Marmara region (81.9%) where the lowest in Eastern and Southeastern Anatolia region (58.0%) in men, and in Mediterranean (10.4%), in Western Black Sea region (0.0%) respectively in women.

**Conclusion:** More than half of all TRCs in Turkey are due to tobacco smoking. Implemented tobacco control programs should be strengthened.

**Keywords:** Population attributable fraction, tobacco related cancers, relative risk

# SO12

# The Determinatıon of Risk Factors for Chronic Kidney Disease with the Associatıon Rule Method

**Gülden Hakverdi**[1], Aslı Suner[1], Timur Köse[1], Mehmet N. Orman[1]

[1]*Ege University Faculty of Medicine Department of Biostatistics and Medical Informatics, İzmir*

g.hakverdi1706@hotmail.com

**Objective:** The prevalence of chronic kidney disease in the world is increasing, and the prevalence of this disease in our country is 15.7%. Since risk factors for the prevention and early diagnosis of the disease are important, the aim of this study was to determine the risk factors of chronic kidney disease by the association rule method.

**Method:** Chronic kidney disease data set was used from UCI database in the study,. In this study, 13 variables such as albumin, sugar, red blood cells, hypertension and anemia were evaluated among the characteristics of 400 individuals. One of the data mining methods, the association rule (apriori) method was used in the study. WEKA 3.9 software was performed in all analyzes.

**Results:** As a result of the analysis, the first cluster containing the most observed parameters was composed of 6 rules, the second set was 9 rules, and the third set consisted of 1 rule. The rules were interpreted according to the size of the support and trust value. In the first rule, there was no leukocyte cluster, no coronary artery disease and lack of bacteria in the blood was found together with 0.97 confidence. In the second rule, the most co-existing variables, lack of leukocyte cluster and lack of bacteria in blood were obtained at a confidence level of 0.97. In the third rule, the most co-existing variables, good nutritional level and lack of bacteria in blood were obtained with confidence of 0.95. In addition, the absence of edema and the lack of coronary artery disease were found together with confidence in 0.93.

**Conclusion:** In the study, it was aimed to reveal the relationships between the parameters in the data set with the association rule. However, no relationship was found between risk factors such as hypertension and diabetes. As a result, association rule for large-scale data can be used to reveal the relationships and save the researchers time.

**Keywords:** Data mining, association rule, chronic kidney diseas

# SO13

# Table of Occupational Exposure in Turkey

**Hatice Gökçe Aktaş**[1]

[1]*Yıldırım Beyazıt University Biostatistics and Medical Informatics, Ankara*

hilalhaticeaktas@gmail.com

**Objective:** With the advancement of industrialization, in almost every line of work, workers are physically or mentally damaged and even deaths are observed due to toxic substances or physical factors. The aim of this study; mapping the diagnosis of occupational illness profession in Turkey and in the diagnostic process is to start as soon as the treatment process without exposure to death.

**Method:** In this study, it resides in Turkey which 182 880 person's gender, employment status, age, occupation, disease diagnosis, lived in city, county, with that institution admitted disease and clinical information randomness principle was filmed with an appropriately Oracle SQL query from the database. For statistical analysis and calculations, IBM SPSS Statistics 21.0 (IBM Corp. Released 2012. IBM SPSS Statistics for Windows, Version 21.0. Armonk, NY: IBM Corp.) was used.

**Results:** Of the individuals in the study, 111575 were female and 71305 were male and the mean age of the individuals was 51.19±18.71, and the minimum age was 15 years. It was observed that the occupational exposure status was statistically significant in terms of age distribution. Disease diagnoses by occupational groups will be analyzed and occupational exposure rates will be calculated.

**Conclusion:** In our country, it has been determined that many individuals have been diagnosed with occupational disease and the quality of life caused by this subject has decreased or even died. In this study, the occupational exposure map will be issued within the scope of big data and a diagnostic map will be created.

**Keywords:** Big data, occupational exposure, occupational disease

# SO14

# Meta-Analysis for Naproxen Bioquivalence Studies

**Emel Doğan Kurtoğlu**[1], Merve Gülşah Ulusoy[1], Onursal Sağlam[1]

[1]*Novagenix Bioanalytical R&D Centre, Ankara*

edogan@novagenix.com

**Objective:** The drug which has the same active ingredient in the same amount and in the same pharmaceutical form with the reference drug and is considered bioequivalent under the different brand name is called generic drug. There are many generic drugs with the same characteristic on the market. Interchangeability between generic drugs is required for switching a patient from one generic drug to another. For this purpose, meta-analysis for average bioequivalence is used between generic drugs based on data obtained from independent bioequivalence studies.

**Method:** Chow and Liu, proposed meta-analysis method to be used in average bioequivalence. Then, Chow and Shao, proposed an alternative method for meta-analysis that relaxes the assumption of having the same inter-subject and intra-subject variances for all studies. This alternative meta-analysis method was applied for Naproxen bioequivalence studies which were analyzed and reported by Novagenix Bio Analytical R&D Centre from 2005 to 2013. The studies which have same reference product name, same sample size and same experimental design were included in the analysis. Seven studies were suitable for these conditions. The 90% confidence intervals for the differences between the means of pharmacokinetic parameters, area under the curve ($AUC_{0\text{-}tlast}$) and maximum plasma concentration ($C_{max}$), were determined for each binary combinations of seven generic drugs.

**Results:** Considering the 90% confidence intervals, 76.2% of the binary combinations for only $C_{max}$ and 66.7% of the binary combinations for only $AUC_{0\text{-}tlast}$ have been concluded as bioequivalent. 47.6% of the binary combinations have been fulfilled the bioequivalence criteria for both $C_{max}$ and $AUC_{0\text{-}tlast}$.

**Conclusion:** Although each generic drug in this study is bioequivalent to the reference drug, approximately 50% of generic drug combinations did not meet the bioequivalence criteria. Therefore, replacement of a generic drug with another can cause concern about the efficacy of the treatment even for the wide therapeutic index drugs like Naproxen.

**Keywords: M**eta-analysis, generic drug, interchangeability

# SO15

# Systematic Review: A Study of the Approaches to Treating Missing Data in the Data Collected with the WHODAS-2.0 Scale

**Duygu Sıddıkoğlu**[1], Beyza Doğanay Erdoğan[1], Derya Gökmen[1], Şehim Kutlay[2]

[1]*Ankara University Faculty of Medicine Department of Biostatistics, Ankara*
[2]*Ankara University Faculty of Medicine Department of Physical Medicine and Rehabilitation, Ankara*

duygu.sddk@gmail.com

**Objective:** The aim of this study was to evaluate whether missing data rates are indicated in the articles which applied WHODAS 2.0 scale (World Health Organization Disability Assessment Schedule 2.0); and to determine if missing data analysis were done and how it was conducted. For this reason, a systematic literature review about missing data analysis was done on articles applying WHODAS 2.0.

**Method:** We conducted a systematic literature review on articles applying WHODAS 2.0 about strategies used for missing data analysis from November 4, 1999, to April 6, 2018. We evaluated whether the rates of missingness is given by domain and item, how the handling missing data strategies proposed in manual for WHODAS 2.0 taken into consideration and other approaches for missing data analysis.

**Results**: A total of 91 abstracts were reviewed in Scopus database 82 articles chosen for full text review and n=32 articles were examined within the scope of inclusion criteria in the final evaluation. Below 30% of articles gave the rate of overall missingness and rates by domain. Only 9.38% of articles gave missing rate for some items that have the highest missing rate. In 50% of articles, the approaches proposed in the WHODAS 2.0 manual (single imputation and multiple imputation) were performed in different ways. The most frequently missing items are in *Life Activities* domain, (D5.8-D5.11, 4 items) that are not answered when patient has no active work/school life and D4.5 (*Sexual activities*) item in domain called *Getting along with people*. By referring to the WHODAS 2.0 manual, n=11(33%) articles defined when respondent has no active work/school life and given responses to the 32 items of scale, the score is used as it is and be comparable to that of the 36-item full version. On the contrary, one study dropped questions about *Household activities* (D5.2-D5.5) in *Life Activities* domain for participants who did not have household responsibilities.

In n=3 (9.38%) articles, no information was given about missing data. In n=7 (21.88%) articles, available data analysis was applied by using full maximum likelihood information. In one of these Rasch analysis was conducted. In n=4 (12.50%) articles, complete case analysis was performed and the respondents with missing data were excluded. In n=6 (18.75%) articles, single imputation approach was applied. n=9 (28.13%) articles applied multiple imputation. According to the manual, one article performed mean imputation by domain when missing rate was below a critical limit otherwise applied listwise deletion. One study used pairwise deletion. Lastly, one study replaced missing answer to item D4.5 on sexuality by *'1=no difficulty'* based on the assumption that nearly all adolescent respondents are at normal sexual developmental status.

**Conclusion:** When analyzing the approaches for missing data analysis among reviewed articles, it is difficult to get a consistent method. In 50% of articles, the approaches proposed in the WHODAS 2.0 manual (single imputation and multiple imputation) were performed in

different ways. In addition to not applicable items related to *work/school life activities* which is mentioned in the manual, a systematic missingness was observed in D4.5 S*exual activities* item.

**Keywords:** WHODAS 2.0, missing data, systematic review

# SO16
# Generalizability and Inter-Rater Reliability of "Intramuscular Injection Pressure Guıdelines" Used in Medical Skills Exams

Yasemin Akşehirli Seyfeli[1], **Ahmet Öztürk**[1], Zeynep Baykan[2], Melis Naçar[2], Elif Deniz Şafak[3]

[1]*Erciyes University Faculty of Medicine Department of Biostatistics, Kayseri*
[2]*Erciyes University Faculty of Medicine Department of Medical Education, Kayseri*
[3]*Erciyes University Faculty of Medicine Department of Family Medicine, Kayseri*

ahmets67@hotmail.com

**Objective:** Scoring consistency and reliability are important points in skill assessments. Many undesirable sources of error in the skill assessment studies can negatively affect the reliability of the scores obtained. In the assessment of skills, reliability is mostly studied by methods based on classical test theory, item response theory and generalizability theory.

The aim of this study is to determine the reliability of the measurement / evaluation results of the ability of intramuscular (IM) injection in the education of first year students in a medical school and to evaluate the scoring reliability by using the generalizability theory in the crossed data by taking into account student, scoring, skill and task interactions.

**Method:** Medical skills examinations in the form of objective structured clinical examinations are recorded at the clinical examination center in our faculty. In the study, the medical skills of the final exam videos were used in the 2015-2016 academic year. The videos of 30 students were randomly selected from the records and re-evaluated by the three faculty members who were teaching the medical skills course separately using the ability to measure intramuscular injection over the video. The results of the faculty members were compared with each other. In order to measure the internal consistency of the observer, the instructor who evaluated the student at the time of the exam was asked to re-evaluate the same 15 students through video. The study was conducted on a single blind. The first scores of the students and the scores of the evaluators were not told to the instructors. The evaluation guideline consists of 17 items and the total of items is scored over 100 points.

The normal distribution of data was tested with Shapiro-Wilk. Kendalls W (Kendallscoefficent of concordans) analysis was performed to test the inter-rater compliance. Intraclasscorrelationcoefficient-ICC analysis was used to test the inter-rater correlation. Correlation (ICC ) analysis was performed to evaluate intraobserver consistency. P <0.05 was considered to be significant.

In our study, 30 individuals were evaluated by seventeen items by three independent raters. In this case, as individuals are the aim of measurement, they are not considered as a facet. The G-theory was modeled by a two-variance welded crossed random effect pattern. In this pattern, experiment is defined as b x m x p, where 'b' the number of individuals, 'm' is the number of ıtems, and 'p' is the number of scorers.

**Results:** There was an agreement between the three faculty members' evaluation scores (p=0.03; Kendalls W=0.115). A statistically significant high agreement was found between the first and second evaluators (p<0.001; ICC=0.829). There was a statistically significant high correlation between the first evaluator and the third evaluator (p<0.001; ICC=0.710) between the second evaluator and the third evaluator (p<0.001; ICC=0.749). The scores of the number

one faculty member during the examination and the scores of the same students were found to be moderate ($p<0.001$; ICC=0.651).

According to G-theory b x m x p pattern; percentage of total variance (25.7%) of variance components of individuals, percentage of total variance of variance components of participants (0.00%), percentage of total variance of variance components of items (3.7%). percentage of total variance explanation of variance components (8.2%), percentage of total variance explanation of variance components of individual and substance common effects (23.8%), percentage of total variance explanation of variance components of scorer and item common effects (0.4% and the total variance of the variance components of the residual variance (individual and item score) was estimated as the percentage of total variance (38.3%). The generalizability coefficient calculated from these variance estimates was 0.84.

ANOVA is used on G-teory b x m x p pattern to compute seven variance variables; there main effects is individuals, items and scorers; theree common effects is individuals-ıtem, individuals-scorers and ıtem- scorers and the end of other effect (b x m x p, e).

**Conclusion:** It was concluded that the guide for the evaluation of the intramuscular injection skill used was generalizable.

**Key words:** Generalizability, reliability, medical students

# SO17

## Nonparametric Analysis of Factorial Designs

**Semiha Ozgul**[1], Timur Kose[1], Aslı Suner[1], Mehmet N. Orman[1]

[1]*Ege University Faculty of Medicine Department of Biostatistics and Medical Informatics, Izmir*

semihaozgul@hotmail.com

**Objective:** The nonparametric analysis of single-factor cross-sectional data is classically performed by Kruskal Wallis method. However, in the case of two or more independent factors, nonparametric methods are not known exactly compared to parametric methods that applied for the same designs. In this study, it was aimed to introduce nonparametric methods used in the analysis of cross-sectional data in the factorial structure and to give information according to the software to which these methods are applicable. Also, one of the methods was presented with the sample application.

**Method:** As an example of two-way factorial design, the differences between the bond strengths of different metal infrastructure materials used in prosthetic tooth treatment with different brands of porcelain were analyzed using the nonparametric method developed by Brunner and Puri. SAS software (Version 9.3; Procedure Mixed; SAS Institute, Cary, NC, USA) was used in all statistical analyzes.

**Results:** According to the results of the analysis, while the effect of different porcelain brands on the bond strength was found statistically significant (p=0.004), the effect of metal production techniques on the bond strength and the interaction of these two factors were not statistically significant (p=0.058 and p=0.713, respectively).

**Conclusion:** It is thought that this study will be useful for introducing the nonparametric methods used in the analysis of experimental designs containing more than one independent factor to the researchers and for planning further studies on this subject.

**Keywords:** Nonparametric method, factorial design, ANOVA

# SO18

# Principle Coordinates Analysis and An Application

**Sıddık Keskin**[1], Birhan Kunter[2], Nurhan Keskin[3]

[1]*Van Yüzüncü Yıl University Faculty of Medicine Department of Biostatistics, Van*
[2]*Van Yüzüncü Yıl University Faculty of Agriculture Department of Horticulture, Van*
[3]*Ankara University Faculty of Agriculture Department of Horticulture, Ankara*

skeskin973@gmail.com

**Objective:** In this study, by giving information about the Principal coordinates analysis which present the similarities among the *p* individuals, variables or objects in reduced dimension graphically by considering the similarity-dissimilarity or distance matrix, an application was performed.

**Method:** Principal coordinates analysis (PCoA) includes the process to show objects in reduced dimension (usually 2 or 3 dimensions). In the analysis, the position of the objects and their similarities are presented visually by changing the coordinate system without a change in the relative position of the objects. While in Principal component analysis, the nxn-dimensional correlation or the covariance matrix is decomposed, in Principal coordinates analysis, nxn-dimensional similarity matrix is decomposed. Thus, in Principal coordinates analysis, the distance matrix is calculated as a centered matrix. This matrix is then decomposed and eigenvalues as well as eigenvectors are obtained. The eigenvectors are standardized by dividing the square root of the respective eigenvalues and used as the principle coordinate axis. Therefore, the analysis method is also called metric multidimensional scaling. As the distance measures between objects, Euclidean and Chi-square are commonly used. As in other ordination methods, the Principle coordinates analysis also seeks new orthogonal dimensions (axes). Each of these dimensions has an eigenvalue, which is related to the accounted variance by the dimensions. The ratio of the corresponding eigenvalue to the sum of eigenvalues indicates the relative importance of each dimension. In this study, mineral (Na, K, Ca, Mn, Fe, Zn, Cu, Mg) contents of 22 grape cultivars grown in Kalecik Viticulture Research Station of Ankara University (Faculty of Agriculture) were used as application materials. Principle coordinates analysis was used to determine the similarities between cultivars in terms of mineral content and to show them in two-dimensional space.

**Results:** In the study, the eigenvalues of the first and second dimensions were found 0.383 and 0.050, respectively. Also, accounted variances by these dimensions were 87.48% and 11.49% respectively. The two dimensions explained 98.97% of the total variation.

**Conclusion:** By using the distance matrix, Principle coordinate analysis, also known as classical scaling or metric multidimensional scaling, is tries to find the graphical configuration of the object's similarities in 2 or 3-dimensional space that best illustrates original distances between objects. Principle coordinates analysis does not require any assumption about the distribution of data and can provide a graphically easy to understand relationship between objects. Thus, it can be concluded that Principle coordinates analysis is advantageous as compared to other multivariate methods.

**Keywords:** Similarity matrix, ordination, configuration, decomposition

# SO19

# Investigation of Causal Structure Using Mediation Analysis

**Nazlı Totik**[1], Sevinç Püren Yücel[1], Zeliha Nazan Alparslan[1]

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

nazlitotik.biostat@gmail.com

**Objective:** In many clinical trials, the purpose of the researchers is not only to understand the causal effects of treatment, but also to understand the process in which the treatment effects the response. Mediation analysis, which has been developed to solve this causal structure, has been used frequently in the recent years. Mediation analysis is based on the examination of situations underlying this effect not taken into consideration in classical methods, to determine the effect of the explanatory variable(s) on the dependent variable. In this context, mediator variables are used to determine the causal structure so that the direct or indirect causal structure is formed as an effect-dependant mechanism. If the effect on the response can better be explained by the presence of mediator variables, taking mediator variables into consideration in addition to the explanatory variable may play an important role in highlighting the underlying hidden structure. In this study, it is aimed to evaluate the effect of considering the causal structure in mediation analysis by evaluating the direct and indirect effects on the response variable.

**Method:** Direct effect on response and mediation analysis based indirect effect on response will be examined with sample datasets. The causal structure will be defined by the mediator variables. Thus, the importance of the use of mediator variables will be indicated.

**Results and Conclusion:** Methods of evaluating the direct effect of explanatory variables on the response variable are frequently found in the literature. Mediation analysis, in which the indirect effect is evaluated, highlights the causal structure. In this context, a contribution may be provided to the understanding of the process affecting the response.

**Keywords:** Mediation analysis, mediator variable, direct effect, indirect effect

# SO20

# Comparison of Two Independent Group Data with Quantitative Data with Three Different Methods

**Nurgül Bulut**[1], Hayriye Ertem Vehid[2], Setenay Öner[3]

[1]*Istanbul Medeniyet University Faculty of Medicine Department of Biostatistics and Medical Informatics, Istanbul*
[2]*Istanbul Bilim University Faculty of Medicine Department of Medical Education and Informatics, Istanbul*
[3]*Eskişehir Osmangazi University Faculty of Medicine Department of Biostatistics, Eskişehir*

nurgul.bulut@medeniyet.edu.tr

**Objective:** Two independent sample groups of quantitative data included Student's t test, Mann Whitney-U test, and Aspin Welch methods. As is known, there are some preconditions in the application of these judging methods, such as the presence / absence of normal distribution or homogeneity of variances. When judicial methods are applied, evaluations that do not comply with these prerequisites lead to erroneous results. This study was conducted to demonstrate that quantitative data for the use of the ineligible trial method is assessed for two independent sample groups, or those methods applied when valid conditions are met, in the same data sets.

**Method:** A 500-hour trial data set was prepared by simulation with n=15, n=20 and n=25 case numbers. 5±2 for the first group and 5±20 for the second group were derived. Differences / similarities were evaluated according to the methods to be applied to these data groups.

**Results:** When three different test results were compared, there were 17 (3.4%) in the sum of n=15 trials, 29 (5.8%) in the sum of n=20 trials and 30 (6%) in the sum of n=25 trials it was observed that the non-significant and incompatible.

**Conclusion:** The result obtained by a single trial method shows that the interpretation is not enough. Therefore, in cases where there is not enough sample size, commenting using 3 different methods will allow clinically more reliable assessments to be made.

**Keywords:** Normal distribution, homogeneity of variance, quantitative data

# SO21

# New Experimental Design Model Based on Mutation: the Umbrella Design

**Yağmur Polat**[1], Mesut Akyol[1], Yavuz Sanisoğlu[1]

[1]*Yıldırım Beyazıt University Faculty of Medicine Department of Biostatistics, Ankara*

ygmur.plat@gmail.com

**Objective:** The objective is this study is to promote the Umbrella Design, which is a clinical trial design that can be designed especially on low prevalence diseases, is quite useful in clinic cancer trials and has effective results.

Drug development and efficacy research is a process of a total of four phases (Phases I – IV). This process is a set of studies that need to be designed carefully and extensively, and it requires experience with extensive knowledge in the field. Phase II and Phase III in cure development process are the phases where the efficacy of the active ingredient on the disease is being examined. As known experimental designs require extensive sampling, researches of many low prevalence diseases, particularly cancer research, cost more time and fund to acquire the satisfactory number for extensive sampling.

One of the problems encountered in cancer research is that patients receive plenty of cures and medicines during chemotherapy and treatment process. For the researchers, this aggravates the determining process of which cure or medicine works better.

**Method:** In the study, the Umbrella Design will be promoted in accordance with information that was obtained from the literature, and will be compared with other designs in the light of the literary information.

**Results:** In the light of the recent cancer researches, new trial designs that can be designed especially on low prevalence diseases, are quite useful in clinic cancer trials and have effective results were designed. The recent cancer researches show that experiment designs based on the genomic variation of the malignant tumour that causes the disease are more efficient than the experiment designs based on the anatomical location where cancer is located.

The Umbrella Design, which is one of the most preferred new trial designs in especially cancer research, is a design that allows the efficacy of the cure on different tumour mutations of the cancer located in a stable anatomical location to be examined. As distinct from other experimental designs, the Umbrella Design allows the researcher to examine the drug efficacy of different drugs simultaneously, and to determine which cure works better.

**Keywords:** Experimental design, clinical cancer trials, phase II clinical studies, the umbrella design

# SO22

# A novel approach for analysing differential cell count of Bronchoalveolar lavage samples: Flow-cytometry

**Cagla Sariturk**[1], Funda Pepedil Tanrikulu[2], Nazan Sen[3], Nurhilal Buyukkurt[4], Ilknur Kozanoglu[5]

[1]*Baskent University Adana Dr Turgut Noyan Research and Training Hospital, Department of Biostatistics, Adana*

[2]*Baskent University Adana Dr Turgut Noyan Research and Training Hospital, Department of İnternal Medicine Division of Hematology, Adana*

[3]*Baskent University Adana Dr Turgut Noyan Research and Training Hospital, Department of Respiratory Medicine, Adana*

[4]*Baskent University Adana Dr Turgut Noyan Research and Training Hospital, Department of Internal Medicine Division of Hematology, Adana*

[5]*Baskent University Adana Dr TurgutNoyan Research and Training Hospital, Department of Physiology, Adana*

caglasariturk@gmail.com

**Objective:** Bronchoalveolar lavage (BAL) is obtaining a sample of fluid containing alveolar cells during the procedure of flexible bronchoscopy. While it aids the diagnosis in many pulmonary diseases, it is also possible to make a certain diagnosis via BAL in some of them. Nowadays, the gold standard method for analysing cells in BAL samples is light microscopy (LM). Flow-cytometry (FCM) is another technique frequently used in the diagnosis of hematologic malignancies, however it is not a standard method for analysing BAL samples. In this technique, cells or particules in the form of a suspension are passed through a compartment illuminated by laser light and the signallings collected from the cells are analysed. In our study, we aimed to show that FCM could also be used instead of LM for analysing differential cellular count of BAL samples.

**Method:** Seventy-six BAL samples accepted to the hematolology laboratory of our center between 2014-2018 were included in our study. All the samples were evaluated via LM by the same hematologist to note the percentages of macrophage, lymphocyte, neutrophile and eosninophile; while FCM analysis were evaluated by another doctor without knowing each other's results.

Considering statistical analysis, categorical measurements were summarised as number and percentage, whereas continuous measurements were summarised as mean and standard deviation (or median and minimun-maximum as needed). Between the results from LM and FCM intraclass correlation coefficient was checked. Linear regression analysis were done to find $R^2$ value for each parameter noted by LM and corresponding percentage noted by FCM. The square of coefficient of multiple correlations is named as multiple $R^2$ and it is expressed as "$R^2$". It shows the degree of explaining dependent variables by independent variables. The value of $R^2$ is between -1 ile +1 and it is desired to be near 1. The kappa value, sensitivity and specificity were calculated for FCM compared to the gold standard method LM, considering the clinically significant cut-off values for each parameter. The package SPSS 24.0 was used for statistical analysis. For all the results statistical significance level were accepted as $<0.05$.

**Results:** We obtained high correlation coefficients for percentages of macrophage, lymphocyte, neutrophile and eosinophile after correlation analysis for each cell type reported by LM and FCM. Analyzing inter class correlation coefficient for macrophage revealed a high correlation between variables (r=0.89);the degree of independent variable explaining dependent variable is

found as $R^2$=0.792. This shows that 79.2% of the result in measurement of macrophage by LM can be explained by the measurement of macrophage by FCM.

Analyzing inter class correlation coefficient for lymphocyte revealed a high correlation between variables (r=0.93); the degree of independent variable explaining dependent variable

is found as $R^2$=0.87. This shows that 87.0% of the result in measurement of lymphocyte by LM can be explained by the measurement of lymphocyte by FCM.

Analyzing inter class correlation coefficient for neutrophile revealed a high correlation between variables (r=0.89); the degree of independent variable explaining dependent variable

is found as $R^2$=0.786. This shows that 78.6% of the result in measurement of neutrophile by LM can be explained by the measurement of neutrophile by FCM.

Analyzing inter class correlation coefficient for eosinophile revealed a high correlation between variables (r=0.78); the degree of independent variable explaining dependent variable

is found as $R^2$=0.615. This shows that 61.5% of the result in measurement of eosiphile by LM can be explained by the measurement of eosinophile by FCM.

In addition, the compatibility between two techniques: LM and FCM were evaluated for determining the pathologically increased cell ratios when normal differential cell counts in healthy individuals were regarded. It was found that for the macrophage the kappa coefficient was 0.82, sensitivity was 100% and specify was 73.3%; for the lymphocytethe kappa coefficient was 0.84, sensitivity was 83.3% and specify was 100%; for the neutrophilethe kappa coefficient was 0.54, sensitivity was 90.6% and specify was 65.9%; for the eosinophilethe kappa coefficient was 0.34, sensitivity was 44.4% and specify was 96.8%.

**Conclusion:** In our study, it has been shown that flow cytometry method for evaluating cell ratios in BAL samples can give similar results with light microscopy method which is accepted as standard method today. Despite its higher cost, FCM allows the examination of much more numbers of cells in a shorter time period. In addition, FCM is less staff dependent and it is possible to decrease the variability between different centers by constituting standardized panels. In the light of our data we believe that, FCM will become an important method for analyzing BAL samples to aid the diagnosis of pulmonary diseases in the future.

**Key-words:** Correlation, sensitivity, specificity

# SO23

## Machine Learning Approaches in Ki-67 Scoring

**Emre Dirican**[1], Zeki Akkuş[2], Erol Kiliç[3], İsmail Yildiz[2], Ömer Satici[2]

[1]*Hatay Mustafa Kemal University Faculty of Medicine Department of Biostatistics, Hatay*
[2]*Dicle University Faculty of Medicine Department of Biostatistics, Diyarbakır*
[3]*Hatay Mustafa Kemal University Faculty of Medicine Department of General Surgery, Hatay*

emredirilbey@gmail.com

**Objective:** The aim of our study is to analyze the classificability of prognostic factors of breast carcinoma patients according to ki-67 scores (≤14% and >14%) by machine learning methods. It has been studied to determine the 14% for ki-67 proliferation might be a threshold value.

**Methods:** Patients treated for breast cancer in our study were retrospectively studied. Radial based support vector machines, random forest and multilayer artificial neural networks are used in machine learning methods for classification.

**Results:** Study was carried out 223 patients. In dataset 74 cases had ki-67 score of ≤14% and 149 cases >14%. The highest classification accuracy for training set was obtained from random forest method 91%. The highest correct classification performance (84%) was obtained with artificial neural networks in test set with models established according to learned information.

**Conclusions:** Threshold classification with high performance of artificial neural networks supports the 14% for ki-67 score as a critical value.

**Key words:** Machine learning, classification, breast cancer, ki-67 proliferation.

# SO24

# Using a Feature Selection Method for Machine Learning Algorithms on GWAS Data

**Ragıp Onur Öztornacı**[1], Bahar Taşdelen[1], Mehmet Tevfik Dorak[2]

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

[2]*Kingston University London School of Life Sciences Pharmacy and Chemistry, UK*

onur.oztornaci@gmail.com

**Objective:** Genome-wide association studies (GWAS) is the most popular methods for investigating to the genetic architecture of human disease. GWAS methodology is used with a large amount of data, in this case sometimes could be problem for researchers in terms of using high thresholds. Genome-wide association studies (GWAS) have long relied on proposed statistical significance thresholds to be able to differentiate true positives from false positives. Although the genome-wide significance *P*-value threshold of $5 \times 10^{-8}$ has become a standard for common-variant GWAS. Therefore, Machine learning approaches could be used instead of classical statistical approaches for prediction. Machine learning approaches doesn't require to use any thresholds for predicting to the case-controls. Nevertheless, computation time is one of the problems for machine learning algorithms. We would like to suggest that the feature selection approach for handling in this problem.

**Method:** We applied machine learning methods on GWAS data frame. We re-analysed an already published GWAS data set for a comparative study. We used case only-study design on 211 sex spesific case-control data. We used case only-study design data which is consist of 211 sex specific case-control data. Case only study design needs to assumption that genotype and sex to be independent in the healthy population, therein lies it could be used for different othozomal chromosom. We followed standart GWAS steps for detecting to the SNPs which were related to disease. We used Chi-square test for feature selection in this way, we were able to use reduced data frame.

**Results:** We obtained 10423 genotype data from the 242323 genotype data and applied machine learning algorithms in the final data set. We would like to say that SVM performance was increased in this method and also, we could avoid of over-fitting problem for machine learning algorithms.

**Conclusion:** Therefore, we can say that it was the advantage for computation time and user-controlled data cleaning for using machine learning. As a result, we can make a good prediction for our case-control study.

**Keywords:** Machine learning, GWAS

# SO25

# Analysis of EEG and Motion Signal Data for Walking Action in a Neurological Disease

**Su Özgür**[1], Turhan Kahraman[2], Mete Eminağaoğlu[3], Ahmet Özkurt[4], Berril Dönmez Çolakoğlu[5], Betül Yürdem[4], Uğur Eliiyi[3], Pembe Keskinoğlu[1], Arzu Genç[6]

[1]*Ege University Faculty of Medicine Department of Biostatistics, İzmir*
[2]*İzmir Katip Çelebi University Faculty of Health Sciences Department of Physiotherapy and Rehabilitation, İzmir*
[3]*Dokuz Eylül University Faculty of Science Department of Computer Science, İzmir*
[4]*Dokuz Eylül University Faculty of Engineering Department of Electrical and Electronics Engineering, İzmir*
[5] *Dokuz Eylül University Faculty of Medicine Department of Neurology, İzmir*
[6]*Dokuz Eylül University School of Physical Therapy and Rehabilitation Department of Neurological Physiotherapy-Rehabilitation, İzmir*

suozgur35@gmail.com

**Objective:** The excitations of the brain and movement system and its responses to these stimuli can be evaluated with EEG and EMG. Very high dimensional data sets occur after the processing of EEG and EMG signal data. Classical approaches are insufficient for analyzing these big matrices. EEG and EMG are frequently used for diagnosis and prognosis in neurodegenerative diseases such as Parkinson. Parkinson's disease affects the nerve cells in the brain that control muscle movements. Although, the freezing of gait episodes are usually transient and short timed, there is an increase in the incidence of the freezing of gait episodes as the disease progresses. There are few computerized methods presented in the literature for the detection of freezing of gait. These are methods of electromyography signals, methods based on goniometer and three-dimensional motion analysis systems and methods based on foot pressure analysis and methods of analyzing motion signals obtained from accelerometers and gyroscopes. There was no study that evaluated the signals coming from the brain together with motion system (EEG) and showing the pathologies while patients are walking. This research is a preliminary study of the project where kinetic motion and EEG signals will be evaluated together. The aim of this study was to determine the pathological stop with using the data which has limited number of repetitive patients.

**Method:** The data which was collected from the EEG sensors with wearable 6 probes during patients walking were evaluated via classification algorithms. The data were obtained from two patients with Parkinson's disease by performing 35 repetitive trials. Subjects were evaluated: walking 15 meters (forth through a doorway) and turning 180 degrees. The involuntary stops during walking are marked by the neurologist in the EEG signal data. An unbalanced structure is formed in the classification since the measurements marked as involuntary stop are less frequent (11.0%). Therefore, performance of the methods were evaluated with Kappa>0.60, the AUC value and Macro Avg.(MA) of true positive and true negative>0.50. The EEG signal data corresponds to a 99501x7 dimensional matrix containing six channels and measurement time. Min-max standardization was applied considering the difference of signal-time data structures and 11 different classification algorithms were performed. Results of the three best performance algorithms are presented. Because of possible correlation between EEG six channel data, the worst performance is the Naive Bayes method.

**Results:** Pathological stagnation in the 35 repetitive trials was determined as 21 times by the neurologist. The best estimation of EEG signals from these stops are rule-based algorithms:

OneR, Ridor and JRIP. The Kappa values of the methods were respectively: 0.999, 0.998 and 0.998, AUC values respectively: 0.999, 0.999 and 0.999, MA respectively; 0.999, 0.999 and 0.998 and high sensitivity (1.000) was determined.

**Conclusion:** Pathological stagnation during walking in patients with Parkinson's disease, were estimated with high sensitivity by rule base classification using EEG signal data. High estimates were also found in other methods. Pathological signals can be detected by processing wearable sensor signals. Subsequent studies should be focused on pre-pathological processes to prevent stops.

**Keywords:** Classification algorithms, data mining, Parkinson's disease, EEG signal data

This research has been produced from the Project which is supported by Dokuz Eylül University BAP:2018.KB.SAG.005

# SO26

# The Multi-Criteria Decision Methods in Precision Medicine-Individual Treatment

**Aslı Boz**[1], Gülşah Seydaoğlu[1], Özge Vural[1]

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

asliboz.stat@gmail.com

**Objective:** Today the mortality rate from cancer is decreasing in cancer patients but unfortunately, the incidence of cancer is increasing. According to World Health Organization, it is estimated that in 2030, 21 million new cancer cases will be occured and 13 million people will die from cancer. In the present day, which is thought to be a deadly disease, cancer which is considered as a complex, multi-factorial disease has become a treatable disease with a large increase in treatment options. Increasing knowledge in tumor biology, more effective and targeted "individual treatment" approach in the field of oncology in the treatment of disease is becoming gradually widespread. In this process, it will be useful to use Multi Criteria Decision Making (MCDM) techniques that clinicians can use to select the most appropriate treatment for the individual. MCDM; determination of objectives, creation of criteria, selection of alternatives, evaluation of alternatives according to criteria, general evaluation and decision, examination of decision and is in the form of turning back. Along with the large number of techniques found in the solution of multi-criteria problems, thanks to the developing technology, researchers, clinicians, administrators and decision-makers who are trying to solve the problem through computer programs suitable for the application of these techniques bring great convenience. In this study; it is aimed to utilize the MCDM methods to determine the selection, classification and ordering of the "individual treatment" options that are becoming increasingly widespread nowadays.

**Method:** Multi-criteria decision making methods will be applied to the data sets obtained from the Oncology department. In this context, the weights of the criteria in the sample datasets were determined using Entropy, CRITIC, Standard Deviation, Statistical Variance Procedure Methods and individual treatment options, risk etc. AHP (Analytic Hierarchical Process), VIKOR, ELECTRE and TOPSIS methods are used for the selection and ranking of the selection. During the analyzes, EXCEL, SPSS and various package programs are used.

**Results:** MCDM methods were applied to the data sets and simulated data obtained from the Oncology department. Age, stage, hematological parameters, genetic characteristics, response to treatment, recurrence, prognosis and risk factors were used. The weights of the criteria were determined by Entropy, CRITIC, Standard Deviation, Statistical Variance Procedure Methods. In the next stage, AHP (Analytic Hierarchical Process), VIKOR, ELECTRE and TOPSIS methods were applied comparatively in the selection and ranking of individual treatment options.

**Conclusion:** It seems that to apply MCDM methods can be able to determine the selection, classification and ranking of individual treatment options which becoming more common nowadays.

**Keywords:** Individual treatment, multi-criteria decision making, oncology

## SO27

# Two New Measures to Evaluate the Accuracy of Screening Tests in Diagnostic Studies Involving Paired Organs: Binocular Sensitivity and Binocular Specificity

**Afra Alkan**[1], Pervin Demir[1], Selcen Yüksel[1]

[1]*Ankara Yildirim Beyazit University Faculty of Medicine Department of Biostatistics, Ankara*

afra.alkan@gmail.com

**Objective:** The performance of a screening test is evaluated by sensitivity (Sen) and specificity (Spe) estimated with respect to a gold standard test (GS). In practice, patients may be subject to the several screening tests or only one test on different occasions. This is the case when the test is reader-based, at least two readers are needed to eliminate the reader bias. The test result obtained in such studies cannot be considered completely independent. Thus, the methods accounting within-subject correlation are required to estimate sensitivity, specificity and their standard errors.

Several methods have been developed to calculate the accuracy measures for the screening tests used in diagnostic studies that involves paired organs (eyes, ears, etc.), which consider the correlation between the fellow organs. Besides, binocular Sen and Spe have been defined for the screening tests, whose positive results are confirmed by a more reliable test. The aim of this study is to introduce the concept and the extended common correlation model (ECCM) used to calculate binocular accuracy measures.

**Method:** Binocular Sen is defined as the probability of at least one correct positive test result in patients with at least one diseased organ. Binocular Spe is the probability of two correct negative test results in patients with two un-diseased organs. ECCM is used to estimate the binocular accuracy measure for the case of exchangeable organs. The data set used in this study is obtained from the study conducted for the glaucoma prevalence. The diagnosis made by clinicians and the intraocular pressure IOP>21 mm-hg are considered as GS and the screening test, respectively. Sen and Spe measures are calculated for only left eye, only right eye, overall considering all eyes are independent, overall considering the intra-pair correlation and binocular approach. Standard errors of estimates are given as well.

**Results:** The overall agreement between IOP and GS is -0.059±0.015. The agreement between pair eyes is moderate for IOP and strong for GS. The (Sen,Spe) are estimated as (0.074±0.009, 0.880±0.006) for the left eye and (0.083±0.009, 0.866±0.007) for the right eye. The overall (Sen, Spe) is (0.078±0.006, 0.873±0.005) when eyes are considered completely independent. The conventional and binocular (Sen, Spe) are estimated from ECCM as (0.077±*, 0.873±*) and (0.104*, 0.820±*), respectively. (*The calculation is under process)

**Conclusion:** We found that the difference between conventional and binocular accuracy measures are small, since the overall correlation between IOP and GS was very week, and intra-pair correlation of IOP was similar to the one of GS. Perara (2013) reports in the simulation study that binocular accuracy measures were mostly less than the conventional accuracy measures except some cases.

**Key words:** Paired organ, binocular sensitivity, binocular specificity

# SO28

# Data Envelopment Analysis Approach for Measuring the Efficiency of University Hospital Units

**Meryem Göral Yıldızlı**[1], Zeliha Nazan Alparslan[1]

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

meryem@cu.edu.tr

**Objective:** Hospitals as the most important components of health institutions, must measure their both medical and financial performance. Data Envelopment Analysis (DEA), which is widely used in the performance measurement of health institutions is a non-parametric efficiency measurement method that can be applied with multiple similar input and output variables. The aim of this study is to evaluate the applicability of the DEA method to the Hospital Information Management System (HIMS) data in the process of determining the efficiency of the departments of Çukurova University Balcalı Hospital.

**Method:** The data envelopment analysis method was applied on the data set obtained from HIMS. Since the difference between clinical units in terms of treatment procedures influences the selection of the variables analyzed, the analysis was performed on two separate groups based on surgical and internal departments. Input variables were: the number of physicians, the number of residents, the number of nurses, the number of beds and expenses. Joint output variables were the number of out-patients, the number of in-patients, bed occupation percentages and income amounts. The number of operations were also outputs for the surgical branches. Efficiency Measurement System (EMS) version 1.3 was used for the analysis.

**Results:** Different analytical procedures applied to the 2017 data showed that there were variations between the efficiency scores of the departments examined in two groups as internal and surgical departments.

**Conclusion:** The efficiency percentages of the departments were determined according to the input and output variables analyzed and indications of the model to increase the performance of the low efficiency departments were evaluated. DEA on the HIMS data can also be used to determine the financial efficiency coefficients of the departments. These coefficients can be a guide in determining the quality efficiency coefficient, which is a component of the additional payment contribution calculations applied in the university hospitals.

**Keywords:** Data envelopment analysis, performance in hospitals, efficiency measurement, EMS

# SO29

# The Frequency of Obesity and Related Factors in Kindergarden, Primary and Middle School Students in Mersin

Serdar Deniz[1], **Mehmet Kıvrak**[2], Sarp Üner[3]

[1]*Malatya Provincial Health Directorate, Malatya*
[2]*Inonu University Faculty of Medicine Biostatistics and Medical Informatics, Malatya.*
[3]*Hacettepe University Institute of Public Health, Ankara.*

mehmetkivrak@tuik.gov.tr

**Objective:** Obesity is an energy metabolism disorder that can cause physical and mental problems caused by excessive fat storage in the body. It can be seen not only in advanced ages, but from very young ages. Many genetic and environmental factors, such as heredity, gender, ethnic origin, sedentary lifestyle and lack of physical activity, are responsible for obesity. With this study, the calculation and classification of Body Mass Index (BMI) values of the children in kindergartens, elementary and secondary schools in the province of Mersin in relation to the classification of very weak, weak, normal weight, overweight, obese, children and children It is intended to be classified by the Random Forest (RF) classifier.

**Method:** The BMI values calculated by the height and weight data of 205.605 children who are studying in kindergartens, elementary schools and secondary schools in Mersin; It is classified according to the criteria accepted by World Health Organization (z-score). As a result of this classification; the incidence of overweight and obese was 34.5 %.The sample size was calculated as 347 by selecting the error margin of 5 %, confidence interval 95 %, overweight and fat expected frequency 33.9 %. A list of students whose BMI is calculated by measuring their height and weight while sampling; BMI is sorted by class and gender. In this study, by using the open source R program, the frequency of obesity and the classification of children and some variables belonging to children were measured by the frequency of obesity in kindergarden, primary and secondary school students.

**Results:** Because the RF algorithm uses a large number of trees and each trait has to be adequately trained to have the chance to appear in several models, the model is divided into 10 times repetitive, 10-split cross validation method (k=10) and training and test data. According to the training data, the overall accuracy of the model is 79.5 %. The overall error rate generated by the estimates of individual trees is 20.5 %. According to the average Gini reduction method, which is an indicator of how important each variable is in classifying data, it is seen that the variable that contributed most to the purity in each node is Weight (21.981032), while the least contributing variable is the total number of children in the household (1.451363). An estimate was made on the test data to see how the model predicted the data it did not see before. According to this estimate, the model gives a high accuracy (96%) ratio in a narrow confidence interval (95% CI: (0.951, 0.978)).

**Conclusion:** As in many areas, RF method is preferred as modeling technique in health field studies. Among the reasons for being preferred in health sciences are reasons such as the ability to work with both categorical and continuous data, to calculate the importance of the variables. In this context, considering the results of the study, it is thought that modeling by using RF algorithm as an alternative method will be successful in similar studies to be performed.

**Keywords**: Ensemle learning, classification and regression trees, random forest trees

# SO30

# Marginal Modeling of Clustered Data with MAR Mechanism

**Betül Dağoğlu Hark**[1], Zeliha Nazan Alparslan[2]

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

betuldagoglu@gmail.com

**Objectives:** Clustered data is defined as the combination of data with groups of specific characteristics. In this data structure, the effects on the outcome variable of both individual level factors cluster level factors need to be examined. GEE is the most commonly known approach for marginal modeling of these effects in clustered data. Frequently however, not all of the planned measurements of the response variable are actually observed, and a missing-data problem arises. A missing-data problem is handled under three mechanisms: missing completely at random (MCAR), missing at random (MAR) and missing not at random (MNAR). The basic form of the GEE approach is valid under the MCAR assumption. If the missing data mechanism is MAR, the marginal model is fitted with WGEE, multiple imputation or DR-GEE approaches. This study aims to evaluate the use of WGEE, multiple imputation and DR-GEE approaches for the MAR mechanism.

**Method:** The dataset of young people who were subjected to school-based and television-based programs to abandon their smoking habits included 1600 seventh grade students at 28 Los Angeles schools. This dataset is dealt with different missing rates in the MAR mechanism and the performance comparisons of GEE, WGEE, multiple imputation and DR-GEE approaches are evaluated according to their standard errors of the estimates.

**Results and Conclusion:** The dataset was first considered under the MAR assumption with a missing rate of 10%. Parameter estimations of two explanatory variables and standard errors of parameter estimations were calculated. In general, the methods developed under the MAR mechanism (WGEE, multiple imputation and DR-GEE) were found to have lower standard errors compared to the GEE approach. Similar results were obtained with different missing rates.

**Keywords:** Clustered data, MAR, GEE, WGEE, multiple imputation, DR-GEE

# SO31

# Comparison of Variable Selection Criteria: A Simulation Study

**Ceren Efe**[1], İlker Ünal[1]

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

cefe@cu.edu.tr

**Objective:** Construction of an eligible regression model necessitates a thorough variable selection. Including too many regressors would increase both the variance of the response variable and the cost of data collection; therefore the number of regressors should be kept at minimum as possible. On the other hand, the more regressors are in the model, the more information about the response is stored in it. Thus; the number of regressors should also be as many as possible. In order to overcome this confliction, a compromise is needed both ways. Researchers, with the help of experience, try to determine the best subset of variables amongst several options. However, most of the time, more than one "best subset" choice is obtained due to the usage of different criteria. At this point, a problem of determination of the proper criteria arises.

In this study, with the purpose of finding a solution to this problem, some commonly used variable selection criteria will be compared by their choice of the best subset model, using simulated data.

**Method:** The criteria in question will be AIC, BIC, SBC, Mallows' Cp, Jp (Final Prediction Error - FPE), and GMSEP (Estimated MSE of Prediction) and their variable selection performances will be discussed under various scenarios (different sample sizes, variances, and such).

**Results and Conclusion:** Comparison studies for these criteria are not commonly encountered in the literature. The method coming closest to the best model will be determined at the end of the study, and the results are expected to be used as guideline for the future studies.

**Keywords**: AIC, BIC, Mallows' Cp, GMSEP variable selection criteria, best subset, SBC

# FULL TEXT PROCEEDINGS

# O17

# A Two-step Approach for Modeling Longitudinal and Survival Data

**Ebru Turgal**[1], Beyza Doğanay Erdoğan[2]

[1]*Hitit University Faculty of Medicine Department of Biostatistics, Çorum*
[2]*Ankara University Faculty of Medicine Department of Biostatistics, Ankara*

ebruturgal@hitit.edu.tr

## 1. INTRODUCTION

The purpose of studies including longitudinal data is usually to examine how the mean response profiles differ among groups as well as the time course of responses. The answers to these questions can be found with the help of various statistical models. Longitudinal data modeling can be examined in 4 main sections: Marginal models [9], Transition models [7], Random effects models [8] and Multilevel (hierarchical) models [18]. Mixed effect models keep both fixed and random effects together. In mixed effect models, fixed effects are taken as population effects, and random effects are taken as individual effects that show the deviation of individuals from the population. Modeling is done by taking into account the relationship between measurements that share the same level of random effect [21]. For longitudinal data, the random intercept model may remain at a simple level. Firstly, it would be wrong to assume that the change in time for all individuals is the same. It is more likely that each individual does not change at the same rate, that is, individuals differ in their trend over time. For these reasons, a mixed effect model which assumes that both the intercept and the slope in time varies between individuals may be more appropriate [5].

### *1.1. Modelling of Survival Data*

#### *1.1.1. Cox Proportional Hazard Regression Model*

The purpose of modeling survival data is to identify the explanatory variables that affect the hazard function and to obtain the hazard function. A proportional hazard model is usually used to model survival data. As suggested by Cox (1972), it is also called the Cox regression model. The proportional hazard assumption implies that the hazard ratio is constant over time or that a group of hazard functions is proportional to the hazard function of the other group [19].

#### *1.1.2. Time-Dependent Explanatory Variables in the Survival Process*

In the Cox regression model, time dependent covariates can also be modeled. If there are time-dependent variables in the model, the Cox regression model can be used but nonproportional hazard arise when the observation period is long. In this case, an extended Cox regression model is used [19].

In addition to the extended Cox regression model, a two-stage approach and joint modeling approaches can be used in cases where long-term observations affect the survival process [14].

### *1.2. Two-Stage Approach*

In the two-stage approach, the parameters of the longitudinal and survival processes are separately estimated. Using the mixed effect model in the first step of the previously proposed two-step approach, the predicted longitudinal response value for each observation time is taken as the explanatory variable in the extended Cox model in the second step. The idea of this approach is to use subject-specific predictions of the true, unobserved biomarker values obtained by using the mixed effect model instead of the observed data, by eliminating the measurement error of the observed longitudinal responses [17].

This study suggests a new two-step approach for evaluating longitudinal and survival data together.

## 2. METHOD

In this study, as a first step of the two-step approach, error-free longitudinal data estimates from the regression tree analysis for longitudinal data are obtained and used as a covariate in the extended Cox regression model in the second step.

### *2.1. First Step: Regression Trees*

In the analysis of longitudinal data, the mixed effect model is often an effective approach to apply, as well as some limitations.

In the linear mixed effect model, the fixed effects are modeled in a simple parametric form. If the number of fixed effects increases, their interactions with each other that are added to the model will also increase and the modeling of this complex structure will be difficult. In addition, when there are a lot of fixed effects, model selection will be difficult. Linear models cannot handle variables that contain missing data as easily as data mining methods. For this reason, the regression tree approach will be considered for the analysis of longitudinal data.

#### *2.1.1. Regression trees for longitudinal data*

A regression tree is a binary tree, where each non-terminal node is split into two nodes based on the values of a single predictor. The regression tree approach can handle high-level interactions between variables. Efforts to generalize regression trees over longitudinal data are often based on multivariate approaches. Repeated observations from a specific person are thought of as multivariate data, and the nodes in the tree are constructed accordingly.

Random effects expectation maximization (RE-EM) trees:

$$y_{it} = Z_{it}b_i + f(x_{it1}, \dots, x_{itK}) + \varepsilon_{it} \quad \text{(Equation 1)}$$

When random effects are known, we construct the regression tree for $y_{it} - Z_{it}b_i$ to estimate $f$ (ie, the coefficients of the function of fixed effects). When we know the coefficients of fixed effect in function f, we estimate $b_i$ random effects with the mixed effect model. Thus, with the regression tree, $\hat{y}_{it}$ estimates corresponding to each measurement time are obtained. Since this method is based on the EM algorithm, the name of the regression tree is also called the random effects expectation maximization.

### *2.2. Second step: Extended Cox Regression*

Estimated longitudinal values in the first step are taken as explanatory variables in the extended Cox model in the second step.

$$h_i(t) = h_0(t)\exp\{\gamma^T w_i + \alpha\hat{y}_i(t)\} \quad \text{(Equation 2)}$$

The *REEMtree* [13] and *nlme* [11] packages in R version 3.4.3 were used to implement the first step of proposed two-stage model. Package *survival* was used for the second step of the model [19]. A p-value of less than 0.05 was considered to be statistically significant.

### *2.3. Application Data*

When acute respiratory failure develops, using *Noninvasive Positive-Pressure Ventilation* (*NPPV*) is the first option after optimum medical treatment. Non-invasive positive pressure mechanical ventilation (NPPV) is mechanical ventilation without any artificial respiratory route. This study includes 58 patients who came at Ankara University Faculty of Medicine, Department of Chest Diseases, Ankara, Turkey. Acute hypercapnic respiratory

failure was diagnosed in intensive care unit in 2013 and 58 patients older than 40 years were followed up.

Random intercept and slope model are used in the longitudinal process. Apache 2 Score (apache), Cat Score (cat) and Respiratory Rate were taken as independent variables, Partial Oxygen Oxygen Fraction Ratio ($PaO_2/FiO_2$) dependent variable. Partial oxygen Oxygen Fraction Ratio ($PaO_2/FiO_2$), Charlson Comorbidity Index (charlson), Glasgow Coma Score (glasgow) and Cat Score (cat) were obtained as explanatory variables in the extended Cox regression model as a dependent variable in the survival process.

## 3. RESULTS

Descriptive statistics of data are shown in Table 1.

**Table 1. Descriptive Statistics**

| | **Patients (n=58)** | |
|---|---|---|
| | **X±SS** | **min-max** |
| **Respiratory Rate** | 22.79±4.551 | 13-40 |
| **Glasgow score** | 14.72 ±0.670 | 12-15 |
| **Cat score** | 30.93±4.426 | 23-40 |
| **Charlson score** | 1.65±1.036 | 1-6 |
| **Apache score** | 17.18±3.505 | 10-27 |
| **0.Hour $PaO_2/FiO_2$** | 210.211±25.980 | 159-260 |
| **1.Hour $PaO_2/FiO_2$** | 218.649±23.197 | 165-260 |
| **2.Hour $PaO_2/FiO_2$** | 223.649±23.776 | 170-265 |
| **4.Hour $PaO_2/FiO_2$** | 227.983±23.804 | 168-267 |
| **6.Hour $PaO_2/FiO_2$** | 230.649±24.081 | 170-264 |
| **12.Hour $PaO_2/FiO_2$** | 234.211±26.015 | 173-300 |
| **24.Hour $PaO_2/FiO_2$** | 237.456±27.606 | 175-320 |
| **48.Hour $PaO_2/FiO_2$** | 240.877±28.191 | 175-320 |
| **72.Hour $PaO_2/FiO_2$** | 243.45±27.430 | 180-320 |
| **96.Hour $PaO_2/FiO_2$** | 245.281±26.687 | 190-320 |
| **120.Hour $PaO_2/FiO_2$** | 246.768±27.386 | 182-320 |

**Table 2. Two-stage and proposed two-stage approach results**

| | | **Two-stage model** | | | | **Proposed two-stage model** | | | | **Proposed two-stage model (AR(1) model)** | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Survival sub model** | | $h(t) = h_0(t)exp(\beta_1 charlson + \beta_2 glasgow + \beta_3 cat + \beta_4 PaO_2\widehat{/FiO_2}(t))$ | | | | $h(t) = h_0(t)exp(\beta_1 charlson + \beta_2 glasgow + \beta_3 cat + \beta_4 PaO_2\widehat{/FiO_2}(t))$ | | | | $h(t) = h_0(t)exp(\beta_1 charlson + \beta_2 glasgow + \beta_3 cat + \beta_4 PaO_2\widehat{/FiO_2}(t))$ | | |
| | | **Parameter estimate** | | | | **Parameter estimate** | | | | **Parameter estimate** | | |
| | | coefficient | standard error | p-value | | coefficient | standard error | p-value | | coefficient | standard error | p-value |
| | $\boldsymbol{\beta_1}$ | 0.089 | 0.109 | 0.419 | $\boldsymbol{\beta_1}$ | 0.015 | 0.094 | 0.872 | $\boldsymbol{\beta_1}$ | 0.028 | 0.093 | 0.763 |
| | $\boldsymbol{\beta_2}$ | -1.573 | 0.268 | <0.001 | $\boldsymbol{\beta_2}$ | -1.384 | 0.260 | <0.001 | $\boldsymbol{\beta_2}$ | -1.219 | 0.244 | <0.001 |
| | $\boldsymbol{\beta_3}$ | 0.239 | 0.069 | <0.001 | $\boldsymbol{\beta_3}$ | 0.334 | 0.061 | <0.001 | $\boldsymbol{\beta_3}$ | 0.315 | 0.060 | <0.001 |
| | $\boldsymbol{\beta_4}$ | -0.042 | 0.011 | <0.001 | $\boldsymbol{\beta_4}$ | -0.035 | 0.010 | <0.001 | $\boldsymbol{\beta_4}$ | -0.045 | 0.011 | <0.001 |

Two-Stage and Proposed Two-Stage Model results are given in Table 2.

The log-likelihood values of the models for the $PaO_2$ / $FiO_2$ variable are:

- RE-EM tree -2logL=-2269.9
- RE- EM tree AR(1) -2logL=-2204.605
- Linear mixed effects model -2logL=-2309.96

## 4. CONCLUSION

In this study, a two-step approach previously proposed by Albert, P. S. and Shih (2010) was re-modeled in modeling longitudinal data and survival data, and a regression tree was used for longitudinal data at the first stage and a new approach was presented. In the light of the foregoing findings, the proposed approach for non-invasive positive pressure mechanical ventilation (NPPV) data has a lower standard error. This result can be supported by further simulations. In addition, regression trees in longitudinal modeling can be preferred because of offering a more flexible solution to the researcher.

## REFERENCES

1. Aktürk, D., Bayramoğlu, Z., & Savran, F. (2012). Sınıflandırma ve Regresyon Ağacı Yönteminin Örnek Veri Seti ile Uygulaması.
2. Albert, P. S., & Shih, J. H. (2010). On estimating the relationship between longitudinal measurements and time-to-event data using a simple two-stage procedure. *Biometrics*, *66*(3), 983-987.
3. Breiman, L. (1984). Friedman; JH; Olshen, RA; Stone, CJ. *Classification and regression trees*.
4. Cox D. (1972). Regression models and life-tables (with discussion). *J Roy Stat Soc B.;* 34: 187–220.
5. Doğanay, B. (2007). Uzunlamasına Çalışmaların Analizinde Karma Etki Modelleri. *Yüksek Lisans Tezi*, Ankara Üniversitesi, Sağlık Bilimleri Enstitüsü, Ankara.
6. Fu, W., & Simonoff, J. S. (2015). Unbiased regression trees for longitudinal and clustered data. *Computational Statistics & Data Analysis*, *88*, 53-74.

7. Kenward, M. G. (2005). Transition Models for Longitudinal Data. *Encyclopedia of Biostatistics*, *8*.
8. Laird, N. M., & Ware, J. H. (1982). Random-effects models for longitudinal data. *Biometrics*, 963-974.
9. Long, J. D., Loeber, R., & Farrington, D. P. (2009). Marginal and random intercepts models for longitudinal binary data with examples from criminology. *Multivariate behavioral research*, *44*(1), 28-58.
10. Özkan, K. (2012). Sınıflandırma ve regresyon ağacı tekniği (SRAT) ile ekolojik verinin modellenmesi. *Turkish Journal of Forestry*, *13*(1), 1-4.
11. Pinheiro J, Bates D, DebRoy S, Sarkar D and R Core Team (2018). *nlme: Linear and Nonlinear Mixed Effects Models*. R package version 3.1-137.
12. R Core Team (2013). R: A language and environment for statistical computing, http://www.r-project.org/.
13. RE-EM Trees: A Data Mining Approach for Longitudinal and Clustered Data: Machine Learning, 2011, vol. 86, pp.169-207.
14. Rizopoulos D. (2016). The R Package Jmbayes For Fitting Joint Models For Longitudinal And Timeto-Event Data Using Mcmc. *Journal Of Statistical Software*, 72(7), 1–45.
15. Sela, R. J., & Simonoff, J. S. (2011). REEMtree: Regression trees with random effects. *R package version 0.90*, *3*, 741-749.
16. Sela, R. J., & Simonoff, J. S. (2009). RE-EM trees: a new data mining approach for longitudinal data. NYU Stern Working Paper SOR-2009-03.
17. Self, S., & Pawitan, Y. (1992). Modeling a marker of disease progression and onset of disease. AIDS epidemiology, 231-255, Birkhäuser Boston.
18. Steele, F. (2008). Multilevel models for longitudinal data. *Journal of the Royal Statistical Society: series A (statistics in society)*, *171*(1), 5-19.
19. Therneau, T. M., & Grambsch, P. M. (2013). *Modeling survival data: extending the Cox model*. Springer Science & Business Media.
20. Verbeke, G. (1997). Linear mixed models for longitudinal data. In *Linear mixed models in practice* (pp. 63-153). Springer, New York, NY.
21. Vittinghoff, E., Glidden, D. V., Shiboski, S. C., & McCulloch, C. E. (2011). *Regression methods in biostatistics: linear, logistic, survival, and repeated measures models*. Springer Science & Business Media.

# O19

# Robust Chart For Monitoring The Contaminated Skew Normal Process Under Ranked Set Sampling

**Derya Karagöz**[1], Nursel Koyuncu[1]

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

deryacal@hacettepe.edu.tr

## 1. INTRODUCTION

The classical Shewhart $\bar{x}$ and R charts are based on the assumption that the distribution of the quality characteristic is normal or approximately normal. In many situations, however, the normality assumption of process population is not valid. For example, the distributions of measurements in chemical processes, semiconductor processes, cutting tool wear processes and observations on lifetimes in accelerated life test samples are often skewed. For skewed populations, Type I risk probabilities grow larger as the skewness increases as a result of variability in population.

Choobineh and Ballard (1987) proposed weighted variance method based on semivariance approximation of Choobineh and Branting (1986). They obtained asymmetric control limits for $\bar{x}$ and R charts for skewed distributions based on the standard deviation of sample means and ranges. Bai and Choi (1995) proposed a simple heuristic method of constructing $\bar{x}$ and R charts using the weighted variance (WV) method.

The most commonly used procedures of statistical quality control (SQC), control charts acceptance sampling plans, are often implemented under the assumption of normal data, which rarely holds in practice. The type of data moderate to strong asymmetry as well as light to heavy tails. This study focus on parametric family of skew-normal distributions introduced by O'Hagan and Leonard (1976) and investigated with more detail by Azzalini and Valle (1996).

Definition 1.1: A random variable X is said to have a location-scale skew-normal distribution, with location λ scale δ and shape parameter α, being denoted X ~ SN (λ, δ2, α) , if its probability density function (pdf) is given by

$$f(x;\lambda,\delta,\alpha) = \frac{2}{\delta}\varphi\left(\frac{x-\lambda}{\delta}\right)\Phi(\alpha\frac{x-\lambda}{\delta})$$

where φ and Φ denote the standard normal probability function and its cumulative distribution function (cdf), respectively. If λ = 0 and δ = 1, we obtain the standard skew-normal distribution, denoted SN (α). Let a random variable $Z = \frac{x-\lambda}{\delta}$ is said to be skew-normal with parameter α, written Z ~ SN (α), if its density function is

$$f(z;\alpha) = 2\varphi(z)\Phi(\alpha z)$$

The aim of this paper is to consider the $\bar{X}$ chart based on Modified Shewhart (MS), Modified Weighted Variance (MWV) and Modified Skewness Correction (MSC) methods are applied to monitor the process variability under the contaminated skew-normal distribution by using ranked set sampling design. Type I risk probabilities of these control charts are compared with respect to different subgroup sizes for contaminated skew-normal distribution.

## 2. METHODS

The robust methods are one of the most commonly used statistical methods then the underlying normality assumption is violated. These methods offer useful and viable alternative to the traditional statistical methods and can provide more accurate results, often yielding greater statistical power and increased sensitivity and yet still be efficient if the normal assumption is correct.

Karagöz (2018) proposed modifications to the Shewhart and weighted variance methods using simple robust estimators to construct skewed and contaminated process. In this section, the control limits of $\bar{X}$ and R control charts for skewed populations under the MS and MWD methods are given. The $\mu_x$, $\mu_R$ and $P_X$ are estimated by using robust estimators. The $\mu_x$ is estimated using the trimmed mean of the subgroup trimmed means $TM_\alpha$ and $\mu_R$ is estimated using the mean of the subgroup interquartile ranges $\overline{IQ}$ R. The control limits are derived by assuming that the parameters of the process are unknown.

The control limits of the $\bar{X}$ chart for MS method are defined as follows:

$$UCL_{\bar{x}\_MS} = \overline{TM}_\alpha + 3\frac{\overline{IQR}}{d_2^r} \text{ and } LCL_{\bar{x}\_MS} = \overline{TM}_\alpha - 3\frac{\overline{IQR}}{d_2^r}$$

where $d_2^r$ is a constant that depends on the subgroup size n and it is calculated when the distribution is skewed.

We consider the mean of the sample trimmed means. Let $X_{i1},...,X_{in}$ represent observations on a variable from ith random sample. We start by ordering the values of $X_{ij}$ from the lowest to highest for each sample and determining the desired amount of trimming, $0 = \alpha < 0.5$ the mean is then calculated for all observations of each samples except the g smallest and largest observations $g = \frac{n\alpha}{2}$, where $\frac{n\alpha}{2}$ is rounded to the nearest integer. The trimmed mean is

$$\overline{TM}_\alpha = \frac{1}{k}\sum_{i=1}^{k}\overline{TM}_{vi} \ ,$$

here TM(vi) denotes the vth ordered value of the sample trimmed means defined by

$$TM_{vi} = \frac{1}{n-2\lceil n\alpha\rceil}\sum_{j=\lceil n\alpha\rceil+1}^{n-\lceil n\alpha\rceil} X_{ij} \ ,$$

where α denotes the percentage of samples to be trimmed, ⌈nα⌉ denotes the ceiling function, i.e., the smallest integer not less than nα. We consider the 20% trimmed mean, which trims the three smallest and the three largest sample trimmed means when k=30.

The mean of the sample interquartile ranges (IQRs) is defined by

$$\overline{IQR}_i = \frac{1}{k}\sum_{i=1}^{k} IQR_i \ ,$$

where $IQR_i$ is the interquartile range of sample i:$IQR_i = Q_{75,i} - Q_{25,i}$; $Q_{r,i}$ is the rth percentile of the values in sample i. The second method investigated is the WV method proposed by Karagöz (2018). The WV method decomposes the skewed distribution into two parts at its mean and both parts are considered symmetric distributions which have the same mean and different standard deviation. In this method, $\mu_X$ and $\mu_R$ are normally estimated using the grand mean of the subgroup means and mean of the subgroup ranges.

The control limits of $\bar{X}$ chart for MWV method are defined as follows:

$$UCL_{\bar{x}\_MWV} = \overline{TM}_{\alpha} + 3\frac{\overline{IQR}}{d_2^{rob}}\sqrt{2\hat{P}_{x\_rob}} \text{ and } LCL_{\bar{x}} = \overline{TM}_{\alpha} - 3\frac{\overline{IQR}}{d_2^{rob}\sqrt{n}}\sqrt{2(1-\hat{P}_{x\_rob})}$$

where $d_2^{rob}$ is the control chart constant for MWV method. $\hat{P}_{X\_rob}$ can be estimated by using the number of observations less than or equal to $\overline{TM}$ as $\hat{P}_{X\_rob} = \frac{\sum\sum \delta\left(\overline{TM} - x_{ij}\right)}{nk}$, here k and n are the number of samples and the number of observations in a subgroup, and $\delta(X) = 1$ for $X \geq 0$ and 0 for $X < 0$.

## 3. SAMPLING DESIGNS

### *3.1. Simple Random Sampling Design*

Let $U = (u_1, u_2, \cdots, u_N)$ denote a finite population of size N. Let Y be the study; X be auxiliary variable associated with each unit $u_j\,(j = 1, \cdots, N)$ of the population. A sample of size n is drawn without replacement from the population. Let $(X_1, Y_1), (X_2, Y_2), \cdots, (X_n, Y_n)$ denote the observed values of X and Y. Moreover let $\mu_y, \mu_x$, $\bar{y}_{SRS} = \frac{\sum_{i=1}^{n} Y_i}{n}$, $\bar{x}_{SRS} = \frac{\sum_{i=1}^{n} X_i}{n}$ be the population and sample means of study and auxiliary variable respectively.

### *3.2. Ranked Set Sampling Design*

Ranked Set Sampling (RSS) is introduced by McIntyre(1952). RSS design can be described as follows:

i. Select a simple random sample of size $n^2$ units from the target population and divide them into n samples each of size n.

ii. Rank the units within each sample in increasing magnitude by using personal judgment, eye inspection or based on a concomitant variable.

iii. Select the ith ranked unit from the ith $(i = 1, 2, \cdots, n)$ sample

iv. Repeat steps (i) through (iii) m times if needed to obtain a RSS of size N=nm.

Let $(X_1, Y_1), (X_2, Y_2), \cdots, (X_n, Y_n)$ be a simple random sample of size n, then the measured ranked set sampling units are denoted by $\left\{\left(X_{(i)j}, Y_{[i]j}\right), i = 1, \cdots, n,\ j = 1, \cdots, m\right\}$ where $\left(X_{(i)j}, Y_{[i]j}\right)$ is the ith ranked unit from the jth cycle of auxiliary variable and study variable respectively, $(\,)$ and $[\,]$ indicate that the ranking of $x$ is perfect and ranking of Y has errors. The sample means of variables can be defined as in RSS:

$$\bar{y}_{RSS} = \frac{1}{nm}\sum_{j=1}^{m}\sum_{i=1}^{n} Y_{[i]j}, \qquad \bar{x}_{RSS} = \frac{1}{nm}\sum_{j=1}^{m}\sum_{i=1}^{n} X_{(i)j}$$

## 4. SIMULATION STUDY

In this section, we consider design schemes for the $\bar{X}$ control chart for contaminated skewed distributed data. We use the the trimmed mean estimators of mean and the interquartile range estimators of the standard deviation for the MS and MWV methods. To evaluate the control chart performance we obtain p for moderate sample size (30 subgroups

of 3-10) for each skewed distribution. The simulation consists of two Phases. The steps of each Phase are described as following.

**Table 1. The parameters of skew-normal distribution**

| $k_3$ | $\Omega$ | $\alpha$ | $\delta^2$ |
|---|---|---|---|
| 0.11 | 0.1 | 3 | 0.047 |
| 0.42 | 0.5 | 10 | 0.99 |
| 0.81 | 0.8 | 20 | 0.79 |
| 0.99 | 0.9 | 100 | 0.89 |

Phase I:

1.a. Generate correlated finite population using bivariate skew-normal distribution with the parameters as given in Table 1 of size N = 10000.

1.b. Select samples size n from bivariate skew-normal distribution for n=3, 5, 7, 10 using SRS and RSS sampling schemes .

1.c. Repeat step 1.b 30 times (k=30).

1.d. By using classic estimators compute the control limits for Shewhart and the WV methods. By using robust estimators compute the control limits for the MS and the MWV methods.

Phase II:

2.a. Generate correlated bivariate skew-normal distribution with the parameters as described in step 1.a.

2.b. Select samples size n from bivariate skew-normal distribution varieties for n=3, 5, 7, 10 using SRS and RSS sampling schemes as in step 1.b.

2.c. Repeat step 2.b 100 times (k=100)

2.d. Compute the sample statistics for $\bar{X}$ chart for the Shewhart and WV methods. Compute the robust estimator interquartile range IQR for the MS and MWV methods.

2.e. Record whether or not the sample statistics calculated in step 2.c are within the control limits of step 1.c. for all methods.

2.f. Repeat steps 1.a through 2.d, 100.000 times and obtain p and ARL values for each method.

In the simulation study, we consider non-contaminated and contaminated data set in Phase I and Phase II. We consider the 20% trimmed mean, which trims the six smallest and the six largest sample trimmed means when k=30.

Contaminated case: The more extreme case of 10% of outliers placed at 50.

The simulation results of p for the $\bar{X}$ control chart for contaminated data under skew-normal distribution are given in Table 2. We can sum up the results from the tables as following: When the distribution is approximately symmetric ($k_3$=0.11, $k_3$=0.42): The MWV mean chart under the RSS design has the best performance. When the skewness of distribution increases ($k_3$=0.81, $k_3$=0.99), the p values of the the p values of the under the RSS design MS and MWV mean charts are more or less similar as seen in Table 2.

**Table 2. The results of p for mean charts based on different sampling designs**

| | $k_3 = 0.11$ | | | | $k_3 = 0.42$ | | | |
|---|---|---|---|---|---|---|---|---|
| | MS | | MWV | | MS | | MWV | |
| n | SRS | RSS | SRS | RSS | SRS | RSS | SRS | RSS |
| 3 | 0.21551 | 0.20190 | 0.21007 | 0.18972 | 0.21668 | 0.20657 | 0.20020 | 0.17758 |
| 5 | 0.08596 | 0.03758 | 0.08609 | 0.03748 | 0.08668 | 0.03788 | 0.08685 | 0.03749 |
| 7 | 0.14033 | 0.09297 | 0.13928 | 0.09297 | 0.14123 | 0.09298 | 0.13755 | 0.09297 |
| 10 | 0.07020 | 0.01486 | 0.07013 | 0.01486 | 0.07053 | 0.01486 | 0.07013 | 0.01486 |
| | $k_3 = 0.81$ | | | | $k_3 = 0.99$ | | | |
| | MS | | MWV | | MS | | MWV | |
| n | SRS | RSS | SRS | RSS | SRS | RSS | SRS | RSS |
| 3 | 0.21703 | 0.20903 | 0.18877 | 0.14386 | 0.21594 | 0.20744 | 0.18651 | 0.13394 |
| 5 | 0.08705 | 0.03818 | 0.08769 | 0.03759 | 0.08668 | 0.03830 | 0.08814 | 0.03752 |
| 7 | 0.14254 | 0.09298 | 0.13534 | 0.09286 | 0.14233 | 0.09298 | 0.13112 | 0.09202 |
| 10 | 0.07110 | 0.01492 | 0.07055 | 0.01488 | 0.07129 | 0.01492 | 0.07053 | 0.01487 |

## 5. CONCLUSION

In this paper, the construction of the mean charts limits based on MS and modified MWV methods under the SRS and RSS designs are considered for contaminated bivariate skew-normal process. We have studied the effect of the skewness on mean charts under skew-normal distributed data for small and large sample sizes by using these two sampling designs. The Monte-Carlo simulation study is run to compare the mean charts under SRS and RSS sampling designs. In the simulation study, the p values of these mean charts are obtained to compare the methods and designs. The results can be summed up as follows: When the distribution is approximately symmetric ($k_3$=0.11, $k_3$=0.42): The MWV mean chart under the RSS design has the best performance. When the skewness of distribution increases ($k_3$=0.81, $k_3$=0.99), the MS and MWV mean charts can be used as alternative to each other.

## REFERENCES

1. Azzalini, A., Valle, A. D. (1996). The multivariate skew-normal distribution. Biometrika, 83(4), 715-726.
2. Bai, D. S., & Choi, I. S. (1995). X and R control charts for skewed populations. Journal of Quality Technology, 27(2), 120-131.
3. Choobineh, F., & Branting, D. (1986). A simple approximation for semivariance. European Journal of Operational Research, 27(3), 364-370.
4. Choobineh, F., Ballard, J. L. (1987). Control-limits of QC Charts for skewed distributions using weighted-variance. IEEE transactions on reliability, 36(4), 473-477.
5. Karagöz, D. (2018). Robust X control chart for monitoring the skewed and contaminated process. Hacettepe Journal of Mathematics and Statistics, 47(1), 223-242.
6. Karagöz, D., Koyuncu, N. (2017). New ranked set sampling schemes for range charts limits under bivariate skewed distributions. Soft Computing, 1-15.

7. Koyuncu, N., Karagöz, D. (2017). Shewhart X-bar ($\bar{X}$) and R control charts using new ranked set sampling schemes, Journal of the Turkish Statistical Association (IJTSA), 10(3), 63-71.
8. Koyuncu, N., Karagöz, D. (2018). New mean charts for bivariate asymmetric distributions using different ranked set sampling designs. Quality Technology & Quantitative Management, 15(5), 602-621.
9. McIntyre, G. A. (1952). A method for unbiased selective sampling, using ranked sets. Australian Journal of Agricultural Research, 3(4), 385-390.
10. O'hagan, A., Leonard, T. (1976). Bayes estimation subject to uncertainty about parameter constraints. Biometrika, 63(1), 201-203.

# O20

# Optimization of Epicardial Fat Thickness Change in Obese Patients with Weight Loss by Bariatric Surgery Using Central Composite and Box-Behnken Design

**Mustafa Agâh Tekindal**[1], Cihan Altin[2], Can Ateş[3], Varlık Erol[4]

[1]*Selçuk University Faculty of Veterinary Medicine Department of Biostatistics, Konya*
[2] *Başkent University Faculty of Medicine Zübeyde Hanım Research and Application Center Cardiology Department, İzmir*
[3] *Van Yüzüncü Yıl University Faculty of Medicine Department of Biostatistics, Van*
[4] *Başkent University Faculty of Medicine Zübeyde Hanım Research and Application Center, Department of Genetic Surgery, İzmir*

matekindal@gmail.com

## 1. INTRODUCTION

Although factorial trials may be applied in all areas of research, biology is particularly common in medical practice. Because biological events are under the influence of multiple factors. Therefore, we need to examine the effects together in any biological event to get closer to the reality. In factorial trials, different levels of multiple factors are studied at the same time and the status of a factor can be addressed at different levels of other factor or factors.

None of the combinations tried may be the best. In other words, the highest yielding combination can be found in or outside the trials. Therefore, a large number of factor combinations are needed in factorial trials. However, it is very expensive to do such trials and in addition, as the number of factors increases, it is difficult to find the homogeneous test material necessary to test all combinations. Therefore, to find the most appropriate combination of factors, statistical methods that do not require the conduct of trials involving all combinations have been developed. These methods basically carry out the first attempt by designing a relatively limited trial area, taking advantage of previous studies or similar trials, and combinations that only determine the points in this area. From the results of this experiment, firstly the point of the highest yielding factor levels is estimated, then the point where the actual optimum point is reached or by using the coefficients of the second degree response surface function of the first trial results (steepest ascent method). The highest optimization is attempted to be found [1].

This study was approved by Baskent University Institutional Review Board and Ethics Committee and supported by Baskent University Research Funding (Approval number: KA16/281). The study data consisted of 40 obese patients who lost weight by bariatric surgery between February 2015 and December 2016. Body Mass Index (BMI), Age and HOMA values were evaluated in 3 categories and 3 levels, and response variable was the change in Epicardial Fat Thickness (ΔEFT).

## 2. METHODS

A first order model should have to be a linear structure. However, the curvature test may reveal the presence of curvature. In this case, second order response surface analysis should be used.

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{11} x_1^2 + \beta_{22} x_2^2 + \beta_{12} x_1 x_2 + \varepsilon \quad (2.1)$$

This model is called the second order response surface model. This trial order has some characteristics [2,3,4].

i) Each factor must have at least 3 levels.

ii) The model shall have at least 1 + 2k + k (k-1) 2 different parameters. As a result, the trial order 1 + 2k + k (k-1) should contain data from 2 different points.

In these experiments, the point where the dependent variable takes its maximum or minimum value is called ini stationary point veya [5]. This point is located at the center of the system shown as ellipses. In some cases, the central point in the center indicates neither the maximum nor the minimum value. In this case, the static point is called the saddle point, and the system is called saddle system. One of the most important points in the method of second order response surfaces are stationary points. 3D graphics (response surface and contour graph) helps to determine these points.

### *2.1. Calculation of constant points*

The determination of the components in the second order response surface method depends on the size of the coefficients given in the regression equation. The steps to calculate the constant points are as follows.

i) A second order response surface model is estimated by the help of the data obtained from the experiment.

ii) For each of the factors included in a model, partial derivatives are taken and equalized to zero.

$$\frac{\partial \hat{Y}}{\partial \chi_1} = \frac{\partial \hat{Y}}{\partial \chi_2} = \ldots = \frac{\partial \hat{Y}}{\partial \chi_j} = 0 \tag{2.2}$$

iii) Equivalent which obtained in step ii. (2.2) solves the equation system. A value will be obtained for each factor. These values are substituted in the model and the predicted variable value is obtained for the stationary points.

It is also possible to obtain constant points with matrices. If the given model is expressed in matrices;

$$\hat{Y} = b_1 + x'b + x'\hat{B}\chi \tag{2.3}$$

At Equation 2.3 $b_1$ shows the model constantb and $\hat{B}$ indicates the expected values of second order model coefficients.

$$x' = \left[x_1, \ldots x_j\right] \tag{2.4}$$

$\hat{B}$ is a symmetric matrix.

$$\hat{B} = \begin{bmatrix} b_{11} & \frac{1}{2}b_{12} & \frac{1}{2}b_{1q} \\ \frac{1}{2}b_{12} & \frac{1}{2}b_{22} & \frac{1}{2}b_{2q} \\ \\ \\ \frac{1}{2}b_{1q} & \frac{1}{2}b_{2q} & b_{qq} \end{bmatrix} \tag{2.5}$$

Constant Points

$$x_S = \frac{1}{2}\hat{\beta}^{-1}b \tag{2.6}$$

Equation is available from 2.6. If we replace the static points in the main equation;

$$\hat{Y}_S = b_0 + x'_S\, b + x'_S\, \hat{B}x_S$$

$$\hat{Y}_S = b_0 + \frac{1}{2}x'_S\, b \tag{2.7}$$

Equation 2.7 will be obtained.

$\hat{Y}_S$ is the predicted value of the response variable from the constant point [4,6,7,8].

### *2.2. Structure of Constant Point (Canonical Analysis)*

When a quadratic equation is found to be sufficient, Canonical analysis is applied to decide about the location and structure of stationary points. The structure of the stationary point determines the marks of the eigenvalues obtained by the matrix $\hat{B}$ . For this, it is possible to write a new equation containing canonical variables.

$$\hat{Y} = \hat{Y}_S + \sum_{j=1}^{k} \lambda_j W_j^2 \tag{2.8}$$

Equation 2.8 shows the eigenvalues to be derived from the $\lambda_1, \lambda_2, ..., \lambda_k$ $\hat{\beta}$ vector, while $W_1, W_2, ..., W_k$ is called canonical variables. It is possible to understand the properties of the constant points obtained with the help of Equation 2.8.

i) If all $\lambda_1, \lambda_2, ..., \lambda_k$ are negative, the static point is showing the maximum,

ii) If all $\lambda_1, \lambda_2, ..., \lambda_k$ are positive, the static point represents the minimum and,.

iii) If the signs of $\lambda_1, \lambda_2, ..., \lambda_k$ eigenvalues are mixed, the static point denotes the saddle point [4, 8, 9].

### *2.3. Central Composite Design*

Central composite trial order (CCD) is one of the most popular methods for creating a second order response level model. The CCD is composed of $2^k$ number of two-level factorial trials, with 2k number of axes or star point. Also nc contains a number of central points. The factors in the model must be at least two-level. The placement of the axis points in the trial layout is given in Table 2.1. The main effects of the second order model and the first-order interaction effects are obtained from the $2^k$ experiment, while the curvature of the system is tested with the help of the center points. Quadratic terms in the model with the help of axis points are estimated [4, 6, 8, 9, 10].

**Table 2.1. Central Composit Design**

| $x_1$ | $x_2$ | | $x_k$ |
|---|---|---|---|
| $-\alpha$ | 0 | | 0 |
| $+\alpha$ | 0 | | 0 |
| 0 | $-\alpha$ | | 0 |
| 0 | $-\alpha$ | | 0 |
| 0 | 0 | | $-\alpha$ |
| 0 | 0 | | $+\alpha$ |

### *2.4. Box-Behnken Design*

These experimental schemes laid out by Box and Behnken in 1980 are an effective method to create model of second order response surfaces. It is a method built on unbalanced block trials. The factors to be included in the model must have at least three levels. Let us try to explain the structure of the experiment with the help of a 3-factor experiment. In the Box-Behnken layout, the value of one of the factors is fixed at the central value and combinations of all the other factors are applied [4,9,11,12,13]. As can be seen in Table 2.2, combinations of all levels of A and B factors were applied at the first level of C factor. The most recent columns of the layout matrix are center point values.

**Table 2.2. Three Factors Box-Behnken Desgin**

| **Order** | **Box-Behnken Design** | | |
|---|---|---|---|
| | **A** | **B** | **C** |
| 1 | -1 | -1 | 0 |
| 2 | 1 | -1 | 0 |
| 3 | -1 | 1 | 0 |
| 4 | 1 | 1 | 0 |
| 5 | -1 | 0 | -1 |
| 6 | 1 | 0 | -1 |
| 7 | -1 | 0 | 1 |
| 8 | 1 | 0 | 1 |
| 9 | 0 | -1 | -1 |
| 10 | 0 | 1 | -1 |
| 11 | 0 | -1 | 1 |
| 12 | 0 | 1 | 1 |
| 13 | 0 | 0 | 0 |
| 14 | 0 | 0 | 0 |
| 15 | 0 | 0 | 0 |

## 3. RESULTS

### *Data set*

The study data consisted of 40 obese patients who lost weight by bariatric surgery between February 2015 and December 2016. Body Mass Index (BMI), Age and HOMA values were evaluated in 3 categories and 3 levels, and the change in Epicardial Fat Thickness (ΔEFT) has been chosen as response variable.. The trial set-up was planned before working and the study data were determined in accordance with the trial order.

**Table 3.1. Data design for three-factor design**

| Age | BMI | HOMA | ΔEFT |
|---|---|---|---|
| 18-35 | 35-45 | 2.5-15 | 2.473 |
| | | 15.1-35 | 1.75 |
| | | 35.1-45 | 2.5 |
| | 45.1-55 | 2.5-15 | 4.5 |
| | | 15.1-35 | 2.5 |
| | | 35.1-45 | 2.14 |
| | 55.1-65 | 2.5-15 | 2 |
| | | 15.1-35 | 2.5 |
| | | 35.1-45 | 2.71 |
| 36-45 | 35-45 | 2.5-15 | 1.72 |
| | | 15.1-35 | 2.5 |
| | | 35.1-45 | 2 |
| | 45.1-55 | 2.5-15 | 2 |
| | | 15.1-35 | 3.5 |
| | | 35.1-45 | 2.3 |
| | 55.1-65 | 2.5-15 | 4 |
| | | 15.1-35 | 2.14 |
| | | 35.1-45 | 3.4 |
| 45-66 | 35-45 | 2.5-15 | 1.83 |
| | | 15.1-35 | 2.21 |
| | | 35.1-45 | 2.37 |
| | 45.1-55 | 2.5-15 | 1.98 |
| | | 15.1-35 | 1 |
| | | 35.1-45 | 2 |
| | 55.1-65 | 2.5-15 | 2 |
| | | 15.1-35 | 2.02 |
| | | 35.1-45 | 2.42 |

First of all, $3^3$ trial designs were determined by CCD design and contour and response graphs were drawn according to the most radical increase or decrease.While planning the CCD design, 3 factors were established as α=1.633, 6 axial points, 4 central points and 2 axial central points.

**Table 3.2. CCD design analysis results**

| | DF | Adj SS | Adj MS | F-Value | P-Value |
|---|---|---|---|---|---|
| Model | 11 | 5.495 | 0.500 | 4.65 | 0.002 |
| Blocks | 2 | 2.921 | 1.461 | 4.90 | 0.012 |
| Linear | 3 | 1.314 | 0.438 | 4.57 | 0.041 |
| Age | 1 | 0.925 | 0.925 | 4.20 | 0.005 |
| BMI | 1 | 0.030 | 0.030 | 6.04 | 0.049 |
| HOMA | 1 | 0.359 | 0.359 | 4.47 | 0.014 |
| Square | 3 | 1.185 | 0.395 | 4.51 | 0.085 |
| Age*Age | 1 | 1.014 | 1.014 | 4.32 | 0.084 |
| BMI*BMI | 1 | 0.053 | 0.053 | 4.07 | 0.001 |
| HOMA*HOMA | 1 | 0.213 | 0.213 | 4.28 | 0.013 |
| 2-Way Interaction | 3 | 0.075 | 0.025 | 5.03 | 0.042 |
| Age*BMI | 1 | 0.029 | 0.029 | 4.04 | 0.041 |
| Age*HOMA | 1 | 0.006 | 0.006 | 2.01 | 0.031 |
| BMI*HOMA | 1 | 0.040 | 0.040 | 4.05 | 0.026 |
| Error | 8 | 6.161 | 0.770 | | |
| Lack-of-Fit | 5 | 6.161 | 1.232 | * | * |
| Pure Error | 3 | 0.000 | 0.000 | | |
| Total | 19 | 11.656 | | | |

When Table 3.2 is examined, the interaction effects of Age and BMI, Age and HOMA and BMI and HOMA variables were significant (p values respectively; 0.041, 0.031 and 0.026 ). According to these results, optimum ΔEFT combinations were determined by drawing contour and response graphs and Model Equation was expressed as 3.1. The $R^2$ value of the model was found as 87.75%.

The model was formed in accordance with the Equation 3.1.

ΔEFT=2.03+ 0.0198 Age+ 0.013 BMI + 0.0078 HOMA - 0.000481 Age*Age - 0.00028 BMI*BMI- 0.000282 HOMA*HOMA + 0.000168 Age*BMI + 0.000055 Age*HOMA + 0.000221 BMI*HOMA (3.1)

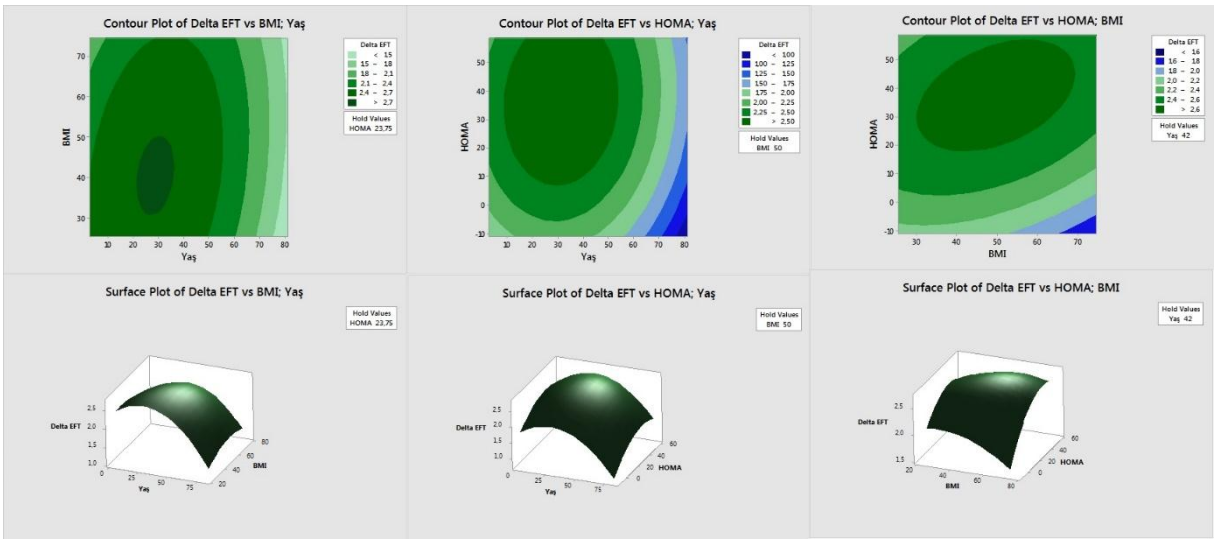


**Figure 3.1. Contour and response surface graphs for CCD design**

As a result of CCD analysis, it was determined that the optimum EFT=2.571 was determined as Age=30.52, BMI=45.30, HOMA=34.62.

BBD design is planned as 3 factors and 3 central points.

**Table 3.3. BBD design analysis results**

| | DF | Adj SS | Adj MS | F-Value | P-Value |
|---|---|---|---|---|---|
| Model | 9 | 7.087 | 0.788 | 4.12 | 0.012 |
| Linear | 3 | 2.351 | 0.784 | 4.10 | 0.027 |
| Age | 1 | 0.053 | 0.053 | 4.21 | 0.045 |
| BMI | 1 | 2.247 | 2.247 | 8.89 | 0.031 |
| HOMA | 1 | 0.050 | 0.050 | 4.20 | 0.075 |
| Square | 3 | 4.291 | 1.430 | 4.66 | 0.046 |
| Age*Age | 1 | 3.736 | 3.736 | 14.78 | 0.012 |
| BMI*BMI | 1 | 0.517 | 0.517 | 4.04 | 0.012 |
| HOMA*HOMA | 1 | 0.442 | 0.442 | 4.75 | 0.043 |
| 2-Way Interaction | 3 | 0.446 | 0.149 | 4.59 | 0.049 |
| Age*BMI | 1 | 0.221 | 0.221 | 4.87 | 0.043 |
| Age*HOMA | 1 | 0.031 | 0.031 | 4.12 | 0.040 |
| BMI*HOMA | 1 | 0.194 | 0.194 | 4.77 | 0.022 |
| Error | 5 | 1.264 | 0.253 | | |
| Lack-of-Fit | 3 | 1.264 | 0.421 | * | * |
| Pure Error | 2 | 0.000 | 0.000 | | |
| Total | 14 | 8.351 | | | |

When Table 3.3 is examined, the interaction effects of Age and BMI, Age and HOMA and BMI and HOMA variables were significant (p values respectively; 0.043, 0.040 and 0.022). In this case, contour and response graphs were drawn, optimum ΔEFT combinations were determined and Model equality was expressed at Equation 3.2.. The $R^2$ value of the model was found to be 91.27%.

ΔEFT =-7.72 + 0.1718 Age + 0.245 BMI + 0.0599 HOMA - 0.001746 Age*Age - 0.00166 BMI*BMI - 0.000766 HOMA*HOMA - 0.000653 Age*BMI + 0.000173 Age*HOMA - 0.000690 BMI*HOMA (3.2)

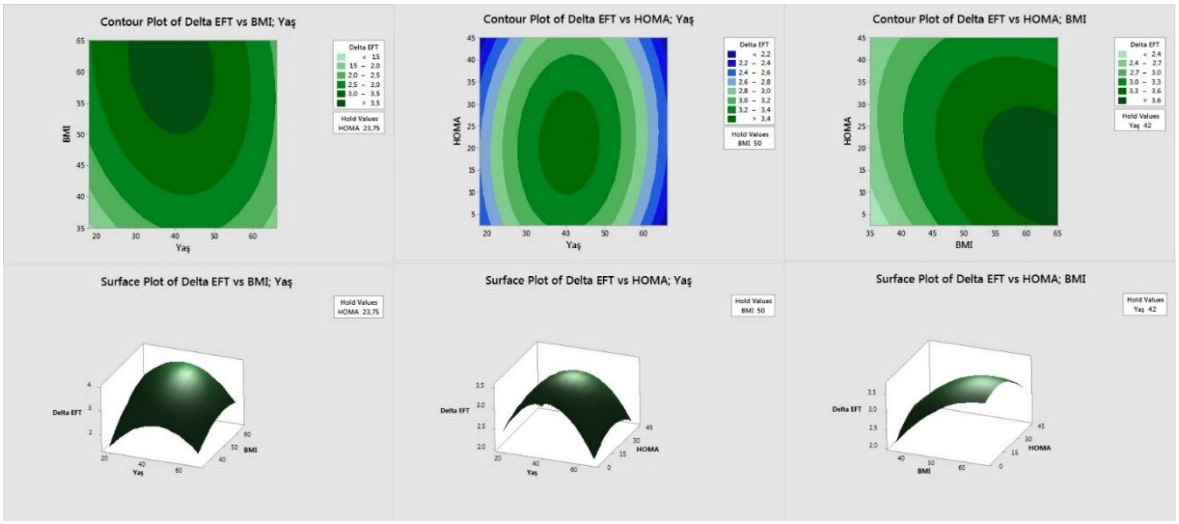


**Figure 3.2. Contour and response surface graphs for BBD experiment design**

As a result of the BBD analysis, when Age=38.36, BMI=63.18, HOMA=14.95, optimum ΔEFT=3.756.

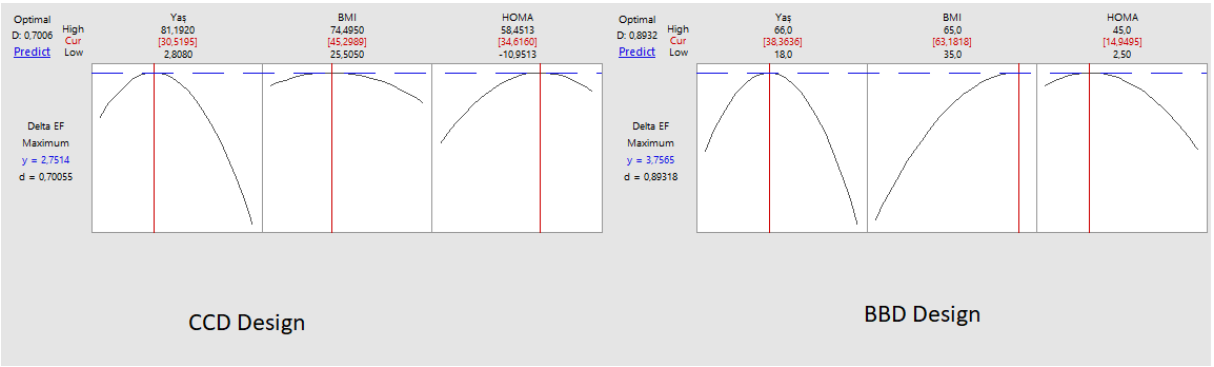


**Figure 3.3. CCD and BBD design optimal point**

As can be seen from Figure 3.3, there are clear optimal differences between CCD and BBD designs.

## 4. DISCUSSION AND CONCLUSION

One of the goals of response surface studies is to determine an appropriate function (or model) to determine the relationship between the response variable and the input variables in order to accurately predict the future values of the response variable. Another is to investigate the largest or the smallest response value depending on the type of the problem and to determine the values of the input variables that can provide this value. Finally, the contribution of a mechanism to an understanding of the mechanism underlying a response system. As a result of CCD analysis, Age=30.52, BMI=45.30, HOMA=34.62, the optimum ΔEFT=2.571. As a result of the BBD

analysis, Age=38.36, BMI=63.18, HOMA=14.95, the optimum, ΔEFT=3.756. OptimumΔ EFT is modeled with Contour and Response surface. According to the results of the analysis, it was found that BBD analysis for optimum ΔEFT was much more positive than CCD and optimum Age, BMI and HOMA combinations were determined to reach maximum ΔEFT.

## REFERENCES

1. Altun, A., "Optimum Verimi Sağlayan Bitki Besin Maddeleri Kombinasyonlarının Etki Yüzey (Response Surface) Metoduyla Saptanması", Doktora Tezi, Ankara Üniversitesi Fen Bilimleri Enstitüsü, Ankara, 1-3, 22-28 (1988).
2. Kul, S., "Cevap Yüzeyi Yöntemleri", Yüksek Lisans Tezi, Mersin Üniversitesi, Mersin, 3-16, 35-48 (2004).
3. Mead, R., Pike D.J., "A Biometrics Invited Paper, A Reviev Of Response Surface Methodology From A Biometric Viewpoint", Biometrics, 31: 803-851 (1975).
4. Doğan, İ. S., Okut, H., "As a Tool Response Surface Methodology (RSM) in a New Product Development", Itafe-03-International Congress on Information Technology in Agriculture, Food and Environment, İzmir (2003).
5. Güvenç, A., Kapucu, N., Kapucu, H., Aydoğan, Ö., Mehmetoğlu, Ü., " Enzymatic Esterification Of İsoamyl Alcohal Obtained From Fusel Oil: Optimization By Response Surface Methodology", Enzyme and Microbial Technology, 40: 778-785 (2006).
6. Schmidt, S., Launsby, R.A., Air Academy Press: Colorado Springs, Understanding Industrial Designed Experiments, 3rd Edition, , CO (1991).
7. Lucas, J.M., "How to Achieve a Robust Process Using Response Surface Methodology," Journal of Quality Technology, 26, 248-260 (1994).
8. Lawson, J.S., Madrigal, J.L.,"Robust Design Through Optimization Techniques," Qualify Engineering, 6: 593-608 (1994).
9. Alexander, M. T., "Response Surface Optimization Using Jmp Software", SAS Conference Proceedings: SAS Users Group International, San Diego, California, 841-847 (1997).
10. Özler, C. "Optimum Toleransların Belirlenmesinde Cevap Yüzeyi Yöntemlerinin Kullanılması Üzerine Bir İnceleme", D.E.Ü.İ.İ.B.F. Dergisi, 21: 71-83 (2006).
11. Myers, R. H., Montgomery D.C., "Response Surface Methodology Process And Product Optimization Using Designed Experiments", John Wiley and Sons, 17-85, 203-303 (2002).
12. Montgomery, D.C., "Design And Analysis Of Experiments", John Wiley and Sons 427-510 (2001).

# O29

# Evaluation of Receiver Operating Characteristic Analysis to Determine the Minimum Clinically Important Difference on Interpretability of Scales

**Selcen Yuksel**[1], Afra Alkan[1], Pervin Demir[1]

[1]*Ankara Yildirim Beyazit University Medical School, Department of Biostatistics, Ankara*

selcenpehlivan@gmail.com

## 1. INTRODUCTION

A scale can be defined as a tool that is constructed to measure a specific feature and has proved psychometric characteristics (Yüksel, 2012). It is important to recognize whether the change in total score of any scale is perceived by patients as the magnitude of, which is an indicator of the scale's responsiveness. The statistical way of revealing whether the perceived change is clinically important, is determining the minimum clinically important difference (MCID). MCID, the mean change score of patients who improves compared to baseline, was first defined as "the smallest difference in score in the domain of interest which patients perceive beneficial" by Jaeschke (Jaeschke,1989). The anchor-based and distribution-based approaches are widely used to calculate MCID among the several statistical methods (Group ISS, 2013; Copay, 2008; Crawford, 2015; Gum, 2013; Bago, 2009; Wells, 2001; Yüksel, 2018). One of the anchor-based approaches is to estimate MCID by the optimal cut-off point for change score between post treatment and baseline scores, resulted from ROC Analysis. The gold standard for ROC analysis is specified by an anchor question addressed after the treatment. This question determines the patient progress during the treatment. "How has spine surgery changed your back status comparing to your baseline status?" can be given as an example. The anchor question generally has response categories between [-7, +7]. In the literature, the patients are divided into two groups as the patients responding at least "-2" and those responding any category less than "-2" and defined as "improved" and "not improved", respectively. The best cut-off point for change score is calculated by considering these groups as gold standard. The aim of the present study is to examine the accuracy of the cut-off point to determine the MCID by a simulation study with 10 000 MCMC repeats in each 36 scenarios. The data were generated by using *irr*, *moments* and *OptimalCutpoints* packages in R Software (R Development Core Team (2008). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.)

## 2. METHOD

### *2.1. Design of the simulation study*

The baseline person parameters are generated according to person parameters and sample sizes as stated at Table 1. The "improved" and "not improved" groups are constructed by adding and subtracting random positive numbers from uniform distribution to/from these parameters, thus the gold standard groups are acquired. Five-point-likert response patterns for baseline, "improved" and "not improved" groups are obtained by using Partial Credit Model with person parameters from the first step, for 20 items with item parameters being in the interval [-7, +7]. The total scores are calculated via response patterns. The mean change score between post-treatment and baseline scores of "improved" groups is set as real MCID ($MCID_R$). $MCID_{ROC}$ is calculated by the cut-off point for change score that best discriminates between the "improved" and "not improved" groups determined by ROC analysis. MCID values from two methods are compared. The sample size and ability level of persons are set as simulation conditions. The data are generated with 10 000 MCMC repeats for each scenario.

**Table 1. Simulation Scenarios**

| | |
|---|---|
| Sample Size | 10; 30; 50; 100; 150; 200; 250; 300; 350; 400; 450; 500 |
| Person parameters | [-7, +7] [-5,+5] [-2,+2] |

## 3. RESULTS

It is clear that the median MCID values obtained by the ROC analysis in Figure 1 are higher than the median MCID values calculated by the average of the differences. From Figure 2 and Figure 3 one can concluded that MCID values are not affected by the person parameters and sample size, respectively. In the same figures, it is seen that the median MCID values obtained by the ROC analysis for each sample size are higher than the median MCID values calculated by the average of the differences.

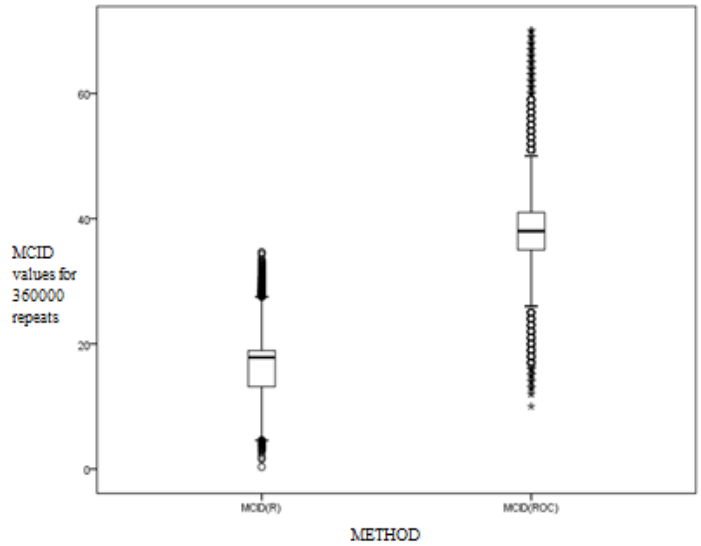


**Figure 1. Summary of results got from 360000 repeats**

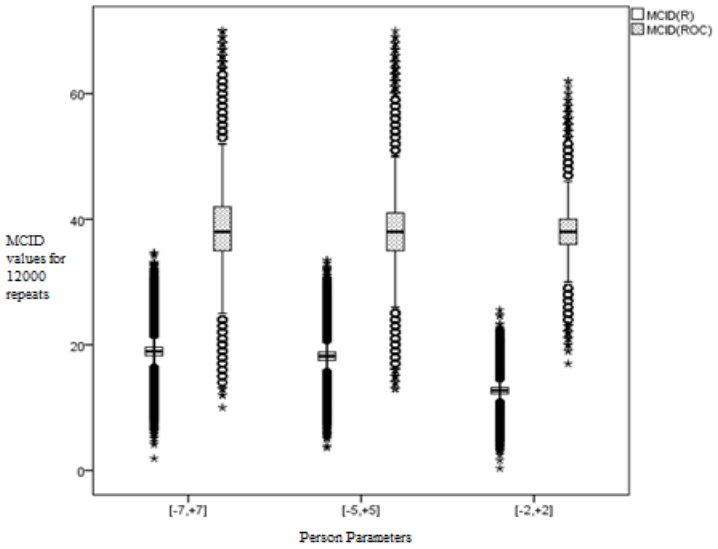


**Figure 2. Summary of results got from 120000 repeat for each person parameter categories**

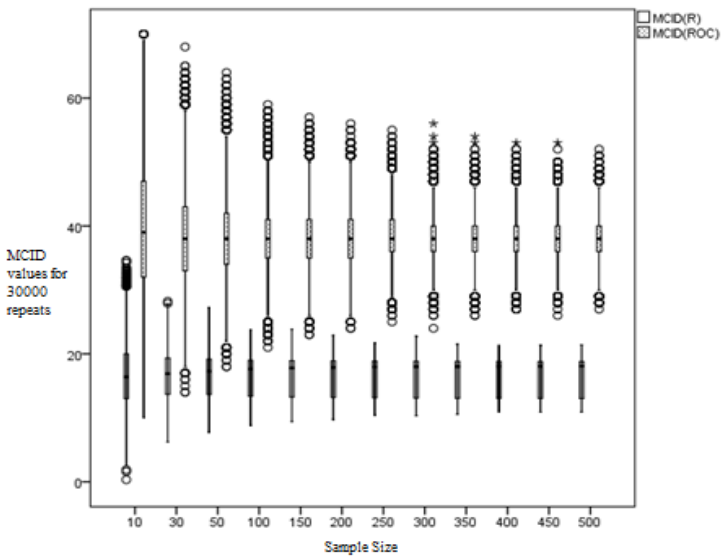


**Figure 3. Summary of results got from 30000 repeat for each sample size**

## 4. CONCLUSION

The aim of this study was to show how far the best cut-off value obtained by the ROC analysis from the real value calculated for the minimum clinically importance of difference in evaluating the interpretability of the scales. For this purpose, data on 36 different simulation scenarios were derived and the results were evaluated. In general, it is observed that the MCID values obtained by the ROC analysis are higher than the real MCID values. When each sample size and person parameter were taken into consideration, the same results were achieved. It was also observed that the result did not change as the sample size increased. In literature, 80% of the scale studies includes the MCID values calculated by the ROC analysis. The fact that the $MCID_{ROC}$ values are higher than the $MCID_R$ is the proof that the ROC analysis will have biased results in the calculations on MCID. The limitation of the study is to include only sample size and person parameters for simulation conditions. There is a need for advanced designs that question the interchangeability of results in different conditions.

## REFERENCES

1. Bago J, Perez-Grueso FJ, Les E, Hernandez P, Pellise F. Minimal important differences of the SRS-22 Patient Questionnaire following surgical treatment of idiopathic scoliosis. *European spine journal : official publication of the European Spine Society, the European Spinal Deformity Society, and the European Section of the Cervical Spine Research Society.* 2009;18(12):1898-1904.
2. Copay AG, Glassman SD, Subach BR, Berven S, Schuler TC, Carreon LY. Minimum clinically important difference in lumbar spine surgery patients: a choice of methods using the Oswestry Disability Index, Medical Outcomes Study questionnaire Short Form 36, and pain scales. *The spine journal : official journal of the North American Spine Society.* 2008;8(6):968-974.
3. Crawford CH, 3rd, Glassman SD, Bridwell KH, Berven SH, Carreon LY. The minimum clinically important difference in SRS-22R total score, appearance, activity and pain domains after surgical treatment of adult spinal deformity. *Spine.* 2015;40(6):377-381.
4. Group ISS, McCarthy I, Bess S, et al. Calculating and Defining Minimally Important Clinical Difference (MCID) and Substantial Clinical Benefit (SCB) Values for Adult Spinal

Deformity (ASD): A Robust Methodology for Consistent Data Reporting. The Spine Journal. 2013;13(9):S75.

5. Gum JL, Glassman SD, Carreon LY. Clinically important deterioration in patients undergoing lumbar spine surgery: a choice of evaluation methods using the Oswestry Disability Index, 36-Item Short Form Health Survey, and pain scales: clinical article. Journal of neurosurgery Spine. 2013;19(5):564-568.
6. Jaeschke R, Singer J, Guyatt GH. Measurement of health status. Ascertaining the minimal clinically important difference. Controlled clinical trials. 1989;10(4):407-415.
7. Wells G, Beaton D, Shea B, et al. Minimal clinically important differences: review of methods. The Journal of rheumatology. 2001;28(2):406-412.
8. Yüksel S, Ayhan S, Nabiyev V, Domingo-Sabat M, Vila-Casademunt A, Obeid I, Perez-Grueso FS, Acaroglu E; European Spine Study Group (ESSG). Minimum clinically important difference of the health-related quality of life scales in adult spinal deformity calculated by latent class analysis: is it appropriate to use the same values for surgical and nonsurgical patients? Spine J. 2018; doi: 10.1016/j.spinee.2018.07.005. [Epub ahead of print]
9. Yüksel,S. Ölçeklerde Saptanan Madde İşlev Farklılığının Karma Rasch Modelleri İle İncelenmesi,Doktora tezi, 2012, Ankara

# O39

# A New Mobile Application for Sample Size Calculations in Medical Research

**Emre Demir**[1], Bedreddin Ali Akça[2], Serhat Hayme[3], Serdal Kenan Köse[3]

[1]*Hitit University Faculty of Medicine Department of Biostatistics, Çorum*
[2]*Hitit University Department of Computer Technology, Osmancık Ömer Derindere Vocational School, Çorum*
[3]*Ankara University Faculty of Medicine Department of Biostatistics, Ankara*

emredemir82@gmail.com

## 1. INTRODUCTION

While most researchers are designing a research, they first answer the question "How many subjects should I take for this study?" Questions such as "With how many experimental animals do you need to work?" or "How many patient polyclinic records do you need to access?" need to be answered in the first phase while planning a study (Ahn, 2014). Accurate determination of sample size, especially in medical research, is one of the first and most important stages of the research. Findings obtained through an inaccurate sample size would be weak, and there might be a need to repeat the study. A study conducted with a smaller sample size may cause low power, while a study conducted with over-sample necessitates additional time and loss of costs.

Sample size is the number of subjects required in order to have certain power to identify clinically significant differences (Lachin, 1977). Identification of sample size is the action of choosing the number of observations to be involved in the statistical sample. Power could briefly be defined as the sensitivity of a study in showing significant differences or relationships in a study. In other words, it is the probability of showing these in the results of the study when a difference or relationship really exists (Chow et al., 2008).

There are four important parameters that determine the sample size for the primary hypothesis in a research (Cohen, 1988);

1) Statistical method
2) Significance level (α)
3) Statistical power (1-β)
4) Effect size

Ideally, the sample size should be large enough to have a high probability of identifying clinically significant differences; if such difference really exists, it should show the statistical significance. If the number of subjects is very high, the clinical study might conclude that a clinically less important effect is statistically significant. On the other hand, if the number of subjects is too little, despite a big clinical difference, the clinical study might not show a statistically significant difference. Choosing an appropriate sample size increases, the probability of identifying a clinically significant effect and makes the study appropriate in terms of both ethics and costs (Ahn 2014).

Scientific research involves two types of errors called Type I (alfa) and Type II (beta) error. α (alfa) is the probability of indicating differences between two groups by mistake while in fact there is none. The alpha value is kept as low as possible in order to bring this case under control, and it is generally taken as 0.05. When the number of subjects in the group is taken more than required, the probability of making Type I error increases. As for B, it is the probability of

indicating no differences while in fact a difference exists. Test power (1- β) is the ability of a test in finding an existing difference; therefore, it is directly related to beta (Chow et al., 2008). The highest value of Type II error is accepted as β=0.20 in literature. Hence, the lowest accepted value of test power is 80% (1 – 0.20=0.80). In other words, with a proportion of 20% at most, the researchers might conclude that there is no difference while in fact there is. Involving insufficient number of subjects in the groups to be compared might cause failure in finding the differences that actually exist, which decreases the test power.

Power analysis which is obtained from the findings after the study is completed identifies the power of the study. However, power of the study could be brought under control by identifying the sample size before the study is conducted, which might prevent the problems of repeating the same study due to low power. Although there are only a few simple mobile applications for sample size calculations in the literature, a free and comprehensive application that researchers can access at any time is not yet available. This study aims to develop a new Android mobile application that calculates sample size for medical researchers.

## 2. MATERIAL AND METHODS

Clinical experiments generally utilize the primary hypothesis of the study for the identification of the sample size. Secondary hypotheses are generally related to other questions regarding the primary hypothesis of the study (Ahn, 2014). Our mobile application enables to obtain sample size for one-sided and two-sided hypotheses. However, formulas for one-sided hypotheses are not given separately here. All the formulas were given for testing two-sided hypotheses. $z_{1-\alpha}$ values should be used in formulas instead of $z_{1-\alpha/2}$ for testing one-sided hypothesis. Normal distribution revisions are performed by multiplying sample size numbers with 1.0472.

The formulas used for calculating the required sample size for the statistical tests included in the mobile application were given in below

### *2.1. One sample t test*

The required sample size for a two sided test for testing $H_0$ versus $H_1$ with a significance level of $\alpha$ and a power of $1-\beta$ is the smallest integer that is larger than or equal to n satisfying the following equation (Ahn 2014);

$H_0: \mu = \mu_0$

$H_1: \mu \neq \mu_0$

$$n = \frac{\sigma^2 * (z_{1-\alpha/2} + z_{1-\beta})^2}{(\mu_0 - \mu)^2} \quad (1)$$

### *2.2. Two independent samples t test*

The required sample size formulas for a two sided test for testing $H_0$ versus $H_1$ with a significance level of $\alpha$ and a power of $1-\beta$ for normally distributed two independent groups with means $\mu_1, \mu_2$ and variances $\sigma_1^2, \sigma_2^2$ was given equation (2) and equation (3) (Ahn 2014).

$H_0: \mu_1 = \mu_2$

$H_1: \mu_1 \neq \mu_2$

Equal sample size;

$$n_1 = n_2 = \frac{(\sigma_1^2 + \sigma_2^2) * \left(z_{1-\alpha/2} + z_{1-\beta}\right)^2}{(\mu_1 - \mu_2)^2} \quad (2)$$

Different sample size;

$$n_1 = \left(1 + \frac{1}{R}\right) * \frac{\left(\frac{\sigma_1^2 + \sigma_2^2}{2}\right) * \left(z_{1-\alpha/2} + z_{1-\beta}\right)^2}{(\mu_1 - \mu_2)^2} \quad (3)$$

$n_2 = R * n_1$

### *2.3. Paired samples t test*

The required sample size for a two sided test for testing $H_0$ versus $H_1$ with a significance level of $\alpha$ and a power of $1 - \beta$ was given the equation (4) (Ahn 2014);

$\mu_d$: mean of difference between the paired samples

$\sigma_d$: standard deviation of difference between the paired samples;

$H_0: \mu_d = 0$

$H_1: \mu_d \neq 0$

$$n = \frac{\sigma_d^2 * (z_{1-\alpha/2} + z_{1-\beta})^2}{(\mu_d)^2} + \frac{(z_{1-\alpha/2})^2}{2} \quad (4)$$

### *2.4. One-Sample Proportion Test*

The required sample size for a two sided test for testing $H_0$ versus $H_1$ with a significance level of α and a power of 1-β was given in the equation (5) (Ahn 2014);

$p$: proportion

$H_0: p = p_0$

$H_1: p \neq p_0$

$$n = \frac{(\, z_{1-\alpha/2} * \sqrt{p_0(1-p_0)} + z_{1-\beta} * \sqrt{p(1-p)}\,)^2}{(p - p_0)^2} \quad (5)$$

### *2.5. Two-Sample Proportion Test*

The required sample size for a two sided test for testing $H_0$ versus $H_1$ with a significance level of α and a power of 1-β was given in the equation (6) (Ahn 2014);

$p_1$: proportion of group 1

$p_2$: proportion of group 2

$H_0: p_1 = p_2$

$H_1: p_1 \neq p_2$

$$n_1 = \frac{(\, z_{1-\alpha/2} + z_{1-\beta})^2}{(p_1 - p_2)^2} * \left(p_1(1-p_1) + \frac{p_2(1-p_2)}{R}\right) \quad (6)$$

$n_2 = R * n_1$

### *2.6. One-way analysis of variance (ANOVA)*

The required sample size for a two sided test for testing $H_0$ versus $H_1$ with a significance level of α and a power of 1-β was given in the equation (7) (Cohen, 1988, Chow, 2008).

$\bar{\mu}$: mean of groups

$\sigma$: standard deviation within a group (same for all groups);

$\lambda$: non-central chi-squared distribution values

$H_0: \mu_i = \mu_j$

$H_1: \mu_i \neq \mu_j, \ \exists i \neq j$

$$\Delta = \frac{1}{\sigma^2}\sum_{i=1}^{k}(\mu_i - \bar{\mu})^2$$

$$n = \frac{\lambda}{\Delta} \quad (7)$$

### *2.7. Correlation analysis*

The required sample size for a two sided test for testing $H_0$ versus $H_1$ (to determine whether a correlation value is different from zero) with a significance level of α and a power of 1-β was given in the equation (8) (Hulley 2007).

$r$: correlation

$H_0: r = 0$

$H_1: r \neq 0$

$$n = \left(\frac{z_{1-\alpha/2} + z_{1-\beta}}{0.5 * \ln\left[\frac{1+r}{1-r}\right]}\right)^2 + 3 \quad (8)$$

The required sample size for a two sided test for testing $H_0$ versus $H_1$ (to determine whether $r_1$ correlation value is different from $r_2$ correlation value) with a significance level of α and a power of 1-β was given in the equation (9) (Hulley 2007).

$r_1$: correlation group 1

$r_2$: correlation group 2

$H_0: r_1 = r_2$

$H_1: r_1 \neq r_2$

$$n_1 = n_2 = \left(\frac{z_{1-\alpha/2}+z_{1-\beta}}{0.5*\ln\left[\frac{1+r_1}{1-r_1}\right] - 0.5*\ln\left[\frac{1+r_2}{1-r_2}\right]}\right)^2 + 3 \quad (9)$$

### *2.8. Chi-square test*

The sample size for chi square can be calculated in two different ways using mobile application. The first way, the sample size can be directly calculated using the effect size (ω=0.1 small, ω=0.3 medium and ω=0.5 large) (Cohen, 1988). In the second way, the sample size can be calculated by finding the effect size using observed and expected values. The required sample size for Chi-square test with a significance level of α and a power of 1-β was given in the following equation (10).

$\omega$: effect size

$\lambda$: non-central chi-squared distribution values

$O$: observed val

$E$: expected values

$$\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$$

$$\omega = \sqrt{\frac{\chi^2}{N}}$$

$$n = \frac{\lambda}{\omega^2} \quad (10)$$

### *2.9. Mobile Application*

Our mobile application is used in devices with Android operating system. The application was produced with Java codes using Android Studio program supported by Google. Minimum SDK version was chosen as 15 so that the application could be used in all mobile devices with Android operating system in the market. Hence, the application works without any problems even in old mobile devices with Android 4.0.3 (Ice Cream Sandwich) operating system.

One of the most important features of our application is that it does not require internet access while working. Once "Sample Size Calculator" application is downloaded from Google Play, the user can easily use the application from mobile devices without needing internet connection. Researchers need to have internet connection only for downloading the application and sending the sample size report through email. The mobile application we developed could be used in Turkish or English, based on the language choice of the mobile device.

Figure 1 shows the screenshot of the mobile application we developed on the mobile phone.


**Figure 1. Mobile application icon**

Figure 2 shows the splash screen of the mobile application according to the selection of the language of use in Turkish and English. The start-up screen of the mobile application consists of 8 menus and an exclamation mark in each menu.


**Figure 2. Mobile application splash screen**

Figure 3 demonstrates the screen shot of sample size calculation section for "one sample *t*-test" in mobile devices. When the exclamation mark in Figure 3 is chosen, sample size will be calculated, and an explanation screen will provide simple and clear instructions that could be understood by researchers from different fields. When the test option is chosen, a sample size calculation screen is opened for calculating the sample size. In this screen, researchers enter a summary of the numerical information required for the test. When they choose the "calculate" option, they see a report screen that shows the required sample size.

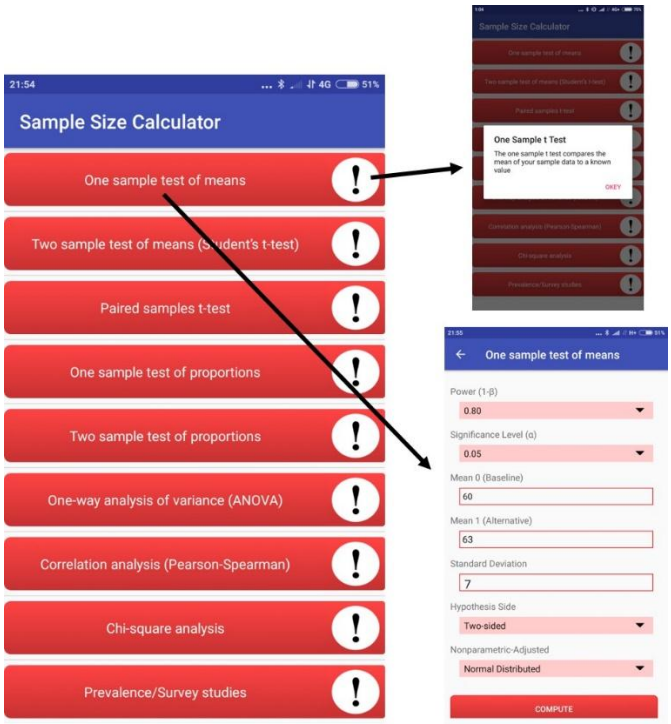


**Figure 3. Example of mobile application usage; sample size calculation section screenshot for one sample t-test**

The report page is shown in Figure 4. In addition to the total sample size to be included in the study, the report also provides summary statements for Ethics committee application or for the report to be used in the article. By choosing the "send an e-mail" option at the bottom of Figure

4, researchers can email the sample size report to themselves directly and use the report for Ethics committee applications once they print it.

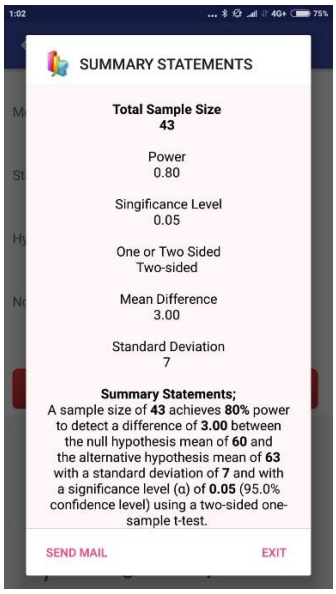


**Figure 4. Mobile application sample size result report**

For international use, the mobile application was designed in English, too. All menus will appear in English for those who already use English in their phones. Figure 5 shows the English screenshot for "one sample *t*-test".

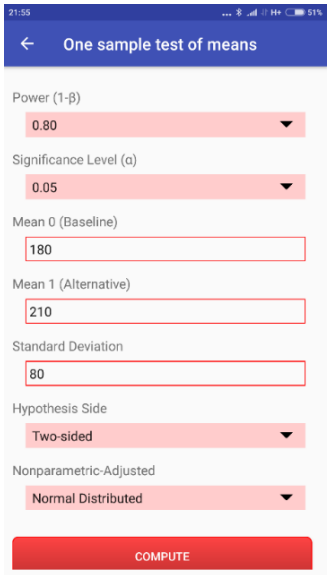


**Figure 5. Example of mobile application usage for English language; sample size calculation section screenshot for one sample t-test**

## 3. RESULTS

With the mobile application that we developed in the application phase of our study and with the G*Power and PASS package programs that are commonly used in literature for calculating sample size and power, sample sizes were calculated for the paired t-test and Chi-square tests whose formulations were given in the second part, and the results are compared.

Sample size analysis results performed with our mobile application showed that totally 16 sample size should be taken in order to reveal a difference 3-unit with endless target population, 0.05 significance level (95% confidence interval) and using two-way paired t test. Screenshots presented in Figure 6. In addition, calculation of the result with the formulation used in obtaining Figure 6 is shown below with equation (11).

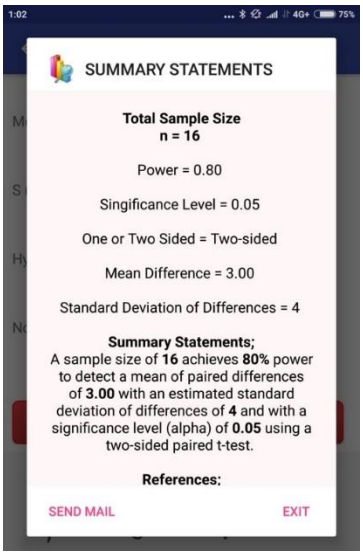


**Figure 6. Mobile application sample size report for paired t-test**

$$n = \frac{4^2(1.960 + 0.842)^2}{3^2} + \frac{1.960^2}{2} \cong 16 \qquad (11)$$

The results of G*Power and PASS and the results of the mobile application were similar. The report outputs of the software are shown in Figure 7.

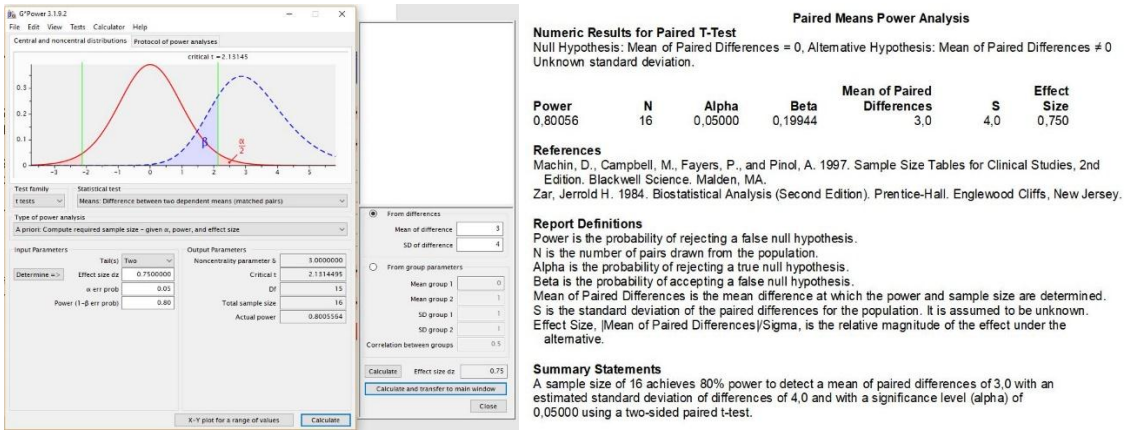


**Figure 7. Sample size report of G*Power and PASS packages for paired t-test**

As a result of sample size analyses performed with our mobile application, the sample size required to detect an effect size of 0.3 with a significance level (alpha) of 0.05 and 80% power using a 1 degree of freedom (df) chi-square test was reported as a total of 88 subjects. Process steps and Report screenshot presented in Figure 8.

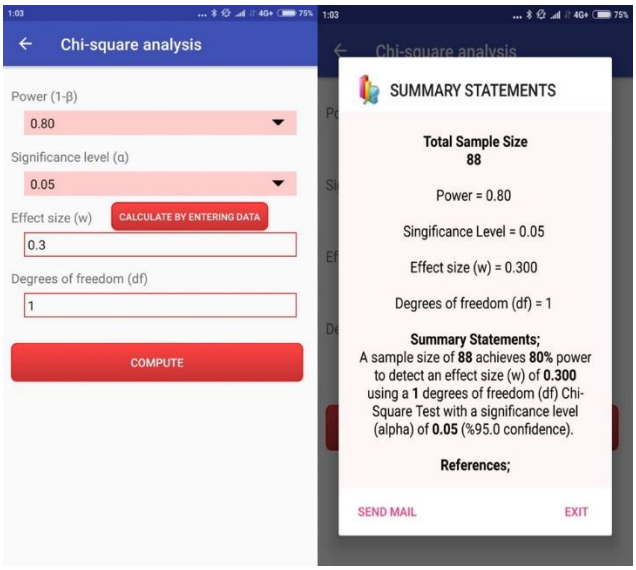


**Figure 8. Mobile application sample size report for chi-square test**

Calculation of the result used for obtaining Figure 8 is shown in equation (12) below. Results are rounded up for sample size.

$$n = \frac{\lambda}{\omega^2} = \frac{7.85}{0.3^2} = \frac{7.85}{0.09} = 87.22 \cong 88 \qquad (12)$$

Non-central chi-square distribution values required for equation (2) were 0.80 and 0.90, with 0.01 power and 0.05 significance level demonstrated in Table 1 (Lachin, 1977).

**Table 1. Non-central chi squared distribution values**

| | $1-\beta = 0.80$ | | $1-\beta = 0.90$ | |
|---|---|---|---|---|
| Degrees of freedom | $\alpha = 0.01$ | $\alpha = 0.05$ | $\alpha = 0.01$ | $\alpha = 0.05$ |
| 1 | 11.68 | **7.85** | 14.88 | 10.51 |
| 2 | 13.89 | 9.64 | 17.43 | 12.66 |
| 3 | 15.46 | 10.91 | 19.25 | 14.18 |
| 4 | 16.75 | 11.94 | 20.74 | 15.41 |
| 5 | 17.87 | 12.83 | 22.03 | 16.47 |
| 6 | 18.88 | 13.63 | 23.19 | 17.42 |
| 7 | 19.79 | 14.36 | 24.24 | 18.29 |
| 8 | 20.64 | 15.03 | 25.22 | 19.09 |
| 9 | 21.43 | 15.65 | 26.13 | 19.83 |
| 10 | 22.18 | 16.25 | 26.99 | 20.54 |

The results of G*Power and PASS and the results of the mobile application were similar. The report outputs of the software are shown in Figure 9.

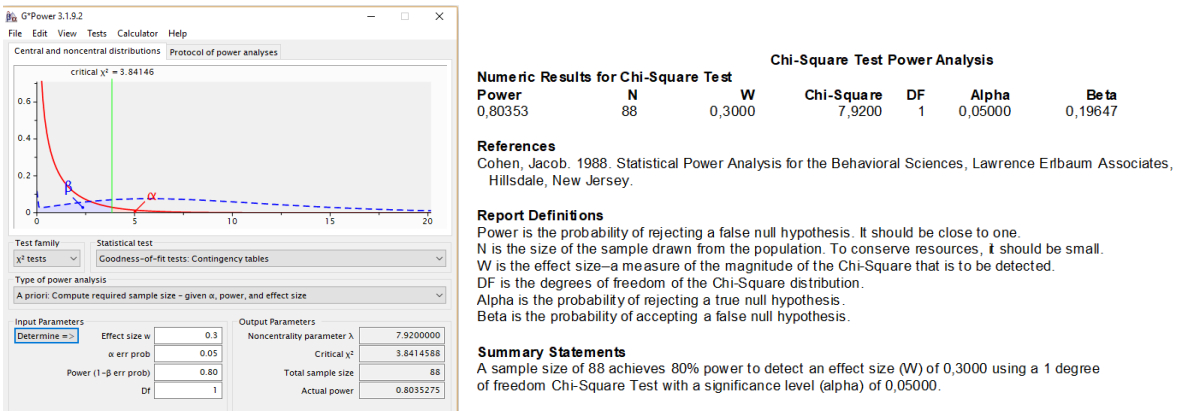


**Figure 9. Sample size report of G*Power and PASS packages for chi-square test**

Effect size in our application is directly calculated with the information entered by the researcher according to related literature, pilot study, or expert opinion. Figure 10 demonstrates the steps of the report and procedures obtained for chi-square test.

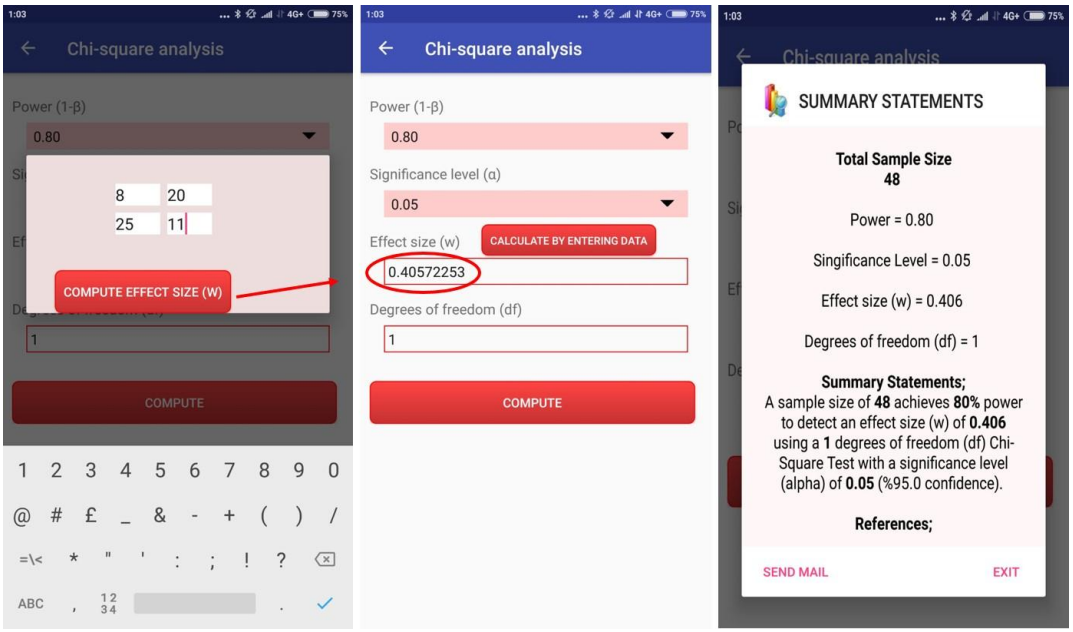


**Figure 10. Effect size calculation and sample size report for chi-square test**

## 4. DISCUSSION

The literature involves a number of package programs such as G*Power, PASS, STATA, and R used for the prediction of sample size. Apart from the package programs, there are also some online web sites (http://powerandsamplesize.com/Calculators/). Review of the related literature showed that there were only two articles related to the mobile applications that can calculate sample size. A mobile application called "GLIMMPSE Lite" developed by Munjal et al. (2014) was developed for one-way ANOVA test, no sample calculation can be performed for other statistical tests. Studies developed by Ngamjarus et al. (2016) could not be found in "n4Studies" in Google play store. As a result of our searches from Play Store with "power analysis" and "sample size" keywords, mobile applications called "SampSize", "Simple Sample Size", "Sample Size", "Statistics and Sample Size", "Sample Size Calculator", "Sample Size Calculators", and "Snap Sample Size Calculator" were found. The application called "Sample Size Calculator" developed by us is more comprehensive than all of these applications, and it works faster. In addition, the result report cannot be obtained through email in other applications. In the application we developed, formation of a result report for Ethics committee

applications is an important superiority. National literature review includes only one mobile application. The application which is called "Güç analizi" includes only two options that compare the means and ratios of two groups. Our sample size application is much more comprehensive in comparison to power analysis application. The application is still being improved to make it an internationally popular program.

## 5. CONCLUSION

Accurate determination of sample size, especially in medical research, is one of the first and most important stages of the research. An accurately calculated sample size enables to make the study have sufficient power, to avoid wastefulness, and to prevent patients from being exposed to unnecessary experimental treatments. With this study, we developed a new mobile application that could be downloaded from Google Play Store and used by researchers both in our country and from other countries. Sample size result report to be obtained from this mobile application is expected to be accepted in ethics committees. All researchers in our country will be able to easily calculate the sample size to use in their studies with this application and they will be able to use the report they receive from the mobile application in the ethics committee applications.

## REFERENCES

1. Ahn C., Heo M., Zhang S., (2014) Sample Size Calculations for Clustered and Longitudinal Outcomes in Clinical Research, New York: Chapman & Hall/CRC; 260 p.
2. Chow SC., Shao J., Wang H., (2008) Sample Size Calculations in Clinical Research (2$^{nd}$ edition), New York: Chapman & Hall/CRC; 449 p.
3. Cohen, J. (1988). Statistical Power Analysis for the Behavioral Sciences (2nd ed.). Hillsdale, NJ Lawrence Erlbaum Associates, Publishers.
4. Hulley, SB., Cummings, SR. Browner, WS. Grady, DG. and Newman TB. (2007). Designing clinical research, 3rd ed. Philadelphia: Wolters Kluwer Health/Lippincott Williams & Wilkins.
5. Lachin, J.M.. (1977). Sample Size Determinations for r x c Comparative Trials, Biometrics 33(2): 315-324.
6. Munjal A, Sakhadeo UR, Muller KE, Glueck DH, Kreidler SM (2014) GLIMMPSE Lite: Calculating Power and Sample Size on Smartphone Devices. PLoS ONE, 9(12): e102082.
7. Ngamjarus, C., Chongsuvivatwong V., McNeil E., (2016), n4Studies: Sample Size Calculation for an Epidemiological Study on a Smart Device. Siriraj Medical Journal, 68 (3); 160-170.

# O40

# Comparison of Fast Regression Methods Used in Model Selection in High Dimensional Data

**Sengul Cangur**[1], Handan Ankaralı[2]

[1]*Duzce University Medical Faculty Department of Biostatistics and Medical Informatics, Duzce*
[2] *İstanbul Medeniyet University Medical Faculty Department of Biostatistics and Medical Informatics, Istanbul*

sengulcangur@gmail.com

## 1. INTRODUCTION

In a study involving a large number of variables, it is very important to get the most accurate predictive statistical model quickly and easily. In particular, algorithms are used to achieve accurate and interpretable models in high dimensional data where there are multicollinearity problems. Algorithms are proposed to find solutions to problems such as multicollinearity, overfitting and outlier which are frequently encountered in application and modeling studies, and they are still being developed.

Dimensional Reduction of Correlation Matrix (DRCM), Nonparametric Dimensional Reduction of Correlation Matrix (DRCM-N), Variance Inflation Factor (VIF) Regression and Robust Variance Inflation Factor (R.VIF) Regression methods were introduced in this study. In a simulation study constructed with different scenarios, it was also evaluated in which situations the approaches are advantageous.

## 2. METHODS

### *2.1. Variance Inflation Factor (VIF) Regression Method*

VIF regression is a method developed by streamwise regression approach using α-investing rule. This method was developed on the basis of the sparsity assumption ($k \ll p$) and controls the marginal False Discovery Rate (mFDR) (Foster & Stine, 2008). VIF regression was developed since the stepwise regression remains unresolved to the multicollinearity problem (Lin, Foster & Ungar, 2011).

### *2.2. Robust Variance Inflation Factor (R.VIF) Regression Method*

This method is a robust approach developed by Dupuis & Victoria-Feser (2013) in response to the disadvantage that the classical VIF regression method was affected by the outlier in the data set. It has all the features of the classic VIF regression method. This algorithm is very robust against small model deviations.

### *2.3. Dimensional Reduction of Correlation Matrix (DRCM) Method*

DRCM method was proposed by Midi & Uraibi (2014). This approach consists of two steps. In the first step, an attempt to reduce the dimension of the correlation matrix by including variables with absolute correlations greater than the threshold determined. In the second step, p values are calculated for the parameter estimates of the potential model.

### *2.4. Nonparametric Dimensional Reduction of Correlation Matrix (DRCM-N) Method*

The DRCM-N method is a nonparametric version of the DRCM method. In this method, the spearman correlation matrix is used as the correlation matrix and all steps in the DRCM

approach are calculated in a similar way. The use of DRCM-N method as fast regression estimator has been first proposed in this study.

The data set was configured to include "normal" and "5% outliers" in order to examine the effect of outliers in the data set. In order to examine the effect of multicollinearity, the correlations between the target variables in the data set were determined as ($corr(L_i, L_j) = \theta, i \neq j, i, j = 1, \dots, k$) $\theta_1 = 0.1$ ($R^2 = 0.20$) and $\theta_2 = 0.85$ ($R^2 = 0.80$). All of the variables in the predictive linear model were derived from the multivariate normal distribution. In each scenario, the number of target independent variables was taken as constant, 5. The noise variables were determined as noise variables which are correlated with target variables and uncorrelated with the target variables. A total of (2x2x6) 24 different scenarios were created by taking two different data types, including normal data set and the data set with 5% outliers, with 50, 100, 250, 500, 750 and 1000 total number of independent variables. The sample size is 5000 and each scenario was repeated 100 times. A total of 9600 models were examined. The root mean square error (RMSE) values obtained with the four methods in each condition were recorded. The simulation was performed in the MATLAB program and by the computer with Intel(R) Core(TM) i7-6500U CPU @ 2.50 GHz, 2592 Mhz, 2 core(s), 4 logical processor(s).

## 3. RESULTS

In the presence of multicollinearity and outliers, the performance values of DRCM, DRCM-N, VIF regression and R.VIF regression methods in case of increase in the number of independent variables considered being included in the model are presented in Table 1. In all scenarios of the simulation study, it was observed that all target variables were selected to final model with four methods. The methods that arrive to the final model within the shortest time are DRCM and DRCM-N, respectively. The time to reach the final model of R.VIF regression method was shorter nearly two times than the time obtained by VIF regression. In all scenarios, the total noise variables to be included in the model were selected by DRCM and DRCM-N methods at most. While RMSE values obtained with four methods were similar within each scenario, it was observed that RMSE values obtained with each method are higher in the data set with outliers than the normal data set, and they tend to decrease with increasing number of variables. It was found that about 2.3% of the total noise variables were selected to the final model with the DRCM and DRCM-N methods when the number of variables was 500 and above. In all scenarios, R.VIF regression method did not include any of the noise variables uncorrelated with target variables in the final model. The times to reach the final model of each method in the data set with outliers were longer than the times obtained with each method in the normal data set. This becomes more evident along with the increase in the number of variables. When the number of variables was above 750 and theta value was 0.85 in both data sets, it was observed that 20% of the noise variables correlated with target variables were included in the final model by the DRCM and DRCM-N methods and 10% of them were included in the final model by the VIF regression method, but none of these variables were included in the final model by the R.VIF regression method.

## 4. DISCUSSION AND CONCLUSION

It has become compulsory to use fast regression algorithms due to the inadequacy of the traditional methods when there are large data sets and outliers in the data set, and the loss of information in the results obtained with these methods. It has been observed that there are a limited number of studies on fast regression methods in the literature and that studies are still going on. In the study of Dupuis & Victoria-Feser (2013) which is one of the studies carried out in this regard in the literature, they proposed to use robust VIF regression method instead of classical VIF regression to achieve fast estimation when the study is performed with large data sets and there are outliers in the data set. Another simulation study was carried out by Midi

& Urabi (2014). In their simulation study, Midi & Urabi (2014) compared Adaptive Lasso, VIF and DRCM methods and found that the DRCM method showed better performance. In this study, the performances of the DRCM, DRCM-N, R.VIF regression and VIF regression methods used in the high dimensional data set were examined in different scenarios. The use of DRCM-N method as fast regression estimator has been first proposed in this study. It was determined that the DRCM and DRCM-N were the fastest methods that reached the final model in all scenarios. It was observed that the number of noise variables correlated with target variable and the number of total noise variables uncorrelated with the target variable that were selected to final model with these two methods were more than those obtained by R.VIF and VIF regression methods. As a result, it was concluded that the R.VIF regression method showed better performance in both the case of normal data and data with outliers.

We propose that predictive models for large data accumulating rapidly with current technology are used as traditional methods and R.VIF regression method should be preferred as fast regression estimator especially in the presence of multicollinearity and outliers in the data set.

## REFERENCES

1. Dupuis, D. J., & Victoria-Feser, M. P. (2003). Robust VIF regression with application to variable selection in large data sets. *The Annals of Applied Statistics, 7(1),* 319-341.
2. Foster, D. P., & Stine, R. A. (2008). Alpha-investing: a procedure for sequential control of expected false discoveries. *Journal of the Royal Statistical Society Series B, 70,* 429-444.
3. Lin, D., Foster, D. P., & Ungar, L. H. (2011). VIF regression: a fast regression algorithm for large data. *Journal of the American Statistical Association, 106,* 232-247.
4. Midi, H., & Uraibi, H. S. (2014). The dimensional reduction of correlation matrix for linear regression model selection. *Mathematical and Computational Methods in Science and Engineering Series, 24,* 166-169. ISBN: 978-960-474-372-8.

**Table 1. Summary of simulation results showing the performances of fast regression methods**

| n=5000 | | 5% Outliers | | | | | | | | Normal | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | $r(l_i, l_j)$=0.20 | | | | $r(l_i, l_j)$=0.80 | | | | $r(l_i, l_j)$=0.20 | | | | $r(l_i, l_j)$=0.80 | | | |
| **p** | **Results** | ***DRCM*** | ***DRCM-N*** | ***R.VIF*** | ***VIF*** | ***DRCM*** | ***DRCM-N*** | ***R.VIF*** | ***VIF*** | ***DRCM*** | ***DRCM-N*** | ***R.VIF*** | ***VIF*** | ***DRCM*** | ***DRCM-N*** | ***R.VIF*** | ***VIF*** |
| **50** | **Avg. time (s)** | 0.186 | 0.213 | 1.003 | 2.186 | 0.153 | 0.210 | 0.965 | 2.094 | 0.118 | 0.140 | 0.569 | 1.304 | 0.106 | 0.130 | 0.570 | 1.305 |
| | **A (n=5) (%)** | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | **B (n=10) (%)** | 0.20 | 0.20 | 0.10 | 0.10 | - | - | - | - | - | - | - | - | - | - | - | - |
| | **C (n=35) (%)** | - | - | - | - | - | - | - | - | 0.06 | 0.06 | - | - | 0.06 | 0.06 | - | - |
| | **D (n=45) (%)** | 0.04 | 0.04 | 0.02 | 0.02 | - | - | - | - | 0.04 | 0.04 | - | - | 0.04 | 0.04 | - | - |
| | **RMSE** | 0.968 | 0.968 | 0.970 | 0.970 | 0.969 | 0.969 | 0.970 | 0.970 | 0.921 | 0.921 | 0.923 | 0.924 | 0.923 | 0.923 | 0.923 | 0.924 |
| **100** | **Avg. time (s)** | 0.315 | 0.486 | 1.983 | 4.258 | 0.330 | 0.462 | 1.882 | 4.149 | 0.182 | 0.250 | 1.092 | 2.480 | 0.159 | 0.224 | 1.119 | 2.487 |
| | **A (n=5) (%)** | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | **B (n=10) (%)** | - | - | 0.10 | 0.10 | - | - | 0.10 | 0.10 | - | - | - | - | - | - | - | - |
| | **C (n=85) (%)** | 1.41 | 1.43 | - | - | 1.62 | 1.55 | - | - | 0.86 | 0.85 | - | - | 1.29 | 1.16 | - | - |
| | **D (n=95) (%)** | 1.26 | 1.28 | 0.01 | 0.01 | 1.45 | 1.39 | 0.01 | 0.01 | 0.77 | 0.76 | - | - | 1.15 | 1.04 | - | - |
| | **RMSE** | 0.967 | 0.967 | 0.966 | 0.966 | 0.967 | 0.967 | 0.966 | 0.966 | 0.921 | 0.921 | 0.919 | 0.920 | 0.920 | 0.920 | 0.910 | 0.920 |
| **250** | **Avg. time (s)** | 1.157 | 1.504 | 4.476 | 9.312 | 1.043 | 1.257 | 4.430 | 9.319 | 0.592 | 0.705 | 2.702 | 5.817 | 0.604 | 0.704 | 2.693 | 5.823 |
| | **A (n=5) (%)** | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | **B (n=10) (%)** | - | - | 0.10 | 0.10 | - | - | 0.10 | 0.10 | - | 9.90 | - | - | - | - | - | - |
| | **C (n=235) (%)** | 1.74 | 1.74 | - | - | 2.08 | 2.04 | - | - | 1.28 | 1.26 | - | - | 1.59 | 1.49 | - | - |
| | **D (n=245) (%)** | 1.67 | 1.67 | 0.004 | 0.004 | 2.00 | 1.96 | 0.004 | 0.004 | 1.22 | 1.61 | - | - | 1.53 | 1.43 | - | - |
| | **RMSE** | 0.953 | 0.954 | 0.953 | 0.953 | 0.951 | 0.95 | 0.952 | 0.953 | 0.907 | 0.906 | 0.904 | 0.904 | 0.907 | 0.906 | 0.904 | 0.904 |
| **500** | **Avg. time (s)** | 2.871 | 3.457 | 8.734 | 19.079 | 2.708 | 3.442 | 8.713 | 19.029 | 1.903 | 2.145 | 5.335 | 11.416 | 1.888 | 1.999 | 5.425 | 11.460 |
| | **A (n=5) (%)** | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | **B (n=10) (%)** | 0.10 | 0.20 | 0.10 | 0.10 | 10 | 10.2 | 0.10 | 0.10 | 0.10 | 0.10 | - | 0.10 | 0.10 | 10 | - | 0.10 |
| | **C (n=485) (%)** | 2.08 | 2.06 | - | - | 2.47 | 2.27 | - | - | 1.86 | 1.65 | - | - | 2.47 | 2.27 | - | - |
| | **D (n=495) (%)** | 2.04 | 2.02 | 0.002 | 0.002 | 2.62 | 2.43 | 0.002 | 0.002 | 1.82 | 1.62 | - | 0.002 | 2.42 | 2.43 | - | 0.002 |
| | **RMSE** | 0.952 | 0.951 | 0.949 | 0.95 | 0.952 | 0.951 | 0.949 | 0.949 | 0.908 | 0.908 | 0.907 | 0.908 | 0.905 | 0.905 | 0.904 | 0.904 |
| **750** | **Avg. time (s)** | 5.835 | 6.683 | 12.745 | 27.248 | 6.214 | 6.585 | 13.242 | 27.412 | 4.190 | 4.482 | 8.000 | 17.004 | 4.386 | 4.811 | 8.842 | 17.944 |
| | **A (n=5) (%)** | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | **B (n=10) (%)** | 0.10 | - | 0.20 | 0.10 | - | 0.10 | 0.20 | 0.10 | - | - | - | - | - | - | - | - |
| | **C (n=735) (%)** | 1.90 | 1.90 | - | - | 2.44 | 2.38 | - | - | 1.77 | 1.64 | - | - | 2.17 | 2.14 | - | - |
| | **D (n=745) (%)** | 1.88 | 1.87 | 0.003 | 0.001 | 2.41 | 2.35 | 0.003 | 0.001 | 1.75 | 1.62 | - | - | 2.14 | 2.11 | - | - |
| | **RMSE** | 0.95 | 0.95 | 0.948 | 0.949 | 0.95 | 0.95 | 0.948 | 0.949 | 0.905 | 0.904 | 0.903 | 0.904 | 0.905 | 0.905 | 0.904 | 0.904 |
| **1000** | **Avg. time (s)** | 9.870 | 10.681 | 17.585 | 37.220 | 9.177 | 10.089 | 17.214 | 36.788 | 6.365 | 6.796 | 10.757 | 22.835 | 6.032 | 6.740 | 11.697 | 23.916 |
| | **A (n=5) (%)** | 100 | | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 | 100 |
| | **B (n=10) (%)** | 19.80 | 19.80 | - | 9.90 | 19.80 | 19.80 | - | 9.90 | 9.90 | 9.90 | - | 9.90 | 19.80 | 19.80 | - | 9.90 |
| | **C (n=985) (%)** | 2.53 | 1.82 | - | - | 2.13 | 2.13 | - | - | 2.47 | 1.94 | - | - | 2.14 | 2.03 | - | - |
| | **D (n=995) (%)** | 2.71 | 2.00 | - | 0.10 | 2.31 | 2.31 | - | 0.10 | 2.54 | 2.02 | - | 0.10 | 2.32 | 2.21 | - | 0.10 |
| | **RMSE** | 0.937 | 0.936 | 0.934 | 0.938 | 0.937 | 0.936 | 0.934 | 0.938 | 0.893 | 0.892 | 0.891 | 0.893 | 0.892 | 0.892 | 0.891 | 0.893 |

Avg: Average, s: second, A: Average number of target variables, B: Average number of noise variables correlated with target variables, C: Average number of noise variables uncorrelated with target variables, D: Average number of total noise variables, RMSE: Root Mean Square Error, DRCM: Dimensional Reduction of Correlation Matrix Method, DRCM-N: Nonparametric Dimensional Reduction of Correlation Matrix Method, VIF: Variance Inflation Factor Regression, R.VIF: Robust Variance Inflation Factor Regression

# O44

# An Extensive Web Interface for Validity and Reliability with RServe

**Ebru Ozturk**[1], Batuhan Bakirarar[2], Sevilay Karahan[1], Ergun Karaağaoglu[1]

[1]*Hacettepe University Department of Biostatistics, Ankara*
[2]*Ankara University Department of Biostatistics, Ankara*

ebru.ozturk3@hacettepe.edu.tr

## 1. INTRODUCTION

Measurement is a systematic and reproducible process that are quantified or classified based on events or objects. Measurement is often acquired by the assignment of numerical values. The most significant and essential compositions of the measurement process are based on validity and reliability. By using different programs such as SPSS, MedCalc or R, the validity and reliability analysis might be conducted. R offers to users working on free software environment for statistical computing and graphics. The validity and reliability analyses are conducted in R with different packages that might be confusing to the users. Therefore, especially for non-R users, the validity and reliability in measurement web interface was developed.

The main motivation of this study is to develop a web interface for validity and reliability via using RServe on Java platform that users can conduct their analysis without knowing how to use R.

## 2. METHOD

The two main components of measurement are validity and reliability. While validity can be defined as a measurement approach that gives the degree of success in evaluating what it is designed to measure, on the other hand; reliability can be defined as the consistency in obtaining similar results upon repeated measurement applications.

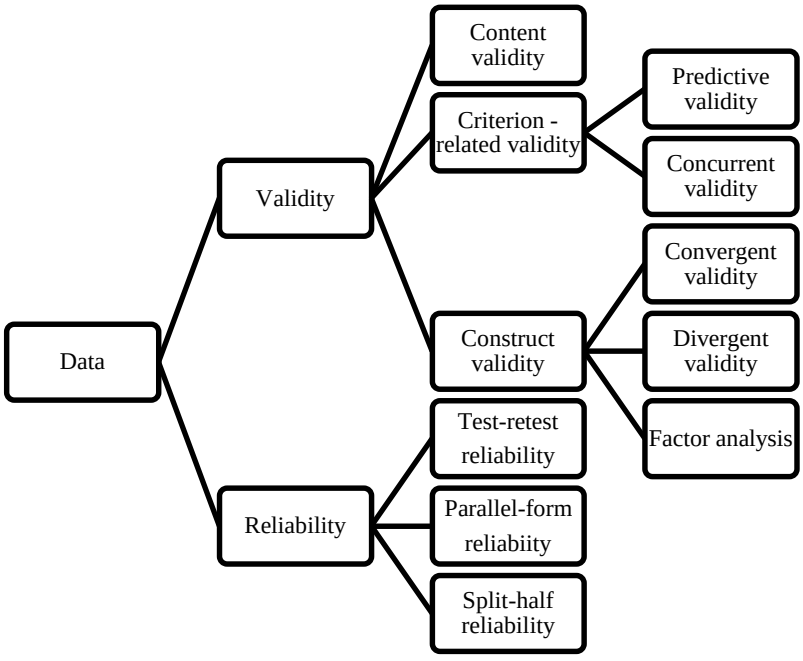


**Figure 1. Structure of Reliability and Validity (Source: Bajpai and Bajpai 2014, Alpar, R. 2018)**

The basic structure of validity and reliability is shown in Figure 1. According to this schema, the details of the statistical analyses in this interface are given in the following.

In this web interface, we offer to users content analysis based on Lawesh method with respect to content validity ratio and content validity index. While for concurrent validity, the correlation coefficient with the scatter plot and Bland-Altman graphics are given, for predictive validity the results of linear regression are given based on cross-validation. To investigate convergent and divergent validity, the correlation coefficient and scatter plot again are examined. For factor analysis, explanatory factor analysis results are given. Moreover, independent two samples tests might also have presented to construct validity. For categorical data, the results of ROC analysis are observed with optimum cut off points can be listed in predictive validity.

In the reliability part of this interface, the paired samples t-test and correlation coefficient are represented in order to analyze parallel-form reliability. For internal consistency the Cronbach's alpha coefficient, intraclass correlation coefficient based on one-way ANOVA, two-way ANOVA and two-way mixed methods and Kendall's coefficient of concordance (Kendall's W) are given in the test-retest method. For Split-Half reliability, the Spearman-Brown, Rulon and Guttman formulas are supplied. The Cohen's kappa and weighted kappa coefficients proposed for the reliability analysis of categorical data.

Except for validity and reliability, item analysis is also a substantial part of the measurement. In this web interface, the results of item analysis of scale are also offered to the users.

In order to develop a web interface, we prefer to use RServe on Java platform. Rserve is "a TCP/IP server which allows other programs to use facilities of R from various languages without the need to initialize R or link against R library" (RServe-Binary RServer, 2018). For the popular languages such as C/C++, PHP, and Java, there exist client-side applications. RServe allows the users authentication, remote connection, and file transfer. Usually, RServe is used to integrate R backend for computation of statistical models, graphics in other applications. Java is a programming language that is open-source, object-oriented and concurrent. It is also a computing platform for application development.

## 3. CONCLUSION

The process that obtaining numerical values regarding research subject can be defined as a measurement. The fundamental elements of the measurement process are validity and reliability. Although there exists the various source of software to conduct validity and reliability, we wanted to develop a web interface that examining this topic under the sub-headings in order to make it easier for the users. To achieve this, we developed this web interface based on RServe on Java platform. Our aim is serving to this interface especially for non-R users to conduct their analysis without R. As for future work, we determine to expand this web interface. We intend to study on confirmatory factor analysis and add this analysis on this interface.

## REFERENCES

1. Alpar, R. (2018). *Uygulamalı istatistik ve geçerlik-güvenirlik: Spor, sağlık ve eğitim bilimlerinden örneklerle*. Ankara: Detay Yayıncılık.
2. Bajpai, R., & Bajpai, S. (2014). Goodness of Measurement: Reliability and Validity. *International Journal of Medical Science and Public Health, 3*(2), 112. doi:10.5455/ijmsph.2013.191120133
3. Rserve - Binary R server. (n.d.). Retrieved from https://www.rforge.net/Rserve/

**XX$^{th}$ National and III$^{rd}$ International Biostatistics Congress**
**26-29 October 2018 GAZIANTEP**

## AUTHORS INDEX

ERAR Aydın
ESER Sultan
EROL Varlık
ERKORKMAZ Ünal
ERTÜRK ZARARSIZ Gözde
ETİKAN İlker
GENÇ Arzu
GÖKÇE AKTAŞ Hatice
GÖKMEN Derya
GÖKSÜLÜK Dinçer
GÖZÜKARA BAĞ G. Harika
GÜNEŞ Sezgin
GÜRTEKİN Başak
HAKVERDİ Gülden
HAYME Serhat
KAHRAMAN Turhan
KAR İrem
KARAAGAOĞLU A. Ergun
KARABULUT Erdem
KARACA Leyla
KARADAĞ Mehmet
KARAGÖZ Derya
KARAHAN Sevilay
KAŞALİ Kamber
KESKİN Nurhan
KESKİN Sıddık
KESKİN Seda
KESKİN D. Denizhan
KESKİNOĞLU Pembe
KILIÇ Erol
KINACI İsmail
KIVRAK Mehmet
KOÇ Senem
KOÇYİĞİT İsmail
KORKMAZ Duygu
KORKMAZ Selçuk
KOYUNCU Nursel
KOZANOĞLU İlknur
KÖREZ M. Kazım
KÖSE S. Kenan
KÖSE Timur
KUL Seval
KUNTER Birhan
KURT Burçin
KURTOĞLU DOĞAN Emel
KUŞ Çoşkun
KUTLAY Şehim
KUTLU Hüseyin
KUZU BULUT Melek
KÜÇÜKDEVECİ A. Ayşe
İPEKTEN Funda
MOLLAHASANOĞLU Emin
NAÇAR Melis
ONUR GÜLEN Elif
ORMAN Mehmet N.
ÖNER Kevser Setenay
ÖRS F. Betül
ÖZDEMİR Pınar
ÖZGÜL Semiha
ÖZGÜR Suriye
ÖZGÜR E. Gökay
ÖZKURT Ahmet
ÖZTORNACI Ragıp Onur
ÖZTÜRK Ahmet
ÖZTÜRK Ebru
PAK Ayşe
PALA Elif
PARMAKSIZ Ayhan
PASİN Özge
PEKGÖR Ahmet

| | |
|---|---|
| POLAT Yağmur | TURGAL Ebru |
| SAĞLAM Onursal | TURHAN Kemal |
| SANİSOĞLU S. Yavuz | TUTKUN ATA Nihal |
| SARAL D. Hazal | UMAROĞLU Mutlu |
| SARITÜRK Çağla | ULUSOY G. Merve |
| SATICI Ömer | USLU Adam |
| SERTDEMİR Yaşar | UZUN S. Begüm |
| SEYDAOĞLU Gülşah | ÜNAL Esin |
| SEYFELİ AKŞEHİRLİ Yasemin | ÜNAL İlker |
| SIDDIKOĞLU Duygu | ÜNER Sarp |
| SUNER Aslı | ÜNLÜSAVURAN Meltem |
| SÜT Necdet | ÜNSAL Serbülent |
| ŞAFAK Çağla | VEHİD ERTEM Hayriye |
| ŞAFAK D. Elif | YAŞAR Şeyma |
| ŞAHİN İbrahim | YAVUZ Zeynep |
| ŞEN Nazan | YILDIZ İsmail |
| TANRIKULU PEPEDİL Funda | YILDIZLI GÖRAL Meryem |
| TAŞDELEN Bahar | YOLOĞLU Saim |
| TATAR Erhan | YUMUK Fulden |
| TEKİNDAL M. Agah | YÜCEL PÜREN Sevinç |
| TENGİZ İ. Funda | YÜKSEL Selcen |
| TOMAK Leman | YÜRDEM Betül |
| TOPRAK Uğur | YÜRÜYEN Merve |
| TOTİK Nazlı | ZARARSIZ Gökmen |
| TUNÇ Zeynep | ZİHNİ B. Nihat |