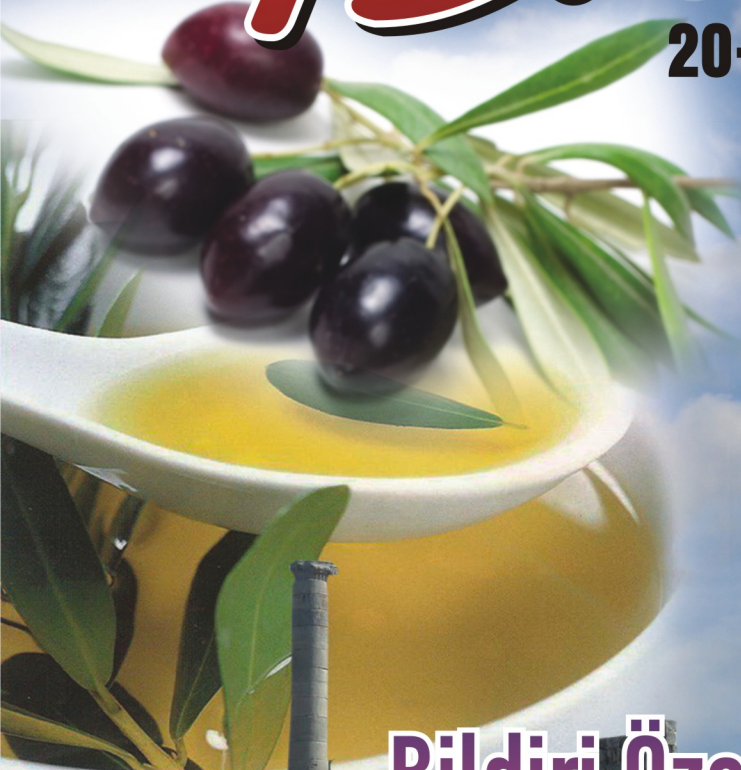
Uluslararası Katılımlı

15. Ulusal Biyoistatistik Kongresi

20-23 Ağustos 2013



Bildiri Özetleri Kitabı

Adnan Menderes Üniversitesi
Didim Sosyal Tesisleri Kültür ve Kongre Merkezi
Didim / AYDIN

ULUSLARARASI KATILIMLI 15. ULUSAL BİYOİSTATİSTİK KONGRESİ

BİLDİRİ ÖZETLERİ KİTABI

20-23 Ağustos 2013 | Didim, Aydın

ÖNSÖZ

Değerli Katılımcılarımız,

Adnan Menderes Üniversitesi Tıp Fakültesi Biyoistatistik ve Tıbbi Bilişim Anabilim Dalı olarak Biyoistatistik Derneği ile 20-23 Ağustos 2013 tarihlerinde "Uluslararası Katılımlı 15. Ulusal Biyoistatistik Kongresi"ni düzenlemenin mutluluğunu yaşıyoruz.

Kongremize yurtiçinden 114, yurtdışından 6 araştırmacı katılmıştır. Kongremizde Pompeu Fabra Üniversitesi'nden Prof. Dr. Michael GREENACRE tarafından "Constrained Biplots" başlıklı konferans ve Biyoistatistik Derneği Başkanı Prof. Dr. Kadir SÜMBÜLOĞLU tarafından "Klinik ve Saha Araştırmalarında Yöntemler ve Yaklaşımlar" kursu düzenlenecektir. Kongre bilimsel programı 42 sözlü sunum ve 37 poster sunum olmak üzere 10 oturumdan oluşmaktadır. Sunulan eserlerin, Biyoistatistik bilim alanına çok önemli kazanımlar ve birikimler sağlayacağından eminiz.

Kongremizin arzu ettiğimiz şekilde düzenlenmesi için katkıda bulunan ve desteklerini esirgemeyen başta Adnan Menderes Üniversitesi Rektörlüğü, Aydın Valiliği İl Kültür ve Turizm Müdürlüğü, Lavinia Travel Oganizasyon Şirketi, Yingari, Falcon, UYTES, Casio-Penta ve STATA'ya teşekkürlerimizi sunarız.

Ayrıca kongremizin düzenlenmesinde ve gerçekleştirilmesinde büyük emeği geçen Kongre Düzenleme Kurulu'na teşekkür ederiz.

Prof. Dr. Mevlüt TÜRE
Kongre Başkanı

Prof. Dr. Kadir SÜMBÜLOĞLU
Biyoistatistik Derneği Başkanı

ONURSAL KURUL

Kongre Onursal Başkanları

Prof. Dr. Mustafa BİRİNCİOĞLU
Adnan Menderes Üniversitesi Rektörü

Prof. Dr. Kadir SÜMBÜLOĞLU
Biyoistatistik Derneği Başkanı

Onur Kurulu

Prof. Dr. Abdulvahit YÜKSELEN
Adnan Menderes Üniversitesi Tıp Fakültesi Dekanı

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

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

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

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

Düzenleme Kurulu

Prof. Dr. Mevlüt TÜRE	Adnan Menderes Üniversitesi	Başkan
Doç. Dr. İmran KURT ÖMÜRLÜ	Adnan Menderes Üniversitesi	Sekreter
Araş. Gör. Merve KATRANCI	Adnan Menderes Üniversitesi	Üye
YL. Öğr. Afra ALKAN	Adnan Menderes Üniversitesi	Üye
YL. Öğr. Ali BENCİK	Adnan Menderes Üniversitesi	Üye

BİLİMSEL DANIŞMA KURULU

Prof. Dr. A. Ergun KARAAĞAOĞLU
Prof. Dr. Ahmet DİRİCAN
Prof. Dr. Atilla Halil ELHAN
Prof. Dr. Beyza AKDAĞ
Prof. Dr. Celal Reha ALPAR
Prof. Dr. Emine Arzu KANIK
Prof. Dr. Hafize SEZER
Prof. Dr. Handan ANKARALI
Prof. Dr. İlker ERCAN
Prof. Dr. İsmet DOĞAN
Prof. Dr. Kazım ÖZDAMAR
Prof. Dr. M. Yusuf ÇELİK
Prof. Dr. Mevlüt TÜRE
Prof. Dr. Mustafa ŞENOCAK
Prof. Dr. M. Nurullah ORMAN
Prof. Dr. Nazan ALPARSLAN
Prof. Dr. Nural BEKİROĞLU
Prof. Dr. Osman SAKA
Prof. Dr. Osman SARAÇBAŞI
Prof. Dr. Ömer SATICI
Prof. Dr. Refik BURGUT
Prof. Dr. Rian DİŞÇİ
Prof. Dr. Saim YOLOĞLU
Prof. Dr. Vildan SÜMBÜLOĞLU
Prof. Dr. Yavuz SANİSOĞLU
Prof. Dr. Yüksel BEK
Prof. Dr. Zeki AKKUŞ

Doç. Dr. Ahmet ÖZTÜRK
Doç. Dr. Ayşe Canan YAZICI
Doç. Dr. Bahar TAŞDELEN
Doç. Dr. Bülent ÇELİK
Doç. Dr. Canan BAYDEMİR
Doç. Dr. Cemil ÇOLAK
Doç. Dr. Derya ÖZTUNA
Doç. Dr. Erdem KARABULUT
Doç. Dr. Ersin ÖĞÜŞ
Doç. Dr. Fezan MUTLU
Doç. Dr. Gülşah SEYDAOĞLU
Doç. Dr. İmran KURT ÖMÜRLÜ
Doç. Dr. İ. Sefa GÜRCAN
Doç. Dr. K. Setenay ÖNER
Doç. Dr. Mehtap AKÇİL
Doç. Dr. Meriç ÇOLAK
Doç. Dr. Necdet SÜT
Doç. Dr. Nurhan DOĞAN
Doç. Dr. Pınar Özdemir GEYİK
Doç. Dr. Semra AKGÖZ
Doç. Dr. S. Kenan KÖSE
Doç. Dr. Sıddık KESKİN
Doç. Dr. Yasemin GENÇ

KONFERANS VE KURSLAR

"CONSTRAINED BIPLOTS"

Prof.Dr. Michael GREENACRE

Pompeu Fabra Üniversitesi, Barcelona, İspanya

Tarih: 19 Ağustos 2013 **Saat:** 16:00-19:00

Konferans detayı: After a general introduction to biplots, he will describe how adding constraints can relate the biplot solution to additional explanatory variables. Constrained biplots (e.g., redundancy analysis and canonical correspondence analysis) are used extensively in ecological research and have potential for use in other areas where many response variables are related to a set of explanatory variables.

"KLİNİK VE SAHA ARAŞTIRMALARINDA YÖNTEMLER VE YAKLAŞIMLAR KURSU"

Prof.Dr. Kadir SÜMBÜLOĞLU

Biyoistatistik Derneği Başkanı

Tarih: 24-25 Ağustos 2013

Kurs detayı: Araştırma görevlileri ve lisansüstü öğrencilere yönelik olarak düzenlenen kurs, klinik ve saha araştırmalarını kapsamaktadır.

İÇİNDEKİLER

ÖNSÖZ----------I

KONFERANS VE KURSLAR----------IV

I. SÖZLÜ BİLDİRİLER ---------- 1

S01. PROSTAT VERİSİNİN SPLAYN DÜZELTME VE REGRESYON SPLAYN YÖNTEMLERİYLE TAHMİNİ 2

S02. İLİŞKİLİ BİLEŞEN REGRESYONU VE UYGULAMASI 5

S03. MODELE YENİ BİR BELİRTEÇ EKLEMENİN SINIFLAMAYA OLAN KATKISI: EĞRİ ALTINDA KALAN ALANA ALTERNATİF İKİ YÖNTEM 6

S04. DNA MİKRODİZİLİM VERİLERİ ÜZERİNDE GENELLEŞTİRİLMİŞ BOOSTED REGRESYON MODELLERİ 8

S05. LOJİSTİK AYRIŞTIRMADA RICHARD EĞRİSİNİN LİNK FONKSİYONU OLARAK KULLANILMASI 10

S06. PROPORTIONAL ODDS REGRESYON MODELİ'NDE UYUM İYİLİĞİ TESTLERİNİN SİMÜLASYON İLE KARŞILAŞTIRILMASI 11

S07. COX REGRESYON MODELİ: ORANSAL HAZARD VARSAYIMININ İNCELENMESİ 12

S08. DOĞRUSAL REGRESYON MODELLERİNDE DEĞİŞKEN SEÇİM KRİTERLERİNİN PERFORMANSLARININ DEĞERLENDİRİLMESİ 14

S09. AĞIRLIKLANDIRILMIŞ KAPLAN-MEIER YÖNTEMİNİN ZAMANA BAĞLI ROC ANALİZİNE UYARLANMASI 15

S10. BİLGİ KURAMI YAKLAŞIMI İLE BİLGİSAYARLI TOMOGRAFİK KORONER ANJİYOGRAFİNİN TANISAL DEĞERİNİN BELİRLENMESİ 16

S11. ISOBOLOGRAM VE ETKİN DOZ TESPİTİ 18

S12. FAKTÖR ANALİZİ İÇİN UYGUN ÖRNEK GENİŞLİĞİNİN BELİRLENMESİ 20

S13. ÇOKLU FAKTÖR ANALİZİNİN TANITIMI VE BİR UYGULAMASI 21

S14. SİMÜLASYON ÇALIŞMALARINDA UYGUN SİMULASYON SAYISININ BELİRLENMESİ 22

S15. YAPAY MADDE İŞLEV FARKLILIĞINA NEDEN OLAN FAKTÖRLERİN BENZETİM ÇALIŞMASI İLE DEĞERLENDİRİLMESİ 24

S16. DRUG/NON-DRUG CLASSIFICATION USING SUPPORT VECTOR MACHINES WITH VARIOUS FEATURE SELECTION STRATEGIES 26

S17. BOOKSTEIN KOORDİNATLARINDA REFERANS LANDMARKLARIN DEĞİŞMESİ DURUMUNDA HOTELLING T^2 TEST SONUCUNDAKİ DEĞİŞİMİN İNCELENMESİ 28

S18. İSTATİSTİKSEL ŞEKİL ANALİZİNDE KULLANILAN İKİ ÖRNEKLEM TESTLERİNDE LANDMARK SAYISINA BAĞLI OLARAK TİP I HATA DÜZEYİNİN İNCELENMESİ 29

S19. ÇOK BOYUTLU SAĞKALIM VERİLERİNDE DENETİMLİ TEMEL BİLEŞENLER ANALİZİNE ALTERNATİF BİR BOYUT İNDİRGEME YAKLAŞIMI 30

S20. BREIMAN ALGORİTMASI İLE HOMOJEN ALT GRUPLARIN BELİRLENMESİ: BİR UYGULAMA 31

S21. KÜMELEME ANALİZİNDE KENDİNİ DÜZENLEYEN AĞAÇ ALGORİTMASI 33

S22. PARAMETRİK OLMAYAN KOVARYANS ANALİZİNDE KULLANILAN YAKLAŞIMLAR 35

S23. PARAMETRİK OLMAYAN KOVARYANS ANALİZİ 36

S24. GENETİK EPİDEMİYOLOJİDE KULLANILAN PAKET PROGRAMLAR: MERLİN İLE BİR UYGULAMA 37

S25. DERMATOLOJİK HASTALIKLARIN TEŞHİSİNDE TEMEL SINIFLANDIRMA ALGORİTMALARININ BAŞARIM ÖLÇÜTLERİNİN DEĞERLENDİRİLMESİ 38

S26. SPOR VE YAŞAM MERKEZLERİNDE VERİ MADENCİLİĞİ ÇALIŞMASI 45

S27. WEB TABANLI İSTATİSTİK ÇÖZÜMLEMELER: FAKTÖRİYEL VARYANS ANALİZİ VE İNTERAKSİYON ETKİSİNİN ÇOKLU KARŞILAŞTIRMA YÖNTEMİ İLE İNCELENMESİ 47

S28. ÇOKLU KALİTE KARAKTERİSTİKLERİNİN EŞZAMANLI OPTİMİZASYONU 48

S29. HOMOJEN OLMAYAN POISSON SÜREÇLERİNDE ŞİDDET FONKSİYONUNUN TAHMİNİ: RADYOLOJİK VERİ ÜZERİNE UYGULAMA 49

S30. MEME KANSERLİ HASTALARDA POST-OP AĞRININ DEĞERLENDİRİLMESİNDE KÖRLEME YÖNTEMLERİNİN ETKİNLİĞİ 50

S31. KONFİGÜRAL FREKANS ANALİZİ 52

S32. ASSOCIATION MAPPING USING A BAYESIAN MODEL FOR FEED EFFICIENCY IN F2 MICE DATASET 53

S33. MIDDLESEX ELDERLY ASSESSMENT OF MENTAL STATE (MEAMS) ÖLÇEĞİNİN DEĞERLENDİRİLMESİ: KARMA RASCH MODEL YAKLAŞIMI 55

S34. ERYTHROMYCIN İLACININ YAN ETKİLERİNİN ARAŞTIRILMASI İÇİN WEB'DEKİ HASTA YORUMLARININ ANALİZİ 57

S35. MINITAB VE FREUD & PERLES KARTİL TAHMİN YÖNTEMLERİNİN ÖRNEKLEME DAĞILIMLARI VE PERFORMANSLARININ KARŞILAŞTIRILMASI: BİR SİMÜLASYON ÇALIŞMASI 59

S36. EN İYİ KANIT SAĞLAYAN KLİNİK DENEMELER: RONALD AYLMER FISHER VE AUSTIN BRADFORD HILL'İN KATKILARI 60

S37. YAKALANMA OLASILIKLARININ HETEROJEN OLDUĞU TOPLULUKLARDA SAMPLE COVERAGE YAKLAŞIMI İLE POPULASYON BÜYÜKLÜĞÜNÜN TAHMİNLENMESİ 62

S38. NORMAL DAĞILIMA AİT DEĞERLENDİRMELERDE UYGUN YÖNTEM SEÇİMİNİN ÖNEMİ 63

S39. TEKRARLI FİZYOLOJİK ÖZELLİKLERDE DÜZEN DEĞİŞİKLİKLERİNİN YARGILANMASI 65

S40. BİYOİSTATİSTİK ANABİLİM DALLARINDA BOLOGNA SÜRECİ BAĞLAMINDA İÇ KONTROL SİSTEMİ UYGULAMALARI 66

S41. MODELE EKLENEN YENİ BİR TAHMİN EDİCİNİN MODEL PERFORMANSINA KATKISININ DEĞERLENDİRİLMESİ: KARDİYOVASKÜLER RİSK TAHMİN MODELİNE İLİŞKİN BİR UYGULAMA 68

II. POSTER BİLDİRİLER ----------70

P01. ALLOMETRİ ÇALIŞMALARINDA TANJANT KOORDİNATLARI VE TEMEL BİLEŞEN SKORLARININ KULLANILDIĞI MODELLERİN PERFORMANSLARININ İNCELENMESİ 71

P02. VETERİNER HEKİMLERİN BİYOİSTATİSTİK BİLGİSİ: ULUSLARARASI WEB ANKET 72

P03. STEREOLOJİ ÇALIŞMALARINDA ARAŞTIRMA TASARIMI VE ÖRNEKLEME YÖNTEMİ 73

P04. ÇOKLU UYUM ANALİZİ YÖNTEMİ İLE TIP FAKÜLTESİ ÖĞRENCİLERİNİN MESLEK SEÇİMİNİ ETKİLEYEN BAZI FAKTÖRLERİN İNCELENMESİ: BAŞKENT ÜNİVERSİTESİ ÖRNEĞİ 74

P05. VERİ KÜMELERİNDEN BİLGİ KEŞFİ: VERİ MADENCİLİĞİ 75

P06. TIP BİLİMLERİNDE YAYINLANAN MAKALELERİN İSTATİSTİKSEL HATALAR BAKIMINDAN İNCELENMESİ 77

P07. META ANALİZİNİN VETERİNER HEKİMLİKTE UYGULANMASI 78

P08. BOX-BEHNKEN DENEME DÜZENİ İLE ENGELLİ ÇOCUĞA SAHİP ANNELERİN RUHSAL ALAN YAŞAM KALİTESİNİN DEĞERLENDİRİLMESİ 79

P09. YANIT DEĞİŞKENİNİN İKİLİ OLDUĞU KÜMELENMİŞ VERİLERDE KULLANILABİLECEK İSTATİSTİKSEL YÖNTEMLER 81

P10. ÇARPIK DAĞILIMLARDA ROC EĞRİSİ ALTINDA KALAN ALAN TAHMİNİ 82

P11. HİPERLİPİDEMİ HASTALARINDA KOLESTEROL SEVİYESİNİN YAŞA GÖRE DEĞİŞİMİNİN DEĞİŞİK REGRESYON MODELLERİYLE İNCELENMESİ 83

P12. GENETİK EPİDEMİYOLOJİDE İSTATİSTİKSEL YAKLAŞIMLAR: HEMATOLOJİK KANSERLERDE BİR UYGULAMA 84

P13. TABANUS BROMIUS LINNAEUS SİNEK TÜRÜNÜN KANAT YAPISININ İSTATİSTİKSEL ŞEKİL ANALİZİ İLE İNCELENMESİ 85

P14. TİP-2 DİYABETLİ HASTALARDA ALBUMİNURİ DÜZEYİNİ TAHMİN ETMEDE SINIFLANDIRMA YÖNTEMLERİNİN PERFORMANSLARININ KARŞILAŞTIRILMASI . 86

P15. DİŞ HEKİMLERİNİN BİYOİSTATİSTİK BİLGİSİ: ULUSLARARASI WEB ANKET .. 87

P16. ROC EĞRİSİ ÇÖZÜMLEMESİNDEKİ YENİ YAKLAŞIMLARIN ÖRNEKLER ÜZERİNDE İNCELENMESİ 88

P17. SEPSİS GELİŞEN YENİDOĞAN BEBEKLERDE VİTAL BULGULAR KULLANILARAK OLUŞTURULAN LOJİSTİK REGRESYON MODELİ İLE DAHA ERKEN TEŞHİS MÜMKÜN MÜ?..... 89

P18. PUBMED KLİNİK DENEME META ANALİZLERİNDE YAYIN YANLILIĞI DEĞERLENDİRİLME YÖNTEMLERİ 91

P19. PEDİATRİK YOĞUN BAKIM BİRİMLERİNİN HİZMET KALİTESİ DEĞERLENDİRİLMESİ İÇİN BİR ÖNERİ PRISM-II VE PIM-II RİSK ÖLÇÜTLERİ 93

P20. BOZULMUŞ NORMAL DAĞILIM ALTINDA KONUM VE DAĞILIM PARAMETRE TAHMİN EDİCİLERİNİN ETKİNLİKLERİNİN BENZETİM YOLUYLA KARŞILAŞTIRILMASI 94

P21. SIRALI KÜME ÖRNEKLEMESİ İLE TEK VE İKİ YIĞIN ORTALAMASI İÇİN HİPOTEZ TESTLERİNİN GÜÇ VE GEREKLİ ÖRNEK ÇAPI BAKIMINDAN İNCELENMESİ 96

P22. 0-5 YAŞ ARASI ÇOCUKLARIN BÜYÜME EĞRİLERİ VE REFERANS EĞRİLERİYLE KARŞILAŞTIRILMASI 98

P23. SAĞKALIM ANALİZ YÖNTEMLERİ VE KARACİĞER NAKLİ VERİLERİ İLE BİR UYGULAMA..... 100

P24. GÜVENİRLİK ANALİZİNDE SINIF İÇİ İLİŞKİ KATSAYILARININ UYGULAMA ÖRNEKLERİ ÜZERİNDE İRDELENMESİ..... 101

P25. DİJİTAL SİNYAL İŞLEME YÖNTEMLERİYLE ELDE EDİLEN VERİLER YARDIMIYLA OBSTRÜKTİF UYKU APNESİNİNİN SINIFLANDIRILMASINDA ÇOK DEĞİŞKENLİ İSTATİSTİKSEL ANALİZ YÖNTEMLERİ VE YAPAY SİNİR AĞLARI 102

P26. FAKTÖR ANALİZİ VE LOJİSTİK REGRESYON ÇÖZÜMLEMESİNİN BİRLİKTE KULLANIMININ ÖNEMİ: POSTER SUNUMU..... 104

P27. HİYERARŞİK KÜMELEME YÖNTEMLERİNİN KARŞILAŞTIRILMASI 105

P28. TÜRKİYE'DE 1987-2011 YILLARI ARASINDAKİ İNTİHAR OLAYLARININ JOINPOINT REGRESYON ANALİZİ İLE DEĞERLENDİRİLMESİ..... 106

P29. RANDOLPH'UN ÇOK DEĞERLENDİRİCİLİ BAĞIMSIZ MARJİNAL KAPPASI (K_{FREE}): FLEISS'İN KAPPA'SINA ALTERNATİF BİR YAKLAŞIM..... 108

P30. FARKLI ÖRNEKLEM GENİŞLİKLERİ VE DAĞILIMLARDA NORMAL DAĞILIM TESTLERİNİN MONTE CARLO SİMULASYONU İLE KARŞILAŞTIRILMASI..... 109

P31. BOLOGNA SÜRECİ VE BİYOİSTATİSTİK PROGRAMLARININ UYUMU 110

P32. LMSP METHOD TO CONSTRUCT WRIST CIRCUMFERENCE REFERENCE CURVES OF 6-17 YEAR-OLD TURKISH CHILDREN 112

P33. YATAN HASTALARDA TOPLUM-KÖKENLİ PNÖMONİ TEDAVİSİNDE SEFTRİAKSONA KARŞI SEFUROKSİMİN MALİYET-ETKİNLİK ANALİZİ 114

P34. İKİ PARALEL DENEY TASARIMINDA HİPOTEZ YARGILAMASINA GÖRE ÖRNEKLEM SAYISI HESAPLARI 115

P35. VERİ MADENCİLİĞİ ANALİZİNDE KULLANILAN YÖNTEMLERDEN HANGİLERİNİN TIPTA KULLANIMI UYGUNDUR? 116

P36. DNA MİKROARRAY ÇALIŞMALARINDA KNORM KORELASYON KATSAYISININ KULLANILMASI 117

P37. GUIDELINES FOR THE DESIGN AND STATISTICAL ANALYSIS OF BIOLOGICAL EXPERIMENTS 118

I. SÖZLÜ BİLDİRİLER

S01

PROSTAT VERİSİNİN SPLAYN DÜZELTME VE REGRESYON SPLAYN YÖNTEMLERİYLE TAHMİNİ

Özlem Aksoy[1], Dursun Aydın[1]

[1]*Muğla Sıtkı Koçman Üniversitesi, Fen Fakültesi, İstatistik Bölümü, MUĞLA*

ozlemaksoy@mu.edu.tr

Amaç: Regresyon analizi istatistiğin en yaygın uygulama alanlarından biridir. Son zamanlarda yarı parametrik (semiparametric) regresyona ilgi artmıştır. Çünkü yarı parametrik regresyon tüm parametrik olmayan yöntemlerin ciddi dezavantajlarını ve parametrik model ile ilgili yanlış belirlenme riskini azaltır. Bu çalışmada temel amaç, prostat kanseri olan bireylerde spesifik prostat antijenini (prostate specific antigen-PSA) etkileyen değişkenler dikkate alınarak her iki yöntem bulguları değerlendirilmiştir.

Yöntem: Bir yarı parametrik regresyon modeli;

$$LPSA_i = \beta_0 + \beta_1 LCAVOL_i + f(LWEİGHT_i) + \varepsilon_i,\ i = 1,...,n$$
$$veya \qquad (1)$$
$$y_i = \beta_0 + \beta_1 x_i + f(t_i) + \varepsilon_i$$

Burada $f(.)$ pürüzsüz (smooth) parametrik olmayan fonksiyon ve ε_i bağımsız ve normal dağılan, sıfır ortalamalı ve σ_ε^2 ortak varyanslı hata terimidir. Eşitlik (1)'in vektör ve matris formu;

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{f} + \boldsymbol{\varepsilon} \qquad (2)$$

Burada $\boldsymbol{\beta} = [\beta_0, \beta_1]$, $\mathbf{y} = [y_1,...,y_n]^T$, $\mathbf{f} = [f(t_1),...,f(t_n)]^T$, $\boldsymbol{\varepsilon} = [\varepsilon_1, \varepsilon_2,...,\varepsilon_n]^T$ ve $\mathbf{X} = [\mathbf{1}\ x_i]_{1\le i\le n}$ `dir.

Yarı parametrik regresyon modelinin splayn düzeltme ile tahmini: Bu yöntemde, $\boldsymbol{\beta}$ parametre vektörü ve f fonksiyonunun $t_1,...,t_n$ örnek noktalarında ki değerleri cezalı en küçük kareleri (CEKK) minimum yapan değerler olarak tahmin edilir:

$$CEKK(\boldsymbol{\beta}, \mathbf{f}) = \sum_{i=1}^{n} \left\{ y_i - x_i^T\boldsymbol{\beta} - f(t_i) \right\}^2 + \lambda \int_a^b \left(f''(t) \right)^2 du \qquad (3)$$

Burada $f \in C^2[a,b]$ pürüzsüz bir fonksiyondur. $\boldsymbol{\beta} = 0$ olduğunda, f `in tahmini, $\hat{\mathbf{f}} = \left(\hat{f}(t_1),...,\hat{f}(t_n) \right) = S_\lambda \mathbf{y}$, olarak elde edilir. Burada S_λ, λ düzeltme parametresi ve $t_1,...,t_n$ düğüm noktalarına bağlı olan pozitif tanımlı simetrik bir düzeltme matrisidir. Çapraz geçerlilik ve genelleştirilmiş çapraz geçerlilik gibi bazı kriter fonksiyonları minimum yapan λ değeri seçilir.

Yarı parametrik regresyon modelinin regresyon splayn ile tahmini: n örneklem hacmi büyük olduğunda n tane de düğüm noktası kullanıldığı için splayn düzeltme daha az tercih edilir. Bu durumda splaynı

tahmin etmek için genel yaklaşım regresyon splayn yöntemini kullanmaktır. Regresyon splayn, sıfırdan farklı en yüksek dereceden türevi sabit düğüm noktalarını içeren parçalı polinom bir fonksiyondur. Genellikle regresyon splayn yöntemi, gereksiz düğüm noktalarını silerek fonksiyonu tahmin eder. Bu durumda eşitlik (1)`deki f aşağıda verilen m sayıda düğüm noktası içeren p. dereceden splayn fonksiyonu (budanmış üstel taban fonksiyon) ile tahmin edilir:

$$f(LWEIGHT_i) = f(LWEIGHT_i, b) = b_0 + b_1 LWEIGHT_i + ... + b_p LWEIGHT_i^p + \sum_{j=1}^{m} \beta_k (LWEIGHT_i - k_j)_+^p, \; i = 1,...,n \tag{4}$$

Burada k_j ,j.düğüm noktası, $\boldsymbol{\beta} = \left(b_0,...,b_p,\beta_1,...,\beta_m\right)^T$ regresyon katsayılarının dizisi ve $(a)_+ = \max(0,a)$ `dır.

Bulgular: Prostat verisi için splayn düzeltme yöntemiyle elde edilen bulgular;

	Df	Npar	Df	Npar	F	Pr(F)
lcavol		1				
s(lweight, 9)	1		8		2.141	0.04014 *

Katsayılar:

	Estimate	Std. Error	t value	Pr(>|t|)
lcavol	0.67761	0.06340	10.69	<2e-16 ***
s(lweight, 9)	0.43378	0.03103	13.98	<2e-16 ***

$R^2_{Adj.}$ = 0.9315,

Regresyon splayn yöntemiyle elde edilen bulgular ise;

Katsayılar:

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	1.58756	0.11574	13.717	< 2e-16 ***
lcavol	0.65987	0.06697	9.853	7.94e-16 ***

Approximate significance of smooth terms:

	edf	Ref.df	F	p-value
s(lweight)	7.848	9.676	2.552	0.0101 *

$R^2_{Adj.}$ =0.62.

Sonuç: Bu çalışmada, R paket programı kullanılarak PSA`yı etkileyen prostat uzunluğu ve kanser hacmi değişkenleri incelenmiş ve splayn düzeltme yöntemi ile PSA`nın daha iyi açıklandığı sonucuna varılmıştır.

Anahtar sözcükler: Yarı Parametrik Regresyon, Splayn Düzeltme, Regresyon Splaynı

Kaynaklar

Aydın, D., (2005), "Semiparametrik Regresyon Modellemede Splayn Düzeltme Yaklaşımı ile Tahmin ve Çıkarsamalar, Doktora Tezi, Anadolu Üniversitesi.

Aydın, D., and Tuzemen Ş., (2010), "Estimation in semi-parametric and additive regression using smoothing spline and regression spline", Second International Conference on Computer Research and Development, IEEE Computer Society, Kuala Lumpur, Malaysia, 7-10 May 2010.

Enayetur Raheem., S.M., (2012), "Absolute Penalty and Shrinkage Estimation Strategies in Linear and Partially Linear Models", Doktora Tezi, Windsor Universitesi.

Hastie, T., Tibshirani, R., and Friedman, J. (2009), "The Elements of Statistical Learning: Data Mining, Inference and Prediction", Springer.

Stamey, T.A., Kabalin, J.N., McNeal, J.E., Johnstone, I.M., Freiha, F., Redwine, E.A., Yang, N.,(1989), "The Journal of Urology, 141(5):1076-1083, Division of Urology, Stanford University Medical Center, California 94305-5118.

S02

İLİŞKİLİ BİLEŞEN REGRESYONU VE UYGULAMASI

Sıddık Keskin[1], **Sadi Elasan[1]**

[1]*Yüzüncü Yıl Üniversitesi, Tıp Fakültesi, Biyoistatistik Anabilim Dalı, Kampüs-Van*

sadi0665@hotmail.com

Amaç: Açıklayıcı değişken sayısının, örnek genişliğine yaklaştığı veya örnek genişliğini geçtiği durumlarda, diğer bir ifade ile yüksek boyutlu veri setlerinde, regresyon modelleri ile tahminde, güvenilirliğin nasıl artırılacağı önemli sorunlardan birisidir. Bu amaçla kullanılabilecek yeni yöntemlerden birsi, İlişkili Bileşen Regresyonu (Correlated Component Regression) dur. Bu çalışmada, İlişkili Bileşen Regresyonu hakkında bilgi verilerek, bir uygulama ile birlikte tanıtılması amaçlanmıştır.

Yöntem: Regresyon analizi yapılması gereken bilimsel çalışmalarda, açıklayıcı değişken sayısı; örnek genişliğine yaklaştığında veya örnek genişliğini geçtiğinde (yüksek boyutlu veri setlerinde), standart regresyon analizi yöntemiyle yapılacak tahminlerde, tahmin edilen katsayılar, çoklu bağlantı (kovaryans matrisinin tekil olması) nedeniyle değişkenlik göstermektedir. Bu tip durumların çözümünde: 1) Sınırlı model yaklaşımları; örneğin cezalı regresyon yöntemleri (penalized regression methods, lasso ve elastik ağı), 2) Boyut indirgeme yaklaşımları; örneğin temel bileşenler regresyonu ya da kısmi en küçük kareler regresyonu, bunlar boyut sayısını K<Min (P, N-1) olarak indirir), kullanılabilse de, genel bir ifade ile yüksek boyutlu verilerin önemli düzeyde düzenlileştirilmesine (regularization) ihtiyaç vardır. Bunun için alternatif bir yöntem olarak, "İlişkili Bileşen Regresyonu" (İBR) problemin çözümüne yardımcı olabilir. İBR, ilk defa Magidson (2010) tarafından geliştirilmiş bir yöntem olup, yüksek boyutlu verilerde, standart regresyon analizi yöntemleri ile karşılaşılabilecek; çoklu bağlantı, uyum eksikliği veya aşırı uyum gibi sorunların çözümü için önerilen alternatif bir yaklaşımdır. Cevap değişkeninin tipine göre değişen İBR' nin farklı algoritmaları mevcuttur. Cevap değişkeninin sürekli olması durumunda, İBR-Doğrusal Regresyon ve ikili (binary) olması durumunda, İBR-Logistik Regresyon kullanılabilirken, sağkalım verilerinde İBR-Cox Regresyonu kullanılabilir. Yöntem, K adet ilişkili bileşenleri kullanır. Bu ilişkili bileşenler, program tarafından belirlenebileceği gibi araştırıcı tarafından da belirlenebilir. Temel Bileşenler Analizinde olduğu gibi her bileşen, orijinal değişkenlerin doğrusal kombinasyonudur ve bu bileşenler tahmin modelinde, orijinal değişkenlerin yerine düzenlileştirme modeli olarak kullanılır.

Sonuç: Regresyon analizlerinde karşılaşılabilecek; çoklu bağlantı, uyum eksikliği veya aşırı uyum gibi sorunların çözümü için İBR'nin kullanılabileceği ve daha yüksek gücü yakalayabileceği söylenebilir.

Anahtar kelimeler: Bileşen, çoklu bağlantı, uyum eksikliği, yüksek boyutlu veri

S03

MODELE YENİ BİR BELİRTEÇ EKLEMENİN SINIFLAMAYA OLAN KATKISI: EĞRİ ALTINDA KALAN ALANA ALTERNATİF İKİ YÖNTEM

Eda Karaismailoğlu[1], A. Ergun Karaağaoğlu[1]

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

eda.ozturk@hacettepe.edu.tr

Giriş: Sağlık alanında, testlerin ya da tahmin modellerinden elde edilen risklerin doğru değerlendirilmesi, hasta ve sağlıklıları belirlemede önemlidir. Testlerin performanslarını değerlendirmede birçok ölçü bulunmakla birlikte, en sık kullanılanı ROC (Receiver Opperating Characteristic) eğrisinden elde edilen eğri altında kalan alandır. Bu yöntem, elde edilmesi ve yorumlaması kolay olmasına rağmen, özellikle, mevcut risk modeline yeni bir değişken eklendiğinde, yeni modelin performansını değerlendirmede duyarlı bir ölçü değildir. Pencina ve arkadaşları (2008) bu sorunun üstesinden gelebilmek için iki yeni yöntem önermişlerdir. Bunlardan ilki; modele yeni bir değişken eklendiğinde bireylerin ne kadarının doğru sınıflandığını gösteren Net Yeniden Sınıflandırma İyileştirmesi-NYSI (Net Reclassification Improvement-NRI); diğeri ise, modele yeni bir değişken eklendiğinde, risk tahmin olasılıklarının hasta ve sağlıklıları birbirinden ne kadar ayırdığını ölçen Birleştirilmiş Ayrımsama İyileştirmesi-BAI (Integrated Discrimination Improvement-IDI) yöntemleridir. Bu çalışmanın amacı; risk modeline eklenen değişkenin model performansına etkisinin; ROC eğrisi altında kalan alan (EAA), NYSI ve BAI ölçüleriyle değerlendirilmesidir.

Gereç ve Yöntem: Bu çalışmada, 1948-1958 yılları arasında, Massachusetts eyaletinin Framingham şehrinde yürütülmüş olan Framingham Kalp Çalışması'ndan alınan toplanan veriler incelenmiştir. Çalışmaya 2588 kişi dahil edilmiştir. Bağımlı değişken olarak koroner arter hastalığı geçirme durumu alınmıştır. Bireylerin; cinsiyet, yaş, diyabet varlığı, HDL (High-density lipoprotein) değeri ve sistolik kan basıncı değişkenleri ile oluşturulan temel lojistik modeline, toplam kolesterol değişkeni eklenerek yeni model elde edilmiştir. Buna göre; toplam kolesterolün model performansına etkisi; EAA, NYSI ve BAI yöntemleri ile değerlendirilmiştir.

Bulgular: Temel ve yeni modelden elde edilen risk tahminlerine göre; temel modelin EAA'sı 0.726, toplam kolesterol eklenerek elde edilen yeni modelin EAA'sı ise 0.733 olarak bulunmuştur. Eğri altında kalan alanların farkı istatistiksel olarak anlamlı değildir (p=0.143). NYSI değeri 0.172 (p=0.003), BAI değeri ise 0.006 (p=0.002) olarak bulunmuştur. NYSI ölçüsünden elde edilen sonuca göre, modele toplam kolesterol eklendiğinde koroner arter hastalığı olanların sağlıklılara göre bir üst risk kategorisine geçme olasılığı, bir alt kategoriye geçme olasılığından %17 daha fazladır. BAI ölçüsünden elde edilen sonuca göre ise; toplam kolesterol değişkeni modele eklendiğinde; hasta ve kontroller arasında kestirilen ortalama risk farkı 0.006 birim artmaktadır.

Sonuç: Temel bir modele yeni bir değişken eklenerek bu değişkenin modeldeki performansı incelenmek istendiğinde - özellikle temel modelde etkili risk faktörleri varsa - EAA bu performansı değerlendirmede yetersiz bir yöntemdir. Bu yüzden yeni değişkenin modelde yer alıp almamasına karar verirken EAA'nın yanı sıra NYSI ve BAI ölçüleri de dikkate alınmalıdır.

Anahtar kelimeler: ROC eğrisi, Net Yeniden Sınıflandırma İyileştirmesi, Birleştirilmiş Ayrımsama İyileştirmesi

S04

DNA MİKRODİZİLİM VERİLERİ ÜZERİNDE GENELLEŞTİRİLMİŞ BOOSTED REGRESYON MODELLERİ

Selen Yılmaz Işıkhan[1], Erdem Karabulut[2], C. Reha Alpar[2]

[1]*Hacettepe Üniversitesi Sosyal Bilimler Meslek Yüksekokulu, Ankara*

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

seleny@hacettepe.edu.tr

Amaç: Mikrodizilim Teknolojisi, biyolojik bir örnekteki binlerce genin ekspresyon düzeyini aynı anda inceleyebilen bir (1995) yöntemdir. Genetik araştırmalarda az sayıda hastaya ait binlerce gen verisinin bulunması, klasik istatistiksel yöntemlerin (Lojistik regresyon,Varyans analizi, Doğrusal regresyon analizi vb.) kullanımında sorunlar ortaya çıkarmaktadır. Mikrodizilim gen ifade çalışmalarında çok fazla sayıdaki genin aynı anda analizi veri madenciliği yöntemlerinin de kullanılmasıyla mümkün hale gelmiştir. İlk olarak sınıflama problemlerinin çözümü için geliştirilen bu yöntemler daha sonra regresyon yöntemlerine özellikle doz-yanıt çalışmalarına da uyarlanmıştır. Bu analizlerin önemli bir bölümü; genlerin sınıflandırılması, önemli genlerin bulunması, hastaların sınıflandırılması, genlerin kümelenmesi ve genlerden doz, metabolik sendrom vb. nicel ölçümlerin kestirimi gibi amaçlar taşımaktadır. Bu çalışmanın amacı, en çok tercih edilen regresyon yöntemlerinden biri olan karar ağacı (KAR) regresyonu ve bu yöntemin "Boosting" topluluk (ensemble) metodu ile birlikte ele alındığı boosted regresyon ağacı yönteminin (BRA) mikrodizilim verileri üzerindeki kestirim performanslarını karşılaştırmaktır.

Gereç ve Yöntem: Çalışmada sözü edilen yöntemler Gene Expression Omnibus (GEO) (http://www.ncbi.nlm.nih.gov/sites/GDSbrowser) online veri tabanından sağlanan sekiz mikrodizilim gen ifade verisine uygulanmıştır. Analizler R- 2.14.0 (http://www.r-project.org/) yazılımından yararlanılarak yapılmıştır. Model sonuçları, belirtme katsayısı (coefficient of determination-R^2), hata kareler ortalamasının karekökü (root mean square error-RMSE) gibi model performans kriterleri ile karşılaştırılmıştır.

Bulgular: Çalışmada GEO'dan ulaşılan sekiz mikrodizilim gen ifade verisi ile modeller test edilmiştir. Çünkü orijinal veri setlerinin kullanımı yeni yöntemlerin test edilmesinde en çok tercih edilen yaklaşımlardandır. Çalışmadaki en önemli bulgu boosted regresyon ağacının (BRA), KAR'dan daha küçük kestirim hatasına sahip olmasıdır. Ayrıca ağaç sayısı ve etkileşim derecesi arttıkça RMSE değerleri giderek azalmaktadır.

Sonuç: Boosted Regresyon Ağaçları, klasik regresyon ağaçlarına göre kullanılan her veri seti için daha küçük varyans tahmini ile sonuçlanmıştır. Çalışmamızda BRA analizinin sadece ekolojik veriler

ve az sayıda açıklayıcı değişken içeren çalışmalarda değil (Hastie ve Leatwick, 2008) yüksek boyutlu gen ifade verileri için de kullanılabilir olduğu gösterilmiştir.

Anahtar kelimeler: Mikrodizilim Gen İfade Verisi, Karar Ağacı Regresyonu, Boosted Regresyon Ağacı

S05

LOJİSTİK AYRIŞTIRMADA RICHARD EĞRİSİNİN LİNK FONKSİYONU OLARAK KULLANILMASI

Mehmet Gürcan[1], **Mehmet Onur Kaya**[2], İlker Ercan[2]

[1]*Fırat Üniversitesi, Fen Fakültesi, İstatistik Bölümü, Elazığ*

[2]*Uludağ Üniversitesi, Tıp Fakültesi, Biyoistatistik AD, Bursa*

monurkaya@gmail.com

Amaç: Çalışmada Richard link fonksiyonları kullanılarak lojistik ayrımsamanın nasıl yapılacağı incelenmiştir. Diferansiyel denklemin çözümünden elde edilen lojistik bir eğrinin parametresi integral sabitidir. İntegral sabitinin doğrusal olasılık modeline bağlı olarak modele yerleştirilmesi yanı sıra diferansiyel denklemi oluşturan fonksiyonların seçiminde bulunması da mümkündür. Lojistik ayrımsamada kullanılan Richard link fonksiyonunda bu şekilde bulunan parametrenin hesaplanması amaçlanmıştır.

Yöntem: Büküm parametresi hatalı sınıflanmış gözlemler incelenerek bağımsız değişken değerleri ile ilişkilendirilmiş ve lojistik ayrımsama dinamik bir yapıya oturtturulmuştur. Yapılan uyarlama anksiyete bozukluğu olgularına ait 20, 40 ve 60 birimlik veri setlerinde değerlendirilerek yöntemin uygulanabilirliği vurgulanmıştır.

Bulgular: Bağımlı değişkenin ikili olduğu Lojistik modelde doğru sınıflama oranları 20, 40 ve 60 birimlik olgular için sırasıyla, %72,5, %66,3 ve %67,5 olarak bulunmuşken, Richard link fonksiyonu kullanıldığında, hatalı sınıflanan gözlemler için eğrinin büküm parametresinin uygun seçilebilmesiyle bu oranlar sırasıyla, %82,5, 78,75 ve % 80' e kadar çıkartılabilmiştir.

Sonuç: Analizden de görüldüğü gibi Richard link fonksiyonu dinamik bir yapıya sahiptir. Büküm parametresinin değiştirilebilmesi doğrusal olasılık modeline dokunulmaksızın yapılabilmekte ve bu sayede hatalı bulunan gözlemlerin içerisinden doğru sınıflama yapılabilmektedir. Büküm parametresinin bağımsız değişken değerlerine bağlı olarak hesaplanması olasılık değerini etkileyecek artımlarda link fonksiyonunun bu değişimi doğru bir şekilde tolere etmesini sağlar. Bu sayede lojistik ayrımsama modeli dinamik bir yapıya kavuşturulmuş olur.

Anahtar Kelimeler: Lojistik Model, Richard Model, Ayrımsama Sorunu.

S06

PROPORTIONAL ODDS REGRESYON MODELİ'NDE UYUM İYİLİĞİ TESTLERİNİN SİMÜLASYON İLE KARŞILAŞTIRILMASI

Gözde Ertürk[1], M.Yasemin Akşehirli Seyfeli[2], Gökmen Zararsız[2], Yaşar Sertdemir[1], Ferhan Elmalı[2]

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

[2]*Erciyes Üniversitesi Tıp Fakültesi Biyoistatistik ve Tıp Bilişimi Anabilim Dalı, Kayseri*

gozderturk@gmail.com

Giriş ve amaç: Proportional odds regresyon modeli, sıralı lojistik regresyon analizinde kullanılan bir yöntemdir. Kümulatif logit veya paralel-line modeli olarak da bilinmektedir. Uyum iyiliği, bağımlı değişkeni açıklamak için oluşturulan en iyi modelin etkinliğinin bir ölçüsünü göstermektedir. Uyum iyiliğinin belirlenmesinde Hosmer Lemeshow, Pulksteniz – Robinson, Lipsitz testleri gibi yöntemler kullanılmaktadır. Bu çalışmada Hosmer Lemeshow, Pulksteniz – Robinson, Lipsitz testlerinin performanslarının karşılaştırılması amaçlanmıştır.

Gereç ve Yöntem: Hosmer Lemeshow, Pulksteniz – Robinson, Lipsitz testlerinin performansları birim sayıları n= 100, 200, 300, 400, 800 olan veri setlerinde; tekrar sayısı r=1000 olmak üzere R paket programında simülasyon çalışması ile karşılaştırıldı. Hosmer Lemeshow test istatistiğinin onlu risk grupları yöntemi kullanıldı.

Bulgular ve Sonuç: Çalışmanın bulgularına göre Hosmer Lemeshow test yönteminin C8, C10 ve C12 test sonuçlarının tip 1 hataları birbirine çok yakın çıkarken, Lipsitz ve Pulksteniz-Robinson birim sayısı arttıkça tip 1 hataları artmaktadır. Sonuç olarak Proportional odds regresyon modelinde uyum iyiliğini belirlerken Hosmer Lemeshow test istatistiğinin kullanılması önerilebilinir.

Anahtar kelime: Hosmer Lemeshow test istatistiği, Pulksteniz – Robinson test istatistiği, Lipsitz test istatistiği

S07

COX REGRESYON MODELİ: ORANSAL HAZARD VARSAYIMININ İNCELENMESİ

Merve Gülşah Ulusoy[1], Soner Duman[2]

[1]*Ege Üniversitesi, Tıp Fakültesi, Biyoistatistik ve Tıbbi Bilişim Anabilim Dalı. Bornova, İzmir*

[2]*Ege Üniversitesi, Tıp Fakültesi, İç hastalıkları Anabilim Dalı. Bornova, İzmir*

mgulsahulusoy@hotmail.com

Amaç: Salgın hastalık ve kronik hastalıklara ilişkin verilerin incelenmesinde ve bu hastalıkları etkileyen faktörlerin belirlenmesinde Yaşam Çözümlemesi: Oransal Cox Modeli yaygın bir şekilde kullanılmaktadır. Modelin kullanılabilmesi için "oransal hazard (tehlike)" varsayımını sağlaması gerekmektedir. Uygulamalarda bu varsayım incelenmeden doğru olduğu kabul edilerek çözümlemeler yapılmaktadır. Ancak, varsayım bozulumu durumunda modelin kullanılması uygun olmamaktadır. Çalışmamızda, gerçek veri seti üzerinden varsayım kontrolünün yapılması amaçlanmaktadır.

Yöntem: Ege Üniversitesi Hastanesi veritabanında kayıtlı, ortalama 4 yıllık izlemde olan Otozomal Dominant Polikistik Böbrek Hastalık teşhisi almış hastalar retrospektif olarak incelenmiştir. Hastaların tanı konulduğu tarihten, böbrek ölümünün gerçekleştiği tarihe kadar olan sürede demografik ve laboratuar verileri çekilmiştir. Hastaların renal replasman tedavisine ihtiyaç duyduğu zaman "böbrek ölümü" olarak değerlendirilmiştir.

Bulgular: Cox modeli için yaş, cinsiyet, serum kreatinin (Scr), 1/Scr ve eGFR değişkenleri tanımlanmıştır. Model için anlamlı olan değişkenlerin oransal hazard varsayımı incelenmiştir. Varsayım değerlendirmesinde grafiksel yöntemler ve artık analizi kullanılmıştır. Grafiksel yöntemlerden log-(-log) grafiği çizilmiş ve Schoenfeld, Martingale ve sapma (deviance) artıkları ile de test edilmiştir.

Sonuç: Uygulamalarda, oransal hazard varsayımı incelenmesi yapılmadan Cox Analizi yapılmaktadır. Fakat bu durum yanlış sonuçlara yol açabilmektedir. Bunun için, incelenen hastalığın sağ kalım süresinde etki eden faktörler belirlendikten sonra, oransal hazard varsayımını sağlanıp sağlanmadığı incelenmelidir.

Anahtar Kelimeler: Yaşam Çözümlemesi, Cox Regresyon Modeli, Orantısız Hazard, Schoenfeld Artıkları, Martingale Artıkları, Sapma Artıkları, Otozomal Dominant Polikistik Böbrek Hastalığı

Kaynakça:

1. Terzi Y., Cengiz M.A., Bek Y., Cox Regresyon Modelinde Oransal Hazard Varsayımının Artıklarla İncelenmesi Ve Akciğer Kanseri Hastaları Üzerinde Uygulanması. Türkiye Klinikleri J Med Sci 2005, 25:770–775

2. Ata N., Karasoy D.S., Sözer M.T., Orantılı Tehlike Varsayımının İncelenmesinde Kullanılan Yöntemler Ve Bir Uygulama. Eskişehir Osmangazi Üniversitesi Müh. Fak. Dergisi C.XX, S.1, 2007
3. Yay M., Çoker E., Uysal Ö., Yaşam Analizinde Cox Regresyon Modeli Ve Artıkların İncelenmesi. Cerrahpaşa Tıp Dergisi 2007; 38: 139–145
4. Aitkin M., Francis B., Hinde J., Darnell R., Statistical Modelling in R. Oxford University Press Inc., New York, First Edition, 2009

S08

DOĞRUSAL REGRESYON MODELLERİNDE DEĞİŞKEN SEÇİM KRİTERLERİNİN PERFORMANSLARININ DEĞERLENDİRİLMESİ

Kıvanç Yüksel[1], Nur Özge Sezgin[1]

[1]*Ege Üniversitesi İlaç Geliştirme & Farmakokinetik Araştırma Uygulama Merkezi (ARGEFAR), Bornova, İzmir*

kivancyuksel@yahoo.com

Amaç: Bu çalışmada, hata teriminin normal dağılımdan sapmalar gösterdiği ve örneklem sayısının küçük olduğu durumda çoklu doğrusal regresyon (ÇDR) modellerinde kullanılan değişken seçim kriterlerinin (Akaike Bilgi Kriteri (AIC), Düzeltilmiş Akaike Bilgi Kriteri (AIC-c), Sawa Bayes Bilgi Kriteri (BIC), Schwarz Bayes Kriteri (SBC), Mallow Cp Kriteri (Cp), Modifiye edilmiş Mallow Cp Kriteri (MCp), Düzeltilmiş R^2 ($\mathbf{R^2_{adj}}$) performansları değerlendirilmiş ve karşılaştırılmıştır.

Yöntem: Seçim kriterlerini karşılaştırmak için Monte Carlo simülasyonu ile elde edilen verilere bağlı olarak oluşturulmuş ÇDR modelleri kullanılmıştır. Modellere ait hata terimleri normal dağılımdan sapmalar gösterecek şekilde, farklı çarpıklık ve basıklık değerleri için "Fleishman's Power Transformation" yöntemi ile elde edilmiştir. Değişken seçimi olası bütün regresyon modelleri (OBRM) prensibine göre gerçekleştirilmiştir.

Bulgular: Hata varyansının (s) 1 alındığı durumda en iyi performans gösteren değişken seçim kriterinin küçük örneklemlerde BIC, örneklem sayısı arttıkça SBC olduğu görülmektedir. En başarısız değişken seçim kriteri ise $\mathbf{R^2_{adj}}$'dir. Hata varyansı 2 alındığında AIC-c ve BIC' nin performansının örneklem sayısı ile birlikte arttığı ve performans sıralamasında en üstte yer aldıkları görülmektedir. Örneklem sayısının artmasıyla birlikte en düşük performansı sergileyen değişken seçim kriterinin $\mathbf{R^2_{adj}}$ olduğu gözlemlenmiştir. Hata varyansı arttırılıp 4 ve 6 olarak alındığında ise en iyi performans gösteren değişken seçim kriterinin $\mathbf{R^2_{adj}}$ olduğu görülmektedir. Örneklem sayısının artmasıyla birlikte en başarısız değişken seçim kriterinin ise SBC olduğu gözlemlenmiştir.

Sonuç: En yüksek ve en düşük performans gösteren değişken seçim kriteri hata varyansının değerine bağlı olarak değişkenlik göstermektedir. Hata varyansının 2, 4 ve 6 alındığı durumda örneklem sayısı arttıkça tüm değişken seçim kriterlerinin performanslarının arttığı gözlemlenmiştir. Buna ek olarak, hata varyansının artmasıyla değişken seçim kriterlerinin performansların da düşüş olduğu belirlenmiştir.

Anahtar Kelimeler: Doğrusal Regresyon, Değişken Seçimi, AIC, AIC-c, BIC, SBC, Cp, MCp, $\mathbf{R^2_{adj}}$

S09

AĞIRLIKLANDIRILMIŞ KAPLAN-MEIER YÖNTEMİNİN ZAMANA BAĞLI ROC ANALİZİNE UYARLANMASI

Deniz Sığırlı[1], İlker Ercan[1]

[1]*Uludağ Üniversitesi Tıp Fakültesi, Biyoistatistik Anabilim Dalı, Bursa*

sigirli@uludag.edu.tr

Amaç: Kaplan-Meier yöntemi zamana bağlı ROC eğrisi (Receiver Operating Characteristic curve) analizinde kullanılmakta olan bir yöntem olup, bu çalışmada ağırlıklandırılmış Kaplan-Meier yöntemi zamana bağlı ROC analizine uyarlanmıştır. Çalışmada, Kaplan-Meier ve ağırlıklandırılmış Kaplan-Meier yöntemlerinin uyarlandığı zamana bağlı ROC analizi yöntemlerinin performanslarının incelenmesi amaçlanmıştır. Ayrıca gerçek veri seti üzerinde uygulama yapılmıştır.

Yöntem: Simülasyon çalışması, farklı örneklem büyüklüklerinde, tanı testi sonucu ve yaşam süresi arasındaki farklı korelasyon düzeyleri için ve farklı tamamlanmamış olgu oranları için 1000 tekrar yapılarak gerçekleştirildi.

Bulgular: Bütün durumlar için, ağırlıklandırılmış Kaplan-Meier yönteminin zamana bağlı ROC analizine uyarlamasından elde edilen AUC (eğri altında kalan alan-area under the curve) değerlerinin, gerçek AUC değerine daha yakın olduğu ve diğer yönteme göre daha iyi sonuçlar verdiği gözlendi.

Sonuç: Tamamlanmamış olguların yer aldığı durumlarda ROC analizi çalışmalarında, ağırlıklandırılmış Kaplan-Meier yönteminin uyarlandığı zamana bağlı ROC analizi kullanılabilir.

Anahtar kelimeler: Kaplan–Meier; ROC eğrisi; Tanı testi

S10

BİLGİ KURAMI YAKLAŞIMI İLE BİLGİSAYARLI TOMOGRAFİK KORONER ANJİYOGRAFİNİN TANISAL DEĞERİNİN BELİRLENMESİ

Mustafa Kılıçkap[1], Ahmet Ergun Karaağaoğlu[2], Kerim Esenboğa[1], Çağlar Uzun[3], Çetin Atasoy[3], Deniz Kumbasar[1], Çetin Erol[1]

[1]*Ankara Üniversitesi Tıp Fakültesi Kardiyoloji AD, Ankara*

[2]*Hacettepe Üniversitesi Tıp Fakültesi Biyoistatistik AD, Ankara*

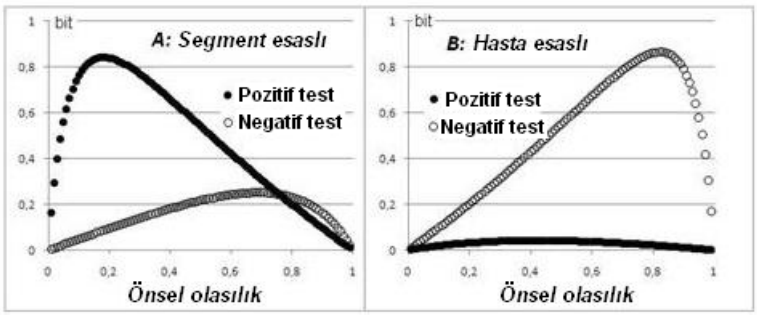
[3]*Ankara Üniversitesi Tıp Fakültesi Radyodiagnostik AD, Ankara*

mkilickap@yahoo.com

AMAÇ: Tanı testlerinin hangi önsel olasılık değerlerinde en fazla bilgi verdiğinin bilinmesi tanısal yaklaşım açısından önemlidir. Bilgi kuramı (*information theory*) yöntemi ile bir testin verdiği bilgi önsel olasılık değerlerine göre "bit" cinsinden nicel olarak belirlenebilmektedir. Bu çalışmada bilgisayarlı tomografik koroner anjiyografinin (BTKA) tanısal değerinin bilgi kuramı yaklaşımı ile değerlendirilmesi amaçlandı.

YÖNTEM: Konvansiyonel koroner anjiyografi altın standart test olarak alındı ve darlığın %50 ve üzerinde olması ciddi darlık olarak değerlendirildi. Testin genel performans ölçüsü olan bilgi içeriği (*mutual information*), önsel olasılığın 0,01-0,99 arasındaki değerlerinde, hem BTKA hem de duyarlılığı ve özgüllüğü 0,999999 olarak alınan mükemmel bir test için hesaplandı. Elde edilen değerlerin eğri altındaki alanlarının oranı (BTKA/mükemmel) BTKA'nın mükemmel bir testin sağlayabileceği bilginin ne kadarını sağlayabildiğini göstermede kullanıldı. Ayrıca testin pozitif ve negatif sonucunun sağladığı bilgiyi ayrı ayrı olarak belirlemek için göreli entropi (*relative entropy*) değerleri hesaplanarak grafiksel olarak ifade edildi. Değerlendirmeler hasta esaslı analiz (n=69) ve segment esaslı analiz (n=1076) olarak yapıldı.

BULGULAR: Segment esaslı analizde elde edilen bilginin hasta esaslı analize göre daha fazla olduğu saptandı (BTKA/mükemmel oranı sırasıyla %33 ve %15). Maksimum bilgi, intermedier arter (BTKA/mükemmel oranı %91) ve proksimal segmentler (BTKA/mükemmel oranı %39) için elde edildi. Göreli entropi değerlerine göre segment esaslı analizde pozitif BTKA sonucunun (Şekil 1A), hasta esaslı analizde ise negatif test sonucunun daha fazla bilgi verdiği görüldü (Şekil 1B). Ayrıca pozitif test sonucunun segment esaslı analizde önsel olasılığın yaklaşık %20 değerinde, negatif test sonucunun ise hem segment esaslı hem de hasta esaslı analizde önsel olasılığın yaklaşık %80 değerlerinde maksimum bilgi verdiği görüldü (Şekil 1A ve 1B).



Resim 1: Segment esaslı analiz (A) ve hasta esaslı analizde pozitif ve negatif BTKA sonucunun göreli entropi değerleri

SONUÇ: Göreli entropi grafikleri BTKA'nın hangi önsel olasılıkta daha fazla bilgi verdiğini görsel olarak göstermesi açısından değerlidir. Bu çalışmada hasta esaslı analizde negatif test sonucunun, segment esaslı analizde ise pozitif test sonucunun verdiği bilginin daha yüksek olduğu görüldü.

Anahtar Kelimeler: Bilgisayarlı tomografi; koroner anjiyografi; bilgi kuramı; göreli entropi

S11

ISOBOLOGRAM VE ETKİN DOZ TESPİTİ

Leman Tomak[1], Süleyman Sırrı Bilge[2], Yüksel Bek[1]

[1]*Ondokuz Mayıs Üniversitesi Tıp Fakültesi Biyoistatistik ve Tıbbi Bilişim A.D. Samsun*

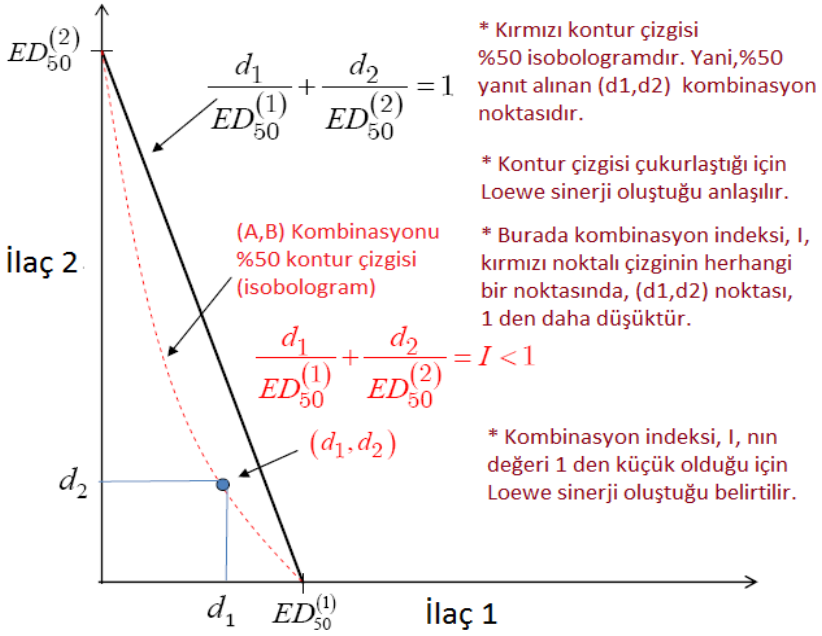
[2]*Ondokuz Mayıs Üniversitesi Tıp Fakültesi Farmakoloji A.D. Samsun*

lemant@omu.edu.tr

Amaç: Isobologram analizi kombine edilerek kullanılan ilaçların etkilerini görmek ve değerlendirmek için kullanılan bir istatistik yöntemdir. Bu çalışmanın amacı, iki ilacın bir arada kullanıldığı durumlarda ikisi arasındaki etkileşimin sinerjik ya da antogonist davranıp davranmadığını saptamaktır.

Yöntem: Isobologram analizinde iki ilacın etkileşiminin tabiatına uygun şekilde bir grafik ile gösteriş sağlanır. Kullanılan grafiksel yöntemle ilaç kombinasyonlarının etkileri saptanmış olur. Sinerji interaksiyon indeksi saptanır.

Bulgular: Şekil 1'de iki ilaç için elde edilen isobologram verilmiştir. Buna göre % 50 isobologramı veren kontur çizgisi çukurlaştığı için Loewe sinerjiden bahsedilebilir.



Şekil 1. Isobologram

Kombine ilaç sinerjisini düşündüğümüzde iki veya daha fazla ilacın bir araya getirilerek beklenenden daha yüksek yanıt alma amaçlanmaktadır. Diğer taraftan iki ilacın birleşiminden yanıt sinerjisi doğmasa da "doz indirgeme potansiyeli" ortaya çıkabilir. Diğer bir ifade ile ilaçları birleştirmekle daha

iyi yanıt beklemeyebiliriz ama her ilaçtan daha az kullanarak, tek olarak kullanıldıklarında sağlanan eşdeğer etkiyi elde edebiliriz.

Sonuç: Sonuç olarak birden fazla ilacın kullanıldığı ilaç doz denemelerinde, ilaçların tek tek etkileri karşılaştırıldığı gibi bir arada kullanımlarıyla oluşan etkileşimin (etkin doz, toksik doz, ölümcül doz gibi ED, TD, LD etkileri) saptanmasında da, sinerjik ve antagonistik etkilerin belirlenmesini ve doğru karar verilmesini sağlayan, bu etkili ancak çizimi zor olduğu için (özel paket programlar veya R-paketi ile çizilebilmekte) diğer yöntemler kadar sık kullanılmayan bir yöntem olan isobologram analizi irdelendi.

Anahtar Kelimeler: Isobologram, effective dose, dose response, drug synergism

Kaynaklar

Niyazi, M; Belka C. 2006. Isobologram Analysis of Triple Therapies. Radiation Oncology, 1:39.

Albert P. Li. 2007. Drug-Drug İnteractions in Pharmaceutical Development. John Wiley & Sons, Inc., Hoboken, New Jersey. Pp:245.

Ronald J. Tallarıda. 2001. Drug Synergism: Its Detection and Applications. The Journal of Pharmacology And Experimental Therapeutics, JPET 298:865–872, 2001

Ronald J. Tallarida, 2000. Drug Synergism and Dose-Effect Analysis. Chapman & Hall/CRC, Pp:247.

S12

FAKTÖR ANALİZİ İÇİN UYGUN ÖRNEK GENİŞLİĞİNİN BELİRLENMESİ

Ahmet Mollaoğulları[1], Mehmet Mendeş[2], Mehmet Koçak[3]

[1]*Çanakkale Onsekiz Mart Üniversitesi, Fen Edebiyat Fakültesi, Matematik Bölümü, Çanakkale*

[2]*Çanakkale Onsekiz Mart Üniversitesi, Ziraat Fakültesi, Zootekni Bölümü, Biyometri ve Genetik AD-Çanakkale*

[3]*Tennessee Üniversitesi, Halk Sağlığı Merkezi, Koruyucu Tıp Bölümü, Memphis, TN, USA*

ahmet_m@comu.edu.tr

Amaç: Bir araştırma ve denemelerden elde edilecek sonuçların güvenilirliği, dikkate alınan örnek genişliği ile oldukça yakından ilişkilidir. Dolayısıyla araştırma ve denemelerin yürütülmesine başlanmadan önce gerekli olan örnek genişliğinin uygun bir şekilde belirlenmesi gerekir. Bu simulasyon çalışmasında başta sosyal bilimler olmak üzere, fen ve sağlık bilimlerinde de yaygın olarak kullanılan Faktör Analizi için uygun örnek genişliğinin belirlenmesine çalışılmıştır.

Materyal ve Yöntem: Bu çalışmanın materyalini Fortran Powerstation Developer Studio programının IMSL kütüphanesinden yararlanarak çok değişkenli dağılımlardan üretilen tesadüf sayıları oluşturmaktadır. Uygun örnek genişliğinin belirlenmesi amacıyla Monte Carlo simulasyon tekniği kullanılmıştır.

Bulgular ve Sonuç: Yapılan simulasyon denemeleri sonucunda gerekli örnek genişliğinin dağılım şeklinden ziyade communality düzeyi, değişken sayısı, KMO değeri, açıklanan varyans ve dikkate alınan değişken sayısının belirlenen faktör sayısına oranından etkilendiği görülmüştür. Simulasyon çalışmaları sonucunda örnek genişliğinin, dikkate alınan değişken sayısının en az 7 katı kadar olması durumunda örneklerden elde edilen değerlerin populasyondan elde edilen değerlere oldukça yakın olduğu görülmüştür. Dolayısıyla Faktör Analizinden beklenilen yararların elde edilebilmesi için çalışılacak örnek genişliğinin en az dikkate alınan değişken sayısının 7 katı kadar olması gerektiği sonucuna varılabilir.

Anahtar Kelimeler: Faktör analizi, örnek genişliği, communality, KMO

S13

ÇOKLU FAKTÖR ANALİZİNİN TANITIMI VE BİR UYGULAMASI

Sıddık Keskin[1], Sadi Elasan[1]

[1]*Yüzüncü Yıl Üniversitesi, Tıp Fakültesi, Biyoistatistik Anablim Dalı, Kampüs-Van*

skeskin973@hotmail.com

Amaç: Bu çalışmada, aynı deney ünitelerinden veya nesnelerden elde edilen verileri içeren çoklu tabloları analiz etmede kullanılan istatistik yöntemlerinden birisi olan Çoklu Faktör Analizi (ÇFA) hakkında bilgi verilerek bir uygulama ile birlikte tanıtılması amaçlanmıştır.

Yöntem: ÇFA, aynı deney ünitelerinden ya da nesnelerden elde edilmiş verilerden oluşan çoklu tabloların, birleştirilerek analiz edilmesine olanak sağlayan çok değişkenli analiz yöntemlerinden birisidir. Yöntem, aynı zamanda "Çoklu Faktöriyel Analiz" olarak da bilinmekte olup, Temel Bileşenler Analizi'nin genişletilmiş hali olarak düşünülebilir. ÇFA; tıp, sağlık, genetik, istatistik kalite kontrolü, ekonomi, moleküler biyoloji, duyusal analizler, tüketici tercihleri, kemometri, süreç (proses) izleme, ekoloji, ziraat, meteoroloji-hava tahminleri ve jeoloji gibi bir çok alanda kullanılabilir. Yöntem, üç adımdan oluşur: Birinci adımda; veri tablolarının her birisine Temel Bileşenler Analizi uygulanır ve her tablodaki veriler, kendi birinci tekil değerine (birinci özdeğerin karekökü) bölünerek normalize edilir. İkinci adımda; her tablo birleştirilerek, genel yada global matris olarak adlandırılan tablo oluşturulur ve bu global matris, (tablo) Temel Bileşenler Analizine tabi tutulur. Üçüncü adımda; ortak iki boyutlu uzayda gösterilmek üzere, nesnelere ve değişkenlere ait faktör skorları hesaplanır.

Sonuç: Çoklu Faktör Analizi, ilgilenilen değişkenler için aynı deney ünitelerinden ya da nesnelerden elde edilmiş veri grupları ile çalışır. Alternatif olarak kullanılabilecek diğer yöntemlere göre kısmen basitliği ve sonuçların kolay yorumlanabilir olması, çoklu tabloların veya büyük veri setlerinin analizinde bu yöntemi çoğunlukla tercih edilebilir hale getirmektedir.

Anahtar kelimeler: Tekil değer ayrışımı, kanonik analiz, çoklu tablo, faktör analizi

S14

SİMÜLASYON ÇALIŞMALARINDA UYGUN SİMULASYON SAYISININ BELİRLENMESİ

Mehmet Koçak[1], Ahmet Mollaoğulları[2], **Mehmet Mendeş[3]**

[1]*Tennessee Üniversitesi, Halk Sağlığı Merkezi, Koruyucu Tıp Bölümü, Memphis, TN, USA*

[2]*Çanakkale Onsekiz Mart Üniversitesi, Fen Edebiyat Fakültesi, Matematik Bölümü Çanakkale*

[3]*Çanakkale Onsekiz Mart Üniversitesi, Ziraat Fakültesi, Zootekni Bölümü, Biyometri ve Genetik ABD Çanakkale*

ahmet_m@comu.edu.tr

Amaç: Aynı amaçla kullanılan testlerin performansları bakımından (1.tip hata ve testin gücü) karşılaştırılmasında, parametrik testlerin varsayımlarının yerine gelmemesinin elde edilecek sonuçlara etkilerinin değerlendirilmesinde, uygun örnek genişliğinin belirlenmesinde, yeni geliştirilen bir test istatistiğinin ampirik olarak örnekleme dağılımının oluşturulmasında, herhangi bir istatistik testi için kritik tablo değerlerinin elde edilmesinde ve matematiksel olarak çözümü çok zor ya da mümkün olmayan problemlerin ampirik olarak çözülmesinde simülasyon çalışmalarından oldukça yaygın bir şekilde yararlanılmaktadır. Literatürler incelendiğinde aynı amaçla yürütülen simülasyon çalışmalarında bile dikkate alınan simülasyon sayılarındaki farklılıklardan (1,000-1,000,000 arasında) dolayı elde edilen sonuçlar arasında oldukça büyük farklılıkların bulunduğu görülür. Halbuki, gerek elde edilecek sonuçların güvenilirliğinde gerekse de yapılacak parametre tahminlerinin güvenilir olması ve stabil kalmasında dikkate alınan simülasyon sayısının büyük etkisi vardır. Buna karşın birçok simülasyon çalışmasında dikkate alınan simülasyon sayısının oldukça düşük düzeylerde (1,000-5,000 arasında) tutulduğu dikkati çekmektedir. Dikkate alınacak simülasyon sayısı; çalışmanın amacına ve istenilen hata marjinine göre farklılık gösterir. Bundan dolayı her koşul için gerekli olan simülasyon sayısının aynı olması beklenmez. Bu çalışma; Tek Yönlü Varyans Analizi için 1.tip hata olasılıklarının tahmin edilmesi, Pearson korelasyon katsayısı için 1.tip hata olasılıklarının tahmin edilmesi ve kritik tablo değerlerinin elde edilmesi ve Cox-Oransal Hazard Modellerinde Hazard Oranlarının tahmin edilmesi için gerekli olan minimum simülasyon sayılarının elde edilmesi amacıyla yürütülmüştür.

Materyal ve Yöntem: Bu çalışmanın materyalini Fortran Powerstation Developer Studio programının IMSL kütüphanesinden yararlanarak çok değişkenli dağılımlardan üretilen tesadüf sayıları oluşturmaktadır. Uygun simülasyon sayısının belirlenmesi amacıyla Monte Carlo simülasyon tekniği kullanılmıştır.

Bulgular ve Sonuç: Yapılan simülasyon denemeleri sonucunda genel olarak 1.tip hata olasılıkları için uygun simülasyon sayısının 50.000, kritik tablo değerlerinin elde edilmesi için ise 350.000 olması gerektiği görülmüştür. Hazard oranlarının elde edilmesi için gerekli simülasyon sayısının ise aynı

zamanda dikkate alınan örnek genişliğine de bağlı olmak koşuluyla genel olarak 40.000'in altına düşürülmemesi gerektiği görülmüştür.

Anahtar Kelimeler: Simülasyon sayısı, 1.tip hata, testin gücü, ANOVA

S15

YAPAY MADDE İŞLEV FARKLILIĞINA NEDEN OLAN FAKTÖRLERİN BENZETİM ÇALIŞMASI İLE DEĞERLENDİRİLMESİ

Pervin Demir[1], Selcen Yüksel[1], Atilla Halil Elhan[2], S. Yavuz Sanisoğlu[1], Mesut Akyol[1]

[1]*Yıldırım Beyazıt Üniversitesi Tıp Fakültesi Biyoistatistik ve Tıbbi Bilişim AD, Ankara*

[2]*Ankara Üniversitesi Tıp Fakültesi Biyoistatistik AD, Ankara*

pervin.demr@gmail.com

Giriş ve Amaç: Ölçme, genel anlamıyla herhangi bir niteliği gözlemek ve gözlem sonuçlarını sayılarla veya başka sembollerle ifade etmektir. Ölçek, belirli bir özelliği ölçmek üzere geliştirilmiş, psikometrik özellikleri (geçerlik, güvenirlik, değişime duyarlılık) belirlenmiş ölçme aracı olarak tanımlanabilir. Ölçekler özellikle sağlık, eğitim ve sosyal bilimler alanlarında doğrudan ölçüm yapılamayan, gizli (latent) özelliklerin ölçülmesinde kullanılır. Zekâ, yetenek ve özürlülük gibi özellikler soyut, gizli özelliklerdir. Bireylerin gizli özelliklerinin ölçüldükten sonra sınıflandırılmalarını sağlayan modellerden biri Rasch Modelidir (RM). Rasch modeli varsayımlarından biri ölçeği oluşturan maddelerin tümünün tek bir özelliği ölçmesi yani ölçülen özelliğin tek boyutlu olmasıdır. Eğer tek boyutluluk sağlanmıyorsa doğru ölçüm yapılamadığı varsayılır. Bu varsayımın bozulmasına neden olan faktörlerden biri maddelerin ölçeğin uygulandığı bireylerin oluşturduğu alt gruplarda işleyişinin farklı olmasıdır. Literatürde bu soruna Madde İşlev Farklılığı (MİF) denir. Başka bir deyişle MİF, aynı yetenek düzeyine sahip olan ancak farklı gruplardan gelen bireylerin bir maddeye yanıt verme olasılıklarının farklı olmasıdır. Eğer bir madde gözlenen bir grup açısından MİF gösteriyorsa, bu gözlenen grubun alt kategorileri için kestirilen madde zorlukları arasındaki mutlak fark MİF büyüklüğü olarak tanımlanmaktadır. Rasch analizi sonucunda eğer bir madde belirli bir grup açısından MİF gösteriyorsa bu sorunu gidermek için öngörülen çözümlerden biri belirlenen bu maddeyi ölçekten çıkarmaktır. Diğer bir çözüm ise, MİF' e neden olan gözlenen grubun alt gruplarına göre madde parametresini ayrı ayrı kestirerek her birey için doğru yanıt olasılıklarını bu parametreleri kullanarak hesaplamaktır. Bu Rasch çözümlemesine MİF analizi denir. Bir ölçekte yer alan maddelerin birden fazlasında MİF olduğu varsayılsın. En güçlü MİF saptanan maddede alt gruplara göre ayrı parametre kestirimi yapıldığında, diğer madde veya maddelerdeki MİF ortadan kalkıyorsa bu durumda yapay MİF' den şüphelenilir. MİF gösteren maddeler içinden MİF büyüklüğü en yüksek olan madde gerçek MİF olan madde iken, alt gruplara göre yeniden kestirim yapıldığında MİF' i ortadan kalkan madde(ler) yapay MİF olarak adlandırılır. Literatürde MİF büyüklüğünün yapay MİF' in oluşmasına etki eden bir faktör olduğunu öngören çalışmalar vardır.

Bu çalışmanın amacı, yapay MİF'e neden olabilecek faktörlerin bir benzetim çalışması ile değerlendirilmesidir. Bu amaca yönelik MİF büyüklüğü ve ölçekte MİF gösteren madde sayısı benzetim faktörleri olarak ele alınmıştır.

Yöntem: Sağlık alanındaki uygulamalarda maddelerin %30'undan fazlasının MİF gösterme olasılığı düşük olduğu için bu çalışmada sadece %10 ve %30 oranları dikkate alınarak çalışma yürütülecektir. MİF büyüklüğü de 0.5 (düşük), 1 (orta) ve 2 (yüksek) olarak alınacaktır. Sabit benzetim faktörleri ise örneklem büyüklüğü olarak 300 ve ölçekteki madde sayısı 20 olarak öngörülmüştür. Sağlık alanında büyük örneklemler ile çalışmak her zaman mümkün olmadığı için bu çalışmada 300 bireyin iki sonuçlu yanıt desenleri türetilecektir. Benzetim faktörlerin kombinasyonlarından elde edilen altı farklı koşul için veri türetilecektir (Tablo 1). Her bir koşul için 1000 Monte Carlo (MC) tekrarı yapılacaktır.

Tablo 1. Benzetim koşulları

Ölçme aracındaki madde sayısı	**MİF gösteren madde yüzdesi**	**MİF büyüklüğü**	**Örneklem büyüklüğü**
20	%10	0.5	300
20	%10	1	300
20	%10	2	300
20	%30	0.5	300
20	%30	1	300
20	%30	2	300

Anahtar Kelimeler: Rasch model, yapay madde işlev farklılığı, iki sonuçlu madde

S16

DRUG/NON-DRUG CLASSIFICATION USING SUPPORT VECTOR MACHINES WITH VARIOUS FEATURE SELECTION STRATEGIES

Selçuk Korkmaz[1], Gökmen Zararsız[1], Erdal Coşgun[1], Sevim Dalkara[2], Osman Saraçbaşı[1]

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

[2]*Hacettepe University Faculty of Pharmacy Department of Pharmaceutıcal Chemistry, Ankara*

selcukorkmaz@hotmail.com

Objective: In drug discovery, it is very important to find a promising molecule that could become a drug in the future. Virtual screening of new drugs relies heavily upon various filters aimed at retaining promising drug-like compounds while throwing away those unlikely to be drugs. Thus, fast classification methods are developed for identification of potentially undesired molecules in very large compound collections.

Methods: In this study, we used support vector machines (SVM) for classification task. SVM is a powerful classification tool that is becoming increasingly popular in various machine learning applications. Our data set consists 34 variables belonging to 631 compounds in which 311 drug-like and 320 nondrug-like compounds for training set. A different set of 98 drug-like and 118 nondrug-like compounds collected for validation set. In pre-processing step, we applied logarithmic transformation in order to rectified skewness. Then, a Pearson correlation matrix was computed for the entire pairs of variables. For all cross terms that were either strongly positive correlated ($r>0.90$) or strongly negative correlated ($r<-0.90$) with each other, we have dropped one variable based on t-test. At the end, we used 11 variables to build ordinary SVM model. We have also investigated the performance of SVM with different feature selection strategies.

Results: We compare the performance of SVM with and without feature selection. Since feature selection removes irrelevant features and increases efficiency of learning tasks it is a useful technique for increasing classification accuracy.

Conclusion: We tested SVM as a classification tool in a real life drug discovery problem and results revealed that it can be a useful method for classification task in virtual screening.

Keywords: Support Vector Machines, Drug Discovery, Virtual Screening.

References

García-Sosa, A. T., Oja, M., Hetényi, C., & Maran, U. (2012). DrugLogit: logistic discrimination between drugs and nondrugs including disease-specificity by assigning probabilities based on molecular properties. Journal of chemical information and modeling, 52(8), 2165-80.

Byvatov, E., Fechner, U., Sadowski, J., & Schneider, G. (2003). Comparison of support vector machine and artificial neural network systems for drug/nondrug classification. Journal of Chemical Information and Computer Sciences, 43(6), 1882-1889. American Chemical Society.

Zernov, V. V., Balakin, K. V., Ivaschenko, A. A., Savchuk, N. P., & Pletnev, I. V. (2003). Drug discovery using support vector machines. The case studies of drug-likeness, agrochemical-likeness, and enzyme inhibition predictions. Journal of Chemical Information and Computer Sciences, 43(6), 2048-2056.

S17

BOOKSTEIN KOORDİNATLARINDA REFERANS LANDMARKLARIN DEĞİŞMESİ DURUMUNDA HOTELLING T^2 TEST SONUCUNDAKİ DEĞİŞİMİN İNCELENMESİ

Güven Özkaya[1], Deniz Sığırlı[1], İlker Ercan[1]

[1]*Uludağ Üniversitesi, Tıp Fakültesi, Biyoistatistik AD., Bursa*

guvenozkaya@gmail.com

Amaç: Pek çok biyolojik ve biyomedikal araştırmalarda, bir bütün olarak biyolojik organların ya da organizmaların formlarının analiz edilmesinde en etkili yol, landmark noktalarının geometrik konumlarının kaydedilmesidir. Eğer şekiller karşılaştırılmak isteniyorsa, birimlerin standart bir pozisyon ve yöne ötelenmesi ve döndürülmesi gerekmektedir. Ayrıca, birimler standart bir büyüklüğe ölçeklenmelidir. Bunu gerçekleştirmek için Bookstein, referans olarak iki landmarkın seçilmesini önermiştir. Böylece her bir birim bu referans landmarklara göre ötelenir, döndürülür ve ölçeklenir. Çalışmamızın amacı, Bookstein koordinatlarını temel alarak iki grup şekil konfigürasyonlarının karşılaştırmasında yaygın olarak tercih edilen Hotelling T^2 testi kullanıldığında, farklı baseline landmarkların seçimi durumunda p değerlerindeki değişimin incelemesine yöneliktir.

Yöntem: Bu amaçla çalışmamızda farklı landmark sayılarından oluşan şekil konfigürasyonlarında mümkün olan tüm ikili landmark kombinasyonları baseline olarak alınarak farklı varyans düzeylerinde ve örneklem büyüklüklerinde p değerlerindeki değişim incelendi.

Bulgular: Çalışmamızda, landmark sayısındaki artışla birlikte seçilebilecek referans landmark kombinasyonu da arttığı, bundan dolayı da p değerlerinde oldukça farklılıklar meydana geldiği görüldü.

Sonuç: Bu nedenle Bookstein koordinatlarının kullanılması durumunda referans landmark olarak hangi landmarkların alınması gerektiği önemli bir sorundur.

Anahtar Kelimeler: Landmark, Bookstein koordinatları, Hotelling T^2 testi, İstatistiksel Şekil Analizi

S18

İSTATİSTİKSEL ŞEKİL ANALİZİNDE KULLANILAN İKİ ÖRNEKLEM TESTLERİNDE LANDMARK SAYISINA BAĞLI OLARAK TİP I HATA DÜZEYİNİN İNCELENMESİ

Gökhan Ocakoğlu[1], İlker Ercan[1]

[1]*Uludağ Üniversitesi Tıp Fakültesi Biyoistatistik Anabilim Dalı, Bursa*

ocakoglu@uludag.edu.tr

Amaç: Tıp alanında yapılan araştırmalarda çıkarsamalar için yapılan istatistiksel analizlerde veri setleri kantitatif veya kalitatif ölçüm değerlerinden oluşurken, günümüzde görüntüleme tekniklerindeki gelişmeyle bir organın veya organizmanın görüntüsü veya şekli de veri girdisi olarak kullanılmaya başlamıştır. Görüntüleme tekniklerindeki gelişimle birlikte modern geometrik morfometrik yöntemler metodolojisindeki ve uygulamalarındaki ilerlemeler de dikkate değer düzeydedir. Geometrik morfometride iki örnekleme ait şekil karşılaştırılmasının yapılması için kullanılan testler mevcut olup bunların bir kısmı oldukça yenidir. Çalışmamızda, istatistiksel şekil analizinde iki örnekleme ait şekil karşılaştırmasında kullanılan testlerin, çalışmada alınan landmark sayısına bağlı olarak tip I hata düzeylerinin incelenmesi amaçlanmaktadır.

Yöntem: Çalışmada, istatistiksel şekil analizinde iki örnekleme ait şekil karşılaştırmalarında kullanılan Hotelling T^2, James F_J ve Goodall F testleri için farklı landmark sayısı (k=3, 4, 5, 6, 7, 8) ve örneklem genişliklerinde (n=20, 50, 100, 500), isotropik modelleme altında düşük ve yüksek varyans dizeyleri ($\sigma^2 = 0.001$, $\sigma^2 = 737$) için ve şekil uzayı olarak tanjant uzayının kullanıldığı durum dikkate alınarak tip I hata oranları incelenmiştir. İlgili testlerin klasik(tabular) versiyonlarına ek olarak permutasyon ve bootstrap versiyonları da dikkate alınarak landmark sayısının tip I hata düzeyi üzerindeki etkisini belirlemeye yönelik olarak simülasyon çalışması yapılmıştır.

Bulgular: Testlerin klasik ve permutasyon uygulamalarında tip I hata düzeyinin belirlenen nominal düzeye yakın sonuçlar verdiği görülmektedir. Bootstrap yönteminde ise belirlenen nominal düzeyde farklılıklar gözlenmektedir.

Sonuç: Bootstrap yönteminde, örneklem büyüklüğünün az olması durumunda ve landmark sayısının artışı ile birlikte tip I hata düzeyi belirlenen nominal düzeyin altında sonuçlar vermektedir.

Anahtar Kelimeler: İstatistiksel şekil analizi, iki örneklem testleri, tip I hata oranı.

S19

ÇOK BOYUTLU SAĞKALIM VERİLERİNDE DENETİMLİ TEMEL BİLEŞENLER ANALİZİNE ALTERNATİF BİR BOYUT İNDİRGEME YAKLAŞIMI

Elvan Aktürk Hayat[1], Mevlüt Türe[2], Şanslı Şenol[3]

[1]*Sinop Üniversitesi, Fen-Edebiyat Fakültesi, İstatistik Bölümü, Sinop*

[2]*Adnan Menderes Üniversitesi, Tıp Fakültesi, Biyoistatistik ve Tıbbi Bilişim Anabilim Dalı, Aydın*

[3]*Ege Üniversitesi, Fen Fakültesi, İstatistik Bölümü, İzmir*

elvanakturk@gmail.com

Amaç: Bu çalışmada, boyut indirgemede kullanılan denetimli temel bileşenler analizi (D-TBA) ile bu yönteme alternatif bir yaklaşım olarak önerilen sağkalım ağacıyla gen seçerek uygulanan, yapay sinir ağlarıyla doğrusal olmayan temel bileşenler analizinin (Sağkalım ağacı temelinde YSA-DOTBA) performanslarının karşılaştırılması amaçlandı.

Yöntem ve gereç: Çalışmada Rosenwald ve diğerlerinin (2002) çalışmasından elde edilen yaygın B-hücreli lenfoma (DLBCL-diffuse large B-cell lymphoma) hastası 240 bireye ilişkin gen verileri kullanıldı. D-TBA'da çok boyutlu gen ekspresyon verilerinden önemli genlerin belirlenmesinde Cox skorlar kullanılırken, Sağkalım ağacı temelinde YSA-DOTBA yaklaşımında önemli genlerin belirlenmesinde sağkalım ağacının önemlilik değerleri kullanıldı. Cox skorlara göre önemli olduğu belirlenen genler tekil değer ayrışması ile 3 temel bileşene indirgendi. Sağkalım ağacıyla önemli bulunan genler YSA'da girdi değişkeni olarak alınarak, 3 temel bileşene indirgendi. D-TBA ve sağkalım ağacı temelinde YSA-DOTBA'nın performansları Cox regresyon modeli (CRM) ile karşılaştırıldı. Elde edilen Cox regresyon modellerini karşılaştırmak için Harrell'in C indeksi hesaplandı.

Bulgular: Cox skorlara göre 121 adet genin, sağkalım ağacının önemlilik değerlerine göre 114 adet genin önemli genler olduğu belirlendi. D-TBA'nın varyans açıklama oranı %18.2, sağkalım ağacı temelinde YSA-DOTBA'nın varyans açıklama oranı %35.1 bulundu. Harrell'in C indeksi CRM-1 için 0.726, CRM-2 için 0.687 olarak hesaplandı.

Sonuç: Sonuç olarak D-TBA, sadece doğrusal ilişkileri göz önüne alırken, sağkalım ağacı temelinde YSA-DOTBA, doğrusal olmayan ilişkileri de dikkate alması ve daha fazla varyans açıklama oranına sahip olması açısından D-TBA'ya alternatif bir yöntem olarak değerlendirilmelidir.

Anahtar kelimeler: Boyut indirgeme, Denetimli temel bileşenler analizi, Sağkalım ağacı, Sinir ağları, Cox regresyon analizi, Gen ekspresyon verisi

S20

BREIMAN ALGORİTMASI İLE HOMOJEN ALT GRUPLARIN BELİRLENMESİ: BİR UYGULAMA

Özge Akşehirli[1], Handan Ankaralı[1], Şengül Cangür[1], Mehmet Ali Sungur[1]

[1]*Düzce Üniversitesi Tıp Fakültesi, Biyoistatistik ve Tıbbi Bilişim Anabilim Dalı, Düzce*

ozge_yilmaz85@hotmail.com

Amaç: Çalışmada, danışmansız öğrenme tekniklerinden biri olan kümeleme yöntemine ait, Breiman algoritmasının işleyiş adımlarının tanıtılması ve gerçek bir araştırmadan elde edilen veri seti üzerinde uygulama adımlarının gösterilmesi, sonuçlarının yorumlanması amaçlanmıştır.

Yöntem: Breiman, verideki kümeleri bulmanın yoğunluk tahmini ile ilişkili olabileceğini ve birçok verinin birbirine yakın olarak toplandığı "yüksek yoğunluklu" alanları bularak verilerin kümelenebileceğini söylemiştir. İşlemlerin ilk adımında, gerçek veri seti kopyalandıktan sonra bu kopya set üzerinde bütün örüntülerin yok olduğu bir yapay veri seti elde edilir. Örüntüler, veri setindeki her bir değişken değerlerinin her defasında diğer değişkenler yok varsayılarak rasgele dağıtımıyla yok edilir. Böylece yapay veri setinde, değişkenler arasındaki bağımlılıklar ortadan kaldırılmış olur. Bu aşamada problem, gerçek veri seti ve yapay veri setinden oluşan iki sınıf problemi olarak ele alınır. Gerçek veri setinde kümelenme olup olmadığı incelemek için, ayırma analizi, lojistik regresyon, CART, RF, vb. sınıflama yöntemlerinden birisi kullanılabilir. Eğer bu iki veri seti başarılı bir şekilde ayrılabiliyorsa, gerçek veri setinde kümelenme yapılarının olduğundan bahsedilir. Çalışmanın uygulama bölümünde, hastaneye gece yeme sendromu şikâyetiyle başvuran 433 kişiye ilişkin sosyo-demografik ve klinik özelliklerden yararlanarak veri setinde farklı yapıda kümelenmeler olup olmadığı Breiman kümeleme algoritmasıyla incelenmiştir.

Bulgular: Gece yeme alışkanlığının sorgulandığı araştırmadan elde edilen gerçek ve yapay veri setleri, CART algoritması ile sınıflandırılmıştır. Optimum ağaçta toplam 31 karar noktası bulunmuş ve bunların 14' ünde yer alan denekler kendi içinde kümelenme göstermiştir. Geriye kalan 17 karar noktasında ise herhangi bir kümelenmenin olmadığı belirlenmiştir. Çalışmada yer alan toplam 433 kişiden 350'si oluşturulan 14 küme içine girmiştir. Geriye kalan 83 kişinin dağılımının ise ölçülen veya sorgulanan özellikleri bakımından rasgele bir dağılım gösterdiği belirlenmiştir. 350 kişiden 273 (%78)'ü klinik olarak gece yeme alışkanlığı yoktur tanısı; 77 (%22)'si ise gece yeme alışkanlığı vardır tanısı almıştır. Elde edilen 14 kümenin 12'sinde yer alan kişilerin ağırlıklı olarak gece yeme alışkanlığı yok tanısı alanlardan oluştuğu ve bu sonuca göre, bu veri setinden elde edilen kümelerin, genel olarak gece yeme alışkanlığı olmayan bireyleri ayırt edebildiği söylenebilir.

Sonuç: Hedef veya bağımlı değişkenin bilinmediği durumlarda veri setinde var olan homojen alt grupların belirlenebilmesi için, değişkenlerin dağılım şekli ve tipinden etkilenmeyen ve danışmansız öğrenme teknikleri kapsamında ele alınan kümeleme analizinden etkin bir şekilde yararlanılabilir. Bu çalışmada incelenen Breiman algoritması, kümeleme amacıyla kullanılan oldukça yeni bir algoritma olup mevcut kümeleme yöntemlerinin birçok dezavantajlı noktasını ortadan kaldırmakta ve veri analizi sonucunda detaylı bilgiler sunmaktadır.

Anahtar kelimeler: Veri madenciliği, danışmansız öğrenme, kümeleme analizi, Breiman algoritması, CART

S21

KÜMELEME ANALİZİNDE KENDİNİ DÜZENLEYEN AĞAÇ ALGORİTMASI

İlker Etikan[1], Osman Demir[1], Yunus Emre Kuyucu[1], Gökmen Zararsız[2], Osman Saraçbaşı[2]

[1]*Gaziosmanpaşa Üniversitesi Tıp Fakültesi Biyoistatistik Anabilim Dalı, Tokat*

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

ietikan@gmail.com

Amaç: Veri madenciliği içerisinde geniş bir kullanım alanı bulan sinir ağları birçok alanda kullanılır hale gelmiştir. Özellikle sağlık alanında kümeleme analizleri ile yapılan uygulamalarda hastalıklarla ilişkili değişken kümelerinin elde edilmesi ve hastalık alt sınıflarının belirlenmesi gibi problemler ile ilgili birçok kullanımı vardır. Kendini düzenleyen ağaç algoritması [1] (Self-Organizing Tree Algorithm-SOTA) 1997 yılında Dopazo ve arkadaşları tarafından geliştirilen ve kendini düzenleyen haritalar [2] (Self-Organizing Maps-SOM) ile büyüyen hücre yapıları [3] (Growing Cell Structures) yöntemlerinin üstün yanlarını birleştirerek kümeleme yapan hibrit bir sinir ağı algoritmasıdır. Bu çalışmada SOTA algoritması ve sağlık alanında bir uygulaması gösterilecektir.

Yöntem: Çalışmada UCI Machine Learning veri ambarından elde edilen Wisconsin meme kanseri veri seti kullanıldı. Veri seti 569 gözlem ve 32 değişkenden oluşmaktadır. Veri setindeki değişkenlerin birim farklılıklarını gidermek için veri setine önce z standartlaştırması uygulandı. Daha sonra, sınıf etiketleri saklanarak veri seti SOTA algoritması 2 kümeye ayrıldı. Performans karşılaştırması için ayrıca k-means, pam, som algoritmaları kullanıldı. Performans ölçümü olarak ise kümeleme sonuçlarının, gerçek sınıf etiketleri ile uyumunu ölçen Rand, düzeltilmiş Rand ve diğer dışsal geçerlilik ölçüleri kullanıldı. Ayrıca elde edilen kümeleme sonuçlarının kalitelerinin değerlendirilmesi için bağlantısallık (connectivity), Dunn, Silhoulette içsel ölçümleri ile üst üste olmayan ortalama oranı (average proportion of non-overlap- APN), ortalama uzaklık (average distance-AD), ortalamalar arasındaki ortalama uzaklık (average distance between means-ADM) ve değer şekli (figure of merit-FOM) kararlılık ölçümlerinden yararlanıldı. Verilerin analizi R 3.0.0 yazılımının clValid, validator paketleri ile yapıldı.

Bulgular: Uzaklık ölçülerinin kullanıldığı kümeleme yöntemleri değişkenler arası birim farklılıklarından etkileneceğinden öncelikle veri standartlaştırılmıştır. SOTA algoritması için en iyi bağlantısallık (64,1004) ve Dunn (0,0849) ölçülerine göre küme sayıları 2 ve 4 olarak önerilmiştir. Pam algoritması için Silhouette ölçüsüne göre ise küme sayısı 2 olarak önerilmiştir. SOTA algoritması için en iyi APN (0,0101) ve ADM (0,0717) değerine göre küme sayısı 2 önerilmiştir. Som algoritmasına için ise en iyi AD (5,0410) ve FOM (0,7143) değerine küme sayısı 6 önerilmiştir. SOTA algoritması için en iyi BHI (0,8988) ve BSI (0,8391) değerlerine göre küme sayısı 6 ve 2 önerilmiştir.

SOTA algoritması için Rand (0,8630), düzeltilmiş Rand (0,7248) değerleri, kmeans, som ve pam algoritmalarından elde edilen değerlerden daha iyi olduğu gösterilmiştir.

Sonuç: Genel olarak SOTA daha yüksek dereceli ilişkiler yansıtan bir topoloji oluşturmaktadır. Ayrıca gürültüye daha dayanıklıdır ve daha iyi işleyiş süresi özellikleri vardır. Ayrıca esnek bir kümeleme kavramı bulunmaktadır. Bu çalışmada kullanılan SOTA algoritmasının kullanımının ön plana çıkarılması amaçlanmıştır. Literatürde özellikle biyoenformatik verilerinin kümelenmesinde [4] daha çok kullanılan bu algoritmanın bu veriler dışında klinikte elde edilen verilerde de kullanılabilmektedir.

Anahtar Kelimeler: Kümeleme, kendini düzenleyen ağaç algoritması, kendini düzenleyen haritalar, büyüyen hücre yapıları.

Kaynaklar

[1] Dopazo, J. and J. M. Carazo (1997). "Phylogenetic Reconstruction Using an Unsupervised Growing Neural Network That Adopts the Topology of a Phylogenetic Tree." Molecular Evolution **44**: 226-233.

[2] Kohonen, T. (1990). "The Self-Organizing Map." Proceedings of IEE **78**(9): 1464-1480.

[3] Fritzke, B. (1994). "Growing Cell Structures- a self-organizing network for unsupervised and supervised learning." Neural Networks **7**: 1141-1160.

[4] Wang, H. C., J. Dopazo, et al. (1998). "Self-Organizing Tree Growng Network for Classifying Amino Acids." Bioinformatics Applications Note **14**(4): 376-377.

S22

PARAMETRİK OLMAYAN KOVARYANS ANALİZİNDE KULLANILAN YAKLAŞIMLAR

Sengül Cangür[1], Mehmet Ali Sungur[1], Handan Ankaralı[1]

[1]*Düzce Üniversitesi Tıp Fakültesi, Biyoistatistik ve Tıbbi Bilişim Anabilim Dalı, Düzce*

sengulcangur@duzce.edu.tr

Amaç: Parametrik kovaryans analizi (ANCOVA) varsayımlarının sağlanamaması ve/veya bağımlı değişkenin iki değerli/sıralayıcı ölçekli olması durumunda, parametrik olmayan ANCOVA yaklaşımlarından yararlanılmaktadır. Parametrik olmayan ANCOVA metodolojisinde, Quade, Puri & Sen, Conover & Iman, Rutherford ve McSweeney & Porter metotları, Ranklı ANCOVA yöntemleri olarak bilinmektedir. Ancak yaygın kullanılan programlarda, bu metotların uygulanmasına yönelik modül(ler) bulunmamaktadır. Bu çalışmada, Ranklı ANCOVA yaklaşımları tanıtılarak, bu yaklaşımların avantajlarından bahsedilmiştir. Ayrıca bir veri seti üzerinde uygulaması yapılmıştır.

Yöntem: Ranklı ANCOVA yaklaşımlarının teorik özellikleri tanımlanmış ve her bir yaklaşımın uygulanmasına yönelik web tabanlı bir program oluşturulmuştur. Ayrıca simule veriler üzerinde uygulamasına da yer verilmiştir.

Bulgular: Her ne kadar, bu çalışmada açıklanan yaklaşımlar için yaygın kullanılan istatistik programlarında özel bir modül olmasa da, bu yaklaşımların kolaylıkla uygulanabileceği gösterilmiş ve avantajlı yönleri ortaya konulmuştur.

Sonuç: Birçok araştırmada örneklem genişliğinin kısıtlı olması ve/veya bağımlı değişkenin normal dağılım göstermemesi durumunda, faktöriyel modeller için parametrik yöntemlerin kullanılması, Tip I hatanın artmasına ve testin gücünün azalmasına neden olmaktadır. Bu hatayı azaltmak için çalışmada önerilen yaklaşımların kullanılması tavsiye edilmektedir. Bu yaklaşımların, web tabanlı program sayesinde de yaygın kullanıma ulaşacağı düşünülmektedir.

Anahtar Kelimeler: Ranklı ANCOVA, Quade, Puri & Sen, Conover & Iman, Rutherford, McSweeney & Porter

S23

PARAMETRİK OLMAYAN KOVARYANS ANALİZİ

Şirin Çetin[1], S. Kenan Köse[1], Beyza Doğanay Erdoğan[1]

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

sirincetin@ankara.edu.tr

Bu çalışmada Kovaryans analizinin (ANCOVA) varsayımları sağlanmadığında parametrik olmayan kovaryans analizi için çözüm önerileri sunulmuştur. Bu çözüm önerilerinden Rank Transformasyonu, Randomization-based Yöntemi ve Robust Linear Model analizi yöntemlerinin incelenmesi amaçlanmıştır.

Kovaryans Analizinin varsayımları sağlanmadığında; Koch ve ark. (1982,1990) parametrik olmayanANCOVA'nın çözümünü; Quade(1967) 'nın önerdiği rank ANCOVA ile Mantel-Haenszel istatistiğini birleştirilerek tanımlamışlardır. Bu yöntem kovaryet etkilerini düzelttikten sonra gruplar arasındaki karşılaştırmayı Mantel-Haenszel istatistiğini kullanarak yapar. Bu yöntem bir ya da daha fazla kovaryet olduğunda gruplar arasındaki karşılaştırmayı kolayca yapabilmektedir. Randomazition-based yöntemi Parametrik olmayan kovaryans analizine Koch ve ark. tarafından 1998 yılında uygulanmıştır. Bu yöntem veriler sürekli, iki şıklı, sıralı veya zamana bağlı sürekli veriler olduğunda kullanılabilir. Yöntem sağlık alanından elde edilen gerçek bir veri seti üzerinde uygunacaktır. Verilerin ranklarını alıp, ANCOVA hesabı çok kolay bir şekilde yapılabilir. Ancak bulunan sonuçlar orijinal veriyi yeterince yorumlayamaz. Fakat RobustLinear Model (RGLM) ve Randomization-based yöntemi pek çok avantaj sağlar. Bu yöntemlerle güven sınırlarını elde etmek ve orijinal verinin tabiatına yakın sonuçlar elde etmek mümkündür. Eğer aykırı gözlem varsa parametrik olmayan ANCOVA, ANCOVA'ya göre daha güçlü analizler yapar.

Anahtar Kelimeler: Parametrik Olmayan Kovaryans Analizi, Rank ANCOVA, Randomization-based, Robust ANCOVA, Nonparametrik ANCOVA

S24

GENETİK EPİDEMİYOLOJİDE KULLANILAN PAKET PROGRAMLAR: MERLİN İLE BİR UYGULAMA

Yusuf Kemal Arslan[1], Pınar Özaltun[1], Selen Begüm Uzun[1], Refik Burgut[1]

[1]*Çukurova Üniversitesi Tıp Fakültesi, Biyoistatistik ve Tıbbi Bilişim ABD, Adana*

ykarslan@cu.edu.tr

Genetik epidemiyoloji, hastalıkların genetik kökeni olup olmadığından başlayıp hastalığın genetik nedenine ve bu nedenin nasıl çalıştığına dair analizlerin bütünüdür. Genlerin ve çevrenin hastalıklar üzerindeki etkisinin, bu genlerin popülasyon varyasyonlarının ve diğer faktörlerle etkileşiminin nasıl olduğunun araştırılması gereklidir. Bu araştırmalarda kullanılan Merlin, Mendel, S.A.G.E, Genehunter gibi bir çok paket program bulunmaktadır. Bu programlarla bağlantı analizi, IBD tahmini, regresyon analizi, tekrarlı ölçümler, ilişki analizi, Hardy Weinberg dengesi gibi genetiğe özgü analizler gerçekleştirilebilmektedir. Çalışmamızda bazı paket programlar ve bu programlarla gerçekleştirilebilecek genetiğe özgü bazı analizler belirtilecek ve Merlin paket programı ile uygulamalar yapılacaktır. Merlin, soy ağacı içindeki kişilerin gen akışlarını hızlı ve grafiksel bir biçimde analiz eden paket programıdır. Bu program, nicel özelliklerde parametrik ve parametrik olmayan bağlantı analizinde, regresyon tabanlı bağlantı analizinde veya ilişki analizinde; ayrıca IBD ve akrabalık tahmininde, hata belirlemede ve simülasyonda kullanılabilir. Bağlantı dengesizliği Merlin programındaki bir çok analizde yer almaktadır. Literatürde verilen örnek veri setleri kullanılarak bu paket programda bazı analizler gerçekleştirilecek ve sonuçlar değerlendirilecektir. Merlin ve diğer paket programlar karşılaştırılarak avantaj ve dezavantajlar gösterilecektir.

Anahtar Kelimeler: Merlin, Genetik epidemiyoloji, Biyoistatistiksel yöntemler

S25

DERMATOLOJİK HASTALIKLARIN TEŞHİSİNDE TEMEL SINIFLANDIRMA ALGORİTMALARININ BAŞARIM ÖLÇÜTLERİNİN DEĞERLENDİRİLMESİ

Evaluation of Classifiers' Performance for Diagnosis of Dermatologic Diseases

Emel Koç[1], Aslı Uyar[1], Yasemin Atılgan Şengül[2]

[1]*Bilgisayar Mühendisliği Bölümü, Okan Üniversitesi, İstanbul*

[2]*Bilgisayar Programcılığı Bölümü, Doğuş Üniversitesi, İstanbul*

emel.koc@okan.edu.tr

Özetçe—Tıbbi teşhis ve tedavi süreçlerinde sağlık uzmanlarına karar desteği sağlayan makine öğrenmesi tabanlı tahmin modellerinin kullanımı son yıllarda artış göstermektedir. Bu kapsamda, önemli bir uygulama alanı olarak dermatolojik hastalıklara ilişkin verileri içeren açık erişimli bir veri tabanı oluşturulmuş ve bu hastalıkların otomatik olarak sınıflandırılması konusunda çalışmalar yapılmıştır. Öte yandan, güvenilir klinik karar destek sistemleri oluşturulabilmesi için tahmin modellerinin başarım ölçütlerinin tutarlı ve yansız bir bicimde değerlendirilmesi gerekmektedir. Bu çalışmada, 10 kat çapraz doğrulama ve ROC performans analizi yöntemleri kullanılarak temel sınıflandırma algoritmalarının dermatolojik hastalıkların teşhisindeki başarımları karşılaştırmalı olarak değerlendirilmiştir. Deneysel sonuçlara göre, Naïve Bayes ve Çok Katmanlı Perseptron yöntemleri, k-NN ve Karar Ağacı yöntemlerine kıyasla anlamlı olarak daha yüksek performans sergilemektedir.

Anahtar Kelimeler — Dermatolojik Hastalıklar, Sınıflandırma, ROC, k-NN, Naïve Bayes, Karar Ağacı, MLP.

Abstract— In recent years there is a significant increase in a machine learning based predictive models in medical care as a strong tool for providing decision support for healthcare professionals during diagnosis and treatment. In this context, an open access data repository related with dermatologic diseases has been generated and some studies have been conducted in the field of classification of dermatologic diseases. On the other hand, there is a need for consistent and objective evaluation of these models for reliable clinical decision support systems. In this study, the success rate of classification algorithms for diagnosis of dermatologic diseases have been comparatively evaluated using 10 fold cross validation and ROC performance analysis methods. According to experimental results, Naïve Bayes and Multi Layer Perceptron methods show significantly higher performance compared to k-NN and Decision Tree methods.

Keywords — Dermatologic Diseases, Classification, ROC, k-NN, Naïve Bayes, Decision Tree, MLP.

GİRİŞ

Sağlık sektöründe Elektronik Hasta Kayıt sistemlerinin kullanılmasının yaygınlaşmasıyla birlikte tıbbi verilerin elektronik veritabanlarında saklanması mümkün hale gelmiş ve bunun sonucunda tıp alanında kayıtlı veriler kullanılarak oluşturulan karar destek sistemlerinin kullanımı yaygınlaşmıştır. Yapılan sistematik bir araştırmanın sonuçlarına göre, klinik karar destek sistemleri kullanımı sağlık uzmanlarının performansını arttırmaktadır; ancak hastalara uygulanan tedavi sonuçları üzerine yapılan araştırmaların sonuçları tutarlı değildir [1]. Akademik çalışmaların seyri göz önünde bulundurulduğunda, yakın gelecekte makine öğrenmesi tabanlı teşhis ve tedavi karar destek sistemlerinin, elektronik sağlık süreçlerinin bir parçası olacağı öngörülmektedir. Bu alanda güvenilir ve

karsılaştırılabilir sonuçlar elde edilmesi için, makine öğrenmesi yöntemlerinin kapsamlı ve yansız bir biçimde değerlendirilmesi gerekmektedir.

Klinik karar destek sistemleri kapsamında önemli bir uygulama alanı olarak dermatolojik hastalıklara ilişkin verileri içeren açık erişimli bir veri tabanı oluşturulmuş ve bu hastalıkların otomatik olarak sınıflandırılması konusunda çalışmalar yapılmıştır [2]. Bu çalışmada, eryhemato-squamous hastalığının farklı türlerinin teşhisinde temel sınıflandırma algoritmalarının performansları karşılaştırmalı olarak değerlendirilmiştir.

Bildirinin ikinci bölümünde seçilen veri kümesi üzerinde uygulanan işlemler ve çalışmada kullanılan performans analizi yöntemleri kısaca tanımlanmış, üçüncü bölümde ise deneysel sonuçlar sunulmuştur. Dördüncü bölümde elde edilen bulgulara dayanılarak sonuçlar değerlendirilmiştir.

YÖNTEM

Sınıflandırma Algoritmalarına Uygulanan Önişlemler

Boyutları büyük olan ve çeşitlilik gösteren tıbbi veriler hasta öyküleri, tahlil ve rapor sonuçları vb. içeren metin verisi, laboratuar sonuçları gibi nümerik ölçümler, elektrokardiyografi gibi kayıtlı sinyaller ve radyoloji görüntüleri, filmler, mikroskop verileri, ultrason verileri, kamera verileri gibi örüntü verilerinden oluşmaktadır. Veriler arasında eksik, geçersiz, önemsiz veri bulunabileceği gibi veriler gürültülü, dengesiz ve tutarsız sınıf dağılımına sahip de olabilir. Eksik değerleri kaldırmak, pürüzsüz gürültüsüz veri tanımlamak ve tutarsızlıkları gidermek için veriler üzerinde veri önişleme işlemi uygulanmaktadır.

Veri dönüştürme (normalizasyon) sınıflandırma sürecinde başarıya doğru katkı sağlaması beklenen bir veri önişleme yöntemidir. Öznitelik seçimi, makine öğrenmesinin önemli bir problemi olan boyut fazlalılığını ortadan kaldırmayı hedeflemektedir. Amacı, bir makine öğrenme sisteminde öğrenmeyi olumlu yönde etkileyen nitelikleri seçip, olumsuz yönde etkileyen, sisteme zarar veren nitelikleri de ortadan kaldırmaktır. Temel bileşen analizi (Principal Component Analysis - PCA) ise özniteliklerin lineer kombinasyonları kullanılarak veri kümesinin içerdiği varyansı büyük oranda koruyarak boyut indirgemeyi hedeflemektedir.

Sınıflandırıcı Yöntemler

Çalışmada Dermatoloji veri seti üzerinde farklı matematiksel temellere dayalı k En Yakin Komsu (k-Nearest Neighbor - k-NN), Naïve Bayes, Karar Ağacı ve Çok Katmanlı Perseptron (Multi-Layer Perceptron - MLP) sınıflandırma algoritmalarının basarım oranları test edilmiştir.

Performans Analizi

Yapay öğrenme sınıflandırma algoritmalarının yeterliliğini değerlendirmek için kullanılan çeşitli performans kriterleri bulunmaktadır. İyi bir başarım oranının yüksek doğruluk (accuracy), duyarlılık veya hassasiyet (sensitivity), özgüllük veya seçicilik (specificity) tahmin değerlerine sahip olması beklenir. Duyarlılık aynı zamanda doğru pozitif (DP) ve [1 - seçicilik] ise yanlış pozitif (YP) olarak adlandırılmaktadır. ROC (Receiver Operating Characteristics) eğrisi ise; duyarlılık ve seçicilik birleştirilerek tek bir ölçüt (ROC eğrisi altında kalan alan – Area Under the ROC Curve (AUC)) ile testin ayırt etme gücünün belirlenmesine olanak sağlar. En faydalı test, doğru pozitiflik oranı yüksek ve yanlış pozitiflik oranı düşük olan testtir. Mükemmele yakın bir tanı testi, hemen hemen dikey (0,0)'dan (0,1)'e ve sonra yatayda (1,1)'den geçen bir ROC eğrisine sahip olmalıdır [3]. Çalışma kapsamında kullanılan dermatoloji veri

kümesine temel sınıflandırma ve öznitelik seçme algoritmaları uygulanırken 10-kat çapraz doğrulama (10-fold cross validation) öğrenme ve test yöntemi kullanılmıştır. 10-kat çapraz doğrulama yöntemi ile veri kaynağı 10 bölüme ayrılmakta ve her bölüm bir kez test kümesi, kalan diğer 9 bölüm ise öğrenme kümesi olarak kullanılmaktadır. Algoritmaların performansları ortalama doğruluk, DP, YP ve ROC değerleri göz önünde bulundurularak değerlendirilmiştir. ROC analizi temelde iki sınıflı problemler için performans analizi sağlayan bir yöntemdir. Bu çalışmada incelenen araştırma problemi 6 sınıflı problem olduğundan, YP, DP ve ROC alanı değerleri hesaplanırken, her bir sınıf için diğer 5 sınıf ortak tek bir sınıf gibi değerlendirilmiş, dolayısıyla ROC analizi 6 adet iki sınıflı problem şeklinde gerçekleştirilmiş ve sonuçta ortalamalar sunulmuştur.

Dermatolojik Hastalıkların Sınıflandırılması

Dermatolojik hastalıkların teşhisi için kullanılan veriler UCI Machine Learning Repository [4] veritabanından elde edilmiştir. Sınıflandırma deneyleri Waikato Environment for Knowledge Analysis (WEKA) [5] programı kullanılarak gerçekleştirilmiştir.

Dermatoloji Veri Kümesi

Eryhemato-squamous hastalığının belirlenmesi amacıyla oluşturulan dermatoloji veri kümesi içerisinde hastalığın 6 sınıfı; psoriasis, seboreic dermatitis, lichen planus, pityriasis rosea, chronic dermatitis ve pityriasis rubra pilaris olarak isimlendirilmektedir.

Veri kümesi 366 adet örnek, 34 adet öznitelik içermektedir. Öznitelikler kategorik ve tam sayı olmak üzere 2 farklı biçimden oluşmaktadır. Veri kümesi içerisindeki bütün bileşenlerin değerleri [0…3] aralığında kademeli ölçek ile eşleştirilmiştir. 0 değeri ilgili karakteristikten yoksunluğunu, 3 değeri ise güçlü bir karakteristiğe sahip olduğunu, 1 ve 2 ise ara değerleri ifade etmektedir. Ayrıca family_history özniteliğinin 1 değerini alması hastalığın aile öyküsü olduğu, 0 değerini alması ise hastalığın aile öyküsü olmadığını ifade etmektedir. Veri setine ait 6 sınıf ve sınıflara ait örnek sayıları:

- Psoriasis 112
- seboreic dermatitis 61
- lichen planus 72
- ptyriasis rosea 49
- cronic dermatitis 52
- pityriasis rubra pilaris 20

şeklindedir. Veri kümesinde yaş özniteliğine ait 8 tane kayıp değer bulunmaktadır. Bu kayıp değerler için uygulama içerisinde bulunan varsayılan eksik değer giderme yöntemi kullanılmıştır. Bu yöntemde, sayısal öznitelikler için eksik veriler yerine o özniteliğin eğitim setindeki ortalaması eklenir.

Literatürde Dermatoloji Veri Kümesi ile Yapılan Çalışmalar

Tablo 1'de Dermatoloji veri seti ile yapılmış, literatürde yer alan çalışmaların özeti sunulmuştur. Çoğunlukla kullanılan yöntemler Genetik Algoritma ve C4.5 Karar Ağacı yöntemleridir. Sonuçlar, doğruluk ya da bütünleyeni olan hata oranı ile belirtilmiştir. Çalışmalar, %84.2 ile %100 arasında değişen doğruluk değerleri sunmaktadır. Öte yandan,

sonuçların doğrudan karşılaştırılabilir olması için önişleme yöntemlerinin, öğrenme ve test süreçlerinin ve model parametresi seçiminin uyumlu olması gerekmektedir.

Tablo 1: Dermatoloji veri seti kullanılarak literatürde yapılan çalışmalar

Yazar	Kullanılan Algoritmalar	Performans Sonuçları
Demiröz et al. (1998), [2].	**VFI5**	**Doğruluk %99.2**
Pappa et al. (2002), [6].	**Çokdeğişkenli Genetik Algoritma C4.5**	**Hata Oranı (%) C4.5+GA / C4.5 5.5 ± 1.46 / 4.2 ± 0.96**
Pappa et al. (2002), [7].	**Çokdeğişkenli Genetik Algoritma C4.5**	**C4.5 Ağaç Boyutu / Hata Oranı / Total 29.0 ± 3.65/15.95 ± 1.43/1.11 ± 0.11**
Parpinelli et al. (2001), [8].	**Ant Miner C4.5**	**AntMiner vs. C4.5 AntClass Doğruluk / Kural Sayısı / Seçenek Sayısı 84.21% ± 6.34/6.00 ± 0.00/79.00 ± 3.46 C4.5 Doğruluk / Kural Sayısı / Seçenek Sayısı 89.05% ± 0.62/23.2 ± 1.99/91.7 ± 10.64**
Fidelis et al. (2000), [9].	**GA**	**Doğruluk 95%**
Güvenir. (1998), [10].	**k-NN VFI5**	**Doğruluk 100%**
Parpinelli et al. (2001), [11].	**AntMiner Algoritması C4.5**	**AntMiner Doğruluk (%) / Kural Sayısı 86.55 ± 6.13 / 7.00 ± 0.00 C4.5 Doğruluk (%) / Kural Sayısı 89.05 ± 0.62 / 23.2 ± 1.99**

DENEYLER VE BULGULAR

Dermatoloji veri setine 10 – kat çapraz doğrulama metodu ile uygulanan sınıflandırma algoritmalarının performans sonuçları Tablo 2'de sunulmuştur. Algoritmaların model parametreleri olarak Weka ortamındaki varsayılan değerler kullanılmıştır.

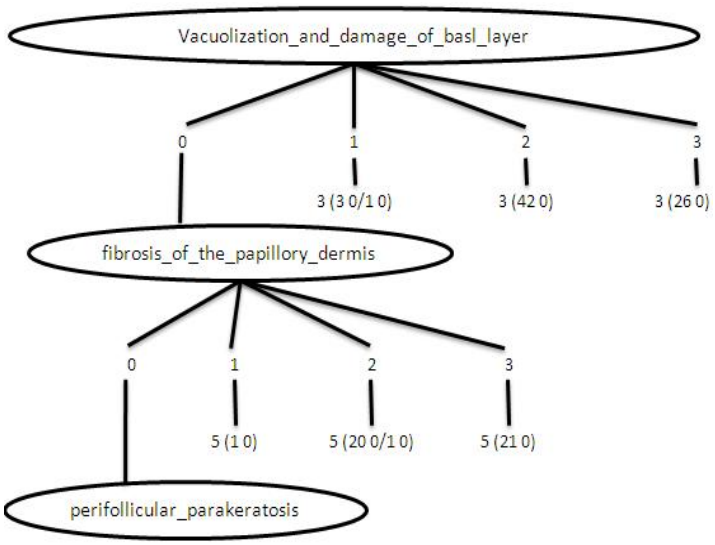
Sonuçlar incelendiğinde, değerlendirmede kullanılan bütün performans kriterlerine göre en başarılı sonuçların Naïve Bayes algoritması ile elde edildiği görülmektedir. Bayes teoremine göre istatistiksel kestirim yapan Naïve Bayes sınıflayıcı bütün örneklerin sınıf üyelik olasılığını tahminleyerek başarılı sonuçlara ulaşmıştır. Naïve Bayes

algoritmasını %96.1'lik doğruluk oranı ile MLP sınıflandırma algoritması izlemektedir. Dermatoloji veri setine k-NN algoritmasının k=1 değeri ile uygulanması sonucu oluşan doğruluk oranı ise %94.5'dir.

Tablo 2: Dermatoloji veri setine 10-kat çapraz doğrulama metodu ile uygulanan sınıflandırma algoritmalarının performans sonuçları

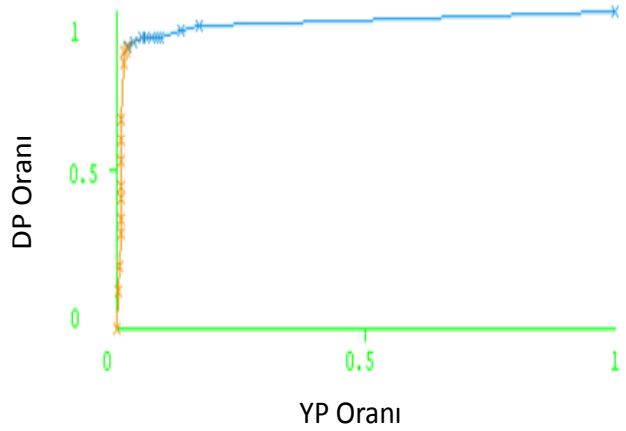
Yöntem	Doğruluk(%)	Ortalama DP Oranı (%)	Ortalama YP Oranı (%)	Ortalama ROC Alanı
K-nn	**94.5 ± 0.397**	**94.5 ± 0.401**	**1.0 ± 29.57**	**0.969 ± 0.004**
Naïve Bayes	**97.2 ± 0.330**	**97.3 ± 0.309**	**0.5 ± 0.056**	**0.999 ± 0**
Karar **Ağacı**	**93.9 ± 0.440**	**94.0 ± 0.428**	**1.5 ± 0.182**	**0.977 ± 0.002**
MLP	**96.1 ± 0.398**	**96.2 ± 0.427**	**0.7 ± 0.081**	**0.997 ± 0.001**

Son olarak; dermatoloji veri setine karar ağacı algoritmalarından J48 uygulanmıştır. Algoritmanın doğruluk oranı %93.9 olarak gözlemlenmiştir. Dermatoloji veri seti ile J48 karar ağacı algoritması kullanımı sonucu 25 yapraktan oluşan budanmış bir ağaç ortaya çıkmış bulunmaktadır. Sekil 1'de bir bölümü belirtilen ağacın en önemli özelliği kökü olarak belirtilen vacuolization_and_damage_of_basl_layer'dır.



Sekil 1. Dermatoloji veri setine J48 sınıflandırma algoritması uygulanması sonucu oluşan karar ağacından bir kesit.

Uygulanan dört sınıflandırma algoritması içinde model başarım oranı en yüksek olan Naïve Bayes algoritması değerlendirildiğinde gerçek dermatolojik hastalığı doğru olarak tahmin edilen örneklerin oranını belirten DP oranı %97.3, gerçek hastalıktan farklı bir dermatolojik hastalık teşhisi konulan örnekleri ifade eden YP oranı ise %0.5 olarak gözlemlenmiştir. Klinik durumlarda optimum eşik değerine bağlı olarak değişen DP ve YP oranlarının tamamını tek bir ölçüt ile belirtme imkanı sağlayan ROC analizi sonuçları ise 1'e yakın AUC değerleri ile mükemmel seviyeye yaklaşmış durumdadır. Dermatoloji veri setine J48 sınıflandırma algoritması uygulanması sonucu oluşan sınıf 2'ye ait ROC eğrisi (AUC = 0.958) Şekil 2'de gösterilmiştir.



Sekil 2. Dermatoloji veri setine J48 sınıflandırma algoritması uygulanması sonucu oluşan sınıf 2'ye ait ROC eğrisi (AUC = 0.958).

Sınıflandırma algoritmalarının ortalama ROC alanları arasındaki farklılıklar ANOVA yöntemi ile test edilmiş ve sonuçlar arasında anlamlı farklılık olduğu belirlenmiştir ($p < 0.05$). ANOVA sonrası karsılaştırma sonuçlarına göre ise, Naïve Bayes ve MLP yöntemleri, k-NN ve Karar Ağacı yöntemlerine kıyasla anlamlı olarak daha yüksek başarım oranı sergilemektedir.

Sınıflandırma problemlerinde, öznitelik seçimi ya da boyut indirgeme önişlemlerinin gereksiz veriyi eleyerek sınıflandırma performansını arttırması beklenmektedir. Öte yandan, bütün öznitelikler kullanılarak elde edilen başarım oranının oldukça yüksek olduğu dermatolojik hastalıkların sınıflandırılması probleminde öznitelik seçimi ve boyut indirgeme yöntemlerinin sonuçlar üzerindeki etkisi de incelenmiştir. İlgili deneyler sonucunda elde edilen bulgular Tablo 3'te belirtilmektedir. Seçilen 19 öznitelik kullanıldığında sadece k-NN sınıflandırma başarımında artış gözlemlenmektedir. Veri kümesindeki varyansın %95'ini içeren 71 adet temel bileşen kullanılarak gerçekleştirilen tahmin sonuçları ise bütün sınıflandırıcıların başarım oranında azalmaya neden olmuştur.

Tablo 3: Dermatoloji veri setine 10-kat çapraz doğrulama metodu ve öznitelik seçimi yöntemi ile uygulanan sınıflandırma algoritmalarının ROC alanı performans sonuçları

Yöntem	**Bütün Öznitelikler**	**Öznitelik Seçimi**	**Temel Bilesen Analizi**
k-NN	**0.969 ± 0.004**	**0.99 ± 0.001**	**0.944 ± 0.009**
Naïve Bayes	**0.999 ± 0**	**0.999 ± 0.001**	**0.995 ± 0.001**
Karar Ağacı	**0.977 ± 0.002**	**0.975 ± 0.004**	**0.97 ± 0.006**
MLP	**0.997 ± 0.001**	**0.997 ± 0.001**	**0.995 ± 0.001**

SONUÇLAR

Hayati önem taşıyan teşhis ve tedavi karar destek sistemlerinin güvenilirliği için makine öğrenmesi yöntemlerinin tutarlı bir şekilde uygulanması ve performanslarının yansız olarak değerlendirilmesi kritik önem taşımaktadır. Bu çalışmada, dermatolojik hastalıkların sınıflandırılması problemi dört farklı sınıflandırma algoritması ile karşılaştırmalı olarak analiz edilmiştir. Elde edilen sonuçlar başarım ölçütleri açısından literatür ile uyumludur. Sonuçlar üzerinde yapılan istatistiksel testlere göre Naïve Bayes ve MLP algoritmaları dermatolojik hastalık teşhisinde görece daha

yüksek başarım oranı sergilemektedir. Aynı zamanda öznitelik seçimi ve temel bileşen analizi boyut indirgeme yöntemlerinin tahmin performansını arttırmadığı belirlenmiştir.

Elde edilen bulguların yanı sıra, deneylerde kullanılan açık erişimli veri kümesi üzerinde ilk defa bu kadar kapsamlı bir çalışma yapılmış ve tekrar edilebilir, karşılaştırılabilir sonuçlar sunulmuştur. Elektronik sağlık sistemlerine entegre olarak çalışması hedeflenen dermatoloji karar destek modülünün algoritmik altyapısı oluşturulmuştur.

KAYNAKÇA

[1] Garg AX, Adhikari NK, McDonald H, Rosas-Arellano MP, Devereaux PJ, Beyene J, Sam J, Haynes RB: Effects of computerized clinical decision support systems on practitioner performance and patient outcomes: a systematic review. *JAMA*, 293:1223-1238, 2005.

[2] Demiroz G., Güvenir H.A., İlter N., "Learning differential diagnosis of erythemato- squamous diseases using voting feature", *Artificial Intelligence in Medicine*, 147-165, 1998.

[3] Metz C.E., "Receiver operating characteristic analysis: a tool for the quantitative evaluation of observer performance and imaging systems", *J Am Coll Radiol*, 413-422, 2006.

[4] A. Frank ve A. Asuncion, UCI Machine Learning Repository, University of California, *School of Information and Computer Science*, 2010.

[5] http://www.cs.waikato.ac.nz/ml/weka/.

[6] Pappa L.G., Freitas A.A., KaestnerA.A.C., "Attribute Selection with a Multiobjective Genetic Algorithm", *Advances in Artificial Intelligence*, 2002.

[7] Pappa L.G., Freitas A.A., KaestnerA.A.C., "A Multiobjective Genetic Algorithm for Attribute Selection", 116-121, 2002.

[8] Parpinelli S.R., Lopes S.H., Freitas A.A., "An Ant Colony Based System for Data Mining: Applications to Medical Data", *Proceeedings of the Genetic and Evolutionary ComputationConference*, 791-797, 2001.

[9] Fidelis M.V., Lopes S.H., Freitas A.A., "Discovering Comprehensible Classification Rules with a Genetic Algorithm", 205, 805-811, 2000.

[10] Güvenir A.H., "A Classification Learning Algorithm Robust to Irrelevant Features", Artificial Intelligence: Methodology, Systems, and Applications Lecture Notes in Computer Science, 1998.

[11] Parpinelli S.R., Lopes S.H., Freitas A.A., "An Ant Colony Algorithm for Classification Rule Discovery", *IDIAP*, 00-25.

S26

SPOR VE YAŞAM MERKEZLERİNDE VERİ MADENCİLİĞİ ÇALIŞMASI

Melih Keskin[1], Pınar Yıldırım[2]

[1]*Okan Üniversitesi, Fen Bilimleri Enstitüsü, Bilgisayar Mühendisliği Yüksek Lisans Programı, İstanbul*

[2]*Okan Üniversitesi, Mühendislik ve Mimarlık Fakültesi, Bilgisayar Mühendisliği Bölümü, İstanbul*

melih@senkron.net

Giriş ve Amaç: Son yıllarda bilgi ve iletişim teknolojilerindeki gelişmeler ile birlikte veri madenciliği hemen her alanda uygulanmaktadır. Buna rağmen ülkemizde yeni gelişen bir sektör olan spor ve yaşam merkezleri üzerine yapılan çalışma sayısı çok azdır. Spor sektöründe, sadece bireysel oyuncu performansları, takım başarı istatistiği ve kulüp yönetimi karar alma mekanizmasına yardımcı çalışmalar yapılmıştır. Bu çalışmanın amacı, spor ve yaşam merkezi sektöründe yaşanılan problemler ve karar alma süreçleri için veri madenciliği yöntemleri ile çözüm sunmaktır. Üyelik yenileme ve müşteri sadakatini arttırma, fiyatlandırma politikası, PR yönetimi (Public Relations) ve buna bağlı pazarlama stratejilerine yardımcı olacak kararların çıkarılması hedeflenmiştir.

Yöntem: Bu çalışma kapsamında bir spor merkezinin müşteri bilgileri, üyelik bilgileri, tesis kullanım bilgileri ve üyelik dışı satılan özel ders ve spa hizmet bilgileri göz önüne alınmıştır. Yapılan inceleme sonrası oluşturulan veri seti ile kişilerin aldıkları üyelikler ve yenileme durumları, havuz kullanımı, grup ders kullanımı, özel ders ve spa hizmeti satın alım sayıları ayrı ayrı hesaplanarak bir tablo oluşturulmuştur. Çalışmada Sql Business Intelligence ve Weka yazılımları kullanılmıştır.Oluşturulan veri setinde 500 adet kayıt ve 13 adet nitelik bulunmaktadır. Veri seti kümeleme ve sınıflandırma algoritmaları ile analiz edilmiş ve elde edilen sonuçlar karşılaştırılmıştır. Kümeleme analizi için Simple K-Means, Scalable K-Means algoritmaları kullanılmış ve kümeleme sayısı 5 olarak belirlenmiştir. Sınıflandırma tekniklerinde ise J48 (C4.5), IB1, IBk, Naive Bayes algoritmaları kullanılmıştır.

Bulgu: Sınıflandırma algoritmaları ile yapılan analizde J48 için % 96, KStar için % 82, Naive Bayes için % 80, IB1 için % 73 ve IBk için % 73 başarı oranları elde edilmiştir. Ayrıca J48 algoritmasının sonucunda elde edilen karar ağacından kurallar oluşturulmuş. Öne çıkan nitelikler, tesise geliş sayıları ile havuz girişleri olmuştur. Bu nitelikler üyelik yenilenmesine doğrudan etki eden faktörlerdir. Kümeleme algoritmalarında ise veri seti içinde doğrudan görülemeyen benzer özelliklere sahip müşteri kümeleri ortaya çıkarılmıştır.Bunun sonuçlarında lokasyon niteliği baskın çıkmıştır.

Sonuç: Bu çalışmada, spor ve yaşam merkezlerinde kullanılan içerik yönetim sisteminden elde edilen ve müşteri bilgileri ile oluşturulan veri seti, hem sınıflandırma hem de kümeleme algoritmaları ile

analiz edilmiştir. Sektörün karar destek sistemine katkı sağlayacak yararlı bilgiler keşfedilmeye çalışılmıştır.

Bu çalışma; üyelik yenileme analizi dışında, müşteri hizmet eşleştirmesi, personel ve hizmet performans değerlendirmesi ile fiyatlandırma ve pazarlama stratejileri oluşturulmasına katkı sağlar.

Anahtar kelimeler: Spor ve yaşam merkezlerinde veri madenciliği, Kulüp üyeliğinde müşteri sadakati, Spor merkezlerinde müşteri yaşam döngüsü.

S27

WEB TABANLI İSTATİSTİK ÇÖZÜMLEMELER: FAKTÖRİYEL VARYANS ANALİZİ VE İNTERAKSİYON ETKİSİNİN ÇOKLU KARŞILAŞTIRMA YÖNTEMİ İLE İNCELENMESİ

Mehmet Ali Sungur[1], Şengül Cangür[1], Handan Ankaralı[1]

[1]*Düzce Üniversitesi, Tıp Fakültesi, Biyoistatistik ve Tıbbi Bilişim Anabilim Dalı, Düzce*

malisungur@yahoo.com

Amaç: Mevcut teknolojik koşullar ve alternatif paket program çokluğuna karşın halen birçok paket programda yer almadığı için elle hesaplanması gereken istatistik yöntemler bulunmaktadır. Bu yöntemlerden birisi de faktöriyel denemelerde interaksiyona ait sıfır hipotezinin ret edilmesi durumunda, interaksiyon etkisini değerlendirmek için kullanılan çoklu karşılaştırmalardır. Bu çalışmada, söz konusu probleme çözüm getirmek amacıyla faktöriyel bir denemenin analizi ve interaksiyon etkisini değerlendirmek için çoklu karşılaştırma yapan web tabanlı bir hesaplama modülü geliştirmek amaçlanmıştır. Ayrıca çeşitli tanımlayıcı istatistiklerin hesaplanması ve tanımlayıcı istatistiklerden yararlanılarak basit grup karşılaştırmalarının yapılması da bu çalışma kapsamında ele alınmıştır. Böylece sahip olduğumuz paket programlarda yer almayan istatistik çözümleme yöntemlerine, web tabanlı çözümleme getirilebileceği tartışılmıştır.

Yöntem: Bu çalışmada, PHP dili ve JavaScript ile iki yönlü faktöriyel varyans analizi modelinde veri matrisinin oluşturulması ve veri girişinin yapılmasının ardından tek bir tuş yardımıyla varyans analizi tablosunun elde edilebildiği web tabanlı bir hesaplama uygulaması yapılmıştır. Ayrıca varyans analizi sonucunda esas etkilerin ve/veya interaksiyon etkisinin istatistiksel olarak anlamlı bulunması durumunda, çoklu karşılaştırmaların yapılması sağlanmıştır. Bunlara ilaveten tanımlayıcı istatistik modülleri de geliştirilmiştir.

Bulgular: Karmaşık kod yapısı nedeniyle hazırlık aşamaları zor ve zaman alıcı olan PHP dili ile form oluşturma, hesaplama yapma gibi çalışmaların dezavantaj gibi görünmesine karşın bir kez hazırlandıktan sonra sadece Internet Explorer gibi tarayıcılar aracılığıyla çalıştırılabilir olması en büyük avantajlarından biridir. Oluşturduğumuz web tabanlı hesaplama modülüne kolaylıkla ulaşılabilmekte ve veri girişi yapıldıktan sonra hesaplamalar sadece bir tuş aracılığıyla yapılabilmektedir. Bu şekilde karmaşık ve uygulaması zor olan, ayrıca hesaplamaları zaman alan istatistik yöntemler kolay erişilebilir, uygulanabilir ve yorumlanabilir hale gelmiştir.

Sonuç: Bu çalışmanın sonunda, matematiksel modeli bilinen ancak mevcut paket programlar aracılığıyla yapılamayan veya henüz paket programlara girmemiş herhangi bir istatistiksel analiz yöntemine ait hesaplamaların, web tabanlı ve kullanımı kolay ara yüzlerle yapılabileceği vurgulanmıştır.

Anahtar Kelimeler: Web tabanlı istatistik hesaplamalar, İki yönlü varyans analizi, Çoklu karşılaştırma, PHP

S28

ÇOKLU KALİTE KARAKTERİSTİKLERİNİN EŞZAMANLI OPTİMİZASYONU

Muzaffer Cenk Zöngür[1], Timur Köse[2], Fikret İkiz[2]

[1]*Uludağ Üniversitesi Tıp Fakültesi, Biyoistatistik Anabilim Dalı, Bursa*

[2]*Ege Üniversitesi Tıp Fakültesi, Biyoistatistik ve Tıbbi Bilişim Anabilim Dalı, İzmir*

cenkzongurr@hotmail.com

Amaç: Bu çalışmanın amacı endüstriyel alanda, birden fazla kalite karakteristiğinin (değişkeninin) eşzamanlı olarak optimize edilmesinde kullanılan bazı Taguchi esaslı yöntemleri incelemektir.

Yöntem: Çalışmada Taguchi'nin deneme tasarımı ve analizi stratejisini temel alan ve çoklu kalite karakteristiklerinin eşzamanlı optimizasyonu için önerilen yöntemlerin içinden çoklu cevap sinyal/gürültü oranı, kalite kaybının yüzdesel azalımı ve faktör seviyesine dayalı ağırlıklı sinyal/gürültü oranı yöntemleri kullanılmıştır. Ele alınan bu yöntemler Madhav Phadke tarafından 1989 yılında yayınlanan "Quality Engineering Using Robust Design" kitabında verilen polisilikon koruma ve geliştirme çalışmasında elde edilmiş veriler üzerinde denenmiş, farklı yöntemlerle bulunan optimum faktör kombinasyonlarının kalite kaybını azaltma oranlarına bakılmıştır.

Bulgular: Çoklu cevap sinyal/gürültü oranı yöntemiyle elde edilen optimum faktör kombinasyonunun kalite kaybını %47,05 oranında, kalite kaybının azalma oranı yöntemiyle bulunan kombinasyonun kalite kaybını %71,54 oranında azalttığı, faktör seviyesine dayalı ağırlıklı sinyal/gürültü oranı yaklaşımıyla bulunan optimum faktör kombinasyonunun ise kalite kaybını %351,71 oranında arttırdığı görülmüştür. Çoklu cevap sinyal/gürültü oranı yönteminde düzeltme faktörleri dikkate alındığında da kalite kaybının azalma oranının %47,05'den %68,78'e çıktığı gözlemlenmiştir.

Sonuç: Çalışma sonuçlarına dayalı olarak ürün ve/veya süreç gelişimine en büyük katkıyı kalite kaybının azalma oranı yönteminin yapabileceği düşünülmektedir. Ancak bu yöntemin uygulanabilir olması için başlangıç koşulu gibi bazı kriterlerin belirlenmesi gerekmektedir. Bu kriterlerin belirlenemediği çalışmalarda düzeltme faktörleri dikkate alınmak şartıyla çoklu cevap sinyal/gürültü oranı yönteminin kullanımı da uygun bulunmuştur. Literatürde ilaç geliştirme uygulamalarında faktörlerin etkileri incelenirken genellikle tam faktöriyel deneme tasarımının kullanıldığı görülmekte, ancak, bu tür çalışmalarda tam faktöriyel denemeden daha az deneysel materyal kullanarak faktörlerin etkisinin incelenmesine olanak sağlayan ortogonal dizilerin kullanımı da alternatif olarak düşünülebilir.

Anahtar Kelimeler: Çoklu kalite Karakteristiği, Eşzamanlı Optimizasyon, Taguchi metodu, Kalite Kaybının azalma oranı

S29

HOMOJEN OLMAYAN POISSON SÜREÇLERİNDE ŞİDDET FONKSİYONUNUN TAHMİNİ: RADYOLOJİK VERİ ÜZERİNE UYGULAMA

Özlem Kalın[1], Serdal Kenan Köse[2], Halil Aydoğdu[3]

[1]*Ankara Üniversitesi, Fen Bilimleri Enstitüsü İstatistik Anabilim Dalı, Ankara*

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

[3]*Ankara Üniversitesi, Fen Bilimleri Enstitüsü, İstatistik Anabilim Dalı, Ankara*

ozlemkalin@gmail.com

Zamanın bir fonksiyonu olarak gerçekleşen olayların sayısını sayan bir sürece sayma süreci denilmektedir. Pratikte bir sayma sürecinden bir $\{X_1, X_2, \dots, X_n\}$ veri kümesi geldiğinde ilk yapılacak şey kullanacak olan modeli belirlemektir. Eğer $\{X_1, X_2, \dots, X_n\}$ rasgele değişkenleri bağımsız ve aynı dağılımlı ise bu veri kümesi için model olarak bir yenileme süreci önerilir. Eğer bu rasgele değişkenlerin dağılımının üstel olduğu saptanırsa modelimiz homojen bir Poisson süreci olacaktır. Fakat, veride bir trend varsa, yani $X_1, X_2, \dots, X_n$ rasgele değişkenleri aynı dağılımlı değilse model olarak bir yenileme süreci önerilemez. Böyle bir trendin varlığında alışıldık bir yaklaşım homojen olmayan bir Poisson süreci kullanmaktır.

Bu projede, uygulamada çok sık kullanılan homojen olmayan bir Poisson süreci üzerinde durulacaktır. Bu sebeple, sayma sürecinden gelen veri kümesinde trend olup olmadığı incelenecek. Bu veri kümesinde trendin varlığı durumunda model olarak homojen olmayan bir Poisson süreci önerilecektir. Bu süreçten gelen $\{X_1, X_2, \dots, X_n\}$ veri kümesine dayalı olarak sürecin $\lambda(t)$ şiddet fonksiyonu değerinin hem parametrik olmayan hem de parametrik tahmini üzerinde durulacak. $\lambda(t)$ bilinmediği zaman parametrik olmayan bir tahmin edicisi verilecek. Bu tahmin edicinin nasıl hesaplanacağı, Radyoloji bölümünden alınan bir uygulamaya göre ardışık bozulmalar arası geçen zaman süreleri veri kümesi üzerinde ifade edilecektir. Diğer bir durumda da, $\lambda(t)$ nin bilinmeyen bazı parametrelere bağlı analitik bir ifadesi bilinirse $\lambda(t)$ nin parametrik tahmini üzerinde durulacaktır. Radyoloji bölümünden alınan diğer bir uygulamaya göre bozulma zamanlarının veri kümesini modellemek için hem Cox-Lewis hem de Weibull süreci modelleri önerilecek. Bu modellerdeki şiddet fonksiyonlarının tahminleri elde edilecek. Önerilen bu iki modelin seçimi için hem grafiksel yöntemle hemde istatistiki tahmin ile hangi modelin daha uygun olduğu radyolojik veriler üzerinde tartışılacaktır.

Anahtar Kelimeler: Homojen olmayan Poisson süreci, Cox-Lewis ve Weibull süreci modelleri, Trend

S30

MEME KANSERLİ HASTALARDA POST-OP AĞRININ DEĞERLENDİRİLMESİNDE KÖRLEME YÖNTEMLERİNİN ETKİNLİĞİ

Seval Kul[1], Mehmet Güler[2], Gülendam Karadağ[3]

[1]*Biyoistatistik AD, Gaziantep Üniversitesi, Gaziantep*

[2]*Medikal Park Hastanesi, Genel Cerrahi Kliniği, Antalya*

[3]*Halk Sağlığı Hemşireliği AD, Gaziantep Üniversitesi, Gaziantep*

sevalkul@gantep.edu.tr

Amaç: Tedavi etkilerinin araştırıldığı klinik denemelerde randomizasyon yapmak her ne kadar birçok yanlılığı minimize etse de özellikle değerlendirilen sonuç değişkeninin sübjektif olduğu durumlarda sonuçlar yanlılıklardan etkilenme riski altındadır. Bu nedenle klinik çalışmalarda yapılan uygulamaların etkinliğini ortaya koymak amacıyla mümkün olduğu kadar körleme yöntemleri (tek-kör ya da çift-kör) kullanılmalıdır. Bu çalışmada meme kanseri cerrahisinden sonra ağrı şiddeti değerlendirilirken, körleme yöntemlerinin hastanın ağrıyı raporlamasındaki etkisi araştırılmıştır.

Yöntem: Pubmed veri tabanından belirtilen amaçla yapılmış ve körleme kullanılmış deneysel çalışmalar taranmıştır. Körleme kullanılmayan deneysel çalışmalar ise kontrol grubu olarak kullanılmıştır. İki farklı yaklaşımda anlamlı fark bulma oranları ve ağrı skorlarının etki büyüklükleri karşılaştırılmıştır.

Bulgular: Arama motoru yardımıyla körleme yapılmış 25 çalışmaya ulaşılmış ve bu çalışmaların 13 tanesi çalışmaya dahil edilmiştir. Kontrol grubunu oluşturmak üzere aynı amaçla düzenlenmiş ama körleme yapılmamış 85 çalışmaya ulaşılmış ve bu çalışmalardan ilgili olanları belirlendikten sonra rasgele olarak 16 sı (yaklaşık %15i) çalışmaya dahil edilmiştir. Çalışmaların büyük kısmında ağrı VAS (visual analog scale) ile değerlendirilmiştir. Körleme kullanılan çalışmaların büyük kısmı placebo kontrollü çalışmalardır ve çift kör uygulanmıştır. Dahil edilen toplam 13 çalışmanın sadece 5 tanesinde post-op ağrıdaki düşüş anlamlı bulunmuştur (%39). Kontrol grubu olarak kullanmak amacı ile seçilmiş körleme yapılmamış çalışmalarda farklı tedavi yönteminin post-op ağrıyı düşürme başarısına bakılmıştır. Dahil edilen 16 çalışmanın 11 inde (%69) tedavilerden biri diğerine göre istatistiksel olarak anlamlı düzeyde etkili bulunmuştur.

Sonuç: Hiçbir tedavinin uygulanmadığı placebo grupları ile tedavi grupları arasındaki farklılığın anlamlı olması beklendiği halde farklılığın büyük kısmı aktif tedavilerin karşılaştırıldığı fakat körlemenin yapılmadığı çalışmalarda gözlenmiştir. Bu bulgu körleme uygulanmayan çalışmalarda fark bulma yönünde yan tutulma eğilimi olduğunu göstermektedir.

Anahtar Kelimeler: Körleme; meme kanseri; placebo; cerrahi

Kaynakça:

1. Jadad, Alejandro R.; Enkin, Murray (2007). Randomized Controlled Trials: Questions, Answers and Musings (2nd ed.). Blackwell. ISBN 978-1-4051-3266-4.
2. Paul J. Karanicolas PJ, Farrokhyar F, Bhandari M. Blinding: Who, what, when, why, how? Can J Surg. 2010 October; 53(5): 345–348.
3. http://www.randomizer.org/form.htm

S31

KONFİGÜRAL FREKANS ANALİZİ

N. Anıl Dolgun[1], Osman Saraçbaşı[1]

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

anilbarak@yahoo.com

Konfigüral frekans analizi - KFA (Configural frequency analysis –CFA), çapraz tabloda yer alan bir gözenin veya gözelerin, altta yatan bir log-lineer modeline göre aykırı olup olmadığının araştırılmasında kullanılan bir yöntemdir [1]. Bu yönüyle çapraz tablolar için aykırı değer incelemesinde kullanılan keşifsel bir veri analiz türüdür.

Konfigüral frekans çözümlemesinde amaç log-lineer ve lojistik regresyon modellerinde olduğu gibi etken/bağımsız değişkenlerin etkilerinin incelenmesinden ziyade çapraz tabloda yer alan örüntülerin araştırılmasıdır. Analizde, eğer çapraz tablodaki bir göze belirtilen model altındaki beklenen sıklıktan daha az sıklık içeriyor ise o gözeye *anti-type*, daha fazla sıklık içeriyorsa da o gözeye *type* adı verilmektedir. Beklenen değerler bağımsızlık modeli altında (1nci derece KFA) hesaplanacağı gibi, kategori kombinasyonlarının eşit olasılığa sahip olduğu model (0ncı derece KFA) gibi çok çeşitli modeller altında hesaplanabilmektedir. Bu şekilde hipotez edilen bir yapıdan istatistiksel olarak önemli kabul edilebilecek kadar sapılıp sapılmadığı araştırılır. Konfigüral frekans analizi uygulamalarına ise daha çok sosyoloji, psikoloji ve davranış bilimlerinde rastlanmaktadır [2 - 4].

Bu çalışmanın amacı, Konfigüral frekans analizini ve bu analizde kullanılan log-lineer yöntemlerini tanıtmaktır. Sağlık bilimlerinden bir örnek veri üzerinde klasik ve Bayesci Konfigüral frekans analizi uygulanacak ve sonuçlar yorumlanacaktır.

Anahtar Kelimeler: Konfigüral frekans analizi, çapraz tablo, log-lineer model, aykırı değer.

Kaynaklar

1- Lienert, G.A ve Krauth J (1975) Configural frequency analysis as a statistical tool for defining types, Educational and Psychological Measurement, 35, pp. 231-238.

2- Alexander von Eye, Christiane Spiel ve Phillip K. Wood Configural Frequency Analysis in Applied Psychological Research Applied Psychology Volume 45, Issue 4, October 1996, Pages: 301–327

3- W.-R. Heilmann, G. A. Lienert and V. Maly Prediction Models in Configural Frequency Analysis, Biometrical Journal Volume 21, Issue 1, 1979, Pages: 79–86.

4- Alexander von Eye, Patrick M. G. Anne Bogat. Prediction models for Configural Frequency Analysis Psychological Science, Vol: 47, 2005 p.342-355.

S32

ASSOCIATION MAPPING USING A BAYESIAN MODEL FOR FEED EFFICIENCY IN F2 MICE DATASET

Burak Karacaören

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

burakkaracaoren@akdeniz.edu.tr

Aim: Recent advances in molecular genetics have provided hundreds of thousands of SNPs (Single Nucleotide Polymorphisms) to detect mutations in genes related with complex traits. Undetected shared ancestry within samples of individuals could lead to the detection of false genomic signals in association mapping. Pedigree-based relationship matrices or genomic relationship matrices could be used in a mixed model to predict and correct for genetic stratifications. The objectives of this study 1) to correct ancestral stratification in an association model similar to the GRAMMAR (Genomewide rapid association using mixed model and regression**)** approach (Aulchenko et al., 2007) using a Bayesian model, and 2) to detect genomic signals by association mapping (Karacaören et al, 2011) and compare results with those obtained by use of pedigree and genomic relationship matrixes for residuals and breeding values.

Methods: An F_2 population (n=661) was created using M16 and ICR mouse lines for studying feed efficiency Ehsani et al., 2012. Genotypes were collected for 1813 SNPs for each animal, including the founders. Bayesian residuals and breeding values were used for population stratification in the association model.

Results: Different from other models; residuals of genomic relationship matrix estimated highest ancestry informative proportions for significant SNPs (with 1000 permutations). SNPs rs6281869 and rs13480815 was in linkage disequilibrium with rs13480530 (Phosphodiesterase 3B) with chi square statistics 190.52 ($p<4.10E-40$) and 288.06 ($p<4.06E-61$). However using additional tests for agreement with inflation factors we could not confirm significance of the SNPs. Additional experiments are necessary to confirm our results.

Conclusion: Bayesian genomic model could be used to assess the genetic and genomic structure in genome wide association studies.

Keywords: Genomics, Bayesian Model, Gibbs Sampling, Association Mapping, Population Stratification, Genetic Analyses

References

Aulchenko SY, de Koning DJ, Haley CS: Genomewide rapid association using mixed model and regression: a fast and simple method for genomewide pedigree-based quantitative trait loci association analysis. Genetics 2007, 177: 577-585

Ehsani A, Sørensen P, Pomp D, Allan M, Janss L (2012) Inferring genetic architecture of complex traits using Bayesian integrative analysis of genome and transcriptome data. BMC genomics 13(1) 456.

Karacaören B, Silander T, Álvarez-Castro J, Haley C, de Koning DJ (2011) Association analyses of the MAS-QTL data set using grammar, principal components and Bayesian network methodologies. BMC Proc (Vol. 5, No. Suppl 3, p. S8)

S33

MIDDLESEX ELDERLY ASSESSMENT OF MENTAL STATE (MEAMS) ÖLÇEĞİNİN DEĞERLENDİRİLMESİ: KARMA RASCH MODEL YAKLAŞIMI

Selcen Yüksel[1], Atilla Halil Elhan[2], Şehim Kutlay[3]

[1]*Yıldırım Beyazıt Üniversitesi Tıp Fakültesi Biyoistatistik ve Tıbbi Bilişim ABD, Ankara*

[2]*Ankara Üniversitesi Tıp Fakültesi Biyoistatistik ABD, Ankara*

[3]*Ankara Üniversitesi Tıp Fakültesi Fiziksel Tıp ve Rehabilitasyon ABD, Ankara*

selcenpehlivan@yahoo.com

Rasch modelleri (RM) yaşam kalitesi, tutum ve özürlülük düzeyi gibi gizli değişkenlerin değerlendirilmesinde kullanılmaktadır. Rasch modellerinin sağlık alanında ölçek geliştirmek amacıyla kullanımı standart bir yöntem haline gelmiştir. Bu nedenle, bu modellere ilişkin çalışmalar hızla artmaktadır. Rasch analizinde sıkça karşılaşılan sorunlardan biri maddelerin işleyişlerinin farklı olabilmesidir. Başka bir anlatımla, incelenen özellik düzeyi sabit tutulduğunda, veri setindeki bazı maddelere verilen yanıtlar, bireylere ait cinsiyet, yaş ve hastalık süresi gibi bazı faktörler açısından yanlılık gösterebilmektedir. Bu yanlılığa Rasch literatüründe madde işlev farklılığı (MİF) adı verilmektedir. Bu sorunu gidermek için öngörülen çözümlerden biri Karma Rasch Modelidir (KRM).

KRM ilk kez, iki sonuçlu maddeler için, 1990 yılında Jürgen Rost tarafından tanımlanmıştır. Öngörülen modelin, RM'nin teorik gücü ile Gizli Sınıf Analizi' nin (GSA) deneysel gücünü birleştirdiği ifade edilmiştir. Modelin, RM' nin ve GSA' nın varsayımlarını ortadan kaldırdığı öngörülmüştür. GSA'da her bir gizli sınıf içinde maddelerin zorluk sırasının aynı olması koşulu vardır. KRM ise farklı gizli sınıflarda madde zorluklarının sırasının da farklılaşmasına olanak sağlar. Dolayısıyla bu modelde farklı gizli sınıflara atanan bireyler için madde zorlukları da farklılaşabilecektir. KRM' de çok sonuçlu maddeler için parametre kestirimleri ise ilk kez von Davier ve Rost (1995) tarafından öngörülmüştür. KRM, RM' nin kesikli karma dağılım modeline genelleştirilmesi ile elde edilir. Kısaca açıklanırsa, KRM, RM' nin tüm popülasyon için uyumlu olmayacağını ancak önceden bilinmeyen, bireylerin alt popülasyonları için uyumlu olacağını varsayar.

c. sınıfta olduğu bilinen i. bireyin j. maddeden x puanı alma olasılığı aşağıdaki Karma Rasch Model eşitliği ile verilir. πc c. sınıfa atanma olasılığıdır. θ_{ic}, c.sınıfta olduğu bilinen i. bireyin gizli özellik düzeyidir. B_{jkc} ise c sınıftaki j maddenin k. eşik zorluk düzeydir.

$$P_{ijx|c} = \sum_{c=1}^{C} \pi_c \frac{\exp \sum_{k=0}^{x} \left(\theta_{ic} - \beta_{jkc}\right)}{\sum_{k=0}^{m_i} \exp \sum_{t=0}^{k} \left(\theta_i - \beta_{jtc}\right)} \qquad , x=0,1,\ldots m_i$$

Bilişsel bozukluğun değerlendirilmesi nörorehabilitasyon uygulamasında oldukça önemlidir. Bilişsel bozukluğu değerlendirmek için uygulanan tarama testlerinden en fazla kullanılanı Mini-Mental State Examination ölçeğidir. (MMSE). MMSE, nörolojik bir hastayı rutin değerlendirirken verdiği değerli, tutarlı ve hızlı sonuçları nedeniyle tercih edilmesine rağmen odak eksikliklerini belirlemedeki kısıtlılıkları ve ön lop hastalıklarındaki duyarsızlığı bilinmektedir. Bu kapsamda bilişsel bozukluğu değerlendirmede kullanılan Middlesex Elderly Assessment of Mental State (MEAMS) ölçeğinin MMSE'ye göre daha kapsamlı olduğu düşünülmektedir. MEAMS yaşlılarda spesifik bilişsel yeteneklerdeki bozulmaları belirlemek için kullanılan bir tarama testidir. Bu testin 12 alt testi vardır. Her bir alt test için bir geçme skoru hesaplanır. 12 alt testten elde edilen tarama skorlarının toplamı ise toplam tarama skoru olarak verilmektedir.

Bu çalışmanın amacı KRM' nin bilişsel bozukluğu değerlendirmede kullanılan MEAMS ölçeği üzerine uyumunu değerlendirmek ve KRM ile MEAMS ölçeğindeki maddelerinin işlev farklılığı gösterip göstermediğini sorgulamaktır. Bu amaca yönelik Ankara Üniversitesi Tıp Fakültesi Fiziksel Tıp ve Rehabilitasyon Anabilim Dalı başvuran 155 beyin hasarlı hastanın içerildiği veri seti KRM ile değerlendirilecektir.

Anahtar Kelimeler: Karma Rasch Model, madde işlev farklılığı, MEAMS

S34

ERYTHROMYCIN İLACININ YAN ETKİLERİNİN ARAŞTIRILMASI İÇİN WEB'DEKİ HASTA YORUMLARININ ANALİZİ

Erhan Tahminciler[1], Pınar Yıldırım[2]

[1]*Okan Üniversitesi, Fen Bilimleri Enstitüsü, Bilgisayar Mühendisliği Yüksek Lisans Programı, İstanbul*

[2]*Okan Üniversitesi, Mühendislik ve Mimarlık Fakültesi, Bilgisayar Mühendisliği Bölümü, İstanbul*

erhan@tahminciler.net

Günümüzde tıp ve biyoloji alanında veri madenciliği uygulamalarının önemi gün geçtikçe artmaktadır ve bu çalışmalar sayesinde hem hastalıkların tanınmasında hem de tedavisinde daha başarılı sonuçlar elde edilmektedir. Özellikle hastalıkların tedavisinde kullanılan ilaçların, hastalar üzerindeki etkilerinin araştırılması önem kazanmıştır. Ülkemizde bu araştırmalar az sayıdadır ve bu çalışmada web üzerindeki çok sayıda gerçek hasta bilgileri kullanılarak ilaç yan etkileri ile ilgili gizli bilgilerin keşfedilmesi üzerine bir veri madenciliği çalışması yapılmıştır.

Amaç: Bu çalışmanın amacı, ilaç kullanımları sonrasında hastalar üzerindeki yan etkilerin araştırması için veri madenciliği yöntemleri uygulanarak tıp uzmanlarının ve araştırmacılarının karar alma süreçlerine yardım sağlaması, ilacı kullanan kişiler için hangi cinsiyet ve yaş aralığında hangi dozaj ve sürelerde kullanımı ile en iyi sonucun bulunması öte yandan ilaç üreten firmalara da araştırmalarında çeşitli katkılar sağlaması hedeflenmiştir.

Yöntem: Bu çalışmada, Helikobakter Pilori ve Mycobacterıum Tuberculosis gibi dünyada yaygın olarak görülen bakterilerin tedavisinde kullanılan Erythromycin ilacına ait www.askapatient.com web sitesi üzerinde bulunan hasta yorumları analiz edilmiştir. Yapılan çalışmada, hasta yaşı, cinsiyeti, kullanım süreleri, kullanım dozları, kullanım nedenleri, kullanım süresi boyunca görülen etkiler ve hastanın ilaca verdiği puan bilgileri ayrı ayrı değerlendirilerek bir veriseti oluşturulmuştur. Genel olarak Sql Server 2012, Visual Studio C#.Net ve WEKA yazılımları kullanılmıştır. Oluşturulan veri setinde 711 kayıt ve 8 nitelik bulunmaktadır. Web üzerindeki hasta yorumları içindeki kullanım nedenleri ve görülen etkiler alanları, metin parçalama yöntemleri ile ayrıştırılarak nitelikler oluşturulmuş ve oluşturulan bu veriseti üzerine Apiori algoritması uygulanarak birlikte görülen örüntüler keşfedilmeye çalışılmıştır.

Bulgu: Erythromycin ilacı hakkında yorum yapan hastaların cinsiyet dağılımı %25 erkek, %75 kadın oranındadır ve bununla birlikte ilacı kullanan kişilerin yaş aralıkları belirlenerek yaş grubunda en az kullanım olan 4 kişi ile 70-75, en çok kullanım olan yaş grubu ise 135 kişi ile 25-30 bulunmuştur. Ayrıca hasta yorumları içinde en çok tekrar eden yan etkilerin frekansları belirlenmiştir. İlaç yaş grubu kullanım sayılarının belirlenmesi ile ilgili bakterilerin yaş grupları üzerindeki etkileri gözlenmiştir.

Apiori algoritması ile veri seti üzerinde doğrudan belirlenemeyen fakat birlikte görülen yan etkiler de saptanmıştır. Örneğin stomach(mide), cramps (kramplar), nausea(mide bulantısı), diarrhea(ishal) sıklıkla birlikte karşılaşılan yan etkilerdir.

Sonuç: Bu çalışma; www.askapatient.com web sitesinden Erythromycin ilacı hakkında yapılan yorumlardan oluşturulan veri seti, kural tabanlı metin parçamala ve Apiori algoritması ile analiz edilmiştir.Analiz sonucunda elde edilen bilgiler, ilaç üreten firmaların araştırmalarına, tıp uzmanları ve araştırmacıları için hastalar üzerinde karar verme sürecine katkı sağlayacak yararlı bilgiler sunmaktadır.

Anahtar kelimeler: Biyomedikal metin madenciliği, Erythromycin, Yan etkiler, Apiori algoritması

S35

MINITAB VE FREUD & PERLES KARTİL TAHMİN YÖNTEMLERİNİN ÖRNEKLEME DAĞILIMLARI VE PERFORMANSLARININ KARŞILAŞTIRILMASI: BİR SİMÜLASYON ÇALIŞMASI

Şengül Cangür[1], **Özge Pasin[1]**, Handan Ankaralı[1]

[1]*Düzce Üniversitesi Tıp Fakültesi, Biyoistatistik ve Tıbbi Bilişim Anabilim Dalı, Düzce*

ozgepasin90@yahoo.com.tr

Amaç: Kartiller, veri setindeki aykırı değerleri belirlemek, çarpık dağılımlarda varyasyon ölçüsünü tahmin etmek ve dağılım şekli hakkında bilgi edinmek amacıyla sıklıkla kullanılmaktadır. Literatürde, kartil hesaplaması için çok sayıda farklı yöntem bulunmaktadır. Birçok yazılım ve programda farklı kartil yöntemleri tanımlandığından, kullanılan programa bağlı olarak aynı veri seti için farklı kartil değerleri elde edilebilir. Bu çalışmada, kartil hesaplama yöntemleri kısaca açıklandı ve Minitab (Minitab, SAS-4) ve Freud & Perles (S-Plus, Microsoft Excel) kartil kestiricilerinin performanslarının karşılaştırılması amaçlandı.

Yöntem: Simülasyon çalışması, tek ve çift sayıda gözlem içeren farklı örneklem büyüklüklerinde (n=9, 10, 99, 100) ve 3 farklı dağılım şeklinde (standart normal, ki-kare ve küçük uç değer dağılımları) tasarlandı. Minitab ve Freud & Perles kartil tahmin yöntemleriyle elde edilen birinci kartil (Q_1), üçüncü kartil (Q_3) ve kartiller arası aralık (interquartile range-IQR) değerlerinin tanımlayıcı özellikleri ve örnekleme dağılımları yardımıyla kartil kestiricilerinin performansları karşılaştırıldı.

Bulgular: Simülasyon çalışması sonucunda tek veya çift sayıda gözlem içeren küçük örneklemlerde hesaplanan Q_1 ve Q_3 değerleri açısından iki yöntem arasında anlamlı fark bulunmasına karşın büyük örneklemlerde yöntemlerin tahminleri benzer çıktı. Freud & Perles yöntemiyle elde edilen Q_1 ve Q_3 değerlerinin %95 güven aralıkları, Minitab yöntemiyle elde edilenlere göre daha dardı. Ayrıca örneklem büyüklüğü arttırıldığında her iki yöntemde de Q_1 ve Q_3 tahmin değerlerinin güven aralıkları daralmaktaydı. Her iki yöntemde de, dağılımın standart sapması büyüdükçe IQR değerlerindeki varyasyonun arttığı gözlendi. Bu varyasyon özellikle küçük örneklemlerde daha fazla idi. Bunun yanı sıra bu iki varyans tahmini arasında doğrusal bir ilişkinin olmadığı belirlendi. Kestiricilerin örnekleme dağılımlarının şekli, türetildiği populasyon dağılımlarının şekline oldukça benzerdi.

Sonuç: Veri dağılımı simetrik ve örneklem büyüklüğü küçük olduğunda Minitab kartil yönteminin tercih edilmesi önerildi. Buna karşın aykırı değerlerin fazla olduğu asimetrik dağılımlarda ve küçük örneklem hacimlerinde Freud & Perles kartil yönteminin kullanılmasının daha uygun olduğu belirlendi. Ayrıca büyük örneklemlerde Q_1 ve Q_3 açısından iki kartil kestiricisinin de benzer performansa sahip olduğu görüldü.

Anahtar Kelimeler: Kartil, Minitab, Freud & Perles

S36

EN İYİ KANIT SAĞLAYAN KLİNİK DENEMELER: RONALD AYLMER FISHER VE AUSTIN BRADFORD HILL'İN KATKILARI

M. Yusuf Çelik[1], İsmail Yıldız[1], Ömer Satıcı[1], Zeki Akkuş[1]

[1]*Dicle Üniversitesi Tıp Fakültesi Biyoistatistik Anabilim Dalı*

myusufcelik@hotmail.com

Araştırma dizaynında kontrol veya deneysel gruba girecek hastaların rasgele atanmaları, istenmeyen varyasyonu elemine etmek için çok önemlidir. Araştırıcı rasgelelikle ilgili yöntemleri uygulayarak sistematik yanlılığı azaltarak iç geçerliliğin artmasını amaçlar. Çalışmaya alınan kontrol ve hasta grubundaki hastalar rasgele atanmamış ise bu durumda araştırmadaki iç geçerliliğin zayıflayacağı ve yanlış sonuçlar elde edileceği bir gerçektir.

Bilindiği gibi istatistiğin kurucusu olan ve teorileri eleştirilemeyen RA Fisher, rastgeleliğin istatistiksel anlamlılık testleri için, nasıl bir teorik destek sağladığını ilk kez gösteren bilim adamıdır. Bu nedenle, rastgele klinik denemelerin değerlendirilmesinde anahtar bir rol oynadığı kabul edilir.

Austin Bradford Hill ile ilk kez tıbbi çalışmalara uyguladığı rastgeleliği, Tıbbi Araştırma Konseyi (Medical Research Council-MRC) ancak daha pragmatik bir nedenle kabul etmiştir.

Konu kronolojik olarak incelendiğinde, bu konuda yayın yapan Peter Armitage and Richard Doll (Bradford Hill'in eski çalışma arkadaşları), Walter Bodmer (Fisher'in eski öğrencisi), ve Harry Marks (klinik denemelerin tarihsel yazarı) yazılarında "Fisher'in bir teorisyen" ve "Austin Bradford Hill'in ise bir pragmatist" olduğunu belirtmişlerdir.

Harry Marks RCT (Random Clinical Treatment), Rasgele Klinik Denemelerin Amerika Birleşik Devletleri'nde farklı bir evrimsel yol açtığını yazdı. Ayrıca, bu durumun modern tıp alanında en iyi kanıt sağlayan istatistiksel yöntemlerin, ne kadar "erdemli" olduğunu gösterdi. Laboratuar bilim dalları klinik denemelerle tıp için "iyi kontrol edilebilir" bir standart oluştururken, klinik araştırmalar ise Rasgele Klinik Denemeleri bilimsel bir üstünlük sağlayan yeni bir standart olarak uyguladılar. Böyle bir standart tedavide güven oluşturan ilaçların üretimini sağladı. Bu durum ilaç endüstrisine yeni bir kalite getirdi.

Sonuç olarak, Rasgele Kontrollü Çalışmaların tedavi veya klinik bir karar için geleneksel olarak bir altın standart sonuçlar oluşturduğu söylenebilir.

And again in *The Design of Experiments*4 (the quotation being from p. 26 of the 6th edition, 1951):

The purpose of randomisation is to guarantee the validity of the test of significance, this test being based on an estimate of error made possible by replication.

Marks argues that while laboratory sciences provided the standard of the 'wellcontrolled' experiment for medicine, during the second half of the century the clinically based randomised controlled trial offered a new standard of scientific excellence.

They stand as evidence of the virtue of statistical tools and quantification in bringing about progress in modern medicine. On the other hand, critics of the methodology and conduct of the clinical trial tend to draw on philosophical or transhistorical traditions to construct their arguments.

The papers in this issue of the *International Journal of Epidemiology*—by Peter Armitage and Richard Doll (formerly colleagues of Bradford Hill), Walter Bodmer (formerly a student of Fisher), and Harry Marks (author of a history of clinical trials)—were commissioned to reflect on this characterization of 'Fisher the theoretician' and 'Bradford Hill the pragmatist'.

Austin Bradford Hill son 50 yılda tıp alanında Nobel ödülü kazananlar arasında çok daha etkin bir kişilikti. Tıbbi istatistik konularında 1937 yılında Lancet'te seri halinda yaptığı yayınları "Principles of Medical Statistics" kitabında topladı. Yayınlarda ilk kez klinik denemeler için "Rastgeleliğin nasıl yapılacağını" tanıttı.

1930-50 yılları arasında başta Hill olmak üzere diğer değerli biyoistatistik uzmanlarının yol göstermeleri sonucu biyoistatistik tıp bilimleri arasına katılmış ve ayrılmaz bir parçası olmuştur. Tıp, eczacılık, diş hekimliği fakültelerinde ve diğer sağlık kuruluşlarında biyoistatistik bölümleri kurulmuştur.

Bradford Hill ilk kez retrospektif ve prospektif çalışmaları düzenlemiş ve tüm araştırıcılara tanıtmıştır. Daha sonra bu yöntemlerin adını Vaka-kontrol ve kohort çalışmalar olarak tanıtmıştır. Bu yöntemlerle birden çok faktörlerin etkileri dikkate alınabilmiştir.

Akciğer kanserine atfedilen ölümlerin artmasının nedenlerini incelemek için düzenlediği araştırmada Hill "vaka-kontrol" ve "kohort" tipi çalışmalar için model geliştirmiş ve elde edilen bir ilişkinin neden-sonuç ilişkisi olup olmadığının saptanması için bir kılavuz hazırlamıştır.

Hill neden-sonuç ilişkisinin incelenmesinde dikkate alınacak dokuz özellik ortaya koydu: Bunlar: "Ilişkinin gücü", "tutarlılık (consistency)", "seçicilik (specificity)", "zaman ilişkileri", "biyolojik değişim (biological gradient)", "biyolojik olasılık (biological plausibility)", "olayın uygunluğu (coherence of evidence)", "deney" ve "karşılaştırma (analogy)". Hill bu özelliklerin hiçbirisinin neden-sonuç ilişkilerine tartışmasız açıklama getiremeyeceğini, yalnızca neden sonuç ilişkilerini tanımlamada yardımcı olabileceğini de belirtmiştir.

S37

YAKALANMA OLASILIKLARININ HETEROJEN OLDUĞU TOPLULUKLARDA SAMPLE COVERAGE YAKLAŞIMI İLE POPULASYON BÜYÜKLÜĞÜNÜN TAHMİNLENMESİ

Selin Bayraktaroğlu[1], Timur Köse[1]

[1]*Ege Üniversitesi, Tıp Fakültesi, Biyoistatistik ve Tıbbi Bilişim AD. Bornova, İzmir*

selin.bayraktaroglu@ege.edu.tr

Amaç: 1700'lü yıllarda kullanılmaya başlanan Yakala-Tekrar Yakala yöntemlerinin ilk ortaya çıkma sebebi hayvan populasyon büyüklüğünü tahmin etmekti. Açık ve Kapalı populasyonların her ikisi için de uygulanılabilen bu yöntemler, çeşitli varyasyon kaynakları içeren modeller ile açıklanmıştır. Geçmişten günümüze bu yöntemlerin insan populasyonları üzerindeki uygulamaları dolayısıyla sağlık alanına uyarlanabilmesi önem kazanmıştır. Sample Coverage Approach populasyon büyüklüğünü tahmin etmek için, literatürde adı geçen yöntemlerden biridir. Bu yöntem altında 3 farklı tahmin edici bulunmaktadır. Her bir tahmin edici ile belli koşullar altında güvenilir tahminler elde edilebilmektedir. Bu çalışmada, literatürde geçen bu yöntem ile kapalı populasyonlar için belli varsayımlar altında parametre tahminleri yapılmış ve güven aralıkları hesaplanmıştır. Sample Coverage (C) ölçütü için literatürde var olan 0.55 sınırı araştırılmıştır.

Yöntem: Tahmin edicilerin performans değerlendirmesi için, N=200 populasyon büyüklüğü ile 3 listeli veri setleri üretilmiştir. Yakalanma olasılıkları ise heterojenliği sağlayabilmek için (p=0.1; 0.3; 0.5; 0.7) değerlerinden istatistiksel paket program yardımı ile rasgele olarak seçilmiştir. Sample Coverage (C) ve bağımlılık ölçütleri (CCV) sayesinde hangi tahminin en iyi olduğuna karar verilmiştir. Ek olarak göreli yan(relative bias) miktarları hesaplanarak ulaşılan sonuçlar desteklenmiştir.

Bulgular: N=200 için üretilen veri setlerinin genelinde gerçek değerden düşük tahminler elde edilmiştir. Sample Coverage (C) ölçütü ortalaması, literatürdeki sınır değeri olan 0.55'in üzerinde hesaplanmıştır. Göreli yan(relative bias) miktarı ortalaması '0'(sıfır) değerine en yakın olan $\hat{N}$ tahmin edicisi olarak bulunmuştur. Hesaplanan standart hatalar ise kabul edilebilir miktarlardır.

Sonuç: $\hat{N}_0$, $\hat{N}$, $\hat{N}_1$ tahminleri göz önüne alındığında, $\hat{N}$ tahmin edicisi en iyi sonuçları vermiştir. $\hat{N}$ tahmin edicisinin göreli yan miktarları ortalaması diğerlerine göre '0' değerine daha yakındır. $\hat{N}_0$ tahmin edicisi ise bağımsızlık varsayımı altında kullanılmaktadır. Yakalanma olasılıklarının farklılığından kaynaklanan bireyler arası heterojenlik, varsayımın bozulmasına sebep olmuştur. Hesaplanan CCV değerleri değerlendirildiğinde veri setlerinin bağımlı olduğu görülmüştür. Bu çalışma genişletilerek düşük yakalanma olasılıkları ile tekrar düzenlenilebilir. Sample Coverage (C) değerinin düşük olduğu durumlarda da sonuçlar tartışılmalıdır.

Anahtar Kelimeler: Yakala-Tekrar Yakala, Sample Coverage Approach, Göreli yan(relative bias)

S38

NORMAL DAĞILIMA AİT DEĞERLENDİRMELERDE UYGUN YÖNTEM SEÇİMİNİN ÖNEMİ

THE IMPORTANCE OF THE CHOICE OF THE APPROPRIATED METHOD FOR NORMALITY

Hayriye Ertem-Vehid[1], Nurgül Bulut[2], Mustafa Şükrü Şenocak[3]

[1]*İstanbul Üniversitesi Çocuk Sağlığı Ens. ve Cerrahpaşa Tıp Fak. Biyoistatistik ve Tıp Bilişimi Anabilim Dalı*

[2]*Yeni Yüzyıl Üniversitesi Biyoistatistik Anabilim Dalı*

[3]*İstanbul Üniversitesi, Cerrahpaşa Tıp Fak. Biyoistatistik ve Tıp Bilişimi Anabilim Dalı*

ertem@istanbul.edu.tr

Bilimsel çalışmalarda elde edilen verilerin değerlendirilmesi için kullanılacak biyoistatistiksel yöntemin seçiminde verilerin normal dağılıma uygun olup, olmaması önem taşımaktadır.

Bilindiği gibi, verilerin normal dağılıma uygunluğunun değerlendirilmesinde Shapiro-Wilk, Kolmogorov-Smirnov, Anderson Darling vb. farklı yöntemler vardır.

Verilerin normal dağılıma uygunluğunu değerlendiren bu testlerin uygulanmasıyla ulaşılan sonuç yargılarda verilerin yer aldığı dağılıma ait teorik dağılımın bilinmesi, örnek sayısı, verilerin çarpıklık ve basıklık durumlarının dikkate alınması önem taşımaktadır.

Öncelikle çalışmamızda, aynı standart sapmaya sahip ortalamaları farklı olan ancak eşit düzeyde simetrik yapıda olan verilerin normalliğinin değerlendirilmesinde bile testlerin normal dağılımlarının farklı olabileceği varsayımıyla N(100 10);(110 10), N(100 10);(120 10) ve N(100 10);(140 10) düzenlerinde veri üretilerek test sonuçlarının farklı olup/olmadığının değerlendirilmesi hedeflendi.
Daha sonra, NCSS paket programında, n=50, n=100 ve n=500 örnek büyüklükleri ve aşağıda gözlenen düzenlerin her biri için;

N (100 10) [75]; N (120 10) [25],
N (100 10) [75]; N (140 10) [25],
N (100 10) [75]; N (80 10) [25],
N (100 10) [75]; N (60 10) [25] simülasyonla 200'er veri türetildi. NCSS paket programında yer alan tüm normal dağılıma uygunluk testlerinden elde edilen sonuç yargıların benzerlik veya farklılığının değerlendirilmesinin hedeflendiği bu çalışma planlandı.

Sağa veya sola çarpıklığın az ve örnek sayısı (n=50) az olduğunda verilerin normalliğe ait uygunluk değerlendirilmelerinin D'Agostino Omnibus dışındaki tüm testler için aynı yönde olduğu

gözlemlenirken sağa ve sola çarpıklığın arttığı durumlarda özellikle Shapiro-Wilk ve Anderson Darling testlerinin normal dağılıma uygun sonuçlar göstermediği gözlendi.

Örnek sayısının artması durumunda ise Martinez Iglewicz ve D'Agostino Kurtosis testlerinin sağa ve sola çarpıklığın az olduğu durumda normal dağılıma uygunluğu red edemedikleri gözlendi.

Sonuç olarak, verilerin normal dağılıma uygunluğunun değerlendirilmesinde örnek sayısının büyüklüğü yanı sıra verilerin sağa veya sola çarpıklığının düzeyinin önemli olduğu gözlendi.

Anahtar kelimeler; normal dağılım, çarpıklık, normal dağılım testleri

S39

TEKRARLI FİZYOLOJİK ÖZELLİKLERDE DÜZEN DEĞİŞİKLİKLERİNİN YARGILANMASI

Mustafa Şenocak[1], **Nurgül Bulut[2]**, Alev Bakır[1], Günalp Uzun[1], Muzaffer Öztürk[3]

[1]*İstanbul Üniversitesi, Cerrahpaşa Tıp Fak. Biyoistatistik ve Tıp Bilişimi Anabilim Dalı*

[2]*İstanbul Üniversitesi, Cerrahpaşa Tıp Fakültesi, Biyoistatistik ve Tıbbi Bilişim ABD*

[3]*İstanbul Üniversitesi, Cerrahpaşa Tıp Fak. Kardiyoloji - (Em.) Anabilim Dalı*

nrgl.bulut@hotmail.com

Amaç: Biyolojide, sağlıklı insanlarda periyodik istemsiz (otonom) hareketlerin ölçümlerindeki aralıkların eşit olması beklenirken ölçümler arası düzen değişiklikleri incelenen zaman aralıklarına bağlı olarak bazen görmezden gelinebilir. Fakat toplamda fark edilmeyen bu değerler daha dar zamanlarda incelendiğinde daha farklı sonuçlardan bahsedebilir.

Yöntem: Örneğin kalp atışı, yutkunma, nefes alıp-verme, peristaltik hareketler periyodik istemsiz hareketlerden en basit olarak bilinenleridir. Tekrarlı ölçüm olarak kabul edilen kalp atışı için ortalama 800 ms, 1 saatlik 5'er dakikalık aralıklar genel referans değere göre 5 dakika az, 10 dakika az, 15 dakika az, 1 dakika az bir kısma göre farklı değişikliklerin ölçümleri elde edilmiştir. 1 saatlik referans değeri, 800±50 ms li, 375 te bir atımlar 12 kez simülasyonla elde edilmiştir. Referans değere göre çeşitli düzenlerde 1000±50 ms lik ve 1200±100 ms lik 6 farklı düzen oluşturulmuştur.

Bulgular: Kalp atımı ölçüm aralıkları ne kadar az sürede belirgin değişiklikleri, hangi ölçütlerde farklılıklar çıktığı incelenmiştir. Değişik atımlar olduğunda belli tip değişikliklerin nasıl yansıdığı incelenmiştir. NNO, SDANN, SDNN gibi indekslerin ölçüm yaklaşımları belli farklılaşma tiplerindeki sonuç farklılaşmaları değerlendirilmiştir.

Sonuç: Kalp atımı değişkenliği üzerinden yapılan değerlendirmede, farklı ölçüm aralıklarında yeterli düzeyde sürekli ölçümler kayıt edilerek titiz inceleme sonucunda farklı tanılamalar elde edilebilmektedir, bu yöntem tüm tekrarlı fizyolojik özelliklerde düzen değişikliği olabilecek durumlar için de geçerli olabilir.

Anahtar Kelime: Tekrarlı Ölçüm, Düzen Değişiklikleri Yargılaması, Fizyolojik Atımlar

S40

BİYOİSTATİSTİK ANABİLİM DALLARINDA BOLOGNA SÜRECİ BAĞLAMINDA İÇ KONTROL SİSTEMİ UYGULAMALARI

İsmail Yıldız

Dicle Üniversitesi, Tıp Fakültesi, Biyoistatistik ve Tıbbi Bilişim Anabilim Dalı

iyildiz21@yahoo.com

Üniversitelerin Uluslararasılaştırılması, Rekabetçi ve dinamik bilgi tabanlı bir Avrupa ekonomisi hedefi doğrultusunda "Avrupa Yükseköğretim Alanı"nı oluşturmak ve bu kapsamda Avrupa boyutunda yükseköğretim kurumlarının yeniden yapılanması ve yükseköğretimde şeffaflık, hareketlilik ve akademik derecelerin tanınmasını sağlamak için Avrupa düzeyinde 29 ülkenin Eğitim Bakanlarının 19 Haziran 1999 tarihli ortak deklarasyonu ile Bologna'da başlatılan ve 2010 yılında tamamlanması öngörülen süreçtir.

Uluslararası düzeyde kabul gören iç kontrol, kurumun hedeflerine ulaşması için makul güvence sağlamak üzere tasarlanmış olan bir sistemdir. Bu sistemin en iyi bilinen modeli olan Committee of Sponsoring Organizations (COSO) çerçevesinde iç kontrol; kurumdaki iş ve eylemlerinin mevzuata uygunluğunu, mali ve yönetsel raporlamanın güvenilirliğini, faaliyetlerin etkililiği ve etkinliği ile varlıkların korunmasını sağlamayı amaçlar. COSO tarafından oluşturulan ve kontrol ortamı, risk yönetimi, kontrol faaliyetleri, bilgi ve iletişim ile izleme bileşenlerinden oluşan iç kontrol sistemi, Uluslararası Sayıştaylar Birliği (INTOSAI), Avrupa Komisyonu ve benzer uluslararası kuruluşlarca da referans olarak kabul edilen bir modeldir.

Biyoistatistik AD tarafından sunulan hizmetlerin ve faaliyetlerin çeşitliliği de dikkate alınarak; Hizmet sunulmasını engelleyecek veya hizmetin kalitesini düşürecek, Hizmet alan, Öğrenci, Tedarikçi ve Çalışan Güvenliğini sarsabilecek, Yolsuzluk yapılmasına meydan verecek, Mevzuata aykırılığa imkan tanıyacak, kurum içinden ya da kurum dışından oluşan her türlü olay Risk olarak adlandırılabilir.

Çalışanların, Müşterilerin, Hissedarların ve Tedarikçilerin beklentilerini dengeli bir şekilde karşılamak üzere, Biyoistatistik Anabilim Dallarının bütün faaliyetlerinin sürekli olarak iyileştirmesini koordine edip; dünya çapında kalite seviyesine ulaşmasını ve rekabet gücünün artması için uygun ortamlar oluşturması bağlamında; Akademik Değerlendirme ve Kalite Geliştirme, Stratejik Yönetim, Kurumsal Değerlendirme, Periyodik İyileştirme ve İzleme, İnovasyon ve Akreditasyon konularında Anabilim Dalı içinde ve dışında koordinasyonu sağlamak amacıyla, Biyoistatistik Anabilim Dallarında Bologna sürecine paralel olarak "5018 Sayılı Kamu Mali Yönetimi ve Kontrol" kanunu kapsamında İç Kontrol Sistemi kurulmalıdır.

Bu çalışmada, Bologna sürecine paralel olarak Biyoistatistik Anabilim Dallarında İç Kontrol Sisteminin kurulmasının ve sürekliliğinin sağlanmasının detaylarıyla anlatılması amaçlandı.

Anahtar Kelimeler: Bologna Süreci, Risk Yönetimi, İç Kontrol, 5018 Sayılı Kanun

S41

MODELE EKLENEN YENİ BİR TAHMİN EDİCİNİN MODEL PERFORMANSINA KATKISININ DEĞERLENDİRİLMESİ: KARDİYOVASKÜLER RİSK TAHMİN MODELİNE İLİŞKİN BİR UYGULAMA

Özlem Güllü[1], Can Ateş[2],Mustafa Agah Tekindal[3], Beyza Doğanay Erdoğan[2], Yasemin Yavuz[2], İ. Safa Gürcan[1], Berkay Ekici[4]

[1]*Ankara Üniversitesi Veteriner Fakültesi, Biyoistatistik Anabilim Dalı, Ankara*

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

[3]*Başkent Üniversitesi Tıp Fakültesi, Biyoistatistik Anabilim Dalı, Ankara*

[4]*Ufuk Üniversitesi Tıp Fakültesi, Kardiyoloji Anabilim Dalı, Ankara*

ozlem.gullu@gmail.com

Amaç: Sağlık alanında yaygın olarak kullanılan tahmin modelleri, hastalığın izlenmesi ve zaman içinde gelişiminin ileri evrelere göre tahmin edilmesi, tanı konması ve uygun tedavi yöntemine karar verilebilmesi için risk gruplarının belirlenmesi gibi amaçlar için kullanılır. Bu çalışmanın amacı, yeni bir tahmin edicinin performansının değerlendirilmesinde kullanılan yeni ölçüler ile klasik ölçülerin karşılaştırılmasıdır.

Yöntem: Çalışmada, var olan bir referans modele yeni bir tahmin edici eklendiğinde bu tahmin edicinin model performansına etkisinin ölçülmesi üzerine durulmuştur. Bu amaçla, çalışmada koroner kalp yetmezliği riskine sahip 293 hastadan elde edilen veri seti kullanılmıştır. Hastalardan elde edilen yaş, cinsiyet, yüksek tansiyon, hiperlipidemi, sigara kullanımı ve toplam kolesterol/hiperlipidemi oranı gibi değişkenler, klinik risk faktörleri olarak alınmış ve referans model oluşturulmuştur. Daha sonra, referans modele MPV (Mean Platelet Volume) değişkeni eklenerek yeni bir model elde edilmiştir. Koroner kalp yetmezliği hastalığına sahip bireylerin tahmin edilmesinde, MPV değişkeninin sınıflandırma gücünün değerlendirilmesi için, veri seti lojistik regresyon yöntemi ile analiz edilmiştir ve tekrar sınıflandırmaya ilişkin indeksler olan NRI (Net Reclassification Improvement Index) ve IDI (Integrated Discrimination Improvement Index) hesaplanmıştır.

Bulgular: Referans modelde eğri altında kalan alan (AUC) 0.77 olarak hesaplanmıştır. Yeni modelde ise, sadece MPV değişkeninin eklenmesi ile bu değer 0.81'e yükselmiştir. İki model arasındaki gözlenen bu artış istatistiksel olarak anlamlı bulunmuştur (p=0.011). Ayrıca yeni modele ilişkin doğruluk oranı da 0.67'den 0.74'e yükselmiştir. Tekrar sınıflandırma yöntemiyle elde ettiğimiz indekslere ilişkin sonuçlar ise sırasıyla NRI=0.1039 (SE = 0.0364, p = 0.004) ve IDI= 0.10388 (SE=0.0358 p=0.003) olarak bulunmuştur. MPV'nin modele eklenmesi ile elde edilen doğru sınıflandırma gücündeki artış, klasik yaklaşım olan eğri altında kalan alanla değerlendirildiğinde yaklaşık %4 iken yeni geliştirilen NRI ve IDI indeksleri bu artışı %10 olarak bulunmuştur.

Sonuç: Sonuç olarak, klinik risk faktörlerini ve MPV değişkenini içeren modelin referans modele göre performansının daha iyi olduğu gözlenmiştir. Bu bağlamda, yeni geliştirilen bu performans indeksleri (NRI ve IDI), klasik yaklaşıma (ROC-AUC) nazaran daha duyarlı sonuçlar üretmektedirler.

Anahtar Kelimeler: Performans ölçüleri, Lojistik regresyon, Tekrar sınıflandırma, Risk değerlendirme, ROC-AUC, NRI, IDI

II. POSTER BİLDİRİLER

P01

ALLOMETRİ ÇALIŞMALARINDA TANJANT KOORDİNATLARI VE TEMEL BİLEŞEN SKORLARININ KULLANILDIĞI MODELLERİN PERFORMANSLARININ İNCELENMESİ

Deniz Sığırlı[1], **İlker Ercan[1]**

[1]*Uludağ Üniversitesi Tıp Fakültesi, Biyoistatistik Anabilim Dalı, Bursa*

iercan@msn.com

Amaç: Bu çalışmada, çok değişkenli regresyon analizine bağımlı değişkenler olarak tanjant koordinatlarının ve tanjant koordinatlarına ilişkin temel bileşen skorlarının alındığı iki farklı modelin performanslarının incelenmesi amaçlanmıştır.

Metod: Büyüme ve allometri çalışmaları hastalıkların ve çevresel faktörlerin organ ya da organizmaların yapısı üzerindeki etkilerinin araştırılması yönünden önemli çalışmalardır. İstatistiksel şekil analizi son dönemde medikal ve biyomedikal çalışmalarda önemli hale gelmiştir. İstatistiksel şekil analizi nesnelerden elde edilen geometrik bilgiyi kullanan yöntemleri içermektedir. Bu çalışmada allometri için iki model incelenmiştir. İlk modelde bağımlı değişken olarak tanjant koordinatları alınırken, ikinci modelde tanjant koordinatlarına ilişkin temel bileşen skorları alınmıştır.

Bulgular: Tanjant koordinatları kullanılarak oluşturulan modelin büyük ve küçük örneklemlerde daha küçük hata kareler ortalaması değerleri verdiği görüldü. Her iki model için de mutlak ortalama sapma ve yanlılık değerlerinin örneklem büyüklüğü arttıkça azaldığı görülmekle birlikte, tanjant koordinatları kullanılan modelde daha fazla azalma olduğu görüldü.

Sonuç: Simülasyon çalışmasının sonuçları, bütün örneklem büyüklüklerinde, bağımlı değişken olarak tanjant koordinatları alınarak oluşturulan modelin, diğer modele göre daha iyi sonuçlar verdiğini göstermiştir.

Anahtar kelimeler: Allometri, İstatistiksel şekil analizi, Procrustes analizi, Tanjant uzayı koordinatları

P02

VETERİNER HEKİMLERİN BİYOİSTATİSTİK BİLGİSİ: ULUSLARARASI WEB ANKET

Mehmet Onur Kaya[1], İlker Ercan[1], Gökhan Ocakoğlu[1], Güven Özkaya[1], Ender Uzabacı[1], Fatma Ezgi Can[1]

[1]*Uludağ Üniversitesi, Tıp Fakültesi, Biyoistatistik AD, Bursa*

monurkaya@gmail.com

Amaç: Uluslararası web tabanlı anket üzerine temellenen çalışmamızın birinci amacı, veteriner hekimlerinin istatistik bilgisi hakkında bilgi elde etmektedir. İkincil amaçlar ise şu sorulara yanıt aramaktadır: Biyoistatistik dersi veteriner hekimliği eğitiminde hangi noktada verilmeli, ve veteriner hekimliği eğitiminde anahtar olabilecek istatistiksel yöntemler nelerdir.

Yöntem: Çalışmamıza 22 ülkeden elektronik posta ile davet edilen 49 veteriner hekim katılmıştır. Veteriner hekimlerine ait veriler web tabanlı anket ile elde edilmiştir. Katılımcılar, veteriner hekimliği dergilerinin taranması sonucundan belirlenmişlerdir.

Bulgular: Sonuçlar veteriner hekimlerinin biyoistatistik eğitimine önem verdiğini göstermekle birlikte katılımcıların büyük bir kısmı biyoistatistik eğitiminin lisans sonrasındaki eğitim düzeyinde alınması gerektiğini belirtmişlerdir.

Sonuç: Veteriner hekimler, biyoistatistik eğitimine önem vermekle birlikte, araştırmalarda biyoistatistik bilgisinin önemli olduğuna inanmaktadır. Biyoistatistik eğitiminde, çalışmanın tasarımı aşamasında önemli bir role sahip olan örnekleme konusuna özellikle önem verilmelidir.

Anahtar Kelimeler: Biyoistatistik, Biyoistatistik Bilgisi, Veteriner Hekimi.

P03

STEREOLOJİ ÇALIŞMALARINDA ARAŞTIRMA TASARIMI VE ÖRNEKLEME YÖNTEMİ

Ayşe Canan Yazıcı[1], Sinan Canan[2], Mustafa Agah Tekindal[1], Ersin Öğüş[1]

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

[2]*Yıldırım Beyazıt Üniversitesi Tıp Fakültesi, Fizyoloji Anabilim Dalı, Ankara*

acyazici@gmail.com

Amaç: Stereoloji, üç boyutlu yapıların iki boyutlu kesitlerinden yararlanılarak orijinal yapının üç boyutlu nicel özelliklerine ilişkin doğru ve etkin hesaplamalar yapılmasına olanak sağlayan yöntemlerin ortak adıdır. Düzgün, simetrik şekle sahip, yahut çevresindeki yapılardan izole edilebilen nesnelerin alanları, hacimleri veya sayıları, basit ölçümler veya matematiksel formüller yardımı ile kolaylıkla hesaplanabilir. Ancak biyolojik yapıların şekillerinin düzgün olmaması ve doğrudan gözlenememeleri gibi nedenlerle, stereolojide ışık ve elektron mikroskobu gibi teknikler ve radyolojik çalışmalardan elde edilen iki boyutlu kesitlerden yararlanılarak bu yapıların hacim hesaplamaları ve morfolojik özelliklerine ilişkin diğer tahminler yapılmaktadır. Üzerinde çalışılan biyolojik yapıya ilişkin yansız ve doğru tahminler yapılabilmesi için alınacak kesitlerin yapıyı doğru temsil edebilmesi dolayısıyla çalışmada değerlendirilecek örneklemin oluşturulması oldukça önemlidir.

Yöntem: Bu çalışmada, biyolojik yapının morfolojik özelliklerine ilişkin yansız ve doğru tahminler yapılabilmesi için parçaların seçimi, kesitlerin örneklenmesi, sayım alanlarının ve alan örneklerinin belirlenmesi gibi, araştırmanın farklı aşamalarında stereolojide kullanılan çeşitli yöntemler incelenerek, güvenilir tahminlerin elde edilebilirliği tartışılmıştır.

Bulgular ve Sonuç: Değerlendirmeye alınacak kesitlerin, söz konusu yapıyı en iyi biçimde temsil edebilmesi için, biyolojik yapının tamamının eşit örneklenme şansına sahip olması istatistik bir zorunluluktur. Optimum sayıda kesit alınarak, kesitlerin yapının her noktasından eşit örneklenmesini sağlaması açısından ve morfolojik yapıdaki niceliksel parametrelerin güvenilir bir biçimde tahmin edilebilmesi açısından sistematik örnekleme yönteminin uygun olduğu görülmüştür. Mevcut stereoloji çalışmalarında da kullanılmakta olan stereolojik metodların temelini bu örnekleme yöntemi oluşturmaktadır.

Anahtar Kelimeler: Stereoloji; Sistematik Örnekleme; Disektör Sayım; Bileşen Hacim Oranı

P04

ÇOKLU UYUM ANALİZİ YÖNTEMİ İLE TIP FAKÜLTESİ ÖĞRENCİLERİNİN MESLEK SEÇİMİNİ ETKİLEYEN BAZI FAKTÖRLERİN İNCELENMESİ: BAŞKENT ÜNİVERSİTESİ ÖRNEĞİ

Ayşe Canan Yazıcı[1], Feride İffet Şahin[2], Erkan Yurtcu[3], Haldun Müderrisoğlu[4], Faik Sarıalioğlu[5]

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

[2]*Başkent Üniversitesi Tıp Fakültesi Tıbbi Genetik Anabilim Dalı, Ankara*

[3]*Başkent Üniversitesi Tıp Fakültesi Tıbbi Biyoloji Anabilim Dalı, Ankara*

[4]*Başkent Üniversitesi Tıp Fakültesi Kardiyoloji Anabilim Dalı ve Tıp Eğitimi Anabilim Dalı, Ankara*

[5]*Başkent Üniversitesi Tıp Fakültesi Pediatri Anabilim Dalı, Ankara*

acyazici@gmail.com

Amaç: Çoklu Uyum analizi (Multiple Correspondence Analysis) yöntemi, çok yönlü tablolarda satırlarda ve sütunlarda yer alan değişken kategorileri arasındaki ilişkileri, benzerlik ve farklılıkları grafiksel gösterimlere dönüştürerek araştıran bir veri analiz yöntemidir. Değişken kategorilerinin birbirleriyle olan ve her bir değişkenin kendi kategorileri arası ilişkileri *daha az boyutlu bir uzayda grafiksel olarak* ortaya konabilmektedir. Bu çalışma, araştırmalardan elde edilen kategorik veya elde edildikten sonra kategorik hale getirilmiş ikiden fazla değişkenin birlikte değerlendirilmesinde Çoklu uyum analizinin yaygın olarak kullanılan klasik istatistik tekniklere göre avantajlarını incelemek amacıyla yapılmıştır. Diğer bir amaç ise uzun ve zorlu bir eğitim süreci gerektiren hekimlik mesleğini seçmeye tıp öğrencilerini yönlendiren faktörleri incelemektir.

Yöntem: Başkent Üniversitesi Tıp Fakültesi Dönem I öğrencilerine 2007-2013 yılları arası, her yıl rutin olarak uygulanan, öğrencileri tanımaya yönelik olarak hazırlanan anketlerin sonuçları Çoklu Uyum Analizi yöntemi ile değerlendirilmiştir. Çalışmada 212 kız, 126 erkek olmak üzere toplam 338 öğrenci yer almıştır.

Bulgular ve Sonuç: Tıp mesleği bir yaşam tarzı, tıp eğitimi ise bu tarzı benimsemekte aşılan uzun ve zorlu bir süreçtir. Hekimlik mesleğini seçen öğrencilerin bu tercihlerinde rol oynayan faktörler nelerdir sorusunun cevabı Çoklu Uyum Analizi yöntemi ile araştırılmıştır. Anket sorularına verilen yanıt değişkenleri arası ilişkiler, kategoriler arası yakınlıklar, benzerlikler Çoklu Uyum Analizinin özellikle *grafiksel gösterimi ile* ayrıntılı bir biçimde ortaya konarak yorumlanmıştır. Öğrencilerimizin Tıp fakültesini tercih etmelerinde rol oynayan faktörler arasında, özellikle ailede hekim yakını olmasının ve hekimlik mesleğinin manevi kazancını önemsemelerinin en etkili faktörler olduğu belirlenmiştir. Ailesinin yönlendirmesi ile bu mesleği tercih edenlerin ise, daha çok mesleğin toplumdaki statüsünü ve maddi kazancını önemsedikleri görülmüştür.

Anahtar Kelimeler: Çoklu Uyum Analizi; Homojenlik Analizi; Özdeğer; Değişim; Tıp Eğitimi

P05

VERİ KÜMELERİNDEN BİLGİ KEŞFİ: VERİ MADENCİLİĞİ

Ersin Öğüş[1], M. Berkay Can[2], Eren Çamur[2], Mine Koru[2], Ömer Özkan[2], Zeynep Rzayeva[2]

[1]*Başkent Üniversitesi Tıp Fakültesi, Biyoistatistik Anabilim Dalı,*

[2]*Başkent Üniversitesi Tıp Fakültesi*

eogus@baskent.edu.tr

Amaç: Dünya üzerinde uydu verileri, tıbbi veriler, alışveriş verileri, otomasyon verileri gibi alanlarda hızla artan veri miktarları bu verilerin toplanması ve saklanması gibi problemleri gündeme getirmiştir. Bu problemlere veri tabanları ve dosya sistemlerindeki gelişmelerle çözüm aranmıştır. Veri tabanlarındaki bu veriler üzerinde analiz yapmak ve karar aşamasında faydalanmak, her hangi bir araç kullanmaksızın olanaksız hale gelmiştir. Bu noktada karşımıza *"Veri Madenciliği" (Data Mining)* bir çözüm olarak çıkmaktadır. Bilişim Teknolojilerinin birçok alanda kullanımı ve uygulamaları her geçen yıl artmaktadır. "Veri Madenciliği"nin de özellikle ticari alanlarda yoğun kullanımından sonra, tıp ve sağlık alanlarında kullanımı da gündeme gelmiştir. Bu çalışmada veri madenciliğinin unsurları, kullanım alanları ve sağlık alanında kullanılması ve yöntemleri incelenmiştir.

Yöntem: Veri madenciliğinin tarihçesinden başlanarak, unsurları, kullanım alanları sağlık alanında kullanılması, hukuksal yönleri incelenmiş, kullanılan teknik ve yöntemler özetlenmiştir. Derleme türünde planlanan bu çalışmada Dünya'da ve Türkiye'de veri madenciliğinin durumu üzerinde de durulmuştur.

Bulgular ve Sonuç: Veri madenciliği teknikleri üzerinde 1950' li yıllarda çalışılmaya başlanmış, 1970' li yıllarda sağlık alanında kullanılması yaygınlaşmıştır. Tıp alanında bulunan mevcut veri oldukça fazla ve hayati öneme sahiptir. Hastane bilgi sistemleri sayesinde bu veriler düzenli olarak tutulmaktadır. Hayati öneme sahip olan bu verilerden daha fazla yararlanmak mümkündür. Hastane Bilgi sistemlerinden veya diğer tıbbi veri toplayan sistemlerden alınan veriler üzerinde yapılan veri madenciliği çalışmaları hem uzmanlar için hem hastane yönetimi için hem de hastaların daha kaliteli bir hizmet almalarında etkin rol alabilir.

Madenciliği yapılacak olan verinin de bazı vasıflara sahip olması gerekmektedir. Bu vasıflar veri ambarı (Data Warehouse) ile sağlanmaktadır. Veri ambarları basit olarak veri madenciliği işleminin yapılacağı verilerin oluşturulduğu özel veri tabanlarıdır. Veri ambarlarının oluşturulması işlemi verinin çeşitli kaynaklardan toplanarak, veriler içerisindeki uyuşmazlıklar ve hatalardan arındırılmasından ibarettir.

Veri madenciliğinde esas olarak bazı istatistik yöntemler kullanılmaktadır, bunların yanı sıra bellek tabanlı yöntemler, yapay sinir ağları ve karar ağaçları yöntemleri kullanılmaktadır. Dünyada gelişmiş

ülkelerde kaliteli veri birikimi yeterli olduğundan, veri madenciliğinden çok fazla yararlanılmaktadır. Türkiye'de ise oldukça yaygın olarak kullanılmaktadır, ancak Sağlık Bakanlığı bünyesinde sağlık kurumlarından standart veri toplama süreci Ocak 2009 yılında başlamıştır, verilerden anlamlı bilgiler keşfedilmesi için en az 5 yıllık verinin bulunması gerekmekte, dolayısıyla ilk sonuçlar 2014 yılında alınabilecektir.

"Veri Madenciliği" tıbbi kullanımı ile daha önce belki de birçok klinik araştırma gerektiren, hem ekonomik hem de insan (veya deney hayvanları) sağlığı açısından sakıncaları olan tıbbi araştırmaların yerini kısmen de olsa doldurarak tıbbi araştırmalar için yeni bir ufuk sağlayacaktır.

Anahtar kelimeler: Veri madenciliği, veri tabanı, veri ambarı, veri madenciliği ve tıp.

P06

TIP BİLİMLERİNDE YAYINLANAN MAKALELERİN İSTATİSTİKSEL HATALAR BAKIMINDAN İNCELENMESİ

İlker Ercan[1], Pınar Günel Karadeniz[1,2], Şengül Cangür[3], Güven Özkaya[1], Hakan Demirtaş[4]

[1]*Uludağ Üniversitesi, Tıp Fakültesi, Biyoistatistik AD, Bursa*

[2]*Marmara Üniversitesi, Tıp Fakültesi, Biyoistatistik ve Tıbbi Bilişim AD, İstanbul*

[3]*Düzce Üniversitesi, Tıp Fakültesi, Biyoistatistik AD, Düzce*

[4]*Illinois University, School of Public Health, Epidemiology ve Biostatistics Division, Chicago, USA.*

iercan@msn.com

Amaç: Bu çalışmanın amacı, bilimsel makalelerdeki istatistiksel hataları; (i) benzer çalışmalar içerisindeki hataların dağılımı bakımından, (ii) SCI veya SCI-Expanded kapsamındaki dergilerde ve SCI veya SCI-Expanded kapsamında olmayan dergilerde yayınlanan makalelerdeki hata oranları bakımından incelemektir.

Yöntem: Çalışmaya, SCI ya da SCI-Expanded kapsamındaki dergilerde yayınlanan 95 makale ile, SCI ya da SCI-Expanded kapsamında bulunmayan dergilerde yayınlanan 122 makale dahil edilmiştir. Makaleler, PubMed ve Bioline veri tabanlarında yer alan, 2004-2010 yılları arasında yayınlanmış makaleler içerisinden seçilmiştir.

Bulgular: Toplam 217 makaleden her birinde en az bir istatistiksel hata bulunmuştur. Hem SCI veya SCI-Expanded kapsamındaki hem de SCI veya SCI-Expanded kapsamında bulunmayan dergilerde yayınlanan makalelerde en sık görülen istatistiksel hatalar "verilerin özetlenmesindeki hatalar"dır. SCI veya SCI Expanded kapsamındaki dergilerde yayınlanan makalelerle, bu kapsamda bulunmayan dergilerde yayınlanan makaleler arasında, "yanlış test kullanımı" ve "istatistiksel sembol hataları" bakımından istatistiksel olarak anlamlı bir fark bulunmuştur.

Sonuç: Bilimsel yayınlardaki istatistiksel hataların ortaya çıkmasını önlemek için araştırmacılar ve editörlerin yerine getirmesi gereken yükümlülükler vardır. Araştırmacılar (i) temel istatistik bilgisine sahip olmalıdır, (ii) çalışmanın planlama, analiz, yorumlama ve raporlama aşamalarında bir biyoistatistikçiye danışmalıdır. Ayrıca, editörler dergilerine gönderilen çalışmaları, değerlendirme aşamasında bir biyoistatistikçiye göndermelidir.

Anahtar Kelimeler: İstatistik, Biyoistatistik, Tıbbi Makaleler

P07

META ANALİZİNİN VETERİNER HEKİMLİKTE UYGULANMASI

Ender Uzabacı[1], Bülent Ediz[2]

[1]*Uludağ Üniversitesi Veteriner Fakültesi, Biyoistatistik AD, Bursa*

[2]*Uludağ Üniversitesi Tıp Fakültesi, Biyoistatistik AD, Bursa*

euzabaci@gmail.com

Amaç: Broyler (etlik piliç) yetiştiriciliğinde uygulanan sürekli ve kesintili aydınlatma programlarının canlı ağırlık üzerindeki etkisinin incelenmesi amacı ile yapılmış araştırmaların sonuçlarını meta analizi ile birleştirerek tutarlı tahminler elde etmeye çalışmak ve ileride bu konuda yapılacak çalışmalara ışık tutmaktır.

Yöntem:1997-2008 yılları arasında sürekli aydınlatmaya karşı kesintili aydınlatmanın kullanılmasının Broylerlerde canlı ağırlığa etkisini araştırmak için yapılmış çalışmaların sonuçları meta analizi yöntemi ile birleştirilmiştir. Çalışmalara ait canlı ağırlık değerlerinin ortalama farkları, bu değerlerin standart hataları ve etki büyüklüğü olarak kullanılacak olan standartlaştırılmış ortalama fark değerleri bulunmuştur. Analize toplam 5893 birim içeren 11 çalışma katılmış ve elde edilen etki büyüklükleri kesintili ve sürekli aydınlatma tipleri arasında karşılaştırılmıştır.

Bulgular: Çalışmalarda canlı ağırlık ölçümünün yapıldığı 1, 7, 14, 21, 28, 35 ve 42. günlerdeki sonuçların hepsinde kesintili ve sürekli aydınlatma programlarının canlı ağırlık bakımından anlamlı farklılığa yol açtığı bulunmuştur ($p<0.001$).

Sonuç: Broylerlerde canlı ağırlık bakımından kesintili aydınlatmanın sürekli aydınlatmaya göre daha olumlu ve anlamlı bir etkisi olduğu tespit edilmiştir.

Anahtar sözcükler: meta analizi, etki büyüklüğü, kesintili aydınlatma, Broyler.

P08

BOX-BEHNKEN DENEME DÜZENİ İLE ENGELLİ ÇOCUĞA SAHİP ANNELERİN RUHSAL ALAN YAŞAM KALİTESİNİN DEĞERLENDİRİLMESİ

Mustafa Agah Tekindal[1], Melike Tunç[2], Ayşe Canan Yazıcı[1], Serdal Kenan Köse[3], Özlem Güllü[4], Can Ateş[3], Beyza Doğanay Erdoğan[3]

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

[2] *Hacettepe Üniversitesi, İktisadi ve İdari Bilimler Fakültesi, Sosyal Hizmet Anabilim Dalı, Ankara*

[3] *Ankara Üniversitesi, Tıp Fakültesi, Biyoistatistik Anabilim Dalı, Ankara*

[4] *Ankara Üniversitesi Veteriner Fakültesi, Biyoistatistik Anabilim Dalı, Ankara*

matekindal@gmail.com

Amaç: Box-Behnken (BDD) deneme düzeni faktöriyel düzende sabit etkili deneme düzenleri içinde oldukça güvenilir sonuçlar veren bir yöntemdir. Çalışmamızın amacı, Yanıt Yüzeyi Metodlarından BDD yöntemini kantitatif olarak uygulamak ve 3 yönlü faktöriyel varyans analizi ile karşılaştırmaktır.

Yöntem: Çalışmamızda Dünya Sağlık Örgütü tarafından geliştirilen, geçerlik ve güvenilirliği Eser ve arkadaşları tarafından yapılmış olan, Türkiye için ulusal soru eklenen, kısa form WHOQOL-BREF-TR ölçeği kullanılmıştır. Ölçekten hesaplanan ruhsal alan toplam puan ortalamaları hesaplanmıştır. Bu kapsamda, zihinsel engelli çocuğa sahip annelerin ruhsal alan yaşam kalitesindeki optimal değişken kombinasyonunun belirlenmesi amaçlanmıştır. Anneler arasındaki ruhsal alan toplam puan varyansını minimuma indirmek için her kategoride, daha önceden belirlenen birden çok anneye ulaşılmış ve ortalaması alınmıştır. Toplamda (Ocak 2010-Mart 2010) 540 zihinsel engelli çocuğa sahip anneye ulaşılmıştır. Yanıt yüzeyi metodunun 3^3 BDD deneme düzeni kullanılarak, dengeli tamamlanmamış blok tasarımı düzeninde, çalışma verileri toplanmadan önce model belirlenmiştir. Elimizdeki sabit modele göre veriler toplanmıştır. Ayrıca verilere 3 yönlü faktöriyel varyans analizi uygulanmıştır. BDD deneme düzeni ile karşılaştırılarak değerlendirilmiştir.

Bulgular ve Sonuçlar: Veriler ikinci derece yanıt yüzeyi modeli için hesaplandığında R^2 = %90.33 olduğu görülmüş, modelin bağımlı değişkendeki değişimin büyük bir bölümünü açıkladığı saptanmıştır. Optimum olarak değişken kombinasyonları belirlenmiştir. Ayrıca, aynı verilere 3 yönlü varyans analizi uygulanmış, hem ana etkiler hemde etkileşim etkileri anlamsız çıkmıştır. Sabit olarak belirlenen faktöriyel denemelerde, bağımsız değişkenler arasındaki farklılıklar ve optimum kombinasyonlar araştırılmak istendiğinde, dengeli tamamlanmamış blok tasarımını temel alan yanıt yüzeyi metodunun BDD düzeni, yaygın olarak kullanılan faktöriyel denemelerdeki varyans analizine göre daha belirleyici ve açıklayıcı sonuçlar vermiştir. Özellikle ölçme değerlendirme alanında yaygın olarak kullanılan, faktöriyel denemelerdeki varyans analizinin yerine yanıt yüzeyi metodu ciddi bir alternatif oluşturabilir.

Anahtar Kelimeler: Yanıt Yüzeyi Metodu, Box-Behnken Deneme Düzeni (BDD), WHOQOL-BREF-TR-Ruhsal Alan

P09

YANIT DEĞİŞKENİNİN İKİLİ OLDUĞU KÜMELENMİŞ VERİLERDE KULLANILABİLECEK İSTATİSTİKSEL YÖNTEMLER

Sevilay Karahan[1], Harika Gözükara Bağ[2], Osman Saraçbaşı[1]

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

[2]*İnönü Üniversitesi, Tıp Fakültesi, Biyoistatistik ve Tıp Bilişimi Anabilim Dalı, Malatya*

sevilaykarahan@gmail.com

Amaç: Kümelenmiş veriler sağlık alanında sıkça karşılaşılan bir veri yapısıdır. Aynı hastanın farklı organlarından (göz, kulak, diş vb.) elde edilmiş veriler buna bir örnektir. Yanıt değişkenine etki eden faktörler değerlendirilirken, küme içi değişimler de göz önünde bulundurulmalıdır. Bu çalışmada kümelenmiş veriler için geliştirilen iki yöntemin örnek bir veri kümesinde uygulamasının gösterilmesi amaçlanmaktadır.

Yöntem: Literatürde küme içi değişimleri göz önünde bulunduran farklı yöntemler bulunmaktadır. Bunlardan sıklıkla kullanılan ikisi genelleştirilmiş eşitlik kestirimleri (GEK) ve alternatif lojistik regresyon (ALR)'dir. GEK; bağımlılık yapısını korelasyon katsayısı veya odds oranını kullanarak çözümler. GEK ile elde edilen regresyon katsayıları kestirimleri tutarlıdır, asimptotik olarak normal dağılım gösterir. Ancak küme genişliklerinin büyük olması durumunda varyans kovaryans matrisinin boyutu büyüyeceğinden hesaplamalar zorlaşmaktadır. ALR yöntemi bu sorunun üstesinden gelmektedir. ALR verilen α katsayısı için β katsayılarını, ardından elde edilen β katsayısı için α katsayısını iterasyonla hesaplama mantığına dayalıdır. Bu çalışmada bu iki yöntem tanıtılacak ve sağlık alanından bir veriye bu yöntemler uygulanacaktır.

Bulgular: Bu iki yöntemin sonuçlarına ilişkin uygun istatistikler gösterilecektir.

Sonuç: Bu iki yöntemin istatistikleri karşılaştırılacak ve gelecek uygulamalar için önerilerde bulunulacaktır.

Anahtar kelimeler: Kümelenmiş veri, genelleştirilmiş eşitlik kestirimleri, alternatif lojistik regresyon

P10

ÇARPIK DAĞILIMLARDA ROC EĞRİSİ ALTINDA KALAN ALAN TAHMİNİ

İlker Ünal[1], Refik Burgut[2], Yaşar Sertdemir[2]

[1]*İzmir Üniversitesi Tıp Fakültesi Biyoistatistik ve Tıp Bilişimi AD, İzmir, Türkiye*

[2]*Çukurova Üniversitesi Tıp Fakültesi Biyoistatistik ve Tıp Bilişimi AD, Adana, Türkiye*

ilker.unal@izmir.edu.tr

ROC eğrisi altında kalan alan tahmini parametrik ve parametrik olmayan yöntemler kullanılarak yapılabilmektedir. Bu yöntemlerin tercihinde incelenen sayısal ölçümün hasta ve sağlıklı grubundaki dağılım bilgisi ön plandadır. Parametrik yaklaşımlarda incelenen sayısal ölçüm hasta ve sağlıklı gruplarında normal dağılım gösterdiği veya transformasyon ile dağılımın normal dağılıma dönüştürülebileceği varsayılmaktadır. Parametrik olmayan yaklaşımlarda ise eğri altında kalan alan geometrik olarak hesaplanmaktadır. Bu hesaplamada eğri altında kalan alan yamuklara bölünebildiği gibi, eğrinin çekirdek düzgünleştirme (kernel smoothing) yöntemleri ile elde edilen olasılık yoğunluk fonksiyonu da kullanılabilmektedir. Bu çalışmada incelenen sayısal ölçümün çarpık bir dağılım gösteriyor olması halinde parametrik olmayan yöntemlerin tercih edilmesi yerine incelenen ölçümün transformasyonlar sonucunda normal dağılıma yaklaştırılması ve ardından alan tahmini için parametrik yaklaşımların kullanılmasının avantaj ve dezavantajları araştırılacaktır. Bu amaçla, çalışmada normal dağılım göstermeyen çarpık dağılımlarda ROC eğrisi altında kalan alan tahmininde kullanılan yöntemler üretilmiş ve gerçek veri seti kullanılarak karşılaştırılacaktır.

Anahtar Kelimeler: ROC eğrisi, Eğri altında kalan alan, Çarpık dağılımlar

P11

HİPERLİPİDEMİ HASTALARINDA KOLESTEROL SEVİYESİNİN YAŞA GÖRE DEĞİŞİMİNİN DEĞİŞİK REGRESYON MODELLERİYLE İNCELENMESİ

Emre Dirican[1], Cemil Çolak[1], **Saim Yoloğlu[1]**

[1]*İnönü Üniversitesi Tıp Fakültesi Biyoistatistik ve Tıp Bilişimi AD Malatya, Türkiye*

saim.yologlu@inonu.edu.tr

Amaç: Bu araştırmada, İnönü Üniversitesi Turgut Özal Tıp Merkezi Kardiyoloji polikliniğine müracaat eden hiperlipidemi hastalarında Kolesterol seviyesinin yaşa göre değişiminin değişik regresyon modelleriyle tahmin edilmesi amaçlanmıştır.

Yöntem: Doğrusal regresyon modeli ile doğrusal olmayan regresyon modellerinden Gompertz ve Lojistik büyüme eğrisi modelleri yardımıyla analizler yapılmıştır.

Bulgular: Tahminlenen modellerin uyumu, Hata Kareler Ortalaması, Bilgi kriterleri ve açıklayıcılık katsayısı değerleri kullanılarak yapılmıştır. Modeller uyum kriterlerine göre değerlendirildiğinde, erkek ve kadın hiperlipidemi hastaları için en iyi uyuma sahip modeller sırasıyla; Gompert, Lojistik ve doğrusal regresyon modelleri olarak elde edilmiştir.

Sonuç: İncelenen doğrusal ve doğrusal olmayan modellerin kolesterol seviyesinin tahmininde başarılı oldukları belirlenmiştir.

Anahtar Kelimeler: Bilgikriterleri, Doğrusal regresyon modeli, Gompertzmodeli, Lojistik modeli.

P12

GENETİK EPİDEMİYOLOJİDE İSTATİSTİKSEL YAKLAŞIMLAR: HEMATOLOJİK KANSERLERDE BİR UYGULAMA

Selen Begüm Uzun[1], Yusuf Kemal Arslan[1], Pınar Özaltun[1], Refik Burgut[1]

[1]*Çukurova Üniversitesi Tıp Fakültesi, Biyoistatistik ve Tıbbi Bilişim ABD, Adana*

selenbegumuzun@gmail.com

Genetik ve çevrenin, hastalıkların sürecine katkısının bilinmesi kanser türlerinin, kronik hastalıkların, alerjilerin ve psikiyatrik bozuklukların etiyolojisini daha karmaşık hale getirmiştir. Genetik epidemiyoloji, hastalık etiyolojisinde genetik faktörlerin oynadığı rolü araştırmada klasik epidemiyolojik tasarım ve yöntemleri kullanan yeni bir çalışma alanıdır. Genetik epidemiyolojik yaklaşımlarla aşağıdaki sorulara yanıtlar aranır. Bunlar:

1. Ailede bir kümelenme var mı ve bu durum gen ya da çevre etkisi ile açıklanabilir mi?
2. Bu spesifik gen çevresel etki ile birleştiğinde hastalığın riski nasıl kontrol edilir?
3. Yatkın genlerin yerini belirleme şansını arttıracak genetik olarak homojen alt gruplara sahip hastaları nasıl belirleyebiliriz?
4. Ailesel korelasyon büyüklüğü ve modeli en iyi şekilde nasıl ölçülür ?

Bu çalışmada özellikle genetik epidemiyoloji sürecinin üç başlığı üzerinde durulacaktır: Bağlantı analizi, ilişki analizi ve ailesel kümelenme. Bu başlıklar, tanımları ve örnekleri ile açıklanacak, amaca uygun olarak biyoistatistiksel yöntemler ve uygun modeller "Hematolojik kanserlerde ailesel kümelenme var mı?" sorusuna yanıtta kullanılacaktır.

Anahtar Sözcükler: Genetik epidemiyoloji, Ailesel kümelenme, Bağlantı analizi, İlişki analizi, Hematolojik kanser

P13

TABANUS BROMIUS LINNAEUS SİNEK TÜRÜNÜN KANAT YAPISININ İSTATİSTİKSEL ŞEKİL ANALİZİ İLE İNCELENMESİ

Gökhan Ocakoğlu[1], Yakup Afacan[2], İlker Ercan[1], Ferhat Altunsoy[2]

[1]*Uludağ Üniversitesi Tıp Fakültesi Biyoistatistik Anabilim Dalı, Bursa*

[2]*Anadolu Üniversitesi Fen Fakültesi Biyoloji Bölümü, Eskişehir*

ocakoglu@uludag.edu.tr

Amaç: Dünyada, Avrupa'da ve Türkiye'de yaygın olarak görülen *Tabanus bromius Linnaeus* sinek türünün kanat yapısının habitat farklılıklarına göre biyolojik şeklinin ve bazı morfolojik karakteristiklerinin istatistiksel şekil analizi ile incelenmesi amaçlanmıştır.

Yöntem: Bu amaçla iklimsel özelliklere göre farklılık gösteren 4 bölgeden(Ege, İç Anadolu, Karadeniz, Marmara) toplanan sinek örneklerine ait 211adet sinek kanadı üzerinde çalışma yapılmıştır. Sinek kanatlarının sınıflandırılması için kümeleme analizi kullanılmıştır. Bu amaçla Procrustes uzaklığı kullanılarak kanatlar arasındaki şekilsel benzerlik hesaplanmıştır. Ortalama şekillere göre kümeler arasındaki farklılıkları incelemek için Genelleştirilmiş Procrustes Analizi (GPA) uygulanmıştır. Şekil deformasyonları Thin Plate Spline (TPS) analizi kullanılarak incelenmiştir.

Bulgular: Çalışmada, benzerliklerine göre sinek örnekleri iki kümeye ayrılmış olup iki küme 0,046 benzemezlik düzeyinde tanımlanmıştır. Sinek kanatlarının %90,05'i kümelere dahil edilmiş geriye kalanlar ise uç değer olarak değerlendirilmiştir. İki kümeden bir tanesi toplam deneklerin %19,91(42/211)' ini, diğer küme ise %70,14(148/211)' ini kapsamaktaydı. Kümeler arasında ortalama şekiller bakımından anlamlı farklılık bulunmuştur(p=0,010). İki küme yükseklik, yağış ve rüzgâr düzeyleri bakımından farklılık gösterirken(sırasıyla p=0,012, p=0,007 ve p=0,031), nem düzeyi bakımından kümeler arasında farklılık bulunmamıştır(p=0,656).

Sonuç: Çalışmamızda iklimsel özelliklere göre farklılık gösteren 4 bölgeden (Ege, İç Anadolu, Karadeniz, Marmara) toplanan örneklerin analizi sonucunda iki grupta değerlendirmesi yapıldığında yükseklik, yağış ve rüzgar etkisinin sineklerin kanat şekilleri üzerinde farklılık oluşturduğu belirlenmiştir. Sineklerin kanat şekillerinde nemin ise bir farklılık oluşturmadığı görülmektedir.

Anahtar Kelimeler: Tabanus bromius Linnaeus, istatistiksel şekil analizi, Procrustes uzaklığı.

P14

TİP-2 DİYABETLİ HASTALARDA ALBUMİNURİ DÜZEYİNİ TAHMİN ETMEDE SINIFLANDIRMA YÖNTEMLERİNİN PERFORMANSLARININ KARŞILAŞTIRILMASI

İmran Kurt Ömürlü[1], Mevlüt Türe[1], Mustafa Ünübol[2], Merve Katranci[1], Engin Güney[2]

[1]*Adnan Menderes Üniversitesi Tıp Fakültesi Biyoistatistik ve Tıbbi Bilişim Anabilim Dalı, Aydın*

[2]*Adnan Menderes Üniversitesi Tıp Fakültesi Endokrinoloji ve Metabolizma Hastalıkları Bilim Dalı, Aydın*

ikurtomurlu@gmail.com

Amaç: Bağımlı değişkenin kategorik yapıda olduğu veri setlerinin analizi için birçok yöntem kullanılmaktadır. Bu yöntemler, bağımsız değişkenlerin bağımlı değişken üzerindeki etkilerini değerlendirmek ve birimlerin bağımlı değişkenin kategorilerine göre en az hatayla sınıflandırılması için model oluşturmada araştırmacılara her zaman kılavuz olmaktadır. Bu çalışmada; lojistik regresyon (LR), yapay sinir ağları (YSA), sınıflandırma ve regresyon ağacı (C&RT), esnek ayırma analizi (EAA), EAA/MARS (derece=1) ve EAA/MARS (derece=2) yöntemlerinin tip-2 diyabetli hastalarda albüminuri düzeyini tahmin etmede performanslarının karşılaştırılması amaçlandı.

Yöntem: Tip-2 diyabetli hastalarda albuminuri düzeyi; idrardaki albumin atılımı miktarına göre 0-30 mg/24 saat için nonalbuminuri, 30-300 mg/24 saat için mikro ve ≥300 mg/24 saat için makroalbuminuri olarak sınıflandırılmaktadır. Çalışmamızda albuminuri veri seti, nonalbuminuri (n=190) ve mikro+makro albuminuri (n=76) olmak üzere iki gruba ayrıldı ve toplam 266 birim cinsiyet, kilo (kg), idrar kreatinin miktarı (mg/dl), serum albumin miktarı (mg/dl), yaş, kandaki üre miktarı (mg/dl), kreatinin klirensi (mL/dk), açlık kan şekeri (mg/dl), tokluk kan şekeri (mg/dl) ve HbA1c (mg/dl) bakımından incelendi. Bu bağımsız değişkenlerin albuminuri miktarını tahmin etmek için 10-katlı çapraz geçerlilik yöntemi kullanıldı ve yanlış sınıflandırma oranına dayanarak yöntemlerin performansları karşılaştırıldı.

Bulgular ve Sonuç: Sonuç olarak, tip-2 diyabetli hastalarda albuminuri düzeyini en az hatayla tahmin etmede EAA/MARS (derece=1) yönteminin performansının diğer yöntemlere göre daha iyi olduğu belirlendi.

Anahtar Kelimeler: C&RT, Esnek Ayırma Analizi, Yapay Sinir Ağları, Lojistik Regresyon, Albuminuri

P15

DİŞ HEKİMLERİNİN BİYOİSTATİSTİK BİLGİSİ: ULUSLARARASI WEB ANKET

Gökhan Ocakoğlu[1], İlker Ercan[1], Pınar Günel Karadeniz[2]

[1]*Uludağ Üniversitesi, Tıp Fakültesi, Biyoistatistik AD, Bursa*

[2]*Marmara Üniversitesi, Tıp Fakültesi, Biyoistatistik AD, İstanbul*

ocakoglu@uludag.edu.tr

Amaç: Bilimsel önemliliğe sahip verilerin ve güvenilir sonuçların elde edilebilmesi için istatistiğe araştırmanın her aşamasında ihtiyaç duyulmaktadır. Bu nedenle sağlık bilimlerinde çalışan araştırmacıların biyoistatistik hakkında bilgilendirilmesi gerekmektedir. Uluslararası web tabanlı anket üzerine temellenen çalışmamızın birinci amacı, diş hekimlerinin istatistik bilgisi(I-a), ve bu bilginin uzmanlık alanlarına(klinik/temel) göre nasıl değiştiği (I-b) hakkında bilgi elde etmektedir. İkincil amaçlar ise şu sorulara yanıt aramaktadır: Biyoistatistik dersi diş hekimliği eğitiminde hangi noktada verilmeli (II-a) ve diş hekimliği eğitiminde anahtar olabilecek istatistiksel yöntemler nelerdir (II-b).

Yöntem: Çalışmamıza 46 ülkeden elektronik posta ile davet edilen 252 diş hekimi katılmıştır. Diş hekimlerine ait veriler web tabanlı anket ile elde edilmiştir. Katılımcılar, diş hekimliği dergilerinin taranması sonucundan belirlenmişlerdir.

Bulgular: Sonuçlar diş hekimlerinin biyoistatistik eğitimine önem verdiğini göstermekle birlikte katılımcıların büyük bir kısmı biyoistatistik eğitiminin lisans sonrasındaki eğitim düzeyinde alınması gerektiğini belirtmişlerdir. Bununla birlikte katılımcıların çok azı bu eğitimin lisans düzeyinde verilmesi gerektiğini belirtmişlerdir.

Sonuç: Biyoistatistik eğitiminde, çalışmanın tasarımı aşamasında önemli bir role sahip olan örnekleme konusuna özellikle önem verilmelidir. Üzerinde önemle durulması gereken diğer önemli bir konuda parametrik olmayan testlerdir. Bununla birlikte biyoistatistik eğitiminde, araştırmacıların biyoistatistiksel danışmanlığı çalışmalarının planlama aşamasında almaları gerektiği de vurgulanmalıdır.

Anahtar Kelimeler: Biyoistatistik, Biyoistatistik bilgisi, Diş hekimi.

P16

ROC EĞRİSİ ÇÖZÜMLEMESİNDEKİ YENİ YAKLAŞIMLARIN ÖRNEKLER ÜZERİNDE İNCELENMESİ

Sevda Özel[1], Başak Gürtekin[1], Kamber Kaşali[1], Ahmet Dirican[1], Rian Dişçi[1]

[1]*İ.Ü. İstanbul Tıp Fakültesi, Biyoistatistik ve Tıp Bilişimi Anabilim Dalı, İstanbul*

sevda@istanbul.edu.tr

Amaç: Çalışmamızda, geleneksel ROC eğrisi çözümlemesinde özellikle görüntüleme sistemlerinde, lezyonların varlığı ya da yokluğu, birden çok lezyonun tanımlanması söz konusu olduğunda bunların yerleşim yerlerinin de belirlenmesi için geliştirilen; Lokalize ROC (Localization ROC- LROC), Serbest-yanıt ROC (Free-Response ROC-FROC) ve Alternatif FROC (AFROC) analizi yaklaşımlarının örneklerle açıklanması amaçlanmıştır.

Yöntem: Görüntüleme sistemlerinde, tek bir lezyonun varlığı veya yokluğunu belirlemede ve yine lezyon varsa işaretlemenin doğruluğunu belirlemek amacı ile LROC yaklaşımı, bir görüntülemede birden çok lezyonun olması durumunda ise FROC ve AFROC yaklaşımları kullanılmıştır.

Bulgular ve Sonuç: ROC analizi, klinik karar vermenin geliştirilmesine yardım eden ve tanı performansını en iyi açıklayan güçlü bir yaklaşımdır. Bu yüzden, ROC eğrisi yöntemi tanı değerlendirmelerinde sıklıkla tercih edilir. Ancak, özellikle görüntüleme sistemleri için hastalık tanısı konmasında olası bazı ihmallerin sonuçların yorumunda önemli farklar yaratabileceği ve bu farklılıkların geliştirilen LROC, FROC ve AFROC yaklaşımları ile çözümlenebileceği saptanmıştır. Yine bu yaklaşımların geleneksel ROC eğrisi yönteminden istatistiksel yönden daha güçlü olduğu yönünde bulgular mevcuttur.

Anahtar Kelimeler: ROC eğrisi, LROC, FROC, AFROC, Lezyon

P17

SEPSİS GELİŞEN YENİDOĞAN BEBEKLERDE VİTAL BULGULAR KULLANILARAK OLUŞTURULAN LOJİSTİK REGRESYON MODELİ İLE DAHA ERKEN TEŞHİS MÜMKÜN MÜ?

Yaşar Sertdemir[1], Hacer Yapıcıoğlu[2], Ferda Özlü[2]

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

[2]*Çukurova Üniversitesi Tıp Fakültesi Neonatoloji Bilim Dalı, Adana*

yasarser@cu.edu.tr

Giriş: Yenidoğan Yoğun Bakım Ünitelerinde rastlanılan en ağır sorunlardan biri olan sepsis yenidoğan ölümlerin de önemli bir kısmından sorumludur. Sepsisin en önemli nedeni bakteriyemidir. Sepsis tanısı çoğu zaman enfeksiyonun ileri evrelerinde konulabilmekte; bağışıklık sistemi deprese olan yenidoğanlarda olay çok hızlı ilerleyebilmekte ve ölümcül olabilmektedir. Ayrıca bakteriyeminin teşhisinde altın standart kan kültürü olup, kan kültürünün sonucunun alınması en az 24-72 saat sürmekte ve yenidoğan bebeklerde yanlış negatif oranı da %30'a dek yükselmektedir. Yenidoğan yoğun bakım birimlerinde çalışan doktorlara daha erken müdahale etme imkanı sağlayan pratik ve invazif olmayan bir yol bulunabilir ise tedavide başarı oranı önemli ölçüde artacaktır.

Amaç: Yenidoğanların bağlı bulundukları cihazla kayıt edilen nabız, solunum sayısı, vücut ısısı ve kan basıncı gibi vital bulgular kullanılarak sepsis olan yenidoğanların daha erken fark edilmesini sağlayacak bir tahmin modeli oluşturmak.

Yöntem: Araştırma bir hemşirenin 3 bebekten sorumlu olduğu, 18 adet seviye-III yoğun bakım yatağı bulunan 42 yataklı Yenidoğan Yoğun Bakım biriminde yapılmıştır. Çalışmaya, geldikleri günden itibaren takip edilen 49 hasta (sepsis) ve her hasta için yaş ve cinsiyete göre eşleştirilmiş 2 kontrol alınmıştır. Geldiğinde hasta olan veya ilk 5 günde hastalığı ortaya çıkan yenidoğanlar araştırma dışında bırakılmıştır. Tanı öncesi 5. 4. 3. 2. ve 1. gün vital bulgular lojistik regresyon uygulanarak analiz edilmiştir. Regresyon modeli elde edilirken veri setinin ilk yarısı analiz, diğer yarısı ise geçerliliğinin test edilmesi için kullanılmıştır.

Bulgular: Tablo 1 Vital bulgular ile elde edilen regresyon modellerinin tanı öncesi güne göre performans değerleri

Tanı öncesi Gün	Sensitivite	spesifite	Genel	Nagelkerke R Square
5	54.5	86.0	75.4	0.388
4	66.0	87.1	80.0	0.502
3	81.6	94.9	90.5	0.721

2	91.8	96.9	95.2	0.912
1	79.6	91.8	87.8	0.627

Sonuç: Vital bulgular yardımıyla tanıyı 3 gün daha erken koyarak sepsis olan yenidoğanların %80'den fazlasının tedavi başarısını artırmak mümkündür. Ancak, bu modellerin uygulanmasından doğabilecek olumsuzluk sağlıklı yenidoğanların %5'ine gereksiz antibiyotik verilmesi olabilir.

Anahtar kelime: Lojistik regresyon, Sepsis, Vital bulgu, Erken tanı.

P18

PUBMED KLİNİK DENEME META ANALİZLERİNDE YAYIN YANLILIĞI DEĞERLENDİRİLME YÖNTEMLERİ

Duygu Sıddıkoğlu[1], **Z. Nazan Alparslan**[1]

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

nazan@cu.edu.tr

Aynı konuda yapılmış pek çok araştırmanın sonuçları arasında ortaya çıkan uyumsuzlukların kaynaklarının incelenmesini, bulgular arasındaki ilişkinin incelenmesini ve konu hakkındaki tüm kanıtın toparlanıp tıbbi bilginin kazanılmasını standart protokollerle sağlayan meta analizi; aynı konuda farklı yer, zaman ve merkezlerde yapılmış olan araştırma sonuçlarını niteliksel ve niceliksel olarak birleştirmeye yardımcı olan istatistiksel bir yöntemdir. Ancak yöntemle elde edilen sonuçların güvenirliği yapılan analizlerden öte, derlenen yayınlarla ilgili birçok faktöre de bağlıdır. Genel olarak yayın yanlılığı diye adlandırılan ve sadece yayınlanmış makalelerin kullanılması ile ortaya çıkan bu durum, meta analizinin geçerliliğini farklı nedenlerle tehdit edebilir. Örneğin, sonuçları istatiksel olarak anlamlı olan çalışmaların yayınlanma olasılığı, sonuçları istatistiksel olarak anlamlı bulunmamış olan çalışmalara kıyasla çok daha yüksektir. Ek olarak, anlamlı sonuç veren çalışmalar kendilerine benzeyenlerin sayıca artmasına sebep olurlar. Ya da, ana dili İngilizce olmayan araştırıcılar özellikle de istatistiksel anlamlılığı olmayan çalışmaları yerel dergilere yazabilirler. Bu uygulamaların hepsi yüksek etki büyüklüğüne neden olur.

Yayın yanlılığını anlamanın çeşitli istatistiksel yöntemleri vardır(funnel plot, forest plot, Egger regresyon testi gibi). Yayın yanlılığı değerlendirmesi bir meta analizinin vazgeçilmez bir bölümüdür ve meta analizi raporlaması ile ilgili raporlama formatlarında(QUAROM, PRISMA) da çeşitli alt başlıklarla talep edilir.

Bu çalışma, tıp çalışanlarının kolaylıkla ulaştığı ve sıklıkla başvurduğu, PubMed dizininin meta analizlerindeki yayın yanlılığı değerlendirmelerini gözden geçirmek üzere planlanmıştır. MEDLINE'nın da web arayüzü olan PubMed üzerinden tam metne bağlantı verilmiş ve dili İngilizce olan klinik deneme meta analizlerinin tamamı çalışmanın evrenini oluşturmuştur. Çalışma için yayın yanlılığı konusunda Cochrane Collaboration'ın klavuzundan edinilen bilgiler ışığında bir soru formu oluşturulmuştur. Soru formu alt başlıklar olarak: yayın yanlılığını önlemek ve tahmin etmek için kullanılan: sistematik metotlar (funnel plot, forest plot, Egger regresyon testi), heterojenite testleri, alt grup analizleri, duyarlılık analizleri, analize alınan çalışmaların kalite değerlendirmelerini içermektedir ve değerlendirmeye alınan her meta analizi bu sorular ışığında incelenmektedir. Çalışmada, 01.06.2013 tarihi itibariyle PubMed ile ulaşılabilen ve çalışmanın arama kriterlerine uygun 198 meta analizi bu sorulara yanıt aramak amacıyla değerlendirilmiştir.

Klinisyenler tarafından kolayca ulaşılabilen bir kaynak olan PubMed'den erişilen klinik dal dergilerinde yayınlanan meta analizlerinde rapor edilen 'yayın yanlılığı değerlendirilmeleri' mutlaka göz önünde bulundurulması gereken bir konudur. Zira, çeşitli alt başlıklarda yayın yanlılığı değerlendirmelerini talep eden raporlama formatlarının (QUAROM, PRISMA) klinik dal dergilerinde kurallaştırılması henüz çok yaygın bir uygulama değildir.

Anahtar Kelimeler: Klinik denemeler, meta analizi, yayın yanlılığı, yayın yanlılığının değerlendirilmesi, PubMed.

P19

PEDİATRİK YOĞUN BAKIM BİRİMLERİNİN HİZMET KALİTESİ DEĞERLENDİRİLMESİ İÇİN BİR ÖNERİ PRISM-II VE PIM-II RİSK ÖLÇÜTLERİ

Nazan Alparslan[1], Yaşar Sertdemir[1], Dinçer Yıldızdaş[2], Hacer Yapıcıoğlu[3], Özden Özgür Horoz[2]

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

[2]*Çukurova Üniversitesi Tıp Fakültesi Pediatrik Yoğun Bakım birimi, Adana*

[3]*Çukurova Üniversitesi Tıp Fakültesi Yeni Doğan Bölümü, Adana*

nazan@cu.edu.tr

Giriş: Ölüm riskinin tahmin modelleri yardımıyla belirlenmesi yoğun bakım birimlerine yatırılan hastalar için standart bir işlemdir. Çocuk yoğun bakım birimlerinde mortalite tahmini için iki alternatif skor kullanılmaktadır: Pediatrik Ölüm Riski (PRISM 1, PRISM 2 ve PRISM 3) ve Pediatrik Ölüm Endeksi (PIM ve PIM-II). Bu ölçütlerle yoğun bakıma yatırılan hastaların ölüm olasılıkları hesaplanabilir. Bu ölüm olasılıklarının standart olarak kabul edilmesi durumunda PRISM-II ve PIM-II skorları yoğun bakım birimlerinin değerlendirilmesine yarayabilir.

Amaç: Çukurova Üniversitesi Balcalı Hastanesi çocuk yoğun bakım biriminin PRISM-II ve PIM-II kullanılarak değerlendirilmesi.

Yöntem: Çukurova Üniversitesi Balcalı Hastanesi çocuk yoğun bakım biriminde 01.03.2011 - 29.02.2012 tarihleri arasında tedavi gören 706 hasta değerlendirilmiştir. PRISM-II ve PIM-II ölçütlerinin ölümleri ayırt etmesindeki performansları hesaplanmış, ölçütlerle tahmin edilen ölüm olasılıkları gözlenen ölümlerle karşılaştırılmıştır.

Bulgular: Değerlendirilen 706 hasta için ölüm hızı %6,1(43 hasta)dır. Ölçütlerin ölümlerin ayırt edilmesindeki performansları ROC analizinde eğri altında kalan alan (AUC) ile değerlendirildiğinde PRISM-II için AUC=0,993, %95 GA (0,989-0,998) ve PIM-II için AUC=0,992, %95 GA (0,987-0,9) olarak bulunmuştur. PRISM-II ve PIM-II için elde edilen standartlaştırılmış ölüm hızları ve %95 GA sırasıyla: 0,51(0,38-0,69), ve 0,35(0,26-0,47) olarak hesaplanmaktadır.

Sonuç: PRISM-II ve PIM-II skorları kullanılarak değerlendirildiğinde, Çukurova Üniversitesi Balcalı Hastanesinin çocuk yoğun bakım biriminde tedavi gören hastaların gözlenen ölüm hızlarının beklenen değerinden anlamlı derece düşük olduğu bulunmuştur. Çocuk yoğun bakım birimlerinin PRISM-II ve PIM-II skorları kullanılarak değerlendirilmesi, SMR değeri >1 olanların hizmet kalitesinin arttırılması için uyarılması önerilebilir. Ayrıca bu iki ölçüt, alt gruplarda yapılan analizlerle birimin hangi hasta gruplarında daha başarılı /başarısız olduğu konusunda da veri sağlayabilir.

Anahtar kelimeler: *PIM 2, PRISM 2, Çocuk yoğun bakım, SMR, AUC, Standart hizmet.*

P20

BOZULMUŞ NORMAL DAĞILIM ALTINDA KONUM VE DAĞILIM PARAMETRE TAHMİN EDİCİLERİNİN ETKİNLİKLERİNİN BENZETİM YOLUYLA KARŞILAŞTIRILMASI

HülyaYılmaz[1], Hakan Savaş Sazak[2]

[1]*Eskişehir Osmangazi Üniversitesi Tıp Fakültesi, Biyoistatistik ve Tıbbi Bilişim AD, Eskişehir*

[2]*Ege Üniversitesi Fen Fakültesi, İstatistik Bölümü, İzmir*

hulyayilmaz@ogu.edu.tr

Amaç: Bu çalışmada, Bozulmuş Normal Dağılım durumunda Huber'ın M tahmincileri (BS82 ve w24) ve geleneksel konum ve dağılım parametre tahmin edicilerinin (medyan ve medyan mutlak sapma, aritmetik ortalama ve örneklem standart sapması) Hata Kareler Ortalaması (MSE) değerlerinin aritmetik ortalama ve standart sapmaya göre göreceli etkinliklerinin karşılaştırılması amaçlanmaktadır.

Yöntem: (100000/n) Monte Carlo benzetim tekniği ile veri türetimi yapılmıştır. MATLAB paket programından faydalanılarak üç farklı benzetim çalışması yapılmıştır. Ortalamanın 0 , varyansın σ^2 olduğu, bozulmuş simetrik dağılımlar (normal ve düzgün dağılımdan) elde edilmiştir. İlk çalışmada, farklı örneklem ölçümleri (n) altında belirli π olasılık değerine göre $N(0,k^2)$ dağılımından gelen veriler ile $(1-\pi)$ olasılıkla N(0,1) standart normal dağılımdan gelen veri setleri oluşturulup bunların parametre tahminlerini yapan bir algoritma kurulmuştur. İkinci benzetim çalışmasında algoritma, yine farklı örneklem ölçümleri (n) altında belirli π olasılık değerine göre düzgün dağılım U(-k,k) ve $(1-\pi)$ olasılıkla standart normal dağılımdan gelen veriler bütününün konum ve dağılım parametre tahminlerini yapmaktadır. Her iki algoritma; farklı n (n=20,50,100), π (π=0.05, 0.1, 0.2) ve k (k=5, 10, 20) değerleri için çalıştırılarak bahsi geçen konum ve dağılım parametre tahmin edicilerine ilişkin MSE değerleri ve bu MSE değerleri kullanılarak konum parametreleri için ortalamaya ve dağılım parametreleri için ise örneklem standart sapmasına göre göreceli etkinlikler elde edilmiştir. Son olarak ise verilerin sadece standart normal dağılımdan olduğu koşul altında tahmincilerin etkinlikleri araştırılmıştır.

Bulgular: Normal dağılımın bozulmasının hedeflendiği her iki benzetim çalışmasında da tüm örneklem ölçümleri sabit iken π olasılık değeri ve k dağılım katsayı değeri arttıkça veri seti normallikten uzaklaşmıştır. Bu koşullar altında Huber'ın M konum parametre tahmincilerinin ve medyanın, aritmetik ortalamaya göre özellikle $k\geq10$ ve $\pi\geq0.1$ koşulunda en az %230 etkin bir performans sergilediği görülmüştür. Dağılım parametresi tahmin edicilerinden örneklem standart sapması, Huber'ın tahmincileri ve medyanın dağılımını gösteren medyan mutlak sapma (median absolute deviation) değeri karşısında oldukça etkin bir performans sergileyip, tüm simülasyon çalışmalarında diğer tahmincilerin etkinliklerinin %100 ün üstüne çıkmasını engellemiştir. Veri setinin

sadece standart normal dağılımdan olduğu üçüncü benzetim çalışmasında ise en etkin konum ve dağılım parametre tahmincileri doğal olarak aritmetik ortalama ve standart sapma olmuştur.

Sonuç: Bu çalışmada elde edilen sonuçlara göre, bozulmuş normal dağılıma ait veri setleri için en uygun konum parametresi tahmin edicisi Huber'ın M tahmincileri, en iyi dağılım parametresi tahmincisi ise örneklem standart sapması olmuştur. Normal dağılım sergileyen veri setlerinde ise en uygun konum ve dağılım parametre tahmincileri beklenildiği gibi sırasıyla aritmetik ortalama ve örneklem standart sapması olarak bulunmuştur.

Anahtar kelimeler: Bozulmuş normal dağılım, M tahmincileri, parametre tahmini

P21

SIRALI KÜME ÖRNEKLEMESİ İLE TEK VE İKİ YIĞIN ORTALAMASI İÇİN HİPOTEZ TESTLERİNİN GÜÇ VE GEREKLİ ÖRNEK ÇAPI BAKIMINDAN İNCELENMESİ

Yaprak Arzu Özdemir[1], Fikri Gökpınar[1]

[1]*Gazi Üniversitesi Fen Fakültesi, İstatistik Bölümü, Ankara*

yaprak@gazi.edu.tr

Amaç: Sıralı küme örneklemesi (SKÖ), birimleri ölçmenin zor ancak sıralamanın kolay olduğu durumlarda basit tesadüfi örneklemeye(BTÖ) göre daha etkin tahmin edicilerin elde edilmesini sağlayan bir örnekleme tekniğidir. Özellikle çevre, ekoloji, tıp ve tarım gibi alanlarda sıkça kullanılmaktadır.

Hipotez testlerinde BTÖ yerine SKÖ nün kullanılması durumunda elde edilen güç değerlerinin daha yüksek olduğu yapılan çalışmalarda gösterilmiştir(Mutlak ve Abu Dayyeh (1998), Pan and Sien(2002), Shen(1994)). Ayrıca Tseng ve Wu (2007) göreli etkinlik değerlerine bağlı olarak normal ve üstel dağılımın ortalamasına ilişkin hipotez testi için Medyan SKÖ ve SKÖ altında kritik değerleri elde etmek üzere bir formül geliştirmişlerdir. Bununla birlikte Özdemir ve Gökpınar (2010), Gökpınar ve Özdemir(2011), SKÖ de yığın ortalamasına ilişkin hipotez testinde Meta Modelleme yöntemini kullanarak kritik değer için örnek çapına bağlı yaklaşık bir formül geliştirmişlerdir.

Bu çalışmada, BTÖ yerine SKÖ kullanılarak elde edilen örnekten yararlanarak yapılacak yığın ortalaması için hipotez testlerinde testin gücü ve gerekli örnek çapı elde edilecektir.

Yöntem: SKÖ altında örnek ortalaması istatistiğinin dağılımı teorik olarak bulunamadığından hipotez testi için gerekli kritik değerlerin elde edilmesi mümkün olamamaktadır. Bu nedenle Monte Carlo yöntemi kullanılarak, tek ve iki grup ortalama farkının hipotez testine ilişkin farklı örnek çapları için kritik değerler elde edilmiştir. Kritik değerler kullanılarak SKÖ tasarımı, BTÖ ye göre I. tip hata ve testin gücü bakımından karşılaştırılmış ve hangi durumlarda daha iyi sonuç verdiği belirlenmiştir. Bunun yanı sıra verilen bir güç değerine göre en az kaç birim örnek gerektiği belirlenmiş ve BTÖ'ye göre karşılaştırılmıştır.

Bulgular ve Sonuç: Elde edilen sonuçlara göre, örnek çapları arttıkça hem tek hem de iki grup için elde edilen kritik değerlerin, küçük örnek çaplarında bile normal dağılım varsayımı altında elde edilen kritik değerlere yakınsadığı gözlemlenmiştir. Ayrıca testin gücü ve örnek çapı için elde edilen sonuçlar incelendiğinde, aynı ortalama farkı ve testin gücü değerlerinde SKÖ için gereken örnek çapının BTÖ için gereken örnek çapının yaklaşık 3 de 1 i kadar olduğu gözlemlenmiştir. Bu durum özellikle yüksek maliyet veya uzun zaman gerektiren ölçümler için SKÖ nün oldukça kullanışlı bir yöntem olduğunu göstermektedir.

Anahtar Kelimeler: Kritik Değer, Sıralı Küme Örneklemesi, Monte Carlo Simülasyonu, Örnek Çapı

Kaynaklar

Mutlak, H. A.,Abu-Dayyeh, W.,(1998). Testing some hypothesis about the normal distribution using ranked set sample: A more powerful test. *Journal of Information and Optimization Sciences*. 19: 1-11.

Özdemir Y.A., Gökpınar, F. (2010) Sıralı Küme Örneklemesi Tasarımında Yığın Ortalamasına İlişkin Hipotez Testi, 7. İstatistik Gümleri Sempozyumu 28-30 Haziran, Bildiri Özetleri Kitabı, 20-21.

Gökpınar F., Özdemir Y.A., (2011) Sıralı Küme Örneklemesi Tasarımlarında İki Ortalama Farkına İlişkin Hipotez Testi, 12. Uluslararası Ekonometri, Yöneylem Araştırması ve İstatistik Sempozyumu, 26-29 Mayıs 2011.

Pan, Y.J., Sien, W.H.,(2002). Tests for normal parameters based on a ranked set sample. *Tunghai Management review*. 4:1-16.

Shen, W.H.,(1994), Use of ranked set sampling for test of a normal mean. *Calcutta Statistical Association Bulletin*. 44: 183-193.

Sinha, B.K., Sinha, B.K., S. Purkayasta,(1996). On some aspects of ranked set sampling for estimation of normal and exponential parameters. *Statistical Decisions* 14: 223-240.

Tseng, Y., Wu, S., (2007). Ranked- Set- Sample- based Tests for Normal and Exponential Means. *Communication in Statistics:Simulation and Computation*.36: 761-782.

P22

0-5 YAŞ ARASI ÇOCUKLARIN BÜYÜME EĞRİLERİ VE REFERANS EĞRİLERİYLE KARŞILAŞTIRILMASI

Nilüfer G. Özbaturlar[1], Timur Köse[1]

[1]*Ege Üniversitesi Tıp Fakültesi Biyoistatistik ve Tıbbi Bilişim A.D., İzmir*

nozbaturlar@gmail.com

Amaç: Büyüme eğrilerinin kullanımı, çocukların büyüme ve gelişmelerinin değerlendirilmesinde büyük öneme sahiptir. Türkiye'de 0-5 yaş grubu, sağlıklı Türk çocukları için Prof. Dr. Olcay Neyzi ve arkadaşlarının oluşturduğu referans büyüme eğrileri kullanılmakta olup, elde edilen boy uzunluğu ve vücut ağırlığı verileri İstanbul ve çevresindeki çocuklara aittir. İzmir ve çevresine ait çalışma olmadığından bu araştırmamızda İzmir ve çevresinde yaşayan 0-5 yaş grubu çocukların boy uzunluğu ve vücut ağırlıklarıyla oluşturulmuş "Persantil Eğrileri ve Tablolarının" Dünya Sağlık Örgütü (DSÖ), Amerika-Hastalık Kontrol Merkezi (ABD-CDC) ve Prof. Dr. Olcay Neyzi ve arkadaşlarının büyüme eğrileriyle karşılaştırılmaları amaçlanmıştır.

Yöntem: Yaptığımız araştırmamızda, İzmir ilinde bulunan Ege Üniversitesi Tıp Fakültesi Çocuk Sağlığı ve Hastalıkları Anabilim Dalı Sağlıklı Çocuk Polikliniği'ne, 1985 - 2011 yılları arasında İzmir ve çevresinden başvuran, 0-5 yaş grubu sağlıklı çocuklara ait veriler alınmıştır. Çalışmamızda boy uzunluğu, vücut ağırlığı ve BMI(beden kitle indeksi)'ye ait veriler kullanılmıştır. İstatistiksel analiz için ACCESS programıyla kaydettiğimiz veriler, Microsoft Excel programına aktarılmış ve SAS 9.2 programıyla LMS Yöntemi kullanılarak persantiller oluşturulmuştur. Referans ve standart büyüme eğrileriyle karşılaştırılmasında ise PASW18 programıyla istatistiksel analiz yapılarak ($p<0.05$) anlamlı kabul edilmiştir.

Bulgu: Araştırmamıza 2269'u (%49,55) kız, 2309'u (%50,45) erkek olmak üzere toplam 4578 sağlıklı çocuklardan elde edilen veriler kullanılmıştır. Persantil eğrileri 3% ile 97% aralığında hesaplanmıştır. Çalışmamızın 0-5 yaş grubu erkek ve kız çocukların 50 persantil ağırlık verileri; Neyzi (2008) referans ve Amerika-CDC 2000 referans büyüme eğrileri arasında anlamlı fark bulunmuştur. DSÖ (2006) standart büyüme eğrileriyle arasında anlamlı aralarında anlamlı fark ($p <0.05$) bulunmuştur.

0-5 yaş grubu erkek çocukların 50 persantil boy uzunluğu verileri; DSÖ(2006) standart ve Amerika-CDC 2000 referans büyüme eğrileri arasında anlamlı fark bulunmuştur. Neyzi(2008) referans büyüme eğrileriyle arasında anlamlı aralarında anlamlı fark ($p <0.05$) bulunmuştur. 0-5 yaş grubu kız çocukların boy uzunluğu verileri; DSÖ (2006) standart, CDC 2000 ve Neyzi(2008) referans büyüme eğrileri arasında anlamlı fark bulunmuştur.

Sonuç: Çalışmamızdan elde ettiğimiz sonuca göre; çocuklarda büyüme ve gelişmenin izlenmesinde İzmir ve çevresine ait persantil eğri ve tabloları oluşturulmuş ve Türk çocuklarının persantil eğrilerinin de kullanılmasının yararlı (referans) olabileceği düşünülmektedir.

Anahtar Sözcük: Büyüme Eğrileri, Boy Uzunluğu, Vücut Ağırlığı, Büyüme Eğrilerinin Karşılaştırılması

P23

SAĞKALIM ANALİZ YÖNTEMLERİ VE KARACİĞER NAKLİ VERİLERİ İLE BİR UYGULAMA

Feyza İnceoğlu[1], **Harika Gözükara Bağ[1]**, Cemil Çolak[1], Saim Yoloğlu[1], Sezai Yılmaz[2]

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

[2]*İnönü Üniversitesi, Tıp Fakültesi, Genel Cerrahi Anabilim Dalı, Malatya*

harika.gozukara@inonu.edu.tr

Amaç: Bu çalışmada amaç; karaciğer nakli sonrası sağkalım üzerine etki eden değişkenlerin belirlenmesi için Yaşam Tablosu, Kaplan-Meier ve Cox regresyon sağkalım analizlerinin sonuçlarının karşılaştırılmasıdır.

Yöntem: Çalışmada, 2002 yılı Mart ayı ve 2012 yılı Kasım ayı arasında Malatya Turgut Özal Tıp Merkezi Hastanesi Karaciğer Nakil Enstitüsünde karaciğer nakli yapılan 894 hastaya ait sağkalım verileri kullanılmıştır. Çalışmada ele alınan değişkenler; alıcının kan grubu (A Rh (+), A Rh (-), B Rh (+), B Rh (-), 0 Rh (+), 0 Rh (-), AB Rh (+), AB Rh (-)), vericinin kan grubu (A Rh (+), A Rh (-), B Rh (+), B Rh (-), 0 Rh (+), 0 Rh (-), AB Rh (+), AB Rh (-)) ve alıcı ile vericinin kan grubu eşleşmesi (gruplar aynı, gruplar farklı) değişkenleri ele alınmıştır.

Bulgular: Alıcının kan grubuna göre oluşturulan sekiz grup arasında, Yaşam Tablosu Wilcoxon, Kaplan – Meier Logrank (Mantel – Cox), Breslow (Generalized Wilcoxon), Tarone Ware testlerine ve Cox Regresyon Analizine göre sağkalım süresi yönünden istatistiksel olarak farklılık bulunmuştur. Vericinin kan grubuna göre ve dikkate alınan bir diğer değişken olan alıcı ile verici kan gruplarının eşleşmesine bağlı olarak elde edilen iki grup için Yaşam Tablosu Wilcoxon, Kaplan – Meier Breslow (Generalized Wilcoxon), Tarone Ware testlerine göre sağkalım süresi yönünden anlamlı bir farklılık bulunurken, Kaplan – Meier Logrank (Mantel – Cox) testi ve Cox Regresyon Analizine göre bir farklılık bulunmamıştır.

Sonuç: Yapılan analizlerde dikkate alınan bazı değişkenler için tüm yöntemler aynı sonucu verirken, bazı değişkenler için yöntemlerin sonuçları arasında farklılıklar gözlenmiştir.

Anahtar Kelimeler: Sağkalım Analizi, Yaşam Tablosu Analizi, Kaplan – Meier Analizi, Cox Regresyon Analizi

P24

GÜVENİRLİK ANALİZİNDE SINIF İÇİ İLİŞKİ KATSAYILARININ UYGULAMA ÖRNEKLERİ ÜZERİNDE İRDELENMESİ

Ahmet Dirican[1], **Ömer Uysal[2]**, Başak Gürtekin[1], Sevda Özel[1], Sevim Purisa[1], M. Fatih Sert[1], Rian Dişçi[1]

[1]İ.Ü. İstanbul Tıp Fakültesi, Biyoistatistik ve Tıp Bilişimi Anabilim Dalı, İstanbul

[2]Bezmialem Vakıf Üniversitesi, Tıp Fakültesi, Biyoistatistik ve Tıp Bilişimi Anabilim Dalı, İstanbul

omeruysal@yahoo.com

Amaç: Bilimsel araştırmalarda ele alınan ölçütlere ait boyutların saptanmasındaki tutarlılık ve tekrar edilebilirlik, araştırmanın vereceği sonuçların doğruluğunun en önemli gereklerindendir. Güvenirlik çalışmalarında temel amaç, aynı denekten alınan tekrarlı ölçümler veya aynı denek üzerinde iki ya da daha fazla sayıda gözlemcinin ölçümleri arasındaki (gözlemci içi veya gözlemciler arası) uyumun değerlendirilmesidir. Bu tip kurgular için, sınıf içi korelasyon katsayısı (SKK) başta olmak üzere çeşitli uyum ve ilişki katsayısı yöntemleri kullanılır.

Yöntem: Çalışmanın başında deneyleme düzenine uygun olan hesaplama yöntemi ve modelinin belirlenmesi önemlidir. Çözümlemenin amacına, tasarımına ve alınan ölçümün tipine bağlı olarak farklı yaklaşımların (tek yönlü modeller, iki yönlü modeller) kullanıldığı sınıf içi korelasyon katsayısı hesaplamaları ile özgün yorumlara ulaşılabilmektedir. SKK veya diğer ilişki katsayıları ile ölçümün ne kadarının hatadan arınmış olduğunun veya hata düzeyinin ortaya konması ile klinik gerçekler tanımlanmaya ve nedensellikleri açıklanmaya çalışılır.

Bulgular: Hesaplama yöntemlerinin, örneklemin yapısı ve büyüklüğünden çeşitli düzeylerde etkilenmeleri söz konusudur. Sonuçların doğruluğunun artırılabilmesi için, ölçüm değerlerindeki değişebilirlik ve çeşitli hataların etkilerine karşı önlemler alınmalı, sonuç bulguların ve yorumunun en az hata ile sunulması sağlanmalıdır.

Sonuç: Çalışmamızda, sınıf içi ilişki denetlemesinde kullanılacak uygun istatistiksel yöntemlerin seçimi ve hesaplama yaklaşımları, uygulama örnekleri üzerinde ayrıntılı olarak irdelenecektir.

Anahtar Kelimeler: Güvenirlik, Sınıf İçi Korelasyon Katsayısı, Hata

P25

DİJİTAL SİNYAL İŞLEME YÖNTEMLERİYLE ELDE EDİLEN VERİLER YARDIMIYLA OBSTRÜKTİF UYKU APNESİNİNİN SINIFLANDIRILMASINDA ÇOK DEĞİŞKENLİ İSTATİSTİKSEL ANALİZ YÖNTEMLERİ VE YAPAY SİNİR AĞLARI

Necdet Süt[1], İlhan Umut[2], Erdem Uçar[2], Levent Öztürk[3]

[1]*Trakya Üniversitesi Tıp Fakültesi Biyoistatistik ve Tıbbi Bilişim Anabilim Dalı, Edirne*

[2]*Trakya Üniversitesi Mühendislik Fakültesi Bilgisayar Mühendisliği Bölümü, Edirne*

[3]*Trakya Üniversitesi Tıp Fakültesi Fizyoloji Anabilim Dalı, Edirne*

nsut@trakya.edu.tr

Amaç: Çalışmada dijital sinyal işleme yöntemleriyle (Fourier transformation) EEG frekanslarının elde edilmesi, bu veriler yardımıyla obstrüktif uyku apnesininin çok değişkenli istatistiksel analiz yöntemleriyle (Diskriminant Analizi, Lojistik Regresyon Analizi) ve yapay sinir ağlarıyla (Çok Katmanlı Perseptron - MLP) sınıflandırılması amaçlanmıştır.

Yöntem: Veriler Temmuz 2009 ile Haziran 2010 tarihleri arasında Trakya Üniversitesi Sağlık Uygulama ve Araştırma merkezi uyku bozuklukları ünitesinde tedavi görmüş hastaların EEG kayıtlarından retrospektif olarak elde edildi. Ham veriler obstrüktif uyku apneli 58 hastanın ve uyku apnesi bulunmayan 30 kontrolün 8 saatlik (22:00 – 06:00 arası) PSG (polysomnography) kayıtlarında yer alan EEG kanalları üzerinden elde edildi. Verileri elde etmede apne öncesi 5 saniyelik EEG kayıtlar ile kontrol grubundan rastgele seçilen 5 saniyelik EEG kayıtları kullanıldı. EEG kayıtları üzerinde yer alan ilk apneler solunum kayıt kanallarında işaretlenerek ve anormal solunum olayı ile aynı zaman dilimine denk gelen EEG kayıtları C4A1 ve C3A2 kanallarından alınarak EEG dijital sinyal işleme teknikleri (Fourier transformation) kullanılarak delta (0,5 – 4 Hz), teta (4 – 7 Hz), alfa (8 – 12 Hz) ve beta (15 – 30 Hz) frekans bantlarının yüzde değerleri hesaplandı. Çalışmada hastalara ilişkin çeşitli demografik özellikler (yaş, kilo, boy, boyun çapı ve cinsiyet) ile Fourier dönüşümü ile elde edilen alfa, beta, delta ve teta frekans değerleri incelenerek obstrüktif uyku apnesi diskriminant analizi, lojistik regresyon analizi ve yapay sinir ağları (MLP) ile sınıflandırıldı.

Bulgular: Yapılan univariate analizler sonucunda obstrüktif uyku apnesi olanlarla olmayanlar arasında C4A1 ve C3A2 kanallarından alınan delta ve beta EEG frekanslarının ve domografik verilerden kilonun istatistiksel anlamlı farklı olduğu bulundu ($p<0,05$). Anlamlı bulunan bu değişkenlerle yapılan obstrüktif uyku apnesini sınıflandırmada MLP'nin pozitif kestirim değeri ve doğruluk oranının (%87,5 ; %84,0), Diskriminant Analizi (%83,3 ; %75,0) ve Lojistik Regresyon Analizinden (%78,3 ; %78,4) yüksek olduğu bulundu.

Sonuç: Dijital sinyal işleme yöntemleriyle C4A1 ve C3A2 kanallarından alınan EEG frekanslarına ve kiloya dayalı olarak MLP obstrüktif uyku apnesini sınıflandırmada Diskriminant Analizi ve Lojistik Regresyon Analizinden daha yüksek doğrulukta sonuç vermektedir. Bu nedenle MLP'nin obstrüktif uyku apnesini önceden tahmin etmeye yönelik modelleme çalışmalarında kullanılması düşünülmelidir.

Anahtar Kelimeler: Dijital sinyal işleme, Obstrüktif uyku apnesi, Yapay sinir ağları

P26

FAKTÖR ANALİZİ VE LOJİSTİK REGRESYON ÇÖZÜMLEMESİNİN BİRLİKTE KULLANIMININ ÖNEMİ: POSTER SUNUMU

Arzu Çalışgan[1], Yeşim Tunç[1], Alev Bakır[1], Mustafa Şükrü Şenocak[1]

[1]*İstanbul Üniversitesi, Cerrahpaşa Tıp Fakültesi Biyoistatistik ve Tıp Bilişimi Ana Bilim Dalı*

arzucalisgan@gmail.com

Faktör analizi çok sayıda değişkenin sorgulandığı klinik çalışmalarda sonuca gerçekten etki eden değişkenlerin birbiriyle yüksek korelasyona sahip olanlarını birleştirerek faktörler oluşturur. Sayısı değişken sayısından daha az olan faktörler klinik çalışmanın analizinde kolaylıklar sağlar. Başka bir deyişle faktör analizi, birçok değişkenin birkaç başlık altında toplanması tekniğidir. Değişkenlerin faktörler üzerindeki ağırlıklarını hesaplayan bir yöntemdir.

Faktör analizi sonucu olarak tek faktörün çıktığı durumlarda soruların faktör üzerine ağırlıkları hakkında bir karar verilemez. Böyle durumlarda birlikte kullanılan Lojistik Regresyon'un önemi ortaya çıkmaktadır.

Faktör analizi ile lojistik regresyonun birlikte kullanımının önemini test etmek için 8 sorudan oluşan bir ölçek sanal olarak üretildi. Sorulara verilen puanlar (1-5) özel bir yazılım kullanılarak 2 soruya yüksek, diğer sorulara rassal olarak dağıtıldı. Sorulardan elde edilen skorlardan klinisyen tarafından kesim (cut-off) noktasına göre belirlenen belli bir olayın varlığı / yokluğunu belirten dikotom bir değişken üretildi. Bu değişken bağımlı değişken olarak alınarak soruların bağımlı değişken üzerine etkisi lojistik modelle incelendi. Ortaya çıkan anlamlı lojistik modelde kasıtlı olarak yüksek puan atanan değişkenlerin Odds Ratio değerleri diğer sorulara göre daha yüksek çıktı. İlginç olan bir sonuç bazı soruların bağımlı değişken üzerine etkisi modelde anlamsız olarak bulundu. Anlamsız çıkan değişkenleri göz ardı ederek faktör analizini uyguladığımızda açıklanan varyans oranının arttı.

Böylece faktör analizi tek başına uygulandığında tüm soruları analize dahil etmemiz gerektiği sonucuna ulaşmamıza rağmen, lojistik regresyonla beraber uygulandığında bazı soruları göz ardı edebileceğimiz sonucuna vardık. Bu yöntem gerçek bir klinik çalışmadan elde edilen veriler üzerinde de denenmiş ve benzer sonuçlar bulunmuştur.

Anahtar Kelimeler: Faktör analizi, Ölçek, Lojistik regresyon analizi

P27

HİYERARŞİK KÜMELEME YÖNTEMLERİNİN KARŞILAŞTIRILMASI

Sinan Saraçlı[1], Nurhan Doğan[2] , **İsmet Doğan[2]**

[1]*Afyon Kocatepe Üniversitesi Fen Edebiyat Fakültesi İstatistik Bölümü*

[2]*Afyon Kocatepe Üniversitesi Tıp Fakültesi Biyoistatistik ve Tıbbi Bilişim AD*

idogan@aku.edu.tr

Amaç: Bu çalışmada, Cophenetic korelasyon katsayısı kullanılarak, farklı koşullarda hangi hiyerarşik kümeleme yönteminin en iyi olduğunun belirlenmesi amaçlanmıştır.

Yöntem: Çalışmada, MATLAB paket programı kullanılarak simülatif olarak türetilen veriler kullanılmıştır. Verilerin türetilmesinde $\bar{X} = 0, \sigma^2 = 1$olan çok değişkenli normal dağılımdan yararlanılmıştır. Çok değişkenli normal dağılımdan veri türetilirken, aykırı değer içermeyen ve %5 aykırı değer içeren iki farklı durum dikkate alınmıştır. Değişken sayısının 3, 5 ve 10 olarak belirlendiği, her bir değişken için 10, 50 ve 100 gözlemin yer aldığı çalışmada, yedi farklı kümeleme yöntemi dokuz farklı uzaklık ölçüsü kullanılarak değerlendirilmiştir. Her bir durum için simülasyon sayısının $100000/n$ olarak belirlendiği çalışmada 2x3x3x7x9 = 1134 adet farklı senaryo planlanmıştır. Her bir senaryo için $100000/n$ kez elde edilen Cophenetic korelasyon katsayılarının ortalaması dikkate alınarak, ortalamanın en yüksek değere ulaştığı durumdaki kümeleme yöntemi en iyi yöntem olarak belirlenmiştir.

Bulgular: Çalışmada n=10 için tüm durumlarda en iyi hiyerarşik kümeleme yönteminin Ortalama Bağlantı Kümeleme Yöntemi (Average Linkage Method), n=50 ve n=100 için ise Küresel Ortalama Bağlantı Kümeleme Yöntemi (Centroid Linkage Method) olduğu belirlenmiştir.

Sonuç: Literatürde, hiyerarşik kümeleme yöntemlerinin karşılaştırıldığı çalışmalar bulunmaktadır. Ancak çalışmalarda her durumda özellikle öne çıkan bir yöntem söz konusu değildir. Dolayısıyla, hiyerarşik kümeleme yöntemleri kullanılarak yapılacak sınıflandırmalarda en uygun sınıflandırma yönteminin belirlenmesi çalışmadan elde edilecek sonuçların bilimselliği açısından önem taşımaktadır.

Anahtar Kelimeler: Sınıflandırma, Hiyerarşik Kümeleme Analizi, Cophenetic Korelasyon

P28

TÜRKİYE'DE 1987-2011 YILLARI ARASINDAKİ İNTİHAR OLAYLARININ JOINPOINT REGRESYON ANALİZİ İLE DEĞERLENDİRİLMESİ

Nurhan Doğan

Afyon Kocatepe Üniversitesi Tıp Fakültesi Biyoistatistik ve Tıbbi Bilişim AD

nurhandogan@hotmail.com

Amaç: Bu çalışmada, Joinpoint Regresyon Analizi (JRA) kullanılarak, 1987-2011 yılları arasında Türkiye'de yaşanan intihar olaylarının yıllar itibari ile değişiminin yaş ve cinsiyet bazında ortaya konması amaçlanmıştır.

Yöntem: Çalışmada kullanılan veriler Türkiye İstatistik Kurumu tarafından yayınlanan intihar istatistikleri kitapçıklarından elde edilmiştir. JRA, intihar olayları ile ilgili trendlerdeki anlamlı değişikliklerin olduğu noktaların belirlenmesi amacıyla kullanılmıştır. Trendlerin incelenmesinde yaş ve cinsiyet ayrı ayrı dikkate alınmıştır. Her bir cinsiyet için yaşa göre standartlaştırılmış intihar oranları Dünya standart nüfusuna göre direkt standartlaştırma yöntemi kullanılarak elde edilmiştir. Oranlardaki değişimleri tanımlamak için yıllık yüzde değişim (APC) hesaplanmıştır. Verilerin değerlendirilmesinde Joinpoint Regression Program, Version 4.0.4 paket programından yararlanılmıştır.

Bulgular: Türkiyede 1987-2011 yılları arasında 25 yıllık bir dönemde gerçekleşen intihar sayısı 50642'dir. İntihar edenlerin 18373 (%36,28)'ü kadın 32269 (%63,71)'u ise erkektir. Türkiye'de intihar edenler içerisindeki en büyük oran; kadınlarda %43,86 ile 15-24 yaş grubunda iken erkeklerde %40,83 ile 25-44 yaş grubunda yoğunlaştığı görülmektedir. Ortalama ilk beş yıllık (1987-1991; %2,4) ve son beş yıllık (2007-2011; %3,6) periyotta yaşa göre standartlaştırılmış intihar oranı ortalama %52,5 artmıştır (her 100.000 kişide).

JRA sonucuna göre, intihar oranları yaş ve cinsiyet ayrımı yapılmadan genel olarak değerlendirildiğinde permütasyon testine göre tek joinpointli (tek kırılma noktalı) model en iyi model olarak tanımlandı. 1987-2006 periyodunda APC:%3,1, 2006-2011 periyodunda ise APC:%-1,7 olarak elde edilmiştir. Erkeklerde, en iyi model sıfır joinpointli model (Trendde anlamlı bir değişiklik olmadığını gösterir) olarak 1987-2011 periyodunda APC:%2.8 olarak belirlenmiştir. Kadınlarda, dört joinpointli model en iyi model olarak elde edilmiştir. Bunlar sırası ile 1987-1992 (APC:%-1,8), 1992-1997 (APC:%8,4), 1997-2000 (APC:%-5,3), 2000-2003 (APC:%15,33) ve 2003-2011 (APC:%-5,41) olarak elde edilmiştir. Yaşlara göre değerlendirildiğinde, erkeklerde 15-24 yaş grubunda iki joinpointli model en iyi model olarak elde edilirken [1987-2005 (APC:%4,9) ve 2005-2011 (APC:%-2,5)] diğer yaş gruplarında sıfır joinpointli model (dönem boyunca %1,5-3,3 aralığında sürekli bir artış

gözlenmiştir) en iyi model olarak belirlenmiştir. Kadınlarda ise 15-24 yaş grubunda dört joinpointli model en iyi model olarak elde edilmiş [1987-1991, APC:%-2,3; 1991-1997 (APC:%11,3), 1997-2000 (APC:%-1,6), 2000-2004 (APC:%14,3) ve 2004-2011 (APC:%-9,7)], 25-34 yaş grubunda ise iki joinpointli model en iyi model olarak elde edilmiştir [1987-2003 (APC:%4,6), 2003-2011 (APC:%-4,6)].

Sonuç: İntihar, Dünya'da en önemli sosyo-medikal problemler arasında yer almaktadır. Dünya Sağlık Örgütü verilerine göre intihar, 15-44 yaş grubunda ölüm nedenleri arasında üçüncü sırada, 15-19 yaş grubunda ise ikinci sırada yer almaktadır. Dünya'da en yüksek intihar oranına sahip ülkeler arasında Greenland, Güney Kore ve Litvanya yer alırken en düşük oranına sahip ülkeler ise Jamaika, Suriye ve Etiyopya'dır.

Bu çalışmanın sonucunda ülkemizde erkeklerde intihar oranının kadınlardan, yaş gruplarına göre ise gerek erkeklerde gerekse kadınlarda diğer yaş gruplarına göre 15-24 yaş grubunda yüksek olduğu tespit edilmiştir. Bu yaş grubundaki erkeklerde 2005 yılı, kadınlarda ise 1991, 1997, 2000 ve 2004 yılları, ayrıca 25-34 yaş grubu kadınlarda ise 2003 yılı kırılma noktası olarak belirlenmiştir.

Anahtar Kelimeler: Joinpoint Regresyon, İntihar İstatistikleri, Türkiye

P29

RANDOLPH'UN ÇOK DEĞERLENDİRİCİLİ BAĞIMSIZ MARJİNAL KAPPASI (K_{FREE}): FLEISS'İN KAPPA'SINA ALTERNATİF BİR YAKLAŞIM

Sevim Purisa

İstanbul Üniversitesi, İstanbul Tıp Fakültesi, Biyoistatistik ve Tıp Bilişimi Anabilim Dalı, İstanbul

purisa@istanbul.edu.tr

Fleiss'in çok değerlendiricili Kappa istatistiği (1971), kategorik değişkenlerin ikiden fazla değerlendirici arasındaki uyumunu analiz etmek için sağlık ve sosyal bilimlerde sıklıkla kullanılmaktadır. Fleiss Kappa (F_K), Scott Pi istatistiğinin çok değerlendiricili genişletilmiş bir versiyonudur. Fleiss Kappa prevelans, yanlılık (bias) ve değerlendirmelerin bağımsız olmaması özelliklerinden oldukça etkilenmektedir. Bu durum, uyum çalışmalarında yüksek prevelans etkisi olduğunda Kappa değerini düşürürken, yüksek yanlılık etkisi olduğunda da Kappa değerini yükseltmektedir.

Fleiss Kappa'nın bu dezavantajından dolayı eğer çalışmada prevelans ve yanlılık etkisi sözkonusu ise Randolph çok değerlendiricili bağımsız marjinal kappanın kullanımının daha doğru sonuç vereceğini ifade etmektedir. Randolph Kappa (K_{free}), Bennett S Kappa'nın çok değerlendiricili genelleştirilmiş şeklidir. K_{free} iki değerlendiricili bağımsız marjinal Kappa'lara (PABAK, S, RE) benzer ve tipik uyum çalışmaları için uygundur. Randolph Kappa'da tesadüfi uyum oranı kategori sayısına göre belirlenmektedir. Bu nedenle prevalans ve yanlılıktan etkilenmez.

Uyum çalışmalarında prevelans ve yanlılık etkisi sözkonusu ise Fleiss Kappa yerine Randolph Kappa'nın kullanılmasının daha uygun olacağı düşünülmekte ve önerilmektedir.

Anahtar kelimeler: Fleiss Kappa, Randolph Kappa, prevelans etkisi, bias etkisi, bağımsız marjinallik

P30

FARKLI ÖRNEKLEM GENİŞLİKLERİ VE DAĞILIMLARDA NORMAL DAĞILIM TESTLERİNİN MONTE CARLO SİMULASYONU İLE KARŞILAŞTIRILMASI

Mustafa Çağatay Büyükuysal[1], Muzaffer Bilgin[2], Fürüzan Köktürk[1]

[1]*Bülent Ecevit Üniversitesi, Biyoistatistik Anabilim Dalı, Zonguldak*

[2]*Eskişehir Osmangazi Üniversitesi, Biyoistatistik Anabilim Dalı, Eskişehir*

cbuyukuysal@gmail.com

Amaç: Bu çalışmada farklı örneklem genişlikleri ve farklı dağılımlarda normal dağılım testlerinin tip-1 hata ve güçleri bakımından karşılaştırılması amaçlanmıştır.

Yöntem: Bu amaç doğrultusunda Shapiro-Wilk, Kolmogorov-Smirnov, Anderson-Darling ve Jarque-Bera testleri incelenmiştir. Farklı örneklem genişlikleri (n=10, 30, 50 ve 100) için simetrik ve simetrik olmayan dağılımlardan Monte Carlo simülasyonu 100.000'er örneklem türetilerek belirtilen testlerin tip-1 hata ve güçleri karşılşatırılmıştır. Simülasyon kodları SAS paket programında yazılmıştır.

Bulgular: Yapılan simülasyon çalışması sonunda farklı dağılım tiplerinde ve örneklem genişliklerinde Nornadiah M.R. ve ark.'larının 2011yılında yaptıkları çalışma sonuçlarında da olduğu gibi Shapiro Wilk testi diğer testlere göre hep daha güçlü bulunmuştur. İkinci en güçlü test Anderson Darling olurken bu testleri sırasıyla Kolmogorov-Smirnov ve Jarque Bera testleri takip etmiştir. Tüm normal dağılım testleri için geçerli olmak üzere örneklem genişliği azaldıkça testlerin güçlerinin düştüğü saptanmıştır.

Sonuç: Sonuç olarak 4 test kıyaslandığında Shapiro Wilk testi en iyi sonucu verirken, örneklem genişliği azaldıkça Anderson Darling testi de Shapiro Wilk testi kadar iyi sonuçlar vermiştir. Örneklem genişliği azaldıkça tüm testlerin güçlerinde anlamlı bir düşüş olduğu ve bu durumda sadece test sonuçlarıya değil, grafiksel yöntemlerle de desteklenmesini tavsiye ediyoruz. Basıklık ve çarpıklık katsayılarının düşük örneklem genişliklerinde yorumlanması tavsiye edilmemektedir çünkü bu katsayılar örneklem genişliğinden etkilenmektedir.

Anahtar Kelimeler: Tip-1 Hata, Güç, Normal dağılım testleri, Monte Carlo

P31

BOLOGNA SÜRECİ VE BİYOİSTATİSTİK PROGRAMLARININ UYUMU

Gülşah Seydaoğlu[1], **Pınar Özaltun[1]**

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

gseyda@cu.edu.tr

Giriş: 2010 yılına kadar Avrupa Yükseköğretim Alanı yaratmayı hedefleyen Bologna Süreci, ülkemizde de hayata geçirilen bir reform sürecidir. Toplam 47 üye ülke tarafından oluşturulan ve sürdürülen, bu zorlu sürece üyelik hükümetler/devletlerarası herhangi bir anlaşmaya dayanmamaktadır. Bologna Süreci kapsamında yayımlanan bildirilerin yasal bir bağlayıcılığı bulunmamaktadır. Süreç tamamen her ülkenin özgür iradeleri ile katıldıkları bir oluşumdur. Bologna Sürecinin temel amacı Avrupa Yükseköğretim Alanı'nın cazip hale getirilmesi, Avrupa Yükseköğretim Alanı içerisinde yer alan ülke vatandaşları, yükseköğrenim görmek ya da çalışmak amaçları ile Avrupa'da kolayca dolaşabilmeleri gerek yükseköğretim ve gerekse iş imkanları açısından dünyanın diğer bölgelerinden kişiler tarafından tercih edilir hale getirilmesidir.

Amaç: Bu çalışmanın amacı biyoistatistik alanında Türkiye'de uygulanan programların "kolay anlaşılır, birbirleriyle ve Avrupa'da uygulanan programlar ile karşılaştırılabilirliği" konusunda durum saptaması yapmaktır.

Yöntem: Türkiye'de eğitim veren fakültelerdeki (Ankara, Hacettepe, Dicle, Çukurova, Mersin, İnönü, Düzce, Osmangazi, İstanbul, Marmara Ün. vb.) lisans, yüksek lisans, doktora Biyoistatistik programları amaç, öğrenim kazanımları ve ders başlıkları, ders içerikleri ve ders kredileri-AKTS (Avrupa Kredi Transfer Sistemi-*European Credit Transfer System, ECTS)* açısından karşılaştırılmıştır. Bu veriler, bilgilerine ulaşılabilen Avrupa üniversite programları ile karşılaştırılmıştır. Karşılaştırma grubu olarak Bologna süreci içinde olmayan ABD, Kanada gibi ülkelerdeki köklü üniversitelerin programlarının içerikleri de analiz edilmiştir.

Bulgular: Programların detaylarına ulaşılan fakültelerde ortak ve zorunlu derslerin başlık ve içeriklerinin, ders saatlerinin ve ders kredilerinin uyumunun zayıf olduğu saptanmıştır. Bu durumun Avrupa ülkelerindeki fakülteler için de benzer olduğu izlenmiştir.

Ayrıca programlarda Bologna süreci gereği programın sosyal boyutunun güçlendirilmesi, öğrenci ve öğretim görevlilerinin hareketliliği ve Avrupa dışındaki ülkelerle işbirliğinin sağlanması ve güçlendirilmesi hedefleri konusunda stratejik hedeflerin tam olarak tanımlı olmadığı saptanmıştır.

Sonuç: Bologna sürecinde en istenmeyen şey üye ülkelerin eğitim sistemlerinin tek tip yükseköğretim sistemi haline getirilmesidir. Avrupa Yükseköğretim Alanı'nda, çeşitlilik ile birlik arasında bir denge

kurulması hedeflenmektedir. Biyoistatistik programlarının kendilerine özgü farklılıkları korunarak birbirleriyle karşılaştırılabilir olması ve uyumlu hale getirilmesi için çalışmalar yapılmasında yarar vardır. Böylece, Avrupa Yükseköğretim Alanı ile Avrupa Araştırma Alanı arasında sinerji yaratılması, programlar arasında geçişin kolaylaşması ve öğrenciler ve öğretim görevlilerin hareketliliği ve istihdamının artırılması sağlanabilir.

Anahtar Kelimeler: Üniversiteler Arası Uyum, Metot Karşılaştırması, Bologna Süreci

P32

LMSP METHOD TO CONSTRUCT WRIST CIRCUMFERENCE REFERENCE CURVES OF 6-17 YEAR-OLD TURKISH CHILDREN

Ahmet Öztürk[1], Betül Çiçek[2], Gökmen Zararsız[1], Gözde Ertürk[1], Yasemin Akşehirli Seyfeli[1], Ferhan Elmalı[1], Mümtaz Mustafa Mazıcıoğlu[3], Selim Kurtoğlu[4]

[1]*Erciyes University, Faculty of Medicine, Department of Biostatistics, Kayseri*

[2] *Erciyes University, Faculty of Health Sciences, Department of Nutrition and Dietetics, Kayseri*

[3]*Erciyes University, Faculty of Medicine, Department of Family Medicine, Kayseri*

[4]*Erciyes University, Faculty of Medicine, Department of Pediatric Endocrinology and Metabolism, Kayseri*

ahmets67@hotmail.com

Aim: Anthropometric indices are frequently preferred by health professionals as valuable tools to determine health and disease, to define nutritional status, to assess growth and development, to evaluate differences in body proportions between populations as well as to optimize diagnosis and treatment. Wrist circumference is a simple anthropometric measure of the skeletal frame size and it has recently been suggested to be associated with a number of diseases. LMSP [1] is a GAMLSS (generalized additive models, location, scale and shape) method which uses a Box-Cox power exponential distribution for model fitting and cubic splines or fractional polynomials for curve fitting. In this study, we aimed to provide wrist circumference references with LMSP method in 6-17 year-old Turkish children and compare our results with other powerful LMS and LMST methods.

Methods: Wrist circumferences of a total of 4330 (1931 boys and 2399 girls) aged 6-17 years were included. Observations were randomly split into training and test sets (70%, 30%). Training set was used to fit the models (for each distribution and each gender) and test set to avoid overfitting, validate models and choose the convenient distribution. LMSP method was applied and its performance was compared with LMS and LMST methods using generalized Akaike information criteria (GAIC) statistic. Analysis were conducted using GAMLSS package of R 3.0.0 software.

Results: Optimum parameters and GAIC statistics for each method and gender are given in Table 1. Best models, in which GAIC values were minimum, were obtained from LMSP method for both girls and boys. LMS method performed the second best fit to the data and LMSP as third. The final model for construction of age-related wrist circumference curves was LMSP (df_μ=2.9, df_σ=1.0, df_ν=0.9, df_τ=0.2, λ=2.51) for boys and LMSP (df_μ=3.6, df_σ=1.1, df_ν=1.0, df_τ=0.6, λ=1.55) for girls.

Conclusion: Wrist circumference references of 6-17 year-old Turkish children were provided in this study. LMSP method was applied to construct these references and outperformed other methods for each gender.

Keywords: Wrist circumference, LMSP, reference curves, Turkish children

Table 1. Comparison of BCT, BCCG and BCPE distributions using GAIC(#=3) for each gender

Distribution	df_μ	df_σ	df_ν	df_τ	λ	GAIC (3)*
Boys						
LMST	2.8	1.0	0.9	0.0	2.20	6.025
LMS	2.9	1.0	1.0	-	2.30	0.317
LMSP	2.9	1.0	0.9	0.2	2.51	0
Girls						
LMST	8.4	1.0	1.0	0.1	0.80	27.385
LMS	9.1	0.8	1.8	-	1.20	23.652
LMSP	3.6	1.1	1.0	0.6	1.55	0

* For clarity; 1707.715 was substracted from GAIC (3) values for boys and 1858.134 for girls for each model

P33

YATAN HASTALARDA TOPLUM-KÖKENLİ PNÖMONİ TEDAVİSİNDE SEFTRİAKSONA KARŞI SEFUROKSİMİN MALİYET-ETKİNLİK ANALİZİ

Merve Katrancı[1], Mevlüt Türe[1], İmran Kurt Ömürlü[1]

[1]*Adnan Menderes Üniversitesi Tıp Fakültesi Biyoistatistik ve Tıbbi Bilişim Anabilim Dalı, Aydın*

mervekatranci@hotmail.com

Amaç: Sağlık hizmetlerinin ekonomik değerlendirilmesinde, en önemli karar alma mekanizması olan *maliyet-etkinlik* analizinden yararlanılmaktadır. Bu yöntemle yapılan ekonomik değerlendirmedeki amaç, farklı sağlık hizmetlerinden elde edilen seçenekler içerisinden daha az maliyetle elde edilenin aynı zamanda daha fazla etkin olanın seçilmesidir. Bu yöntem, seçeneklerin bir maliyet-etkinlik ölçüsü olan maliyet-etkinlik oranlarını (CER) karşılaştırmaktadır. CER her bir başarılı sonuç elde etmenin maliyetidir ve en az CER'e sahip olan seçeneğin maliyet-etkin olarak karar verilmesinde araştırmacılara yol göstermektedir. Bu çalışma, 1-15 yaş grubu yatan çocuk hastalarda toplum-kökenli pnömoni (TKP) tedavisinde kullanılan seftriakson ile sefuroksimin maliyet ve etkinliklerinin karşılaştırılarak, hangi tedavinin maliyet-etkin olduğunun belirlenmesi amacıyla yapılmıştır.

Yöntem: Çalışmamızda hastalar seftriakson verilen grup (n=29) ve sefuroksim verilen grup (n=36) olmak üzere ikiye ayrıldı ve toplam 65 birim cinsiyet, yaş (yıl), ön tanı, maliyet, antibiyotikle ilişkili hastanede kalma süresi (gün) ve ilaçların hastayı tedavi etmedeki başarı durumları açısından incelendi. Hastaların hastanede kalma süresi boyunca hastaneye olan toplam maliyetleri alındı. İlaçların etkinliğini ölçmede, hastaların antibiyotikle ilişkili hastanede kalma süreleri dikkate alındı ve her iki ilacın hastayı tedavi etmedeki başarı durumu maliyetlerinin geometrik ortalamasına dayanarak maliyet-etkinlik oranları karşılaştırıldı.

Sonuç: Karar analizi sonuçlarına göre seftriaksonun maliyet-etkinlik oranı 373,77 TL ve sefuroksimin maliyet-etkinlik oranı 415,95 TL bulundu. Bu sonuç, seftriakson ile sefuroksimin her bir başarılı tedavi sonucunu elde etmenin maliyetleri arasında 42,18 TL 'lik bir fark olduğunu göstermektedir. Sonuç olarak, 1-15 yaş grubu yatan çocuk hastalarda toplum-kökenli pnömoni tedavisinde kullanılan seftriaksonun, sefuroksime göre yüksek başarı oranı, daha kısa antibiyotikle ilişkili hastanede kalma süresi ve düşük maliyet-etkinlik oranı ile daha maliyet-etkin bir tedavi olduğu belirlendi.

Anahtar Kelimeler: Maliyet-Etkinlik Analizi, Pnömoni, Seftriakson, Sefuroksim

P34

İKİ PARALEL DENEY TASARIMINDA HİPOTEZ YARGILAMASINA GÖRE ÖRNEKLEM SAYISI HESAPLARI

Alev Bakır[1], Pınar Ambarcıoğlu[1], Özge Karademir[1], Mustafa Şenocak[1]

[1]*İstanbul Üniversitesi, Cerrahpaşa Tıp Fakültesi, Biyoistatistik ve Tıbbi Bilişim ABD*

alevbakir@yahoo.com

Amaç: Belli bir toplumdan alınan ve toplumu ifade ettiği varsayılan örneklem için örneklem sayısı hesabı, çalışmanın dizaynı göz önüne alınarak hesaplanmaktadır. Farklı çalışma dizaynlarına ait farklı örneklem hesapları bulunmakla birlikte aynı çalışma dizaynında hipotez yargı yorumlamasının farklılıkları için de farklı hesaplama yöntemleri bulunmaktadır. Her vakanın sadece bir grupta yer aldığı randomize dizayn ile tamamlanan "Paralel Dizayn Örnekleme" temel alınarak yürütülmek istenen bir çalışma, iki grubun birbirine denk mi (equality), biri birinden düşük olmama/üstün olma mı (non-inferiority/superiority) ya da birbirine eşdeğer mi (equivalence) olup olmama yargısı için farklı örneklem sayılarına sahip olabilir.

Yöntem: X_{ij} vakası, i hangi gruptan i=1,2 ve j=1,2…….n şeklinde j. kişi olduğunu belirtiyorsa, belirlenmiş bir güvenilirlik ve güç ile 1. ve 2. grubun ortalamaları farkının sıfır olup olmama, bir değerden küçük eşit olup olmama ya da büyük eşit olup olmama durumunun örneklem sayısı (n) hesaplanması.

Bulgular: Bu koşullarda gerçekleştirilen 3 tip farklı hipotez yargılama grubunun farklı formüllerle hesaplandığı ve örneklem sayılarının farklı çıktığı görülmüştür. Bu hipotezler doğrultusunda n örneklem sayıları, %95 güvenilirlik ve %80 güç ile p_1=%60 ve p_2=%80 ayrıca farkın ε=0,20 ve hata payının δ=0,10 olduğu ayrıca her iki grubun örneklem sayılarının eşit olması varsayımları ile hesaplanmış ve sırasıyla; denklik için 78 kişi, düşük olmama için 27 kişi, üstün olma için 246 kişi ve eşdeğerlilik için 341 kişi olarak hesaplanmıştır.

Sonuç: Farklı çalışma dizaynlarında örneklem hesabındaki bu farklılık sonuç yargılarımızın güvenirliliği ve gücü açısından önem taşımaktadır, dolayısı ile örneklem hesabında formülün, hipotez yargısına uygun olarak seçilmesi titizlik isteyen bir durumdur.

Anahtar Kelime: Örneklem hesabı, Paralel deney tasarımı, Farklı hipotez yargılamaları

P35

VERİ MADENCİLİĞİ ANALİZİNDE KULLANILAN YÖNTEMLERDEN HANGİLERİNİN TIPTA KULLANIMI UYGUNDUR?

Setenay Öner[1], Canan Baydemir[1], Ertuğrul Çolak[1], Kazım Özdamar[1]

[1]*Eskişehir Osmangazi Üniversitesi, Tıp Fakültesi, Biyoistatistik ve Tıbbi Bilişim Anabilim Dalı, Eskişehir*

oners@ogu.edu.tr

Tıpta bilginin içeriği ve yapısal durumları hızlı bir şekilde değişim gösterir. Bu değişimde, hastalıkların ortaya çıkışında birçok değişken ve bu değişkenlere ilişkin veriler önemli rol oynamaktadır. Veri madenciliği, bir çok analiz yöntemini kullanarak verinin ilişkilerini kullanarak ve verinin geçmiş analizlerini temel alarak ileriye yönelik tahminler için karar verme modelleri yaratmaktır.

Ortalama yaşam süresinin artması ile kronik hastalıkların görülme sıklığı önemli düzeyde artış göstermektedir. Bir kronik hastalık için sosyal, demografik vb. gibi değişkenler dikkate alınarak, hastalığın ortaya çıkmasında etkin olan değişkenler Faktör Analizi veya Lojistik Regresyon Analizleri ile belirlenmesi mümkündür. Bunun sonucunda da belirlenen bu değişkenlerin hastalık tanımlanması için kullanılacak kesim noktaları belirlenerek, değişkenlerin risk faktörü olarak tanımlanması amaçlanır.

Bu çalışmada, tıpta kullanılan değişkene ve veri tipine göre Veri Madenciliğinde kullanılan klasik yöntemlerden Regresyon Modelleri, K-En Yakın Komşuluk, Kümeleme yöntemleri ile Karar Ağaçları, Sinir Ağları, Kohonen Ağları, Bayesci Ağlar, Diskriminant Analizi, Faktör Analizi, Genetik Algoritmalar ve Temel Bileşenler Analizi yöntemlerinden hangisinin kullanılabileceği tartışılmıştır.

Anahtar Kelimeler: Veri Madenciliği, Regresyon Modelleri, K-En Yakın Komşuluk, Kümeleme, Karar Ağaçları, Sinir Ağları, Kohonen Ağları Bayesci Ağlar, Diskriminant Analizi, Faktör Analizi, Genetik Algoritmalar, Temel Bileşenler Analizi

P36

DNA MİKROARRAY ÇALIŞMALARINDA KNORM KORELASYON KATSAYISININ KULLANILMASI

Canan Baydemir[1], Setenay Öner[1], Kazım Özdamar[1]

[1]*Eskişehir Osmangazi Üniversitesi, Tıp Fakültesi, Biyoistatistik ve Tıbbi Bilişim Anabilim Dalı, Eskişehir*

canand@ogu.edu.tr

DNA mikroarray, hücre ve dokulardaki gen ekspresyonu profillerindeki global değişikliklerin incelenmesinde kullanılan bir teknolojidir. DNA microarray'i cam, plastik veya silikon çip gibi katı bir yüzeye tutturularak sıralı bir şekilde (array) oluşturulmuş mikroskobik DNA spotlarıdır. Gen array deneyleri tipik olarak farklı doku ya da koşullardaki ya da bir uygulamadan sonraki gen ekspresyon düzeyini tayin etme amacıyla kullanılır.

Genetik verilerde çalışırken yüksek boyutta veri setleri ile çalışmak zorunda kalınır. Yüksek boyutta veri setleri değerlendirmelerinde büyük tahmin hataları ve sabit olmayan varyans-kovaryans matrisi tahminleri kaçınılmazdır. Özellikler varyans-kovaryans matrisi ve korelasyon matrisi birim sayısının az olduğu fakat değişkenlerin çok olduğu veri yapılarında normalden daha yüksek çıkma eğilimindedir. Bu etkiyi azaltmak için mikroarray çalışmalarında Pearson korelasyon katsayısı yerine Knorm Korelasyon katsayısı kullanılmaktadır. Knorm Korelasyon katsayısı, genetik verilerde satırlarda yer alan altörneklerin ve iteratif olarak tahmin edilen varyans- kovaryans matrislerinin küçültülmesi ve korelasyonların daha güvenilir olarak tahmin edilmesi amacıyla geliştirilmiştir.

Knorm Korelasyon katsayısı genetik veri setlerinde (arraylerde) tekrar sayılarının az olduğunda Pearson korelasyon katsayısına göre daha sapmasız, asimtotik olarak tutarlı ve daha küçük değişkenliğe sahiptir. Knorm korelasyon katsayısı, basitçe veri setindeki satırlarda yer alan denemelerin kısmi korelasyonlarının ağırlıklandırılmış biçimidir.

Bu çalışmada DNA mikroarray çalışmalarında Pearson Korelasyon Katsayısı yerine kullanılmakta olan Knorm Korelasyon Katsayısı tanıtılmıştır.

Anahtar Kelimeler: DNA, Mikroarray Çalışmaları, Knorm Korelasyon Katsayısı

P37

GUIDELINES FOR THE DESIGN AND STATISTICAL ANALYSIS OF BIOLOGICAL EXPERIMENTS

Mahmoud Hozayn

National Research Centre, Cairo, Egypt

m_hozien4@yahoo.com

Abstract: It is important to design biological experiments well, to analyze the data correctly. An appropriate statistical model properly applied to data allows experimental results to be interpreted accurately and completely. In contrast, inappropriate or incomplete statistical designs can lead to inaccurate, or incomplete, conclusions that may lead to misinterpretation of the results of the study. Furthermore, the materials and methods section should describe the statistical methods used in analyzing the results. These guidelines are provided to help scientific research workers perform their experiments efficiently and analyze their results so that they can extract all useful information from the resulting data. Among the topics discussed are the varying of experiments i.e., simple, factorial with two and three factors, split and strip plot experiments. As well as factorial within split plot and split plot within factorial experiments either these experiments designed in completely randomized (CR) and/or completely randomized block designs (CRBD). Also will be discussed the size of the experiment using power and sample size calculations; screening raw data for obvious errors; selecting the suitable transformation methods using SPSS program, assumption and importance of t-test, analysis of variance (ANOVA), simple correlation and regression, multi comparison i.e., LSD, Duncan, Tukey New LSD testes as parametric analysis; and effective design of graphical data.

Keywords: Experimental design; Statistical analysis, Parametric and nonparametric tests