İSTKON 10

10th ISC-2017

INTERNATIONAL STATISTICS CONGRESS

CELEBRATING 25 Years Turkish Statistical Association

DECEMBER 6-8 , 2017
ANKARA UNIVERSITY

ABSTRACT BOOK

TURKISH STATISTICAL ASSOCIATION
www.turkistatistik.org

ANKARA UNIVERSITY

ANKARA UNIVERSITY
FACULTY OF SCIENCES

ANKARA UNIVERSITY
FACULTY OF APPLIED SCIENCES

www.istkon.net

Congress Mail : kayit@istkon.net
Association Mail : tid@turkistatistik.org

10th INTERNATIONAL STATISTICS CONGRESS

DECEMBER 6-8, 2017

CONTENTS

HONORY COMMITTEE v
SCIENTIFIC COMMITTEE vi
ADVISORY COMMITTEE vii
ORGANIZING COMMITTEE viii
ORGANIZERS ix
SPONSORS x
CONGRESS PROGRAM xi
INVITED SPEAKERS' SESSIONS 1
SESSION I 8
STATISTICS THEORY I 8
APPLIED STATISTICS I 14
ACTUARIAL SCIENCES 21
TIME SERIES I 27
DATA ANALYSIS AND MODELLING 33
FUZZY THEORY AND APPLICATION 39
SESSION II 46
STATISTICS THEORY II 46
APPLIED STATISTICS II 52
APPLIED STATISTICS III 58
PROBABILITY AND STOCHASTIC PROCESSES 64
MODELING AND SIMULATION I 70
OTHER STATISTICAL METHODS I 76
SESSION III 82
TIME SERIES II 82
DATA MINING I 88
APPLIED STATISTICS IV 94
OPERATIONAL RESEARCH I 100
OPERATIONAL RESEARCH II 106
SESSION IV 112
APPLIED STATISTICS V 112
APPLIED STATISTICS VI 117
APPLIED STATISTICS VII 122

OTHER STATISTICAL METHODS II 128
OPERATIONAL RESEARCH III 134
DATA MINING II 139
SESSION V 144
FINANCE INSURANCE AND RISK MANAGEMENT 144
OTHER STATISTICAL METHODS III 151
STATISTICS THEORY III 157
MODELING AND SIMULATION II 163
STATISTICS THEORY IV 169
SESSION VI 175
STATISTICS THEORY V 175
APPLIED STATISTICS VIII 181
OTHER STATISTICAL METHODS IV 187
MODELING AND SIMULATION III 193
OTHER STATISTICAL METHODS V 199
APPLIED STATISTICS IX 205
POSTER PRESENTATION SESSIONS 211

HONORY COMMITTEE

Ankara University

Prof. Dr. Erkan İBİŞ	Ankara University, Rector
Prof. Dr. Selim Osman SELAM	Ankara University, Faculty of Science, Dean
Prof. Dr. Harun TANRIVERMİŞ	Ankara University, Faculty of Applied Sciences, Dean

Founding Board Members of Turkish Statistical Association

Prof. Dr. Fikri AKDENİZ	Çağ University
Prof. Dr. Mustafa AKGÜL	Bilkent University
Prof. Dr. Merih CELASUN	
Prof. Dr. Uluğ ÇAPAR	Sabanci University
Prof. Dr. Orhan GÜVENEN	Bilkent University
Prof. Dr. Cevdet KOÇAK	
Prof. Dr. Ceyhan İNAL	Hacettepe University
Prof. Dr. Tosun TERZİOĞLU	
Prof. Dr. Yalçın TUNCER	

Former Presidents of Turkish Statistical Association

Prof. Dr. Orhan GÜVENEN	Bilkent University
Prof. Dr. Yalçın TUNCER	
Prof. Dr. Ömer L. GEBİZLİOĞLU	Kadir Has University
Prof. Dr. Süleyman GÜNAY	Hacettepe University

SCIENTIFIC COMMITTEE

Prof. Dr. İsmihan BAYRAMOĞLU	İzmir University of Economics, TURKEY
Prof. Dr. Hamparsum BOZDOĞAN	University of Tennessee, USA
Prof. Dr. Orhan GÜVENEN	Bilkent University, TURKEY
Prof. Dr. John HEARNE	RMIT University, AUSTRALIA
Prof. Dr. Dimitirios KONSTANTINIDIS	Egean University, GREECE
Prof. Dr. Timothy O'BRIEN	Loyola University, Chicago, USA
Prof. Dr. Klaus RITTER	University of Kaiserslautern, GERMANY
Prof. Dr. Andreas ROßLER	University of Lübeck, GERMANY
Prof. Dr. Joao Miguel da Costa SOUSA	Technical University of Lisbon, PORTUGAL
Prof. Dr. Maria Antonia Amaral TURKMAN	University of Lisbon, PORTUGAL
Prof. Dr. Kamil Feridun TURKMAN	University of Lisbon, PORTUGAL
Prof. Dr. Burhan TURKSEN	TOBB University of Economics and Technology, TURKEY
Prof. Dr. Gerhard-Wilhelm Weber	Charles University, CZECHIA REPUBLIC
Assoc. Prof. Dr. Carlos Manuel Agra COELHO	Universidade Nova de Lisboa, PORTUGAL
Assoc. Prof. Dr. Haydar DEMİRHAN	RMIT University, AUSTRALIA
Assist. Prof. Dr. Soutir BANDYOPADHYAY	Lehigh University, USA

ADVISORY COMMITTEE

Sinan SARAÇLI	Afyon Kocatepe University
Berna YAZICI	Anadolu University
Birdal ŞENOĞLU	Ankara University
Bahar BAŞKIR	Bartın University
Güzin YÜKSEL	Çukurova University
Aylin ALIN	Dokuz Eylül University
Onur KÖKSOY	Ege University
Zeynep FİLİZ	Eskişehir Osmangazi University
Sinan ÇALIK	Fırat University
Hasan BAL	Gazi University
Erol EĞRİOĞLU	Giresun University
Özgür YENİAY	Hacettepe University
İsmail TOK	İstanbul Aydın University
Rahmet SAVAŞ	İstanbul Medeniyet University
Münevver TURANLI	İstanbul Ticaret University
Türkan ERBAY DALKILIÇ	Karadeniz Teknik University
Sevgi Y. ÖNCEL	Kırıkkale University
Müjgan TEZ	Marmara University
Gülay BAŞARIR	Mimar Sinan Güzel Sanatlar University
Dursun AYDIN	Muğla Sıtkı Koçman University
Aydın KARAKOCA	Necmettin Erbakan University
Mehmet Ali CENGİZ	Ondokuz Mayıs University
Ayşen DENER AKKAYA	Middle East Technical University
Coşkun KUŞ	Selçuk University
Nesrin ALKAN	Sinop University
Cenap ERDEMİR	Ufuk University
Ali Hakan BÜYÜKLÜ	Yıldız Teknik University

ORGANIZING COMMITTEE

Head Of The Organizing Committee

Ayşen APAYDIN	Turkish Statistical Association,President

Members Of The Organizing Committee

A. Sevtap KESTEL	Turkish Statistical Association,Vice President
Süzülay HAZAR	Turkish Statistical Association,Vice President
Furkan BAŞER	Turkish Statistical Association,Vice President
Gürol İLHAN	Turkish Statistical Association,General Secretary
İsmet TEMEL	Turkish Statistical Association,Treasurer
Esra AKDENİZ	Turkish Statistical Association,Member
Onur TOKA	Turkish Statistical Association,Member
Serpil CULA	Turkish Statistical Association,Member
Birdal ŞENOĞLU	Ankara University, Department of Statistics
Fatih TANK	Ankara University,Department of Insurance and Actuarial Sciences
Yılmaz AKDİ	Ankara University, Department of Statistics
Halil AYDOĞDU	Ankara University, Department of Statistics
Cemal ATAKAN	Ankara University, Department of Statistics
Mehmet YILMAZ	Ankara University, Department of Statistics
Rukiye DAĞALP	Ankara University, Department of Statistics
Özlem TÜRKŞEN	Ankara University, Department of Statistics
Sibel AÇIK KEMALOĞLU	Ankara University, Department of Statistics
Nejla ÖZKAYA TURHAN	Ankara University, Department of Statistics
Özlem KAYMAZ	Ankara University, Department of Statistics
Kamil Demirberk ÜNLÜ	Ankara University, Department of Statistics
Abdullah YALÇINKAYA	Ankara University, Department of Statistics
Feyza GÜNAY	Ankara University, Department of Statistics
Mustafa Hilmi PEKALP	Ankara University, Department of Statistics
Yasin OKKAOĞLU	Ankara University, Department of Statistics
Özge GÜRER	Ankara University, Department of Statistics
Talha ARSLAN	Eskişehir Osmangazi University, Department of Statistics

ORGANIZERS

TURKISH STATISTICAL ASSOCIATION

ANKARA UNIVERSITY

FACULTY OF SCIENCE
DEPARTMENT OF STATISTICS

FACULTY OF APPLIED SCIENCE
DEPARTMENT OF INSURANCE AND ACTUARIAL SCIENCES

SPONSORS

NGN TRADE INC.

CENTRAL BANK OF THE REPUBLIC OF TURKEY

CONGRESS PROGRAM

	6 DECEMBER 2017 WEDNESDAY					
09:00-09:30	REGISTRATION					
09:30-11:00	OPENING CEREMONY					
11:00-11:15	Tea - Coffee Break					
11:15-12:30	INVITED PAPER I Ankara University Rectorate 100.Yıl Conference Hall Session Chair: Prof. Dr. Fikri AKDENİZ Prof. Dr. Orhan GÜVENEN Some Comments on Information Distortion, Statistical Error Margins and Decision Systems Interactions					
12:30:13:30	LUNCH					
13:30-14.00	POSTER PRESENTATIONS					
14:00-15:45	25th YEAR SPECIAL SESSION - Bernoulli Hall Session Chair: Prof. Dr. Alptekin ESİN Prof. Dr. Fikri AKDENİZ, Prof. Dr. Ömer L. GEBİZLİOĞLU, Prof. Dr. Orhan GÜVENEN, Prof. Dr. Süleyman GÜNAY, Prof. Dr. Ceyhan İNAL					
15:45-16:00	Tea - Coffee Break					
16:00:17:40	SESSION I					
	Bernoulli Hall	Pearson Hall	Fisher Hall	Gauss Hall	Poisson Hall	Tukey Hall
	STATISTICS THEORY I	APPLIED STATISTICS I	ACTUARIAL SCIENCES	TIME SERIES I	DATA ANALYSIS AND MODELING	FUZZY THEORY AND APPLICATION
	ENG	TR	TR	TR	ENG	TR
	SESSION CHAIR	*SESSION CHAIR*	*SESSION CHAIR*	*SESSION CHAIR*	*SESSION CHAIR*	*SESSION CHAIR*
	Mustafa Y. ATA	Fahrettin ÖZBEY	Murat GÜL	Hülya OLMUŞ	Özlem TÜRKŞEN	Nuray TOSUNOĞLU
	A Genetic Algorithm Approach for Parameter Estimation of Mixture of Two Weibull Distributions Muhammet Burak KILIÇ, Yusuf ŞAHİN, Melih Burak KOCA	Investigation of Text Mining Methods on Turkish Text Ezgi PASİN, Sedat ÇAPAR	Mining Sequential Patterns in Smart Farming Using Spark Duygu Nazife ZARALI, Hacer KARACAN	An Investigation on Matching Methods Using Propensity Scores in Observational Studies Esra BEŞPINAR, Hülya OLMUŞ	Intituionistic Fuzzy TIx (If-TIx): Implementation of Intituionistic Fuzzy Set Theory for Evaluating Subjective Workload Gülin Feryal CAN	Assessment of Turkey's Provincial Living Performance with Data Envelopment Analysis Gül GÜRBÜZ, Meltem EKİZ
	Recurrent Fuzzy Regression Functions Approach based on IID Innovations Bootstrap with Rejection Sampling Ali Zafer DALAR, Eren BAS, Erol EGRIOGLU, Ufuk YOLCU, Ozge CAGCAG YOLCU	Cost Analysis of Modified Block Replacement Policies in Continuous Time Pelin TOKTAŞ, Vladimir V. ANISIMOV	Multivariate Markov Chain Model : An Application To S&P500 And Ftse-100 Stock Exchanges Murat GÜL, Ersoy ÖZ	A Simulation Study on How Outliers Effect The Performance of Count Data Models Fatih TÜZEN , Semra ERBAŞ, Hülya OLMUŞ	Evaluation of Municipal Services with Fuzzy Analytic Hierarchy Process for Local Elections Abdullah YILDIZBASI, Babek ERDEBILLI, Seyma OZDOGAN	Modified TOPSIS Methods for Ranking The Financial Performance of Deposit Banks in Turkey Semra ERPOLAT TAŞABAT
	An Infrastructural Approach to Spatial Autocorrelation Ahmet Furkan EMREHAN, Dogan YILDIZ	Examination of The Quality of Life of OECD Countries Ebru GÜNDOĞAN AŞIK, Arzu ALTIN YAVUZ	Use of Haralick Features for the Classification of Skin Burn Images and Performance Comparison of k-Means and SLIC Methods Erdinç KARAKULLUKÇU, Uğur ŞEVİK	Comparison of Parametric and Non-Parametric Nonlinear Time Series Methods Selman MERMİ, Dursun AYDIN	Analyzing the Influence of Genetic Variants by Using Allelic Depth in the Presence of Zero-Inflation Özge KARADAĞ	A New Multi Criteria Decision Making Method Based On Distance, Similarity and Correlation Semra ERPOLAT TAŞABAT
	A Miscalculated Statistic Presented as An Evidence in A Case and Its Aftermath Mustafa Y. ATA	Multicollinearity With Measurement Error Şahika GÖKMEN, Rukiye DAĞALP, Serdar KILIÇKAPLAN	Learning Bayesian networks with CoPlot approach Derya ERSEL, Yasemin KAYHAN ATILGAN	Regression Clustering for PM10 and SO2 Concentrations in Order to Decrease Air Pollution Monitoring Costs at Turkey Aytaç PEKMEZCİ, Nevin GÜLER DİNCER	Survival Analysis And Decision Theory In Aplastic Anemia Case Mariem BAAZAOUI , Nihal ATA TUTKUN	Ranking of General Ranking Indicators of Turkish Universities by Fuzzy AHP Ayşen APAYDIN, Nuray TOSUNOĞLU
	Estimation of Variance Components in Gage Repeatibility & Reproducibility Studies Zeliha DİNDAŞ, Serpil AKTAŞ ALTUNAY	The Effect of Choosing the Sample on the Estimator in Pareto Distribution Seval ŞAHİN, Fahrettin ÖZBEY	Evaluation of Ergonomic Risks in Green Buildings with AHP Approach Ergun ERASLAN, Abdullah YILDIZBASI	Analysis of a Blocked Tandem Queueing Model with Homogeneous Second Stage Erdinç YÜCESOY, Murat SAĞIR , Abdullah ÇELİK , Vedat SAĞLAM	Determinants of Wages & Inequality of Education in Palestinian Labor Force Survey Ola ALKHUFFASH	Exploring The Factors Affecting The Organizational Commitment in an Almshouse: Results Of A CHAID Analysis Zeynep FİLİZ, Tarkan TAŞKIN
		Application of Fuzzy c-means Clustering Algorithm for Prediction of Students' Academic Performance Furkan BAŞER, Ayşen APAYDIN, Ömer KUTLU, M. Cem BABADOĞAN, Hatice CANSEVER, Özge ALTINTAŞ, Tuğba KUNDUROĞLU AKAR				Fuzzy Multi Criteria Decision Making Approach for Portfolio Selection Serkan AKBAŞ, Türkan ERBAY DALKILIÇ

CONGRESS PROGRAM

	7 DECEMBER 2017 THURSDAY					
	SESSION II					
	Bernoulli Hall	Pearson Hall	Gauss Hall	Poisson Hall	Tukey Hall	Rao Hall
09:30-11:10	STATISTICS THEORY II	APPLIED STATISTICS II	APPLIED STATISTICS III	PROBABILITY AND STOCHASTIC PROCESSES	MODELING AND SIMULATION I	OTHER STATISTICAL METHODS I
	ENG	ENG	TR	TR	TR	TR
	SESSION CHAIR	*SESSION CHAIR*	*SESSION CHAIR*	*SESSION CHAIR*	*SESSION CHAIR*	*SESSION CHAIR*
	Serpil AKTAŞ ALTUNAY	Birdal ŞENOĞLU	Yüksel TERZİ	Halil AYDOĞDU	Sibel AÇIK KEMALOĞLU	Nevin GÜLER DİNCER
	Bayesian Conditional Auto Regressive Model for Mapping Respiratory Disease Mortality in Turkey Ceren Eda CAN, Leyla BAKACAK, Serpil AKTAŞ ALTUNAY, Ayten YİĞİTER	Estimation for the Censored Regression Model with the Jones and Faddy's Skew t Distribution: Maximum Likelihood and Modified Maximum Likelihood Estimation Methods Sukru ACITAS, Birdal SENOGLU, Yeliz MERT KANTAR, Ismail YENILMEZ	Comparison of the Lord's Statistic and Raju's Area Measurements Methods in Determination of the Differential Item Function Burcu HASANÇEBİ, Yüksel TERZİ, Zafer KÜÇÜK	Variance Function of Type II Counter Process with Constant Locking Time Mustafa Hilmi PEKALP, Halil AYDOĞDU	A New Compounded Lifetime Distribution Sibel ACIK KEMALOGLU, Mehmet YILMAZ	Analysing in Detail of Air Pollution Behaviour at Turkey by Using Observation-Based Time Series Clustering Nevin GÜLER DİNCER, Muhammet Oğuzhan YALÇIN
	Joint Modelling of Location, Scale and Skewness Parameters of the Skew Laplace Normal Distribution Fatma Zehra DOĞRU, Olcay ARSLAN	Scale Mixture Extension of the Maxwell Distribution: Properties, Estimation and Application Sukru ACITAS, Talha ARSLAN, Birdal SENOGLU	On Suitable Copula Selection for Tempeature Measurement Data Ayşe METİN KARAKAŞ, Mine DOĞAN, Elçin SEZGİN	Power Series Expansion for the Variance Function of Erlang Geometric Process Mustafa Hilmi PEKALP, Halil AYDOĞDU	A New Modified Transmuted Distribution Family Mehmet YILMAZ, Sibel ACIK KEMALOGLU	Outlier Problem in Meta-Analysis and Comparing Some Methods for Outliers Mutlu UMAROGLU, Pınar OZDEMIR
	Artificial Neural Networks based Cross-entropy and Fuzzy relations for Individual Credit Approval Process Damla ILTER, Ozan KOCADAGLI	Maximum Likelihood Estimation Using Genetic Algorithm for the Parameters of Skew-t Distribution under Type II Censoring Abdullah YALÇINKAYA, Ufuk YOLCU, Birdal ŞENOĞLU	Variable Selection in Polynomial Regression and a Model of Minimum Temperature in Turkey Onur TOKA, Aydın ERAR, Meral ÇETİN	A Plug-in Estimator for the Lognormal Renewal Function under Progressively Censored Data Ömer ALTINDAĞ, Halil AYDOĞDU	Exponential Geometric Distribution: Comparing the Parameter Estimation Methods Feyza GÜNAY, Mehmet YILMAZ	The Upper Limit of Real Estate Acquisition by Foreign Real Persons And Comparison of Risk Limits in Antalya Province Alanya District Toygun ATASOY, Ayşen APAYDIN, Harun TANRIVERMİŞ
	Estimators of the Censored Regression in the Cases of Heteroscedasticity and Non-Normality Ismail YENILMEZ, Yeliz MERT KANTAR	Robust Two-way ANOVA Under Nonnormality Nuri ÇELİK, Birdal ŞENOĞLU	For Raeigly Distribution Simulation with the Help of Kendall Distribution Function Archimedean Copula Parameter Estimation Ayşe METİN KARAKAŞ, Elçin SEZGİN, Mine DOĞAN	Estimation of the Mean Value Function for Weibull Trend Renewal Process Melike Özlem KARADUMAN, Mustafa Hilmi PEKALP, Halil AYDOĞDU	Macroeconomic Determinants and Volume of Mortgage Loans in Turkey Ayşen APAYDIN, Tuğba GÜNEŞ	Comparison of MED-T and MAD-T Interval Estimators for Mean of A Positively Skewed Distributions Gözde ÖZÇIRPAN, Meltem EKİZ
	Functional Modelling of Remote Sensing Data Nihan ACAR-DENIZLI, Pedro DELICADO, Gülay BAŞARIR, Isabel CABALLERO	Linear Contrasts for Time Series Data with Non-Normal Innovations: An Application to a Real Life Data Özgecan YILDIRIM, Ceylan YOZGATLIGİL, Birdal ŞENOĞLU	HIV-1 Protease Cleavage Site Prediction Using a New Encoding Scheme Based on Physicochemical Properties Metin YANGIN, Bilge BAŞER, Ayça ÇAKMAK PEHLİVANLI	First Moment Approximations For Order Statistics From Normal Distribution Asuman YILMAZ, Mahmut KARA	Classification in Automobile Insurance Using Fuzzy c-means Algorithm Furkan BAŞER, Ayşen APAYDIN	Bayesian Estimation for the Topp-Leone Distribution Based on Type-II Censored Data İlhan USTA, Merve AKDEDE

11:30-12:30	INVITED PAPER II- Bernoulli Hall
	SESSION CHAIR: Prof. Dr. Türkan ERBAY DALKILIÇ
	Assoc. Prof. Carlos M. Agra COELHO
	Near-Exact Distributions – Problems They can Solve

CONGRESS PROGRAM

12:30:13:30	LUNCH
13:30-14:00	POSTER PRESENTATIONS

14:00-15:00	INVITED PAPER III- Bernoulli Hall
	SESSION CHAIR: Prof. Dr. Fetih YILDIRIM
	Prof. Dr. Maria Ivette GOMES
	Generalized Means and Resampling Methodologies in Statistics of Extremes

	SESSION III					
	Bernoulli Hall	Pearson Hall	Fisher Hall	Gauss Hall	Poisson Hall	Rao Hall
15:15-16:55	TIME SERIES II	DATA MINING I	APPLIED STATISTICS IV	OPERATIONAL RESEARCH I	OPERATIONAL RESEARCH II	
	TR	ENG	ENG	ENG	TR	
	SESSION CHAIR	*SESSION CHAIR*	*SESSION CHAIR*	*SESSION CHAIR*	*SESSION CHAIR*	
	Fikri ÖZTÜRK	Didem CİVELEK	Ilgım YAMAN	Esra AKDENİZ	Hülya BAYRAK	
	An Overview on Error Rates and Error Rate Estimators in Discriminant Analysis Cemal ATAKAN, Fikri ÖZTÜRK	Recommendation System based on Matrix Factorization Approach for Grocery Retail Merve AYGÜN, Didem CİVELEK, Taylan CEMGİL	A New Hybrid Method for the Training of Multiplicative Neuron Model Artificial Neural Networks Eren BAS, Erol EGRIOGLU, Ufuk YOLCU	A Robust Monte Carlo Approach for Interval-Valued Data Regression Esra AKDENİZ, Ufuk BEYAZTAŞ, Beste BEYAZTAŞ	A comparison of Goodness of Fit Tests of Rayleigh Distribution against Nakagami Distribution Deniz OZONUR, Hatice Tül Kübra AKDUR, Hülya BAYRAK	
	A New VARMA Type Approach of Multivariate Fuzzy Time Series Based on Artificial Neural Network Cem KOÇAK, Erol EĞRİOĞLU	Demand Forecasting Model for New Products in Apparel Retail Business Tufan BAYDEMİR, Dilek Tüzün AKSU	Investigation of The Insurer's Optimal Strategy: An Application on Agricultural Insurance Mustafa Asım ÖZALP, Uğur KARABEY	sNBLDA: Sparse Negative Binomial Linear Discriminant Analysis Dinçer GÖKSÜLÜK, Merve BAŞOL, Duygu AYDIN HAKLI	Generalized Entropy Optimization Methods on Leukemia Remission Times Sevda OZDEMIR, Aladdin SHAMILOV, H. Eray CELIK	
	An Application of Single Multiplicative Neuron Model Artificial Neural Network with Adaptive Weights and Biases based on Autoregressive Structure Ozge Cagcag YOLCU, Eren BAS, Erol EGRIOGLU, Ufuk YOLCU	Comparison of the Modified Generalized F-test with the Non-Parametric Alternatives Mustafa ÇAVUŞ, Berna YAZICI, Ahmet SEZER	Portfolio Selection based on a Nonlinear Neural Network: An Application on the Istanbul Stock Exchange (ISE30) Ilgım YAMAN, Türkan ERBAY DALKILIÇ	Modelling Dependence Between Claim Frequency and Claim Severity: Copula Approach Aslıhan ŞENTÜRK ACAR, Uğur KARABEY	The Province on the Basis of Deposit and Credit Efficiency (2007 – 2016) Mehmet ÖKSÜZKAYA, Murat ATAN, Sibel ATAN	
	A novel Holt's Method with Seasonal Component based on Particle Swarm Optimization Ufuk YOLCU, Erol EGRIOGLU, Eren BAS	Robustified Elastic Net Estimator for Regression and Classification Fatma Sevinç KURNAZ, Irene HOFFMANN, Peter FILZMOSER	A Novel Approach for Modelling HIV-1 Protease Cleavage Site Preferability with Epistemic Game Theory Bilge BAŞER, Metin YANGIN, Ayça ÇAKMAK PEHLİVANLI	Detection of Outliers Using Fourier Transform Ekin Can ERKUŞ, Vilda PURUTÇUOĞLU, Melih AĞRAZ	On the WABL Ddefuzzification Operator for Discrete Fuzzy Numbers Rahila ABDULLAYEVA, Resmiye NASIBOGLU	
	A New Intuitionistic High Order Fuzzy Time Series Method Erol EGRIOGLU, Ufuk YOLCU, Eren BAS	Insider Trading Fraud Detection: A Data Mining Approach Emrah BİLGİÇ, M.Fevzi ESEN	Linear Mixed Effects Modelling for Non-Gaussian Repeated Measurement Data Özgür ASAR, David BOLIN, Peter J DIGGLE, Jonas WALLIN	A perspective on analysis of loss ratio and Value at Risk under Aggregate Stop Loss Reinsurance Başak Bulut KARAGEYİK, Uğur KARABEY	Performance Comparison of the Distance Metrics in Fuzzy Clustering of Burn Images Yeşim AKBAŞ, Tolga BERBER	

16:55-17:00	Tea - Coffee Break

CONGRESS PROGRAM

	SESSION IV					
	Bernoulli Hall	Pearson Hall	Fisher Hall	Gauss Hall	Poisson Hall	Rao Hall
17:00-18:40	APPLIED STATISTICS V	APPLIED STATISTICS VI	APPLIED STATISTICS VII	OTHER STATISTICAL METHODS II	OPERATIONAL RESEARCH III	DATA MINING II
	ENG	ENG	TR	TR	ENG	ENG
	SESSION CHAIR	*SESSION CHAIR*	*SESSION CHAIR*	*SESSION CHAIR*	*SESSION CHAIR*	*SESSION CHAIR*
	Pius MARTIN	Derya KARAGÖZ	Semra ERBAŞ	Cemal ATAKAN	Rukiye DAĞALP	Furkan BAŞER
	Correspondence Analysis (CA) on Influence of Geographic Location to Children Health. Pius MARTIN, Peter JOSEPHAT	Examination Of Malignant Neoplasms And Revealing Relationships With Cigarette Consumption İrem ÜNAL, Özlem ŞENVAR	Structural Equation Modelling About the Perception of Citizens Living in Çankaya District of Ankara Province Towards the Syrian Immigrants Ali Mertcan KÖSE, Eylem DENİZ HOWE	Sorting of Decision Making Units Using Mcdm Through the Weights Obtained With Dea Emre KOÇAK, Zülal TÜZÜNER	Author Name Disambiguation Problem: A Machine Learning Approach Cihan AKSOP	The Effect of Estimation on Ewma-R Control Chart for Monitoring Linear Profiles under Non Normality Özlem TÜRKER BAYRAK, Burcu AYTAÇOĞLU
	Cluster Based Model Selection Method for Nested Logistic Regression Models Özge GÜRER, Zeynep KALAYLIOGLU	Various Ranked Set Sampling designs to construct mean charts for monitoring the skewed normal process Derya KARAGÖZ, Nursel KOYUNCU	Compare Classification Accuracy of Support Vector Machines and Decision Tree for Hepatitis Disease Ülkü ÜNSAL, Fatma Sevinç KURNAZ, Kemal TURHAN	The Health Performances of the Turkey Cities by the Mixed Integer DEA Models Zülal TÜZÜNER, H. Hasan ÖRKCÜ, Hasan BAL, Volkan Soner ÖZSOY, Emre KOÇAK	Deep Learning Optimization Algorithms for Image Recognition Derya SOYDANER	A Comparison Of Different Ridge Parameters Under Both Multicollinearity And Heteroscedasticity Volkan SEVİNÇ, Atila GÖKTAŞ
	Dependence Analysis with Normally Distributed Aggregate Claims in Stop-Loss Insurance Özenç Murat MERT, A. Sevtap SELÇUK KESTEL	Integrating Conjoint Measurement Data to ELECTRE II: Case of University Preference Problem Tutku TUNCALI YAMAN	Effectiveness of Three Factors on Classification Accuracy Duygu AYDIN HAKLI, Merve BASOL, Ebru OZTURK, Erdem KARABULUT	Efficiency and Spatial Regression Analysis Related to Illiteracy Rate Zülal TÜZÜNER, Emre KOÇAK	Faster Computation of Successive Bounds on the Group Betweenness Centrality Derya DİNLER, Mustafa Kemal TURAL	A Comparison of the Mostly Used Information Criteria for Different Degrees of Autoregressive Time Series Models Atilla GÖKTAŞ,Aytaç PEKMEZCİ, Özge AKKUŞ
	Risk Measurement Using Extreme Value Theory: The Case of BIST100 Index Bükre YILDIRIM KÜLEKCİ, A. Sevtap SELÇUK-KESTEL, Uğur KARABEY	Lmmpar: A Package For Parallel Programming In Linear Mixed Models Fulya GOKALP YAVUZ, Barret SCHLOERKE	Evaluation of the Life Index Based On Data Envelopment Analysis: Quality of Life Indexes of Turkey Volkan Soner ÖZSOY, Emre KOÇAK	Forecasting the Tourism in Tuscany with Google Trend Ahmet KOYUNCU, Monica PRATESİ	Clustering of Tree-Structured Data Objects Derya DİNLER, Mustafa Kemal TURAL, Nur Evin ÖZDEMİREL	Comparison of Partial Least Squares With Other Prediction Methods Via Generated Data Atilla GÖKTAŞ,Özge AKKUŞ, İsmail BAĞCI
			Measurement Errors Models with Dummy Variables Gökhan GÖK, Rukiye DAĞALP	A New Approach to Parameter Estimation in Nonlinear Regression Models in Case of Multicollinearity Ali ERKOÇ, M. Aydın ERAR		

CONGRESS PROGRAM

8 DECEMBER 2017 FRIDAY						
	SESSION V					
	Bernoulli Hall	Pearson Hall	Gauss Hall	Poisson Hall	Rao Hall	
09:30-11:10	FINANCE, INSURANCE AND RISK MANAGEMENT	OTHER STATISTICAL METHODS III	STATISTICS THEORY III	MODELING AND SIMULATION II	STATISTICS THEORY IV	
	ENG	TR	TR	TR	TR	
	SESSION CHAIR	*SESSION CHAIR*	*SESSION CHAIR*	*SESSION CHAIR*	*SESSION CHAIR*	
	Ceren VARDAR ACAR	Kamile ŞANLI KULA	Fikri AKDENİZ	Ali Rıza FİRUZAN	Hülya ÇINGI	
	Maximum Loss and Maximum Gain of Spectrally Negative Levy Processes Ceren Vardar ACAR, Mine ÇAĞLAR	Small Area Estımatıon Of Poverty Rate At Province Level In Turkey Gülser Pınar YILMAZ EKŞİ, Rukiye DAĞALP	Linear Bayesian Estimation in Linear Models Fikri AKDENİZ , İhsan ÜNVER, Fikri ÖZTÜRK	The Determination Of Optimal Production Of Corn Bread Using Response Surface Method And Data Envelopment Analysis Başak APAYDIN AVŞAR, Hülya BAYRAK, Meral EBEGİL, Duygu KILIÇ	Cubic Rank Transmuted Exponentiated Exponential Distribution Caner TANIŞ, Buğra SARAÇOĞLU	
	Price Level Effect in Istanbul Stock Exchange: Evidence from BIST30 Ayşegül İŞCANOĞLU ÇEKİÇ, Demet SEZER	Investigation of the CO2 Emission Performances of G20 Countries due to the Energy Consumption with Data Envelopment Analysis Esra ÖZKAN AKSU, Aslı ÇALIŞ BOYACI, Cevriye TEMEL GENCER	Alpha logarihtmic Weibull Distribution: Properties and Applications Yunus AKDOĞAN, Fatih ŞAHİN, Kadir KARAKAYA	A Classification and Regression Model for Air Passenger Flow Among Countries Tuğba ORHAN, Betül KAN KILINÇ	Detecting Change Point via Precedence Type Test Muslu Kazım KÖREZ, İsmail KINACI, Hon Keung Tony NG, Coşkun KUŞ	
	Analysis Of The Cross Correlations Between Turkish Stock Market And Developed Market Indices Havva GÜLTEKİN, Ayşegül İŞCANOĞLU ÇEKİÇ	European Union Countries and Turkey's Waste Management Performance Analysis with Malmquist Total Factor Productivity Index Ahmet KOCATÜRK, Seher BODUR, Hasan Hüseyin GÜL	Binomial-Discrete Lindley Distribution Coşkun KUŞ, Yunus AKDOĞAN, Akbar ASGHARZADEH, İsmail KINACI, Kadir KARAKAYA	On Facility Location Interval Games Osman PALANCI, Mustafa EKİCİ, Sırma Zeynep ALPARSLAN GÖK	Score Test for the Equality of Means for Several Log-Normal Distributions Mehmet ÇAKMAK, Fikri GÖKPINAR, Esra GÖKPINAR	
	Political Risk and Foreign Direct Investment in Tunisia: The Case of the Services Sector Maroua Ben GHOUL, Md. Musa KHAN	Evaluation of Statistical Regions According to Formal Education Statistics with AHP Based VIKOR Method Aslı ÇALIŞ BOYACI, Esra ÖZKAN AKSU	Asymptotic Properties of RALS-LM Cointegration Test Presence of Structural Breaks and G/ARCH Innovations Esin FİRUZAN, Berhan ÇOBAN	Measurement System Capability for Quality Improvement by Gage R&R with An Application Ali Rıza FİRUZAN, Ümit KUVVETLİ	A New Class of Exponential Regression cum Ratio Estimator in Systematic Sampling and Application on Real Air Quality Data Set Eda Gizem KOÇYİĞİT, Hülya ÇINGI	
	Bivariate Risk Aversion and Risk Premium Based on Various Utility Copula Functions Kübra DURUKAN, Emel KIZILOK KARA, H.Hasan ÖRKCÜ	On Sample Allocation Based on Coefficient of Variation and Nonlinear Cost Constraint in Stratified Random Sampling Sinem Tuğba ŞAHİN TEKİN, Yaprak Arzu ÖZDEMİR, Cenker METİN	Transmuted Complementary Exponential Power Distribution Buğra SARAÇOĞLU, Caner TANIŞ	Measuring Service Quality in Rubber-Wheeled Urban Public Transportation by Using Smart Card Boarding Data: A Case Study for Izmir Ümit KUVVETLİ, Ali Rıza FİRUZAN	Alpha Power Chen Distribution and its Properties Fatih ŞAHİN, Kadir KARAKAYA, Yunus AKDOĞAN	
	Linear and Nonlinear Market Model Specifications for Stock Markets Serdar NESLİHANOĞLU					
11:10-11:30	Tea - Coffee Break					

CONGRESS PROGRAM

11:30-12:30	INVITED PAPER IV-Bernoulli Hall SESSION CHAIR: Doç. Dr. Esra AKDENİZ Prof. Dr. Dimitrios G. KONSTANTINIDIS Asymptotic Ruin Probabilities for a Multidimensional Renewal Risk Model with Multivariate Regularly Varying Claims					
12:30:13:30	LUNCH					
13:30-14.00	POSTER PRESENTATION					
14:00-15:00	INVITED PAPER V- Bernaoulli Hall SESSION CHAIR: Prof. Dr. M. Aydın ERAR Prof. Dr. Karl-Theodor EISELE Non-Linear Hachemeister Credibility with Application to Loss Preserving					
15:00-15:15	Tea - Coffee Break					
15:15-16:15	INVITED PAPER VI- Bernoulli Hall SESSION CHAIR: Prof. Dr. Birdal ŞENOĞLU Prof. Dr. Ashis SENGUPTA Directional Statistics: Solving Challenges from Emerging Manifold Data					
16:15-16:20	Tea - Coffee Break					
16:20- 18:00	SESSION VI					
	Bernoulli Hall	Pearson Hall	Fisher Hall	Gauss Hall	Poisson Hall	Tukey Hall
	STATISTICS THEORY V	APPLIED STATISTICS VIII	OTHER STATISTICAL METHODS IV	MODELING AND SIMULATION III	OTHER STATISTICAL METHODS V	APPLIED STATISTİCS IX
	ENG	ENG	TR	ENG	TR	TR
	SESSION CHAIR	*SESSION CHAIR*	*SESSION CHAIR*	*SESSION CHAIR*	*SESSION CHAIR*	*SESSION CHAIR*
	Fatma Zehra DOĞRU	Nimet YAPICI PEHLİVAN	Hüseyin TATLIDİL	Md Musa KHAN	Nejla ÖZKAYA TURHAN	Fikri GÖKPINAR
	Robust Mixture Multivariate Regression Model based on Multivariate Skew Laplace Distribution Y. Murat BULUT, Fatma Zehra DOĞRU, Olcay ARSLAN	Intensity Estimation Methods for an Earthquake Point Pattern Cenk İÇÖZ and K. Özgür PEKER	Word problem for the Schützenberger Product Esra KIRMIZI ÇETİNALP, Eylem GÜZEL KARPUZ, Ahmet Sinan ÇEVİK	Classifying of Pension Companies Operating in Turkey with Discriminant and Multidimensional Scaling Analysis Murat KIRKAĞAÇ, Nilüfer DALKILIÇ	Demonstration Of A Computerized Adaptive Testing Application Over A Simulated Data Batuhan BAKIRARAR, İrem KAR, Derya GÖKMEN, Beyza DOĞANAY ERDOĞAN, Atilla Halil ELHAN	PLSR and PCR under Multicollinearity Hatice ŞAMKAR, Gamze GÜVEN
	Robustness Properties for Maximum Likelihood Estimators of Parameters in Exponential Power and Generalized t Distributions Mehmet Niyazi ÇANKAYA, Olcay ARSLAN	Causality Test for Multiple Regression Models Harun YONAR, Neslihan İYİT	Automata Theory and Automaticity for Some Semigroup Constructions Eylem GÜZEL KARPUZ, Esra KIRMIZI ÇETİNALP, Ahmet Sinan ÇEVİK	A Bayesian Longitudinal Circular Model and Model Selection Onur ÇAMLI, Zeynep KALAYLIOĞLU	A Comparison Of Maximum Likelihood And Expected A Posteriori Estimation In Computerized Adaptive Testing İrem KAR, Batuhan BAKIRARAR, Beyza DOĞANAY ERDOĞAN, Derya GÖKMEN, Serdal Kenan KÖSE, Atilla Halil ELHAN	On the Testing Homogeneity of Inverse Gaussian Scale Parameters Gamze GÜVEN, Esra GÖKPINAR, Fikri GÖKPINAR
	Robust Inference with a Skew t Distribution M. Qamarul ISLAM	Drought Forecasting with Time Series and Machine Learning Approaches Ozan EVKAYA, Ceylan YOZGATLIGİL, A. Sevtap SELCUK-KESTEL	The Structure of Hierarchical Linear Models and a Two-Level HLM Application Yüksel Akay ÜNVAN, Hüseyin TATLIDİL	A Computerized Adaptive Testing Platform: SmartCAT Beyza Doğanay ERDOĞAN, Derya GÖKMEN, Atilla Halil ELHAN, Umut YILDIRIM, Alan TENNANT	Some Relations Between Curvature Tensors of a Riemannian Manifold Gülhan AYAR, Pelin TEKİN, Nesip AKTAN	On An Approach to Ratio-Dependent Predator-Prey System Mustafa EKİCİ, Osman PALANCI
	Some Properties of Epsilon Skew Burr III Distribution Mehmet Niyazi ÇANKAYA, Abdullah YALÇINKAYA, Ömer ALTINDAĞ, Olcay ARSLAN	Stochastic Multi Criteria Decision Making Methods for Supplier Selection in Green Supply Chain Management Nimet YAPICI PEHLİVAN, Aynur ŞAHİN	Credit Risk Measurement Methods and a Modelling on a Sample Bank Yüksel Akay ÜNVAN, Hüseyin TATLIDİL	Educational Use of Social Networking Sites in Higher Education: A Case Study on Anadolu University Open Education System Md Musa KHAN, Zerrin AŞAN GREENACRE	Comparisons of Some Importance Measures Ahmet DEMİRALP, M. Şamil ŞIK	Analysis of Transition Probabilities Between Parties of Voter Preferences with the Ecological Regression Method Berrin GÜLTAY, Selahattin KAÇIRANLAR
	Katugampola Fractional Integrals Within the Class of s-Convex Functions Hatice YALDIZ	Parameter Estimation of Three-parameter Gamma Distribution using Particle Swarm Optimization Aynur ŞAHİN, Nimet YAPICI PEHLİVAN	A Comparison on the Ranking of Decision Making Units of Data Envelopment and Linear Discriminant Analysis Hatice ŞENER, Semra ERBAŞ, Ezgi NAZMAN	An Improved New Exponential Ratio Estimators For Population Median Using Auxilary Information In Simple Random Sampling Sibel AL, Hulya CINGI	Determining the Importance of Wind Turbine Components M. Şamil ŞIK, Ahmet DEMİRALP	Variable Neighborhood – Simulated Annealing Algorithm For Single Machine Total Weighted Tardiness Problem Sena AYDOĞAN

INVITED SPEAKERS' SESSIONS

Some Comments on Information Distortion, Statistical Error Margins and Decision Systems Interactions

Orhan GÜVENEN[1]
gorhan@bilkent.edu.tr

[1]*Department of Accounting Information Systems Bilkent University, Turkey*

Information and statistics are the raw materials of statistical inference, modeling and decision systems. The amount of information and data produced, distributed through modern communication channels are increasing exponentially. Remarkable percentage of this information and data are distorted. That leads to information distortion and statistical error margins. To minimize information distortion, statisical error margins and maximize information security, principles of heurmeneutics must be embraced. A transdisciplinary approach in education and research is required to deal with complex problems of the world. The scope of science and its structure are constantly changing and evolving. As the science progresses over the time it has to deal with more complicated issues and manage to come up with minum error margins to scientific explanations, solutions and to decision systems. To deal with sophisticated questions of high degree complexity, requires the cooperation of multiple scientific disciplines. It needs to be targeted to the problem, analyse, interpret, converge to the solutions with an iterative transdisciplinary approach which endogenize various disciplines. Equally any search for a system optimal requires that 'ethics' must remain constant in the dynamics of time and space at the individuals, institutions, corporations, nation states and international level.

10TH ISC2017 INTERNATIONAL STATISTICS CONGRESS
December 6-8, 2017 ANKARA/TURKEY

Near-Exact Distributions – Problems They Can Solve

Carlos A. COELHO[1]
cmac@fct.unl.pt

[1]*Mathematics Department – Faculdade de Ciências e Tecnologia*
Center for Mathematics and its Applications (CMA-FCT/UNL)
Universidade Nova de Lisboa, Caparica, Portugal

We are all quite familiar with the concept of asymptotic distribution. However, such asymptotic distributions quite commonly yield approximations which fall short of the precision we need and they may also exhibit some problems when the number of variables involved grows large, as it is the case of many asymptotic distributions commonly used in Multivariate Analysis. The pertinent question is thus the following one: What can we do? But before we can answer this question we need to raise one other question: are we willing to handle approximations that may have a little more elaborate structure, anyway keeping it much manageable in terms of allowing for a quite easy computation of p-values and quantiles? If our answer to this question is affirmative, then we are ready to enter the surprising world of "near-exact distributions". [1][3]
Near exact distributions are asymptotic distributions which lie much closer to the exact distribution than common asymptotic distributions. This is so because they are developed under a new concept of approximating distributions. They are based on a decomposition (usually a factorization or a split in two or more terms) of the characteristic function of the statistic being studied, or of the characteristic function of its logarithm, where we then approximate only a part of this characteristic function, leaving the remaining unchanged. [1][2][3][4][5]
If we are able to keep untouched a good part of the original structure of the exact distribution of the random variable or statistic being studied, we may in this way obtain a much better approximation, which not only does not exhibit anymore the problems referred above which occur with most asymptotic distributions, but which on top of this exhibits extremely good performances even for very small sample sizes and large numbers of variables involved, being asymptotic not only for increasing sample sizes but also (opposite to what happens with the common asymptotic distributions) for increasing values of the number of variables involved.[3][4][5]

Keywords: *asymptotic distributions, characteristic functions, likelihood ratio statistics*

References

[1] Coelho, C. A. (2004). *The Generalized Near-Integer Gamma distribution – a basis for 'near-exact' approximations to the distributions of statistics which are the product of an odd number of particular independent Beta random variables*. Journal of Multivariate Analysis, 89, 191-218.

[2] Coelho, C. A., Arnold, B. C. (2014). *On the exact and near-exact distributions of the product of generalized Gamma random variables and the generalized variance*, Communications in Statistics – Theory and Methods, 43, 2007–2033.

[3] Coelho, C. A., Marques, F. J. (2010) *Near-exact distributions for the independence and sphericity likelihood ratio test statistics*. Journal of Multivariate Analysis, 101, 583-593.

[4] Coelho, C. A., Marques, F. J., Arnold, B. C. (2015). *The exact and near-exact distributions of the main likelihood ratio test statistics used in the complex multivariate normal setting*, Test, 24, 386–416.

[5] Coelho, C. A., Roy, A. (2017). *Testing the hypothesis of a block compound symmetric covariance matrix for elliptically contoured distributions*, Test, 26, 308–330.

Generalized Means and Resampling Methodologies in Statistics of Extremes

M. Ivette GOMES[1]
ivette.gomes@fc.ul.pt

[1]*DEIO and CEAUL, Universidade de Lisboa, Lisboa, Portugal*

Most of the estimators of parameters of rare events, among which we distinguish the *extreme value index* (EVI), the primary parameter in statistical extreme value theory, are averages of adequate statistics V_{ik}, $1 \le i \le k$, based on the k upper or lower ordered observations associated with a stationary weakly dependent sample from a parent $F(.)$. Those averages can be regarded as the logarithm of the geometric mean (or Holder's mean-of-order-0) of $U_{ik} := \exp(V_{ik})$, $1 \le i \le k$. It is thus sensible to ask how much Holder's mean-of-order-p is able to improve the EVI-estimation, as performed by [1], among others, for $p \ge 0$, and by [2] for any real p. And new classes of reliable EVI-estimators based on other adequate generalized means, like Lehmer's mean-of-order-p, have recently appeared in the literature (see [5]), and will be introduced and discussed. The asymptotic behavior of the aforementioned classes of EVI-estimators enables their asymptotic comparison at optimal levels (k, p), in the sense of minimal mean square error. Again, a high variance for small k and a high bias for large k appear, and thus the need for bias-reduction and/or an adequate choice of k. Resampling methodologies, like the jackknife and the bootstrap (see, among others, [3] and [4]) are thus important tools for a reliable semi-parametric estimation of the EVI and will be discussed.

Keywords: *Bootstrap, generalized jackknife, generalized means, heavy tails, semi-parametric estimation.*

References

[1] Brilhante, F., Gomes, M.I. and Pestana, D. (2013), A simple generalization of the Hill estimator. *Computational Statistics & Data Analysis* **57**:1, 518-535.

[2] Caeiro, F., Gomes, M.I., Beirlant, J. and de Wet, T. (2016), Mean-of-order-p reduced-bias extreme value index estimation under a third-order framework. *Extremes* **19**:4, 561-589.

[3] Gomes, M.I., Caeiro, F., Henriques-Rodrigues, L. and Manjunath, B.G. (2016), Bootstrap methods in statistics of extremes. In F. Longin (ed.), *Extreme Events in Finance: A Handbook of Extreme Value Theory and its Applications.* John Wiley & Sons, Chapter 6, 117-138.

[4] Gomes, M.I., Figueiredo, F., Martins, M.J. and Neves, M.M. (2015), Resampling methodologies and reliable tail estimation. *South African Statistical Journal* **49**, 1-20.

[5] Penalva, H., Caeiro, F., Gomes, M.I. and Neves, M. (2016), *An Efficient Naive Generalization of the Hill Estimator—Discrepancy between Asymptotic and Finite Sample Behaviour.* Notas e Comunicações CEAUL 02/2016. Available at: http://www.ceaul.fc.ul.pt/notas.html?ano=2016

Asymptotic Ruin Probabilities for a Multidimensional Renewal Risk Model with Multivariate Regularly Varying Claims

Dimitrios G. KONSTANTINIDES[1], Jinzhu LI[2]
konstant@aegean.gr, lijinzhu@nankai.edu.cn

[1]*Department of Mathematics University of the Aegean, Karlovassi, Greece*
[2]*School of Mathematical Science and LPMC Nankai University, Tianjin, P.R. China*

This paper studies a continuous-time multidimensional risk model with constant force of interest and dependence structures among random factors involved. The model allows a general dependence among the claim-number processes from different insurance businesses. Moreover, we utilize the framework of multivariate regular variation to describe the dependence and heavy-tailed nature of the claim sizes. Some precise asymptotic expansions are derived for both _nite- time and in_nite-time ruin probabilities.

Keywords: *asymptotics; multidimensional renewal risk model; multivariate regular variation; ruin probability*

Non-Linear Hachemeister Credibility with Application to Loss Preserving

Karl-Theodor EISELE[1]
eisele@unistra.fr

[1]*Université De Strasbourg Laboratoire De Recherche En Gestion Et Économie Institut De Recherche Mathématique Avancée, Strasbourg Cedex, France*

We present a specific non-linear version of Hachemeister's hierarchical credibility theory. This theory is applied to a multivariate model for loss prediction with several contracts for each accident year. The basic model assumption starts from the idea that there exists a relatively small number of characteristic development patterns as ratios of the loss payments, and that these patterns are independent of the final amount of the claims. In non-linear hierarchical credibility theory, the estimation of the parameters of the coupled variables is tricky task, even when the latter are stochastically independent. Interdependent pseudo-estimators show up which can be resolved by an iteration procedure. The characteristic development patterns are found by an application of the well-known clustering method of k means, where the number k of clusters is chosen by the Bayesian information criterion (BIC). Once an estimation of the development pattern is found for each claim, the final claim amount can be easily estimated.

Directional Statistics: Solving Challenges from Emerging Manifold Data

Ashis SenGupta[1]
amsseng@gmail.com

[1]*Applied Statistics Unit, Indian Statistical Institute, Kolkata*

In this era of complex data problems from multidisciplinary research, statistical analysis for data on manifolds has become indispensable. The emergence of Directional Statistics (DS) for the analysis of Directional Data (DD) has been a key ingredient for analysis of such data as were not encompassed by the previously existing statistical methods. The growth of DS has been phenomenal over the last two decades. DD refer to observations on angular propagation, orientation, displacement, etc. Data on periodic occurrences can also be cast in the arena of DD. Analysis of such data sets differs markedly from those for linear ones due to the disparate topologies between the line and the circle. Misuse of linear methods to analyze DD, as seen in several areas, is alarming and can lead to dire consequences. First, methods of construction of probability distributions on manifolds such as circle, torus, sphere, cylinder, etc. for DD are presented. Then it is shown how statistical procedures can be developed to meet challenges of drawing sensible inference for such data as arising in a variety of applied sciences, e.g. from Astrostatistics, Bioinformatics, Defence Science, Econometrics, Geoscience, etc. and can enhance such work for the usefulness of our society.

Keywords: *Directional data analysis, Cylindrical distribution, Statistical inference*

SESSION I

STATISTICS THEORY I

A Genetic Algorithm Approach for Parameter Estimation of Mixture of Two Weibull Distributions

Muhammet Burak KILIÇ[1], Yusuf ŞAHİN[1], Melih Burak KOCA[1]
mburak@mehmetakif.edu.tr, ysahin@mehmetakif.edu.tr, mbkoca@mehmetakif.edu.tr

[1]*Mehmet Akif Ersoy University, Department of Business Administration, Burdur, Turkey*

A mixture of two Weibull distributions has variety of usage area from reliability analysis to wind speed modelling [1,3]. The existing conventional methods such as Maximum likelihood (ML) and Expectation-Maximization (EM) algorithm for estimating the parameters of the mixture of two Weibull distributions are very sensitive to initial values. In other words, the efficiency of the estimation highly depends on the initial values. The aim of this paper is to present a Genetic Algorithm (GA), which is a class of evolutionary algorithms proposed by [2] and needs a set of initial solutions instead of initial values for parameter estimation. This paper also presents a comparison for parameter estimations of the mixture of two Weibull distributions obtained by three computational methods: ML via Newton Raphson method, EM and the proposed GA respectively. The bias and root mean square error (RMSE) are used as decision criteria for the comparison of the estimations via Monte Carlo simulations. Results of the simulation experiment present the superiority of GA in terms of efficiency. The GA approach is also illustrated through life and wind speed data examples and compared with existing methods in the literature.

Keywords: *Mixture of two Weibull distributions, Genetic Algorithm, Monte Carlo Simulations*

References

[1] Carta, J.A. and Ramirez, B. (2007), *Analysis of two-component mixture Weibull statistics for estimation of wind speed distributions*, Renewable Energy, 32, 518-531.

[2] Holland, J.H. (1975), *Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control and artificial intelligence*, USA, University of Michigan Press.

[3] Karakoca, A., Erisoglu, U. and Erisoglu, M. (2015), *A comparison of the parameter estimation methods for bimodal mixture Weibull distribution with complete data*, Journal of Applied Statistics, 42, 1472-1489.

10TH ISC2017 INTERNATIONAL STATISTICS CONGRESS
December 6-8, 2017 ANKARA/TURKEY

Recurrent Fuzzy Regression Functions Approach based on IID Innovations Bootstrap with Rejection Sampling

Ali Zafer DALAR[1], Eren BAS[1], Erol EGRIOGLU[1], Ufuk YOLCU[2], Ozge CAGCAG YOLCU[3]
alizaferdalar@hotmail.com,eren6061@hotmail.com, erole1977@yahoo.com, varyansx@hotmail.com, ozgecagcag@yahoo.com

[1]*Giresun University, Department of Statistics, Forecast Research Laboratory, Giresun, Turkey*
[2]*Giresun University, Department of Econometrics, Forecast Research Laboratory, Giresun, Turkey*
[3]*Giresun University, Department of Industrial Engineering, Forecast Research Laboratory, Giresun, Turkey*

Fuzzy regression functions (FRF) approaches are tools used for the purpose of forecasting. FRF approaches are data based methods, and they can solve complex nonlinear real world time series data sets. Inputs of the FRF approaches are lagged variables of time series if it is used for forecasting. Moreover, there is no any probabilistic inference in the system, and it ignores random sampling variations. In this study, a new recurrent FRF approach are proposed based on IID innovations bootstrap with rejection sampling. The new method is called bootstrapped recurrent FRF (B-RFRF). B-RFRF is a recurrent system, because lagged variables of residual series are given as inputs to the systems as well as lagged variables of time series. The artificial bee colony algorithm is used to estimate the parameters of system. The probabilistic inference is made by using IID innovations bootstrap with rejection sampling. The bootstrap forecasts, bootstrap confidence intervals, and standard errors of forecasts can be calculated from bootstrap samplings. The proposed method is compared with others by using stock exchange data sets.

Keywords: *forecasting, fuzzy sets, fuzzy inference systems, bootstrap methods, artificial bee colony*

References

[1] Efron, B. and Tibshirani, R. J. (1993), *An Introduction to Bootstrap*, USA, CRC Press.
[2] Karaboga, D. (2010), *Artificial bee colony algorithm*, Scholarpedia, 5(3), 6915.
[3] Turksen, I. B. (2008), *Fuzzy Functions with LSE*, Applied Soft Computing, 8(3), 1178-1188.

An Infrastructural Approach to Spatial Autocorrelation

Ahmet Furkan EMREHAN[1], Dogan YILDIZ[1]
matahmet@gmail.com, dyildizyildiz.edu.tr

[1]*Yildiz Technical University, Istanbul, TURKEY*

As is known, Spatial Autocorrelation is a useful measure to detect the degree of spatial dependency over units in a region. Spatial Autocorrelation can be computed in many ways, like Moran's I and Geary's c. Beyond these statistics, it is an incontrovertible fact that spatial weighting plays an important role for computation of Spatial Autocorrelation Statistics [1]. However it is obvious that many studies in Spatial Autocorrelation literature have tendency to use Standard Spatial Contiguity Weights based on geometric for boundary based models. But geographical objects cannot be confined in standard geometric structures. Because Standard Spatial Contiguity Weighting may be sufficient to make model representing actual phenomenon including man-made infrastructure. In this study, Differentiation in Moran's I, generated by Various Spatial Weightings possessing road property as an infrastructural approach and standard contiguity, for boundary based model, is examined. Provincial Data provided by TUIK is used for application of this study. The results of that differentiation in global and local scale are to be discussed .

Keywords: *Spatial Analysis, Global Spatial Autocorrelation, Spatial Weightings, Moran's I, Provincial Data*

References

[1] Cliff, A.D. and Ord, J.K. (1969), *The Problem of Spatial Autocorrelation*, London Papers in Regional Science 1, Studies in Regional Science, London:Pion, Pg 25-55.

A Miscalculated Statistic Presented as an Evidence in a Case and Its Aftermath

Mustafa Y. ATA
mya@akademikidea.net

akademikidea Community, Ankara,Turkey

Sally Clark was convicted and given two life sentences in November 1999 for she was found guilty of the murder of her two elder sons. However, she and her family never accepted this criminal charge and earnestly continued to defend the innocence of the mother. Their argument was based on that the jury had found her guilty on a miscalculated probability presented to the court as an evidence by Sir Roy Meadow who were then a highly respected expert in field of child abuse, and Emeritus Professor of Paediatrics. The convictions were upheld on appeal in October 2000, but overturned in a second appeal in January 2003. Sally was released from prison having served more than three years of her sentence, but with having developed serious psychiatric problems and died in March 2007 from alcohol poisoning at an age of 43.[1]

After one year than the first appeal, the Royal Statistical Society issued a statement, in October 2001, arguing that there was "no statistical basis" for Meadow's claim, and expressing its concern at the "misuse of statistics in the courts"[2], [3]. Sally's release in January 2003 prompted the Attorney General to order a review of hundreds of other cases resulting in overturning of three similar convictions in which expert witness Meadow had testified about the unlikelihood of coth deaths more than one in a single family.

In this presentation, lessons drawn and achievements to date for each actor in the Sally's tradegy will be discussed.

Keywords: *statistical evidence, statistical literacy, conditional probability, prosecuter's fallacy*

References

[1] Sally Clark: Home Page, http://www.sallyclark.org.uk/. Accessed on Nov. 23rd of 2017.

[2] Royal Statistical Society Statement regarding statistical issues in the Sally Clark case (News Release, 23 October 2001), "Royal Statistical Society concerned by issues raised in Sally Clark case". http://www.rss.org.uk/Images/PDF/influencing-change/2017/SallyClarkRSSstatement2001.pdf, Retrieved on Nov. 23rd of 2017.

[3] Royal Statistical Society Letter from the President to the Lord Chancellor regarding the use of statistical evidence in court cases, (Jan. 23rd 0f 2002) http://www.rss.org.uk/Images/PDF/influencing-change/rss-use-statistical-evidence-court-cases-2002.pdf, Retrieved on Nov. 23rd of 2017.

Estimation of Variance Components in Gage Repeatibility & Reproducibility Studies

Zeliha DİNDAŞ[1] , Serpil AKTAŞ ALTUNAY[2]
zeliha.dindas@sanayi.gov.tr spxl@hacettepe.edu.tr

[1]*Ministry of Science, Industry and Technology, Ankara, Turkey*
[2] *Hacettepe University, Department of Statistics, Ankara, Turkey*

Quality Control which plays an important role in the production process, is one of the tools necessary for companies to increase the quality of their products and services and to meet the expectations of their customers. If quality control is done effectively, quality control provides high levels of productivity and savings in expenses. Contribution to the production process can be achieved by using a quality control system based on a standard such as ISO 9001 published by the International Standards Organization (ISO). In this regard, Gage Repeatability & Reproducibility analysis is a part of the Measurement System Analysis (MSA). Generally, Gage Repeatability & Reproducibility studies are preferred at the beginning of the process in order to determine whether the devices are measuring correctly and to improve the manufacturing process of various companies. For this reason, how to obtain the measurement quality is important for those who will apply quality control. In this study, using the ANOVA, Maximum Likelihood Estimation (ML), Restricted Maximum Likelihood Estimation (REML) and Minimum Norm Quadratic Estimation (MINQUE) methods, how to apply these estimates to the Measurement Systems Analysis (MSA) is discussed. Besides, the advantages and disadvantages of these methods are discussed. Various numerical examples related to the MSA are analysed and the methods are compared by estimating the variance components by different methods.

Keywords: *ANOVA, ML, REML, MINQUE, Measuring System Analysis, Gauge Repeatibility&Reproducibility*

References

[1] Montgomery, D. C., Runger, G. C., *Gauge Capability Analysis and Designed Experiments. Part I: Basic Methods, Qual. Eng.*, 6, 115-135, 1993.

[2] Montgomery, D. C., Runger, G. C., *Gauge Capability Analysis and Designed Experiment, Part II: Experimental Design Models and Variance Component Estimation, Quality Engineering*, 6, 2, 289-305.1993.

[3] Montgomery, D.C., *Statistical Quality Control: A Modern Introduction, sixth ed.*, Wiley, New York, 2009.

[4] Searle, S.R., Casella, G., McCulloch, C.E., *Variance Components*, Wiley, New York. 1992.

[5] Rao, C. R., *Estimation of variance and covariance components MINQUE theory*, J. Multi. Anal., 3, 257-275, 1971.

SESSION I

APPLIED STATISTICS I

10TH ISC2017 INTERNATIONAL STATISTICS CONGRESS
December 6-8, 2017 ANKARA/TURKEY

Investigation of Text Mining Methods on Turkish Text

Ezgi PASİN[1], Sedat ÇAPAR[2]
epasin0@gmail.com, sedat.capar@deu.edu.tr

[1]*The Graduate School of Natural and Applied Science, Department of Statistics, Dokuz Eylül University, İzmir, Turkey*
[2] *Faculty of Science, Department of Statistics, Dokuz Eylül University, İzmir, Turkey*

With the widespread use of the Internet, non-structural data in the virtual environment has increased the amount of data. With increasing amounts of data to analyze and discover valuable information is difficult. In order to analyze such non-structural data, the concept of Text Mining, which is known as the sub-study area of Data Mining, has been defined.

Text mining is a general term used for methods that provide meaningful information from text sources. Social media, which has been rising in 2000 and increasing in use in recent years, has become the most widely used medium of text mining, both as a communication tool and as an information sharing medium.

Text categorization methods are used in order to get the information from the databases which includes text type data. With the increase of the number of the number of documents, classfication has been being made automatically. For this purpose, with the help of the keywords of which categories are determined firstly, text type data can be classfied.

In this study, the texts are classified. In order to work on text classification, news is used as a set of Turkish data.

Keywords: *data mining, text mining, unstructured data, text categorization*

References

[1] Pilavcılar, İ.F. (2007), *Metin Madenciliği ile Metin Sınıflandırma*, Yıldız Teknik University, Pages 6-13.

[2] Weiss, S.M., Indurkhya, N. and Zhank, T. (2010), *Fundamentals of Predictive Text Mining*, London, Springer, Pages 1-9.

[3] Ronen, F. and Sanger, J. (2007), *The Text Mining Handbook: Advenced Approaches in Analyzing Unstructured Data*, Cambridge University Press, U.S.A., Pages 82-92

[4] Oğuz, B. (2009), *Metin Madenciliği Teknikleri Kullanılarak Kulak Burun Boğaz Hasta Bilgi Formlarının Analizi*, Akdeniz University, Pages 7-17.

[5] Karaca, M.F. (2012), *Metin Madenciliği Yöntemi ile Haber Sitelerindeki Köşe Yazılarının Sınıflandırılması,* Karabük University, Pages 14-22.

10TH ISC2017 INTERNATIONAL STATISTICS CONGRESS
December 6-8, 2017 ANKARA/TURKEY

Cost Analysis of Modified Block Replacement Policies in Continuous Time

Pelin TOKTAŞ[1], Vladimir V. ANISIMOV[2]
ptoktas@baskent.edu.tr, vladimir.v.anisimov@gmail.com

[1]*Başkent University, Department of Industrial Engineering, Ankara, Turkey*
[2]*AVZ Statistics Ltd, London, United Kingdom*

Various studies on maintenance policies for the systems having random failures are conducted by many researchers for years. These models can be applied to many areas such as industry, military and health. The systems become more complex with technological developments. Therefore, new technologies, control policies and methodologies are needed. Planning activities to ensure that the components of a system are working is important. Some decisions concerning replacement, repair and inspection are made in the study of maintenance policies.

Replacement decision making involves the problem of specifying a replacement policy which balances the cost of failures of a unit during operation against the cost of planned replacements. One of the most widely used replacement policy in the literature is block replacement. Under block replacement, the system is replaced upon at failure and at times $jT, j = 1,2, ...$ [4].

In this study, cost analysis of three modified multi-component block replacement models (total control, partial control and cyclic control) are considered in continuous time. In all models, there are N components which are subject to random failures. Each failed component is changed with probability α. Replacements are allowed only at times $jT, j = 1, 2, ...$ and $T > 0$ is fixed. The long-run expected cost per unit of time and optimal replacement interval $T^*$are calculated for each model and then model comparisons are made based the long-run expected cost per unit of time.

Keywords: *Cost analysis of replacement policies, block replacement, total control, partial control, cyclic control.*

References

[1] Anisimov V. V. (2005), *Asymptotic Analysis of Stochastic Block Replacement Policies for Multicomponent Systems in a Markov Environment*, Operation Research Letters, 33, s. 26-34.

[2] Anisimov V. V., Gürler Ü. (2003), *An Approximate Analytical Method of Analysis of a Threshold Maintenance Policy for a Multiphase Multicomponent Model*, Cybernetics and Systems Analysis, 39(3), s. 325-337.

[3] Barlow R. E., Hunter L. C. (1960), *Optimum Preventive Maintenance Policies*, Operations Research, 8, s. 90-100.

[4] Barlow R. E., Proschan F. (1996), *Mathematical Theory of Reliability*, SIAM edition of the work first published by John Wiley and Sons Inc., New York 1965.

Examination of The Quality of Life of OECD Countries

Ebru GÜNDOĞAN AŞIK[1], Arzu ALTIN YAVUZ [2]
ebrugundogan@ktu.edu.tr, aaltin@ogu.edu.tr

[1]*Karadeniz Teknik Üniversitesi, İstatistik ve Bilgisayar Bilimleri Bölümü, Trabzon, Türkiye*
[2]*Eskişehir Osmangazi Üniversitesi, İstatistik Bölümü, Eskişehir, Türkiye*

The quality of life index is an index used to measure the quality of life of countries. While this index value is calculated, countries are assessed in terms of multivariate features In recent years, in order to determine the quality of life of a country, a new index was established that includes not only GDP but also variables such as health, education, work life, politics, social relations, environment and trust. While determining the quality of life with so many variables, some subindex values are also calculated. A subindex that constitutes a better quality of life index is the life satisfaction index.

In this study, a classification mechanism has been established with the help of other subindex values constituting the quality of life, taking into account the life satisfaction index values. The validity and reliability of the results obtained in the research are closely related to the use of accurate scientific methods. Various classification methods that can be applied depending on data structure are discussed in the study. Logistic regression, robust logistic regression and robust logistic ridge regression analyzes were used to analyze data and correct classification ratios were calculated. With the help of the correct classification ratios, methods are compared and the most appropriate method for data structure is proposed.

Keywords: *Quality of Life, Logistic Regression, Robust Logistics, Ridge Regression*

References

[1] Akar, S.(2014), *Türkiye'de Daha İyi Yaşam İndeksi: OECD Ülkeleri İle Karşılaştırma*, Journal of Economic Life, 1-12.

[2] Bianco, A., and Yohai, V. (1996), *Robust Estimation in the logistic regression model*, Springer.

[3] Durand, M. (2015), *The OECD Better Life Initiative: How's Life And The Measurement Of Well-Being*, Review of Income and Wealth, 61(1), 4-17.

[4] Hobza, T., Pardo, L., and Vajda, I. (2012), *Robust median estimator for generalized linear models with binary responses*, Kybernetika, 48(4), 768-794.

[5] Hoerl, A. E., and Kennard, R. W. (1970*), Ridge regression, Biased estimation for nonorthogonal problems*, Technomctrics, 12, 69-82.

Multicollinearity with Measurement Error

Şahika GÖKMEN[1], Rukiye DAĞALP[2], Serdar KILIÇKAPLAN[1]
sahikagokmen@gazi.edu.tr , rdagalp@ankara.edu.tr , skilickaplan@yahoo.com

[1]*Gazi University, Ankara, Turkey*
[2] *Ankara University, Ankara, Turkey*

Multicollinearity is a linear relationship between explanatory variables in a regression model. In this case, the unbiased parameter estimates of the regression model are not affected. However, the effectiveness of predictor is affected. Additionally, the least squares estimator has the smallest variance [1]. This is a problem, especially when there is a need for a statistically meaningful model. Because the variances of the estimators are predicted larger and this leads to misleading results in the case of Type I errors. Thus, the parameters that are really statistically significant can be seen as not statistically significant. On the other hand, the explanatory variable (s) of the model with a measurement error, which leads to more serious problems than the multicollinearity problem. Presence of any measurement error in the explanatory variables leads to bias estimation of the parameters and the attenuated regression line. Nowadays, studies on estimation methods of measurement error models are increasing, but the issue of multicollinearity with measurement error has not been worked in the literature at all. Accordingly, how the measurement error affects the multicollinearity will be investigated in this study. For this purpose, most commonly used methods as VIF (Variance Inflation Factor), Tolerance Factor and Condition Index are taken into consideration for detecting multicollinearity and its behaviors against different measurement errors will be examined through simulation studies.

Keywords: *measurement error, multicollinearity, simulation, vif, conditional index*

References

[1] Greene, W. H., 2012, *Econometric Analysis*, England, Pearson Education Limited, 279-282.

[2] Buonaccorsi, J. P., 2010, *Measurement Error: Models Methods and Applications*, USA, Chapman&Hall/CRC, 143-154.

[3] Fuller, W.A. (1987), *Measurement Error Models*, John Wiley and Sons. New York.

10TH ISC2017 INTERNATIONAL STATISTICS CONGRESS
December 6-8, 2017 ANKARA/TURKEY

The Effect of Choosing the Sample on the Estimator in Pareto Distribution

Seval ŞAHİN[1], Fahrettin ÖZBEY[1]
sevalsahin89@gmail.com, fozbey2023@gmail.com

[1] *Bitlis Eren University, Department of Statistics Bitlis, Türkiye*

In this study, firstly methods of generating samples from a certain distribution were given[1-4]. Then new methods to generate samples from a certain distribution were developed. As a result; old and new methods to produce samples from the Pareto distribution were used. Using these samples, parameter estimates with the maximum likelihood method were made. These estimated parameters were compared with the parameters used to construct the sample. Better result with the samples generated by the new method were obtained.

Keywords: *Pareto distribution, Estimator, Sample*

References

[1] Bratley, P. Fox, B. L. and Schrage. L. E. (1987), *A Guide to Simulation*, New York, Springer-Verlang.
[2] Çıngı, H. (1990), *Örnekleme Kuramı*, Ankara, Hacettepe Üniversitesi Fen Fakültesi Basımevi.
[3] Öztürk, F. and Özbek, L. (2004), *Matematiksel Modelleme ve Simülasyon,* Ankara, Gazi Kitabevi.
[4] Shahbazov, A. (2005), *Olasılık Teorisine Giriş*, İstanbul, Birsen Yayınevi.

ISC2017 INTERNATIONAL STATISTICS CONGRESS
December 6-8, 2017 ANKARA/TURKEY

Application of Fuzzy c-means Clustering Algorithm for Prediction of Students' Academic Performance

Furkan BAŞER[1], Ayşen APAYDIN[1], Ömer KUTLU[2], M. Cem BABADOĞAN[2], Hatice CANSEVER[3], Özge ALTINTAŞ[2], Tuğba KUNDUROĞLU AKAR[2]
furkan.baser@ankara.edu.tr, aysen.apaydin@ankara.edu.tr, omerkutlu@ankara.edu.tr, cem@ankara.edu.tr, hatice.cansever@ankara.edu.tr, oaltintas@ankara.edu.tr, tkunduroglu@education.ankara.edu.tr

[1]*Faculty of Applied Sciences, Ankara University, Ankara, Turkey*
[2]*Faculty of Educational Sciences, Ankara University, Ankara, Turkey*
[3]*Student Affairs Department, Ankara University, Ankara, Turkey*

Nowadays, the amount of data stored in educational database is rapidly increasing. These databases contain some information to improve the performance of students, which is influenced by many factors. Therefore, it is essential to develop a classification system so as to identify the difference between students (Oyelade et al., 2010).

The main purpose of clustering is to find out the classification structure of the data. Clustering algorithms based on its structure are generally divided into two types: fuzzy and non-fuzzy (crisp) clustering (Gokten et al., 2017). Fuzzy clustering methods are used for calculating the membership function that determines to which degree the objects belong to clusters and used for detecting overlapping clusters in the data set (De Oliveira and Pedrycz, 2007).

The aim of this study is to illustrate the use of a fuzzy c-means (FCM) clustering approach for application to the grouping of students into different clusters according to various factors. Utilizing a set of records for students who were registered at Ankara University in the academic year 2014 – 2015, it was determined that FCM clustering method gives remarkable results.

Keywords: *academic performance, classification, fuzzy c-means*

References

[1] De Oliveira, J.V. and Pedrycz, W. (2007), *Advances in fuzzy clustering and its applications*, West Sussex, Wiley.

[2] Gokten, P. O., Baser, F., and Gokten, S. (2017). *Using fuzzy c-means clustering algorithm in financial health scoring*. The Audit Financiar journal, 15(147), 385-385.

[3] Oyelade, O. J., Oladipupo, O. O., and Obagbuwa, I. C. (2010), *Application of k-means clustering algorithm for prediction of Students Academic Performance*, International Journal of Computer Science and Information Security, 7(1), 292-295.

SESSION I

ACTUARIAL SCIENCES

Mining Sequential Patterns in Smart Farming using Spark

Duygu Nazife ZARALI[1], Hacer KARACAN[1]
duygusonmez@gazi.edu.tr, hkaracan@gazi.edu.tr

[1]*Gazi University Computer Engineering, Ankara, Turkey*

Smart Farming is a development that emphasizes the use of information and communication technology in farm management. Robots and artificial intelligence are expected to be used more in agriculture. Robotic milking systems are new technologies that reduce the labour of dairy farming and the need for human–animal interactions. Increasing the use of smart machines and sensors on farms increases the amount and scope of farm data. Thus, agriculture processes are increasingly data-driven and data will become more effective. Big Data is used to provide predictive information and to make operational decisions in agricultural operations. [2,3].
In this study, by integrating sequential pattern mining algorithms with a distributed data processing engine Spark, which is an effective cluster computing system that makes data processing easier and faster. A well-known data mining algorithm aiming to find sequential pattern, The PrefixSpan [1], is used to extract patterns from a private dataset. This dataset is obtained from an R&D company for the automation of milking, feeding and cleaning robots used in modern dairy farms. Robots working on farms give various alarms to warn and inform the user. These alarms, which are collected in a centralized system, can be critical alarms that stop the robot operation and important process in the farm, as well as simple warning indications with low urgency level. Sometimes the same alarms generated by robots can be sent back to the farmer repeatedly because there is no intelligent mechanism to prioritize alarms or identify the relationships among them. Therefore, this large data traffic is exhausting the system and the farmer. In this study, past alarm information is analyzed, related alarms and patterns are determined. Alarms and indications are analyzed on a daily basis. Analysis of the 15-day alarm series data took 3.28 seconds with 0,9 minimum support. As a result of the study, it is planned that the actual sources of the alarms can be predicted and the possible problems that may arise based on the past alarm data can be eliminated. With this analysis, it will be possible to minimize significantly costs by early detection of failures that may occur in systems and management of maintenance processes accordingly.

Keywords: *Data Mining, Sequential Pattern Mining, PrefixSpan, Spark, Big Data*

References

[1] Han, J., Pei, J., Mortazavi-Asl, B., Pinto, H., Chen, Q., Dayal, U., and Hsu, M. C. (2001). *Prefixspan: Mining sequential patterns efficiently by prefix-projected pattern growth*. Proceedings of the 17th international conference on data engineering, 215-224.

[2] Holloway, L., Bear, C., & Wilkinson, K. (2014). *Robotic milking technologies and renegotiating situated ethical relationships on UK dairy farms*. Agriculture and human values, 31(2), 185-199.

[3] Wolfert, S., Ge, L., Verdouw, C., & Bogaardt, M.-J. (2017). *Big Data in Smart Farming–A review*. Agricultural Systems, 153, 69-80.

Multivariate Markov Chain Model: An Application to S&P500 and Ftse-100 Stock Exchanges

Murat GÜL[1], Ersoy ÖZ[2]
murat.gul@giresun.edu.tr, ersoyoz@yıldız.edu.tr

[1] *Giresun University, Faculty of Arts and Sciences, Department of Statistics, Giresun, Turkey*
[2] *Yıldız Teknik University, Faculty of Arts and Sciences, Department of Statistics, İstanbul,Turkey*

Markov chains are the stochastic processes that have many application areas. The data that belong to the system being analyzed in the Markov chains come from a single source. The multivariate Markov chain model is a model that is used for the purpose of showing the behaviour of multivariate categorical data sequences produced from the same source or a similar source. In this study we explain the multivariate Markov chain model that is based on the Markov chains from a theoretical standpoint in detail. As for an application, we take on the daily changes that occur in the S&P-500 Index in which the shares of the 500 greatest companies of the United States of America are traded and the daily changes that occur in the UK FTSE 100 Index as two categorical sequences. And we display the proportions that show how much they influence each other via a multivariate Markov chain model.

Keywords: *Markov Chain, Categorical Data Sequences, Multivariate Markov Chain.*

References

[1] Ching W., Fung Eric S. and NG Michael K. (2002), *A Multivariate Markov Chain Model for Categorical Data Sequences and Its Applications in Demand Predictions*, IMA Journal of Management Mathematics, Vol. 13, pp. 187-199.

[2] Ching W., Li L, LI T. and Zhang S. (2007), *A New Multivariate Markov Chain Model with Applications to Sales Demand Forecasting,* International Conference on Industrial Engineering and Systems Management IESM 2007, Beijing – China, May 30-June 2-, pp. 1-8.

[3] Ching W. and NG Michael K. (2006), *Markov Chains: Models, Algorithms and Applications***,** United States of America, Springer Science+Business Media, Inc., 2006

[4] Ross S. (1996), *Stochastic Processes*, Second Edition, New York: John Wiley & Sons Inc.

10TH ISC2017 INTERNATIONAL STATISTICS CONGRESS
December 6-8, 2017 ANKARA/TURKEY

Use of Haralick Features for the Classification of Skin Burn Images and Performance Comparison of k-Means and SLIC Methods

Erdinç KARAKULLUKÇU[1], Uğur ŞEVİK[1]
erdinc.karakullukcu@ktu.edu.tr, usevik@ktu.edu.tr

[1]*Department of Statistics and Computer Sciences, Karadeniz Technical University, Trabzon, Turkey*

Burn injuries require an immediate treatment. However, finding a burn specialist in health centers in rural areas is generally not possible. A solution to deal with the burn injuries is the use of computer aided systems. Color images taken by digital cameras are used as input data. First, the burn color image is segmented, then the segmented parts are classified as skin, burn or background, and finally the depth of the burn is tried to be predicted. The first goal of this work is to extract Haralick and statistical histogram features to train some well-known classification methods to be able to find the best model to classify the skin, burn, and background textures. The second goal is to use this classification model on 7 test images that are segmented by using k-means and simple linear iterative clustering (SLIC) methods.

The proposed system in this work started with the classification process. Texture information was obtained from RGB and LAB color spaces of burn images. Texture was defined by using 13 Haralick features and 7 statistical histogram features. For each texture, 28 gray level co-occurrence matrices (calculated at 0, 45, 90, and 135 degrees) were generated on R, G, B, L, A, B and gray channels, and a total number of 364 Haralick features were extracted from these matrices. Moreover, 49 statistical histogram features were obtained from each texture. 100x100 pixel sized skin, burn, and background textures were randomly sampled from 57 prelabeled burn images. For each class, 600 samples were collected. Well-known supervised pattern classifier methods were trained by using the extracted features. Artificial neural networks obtained the best micro and macro averaged F1 scores (92.02 % and 92.05 %, respectively) to classify the texture images as skin, burn and background. A forward selection algorithm was performed using the artificial neural network classifier. 0.84 % and 0.87 % of performance increases were achieved *in terms of micro and macro averaged F1 scores*, respectively. After the forward selection process, the number of features used in the model decreased from 413 to 10.

In the second part of the proposed system, k-means and SLIC methods were applied on 7 test images. The images were segmented into regions, and each region was classified by the obtained neural network model. The average F1 scores for k-means and SLIC methods were 0.88 and 0.84, respectively.

Keywords: *Haralick features, texture based classification, burn image segmentation, GLCM, SLIC*

References

[1] Acha, B., Serrano, C., Acha, J. I., & Roa, L. M. (2003), *CAD Tool for Burn Diagnosis*, LNCS, 2732, 294–305.

[2] Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., Süsstrunk, S. (2012), *SLIC Superpixels Compared to State-of-the-art Superpixel Methods*. IEEE Transactions on Pattern Analysis and Machine Intelligence.

[3] Haralick, R.M., Shanmugam, K. and Distein, I. (1973), *Textural Features for Image Classification*, IEEE Transactions on Systems, Man, and Cybernetics, SMC-3, 610-621.

Learning Bayesian Networks with CoPlot Approach

Derya ERSEL[1], Yasemin KAYHAN ATILGAN[1]
dtektas@hacettepe.edu.tr, ykayhan@hacettepe.edu.tr

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

Many statistical applications require the analysis of multidimensional dataset which has numerous variables and numerous observations. Generally, methods for visualization of multidimensional data such as multi-dimensional scaling, principal component analysis and cluster analysis, analyze observations and variables separately, and uses the composite of variables instead of original ones. However coplot method, uses original variables and enables to investigate relations between both variables and observations together. Also, potentially inconsequential or important variables for further statistical analysis can be determined. In this study, we use coplot methods results to construct a Bayesian network.

Bayesian Networks (BNs) are effective graphical models to represent probabilistic relationships among variables in a multidimensional dataset. These networks, which have an intuitively understandable structure, provide an effective representation of the multivariate probability distribution of random variables. BNs can be created directly using expert opinion without the need for time-consuming learning processes. If expert knowledge is limited, it would be more appropriate to learn BNs from directly from data.

The aim of this study is firstly to introduce Robcop package which is developed for the graphical representation of multi-dimensional dataset, secondly to demonstrate the benefits of coplot results to construct a BN without expert knowledge. This study uses the data from Turkey Demographic and Health Survey which is carried out by the Institute of Population Studies since 1968, for every 5 years. The opinions of women participating the survey on domestic violence, the equality of women and men, and husband's oppression are evaluated together with selected demographic variables.

Keywords: *Multi-dimensional data, CoPlot, Bayesian networks*

References

[1] Chickering, D., Geiger, D. and Heckerman, D. (1995), *Learning Bayesian networks: Search methods and experimental results*, In Proceedings of Fifth Conference on Artificial Intelligence and Statistics, 112-128.

[2] Kayhan Atılgan, Y. (2016), *Robust CoPlot Analysis*, Communications in Statistics - Simulation and Computation, 45, 1763-1775.

[3] Kayhan Atılgan, Y. and Atılgan, E. L. (2017), *A Matlab Package for Robust CoPlot Analysis*, Open Journal of Statistics, 7, 23 - 35.

Evaluation of Ergonomic Risks in Green Buildings with AHP Approach

Ergun ERASLAN[1], Abdullah YILDIZBASI[1]
erguneraslan@gmail.com, ayildizbasi@ybu.edu.tr

[1]*Ankara Yıldırım Beyazıt University, Ankara, Turkey*

Indoors, built for work and life, where we spend a significant part of our daily lives, pose significant risks in terms of human health, work motivation, productivity, and efficiency [3]. Today, with the increase in the importance of human health, there is an increase in the number of studies and practices aimed at reducing or eliminating the risks seen in closed areas. In recent years, concepts such as green buildings and green ergonomics have been used to detect these risk factors that adversely affect human health. Although many countries, especially the developed countries, have been carrying out certification studies on the features that green buildings should have and their application has seen a rapid increase [1]. It is seen that no work has been done in the field of green ergonomics. We proposed to determine the criteria of ergonomics which can be used in the green building certification system which is not yet fully determined in Turkey. These determined criteria will be prioritized by weighting with the Analytical Hierarchical Process (AHP) approach [2]. With this study, a ranking based on expert opinions will be obtained and the deficiencies and risks in the existing system will be tried to be revealed. In this context, 7 main factors and 26 sub criteria have been defined as ergonomic criteria. As a result, an integrated scoring chart has been proposed that takes into account the green ergonomics for green buildings.

According to the results obtained, the highest priorities were defined as "facility and building security", "safe access" and "laboratory buildings with protection level". "Outdoor lighting" was found as the factor with the lowest weight. Finally, a sample building evaluation was conducted and the study findings were tested. It is aimed to shed light on other work that will take into account the ergonomic risks of green building certification in the future.

Keywords: *AHP, Multi-criteria Decision Making, Green Building, Green Ergonomics*

References

[1] Attaianese E. and Duca G. (2012), "*Human factors and ergonomic principles in building design for life and work activities: an applied methodology*." Theoretical Issues in Ergonomics Science, Vol. 13(2), pp. 187–202.

[2] Saaty, T.L. (2008) "*Decision making with the analytic hierarchy process*", International Journal of Services Sciences, Vol. 1(1), pp. 83–98.

[3] Thatcher, A. and Milnera, K. (2012), "*The impact of a 'green' building on employees' physical and psychological wellbeing*." Work. Vol. 41, pp. 3816-3823. 10.3233/WOR-2012-0683-3816.

SESSION I

TIME SERIES I

10TH ISC2017 INTERNATIONAL STATISTICS CONGRESS
December 6-8, 2017 ANKARA/TURKEY

An Investigation on Matching Methods Using Propensity Scores in Observational Studies

Esra Beşpınar[1], Hülya Olmuş[2]
esrabespnar@gmail.com, hulya@gazi.edu.tr

[1]Gazi University, Graduate School of Natural and Applied Sciences, Department of Statistics, Ankara,Turkey
[2]Gazi University, Faculty of Sciences, Department of Statistics, Ankara,Turkey

In observational studies, the random individuals selected for treatment and control group are out of control of the investigator. In such studies, differences between the units can occur in terms of variables. This case will cause biased estimates. Propensity score is a method to reduce bias in estimating treatment effects on an observational data set. After the propensity score is estimated, matching, statification, covariate/regression adjustment and weighting or some combination of four main methods can be used. Thus, homogeneous groups are obtained using these methods and standard deviation of parameter estimates are reduced. The estimated propensity score, for subject i (i = 1,…, N) is the conditional probability of being assigned to a particular treatment given a vector of observed covariates x_i:

$$e(x_i) = P_r(z_i = 1/x_i)$$

The propensity score can be obtaiend using the logistic regresyon method, discriminant analysis, and clustering analysis. Logistic regression method that does not require any assumption for obtaining the propensity score is more desirable. In propensity score matching, units with similar propensity score in the treatment and control group are matched and all other unmatched units are removed from study. In this study, Nearest Neighbor 1:1 matching, Stratified matching and Caliper matching are discussed by using propensity scores on the real data set with R programming. As a result of these matching, parameter estimation are obtained and the results are interpreted. One of the highlighted results shows that reducing bias in parameter estimation can be important in propensity score matching.

Keywords: *propensity scores, logistic regression, matching, observational studies*

References

[1] Austin, P. C. (2011), *An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies*, Multivariate Behavioral Research, 46, 399–424.

[2] Rosenbaum, P.R. and Rubin, B.R. (1983), *The Central Role of the Propensity Score in Observational Studies for Causal Effects*, Biometrika, 70 (1), 41-55.

[3] Tu, C. and Koh, W.Y. (2015), *A comparison of balancing scores for estimating rate ratios of count data in observational studies*, Communications in Statistics-Simulation and Computation, 46 (1), 772-778.

[4] Demirçelik, Y. and Baltalı O. (2017), *The relationship between the anthropometric measurements of the child and the mother's perception and the affecting factors*, T.C. Ministry of Health Turkish Public Hospitals Institution Izmir Province Public Hospitals Association North General Secretariat University of Health Sciences Tepecik Education and Research Hospital Pediatric Health and Diseases Clinic.

10TH ISC2017 INTERNATIONAL STATISTICS CONGRESS
December 6-8, 2017 ANKARA/TURKEY

A Simulation Study on How Outliers Effect The Performance of Count Data Models

Fatih TÜZEN[1] , Semra ERBAŞ[2] and Hülya OLMUŞ[2]
mfatihtuzen@tuik.gov.tr, serbas@gazi.edu.tr, hulya@gazi.edu.tr

[1]*TURKSTAT, Ankara, Turkey*
[2]*Gazi University, Ankara, Turkey*

In many applications, count data have high proportion of zeros and they are not optimally modelled with a normal distribution. Because the assumptions of the ordinary least-squares regression are violated (homoscedasticity, normality, and linearity), the use of these statistical techniques generally causes biased and inefficient results [1]. Zero-inflated models have been used to cope with excess zeros and overdispersion that occurs when the sample variance exceeds the sample mean. Zero-Inflated Poisson (ZIP) Regression model is one of the zero-inflated regression models. The ZIP regression model was first introduced by Lambert [2], and she applied this model to the data collected from a quality control study, in which the response is the number of defective products in a sample unit. In practice, even after accounting for zero-inflation, the non-zero part of the count distribution is often over-dispersed. In this case, Greene [3] described an extended version of the negative binomial model for excess zero count data, the Zero-Inflated Negative Binomial (ZINB), which may be more suitable than the ZIP. Our study was aimed at comparing the performance of count data models under various outliers and zero inlation situations with simulated data for 500 sample size. Poisson, Negative Binomial, Zero-Inflated Poisson and Zero-Inflated Negative Binomial models were considered to test how well each of the model fits the selected data sets having outliers and excess zeros. We studied three different zero-inflation conditions for the response variable. Also in order to be able to evaluate count data models in a different way, the dependent variable was also designed according to whether it contains outliers or not. Therefore; we examined the count data models in terms of three different outlier magnitudes by creating low, medium and high level of outliers when the outlier ratio is 5%. Finally, the study focused on identifying model(s) which can handle the impact of outliers and excess zeros in count data base on AIC under varying degrees of outliers and zeros with simulated data. We found that Zero-Inflated Negative Binomial (ZINB) models were found to be more successful than other count data models. Also the results indicated that in some scenarios, the NB model outperforms other models in the presence of outliers and/or excess zeros.

Keywords: *count data, zero-inflated data, outliers*

References

[1] Afifi, A.A., Kotlerman, J.B., Ettner, S.L., Cowan, M. (2007), *Methods for improving regression analysis for skewed continuous or counted responses*, Annual Review of Public Health, 28: 95–111.

[2] Lambert, D. (1992), *Zero-inflated Poisson regression with an application to detects in manufacturing*, Technometrics, 34: 1–14.

[3] Greene, W. H., (1994), *Accounting for Excess Zeros and Sample Selection in Poisson and Negative Binomial Regression Models*, NYU Working Paper No. EC-94-10. Available at SSRN: https://ssrn.com/abstract=1293115

10TH ISC2017 INTERNATIONAL STATISTICS CONGRESS
December 6-8, 2017 ANKARA/TURKEY

Comparison of Parametric and Non-Parametric Nonlinear Time Series Methods

Selman MERMİ[1] and Dursun AYDIN[1]
selmanmermi@mu.edu.tr, duaydin@mu.edu.tr

[1]*Muğla Sıtkı Koçman University, Mugla, Turkey*

Modelling and estimating of time series have an important place in many application areas. Non-linear time series models have gained more importance recently due to various restrictions on exposure to observational work and many parametric regime-switching models and non-parametric methods have been developed to demonstrate non-linearity of time series in recent past. Analyses of econometric time series with non-linear models means that certain properties of time series such as mean, variance and autocorrelation vary over time. [1]

Non-linear time series analysis literature was come out as parametric TAR, STAR, SETAR, LSTAR models and these models are improved with various studies. In TAR models, a regime switch happens when the threshold variable crosses a certain threshold. In some cases, regime switch happens gradually in a smooth fashion. If the threshold variable related with TAR models is replaced by a smooth transition function, TAR models can be generalized to smooth transition autoregressive (STAR) models. [2]

Regime switch between regimes happens with an observable threshold variable in TAR and STAR models. In Markov switching models, switching mechanism is controlled by an unobservable state variable contrary to TAR and STAR models. Hence, it is not known exactly which regime is effective at any point in time. [3]

Unlike parametric models, nonparametric regression models do not rely on the calculation of the regression coefficients of a particular model. The nonparametric regression is to provide a model describing the relationship between the two main variables and try to estimate the most appropriate model based on the observations at hand without reference to a particular parametric model. In this work, kernel smoothing and smoothing spline methods are discussed. [4]

The purpose of this work is to model parametric and nonparametric models mentioned above with a financial data set. The obtained models are compared with performance criteria and graphs showing the relation of the real-concordance values of the models. As a result, it is seen that nonparametric methods give much more effective results compared to the parametric models.

Keywords: *nonlinear time series models, TAR model, STAR model, nonparametric methods*

References

[1] Khan, M. Y. (2015), *Advanced in Applied Nonlinear Time Series Modeling*, Doctoral Thesis, Münih Üniversitesi, Münih, 181s.

[2] Zivot, E. ve Wang, J. (2006) *Modelling Financial Time Series with S-PLUS*, 2. Baskı, Springer Science+Business Media, USA, 998s.

[3] Kuan, C. M. (2002) *Lecture On The Markov Switching Model*, Institute of Economics Academia Sinica, Taipei, 40s.

[4] Eubank, L. R. (1999) *Nonparametric Regression and Spline Smoothing*, Marcel Dekker, New York, 337s.

10TH ISC2017 INTERNATIONAL STATISTICS CONGRESS
December 6-8, 2017 ANKARA/TURKEY

Regression Clustering for PM_{10} and SO_2 Concentrations in Order to Decrease Air Pollution Monitoring Costs at Turkey

Aytaç PEKMEZCİ[1], Nevin GÜLER DİNCER[1]
aytac0803@mu.edu.tr, nguler@mu.edu.tr

[1] *Muğla Sıtkı Koçman University, Department of Statistics, Muğla, Turkey*

In this study, parameters of regression models between on weekly PM_{10} and SO_2 concentrations obtained from air pollution monitoring stations at Turkey are clustered. The objective in here is to obtain fewer number of regression models in order to explain the relationship between them and thus get information about all stations by monitoring fewer number of stations. Following procedure to achieve this objective consists of seven steps: i) determining lag lengths according to Akaike Information Criteria (AIC) [1] and Schwarz Information Criterion (SIC) [2], ii) examining the autocorrelations and normality, iii)identifying dependent variable by using Granger causality test [3], iv) determining the regression models being statistically significant, v) determining optimal number of clusters by using Xie-Beni index, vi) clustering of the parameters of regression models being significant and lastly vii) predicting dependent variable by using parameters of regression models obtaining as cluster centres for all stations. When these steps are followed, weekly SO_2 concentrations are determined as dependent variable and it is decided that 80 of 111 stations could be used for predicting. Optimal number of clusters is designated as 5 for 80 stations and Fuzzy K-Medoid Clustering is performed for clustering. SO_2 values are predicted for all stations based on regression parameters as determined cluster centres and weekly PM_{10} concentrations. Prediction results are compared with those of obtained when all stations are predicted separately and it is concluded that one can provide information about all stations by monitoring fewer number of stations.

Keywords: *Regression clustering, Granger causality test, Air pollution prediction*

References

[1] Akaike, H. (1981). *Likelihood of a model and information criteria*. Journal of. Econometrics, 16, 3-14.

[2] Schwarz, G. (1978). *Estimating the dimension of a model*. The Annals of Statistics, 6(2), 461-464.

[3] Granger, C.W.J., Newbold, P., (1977). *Forecasting Economic Time Series*. Academic Pres, London, 333p.

Analysis of a Blocked Tandem Queueing Model with Homogeneous Second Stage

Erdinç Yücesoy[1], Murat Sağır[2] , Abdullah Çelik[3] , Vedat Sağlam[3]
erdincyucesoy@gmail.com , istatistikci_murat@hotmail.com , abdullahcel@gmail.com , vsaglam@omu.edu.tr

[1]*Ordu Üniversitesi Matematik Bölümü, Ordu, Türkiye*
[2]*İskenderun Teknik Üniversitesi Ekonomi Bölümü, İskenderun, Türkiye*
[3]*Ondokuz Mayıs Üniversitesi, İstatistik Bölümü, Samsun, Türkiye*

In this analysed queueing system, the customers arrive the system with parameter Poisson stream. There is a single service unit at first stage which has exponentially distributed service time with parameter μ and no queue is allowed at first stage. There are two service units at second stage and both have exponentially distributed service time with parameter μ_h. In other words the second stage of this queueing system is homogeneous. Also, no queue is allowed at second stage. Upon having service at first stage, if both service units are available, an arriving customer chooses any of two service units at second stage with equal probabilities and leaves the system after completing service. If only one of the service unit is available at second stage, the customer proceeds this service unit and leaves the system after having service. If both of two service units at second stage are busy the customer waits till at least one service unit is empty and hence blocks the service unit at first stage and causes customer loss. The fundamental system measurement in this queueing model is the loss probability.

Keywords: *3-dimensional Markov chain, Poisson stream, Loss probability*

References

[1] Sağlam, V., Sağır, M., Yücesoy, E. and Zobu, M. (2015), *The Analysis, Optimization, and Simulation of a Two-Stage Tandem Queueing Model with Hyperexponential Service Time at Second Stage*, Mathematical Problems in Engineering, Volume 2015, 6 pages.

[2] Alpaslan, F. (1996), *On the minimization probability of loss in queue two heterogeneous channels*, Pure and Applied Mathematika Sciences, Volume 43, Pages 21-25.

[3] Song, X. and Mustafa, M. A. (2009), *A performance analysis of discrete-time tandem queues with Markovian sources*, Performance Evaluation, vol. 66, no. 9-10, pp. 524–543.

[4] Gomez, A. and Martos, M. E. (2006), *Performance of two-stage tandem queues with blocking: the impact of several flows of signals*, Performance Evaluation, vol. 63, no. 9-10, pp. 910–938.

SESSION I

DATA ANALYSIS AND MODELLING

Intituionistic Fuzzy Tlx (If-Tlx): Implementation of Intituionistic Fuzzy Set Theory for Evaluating Subjective Workload

Gülin Feryal CAN[1]
gfcan@baskent.edu.tr

[1]*Başkent University, Engineering Faculty, Industrial Engineering Department, Ankara, Turkey*

The determination of subjective workload (SWL) imposed on an employee plays an important role in designing and evaluating an existing work and work environment system. On the other hand, it is a hard problem since SWL evaluation is typically a multi-dimensional problem involving several work demands on which employee's evaluation is usually vague and imprecise. In this study, NASA TLX (National Aeronautics and Space Administration Task Load Index) method used widely in different work types combined with intuitionistic fuzzy set (IFS) theory to determine SWL in an industrial sailing environment. The integrated method is named as Intuitionistic Fuzzy TLX (IF-TLX). An IFS is a powerful tool to model the uncertainty because of degree of hesitation in human decision system. It is worth pointing out that proposed method also considers work experience effect on SWL evaluation. This improves objectivity of final SWL scores for the whole work. This paper also develops a new intuitionistic evaluation scale for rating of SWL dimensions and work experience. As a result of this study it is determined that industrial salespeople who have more than 15 years of work experience feel the highest SWL with the effect of increasing age.

Keywords: *subjective workload, intuitionistic fuzzy sets, intuitionistic triangular fuzzy numbers, work experience*

References

[1] Hart SG. and Staveland LE. (1988), *Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research*, Advances in psychology, 52, 139-183.

[2] Atanassov KT. (1986), *Intuitionistic fuzzy sets*, Fuzzy sets and Systems, 20(1), 87-96.

[3] Schmidt FL., Hunter, JE. and Outerbridge, AN. (1986*), Impact of job experience and ability on job knowledge, work sample performance, and supervisory ratings of job performance*, Journal of applied psychology, 71(3): 432.

[4] Hussain RJ. and Kumar PS. (2012), *Algorithmic approach for solving intuitionistic fuzzy transportation problem*, Applied mathematical sciences, 6(77-80), 3981-3989.

[5] Mouzé-Amady M, Raufaste E., Prade H. and Meyer JP. (2013), *Fuzzy-TLX: using fuzzy integrals for evaluating human mental workload with NASA-Task Load index in laboratory and field studies*, Ergonomics, 56(5), 752-763.

Evaluation of Municipal Services with Fuzzy Analytic Hierarchy Process for Local Elections

Abdullah YILDIZBASI[1], Babek ERDEBILLI[1], Seyma OZDOGAN[1]
ayildizbasi@ybu.edu.tr, babek.erdebilli2015@gmail.com, seyma.ozdogan6@gmail.com

[1]*Ankara Yıldırım Beyazıt University, Ankara, Turkey*

Since municipalities are the closest institution to society, they are one of the biggest factors for parties to achieve success in local elections. For this reason, mayors must know the wishes of the people well in order to win elections again and provide benefit to the party by making improvements according to the needs of the people [1]. Otherwise, this rule also applies the mayor in the parties. If a person in a municipality management isn't accepted by the people and s/he make dissatisfied the people in terms of their services, the parties can change the person in management. In short, this study covers both parties and mayors. The question is that how mayors know for certain which services are the most important to gain party appreciation by gaining citizen appreciation. It is seen that no work has been done in the field of municipality and elections. In this study, we aimed to reach the solution of the question so that mayor can maintain substantial existing chairmanship. In order to accomplish the result, 4 main factors and 24 sub-criteria have been defined as municipal service criteria. Afterwards, fuzzy analytic hierarchy process (FAHP) approach will be used to weight the criteria which are evaluated by an expert and they will be prioritized according to this weight[2].

According to the results obtained, the highest priority was defined as 'Infrastructure Services' and the lowest one was defined as 'Emergency Services'. Finally, this study was applied to a municipality and results were checked.

Keywords: *Fuzzy AHP, Multi-Criteria Decision Making, Municipal Services*

References

[1] Akyıldız, F. (2012), *Belediye Hizmetleri ve Vatandaş Memnuniyeti: Uşak Belediyesi Örneği,* Journal of Yaşar University, Vol. 26, No. 7, pp. 4415–4436.

[2] Wang, C., Chou, M. and Pang, C. (2012), *Applying Fuzzy Analytic Hierarchy Process for Evaluating Service Quality of Online Auction,* International Journal of Computer and Information Engineering, Vol:6, No:5, pp. 610–617

Analyzing the Influence of Genetic Variants by Using Allelic Depth in the Presence of Zero-Inflation

Özge KARADAĞ
ozgecankaradag21@gmail.com

Hacettepe University, Ankara, Turkey

The influence of genetic variants on a phenotype such as the diastolic blood pressure which measures heart pressure during relaxation, is commonly investigated by testing for association between called genotypes and the quantitative phenotype via fitting statistical models. In genetic association studies, the genetic component is usually obtained as genotype.

As an alternative to genotype, allelic depth can also be used for testing genetic association. The counts of alleles are approximately distributed as a Poisson process and the association can be tested by a standard Poisson regression. However in the sequence data, there is often excess zero. Observations departing on the majority of the data, these zero counts, have a strong influence on standard techniques.

In this study, different testing procedures are compared to evaluate the influence of genetic variants on phenotype, regarding the type-I error rates and the power of association results by considering zero-inflation. Implementation of the models, is evaluated to real sequence data of Hispanic samples for Type 2 Diabetes (T2D).

Keywords: *association test, zero-inflation, allele counts, count data models*

References

[1] Karazsia, B.T., Dulmen M.H.M., (2008), *Regression Models for Count Data: Illustrations using Longitudinal Predictors of Childhood Injury*, Journal of Pediatric Psychology 33(10): 1076-1084.

[2] Lambert, D. (1992), *Zero-Inflated Poisson Regression, with an Application to Defects in Manufacturing*, *Technometrics* 34(1): 1–14.

[3] Satten, G.A., Johnston, H.J., Allen, A.S. and Hu, Y. (2012), *Testing Association without Calling Genotypes Allows for Systematic Differences in Read Depth between Cases and Controls*, Abstracts from the 22nd Annual Meeting of the International Genetic Epidemiology Society, Chicago IL, USA. Page 9. ISBN: 978-1-940377-00-1.

Survival Analysis and Decision Theory in Aplastic Anemia Case

Mariem BAAZAOUI[1], Nihal ATA TUTKUN[1]
baazaoui.mariem@gmail.com, nihalata@hacettepe.edu.tr

[1]*Department of statistics, Ankara, Turkey*

The community of medicine presents rare and dangerous diseases for which the duration of survival is short. The survival analysis is often used in that diseases cases. Survival time varies according to the method of therapy used. The expert or the patient have to choose between therapy methods tacking into consideration some factors, from this perspective it is one of optimization problems.

This study deals with the aplastic anemia as one of very rare disease. The methods of therapy for this disease differ and depend on factors such as the patient age, the current condition of the patient, find a suitable donor, etc. It is important to estimate the value of different states of health assuming different potential lengths of survival.

The complicated choices for the individual decision-making or the group decision-making in the case of aplastic anemia can be summarized as: If all factors are favorable (young, don't suffer from other diseases and find a suitable donor), nothing can guarantee the success of the bone marrow transplantation (BMT). Otherwise, if the majority of factors are unfavorable, no one can confirm the failure of the bone marrow transplantation (BMT). If no (BMT), which kind of therapy can be suitable for each case. The type of therapy chosen is shown as a decision problem. So that, the methods of optimization from the decision theory can be applied in this purpose. In this study, one of these optimization methods called Savage method was applied on the results of survival analysis investigated by Judith (2006).

Keywords: *decision theory, survival analysis, aplastic anemia*

References

[1] Amy, E. and Robert, A. (2011), *Clinical management of aplastic anemia*, *Expert Rev* Hematol., 4(2), 221–230.

[2] Fouladi, M., Herman, R., Rolland-Grinton, M., Jones-Wallace, D., Blanchette, V., Calderwood, S., Doyle, J., Halperin, D., Leaker, M., Saunders, EF., Zipursky, A. and Freedman, MH. (2000), *Improved survival in severe acquired aplastic anemia of childhood*, Bone Marrow Transplantation, 26, 1149–1156.

[3] Hasan, J. and Ahmad, KH. (2015), *Immunosuppressive Therapy in Patients with Aplastic Anemia: A Single-Center Retrospective Study*, Plos One, 10(5), 1-10.

[4] Judith, M. (2006), *Making Therapeutic Decisions in Adults with Aplastic Anemia*, American Society of Hematology, 1, 78-85.

Determinants of wages & Inequality of Education in Palestinian Labor Force Survey

Ola Alkhuffash[1]
Ola.khuffash@yahoo.com

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

The Palestinian Labor Force survey is a household survey has a time series started since 1993, it provides a data of employment and unemployment in Palestine with demographic, social and economic characteristics of the sample which is represented of Palestinian Society, this paper aimed to study the factors which affect on wages for employed Palestinian according to their locality type as a second level by Hierarchical linear model technique, and also to determine the inequality of education over the years 2010-2015 by calculating gini index.

Keywords*: Labor Force, Hierarchical Linear Model, Gini Index*

References

[1] ILO, *Current international recommendations on labour statistics*,2000.

[2] Elqda & Bashayre, 2013.Y*outh and work - an analytical study of the characteristics of the labor force of young people in Jordan*. Amman-Jordan

[3] Palestinian Central Bureau of Statistics, 2016. *Labor Force Survey*, 2015. Ramallah-Palestine

[4] Knight John , Shi Li, Quheng Deng .2010. *Education and the Poverty Trap in Rural China*. Oxford Development Studies.

[5] Palestinian Central Bureau of Statistics, 2007. *Wage Structure and Work Hours Survey 2006: Main Findings.* Ramallah-Palestine.

SESSION I

FUZZY THEORY AND APPLICATION

10TH ISC2017 INTERNATIONAL STATISTICS CONGRESS
December 6-8, 2017 ANKARA/TURKEY

Assessment of Turkey's Provincial Living Performance with Data Envelopment Analysis

Gül GÜRBÜZ[1], Meltem EKİZ[2]
gul.gurbuz@tuik.gov.tr, ozmeltem@gazi.edu.tr

[1]*Türkiye İstatistikKurumu, Malatya, TÜRKİYE*
[2]*Gazi Üniversitesi, Ankara, TÜRKİYE*

Population indicators may denote a country development level. This indicators are effecitve assessment of socio-economic development level. As families and societies socio-economic status slips, living standards are affecting negatively. Aim of this study is investigation of 81 Turkey provinces social and economical level and present inhabitability performance with data envelopment analysis. Data Envelopment Analyses (DEA) is a powerful, non-parametric method , using for measuring performance. [1]. This method indicates a best line and analyses data with the being under the line or above the line situation and compare the efficiency. [2,3]. This method's key feature is assessment availability in case of numerous input and output. In this study classical CCR and BCC methods are used and results are investigated. In this study, 81 provinces socio-economic living performance determined with CCR method by using TÜİK 2015 living satisfaction study datas.Turkey's provinces' socio-economic stuation determined and living performance investigated by CCR and BCC In this study, input variables are unemployement rate, homicide rate, application rate for per doctor, baby mortality rate and rate of people who feels alone when they walk alone at night while output variables are the average daily earning taken basic, middle and upper winning class group household rate, faculty nad high scool graduation rate and social life satisfaction rate. 20 province was active by classical CCR results and 30 province was active by BCC results. As a conclusion we observed that, analysis results with CCR method is more distictive than BCC method.

Keywords: *Data Envelopment Analyses , BCC, CCR*

References

[1] CharnesA. and Cooper, W.W. (1985). *Preface to topics in data envelopment analysis*, Annals of operations research, 2, 59-94.

[2]Bowlin W. F. (1999). *An analysis of the financial performance of defense business segments using data envelopment analysis*, Journal ofaccountingandpublicpolicy,18(4/5), 287-310.

[3] Cooper**,** W.W.,Seiford**,** L. M. and Tone**,** K. (2007). *Data envelopment analysis,*USA, Springer-Verlag.

Modified TOPSIS Methods for Ranking The Financial Performance of Deposit Banks in Turkey

Semra ERPOLAT TAŞABAT[1]
semra.erpolat@msgsu.edu.tr

[1]*Faculty of Science Literature, Department of Statistics, Mimar Sinan Fine Art University, Istanbul, Turkey*

Decision-making, defined as the selection of the best among the various alternatives, is called Multi-Criteria Decision Making (MCDM) when there are multiple criteria. The MCDM methods, which presented solution proposals for the correct and useful decisions that can be made in many areas have begun to develop from the beginning of 1960's. The main purpose of using the methods is to control the decision making mechanism in cases where there are a lot of alternative and criterion numbers and to make the decision result as easy and quick as possible.

There are many multi criteria decision making methods in the literature. One of them is the Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS) introduced by Hwang and Yoon (1981). The method is briefly based on the principle that the selected alternative should have the shortest distance from the positive ideal solution and the farthest distance from the negative ideal solution).

In this study, as an alternative to the Euclidean distance measure used in the calculation of the positive and negative ideal solutions at the traditional TOPSIS method a different approach has been proposed by using Lp Minkowski family and $L1$ Family distance measures. With the modified TOPSIS methods, the financial performance of the deposit bank operating in the Turkish Banking Sector was examined. From the results obtained, it has been tried to emphasize the importance of the distance measures used in the TOPSIS method in order of alternatives.

Keywords: *MCDM; TOPSIS, Lp Minkowski family distance, Ł1 Family distance.*

References

[1] Hwang, C. L., and Yoon, K. (1981). *Multiple Attributes Decision Making Methods And Applications*. Berlin: Springer.

[2] Opricovic S., Tzeng Gwo-H. (2004), "Compromise solution by MCDM methods: A comparative analysis of VIKOR and TOPSIS", *European Journal of Operational Research*, 156, pp 445-455.

[3] Taşabat S.E., Cinemre N., Şen S., (2005), "Farklı Ağırlıklandırma Tekniklerinin Denendiği Çok Kriterli Karar Verme Yöntemleri İle Türkiye'deki Mevduat Bankalarının Mali Performanslarının Değerlendirilmesi", *Social Sciences Research Journal*, Volume 4, Issue 2, 96-110, ISSN: 2147-

10TH ISC2017 INTERNATIONAL STATISTICS CONGRESS
December 6-8, 2017 ANKARA/TURKEY

A New Multi Criteria Decision Making Method Based on Distance, Similarity and Correlation

Semra ERPOLAT TAŞABAT[1]
semra.erpolat@msgsu.edu.tr

[1]*Department of Statistics, Mimar Sinan Fine Art University, Istanbul, Turkey*

Decision making, briefly defined as choosing the best among the possible alternatives within the possibilities and conditions available, is a far more comprehensive process than instant. While decision making process, there are often a lot of criteria as well as alternatives. In this case, methods referred to as Multi Criteria Decision Making (MCDM) are applied. The main purpose of the methods is to facilitate the decision maker's job, to guide the decision maker and help him to make the right decisions if there are too many options.

In cases where there are many criteria, effective and useful decisions have been taken for granted at the beginning of the 1960s for the first time, and supported by day-to-day work. A variety of methods have been developed for this purpose. The basis of some of these methods is based on distance measures. The most known method in the literature based on the concept of distance is, of course, a method called Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS).

In this study, a new multi criteria decision making method that uses distance, similarity and correlation measures has been proposed. In the method, Euclid was used as distance measure, cosine was used as similarity measure and Pearson correlation was used as relation measure. Using the positive ideal and negative ideal values obtained from these measures, respectively a common positive ideal value and a common negative ideal value were obtained. The study also proposed a different ranking index from the ranking index used in the traditional TOPSIS method. The proposed method has been tested on the variables showing the development levels of the countries that have a very important place today. The results obtained were compared with the Human Development Index (HDI) value developed by the United Nations.

Keywords: *MCDM, TOPSIS, Distance, Similarity, Correlation, Human Development Index.*

References

[1] Hepu Deng, *A Similarity-Based Approach to Ranking Multicriteria Alternatives*, International Conference on Intelligent Computing ICIC 2007: Advanced Intelligent Computing Theories and Applications With Aspects of Artificial Intelligence pp 253-262

[2] Hossein Safari, Elham Ebrahimi, *Using Modified Similarity Multiple Criteria Decision Making technique to rank countries in terms of Human Development Index*, Journal of Industrial Engineering and Management JIEM, 2014 – 7(1): 254-275 – Online ISSN: 2013-0953 – Print ISSN: 2013-8423 http://dx.doi.org/10.3926/jiem.837

[3] Hossein Safari, Ehsan Khanmohammadi, Alireza Hafezamini and Saiedeh Sadat Ahangari, *A New Technique for Multi Criteria Decision Making Based on Modified Similarity Method*, Middle-East Journal of Scientific Research 14 (5): 712-719, 2013 ISSN 1990-9233 © IDOSI Publications, 2013 DOI: 10.5829/idosi.mejsr.2013.14.5.335.

Ranking of General Ranking Indicators of Turkish Universities by Fuzzy AHP

Ayşen APAYDIN[1], Nuray TOSUNOĞLU[2]
aapaydin@ankara.edu.tr, nguneri@gazi.edu.tr

[1]*Ankara University, Ankara, Turkey*
[2] *Gazi University, Ankara, Turkey*

Ranking universities by academic performance is important both in the world and Turkey. The purpose of this ranking is to help determine the potential areas of progress for universities. In the world, university ranking systems are based on different conflicting indicators for ranking of university. Rankings are conducted by several institutions or organizations including ARWU-Jiao Tong (China), THE (United Kingdom), Leiden (The Netherlands), QS (United Kingdom), Webometrics (Spain), HEEACT/NTU (Taiwan) and SciMago (Spain).

The first ranking system for Turkish universities is University Ranking by Academic Performance (URAP-TR). URAP-TR ranking system was developed in 2009 by the University Ranking and Academic Performance Research Laboratory in METU. URAP-TR uses multiple ranking indicators to balance size-dependent and size-independent academic performance indicators in an effort to devise a fair ranking system for Turkish universities.

The nine indicators that URAP uses in the overall ranking of Turkish universities for 2016-2017 are: the number of articles, the number of articles per teaching member, the number of citations, the number of citations per teaching member, the total number of scientific documents, the number of scientific documents, the number of doctoral graduates, the ratio of doctoral students, the number of students per faculty member. The nine indicators used in the sequence have equal weight percentages.

In this study, the determination of the weight percentages has been considered as a multi-criteria decision making (MCDM) problem. The aim of the study is to determine the significance of the indicators through the fuzzy AHP. Indicators will be compared using fuzzy numbers and fuzzy priorities will be calculated.

Keywords: *University ranking, ranking indicators, URAP, Fuzzy AHP*

References

[1] Alaşehir, O., Çakır, M.P., Acartürk, C., Baykal, N. and Akbulut, U. (2014), *URAP-TR: a national ranking for Turkish universities based on academic performance*, Scientometrics, 101, 159-178.

[2] Çakır, M.P., Acartürk, C., Alaşehir, O. and Çilingir, C. (2015), *A comparative analysis of global and national university ranking systems*, Scientometrics, 103, 813–848.

[3] Moed, H.F. (2017), A critical comparative analysis of five world university rankings, Scientometrics, 110: 967-990.

[4] Olcay, G. A. and Bulu, M. (2017), *Is measuring the knowledge creation of universities possible?: A review of university rankings*, Technological Forecasting & Social Change, 123, 153–160.

[5] Pavel, A-P. (2015), *Global university rankings-a comparative analysis*, Procedia Economics and Finance, 26, 54-63.

10TH ISC2017 INTERNATIONAL STATISTICS CONGRESS
December 6-8, 2017 ANKARA/TURKEY

Exploring the Factors Affecting the Organizational Commitment in an Almshouse: Results of a CHAID Analysis

Zeynep FİLİZ[1], Tarkan TAŞKIN[1]
zfiliz@ogu.edu.tr, tarkan.ttaskin@gmail.com

[1]*Eskişehir Osmangazi University, Eskişehir, Türkiye*

The purpose of this study is analyzing the connection between factors affecting organizational commitment of the workers in an almshouse using CHAID Analysis method.
In order to measure organizational commitment in the research used Allen and Meyer's "Three-Component Model of Organizational Commitment" [1] questionnaire. The applied questionnaire is taken from the Master graduate thesis of Tuğba ŞEN [3]. Questionnaire was distributed to all almshouse workers.
Reliability Analysis was conducted at the beginning of the analysis, 7th and 15th because the questions were not reliable. The Cronbach Alpha value was calculated as 0,843. The mean age of 200 participants was 31-35 years, of whom 47% (n=94) were female and 53% (n=106) male. 83% (n=88) of male employees are married, while 22% (n=21) of female workers are single. 56.5% (n = 113) of the employees were graduated from high school, 32.5% (n = 32.5) primary school, 10.5% (n = 21) have Bachelor's degree and 1 employees have Master Degree. 54% (n = 109) in Care Services, 24.5% (n = 49) in Health services, 4.5% (n = 9) in Therapy services and 16.5% (n = 33) are working in other services. Factor analysis was performed on the survey results, and the Kaiser-Meyer-Olkin (KMO) value was calculated as 0.843. As a result of analysis, 3 factors were obtained. These factors are; Emotional commitment, Continuance commitment, Normative commitment. A total of 10 variables Chi-squared Automatic Interaction Detector (CHAID) analysis [2] was applied to these three factors, Gender, Age, Marital Status, Number of Children, Education Level, Year of Study and Mission.
As a result of the Chi-squared Automatic Interaction Detector, 51.5% (n = 103) of the employees were found to have positive organizational commitment. The variable that best explains organizational commitment, which is one of the factors obtained, is Emotional Commitment. It was observed that 95% (n = 98) of those under an Emotional Loyalty value of -1 and 58% (n = 43) of those between Emotional Commitment values of -1 and 0,35 were not Organizationally Linked. On the other hand, 80% of those with an Emotional Commitment score of 0.35 or higher have reached the conclusion that their organizational commitment is positive. According to the CHAID analysis, the variable that best describes the Emotional Commitment value between -1 and 0.35, which is one of the factors obtained, is the Continuance Commitment variable. It was observed that organizational commitment of 66% (n = 33) of those who were smaller than Continuance Commitment value of 0.23 was negative. It is seen that the organizational commitment of 66.7% (n = 20) of the ones with Continuance Commitment value greater than 0.23 is positive.

Keywords: *CHAID Analysis, Organizational commitment,*

References

[1] Allen, Natalie J., Meyer, John P. (1990), *The measurement and antecedents of affective, continuance and normative commitment to the organization*, Journal of Occupational Psychology, 63, 1, 11-18.

[2] Kass, G. V. (1980*), An Exploratory Technique for Investigating Large Quantities of Categorical Data*. Applied Statistics, 20, 2, 119-127.

[3] Şen, T. (2008), *İş Tatmininin Örgütsel Bağlılık Üzerindeki Etkisine İlişkin Hızlı Yemek Sektöründe Bir Araştırma*, Marmara Üniversitesi SBE, 130.

Fuzzy Multi Criteria Decision Making Approach for Portfolio Selection

Serkan AKBAŞ[1], Türkan ERBAY DALKILIÇ[1]
serkanakbas@ktu.edu.tr, terbay@ktu.edu.tr

[1] *Department of Statistics and Computer Science, Karadeniz Technical University, Trabzon, TURKEY*

In daily life events, there are many complexities arising from lack of information and uncertainty. For this reason, it is difficult to be completely objective in the decision-making process. Fuzzy linear programming model has been developed to reduce or eliminate this complexity. Fuzzy linear programming is the process of choosing the optimum solution from among the decision alternatives to achieve a specific purpose in cases where the information is not certain.

One of the fields where the lack of information or uncertainty makes it difficult to decide is financial markets. Investors who have a certain amount of accumulations are aiming to increase in various ways as well as protecting the value of their income. While doing this, investors trying to create a portfolio from various securities, encounter the problem of deciding to which investment vehicle they need to invest in what extent. Therefore, investors apply to fuzzy linear programming model to eliminate this uncertainty and to create the optimal portfolio.

In the portfolio selection process suggestions in the literature, the determination of criteria weights is based on triangular fuzzy numbers. In this study, criteria weights were determined based on trapezoidal fuzzy numbers. With the solution of the linear programming model which is based on the determined weights, an alternative solution has been produced to the problem of which investment instrument will be invested at what proportion. The results obtained from the existing methods and the results obtained from the proposed model were compared.

Keywords: *Multi-criteria decision making, Analytic hierarchy process, Trapezoidal fuzzy numbers, Portfolio selection.*

References

[1] Enea, M. (2004). Project Selection by Constrained Fuzzy AHP. *Fuzzy optimization and decision making*, 3(1), pp. 39–62.

[2] Ghaffari-Nasab, N., Ahari, S., & Makui, A. (2011). A portfolio selection using fuzzy analytic hierarchy process: A case study of Iranian pharmaceutical industry. *International Journal of Industrial Engineering Computations*, 2(2), 225-236.

[3] Rahmani, N., Talebpour, A., & Ahmadi, T. (2012). Developing aMulti criteria model for stochastic IT portfolio selection by AHP method. *Procedia-Social and Behavioral Sciences*, 62, 1041-1045.

[4] Tiryaki, F. & Ahlatcioglu, B. (2009) Fuzzy portfolio selection using fuzzy analytic hierarchy process. *Information Sciences*, vol. 179, no. 1–2, pp. 53–69, 2009.

[5] Yue, W., & Wang, Y. (2017). A new fuzzy multi-objective higher order moment portfolio selection model for diversified portfolios. *Physica A: Statistical Mechanics and its Applications*, 465, 124-140.

SESSION II

STATISTICS THEORY II

Bayesian Conditional Auto Regressive Model for Mapping Respiratory Disease Mortality in Turkey

Ceren Eda CAN[1], Leyla BAKACAK[1], Serpil AKTAŞ ALTUNAY[1], Ayten YİĞİTER[1]
cerencan@hacettepe.edu.tr, leylabakacak@hacettepe.edu.tr, spxl@hacettepe.edu.tr, yigiter@hacettepe.edu.tr

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

Spatial analysis is a technique to reveal and characterize the spatial patterns and anomalies over a geographical region by regarding both the attribute information of objects in a data set and their locations. The set of spatial objects on which the data are recorded can be as a form of point, polygon, line or grid. The response variable typically exhibits spatial autocorrelation. Observations from objects close together tend to be more similar than those relating to objects further apart. Although a model includes covariates, spatial autocorrelation cannot be captured explicitly and remains in the residuals of the model. In such cases, the residuals due to the spatial autocorrelation violate the assumption of independence. We use conditional autoregressive (CAR) model to avoid from the residual spatial autocorrelation. In CAR model, spatial autocorrelation is modelled by a set of spatially correlated random effects that are assigned a CAR prior distribution. The R package CARBayes provides a Bayesian spatial modelling with CAR priors for data relating to a set of non-overlapping areal objects. In CARBayes, inference is based on Markov Chain Monte Carlo (MCMC) simulation, using a combination of Gibbs sampling and Metropolis Hastings algorithms. In this study, a number of deaths from respiratory diseases in 81 Provinces of Turkey are used for the illustrative purpose. Each province is defined as a polygon, which is a non-overlapping areal object and some attributes associated with 81 provinces are followed. The distribution of the counts is assumed to come from a Poisson distribution, CARBayes models are applied to data and the disease mapping is performed over calculated risk values.

Keywords: *Spatial autocorrelation, CAR models, CARBayes, MCMC, Respiratory disease.*

References

[1] Bivand, S.R., Pebesma, E. and Gόmez-Rubio, V. (2013), *Applied spatial data analysis with R*, Second Edition, New York, Springer, 405.

[2] Lee, D. (2013), *CARBayes: An R package for Bayesian spatial modelling with conditional autoregressive priors*, Journal of Statistical Software, 55, 13.

[3] Lee, D. (2011), *A comparison of conditional autoregressive models used in Bayesian disease mapping*, Spatial and Spatio-temporal Epidemiology, 2, 79-89.

[4] Lee, D., Ferguson, C. and Mitchell, R. (2009), *Air pollution and health in Scothland: a multi-city study*, Biostatistics,10, 409-423.

[5] Leroux, B., Lei, X. and Breslow, N. (1999), *Estimation of Disease rates in small areas: a new mixed model for spatial dependence*, In: Halloran M., Berry, D. editors, Statistical models in epidemiology, the environment and clinical trials, New York, Springer-Verlag, 179-191.

ISC2017 INTERNATIONAL STATISTICS CONGRESS
December 6-8, 2017 ANKARA/TURKEY

Joint Modelling of Location, Scale and Skewness Parameters of the Skew Laplace Normal Distribution

Fatma Zehra DOĞRU[1], Olcay ARSLAN[2]
fatma.dogru@giresun.edu.tr, oarslan@ankara.edu.tr

[1]*Giresun University, Giresun, Turkey*
[2] *Ankara University, Ankara, Turkey*

The skew Laplace normal (SLN) distribution was proposed by [4] which has wider range of skewness and also more applicable than the skew normal (SN) distribution [1,2]. The advantage of the SLN distribution is that it has the same number of parameters as the SN distribution and also it shows heavier tail behavior than the SN distribution. In this study, we consider the following joint location, scale and skewness models of the SLN distribution

$$\begin{cases} y_i \sim SLN(\mu_i, \sigma_i^2, \lambda_i), \;\; i = 1,2,\dots,n \\ \mu_i = \boldsymbol{x}_i^T \boldsymbol{\beta}\,, \\ \log \sigma_i^2 = \boldsymbol{z}_i^T \boldsymbol{\gamma}\,, \\ \lambda_i = \boldsymbol{w}_i^T \boldsymbol{\alpha}\,, \end{cases} \tag{1}$$

where y_i is the ith observed response, $\boldsymbol{x}_i = (x_{i1}, \dots, x_{ip})^T$, $\boldsymbol{z}_i = (z_{i1}, \dots, z_{iq})^T$ and $\boldsymbol{w}_i = (w_{i1}, \dots, w_{ir})^T$ are observed covariates corresponding to y_i, $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^T$ is a $p \times 1$ vector of unknown parameters in the location model, and $\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_q)^T$ is a $q \times 1$ vector of unknown parameters in the scale model and $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_r)^T$ is a $r \times 1$ vector of unknown parameters in the skewness model. These covariate vectors $\boldsymbol{x}_i, \boldsymbol{z}_i$ and $\boldsymbol{w}_i$ are not needed to be identical. We introduce joint modelling location, scale and skewness models of the SLN distribution as an alternative model for joint modelling location, scale and skewness models of the SN distribution proposed by [5] when the data set includes both asymmetric and heavy-tailed observations. We get the maximum likelihood (ML) estimators for the parameters of joint location, scale and skewness models of SLN distribution using the expectation-maximization (EM) algorithm [3]. The performance of the proposed estimators are demonstrated by a simulation study and a real data example.

Keywords: *EM algorithm, Joint location, scale and skewness models, ML, SLN, SN.*

References

[1] Azzalini, A. (1985), *A class of distributions which includes the normal ones*, Scandinavian Journal of Statistics, 12(2), 171-178.

[2] Azzalini, A. (1986), *Further results on a class of distributions which includes the normal ones*, Statistica, 46(2), 199-208.

[3] Dempster, A.P., Laird, N.M. and Rubin, D.B. (1977), *Maximum likelihood from incomplete data via the EM algorithm*, Journal of the Royal Statistical Society, Series B, 39, 1-38.

[4] Gómez, H.W., Venegas, O. and Bolfarine, H. (2007), *Skew-symmetric distributions generated by the distribution function of the normal distribution*, Environmetrics, 18, 395-407.

[5] Li, H., Wu, L. and Ma, T. (2017), *Variable selection in joint location, scale and skewness models of the skew-normal distribution*, Journal of Systems Science and Complexity, 30(3), 694-709.

Artificial Neural Networks based Cross-entropy and Fuzzy relations for Individual Credit Approval Process

Damla ILTER[1], Ozan KOCADAGLI[1]
damla.ilter@msgsu.edu.tr, ozan.kocadagli@msgsu.edu.tr

[1] *Mimar Sinan Fine Arts University, Istanbul, Turkey*

Credit scoring has continued its popularity in the financial sector for the last few decades, because the number of credit applicants is growing day by day depending on many economic factors. This fact prompts the financial institutions to handle this issue more accurately. Therefore, improving the efficient evaluating procedures is inevitable to overcome the systematic and non-systematic errors that inherently included in the decision process. In the context of individual credit applications, the financial institutes are generally interested in the financial histories of their customers as well as many economic indicators. To make a true decision about whether credit application is worthy or not to approve in the evaluating process, the analysts mostly utilize the decision support systems based the statistical, machine learning, artificial intelligence techniques, etc. In this study, the efficient evaluation procedure that comprises artificial neural networks (ANNs) with cross-entropy and fuzzy relations is proposed. In the implementations, the proposed procedure is applied to Australian and German of benchmark credit scoring data sets and its performance is compared with traditional approaches in terms of evaluation performance and robustness.

Keywords: *Credit Scoring, Artificial Neural Networks, Fuzzy Relations, Cross-entropy, Gradient based Algorithms.*

References

[1] Abdou, H., Pointon, J., El-Masry, A. (2008). Neural nets versus conventional techniques in credit scoring in Egyptian banking. Expert systems with applications, 35, 1275-1292.

[2] Bozdogan, H. (2000). *Akaike's information criterion and recent developments in information complexity*. Journal of mathematical psychology, 44(1), 62-91.

[3] Gorzalczany, M., B. and Rudzinski, F., (2016). *A multi-objective genetic optimization for fast, fuzzy rule-based credit classification with balanced accuracy and interpretability.* Applied Soft Computing, 40, 206220. doi:10.1016/j.asoc.2015.11.037.

[4] Kocadagli, O. (2015). *A Novel Hybrid Learning Algorithm For Full Bayesian Approach of Artificial Neural Networks*, Applied Soft Computing, Elsevier, 35, 1 – 958.

[5] Kocadagli, O. and Langari, R., (2017). *Classification of EEG signals for epileptic seizures using hybrid artificial neural networks based wavelet transforms and fuzzy relations*, Science Direct, 88, 419-434.

10TH ISC2017 INTERNATIONAL STATISTICS CONGRESS
December 6-8, 2017 ANKARA/TURKEY

Estimators of the Censored Regression in the Cases of Heteroscedasticity and Non-Normality

Ismail YENILMEZ[1], Yeliz MERT KANTAR[1]
ismailyenilmez@anadolu.edu.tr, ymert@anadolu.edu.tr

[1] *Department of Statistics, Faculty of Science, Anadolu University, Eskisehir, Turkey*

In the regression model, the dependent variable is restricted in certain ways. These variables, which are referred to as limited dependent variables, can be classified into three categories: i. Truncated regression models, ii. Censored regression models and iii. Dummy endogenous models. In this study, censoring scheme to the left of zero has been examined to determine the frame, particularly. In linear regression, ordinary least squares (OLS) estimates are biased and inconsistent when the dependent variable is censored. To solve a part of this problem, the classical estimation method for censored variable (Tobin's censored normal regression estimator or maximum likelihood estimation for censored normal regression – hereafter, Tobit), was proposed by [4]. However, several potential misspecifications cause inconsistency for the Tobit. Such misspecifications include heteroskedasticity [1] and an incorrect normal assumption [2]. In literature, the partially adaptive estimator (PAE) based on flexible probability density function are used for a comparison with other estimators used in censored regression in case of heteroscedasticity and non-normality [3]. In this study, Tobit and a PAE based on the generalized normal distribution (PAE_{GND}) which is introduced by [5] are examined for the censored regression in the presence of both heteroscedasticity and non-normality. A simulation study is used to analyze the OLS, Tobit and PAE_{GND} estimators' relative performance in the case of different error distributions and the presence of heteroscedasticity. A Monte Carlo study is conducted to compare the considered estimators. The results of the study show that the considered partially adaptive estimator performs better than the Tobit in the cases of non-normal error distributions and it is less sensitive to the presence of heteroscedasticity.

Keywords: *Censored regression model, Partially adaptive estimator, Tobit model, Heteroscedasticity, Non-Normality.*

References

[1] Arabmazar, A. and Schmidt, P. (1981), *Further evidence on the robustness of the Tobit estimator to heteroskedasticity*, Journal of Econometrics 17, 253-258.

[2] Arabmazar, A. and Schmidt, P. (1982), *An investigation of the robustness of the Tobit estimator to non-normality*, Econometrica 50, 1055-1063.

[3] Mcdonald, J.B. and Nguyen, H. (2015), *Heteroscedasticity and Distributional Assumptions in the Censored Regression Model*, Communications in Statistics—Simulation and Computation, 44: 2151–2168.

[4] Tobin, J. (1958), *Estimation of relationships for limited dependent variables*, Econometrica: Journal of the Econometric Society, 24-36.

[5] Yenilmez, I. and Kantar, Y.M. (2017), *A partially adaptive estimator for the censored regression model based on generalized normal distribution*. 3rd International Research, Statisticians and Young Statisticians Congress.

10TH ISC2017 INTERNATIONAL STATISTICS CONGRESS
December 6-8, 2017 ANKARA/TURKEY

Functional Modelling of Remote Sensing Data

Nihan ACAR-DENIZLI [1], Pedro DELICADO [2], Gülay BAŞARIR[1] and Isabel CABALLERO[3]
nihan.acar@msgsu.edu.tr, pedro.delicado@upc.edu, gulay.basarir@msgsu.edu.tr,
isabel.caballero@noaa.gov

[1]*Mimar Sinan Güzel Sanatlar Üniversitesi, Istanbul, Turkey*
[2]*Universitat Politecnico de Catalunya, Barcelona, Spain*
[3] *NOAA National Ocean Service, Silver Spring, USA*

Functional models are used to analyse data defined on a continuum such as dense time interval or space [1]. They consider the continuous structure of the data and have many advantages comparing to the ordinary statistical models [2]. In this paper, the spectral data collected from remote sensors were handled as functional data and the concentration of Total Suspended Solids (TSS) regarding to the area Guadalquivir estuary has been predicted on Remote Sensing (RS) data obtained from Medium Resolution Imaging Spectrometer (MERIS) by using various functional models as an alternative to other statistical models. The predictive performances of the related models were compared in terms of their prediction errors computed based on cross validation in a simulation study. The results show that functional linear models predict the relevant characteristics better on RS data.

Keywords: *functional linear models, functional principal component regression, functional partial least squares regression, remote sensing data.*

References

[1] Acar-Denizli, N., Delicado, P., Başarır G. and Caballero I. (2017), *Functional linear regression models for scalar responses on remote sensing data: an application to Oceanography*. In *Functional Statistics and Related Fields*, Springer, Cham, 15-21.

[2] Ramsay, J.O. and Silverman, B.W. (2005), *Functional Data Analysis*, USA, Springer.

SESSION II

APPLIED STATISTICS II

ISC2017 INTERNATIONAL STATISTICS CONGRESS
December 6-8, 2017 ANKARA/TURKEY

Estimation for the Censored Regression Model with the Jones and Faddy's Skew t Distribution: Maximum Likelihood and Modified Maximum Likelihood Estimation Methods

Sukru ACITAS[1], Birdal SENOGLU[2], Yeliz MERT KANTAR[1], Ismail YENILMEZ[1]
sacitas@anadolu.edu.tr, senoglu@science.ankara.edu.tr, ymert@anadolu.edu.tr, ismailyenilmez@anadolu.edu.tr

[1]*Department of Statistics, Faculty of Science, Anadolu University, Eskisehir, Turkey*
[2] *Department of Statistics, Faculty of Science, Ankara University, Ankara, Turkey*

The ordinary least squares (OLS) estimators are biased and inconsistent in the context of censored regression model. For this reason, Tobit estimators are mostly utilized in estimating the model parameters, see [5]. Tobit estimators are obtained via maximum likelihood (ML) method under the assumption of normality. It is clear that they give inefficient estimators when the normality assumption is not satisfied. Therefore, different error distributions for the censored regression model are considered to accommodate skewness and/or kurtosis, see for example [3]. In this study, we assume that the error terms have Jones and Faddy's skew t (JFST) distribution in the censored regression model. JFST distribution covers a wide range of skew and symmetric distributions and nests well-known Student's t and normal distribution as special and limiting cases, respectively [2]. These properties makes JFST distribution an attractive alternative to normal distribution. In the estimation part of the study, modified maximum likelihood (MML) methodology, introduced by [4], is used, see also [1] in the context of generalized logistic error distribution case. The MML method is easy to implement since it provides the explicit forms of the estimators. The MML estimators are also asymptotically equivalent to the ML estimators and robust to outlying observations. A Monte-Carlo Simulation study is conducted for comparing the performances of the MML estimators with some existing estimators used for censored regression model. The results of the simulation study show that MML estimators work well among the others with respect to mean square error (MSE) criterion.

Keywords: *Censored regression model, maximum likelihood, modified maximum likelihood, efficiency*.

References

[1] Acitas, S, Yenilmez I., Senoglu, B. and Kantar Y.M. (2017), *Modified Maximum Likelihood Estimation for the Censored Regression Model*. The 13th IMT-GT International Conference on Mathematics, Statistics and Their Applications, 4th-7th December 2017, Sintok, Kedah, Malaysia, (Accepted for oral presentation).

[2] Jones, M.C. and Faddy, M.J. (2003), *A skew extention of the t-distribution, with applications*. J.R. Stat. Soc. Ser. B 65, 159-175.

[3] McDonald, J. B. and Xu, Y. J. (1996), *A comparison of semi-parametric and partially adaptive estimators of the censored regression model with possibly skewed and leptokurtic error distributions.* Economics Letter, 51(2), 153-159

[4] Tiku, M. L. (1967), *Estimating the mean and standard deviation from a censored normal sample.* Biometrika, 54, 155-165.

[5] Tobin, J. (1958), *Estimation of relationships for limited dependent variables.* Econometrica: Journal of the Econometric Society, 24-36.

10TH ISC2017 INTERNATIONAL STATISTICS CONGRESS
December 6-8, 2017 ANKARA/TURKEY

Scale Mixture Extension of the Maxwell Distribution: Properties, Estimation and Application

Sukru ACITAS[1], Talha ARSLAN[2], Birdal SENOGLU[3]
sacitas@anadolu.edu.tr, mtarslan@ogu.edu.tr, senoglu@science.ankara.edu.tr

[1]*Department of Statistics, Faculty of Science, Anadolu University, Eskisehir, Turkey*
[2] *Department of Statistics, Faculty of Science, Eskisehir Osmangazi University, Eskisehir, Turkey*
[3] *Department of Statistics, Faculty of Science, Ankara University, Ankara, Turkey*

In this study, we introduce scale mixture extension of the Maxwell distribution. It is defined by the quotient of two independent random variables, namely a Maxwell distribution in the numerator and the power of a Uniform(0,1) distribution in the denominator, see for example [1]. Therefore, the resulting distribution is called as slashed Maxwell. The moments, skewness and kurtosis measures of slashed Maxwell distribution are derived. The maximum likelihood (ML) method is utilized to estimate the location and the scale parameters. The explicit forms of ML estimators cannot be obtained because of the nonlinear functions in the likelihood equations. Therefore, we use Tiku's [2, 3] modified maximum likelihood (MML) methodology in the estimation process. The MML estimators have closed forms since they are expressed as the function of the sample observations. Therefore, they are easy to compute besides being efficient and robust to outlying observations. A real life data is modelled using slashed Maxwell distribution at the end of the study.

Keywords: *Maxwell distribution, slash distribution, kurtosis, modified likelihood, robustness.*

References

[1] Rogers W.H. and Tukey J.W. (1972), *Understanding some long-tailed symmetrical distributions*. Statist. Neerlandica, 26, 211–226.

[2] Tiku, M. L. (1967), *Estimating the mean and standard deviation from a censored normal sample.* Biometrika, 54, 155-165.

[3] Tiku, M. L. (1968), *Estimating the parameters of normal and logistic distributions from censored samples.* Australian Journal of Statistics, 10, 64-74.

10TH ISC2017 INTERNATIONAL STATISTICS CONGRESS
December 6-8, 2017 ANKARA/TURKEY

Maximum Likelihood Estimation Using Genetic Algorithm for the Parameters of Skew-t Distribution under Type II Censoring

Abdullah YALÇINKAYA[1], Ufuk YOLCU[2], Birdal ŞENOĞLU[1]
ayalcinkaya@ankara.edu.tr, varyansx@hotmail.com, senoglu@science.ankara.edu.tr

[1] *Ankara University Department of Statistics, Ankara, Turkey*
[2] *Giresun University Department of Econometrics, Giresun, Turkey*

Skew-t (St), an Azzalini type skew extension of the well-known Student's t distribution, provides flexibility for modelling data sets having skewness and heavy tails, see [1]. Type II censoring is one of the most commonly used type of censoring schemes. It occurs when the smallest r_1 and the largest r_2 units in a sample of size n are not observed. In this study, our aim is to obtain the estimates of the parameters of the St distribution under type II censoring. For this purpose, we use the well-known and widely used Maximum Likelihood (ML) methodology. However, ML estimators of the unknown model parameters do not have closed forms, in other words, they cannot be obtained as explicit functions of the sample observations. We therefore resort to numerical methods. Among these methods, Genetic Algorithm (GA), a popular search technique popularized by [3], is preferred to use. Different than the earlier studies, we benefit from the robust confidence intervals (CIs) to identify the search space of GA, see [5]. In constructing the CIs, Modified Maximum Likelihood (MML) estimators of the parameters are utilized, see [4] for details. Maximum Product Spacing (MPS) which is a powerful and useful method for estimating the unknown distribution parameters is also used, see [2]. We compare the efficiencies of the ML estimators using GA, ML estimators using Nelder-Mead (NM) algorithm and MPS estimators via an extensive Monte Carlo simulation study for different parameter settings, sample sizes and censoring schemes. Finally, we presented a real life example for illustrative purpose.

Keywords: *Skew-t distribution, type II censoring, genetic algorithm, modified maximum likelihood, maximum product spacing*

References

[1] Azzalini, A. (1985), *A class of distributions which includes the normal ones*, Scand. J. Stat., 12, pp. 171-178.

[2] Cheng, R.C.H. and Amin, N.A.K. (1983), *Estimating parameters in continuous univariate distributions with a shifted origin*, Journal of the Royal Statistical Society. Series B (Methodological), pp. 394-403.

[3] Holland, J. (1975), *Adaptation in Natural and Artificial System: an Introduction with Application to Biology, Control and Artificial Intelligence*, Ann Arbor, University of Michigan Press.

[4] Tiku, M.L. (1967), *Estimating the mean and standard deviation from censored normal samples*, Biometrika, 54, pp. 155-165.

[5] Yalçınkaya, A., Şenoğlu, B. and Yolcu, U. (2017), *Maximum likelihood estimation for the parameters of skew normal distribution using genetic algorithm,* Swarm and Evolutionary Computation, http://doi.org/10.1016/j.swevo.2017.07.007.

10TH ISC2017 INTERNATIONAL STATISTICS CONGRESS
December 6-8, 2017 ANKARA/TURKEY

Robust Two-way ANOVA under nonnormality

Nuri Celik[1], Birdal Senoglu[2]
ncelik@bartin.edu.tr, senoglu@science.ankara.edu.tr

[1]*Bartin University, Department of Statistics, 74100, Bartin, Turkey*
[2]*Ankara University, Department of Statistics, 06600, Ankara, Turkey*

It is generally assumed that the error terms in two-way ANOVA are normally distributed with mean zero and the common variance σ^2. Least Squares (LS) methodology is used in order to obtain the estimators of the unknown model parameters. However, when the normality assumption is not satisfied, LS estimators of the parameters lose efficiency and the powers of the test statistics based on them decline rapidly. In this study, we assume the distribution of the error terms in two-way ANOVA as Azzalini's skew normal (SN) (Azzalini, 1985), see Celik (2012) and Celik et. al (2015) in the context of one-way ANOVA. We use maximum likelihood (ML) and the modified maximum likelihood (MML) methodologies to obtain the estimators of the parameters of interest, see Tiku (1967). We also propose new test statistics based on these estimators. The performances of the proposed estimators and the test statistics based on them are compared with the corresponding normal theory results via Monte Carlo simulation study, see also Celik and Senoglu (2017). A real life data is analyzed at the end of the study to show the implementation of the methodology

Keywords: *Two-way ANOVA, Modified Maximum Likelihood, Skew Normal Distribution, Robustness*

References

[1] Azzalini, A. (1985), *A class of distribution which includes normal ones*, Scandinavian Journal of Statistics, 12, 171-178.

[2] Celik, N. (2012), *ANOVA Modellerinde Çarpık Dağılımlar Kullanılarak Dayanıklı İstatistiksel Sonuç Çıkarımı ve Uygulamaları*, Ankara University, Ph. D. Thesis.

[3] Celik, N., Senoglu, B. and Arslan, O . (2015), *Estimation and Testing in one-way ANOVA when the errors are skew normal*, Colombian Journal of Statistics, 38(1), 75-91.

[4] Celik, N., Senoglu, B. (2017), *Two-way ANOVA when the distribution of error terms is skew t*, Communication in Statistics: Simulation and Computation, in press.

[5] Tiku, M.L, (1967), *Estimating the mean and standard deviation from censored normal samples*, Biometrika, 54, 155-165.

Linear Contrasts for Time Series Data with Non-Normal Innovations: An Application to a Real Life Data

Özgecan YILDIRIM[1], Ceylan YOZGATLIGİL[2], Birdal ŞENOĞLU[3]
Ozgecan.Yildirim@tcmb.gov.tr, ceylan@metu.edu.tr, senoglu@science.ankara.edu.tr

[1] *Central Bank of the Republic of Turkey, Ankara, Turkey*
[2]*Middle East Technical University, Ankara, Turkey*
[3]*Ankara University, Ankara, Turkey*

Yıldırım et al. [5] estimated the model parameters and introduced a test statistic in one-way classification AR(1) model under the assumption of independently and identically distributed (iid) error terms having Student's t distribution, see also [4].

In this study, we extend their study to linear contrasts which is a well-known and widely used comparison method when the null hypothesis about the equality of the treatment means is rejected, see [3], [4]. See also [1] and [2] in the context of ANOVA. A test statistic for the linear contrasts is introduced. A comprehensive simulation study is done to compare the performance of the test statistic with the corresponding normal theory test statistic. At the end of the study, a real life data is analysed to show the implementation of the introduced test statistic.

Keywords: *Linear contrasts, One-Way ANOVA, AR(1) model, Student's t distribution*

References

[1] Lund, R., Liu, G. and Shao, Q. (2016), *A new approach to ANOVA methods for autocorrelated data*, The American Statistician, 70(1), 55-62.

[2] Pavur, R. J. and Lewis, T. O. (1982), *Test procedure for the analysis of experimental designs with correlated nonnormal data*, Communications in Statistics-Theory and Methods, 11(20), 2315-2334.

[3] Şenoglu, B. and Bayrak, Ö. T. (2016), *Linear contrasts in one-way classification AR(1) model with gamma innovations*, Hacettepe Journal of Mathematics and Statistics, 45(6), 1743-1754.

[4]Yıldırım, Ö. (2017), *One-way ANOVA for time series data with non-normal innovations: An application to a real life data* (Master's thesis), Middle East Technical University, Ankara, Turkey.

[5]Yıldırım, Ö., Yozgatlıgil, C. and Şenoğlu, B. (2017), *Hypothesis testing in one-way classification AR(1) model with Student's t innovations: An application to a real life data*, 3rd International Researchers, Statisticians and Young Statisticians Congress (IRSYSC), p.272.

SESSION II

APPLIED STATISTICS III

ISC2017 INTERNATIONAL STATISTICS CONGRESS
December 6-8, 2017 ANKARA/TURKEY

Comparison of the Lord's χ^2 Statistic and Raju's Area Measurements Methods in Determination of the Differential Item Function

Burcu HASANÇEBİ[1], Yüksel TERZİ[2], Zafer KÜÇÜK[1]
burcuhsncb@gmail.com, yukselt@omu.edu.tr, zkucuk@ktu.edu.tr

[1]*Karadeniz Technical University, Trabzon, Turkey*
[2]*Ondokuz Mayıs University, Samsun, Turkey*

Test development process consists of numerous procedures and steps. Determining the validity of test is the most important of these procedures and steps. Determining the test and item bias is among the techniques for determining the validity of the test. When subjects who same ability level (θ), but come from different subgroups, the existence of item bias can be mentioned. A biased item contains the Differential Item Function (DIF). The important point here is that the DIF is not a proof of item bias. The difference in the answers to an item is a situation that should happen when the subgroups are due to differences in their ability levels. This is the validity and unbiassed of the item, which is to be expected. If a test is to be applied to a heterogeneous population, the bias analysis becomes the most important component of the item selection process. Because the most important criterion for the researcher here is to get the most fair and accurate results for subjects who come from different subgroups and take the test. In this study, literacy levels of probability theory of 3rd and 4th grade students department of Statistics and Computer Science of Karadeniz Technical University were measured. A literacy test with 20 questions was administered to all 3rd and 4th grade students. The obtained data were converted into a binary data set. Bias analysis was conducted according to the gender of the students and the class of the students according to the responses received. When the bias study was conducted, it was analyzed whether the items have differential item function. To perform differential item function (DIF) analysis, Raju's area measurements and the Lord's χ^2 test were used for methods based on Item Response Theory. R software was used in the analysis of differential item function. Experts were consulted for the items for which the differential item function was determined as the result of the analysis. As a result, in some test items determined according to the expert opinion, there was bias according to the gender and class level variables of the 3rd and 4th grade students.

Keywords: *Item Bias, Differential Item Function, Lord's Chi-square, Raju's Area Measurement*

References

[1] McLaughlin, M.E. and Drasgow, F. (1987), *Lord's chi-square test of item bias with estimated and with known person parameters,* Applied Psychological Measurement, 11, 161-173.

[2] Lord, F.M. (1980), *Applications of item response theory to practical testing problems,* Hillside, NJ: Erlbaum.

[3] Raju, N.S. (1990), *Determining the significance of estimated signed and unsigned areas between two item response functions,* Applied Psychological Measurement, 14, 197-207.

[4] Raju, N.S. (1988), *The area between two item characteristic curves,* Psychometrika, 53, 495-502.

On Suitable Copula Selection for Tempeature Measurement Data

Ayşe METİN KARAKAŞ[1], Mine DOĞAN[1], Elçin SEZGİN[1]
aysekarakas5767@gmail.com, mined685@gmail.com, esezgin@beu.edu.tr

[1]*Bitlis Eren University Depeartment of Statistcs,Bitlis, Turkey*

In this paper, we model the dependence structure between random variables by using copula functions. In connection with this, we define basic properties of copulas, goodness of fit test and their nonparametric methods. The aim of this article is to obtain selected suitable copula function for tempeature measurement data set that is daily maximum and minimum temperatures of Bitlis between 2012-2017 years. For dependence structures of the data set, we calculated Kendall Tau and Spearman Rho values which are nonparametric. Based on this method, parameters of copula are obtained. To explain the relationship between the variables, copula families are used and these are Gumbel, Clayton, Frank, Cuadras Auge, Joe and Placket copula. With he help of nonparametric estimation of copula parameters, Kolmogorov Smirnov test which is goodness of fit test, Maximum likelhood method and Akaike information Criteria, Schwartz information criteria, we find the suitable Archimedean copula family for this data set.

Keywords: *Copula functions, Kendall Tau, Spearman Rho, Maximum likelihood method, goodness of fit test, Akaike information criteria Schwartz information criteria* **.**

References

[1] Genest, C., Rémillard, B., & Beaudoin, D. (2009). Goodness-of-fit tests for copulas: A review and a power study. *Insurance: Mathematics and economics*, *44*(2), 199-213.

[2] Genest, C., & Rémillard, B. (2008). Validity of the parametric bootstrap for goodness-of-fit testing in semiparametric models. In *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques* (Vol. 44, No. 6, pp. 1096-1127). Institut Henri Poincaré.

[3] Genest, C., & Favre, A. C. (2007). Everything you always wanted to know about copula modeling but were afraid to ask. *Journal of hydrologic engineering*, *12*(4), 347-368.

[4] Massey Jr, F. J. (1951). The Kolmogorov-Smirnov test for goodness of fit. *Journal of the American statistical Association*, *46*(253), 68-78

ISC2017 INTERNATIONAL STATISTICS CONGRESS
December 6-8, 2017 ANKARA/TURKEY

Variable Selection in Polynomial Regression and a Model of Minimum Temperature in Turkey

Onur TOKA[1], Aydın ERAR[2], Meral ÇETİN[1]
onur.toka@hacettepe.edu.tr, aydin.erar@msgsu.edu.tr, meral@hacettepe.edu.tr

[1] *Hacettepe University, Faculty of Science, Department of Statistics, Ankara, TURKEY*
[2] *Mimar Sinan Fine Arts University, Department of Statistics, İstanbul, TURKEY*

Existence of many exponent and/or interaction terms in Polynomial Regression causes some troubles in modeling, especially with observed data. One of them is hierarchy problem. Non-hierarchical patterns of classical and variable selection against hierarchical ones will be investigated to obtain best subset model(s) for the minimum temperature.

In this study, the variable selection criteria were compared by relating average minimum temperature in January with latitude, longitude and altitude in Turkey. It was obtained the best model(s) by using the hierarchical and variable selection procedures in polynomial regression. It was compared both hierarchical variable selection procedures with classical ones. In addition, the best subset model of minimum temperature in Turkey was given for January.

Keywords: *variable selection, polynomial regression, outliers, minimum temperature*

References

[1] Cornell, J. A., and Montgomery, D. C., (1996), *Fitting models to data: Interaction versus polynomial? your choice*, Communications in Statistics--Theory and Methods, *25*(11), 2531-2555.

[2] Çetin, M. and Erar, A. (2006), *A simulation study on classic and robust variable selection in linear regression*, Applied Mathematics and Computation, *175*(2), 1629-1643.

[3] Erar, A.,(2001), *Dilemma of Hierarchical and Classical Variable Selection in Polynomial Regression and Modelling of Average January Minimum Temperature in Turkey*, Hacettepe Bulletin of Naturel Sciences and Engineering, Series B Mathemetics and Statistics, 30, 97-114.

[4] Peixoto, J. L. (1987), *Hierarchical variable selection in polynomial regression models*, The American Statistician, 41(4), 311-313.

[5] Ronchetti, E. (1985), *Robust model selection in regression*, Statistics & Probability Letters, *3*(1), 21-23.

For Raeigly Distribution Simulation with the Help of Kendall Distribution Function Archimedean Copula Parameter Estimation

Ayşe METİN KARAKAŞ[1], Elçin SEZGİN[1], Mine DOĞAN[1]
aysekarakas5767@gmail.com,esezgin@beu.edu.tr,mined685@gmail.com

[1]*Bitlis Eren University Depeartment of Statistcs,Bitlis, Turkey*

In this paper, we model the dependence structure between random variables that we generated dependent Raeighly distrubiton using Archimedean copula and Kendall distribution function. In connection with this, we define basic properties of copulas and their nonparametric method. The aim of Kendall distribution function is selected suitable copula function for using data set. For dependence structures of the data set, we calculated Kendall Tau and Spearman Rho values which are nonparametric. Based on this method, parameters of copula are obtained. To explain the relationship between the variables, three Archimedean copula families were used; Gumbel, Clayton and Frank. Nonparametric estimation of copula parameters and we find the suitable Archimedean copula family for this data set.

Keywords: *Copula functions; Kendall Tau, Kendall Distribution Function; Raeighly distribution.*

References

[1] Cherubini U, Luciano E. (2013), *Value-at-risk trade-off and capital allocation with copulas. Economic Notes*, vol. 30, pp. 235-256

[2] Fang, Hong-Bin, Kai-Tai Fang, and Samuel Kotz. (2002). *The meta-elliptical distributions with given marginals*. Journal of Multivariate Analysis vol.82, no., pp. 11-16

[3] Frees EW, Valdez EA. (1998). *Understanding relationships using copulas*. North American Actuarial Journal, vol. 2, pp. 1-25.

[4] Genest C, MacKay J. (1986). *The joy of copulas: bivariate distrubitons with uniform marginal*. The American Statisticien, vol. 40, pp. 280-283.

HIV-1 Protease Cleavage Site Prediction Using a New Encoding Scheme Based on Physicochemical Properties

Metin YANGIN[1], Bilge BAŞER[1], Ayça ÇAKMAK PEHLİVANLI[1]
metin.yangin@msgsu.edu.tr, bilge.baser@msgsu.edu.tr, ayca.pehlivanli@msgsu.edu.tr

[1]*Mimar Sinan Fine Arts University of Statistics Department, İstanbul, Turkey*

AIDS is a fatal disease of the immune system and one of the major global threat to human health today. According to World Health Organization (WHO), 36.7 million people are estimated to be living with HIV in December 2016 [1]. HIV-1 protease is an essential enzyme for the replication of HIV. It cleaves the proteins to their component peptides and generates an infectious viral particle. The design of HIV-1protease inhibitors represents a new approach to AIDS therapy. For this reason, it is crucial to predict the cleavability of a peptide by HIV-1 protease.

In literature, most of the studies used orthogonal encoding method for representing peptides. In this study, unlike previous works, it is given a new approach for encoding peptides which consists of the means of each physicochemical characteristic (566 properties) values constructed by AAIndex for each peptide in the 1625 dataset [2]. Several preprocessing methods are applied to clean the data and the median filtering produced the most promising approach for preprocessing to reduce the possible noise in the data set. In this study, besides machine learning methods are applied to data set constructed by proposed encoding scheme, it is also compared to the most recent studies published in this area [3]. Since Singh and Su used four different encoding methods using the same peptides set and they applied decision tree, logistic regression and artificial neural network methods, the same scheme were applied to our encoded dataset for the sake of comparison. As a result of the comparisons, it is observed that, proposed approach yields higher accuracy in the prediction of cleavage site. In addition to these comparative results, it is also applied the kernel logistic regression with different kernel functions, random forest and adaboost methods after preprocessing. Consequently, the random forest method gives the best performance in predicting the cleavability.

Keywords: *HIV-1 protease, Cleavage sites classification, Median filtering, Physicochemical properties, Machine learning.*

References

[1] URL: http://www.who.int/hiv/data/en (2016) *Accessed date*: 10/11/2017

[2] URL: http://www.genome.jp/aaindex/ (2017) *Accessed date*: 19/10/2017

[3] Singh, O. and Su, E.C. (2016), *Prediction of HIV-1 protease cleavage site using a combination of sequence, structural, and physicochemical features*, BMC Bioinformatics, BioMed Central, 280-289.

SESSION II

PROBABILITY AND STOCHASTIC PROCESSES

10TH ISC2017 INTERNATIONAL STATISTICS CONGRESS
December 6-8, 2017 ANKARA/TURKEY

Variance Function of Type II Counter Process with Constant Locking Time

Mustafa Hilmi PEKALP[1], Halil AYDOĞDU[1]
mpekalp@ankara.edu.tr, aydogdu@ankara.edu.tr

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

A radioactive source emits particles according to a Poisson process $\{N_1(t), t \geq 0\}$ with rate λ. Let consider a counter that registers the particles emitted from this source and assume that a particle arriving at the counter locks the counter for a constant locking time L. An arriving event is registered if and only if no particle is arrived during the preceding time interval of length L. Consequently, the probability that a particle is registered is $e^{-\lambda L}$. Let define random variables $Y_1, Y_2, \ldots$ as the consecutive times between two particles registered. It can be constituted a registration process $\{N_2(t), t \geq 0\}$ based on these random variables where $N_2(t)$ is number of the particles registered up to time t. It is obvious that $Y_1, Y_2, \ldots$ are independent. While the random variable Y_1 has exponential distribution with mean $\frac{1}{\lambda}$, $Y_i's, i = 2,3, \ldots$ have same distribution but different from Y_1. Hence, the counting process $\{N_2(t), t \geq 0\}$ is delayed renewal process. In literature, this process is called a type II counter process. In this study, we remind some properties of delayed renewal process and obtain the variance function of the type II counter process.

Keywords: *delayed renewal process; type II counter process; variance function.*

References

[1] Acar, Ö. (2004), *Gecikmeli Yenileme Süreçleri ve Bu Süreçlerde Ortalama Değer ve Varyans Fonksiyonlarının Tahmini*, Ankara University Graduate School of Natural and Applied Sciences, Master Thesis, Ankara.

[2] Karlin, S., Taylor, H.M. (1975). *A First Course in Stochastic Processes, Academic Press*, New York.

[3] Parzen, E. (1999), *Stochastic Processes*, Holden-Day Inc., London.

10TH ISC2017 INTERNATIONAL STATISTICS CONGRESS
December 6-8, 2017 ANKARA/TURKEY

Power Series Expansion for the Variance Function of Erlang Geometric Process

Mustafa Hilmi PEKALP[1], Halil AYDOĞDU[1]
mpekalp@ankara.edu.tr, *aydogdu@ankara.edu.tr*

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

Geometric process (GP) is a powerful tool to facilitate modelling of many practical applications such as system reliability, software engineering, maintenance, queueing systems, risk and warranty analysis. Most of these applications require knowledge of the geometric function $M(t)$, the second moment function $M_2(t)$ and the variance function $V(t)$. The geometric function $M(t)$ which cannot be obtained in an analytical form is studied by many researchers [1,2,3,4,5]. Even though there are many studies for the geometric function $M(t)$ in the literature, there are limited number of studies for the variance function $V(t)$. These studies depend on the convolutions of the distribution functions which require complicated calculation to obtain the variance function $V(t)$ [1]. In this study, we consider a simple and useful method for computing the variance function $V(t)$ by assuming the first interarrival time X_1 has Erlang distribution. For this purpose, a power series expansion for the second moment function $M_2(t)$ of the GP is derived by using the integral equation given for $M_2(t)$. Some computational procedures are also considered to compute the variance function $V(t)$ after the calculation of $M_2(t)$.

Keywords *geometric process; variance function; power series; Erlang distribution.*

References

[1]Aydoğdu, H. Altındağ, Ö. (2015), *Computation of the Mean Value and Variance Functions in Geometric Process*, Journal of Statistical Computation and Simulation 86:5, 986-995.

[2]Aydoğdu, H. Karabulut, İ. (2014), *Power Series Expansions for the Distribution and Mean Value Function of a Geometric Process with Weibull Interarrival Times*, Naval Research Logistics 61, 599-603.

[3]Aydoğdu, H. Karabulut, İ. Şen, E. (2013), *On the Exact Distribution and Mean Value Function of a Geometric Process with Exponential Interarrival Times*, Statistics and Probability Letters 83, 2577-2582.

[4]Braun, W.J. Li, W. Zhao, Y.Q. (2005), *Properties of the Geometric and Related Processes*, Naval Research Logistics 52, 607-616.

[5]Lam, Y. (2007), *The Geometric Processes and Its Applications*, World Scientific, Singapore.

ISC2017 INTERNATIONAL STATISTICS CONGRESS
December 6-8, 2017 ANKARA/TURKEY

A Plug-in Estimator for the Lognormal Renewal Function under Progressively Censored Data

Ömer ALTINDAĞ, Halil AYDOĞDU[1]
omeraltindag87@gmail.com, aydogdu@ankara.edu.tr

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

Renewal process is a counting process model which generalizes Poisson process. It is widely used in the fields of applied probability such as reliability theory, inventory theory, queening theory etc. In applications related with the renewal process, its mean value function, so-called renewal function, is required. For example, let's consider a unit that must be renewed with an identical one after it fails. In this situation, the number of renewals for a specified period can be predicted with the renewal function. So, estimation of the renewal function is important for practitioners. Its formal definition is given as follows.

Let $X_1, X_2, \ldots$ be a sequence of independent and identically distributed positive random variables with distribution function F. They represent the successive failure times of identical units. The number of renewals in the interval $(0, t]$ based on the sequence $(X_k)_{k=1,2,\ldots}$ is

$$N(t) = \max\{n: S_n \leq t\}, t \geq 0,$$

where $S_0 = 0$ and $S_n = \sum_{k=1}^{n} X_k, n = 1,2, \ldots$. The process $\{N(t), t \geq 0\}$ is called as renewal process and its mean value function is called as renewal function. Formally, the renewal function is defined as $M(t) = E\big(N(t)\big), t \geq 0$. Here, E denotes the expectation.

Let a realization of the renewal process has been observed and denote the observations by $\{X_1, X_2, \ldots, X_n\}$. Estimation of the renewal function is studied in the literature when $\{X_1, X_2, \ldots, X_n\}$ is complete. For the literature, see Frees [3] and Aydoğdu [2]. However, this is not always the case. The data set $\{X_1, X_2, \ldots, X_n\}$ may include censored observations. Altındağ [1] has studied the estimation problem of the renewal function when the observations are right censored. In this study, estimation of the renewal function is considered when F is lognormal and the observations are progressively censored. A plug-in estimator is introduced and its asymptotic properties are investigated. A Monte Carlo simulation is carried out for small sample performance of the estimator.

Keywords: *renewal process, renewal function, plug-in estimator, progressively censored data, lognormal distribution*

References

[1] Altındağ, Ö. (2017), *Estimation in Renewal Processes under Censored Data*, Ph.D. Thesis, Ankara University, 186.

[2] Aydoğdu, H. (1997), *Estimation in Renewal Processes*, Ph.D. Thesis, Ankara University, 158.

[3] Frees, E.W. (1996), *Warranty analysis and renewal function estimation*, Naval Research Logistics, 33(3), 361-372.

10TH ISC2017 INTERNATIONAL STATISTICS CONGRESS
December 6-8, 2017 ANKARA/TURKEY

Estimation of the Mean Value Function for Weibull Trend Renewal Process

Melike Özlem KARADUMAN[1], Mustafa Hilmi PEKALP[1], Halil AYDOĞDU[1]
melikeozlemm3@gmail.com, mpekalp@ankara.edu.tr, aydogdu@ankara.edu.tr

[1]*Ankara University, Ankara, TURKEY*

A stochastic process $\{N(t), t \geq 0\}$ is called counting process if it counts the number of the events that occur as a function of time. The sequence of interarrival times in accordance with this process uniquely determine the counting process. For example, if the interarrival times are independent and identically distributed random variables with a distribution function F, then the renewal process can be used in modelling of this counting process. However, in many maintenance, replacement applications and some analysis in reliability theory, the data set comes from a counting process includes random variables that alter in some systematic way. Systematic changes mean that there is a trend in the pattern of the data set and the interarrival times are not identically distributed. In such cases, a trend-renewal process (TRP) can be used as a model. The TRP is defined as follows. Let $\{N(t), t \geq 0\}$ be a counting process with the arrival times $S_1, S_2, \ldots$. Suppose that $\lambda(t)$ is a non-negative function and write $\Lambda(t) = \int_0^t \lambda(u)du$. Then, the counting process $\{N(t), t \geq 0\}$ is a $TRP(F, \lambda)$ if $\Lambda(S_1)$, $\Lambda(S_2) - \Lambda(S_1)$, $\Lambda(S_3) - \Lambda(S_2)$, ... are independent and identically distributed with distribution function F. The distribution F is called the renewal distribution, and λ is called the trend function of the TRP.
Let $\{N(t), t \geq 0\}$ be a $TRP(F, \lambda)$. The mean value function of TRP is defined by $M(t) = E\big(N(t)\big), t \geq 0$. Some statistical applications of TRP need knowledge of the mean value function $M(t)$. From the definition of TRP, it follows that $\widetilde{N}(t) = N\big(\Lambda^{-1}(t)\big), t \geq 0$ is a renewal process with the interarrival time distribution function F. Then, it is clear that

$$\widetilde{M}(\Lambda(t)) = M(\mathrm{t}), \quad t \geq 0, \tag{1}$$

where $\widetilde{M}$ is the renewal function of the renewal process $\big\{\widetilde{N}(t), t \geq 0\big\}$.
In this study, we take the distribution F and trend function as Weibull distribution with shape parameter α and scale parameter $\beta = 1/\Gamma(1 + \frac{1}{\alpha})$ and $\lambda(t) = abt^{b-1}, t \geq 0; a, b > 0$, respectively. The parameters α, a and b are estimated based on the data set $\{X_1, \ldots, X_n\}$ which comes from TRP. Then, a parametric estimation $\widehat{M}(t)$ of $M(t)$ for each fixed $t \geq 0$ is proposed based on the estimation of the renewal function $\widetilde{M}(t)$ by using the equation (1). Further, some asymptotic properties of this estimator is investigated and its small sample properties are evaluated by a simulation study.

Keywords: *parameter estimation, Weibull-power-law trend-renewal process, mean value function, trend function*

References

[1] Gamiz, M.L. , Kulasekera, K.B., Liminios, N. and Lindqvist B. H., (2011), *Applied Nonparametric Statistics in Reliability*, New York, Springer, 96-100.

[2] Jokiel-Rokita, A. and Magiera, R., (2012), *Estimation of the Parameters for Trend-renewal Processes* , Stat Comput, 22, 625-637.

[3] Franz, J., Jokiel-Rokita, A. and Magiera, R., (2014), *Prediction in Trend-renewal Processes for Repairable Systems* , Stat Comput, 24, 633-649.

10TH ISC2017 INTERNATIONAL STATISTICS CONGRESS
December 6-8, 2017 ANKARA/TURKEY

First Moment Approximations for Order Statistics from Normal Distribution

Asuman YILMAZ[1] Mahmut KARA[1]
asumanduva@yyu.edu.tr, mkara2581@gmail.com

[1]Faculty of Science, Department of Statistics, Yuzuncu Yıl University, Van, Turkey

Let $X_1, X_2, ..., X_n$ be a random sample of size *n* from the normal distribution and $X_{(1:n)} \le X_{(2:n)} \le ... \le X_{(n:n)}$ be the order statistics obtained by arranging the *n* variables $X_{(i)}$, $i = 1, 2, ..., n$ in ascending order.The probability density function of the *ith* order statistics of sample size *n* for normal distribution is ;

$$f(x) = \frac{n!}{(i-1)!(n-i)!} * [F(x)]^{i-1} * [1 - F(x)]^{n-i} f(x) \qquad -\infty < x < \infty \qquad (1)$$

The expected value of the *ith* order statistics of sample of size *n* from normal distribution is as follow:

$$E(X_{(i)}) = \frac{n!}{(r-1)!(n-r)!} \int_{-\infty}^{+\infty} x * [1 - \Phi(x)]^{r-1} (\Phi(x))^{n-r} * \varphi(x) d(x) \qquad (2)$$

A well known approximation for $E(X_{i:n})$ for sufficiently large *n* is provided by

$$E(X_i : n) \approx F^{-1}\left(\frac{i}{n+1}\right), \qquad \alpha, \beta \le 1 \qquad (3)$$

Here , F^{-1} is inverse oft he cumulative distriburion function of *X*.
To select values of the parameters α and β we use the method of least squares to minimize the squared difference betwen the expected values of order statistics and the approximation given

$$Q(\alpha, \beta) = \sum_{i=1}^{n} \left[M_i - \Phi^{-1}\left(\frac{i - \alpha}{n - 2\alpha + 1} \right) \right]^2 \qquad (4)$$

Here, M_i represent the expected values and it is aimed to obtain the smallest value of Q through equation (4). In the literature Filliben, Vogel, Gringorten and Blom proposed different approaches for calculating the expected value of the *ith* statistic from the normal distribution. Also in this study we proposed two new method through the estimation of the α and β parameter for approximate expressions of the first moment of Normal distribution.

Keywords: *Order statistics, Normal distribution, Expected value, Approximation*

References

[1] Fard P. N. M., 2006. First Moment Approximations For Order Statistics From the Extreme [2]Value Distribution, Statistical Methodology. Vol. 2007, p. 196-203.

[2]Lieblein J., 1953. On the Exact Evaluation of the Variances and Covariances of Order Statistics in Samples From the Extreme Value Distribution . vol.24, p.282-287.

[3]ROYSTON P. J., 1982. Expected Normal Order Statistics (Exact and Approximate), Journal of the Royal Statistical Society, vol. 31, p.161-165.

SESSION II

MODELING AND SIMULATION I

A New Compounded Lifetime Distribution

Sibel ACIK KEMALOGLU[1], Mehmet YILMAZ[1]
acik@ankara.edu.tr, mehmetyilmaz@ankara.edu.tr

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

In this paper we introduce a new lifetime distribution with decreasing hazard rate by compounding exponential and discrete Lindley distributions which is named Exponential Discrete Lindley (EDL) distribution. In this context, we propose and improve the statistical properties of the proposed distribution and show that it is suitable to use this distribution for reliability analysis. Some statistical properties such as probability density function and hazard rate function, moments, moment generating function, Rènyi entropy are given in the study. In addition, parameter estimation by using the maximum likelihood method and EM algorithm are presented. Finally, applications on real data sets are presented to show the feasibility and usefulness of the distribution.

Keywords: *lifetime distribution, hazard rate function, EM algorithm*

References

[1] Adamidis K., Loukas S. (1998), *A lifetime distribution with decreasing failure rate*, Statistical Probability Letter 39(1), 35–42.

[2] Gómez-Déniz E., Calderín-Ojeda E. (2011), *The discrete Lindley distribution: properties and applications*, Journal of Statistical Compututaion and Simulation, 81(11):1405–1416.

[3] Rényi, A. 1961, *On measures of entropy and information*, University of California Press, Berkeley. Proc. Fourth Berkeley Symp. on Math. Statist. and Prob. 1:547–561.

[4] Yilmaz, M., Hameldarbandi, M., and Kemaloglu, S. A. (2016), *Exponential-modified discrete Lindley distribution*, *SpringerPlus*, *5*(1), 1660.

A New Modified Transmuted Distribution Family

Mehmet YILMAZ[1], Sibel ACIK KEMALOGLU[2]
mehmetyilmaz@ankara.edu.tr, acik@ankara.edu.tr

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

In this paper, a new transmutation is proposed by modifying the rank transmutation. The range of the transmutation parameter is extended from the interval $[-1,1]$ to the interval $[-1,2]$ with this rank transmutation. Thus, the concerned distribution gets more flexible. This transmutation allows to generate two new distribution families. Some statistical and reliability properties of these families such as probability density function, moments, survival function and hazard rate function are obtained in the study. Applications on real data sets are presented to see the performance of the distribution families. In particular, the results of the second data set show that extending the range of the transmutation parameter is useful for modeling data.

Keywords: *quadratic rank transmutation, modified rank transmutation, transmuted distribution*

References

[1] Abd El Hady, N. E. (2014), *Exponentiated Transmuted Weibull Distribution*, International Journal of Mathematical, Computational, Statistical, Natural and Physical Engineering, 8(6).

[2] Das, K. K. and Barman, L. (2015), *On some generalized transmuted distributions*, Int. J. Sci. Eng. Res, 6, 1686-1691.

[3] Mansour, M. M. and Mohamed, S. M. (2015), *A new generalized of transmuted Lindley distribution*, Appl. Math. Sci, 9, 2729-2748.

[4] Nofal, Z. M., Afify, A. Z., Yousof, H. M., and Cordeiro, G. M. (2017), *The generalized transmuted-G family of Distributions*, Communications in Statistics-Theory and Methods, 46(8), 4119-4136.

[5] Shaw, W.T and Buckley, I.R.C. (2007), *The Alchemy of Probability Distributions: Beyond Gram-Charlier and Cornish-Fisher Expansions, and Skew-Normal or Kurtotic-Normal Distributions*, Research report.

Exponential Geometric Distribution: Comparing the Parameter Estimation Methods

Feyza GÜNAY[1], Mehmet YILMAZ[1]
fgunay@ankara.edu.tr, mehmetyilmaz@ankara.edu.tr

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

The new compound distributions which are started to be used with the study of Adamidis et al. (1998) have still found a place in the studies. Exponential Geometric (EG) distribution which is a flexible distribution for modelling the lifetime datasets, has introduced by Adamidis et al. (1998). They have used Maximum Likelihood Estimation (MLE) with Expectation-Maximization (EM) algorithm to estimate unknown parameters of this distribution. In this study, we use MLE with Expectation-Maximization (EM) algorithm and Least Square Estimation (LSE) methods to estimate the unknown parameters of EG distribution family. Then we compare the efficiencies of these estimators via a simulation study for different sample sizes and parameter settings. At the end of the study, a real lifetime data example is given for illustration.

Keywords: *Exponential geometric distribution, lifetime data, parameter estimation methods*

References

[1] Adamidis, K. and Loukas, S. (1998), *A lifetime distribution with decreasing failure rate*. Statistics & Probability Letters, 39, 35–42.

[2] Kus, C. (2007), *A new lifetime distribution.* Computational Statistics & Data Analysis, 51, 4497 – 4509.

[3] Louzada, F., Ramos, P.L and Perdoná, G.S.C. (2016), *Different Estimation Procedures for the Parameters of the Extended Exponential Geometric Distribution for Medical Data*. Computational and Mathematical Methods in Medicine,8727951, 12.

ISC2017 INTERNATIONAL STATISTICS CONGRESS
December 6-8, 2017 ANKARA/TURKEY

Macroeconomic Determinants and Volume of Mortgage Loans in Turkey

Ayşen APAYDIN[1], Tuğba GÜNEŞ[2]
aapaydin@ankara.edu.tr, gunestugba@gmail.com
[1]*Professor, Department of Insurance and Actuarial Sciences, Ankara University, Ankara, Turkey*
[2]*Phd Student, Department of Real Estate Management and Development, Ankara University, Ankara, Turkey*

Turkish mortgage system was established with the entrance into force of the Housing Finance System Law (No. 5582) in 2007. Even though the main reason for great economic crisis, called as 'financial tsunami' and that started in the USA and spreading around the whole world, is the USA mortgage system, volume of mortgage loans in Turkey has been showing a growing trend with some fluctuations since the very beginning of the system.

This paper investigates the impact of macroeconomic variables on the volume of mortgage loans in Turkey. Prior research has shown that various macroeconomic variables are chosen or included as the determinants of the development of the mortgage market. In this study, even though twelve macroeconomic variables are considered initially, four of them can take place in the model.

Using time series data from January 2007 to December 2016, following methodologies are applied in this study: Stationary Tests, Johansen's Cointegration Test, Johansen's Vector Error Correction Model, Granger Causality Tests and Impulse Response Function and Variance Decomposition Analysis.

The results demonstrate that weighted average of mortgage interest rates has the highest impact on the volume of mortgage loans. As the interest rates are decreasing, people incline to use mortgage loans for the purpose of house purchases. Relationship between the consumer price index and mortgage loans volume is also in negative way which is consistent with theoretical conceptual framework. Even though their affect is smaller compared to the first two variables, gross domestic product and money supply are the other macroeconomic variables explaining the changes in the volume of mortgage loans.

Keywords: *mortgage market, macroeconomic determinants, housing finance, cointegration, Turkey*

References

[1] Brooks, C. (2008), *Introductory Econometrics for Finance*, Second Edition, UK, Cambridge University Press.

[2] Choi, J. H. and Painter, G. (2015), *Housing Formation and Unemployment Rates: Evidence from 1975–2011*, Journal of Real Estate Finance and Economics, Vol.50-4, 549-566

[3] Gujarati, D. N. (2004), *Basic Econometrics*, Fourth Edition, The McGraw-Hill, USA.

[4] İbicioğlu, M. and Karan, M. B. (2012), *Konut Kredisi Talebini Etkileyen Faktörler: Türkiye Üzerine Bir Uygulama*, Ekonomi Bilimleri Dergisi, Vol. 4-1, 65-75

[5] Katipoğlu, B. N. and Hepşen, A. (2010), *Relationship Between Economic Indicators and Volume of Mortgage Loans in Turkey*, China-USA Business Review, Vol.9-10, 30-36.

ISC2017 INTERNATIONAL STATISTICS CONGRESS
December 6-8, 2017 ANKARA/TURKEY

Classification in Automobile Insurance Using Fuzzy c-means Algorithm

Furkan BAŞER[1], Ayşen APAYDIN[1]
furkan.baser@ankara.edu.tr, aysen.apaydin@ankara.edu.tr

[1]*Department of Insurance and Actuarial Science, Faculty of Applied Sciences, Ankara University, Ankara, Turkey*

Classifying risks and setting prices are an essential task in the insurance field from both theoretical and practical views [4]. Different methods of classification can produce different safety incentives, different risk distributions, and different protection against loss [3]. The aim of this study is to illustrate the use of a FCM clustering approach for application to the initial stages of the insurance underwriting process.

Clustering algorithms based on its structure are generally divided into two types: fuzzy and non-fuzzy (crisp) clustering. Crisp clustering algorithms give better results if the structure of the data set is well distributed. However, when the boundaries between clusters in data set are ill defined, the concept of fuzzy clustering becomes meaningful [2]. Fuzzy methods allow partial belongings (membership) of each observation to the clusters, so they are effective and useful tool to reveal the overlapping structure of clusters [5]. FCM clustering algorithm is one of the most widely used method among fuzzy associated models [1].

In the case of automobile insurance, it is common for insurers to use a number of a priori classification variables. In this study, the policy information including gender of the policy holder, car age, sum insured, geographical region, provincial traffic intensity, and no-claims discount level were used. Utilizing a data set of an automobile insurance portfolio of a company operating in Turkey, the FCM clustering method performs well despite some of the difficulties in the data.

Keywords: *automobile insurance, risk classification, fuzzy c-means*

References

[1] Bezdek, J.C. and Pal, S.K. (1992), *Fuzzy Models for Pattern Recognition: Methods that Search for Structure in Data*, New York, IEEE Press.

[2] Nefti, S. and Oussalah, M. (2004), *Probalilistic-fuzzy Clustering Algorithm*, in 2004 IEEE lntemational Conference on Systems, Man and Cybemetics, pp. 4786–4791.

[3] Retzlaff-Roberts, D. and Puelz, R. (1996), C*lassification in automobile insurance using a DEA and discriminant analysis hybrid.* Journal of Productivity Analysis, 7(4), 417-427.

[4] Yeo, A. C., Smith, K. A., Willis, R. J. and Brooks, M. (2001), *Clustering technique for risk classification and prediction of claim costs in the automobile insurance industry,* Intelligent Systems in Accounting, Finance and Management, 10(1), 39-50.

[5] Zhang, Y.J. (1996), *A Survey on Evaluation Methods for Image Segmentation*, Pattern Recognition, 29(8), pp. 1335–1346.

SESSION II

OTHER STATISTICAL METHODS I

10TH ISC2017 INTERNATIONAL STATISTICS CONGRESS
December 6-8, 2017 ANKARA/TURKEY

Analysing in Detail of Air Pollution Behaviour at Turkey by Using Observation-Based Time Series Clustering

Nevin GÜLER DİNCER[1], Muhammet Oğuzhan YALÇIN[1]
nguler@mu.edu.tr, *moguzhanyalcin@mu.edu.tr*

[1]*Muğla Sıtkı Koçman University, Faculty of Science, Department of Statistics, Turkey*

Time series clustering is a special case of clustering and is mostly used in determining correlations between time series, fitting a mutual model for numerous time series and revealing interesting patterns in time series data sets. Time series clustering approaches can be divided into three groups: i) observation-based, ii) feature-based and iii) model based. In literature, feature and model-based approaches are more common used since observation-based approaches have both high computation complexities when time series are long and require that all time series have equal length. However, these approaches lead to information lost since they use any characteristic of time series instead of actual time series observations. In this study, observation-based time series clustering approach is applied to daily PM_{10} concentrations time series in order to identify air pollution monitoring stations having similar behaviour. The objective in here is to reduce monitoring cost by determining centre stations to be monitored. For this objective, Fuzzy K-Medoid clustering algorithm providing centre point of stations behaving similar is used. The major advantage of this study is that clustering process is carried out for each week of 52 weeks separately and thus to provide more detail information about air pollution behaviour at Turkey.

Keywords: *time series clustering, fuzzy k-medoid clustering algorithm, air pollution, particulate matter*

References

[1] A. Gionis, H. Mannila, *Finding recurrent sources in sequences, in: Proceedings of the Seventh Annual International Conference on RESEARCH in Computational Molecular Biology*, 2003, pp. 123–130.

[2] A. Ultsch, F. Mörchen, *ESOM-Maps: Tools for Clustering, Visualization, and Classification with Emergent SOM*, 2005.

[3] F. Morchen, A. Ultsch, F. Mörchen, O. Hoos, *Extracting interpretable muscle activation patterns with time series knowledge mining*, J. Knowl. BASED 9 (3) (2005) 197–208.

Outlier Problem in Meta-Analysis and Comparing Some Methods for Outliers

Mutlu UMAROGLU[1], Pınar OZDEMIR[1]
mutlu.umaroglu@hacettepe.edu.tr

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

Meta-analysis is a statistical method that combining the outcomes from similar separate studies. In meta-analysis, effect sizes calculating from studies are combined to obtain more accurate and more powerful estimate. After obtaining the effect sizes, homogeneity of effect sizes is need to be assessed. The similarity of the effect sizes distribution shows that studies are homogeneous, and the difference shows that studies are heterogeneous. In literature, some studies can be different from the other ones. If an effect size of one study is quite different from the other studies, this study is called an outlier in the meta-analysis. It is also possible if one study has a very small standard error, it can be an outlier.

In a meta-analysis, it is possible to visualize the studies with graphical methods such as forest plot, radial plot, labbe plot. These plots give an idea about existence of outlier(s). Nevertheless, residuals must be examined to detect the outlier(s).

The distribution of effect sizes is more heterogeneous when there is an outlier in a meta-analysis. In this situation, the random effect model is constructed. There exist different variance estimation techniques such as DerSimonian–Laird, Maximum Likelihood, restricted maximum likelihood, Sidik-Jonkman, Empirical Bayes. If there is an outlier in a meta analysis study, it is recommended that researchers use robust mixture method or t-distribution to combine the outcomes.

In this study, we generated effect sizes including some outliers under different scenarios. While the combined effect size is the least affected by the outlier in robust mixture method and t-distribution, the combined effect size is the most affected by the outlier in empirical bayes method. While the confidence interval for combined effect size is the narrowest in the robust mixture method and empirical bayes method, the confidence interval is the largest in t-distribution. DerSimonian-Laird method has the greatest between-studies variance (τ^2). According to log-likelihood value, the best model is robust mixture and the worst model is DerSimonian Laird method.

Keywords: *meta-analysis, outlier, heterogeneity*

References

[1] Baker, R., Jackson, D. (2008), *A new approach to outliers in meta-analysis*, Health care management sciences, Volume 11, 121-131

[2] Beath, K.J. 2014, *A finite mixture method for outliers in meta-analysis*, Research synthesis methods, Volume 5, 285-293

[3] Gumedze, F.N. and Jackson, D. (2011), *A random effects variance shift model for detecting and accommodating outliers in meta-analysis*, BMC Medical Research Methodology, 11:19

[4] Lin, L., Chu, H., Hodges, J.S. (2016), *Alternative Measures of Between-Study Heterogeneity in Meta-Analysis: Reducing the Impact of Outlying Studies*, Biometrics, Volume 73, 156-166

[5] Viechtbauer, W. and Cheung, M. (2010), *Outlier and influence diagnostics for meta-analysis*, Research Synthesis Methods, Volume 1, 112-125

10TH ISC2017 INTERNATIONAL STATISTICS CONGRESS
December 6-8, 2017 ANKARA/TURKEY

The Upper Limit of Real Estate Acquisition by Foreign Real Persons and Comparison of Risk Limits in Antalya Province Alanya District

Toygun ATASOY[1], Ayşen APAYDIN[1], Harun TANRIVERMİŞ[1]
atasoy@ankara.edu.tr, aysen.apaydin@ankara.edu.tr, tanrivermis@ankara.edu.tr

[1]*Ankara University, Ankara, Turkey*

There has been many limitations and prohibitions to foreigners for the ownership acquisition throughout the history of property. Such limitations can be through quantity, quality, place and the type of the real estate as well as the combination of the restiriction both can be legal regulations and implementations. In Turkey the acquisition of real estate by foreigners is limited for the quantity, place and the aim of the use. In the Property Law No. 6302 which is enacted on 03.05.2012 the acquisition of real estate by foreign real person is limited with the provisions that the total area of limited real rights in an independent and permanent nature may be up to 10% of the surface area of the district that is subject to private ownership (in terms of surface area) and upper limit of 30 hectares in nation level.

The purpose of this study is to identfy the upper limit of the real estate acquisitions of foreign real persons and to analyze the risk limit in the Alanya district. These analyzes were carried out using the data provided by the General Directorate of Land Registry and Cadastre of the Ministry of Environment and Urbanization of Turkey. The data includes information of June 2015 - May 2017 Alanya district in the period of foreign real persons in real estate acquired as a result of the sales process. Sales of independent condominium units and the main real estate were examined separately and were created polynomial interpolation. Using the interpolation polynomials, the upper limit of the real estate acquisition of foreigners was determined. In addition, the limits of risk arising in the mentioned period are compared with the period of June 2013 - May 2015.

Keywords: *Real Estate Ownweship, Real Estate Acquisition by Foreigners, Limitation of Real Estate Acquisition and Policy Implication.*

[1] Atasoy, T. (2015*), The Limitation of Real Estate Acquisition by Foreign Real Persons: The Case of Antalya Province, Alanya District*. Master Thesis. Turkey. Ankara University.

[2] Tanrıvermiş, H., Apaydın, A., Erpul, G., Çabuk Kaya, N., Aslan, M., Aliefendioğlu, Y., Atasoy, M., Gün, S., Özçelik, A., Çelik, K., İşlek, B. G., Erdoğan, M. K., Atasoy, T., Öztürk, A., Hatipoğlu, C. E., Keleş, R., Tüdeş, T. (2013). *The Project of Real Estate Acquisition by Foreigners in Turkey and Evaluation Of Its Effects.* The Scientific And Technological Research Council of Turkey (TUBITAK) Project Number: 110G020; Ankara.

[3] Tanrıvermiş, H., Doğan, V., Akipek Öcal, Ş., Kurt, Y., Akyılmaz, S. G., Tanrıbilir, F. B., Dardağan Kibar, E., Başpınar, V., Aliefendioğlu, Y., Apaydın, A., Çabuk Kaya, N., Şit, B., Baskıcı, M. (2013), *The Project of Real Estate Acquisition by Foreigners in Turkey and Evaluation Of Its Effects: Analysis of Real Estate Acquisitions of Foreigners in Historical Development Process in Turkey*, The Scientific And Technological Research Council of Turkey (TUBITAK) Project Number: 110G020; Ankara.

Comparison of MED-T and MAD-T Interval Estimators for Mean of A Positively Skewed Distributions

Gözde ÖZÇIRPAN[1], Meltem EKİZ[2]
gozcirpan@ankara.edu.tr,ozmeltem@gazi.edu.tr

[1]*Ankara University Department of Statistics, Ankara, Turkey*
[2] *Gazi University Department of Statistics, Ankara, Turkey*

Several researchers proposed various interval estimators for estimating the mean of a positively skewed distributions. Banik and Kibria (2007) compared the MED-T and MAD-T confidence intervals with those proposed by various researchers, under similar simulation conditions. In order to compare the performance of these intervals, they used coverage probability, average width and ratio of coverage to width criteria.
In this study, the best performance of MED-T and MAD-T interval estimators are investigated in terms of various distributions, skewness, sample sizes and confidence levels. Towards this aim simulation studies are made by using Matlab R2007b. In general, MED-T interval estimator gave better results in terms of coverage probabilities of confidence interval. Coverage probabilities for MED-T interval estimator were close to $1-\alpha$ confidence levels for low skewness and small sample sizes. In case of moderately skewness it has been observed that the coverage probabilities has given better results for large sample sizes. MAD-T interval estimator has the narrower interval in terms of the widths of confidence intervals.

Keywords: *MED-T interval estimator, MAD-T interval estimator, Confidence intervals, Skewness*

References

[1] Baklizi, A., *Inference About mean of a Skewed Population: A Comparative Study*, Journal of Statistical Computation and Simulation, 78:421-435 (2006)

[2] Baklizi, A.,Kibria, B.M.G., *One and Two Sample Confidence Intervals for Estimating the Mean of Skewed Populations: an Empirical Comparative Study*, Journal of Applied Statistics, 36:601-609 (2009)

[3] Banik, W.S., Kibria, B.M.G., *On Some Confidence Intervals for Estimating The Mean of a Skewed Population*, International Journal of Mathematical Education in Science and Tecnology,38 (3):412-421 (2007)

[4] Banik, W.S., Kibria, M.G., *Comparison of Some Parametric and Nonparametric Type One Sample Confidence Intervals for Estimating the Mean of a Positively Skewed Distribution*, Communications in Statistics- Simulation and Computation,39:361-389 (2010)

ISC2017 INTERNATIONAL STATISTICS CONGRESS
December 6-8, 2017 ANKARA/TURKEY

Bayesian Estimation for the Topp-Leone Distribution Based on Type-II Censored Data

İlhan USTA[1], Merve AKDEDE[2]
iusta@anadolu.edu.tr, merve.akdede@usak.edu.tr

[1]*Faculty of Science, Department of Statistics, Anadolu University, Eskisehir, Turkey*
[2]*Faculty of Arts and Science, Department of Statistics, Usak University, Usak, Turkey*

This paper focuses on the estimation of the shape parameter of the Topp-Leone distribution based on Type-II censored data. Using non-informative and informative priors, Bayes estimators of the shape parameter are obtained under squared error, linear exponential (LINEX) and general entropy loss functions. Furthermore, a performance comparison of the obtained Bayes estimators and the corresponding maximum likelihood estimator is conducted in terms of mean squared error (MSE) and bias through an extensive numerical simulation. It can be deduced from simulation results that the Bayesian estimators using asymmetric loss function show good performance in terms of MSE for most of the considered cases.

Keywords: *Topp-Leone distribution, Type-II censoring, LINEX, mean squared error*

References

[1] Cohen, A. C. (1965), *Maximum Likelihood Estimation in the Weibull Distribution Based on Complete and Censored Samples*. Technometrics, 7(4), 579-588.

[2] Feroze, N., and Aslam, M. (2017), *On selection of a suitable prior for the posterior analysis of censored mixture of Topp Leone distribution*, Communications in Statistics - Simulation and Computation, 46(7), 5184-5211.

[3] Sultan. H, and Ahmad S.P. (2016), *Bayesian analysis of Topp-Leone distribution under different loss functions and different priors*, Journal of Statistics Applications & Probability Letters, 3, 109-118.

[4] Tabassum N., Sindhua, T.N., Saleemb M. and Aslama M. (2013), *Bayesian Estimation for Topp-Leone Distribution under Trimmed Samples*, Journal of Basic and Applied Scientific Research, 3(10), 347-360.

[5] Topp, C. W. and Leone, F. C. (1955), *A family of J-shaped frequency functions*, Journal of the American Statistical Association, 50, 209-219.

SESSION III

TIME SERIES II

ISC2017 INTERNATIONAL STATISTICS CONGRESS
December 6-8, 2017 ANKARA/TURKEY

An Overview on Error Rates and Error Rate Estimators in Discriminant Analysis

Cemal ATAKAN[1], Fikri ÖZTÜRK[1]
catakan@ankara.edu.tr, fikri.ozturk@ankara.edu.tr
[1]*Ankara University. Faculty of Science, Department of Statistics, Ankara, Turkey*

Discriminant analysis is a statistical technique that, when the researcher makes measurements on an individual and wishes to assigns this individual into one of several known populations or categories on the basis of these measurements. It is assumed that the individual can come from a finite number of populations and each population is characterized by the probability distribution of a random vector $\underline{X}$ associated with the measurements. When the probability distributions are completely known, then the problem is reduced to identifying the allocation rule[1,5]. The main goal of discriminant analysis is to obtain an allocation procedure with minimum error. According to this optimization criterion, it is important to know the probability of the misclassification or error rate for the evaluation of the allocation rules. Error rates are usually obtained depending on the distribution of the discriminant function. However, error rates can also be calculated independently of the distribution. There are optimal, actual(conditional) and expected actual (unconditional) error rates for allocation rules. The optimal error rate is the error rate that would ocur when the parameters of the discriminant function are known. The actual error rate is obtained according to the sample discriminant function based on the parameter estimates obtained from the samples when the parameters are not known, and the expected actual error rate is the expected value of the actual error rate over all possible samples. There are many error rate estimators described in the literature for the actual error rate[4, 2].

This study will focus on some error rate estimators for the actual error rate. The aim is to draw attention to the estimation of error rates and error rate estimators.

Keywords: *Discriminant analysis, error rate, eror rate estimators*

References

[1] Anderson, T.W. (1984), *An introduction to multivariate statistical analysis,* Second edition, New York, Jhon Wiley and Sons Inc.

[2] Atakan, C. (1997), *Diskriminasyon ve hata oranları tahmini,* Ankara Üniversitesi, Fen Bilimleri Enstitüsü.

[3] Egbo, I. (2016), *Evaluation of error rate estimators in discriminant anaysis with multivariate binary variables,* American Journal of Theoretical and Applied Statistics, Vol.5, No.4, 173-179.

[4] Lachenbruch, P., A., Mickey, M., R. (1968), *Estimation of error rates in discriminant anaysis,* Technometrics, 10, 1-11.

[5] Johnson, R.,A., Wichern, D.,W. (2007), *Applied multivariate statistical analysis,* 7th edition, New Jersey, Pearson.

10TH ISC2017 INTERNATIONAL STATISTICS CONGRESS
December 6-8, 2017 ANKARA/TURKEY

A New VARMA Type Approach of Multivariate Fuzzy Time Series Based on Artificial Neural Network

Cem KOÇAK[1], Erol EĞRİOĞLU[2]
cemkocak@hotmail.com, erole1977@yahoo.com

[1]*Hitit University, School of Health,Çorum, Turkey*
[2]*Giresun University, Faculty of Arts and Sciences, Department of Statistics, Forecast Research Laboratory, Giresun, Turkey*

Methods fuzzy of the time series analysis have usually made progress so as to be the alternative of the univariate time series analysis. There have also some approaches of multivariate fuzzy time series in the literature. Some of these approaches are [1], [2], [3] and [4] and forecasts of a targeted time series have been tried to obtain via two or more time series in these studies in the literature. In also this study that are differ from other studies in the literature, A new multivariate fuzzy time series forecasting model and a solving method of this model which also includes the lagged variables of errors that more than one time series are forecasted at the same time have proposed. Proposed method has been solved for the real life time series and compared other time series method in the literature.

Keywords: *Fuzzy Time Series, Artificial Neural Network,Multiple Output Artificial Neural Network, Multivariate Time Series Analysis.*

References

[1] Egrioglu, E., Aladag, C.H., Yolcu, U., Uslu, V.R., Basaran, M.A. (2009), *A new approach based on artificial neural networks for high order multivariate fuzzy time series*, Expert Systems with Applications, 36 (7), pp. 10589-10594.

[2] Jilani, T. A., & Burney, S. M. A. (2008). *Multivariate stochastic fuzzy forecasting models*, Expert Systems with Applications, 35, 691–700.

[3] Kamal S. Selim and Gihan A. Elanany (2013), *A New Method for Short Multivariate Fuzzy Time Series Based on Genetic Algorithm and Fuzzy Clustering*, Advances in Fuzzy Systems Volume 2013, Article ID 494239, 10 pages http://dx.doi.org/10.1155/2013/494239

[4] Yu, T. K., Huarng, K. (2008), *A bivariate fuzzy time series model to forecast the TAIEX*, Expert Systems with Applications, 34(4), 2945–2952.

ISC2017 INTERNATIONAL STATISTICS CONGRESS
December 6-8, 2017 ANKARA/TURKEY

An Application of Single Multiplicative Neuron Model Artificial Neural Network with Adaptive Weights and Biases based on Autoregressive Structure

Ozge Cagcag YOLCU[1], Eren BAS[2], Erol EGRIOGLU[2], Ufuk YOLCU[3] *ozgecagcag@yahoo.com, eren.bas@giresun.edu.tr, erol.egrioglu@giresun.edu.tr, ufuk.yolcu@giresun.edu.tr,*

[1]*Giresun University, Department of Industrial Engineering, Giresun, Turkey*
[2] *Giresun University, Department of Statistics, Giresun, Turkey*
[3]*Giresun University, Department of Econometrics, Giresun, Turkey*

Various traditional time series forecasting approaches may fail to analysis of complex real-word time series due to their strict assumptions such as model assumptions, normal distribution, and sufficient number of observation. To overcome this kind of failing, especially in recent years, various artificial neural networks (ANNs) have been commonly utilized for modelling time series. Multilayer perceptron (MLP) introduced by [3] is one of the most widely used ANN. In time series forecasting process via MLP, an essential issue is to determine the number of hidden layers and neurons in the hidden layers since it may affect the prediction performance of ANN. This issue can be called as architecture selection problem. Single multiplicative neuron model (SMNM) proposed by [4] does not contain this kind of problems. The main different features of SMNM than MLP are that having just one neuron, use of multiplicative function as an aggregation function and requiring less parameter. Although SMNM has some advantages in comparison to MLP, it is a fundamental problem that it is a model-based due to having only one neuron. In the forecasting time series with more complex structure, SMNM would be insufficient unlike MLP which may produce outstanding through its high compliance with data by changing its architecture. By considering both advantages and disadvantages of MLP and SMNM, a SMNM with dynamic weights and biases based on autoregressive structure was proposed by [1]. In this method proposed by [1], the weights and the biases of the SMNM are determined by favour of autoregressive equations. By using autoregressive equations in the determining of the weights and the biases, time index of each observations are considered. SMNM, therefore, is converted into a data-based forecasting model. The parameters of autoregressive equations are specified by particle swarm optimization introduced by [2]. In this study, the method proposed by [1] are introduced and to display the performance of this SMNM, various time series are analysed and the obtained results are evaluated.

Keywords: *single multiplicative neuron model, data-based forecasting model, autoregressive equations, time series forecasting, particle swarm optimization.*

References

[1] Cagcag Yolcu, O., Bas, E., Egrioglu E. and Yolcu U. (2017), *Single Multiplicative Neuron Model Artificial Neural Network with Autoregressive Coefficient for Time Series Modelling*, Neural Processing Letters, doi:10.1007/s11063-017-9686-3.

[2] Kennedy, J. and Eberhart, R. (1995), *Particle swarm optimization*, In: Proceedings of IEEE international conference on neural networks. Piscataway, NJ, USA. IEEE Press, 1942-1948.

[3] Rumelhart, E., Hinton, G.E. and Williams, R.J. (1986) *Learning internal representations by error propagation, chapter 8*. Cambridge, The M.I.T. Press, 318-362.

[4] Yadav, R.N., Kalra, P.K. and John, J. (2007) *Time series prediction with single multiplicative neuron model*. Applied Soft Computing 7, 1157-1163.

10TH ISC2017 INTERNATIONAL STATISTICS CONGRESS
December 6-8, 2017 ANKARA/TURKEY

A novel Holt's Method with Seasonal Component based on Particle Swarm Optimization

Ufuk YOLCU[1], Erol EGRIOGLU[2], Eren BAS[2]
ufuk.yolcu@giresun.edu.tr, erol.egrioglu@giresun.edu.tr, eren.bas@giresun.edu.tr

[1]*Giresun University, Department of Econometrics, Giresun, Turkey*
[2] *Giresun University, Department of Statistics, Giresun, Turkey*

Exponential smoothing methods is a class for time series forecasting methods. [1-3] and [5] are early studies in this class. Holt's linear trend method (Holt Method) has been widely and accomplishedly used for prediction time series with trend component and the method was proposed in [3]. In the Holt method, the predictions are obtained by updated trend and level of series. Updating of trend and the next level of series are determined via utilization of previous computed and real values. Although this method produce successful prediction results for time series with trend component, many encountered time series include seasonal component as well as trendy. In this study, a new model in Holt method which contains a seasonal component is proposed. The proposed model, therefore, has some new smoothing parameters regarding to seasonal component. Model of the proposed Holt method can be given as follow.

$$\hat{X}_{t+1} = \lambda_1(L_t + B_t) + (1-\lambda_1)(L_{t-s} + B_{t-s})$$
$$L_t = \lambda_2\left(\lambda_3 X_t + (1-\lambda_3)(L_{t-1} + B_{t-1})\right) + (1-\lambda_2)\left(\lambda_4 X_{t-s} + (1-\lambda_4)(L_{t-s} + B_{t-s})\right)$$
$$B_t = \lambda_5(L_t - L_{t-1}) + (1-\lambda_5)B_{t-1}$$

where B_t and L_t represent trend and the level of time series at t time. Moreover, $\lambda_j,\ j = 1,2,\ldots,5$ represents the smoothing parameters. These smoothing parameters of the proposed method are estimated by using particle swarm optimization. Particle swarm optimization is firstly proposed in Kennedy and Eberhart (1995) and it is a good tool for numerical optimization problem. To evaluate the performance of the proposed method, various real-word time series are analysed. The results are evaluate with some time series prediction tools' results.

Keywords: *exponential smoothing methods, predictions, seasonal component, particle swarm optimization*

References

[1] Brown, R.G. (1959), *Statistical Forecasting for inventory control*, New-York, the country for pressing, McGraw-Hill.

[2] Brown, R.G. (1963), Smoothing, forecasting, prediction *of discrete time series*, Engle-wood Cliffs, N.J.:Prentice-Hall

[3] Holt, C.C. (1957), *Forecasting trends and seasonal by exponentially weighted moving averages*, Office of Naval Research, Research Memorandum, Carnegie Institute of Technology, No:52.

[4] Kennedy, J. and Eberhart, R. (1995), *Particle swarm optimization*, In: Proceedings of IEEE international conference on neural networks. Piscataway, NJ, USA. IEEE Press, 1942-1948.

[5] Winters, P.R. (1960), *Forecasting sales by exponentially weighted moving averages*, Managment Science, 6, 324-342.

A New Intuitionistic High Order Fuzzy Time Series Method

Erol EGRIOGLU[1], Ufuk YOLCU[2], Eren BAS[1]
ufuk.yolcu@giresun.edu.tr, erol.egrioglu@giresun.edu.tr, eren.bas@giresun.edu.tr

[1]*Giresun University, Department of Statistics, Giresun, Turkey*
[2] *Giresun University, Department of Econometrics, Giresun, Turkey*

Intuitionistic fuzzy sets are general form of type 1 fuzzy sets. There is a second order uncertainty approach by using hesitation degrees in intuitionistic fuzzy sets. The summation of memberships and non-membership values can be less than one for an intuitionistic fuzzy set. In this study, a new forecasting method is proposed based on intuitionistic fuzzy sets. The intuitionistic fuzzy time series definition is made in the study. The fuzzification is done by using intuitionistic fuzzy c-means algorithm, pi-sigma artificial neural network is used to define fuzzy relations. Artificial bee colony algorithm is used as an optimization algorithm in the proposed method. Real-world time series applications has been made for exploring performance of the proposed method.

Keywords: *Intuitionistic fuzzy sets, forecasting, artificial bee colony, intuitionistic fuzzy c-means, pi-sigma artificial neural network.*

References

[1] Atanassov K. T. (1986), *Intuitionistic fuzzy sets*. *Fuzzy Sets and Systems*, *20*(1), 87–96.

[2] Chaira T. (2011), *A novel intuitionistic fuzzy C means clustering algorithm and its application to medical images*, Applied Soft Computing, 11(2), 1711–1717.

[3] Shin Y., Gosh J. (1991), *The Pi-sigma network: An efficient higher-order neural network for pattern classification and function approximation.* In Proceedings of the International Joint Conference on Neural Networks.

[4] Karaboga D., Akay B. (2009), *A comparative study of artificial bee colony algorithm*, Applied Mathematics and Computation, 214, 108-132.

SESSION III

DATA MINING I

Recommendation System based on Matrix Factorization Approach for Grocery Retail

Merve AYGÜN[1], Didem CİVELEK[1], Taylan CEMGİL[2]
merve.aksoz@obase.com,didem.civelek@obase.com

[1]*OBASE, Department of Project Innovation Lab, İstanbul, Turkey*
[2] *Boğaziçi University, Department of Computer Engineering, İstanbul, Turkey*

In the new big data era, the data being produced in all areas of the retail industry is growing exponentially, creating opportunities for those analysing this data to gain a competitive advantage. As digitalization accelerate, the physical shops have to cope with new competitors, the e-commerce actors. E-commerce sites like Amazon have defined new purchasing strategies: faster, sometimes cheap, and more targeted. Today's new purchasing strategy needs personalized recommendations to improve customer satisfaction by matching customers with relevant products at the specific time and conditions thanks to Recommender System Applications.

The following study has proposed a recommendation system for an on-line grocery store by discovering prominent dimensions that encode the properties of items and users' preferences toward them. These dimensions are in implicit form such as shopping history, browse logs, etc.; in addition, customer demography, product hierarchy, product attributes information has used in order to enhance the data content.

We have developed a recommendation system based on latent factor model with Matrix Factorization (MF) method to incorporate personalized purchase behaviours with product/item attributes. MF methods are known to have good performance for implicit datasets [1,2]. It is developed two algorithm based on matrix factorization: mix and discover. Discover algorithm makes recommendation from not purchased products by customer till now whereas mix algorithm makes recommendation from both of purchased and not purchased products by the customer.

The success of the proposed recommendation system has measured by applying and benchmarking with two other algorithms: random and nopcommerce. Random algorithm makes randomly product recommendation from on sale products. The second competitor algorithm, nopcommerce, makes recommendation based on association rule mining: cross- sell product approach "Customers who bought this item also bought...".

Performance outputs has been measured for one year (2016 December- 2017 November). The results show that developed recommendation system included of two algorithms based on latent factor model outperform statistically better than other 2 competitor algorithms. Click to purchase rate for mix and discover algorithms are about %35 for both of them, while for nopcommerce and random algorithms is respectively %21 and %13. Another used performance metric is purchase amount. Purchase amount for two proposed algorithm is %52 higher than sum of two competitor algorithms.

Keywords: *Recommendation system, latent factor model, matrix factorization, machine learning, grocery retail*

References

[1] He, R. and McAuley J. (2016), VBPR: *Visual Bayesian Personalized Ranking from Implicit Feedback*, Association for the Advancement of Artificial Intelligence

[2] Koren, Y., Bell, R. and Volinsky, C. (2009), *Matrix Factorization Techniques for Recommender Systems*, IEEE Computer Society.

10TH ISC2017 INTERNATIONAL STATISTICS CONGRESS
December 6-8, 2017 ANKARA/TURKEY

Demand Forecasting Model for new products in Apparel Retail Business

Tufan BAYDEMİR[1], Dilek Tüzün AKSU[2]
tufan.baydemir@obase.com, dtuzun@yeditepe.edu.tr

[1]*R&D Team Manager, İstanbul, Turkey*
[2]*Yeditepe University, Department of Industrial and Systems Engineering, İstanbul, Turkey*

Demand forecasting plays an important role for planning in many industries. Especially in apparel retail, merchandising planners plan their budgets for the upcoming seasons in a year advance. Because of the long lead times, they have to decide which product and how much to be produced months before the selling season starts. Merchandising managers plan their budgets under some uncertainties like "which products the customers will likely to buy?"," which color will be popular?". Besides, in apparel retail business, products are changed dramatically in every selling season. Generally, many of the products sold during a selling season have no historical information. Lack of information about customer's tastes and not the existence of historical sales data cause great uncertainty about demand planning.

For these reasons, accurate sales forecasting in the apparel industry is the most important input for many decision-making processes. To generate better forecasting algorithms, some should well understand the dynamics behind the purchasing decision in apparel. Purchasing decision of a customer is generally related to the price of the product.

Since ordinary apparel retailers have thousands of products, manual forecasting is not an easy job. Besides, characteristics of the demand are very complex in apparel retail business. To deal with this sophisticated problem merchandising planners need a decision support tool to forecast the future demand.

In this study, a data-driven demand forecasting model was proposed. Because many products have no historical sales information, clustering approach was proposed by [1] to group similar products. Based on the historical information of the grouped products, multivariate regression analysis was applied. In their study [1], Smith & Achabal pointed out that if some colors or sizes of a product was not on display that would cause a decrease in sales. Therefore, demand was formulated as a function of price, time and inventory. Demand forecasting model was applied on a well-known apparel retailer's data and results were evaluated.

Keywords: *Demand, Forecasting, Apparel, Retail*

References

[1] Smith, S. A., McIntyre, S. H., & Achabal, D. D. (1994). *A two-stage sales forecasting procedure using discounted least squares*, Journal of Marketing Research, 44-56.

[2] Thomassey. S. and Fiordaliso, A. (2005), *A hybrid sales forecasting system based onclustering and decision trees*, Decision Support Systems, 42, 408-421.

Comparison of the Modified Generalized F-test with the Non-Parametric Alternatives

Mustafa ÇAVUŞ[1], Berna YAZICI[1], Ahmet SEZER[1]
mustafacavus@anadolu.edu.tr, bbaloglu@anadolu.edu.tr, a.sezer@anadolu.edu.tr

[1]*Anadolu University, Department of Statistics, Eskişehir, Turkey*

Classical methods are used for testing equality of group means but they lose their power when the assumptions are violated. In case of variance heterogeneity, there are many powerful methods are proposed such as Welch, Brown-Forsythe, Parametric Bootstrap and Generalized F-test. However, power of these tests are affected negatively under non-normality. Cavus et al. (2017) proposed modified generalized F-test which is used under both heteroscedasticity and non-normality. The efficieny of this method to over other parametric methods was shown in Cavus et al. (2017). In this study, modified generalized F-test is compared with non-parametric alternatives such as Brunner-Dette-Munk, Kruskal Wallis and Trimmed Test in terms of their power and type I error rate. Under different scenarios, the performances of these methods are investigated with Monte-Carlo simulation results.

Keywords: *heteroscedasticity, non-normality, outlier, non-parametric test*

References

[1] Brunner, E., Dette, H. and Munk, A. (1997), *Box-type approximations in nonparametric factorial designs*, Journal of the American Statistical Association, 92, 1494-1502.

[2] Cavus, M., Yazıcı, B. and Sezer, A. (2017), *Modified tests for comparison of group means under heteroskedasticity and non-normality caused by outlier(s)*, Hacettepe Journal of Mathematics and Statistics, 46(3), 492-510.

[3] Wilcox, R. R. (2005), *Introduction to robust estimation and hypothesis testing*, Burlington, Elsevier.

Robustified Elastic Net Estimator for Regression and Classification

Fatma Sevinç KURNAZ[1], Irene HOFFMANN[2], Peter FILZMOSER[2]
fskurnaz@yildiz.edu.tr, Irene.Hoffmann@tuwien.ac.at, P.Filzmoser@tuwien.ac.at

Yildiz Technical University, Istanbul, Türkiye[1]
Vienna University of Technology, Vienna, Austria[2]

Elastic net estimators penalize the objective function of a regression problem by adding a term containing the L1 and L2 norm of the coefficient vector. This type of penalization achieves intrinsic variable selection and similar coefficient estimates for highly correlated variables. We propose fully robust versions of elastic net estimator for linear and logistic regression. The algorithm searches for outlier-free subsets on which the classical elastic net estimators can be applied. A final reweighting step is added to improve the statistical efficiency of the proposed methods. An R package, so called enetLTS, is provided to compute the proposed estimators. Simulation studies and real data examples demonstrate the superior performance of the proposed methods.

The work was supported by grant TUBITAK 2214/A from the Scientific and Technological Research Council of Turkey and by the Austrian Science Fund (FWF), project P 26871-N20.

Keywords: *elastic net penalty, least trimmed squares, C-step algorithm, high dimensional data, robustness, sparse estimator*

References

[1] A. Alfons, C. Croux, S. Gelper, Sparse least trimmed squares regression for analyzing high-dimensional large data sets, The Ann. of Apl. Stat.

[2] Friedman, J. and Hastie, T. and Tibshirani R., Regularization paths for generalized linear models via coordinate descent, Journal of Statistical Software.

[3] Maronna, R.A. and Martin, R.D. and Yohai, V.J. (2006). Robust Statistics: Theory and Methods, Wiley, New York.

[4] Rousseeuw, P. J. and Van Driessen, K. (2006) Computing LTS regression for large data sets, Dat. Min. and Know. Disc.

[5] Serneels S., Croux C., Filzmoser P., Espen, P. J. V. (2005). Partial Robust M-Regression, Chem. and Int. Lab. Sys.

ISC2017 INTERNATIONAL STATISTICS CONGRESS
December 6-8, 2017 ANKARA/TURKEY

Insider Trading Fraud Detection: A Data Mining Approach

Emrah BİLGİÇ[1], M.Fevzi ESEN[2]
e.bilgic@alparslan.edu.tr, fevzi.esen@medeniyet.edu.tr

[1]*Muş Alparslan University, Muş, Turkey*
[2]*Istanbul Medeniyet University, İstanbul, Turkey*

Prior researches provide evidence that insiders generate significant profits by trading on private information which is unknown to the market. Separating opportunistic insider trades from routine ones is highly important for detecting a fraud. In the literature, there is only a few studies on fraud detection of insiders' trades [1][2]. In this study, Outlier Detection approach will be used to detect potential frauds. Outlier detection, with other words; anomaly or novelty detection is the task of finding patterns that do not conform to the normal behaviour of the data. This study is organized to detect outliers with data mining approach, then inspect outlying transactions' portfolio by estimating abnormal returns to flag potential fraudulent transactions. Outlier detection is the first step in many data mining applications, as in our case. A clustering-based outlier detection method called "peer group analysis" will be used in this paper. Peer group analysis is first introduced by Bolton and Hand [3] which detects individual objects that begin to behave in a way distinct from similar objects over time. Although the logic behind Bolton & Hand's and this study is same, analysis in this study differs from Bolton & Hand's since they consider time concept additionally. The procedure for this paper searches unusual cases (outliers) based on deviations from the norms of their (cluster) groups. The clustering mentioned here is based on input variables such as volume or price of the trade. After clusters called "peer groups" are produced, anomaly indices based on deviations from peer group norms are calculated. SPSS is used for outlier detection with peer group analysis. A dataset is obtained from Thomson Reuters Insider Filings, containing 1,244,815 transactions belong to 61,780 insiders during the period of January 2010 - April 2017 in NYSE. First of all, NPR and NVR values are calculated for each transaction. Note that an insider may have hundreds or even thousands of transactions between that periods. Then, outlier detection with peer groups analysis is performed using the purchase and sale transaction data separately. 16,362 outliers have been found for purchases data which contain 328,112 transactions, however 4 of them significantly differ from their peer group. The primary reason of these 4 outliers are NVR values, and for other NPR values. Furthermore, outliers in sales data are also inspected and 27,190 outliers are obtained out of 916,703 transactions, again 4 of them significantly differs from their peer group. The primary reason of these 4 outliers is NVR values of the transactions and for others NPR values as in the case for purchase transactions. Since insiders' purchases and sales have different characteristics, future work will focus on measuring returns of purchase and sale portfolios separately for each outlier.

Keywords: *financial fraud detection, data mining, outlier detection, event study methodology*

Reference

[1] Tamersoy, A., Khalil, E., Xie, B., Lenkey, S. L., Routledge, B. R., Chau, D. H., & Navathe, S. B. (2014). *Large-scale insider trading analysis: patterns and discoveries.* Social Network Analysis and Mining, 4(1), 201.

[2] Goldberg, H.G., Kirkland, J.D., Lee, D., Shyr, P., Thakker, D. (2003). *The NASD securities observation, new analysis and regulation system (SONAR).* In: Proceedings of the Conference on Innovative Applications of Artificial Intelligence

[3] Bolton, R. J., & Hand, D. J. (2001). *Peer group analysis–local anomaly detection in longitudinal data.* In: Technical Report, Department of Mathematics, Imperial College, London.

SESSION III

APPLIED STATISTICS IV

10TH ISC2017 INTERNATIONAL STATISTICS CONGRESS
December 6-8, 2017 ANKARA/TURKEY

A New Hybrid Method for the Training of Multiplicative Neuron Model Artificial Neural Networks

Eren BAS[1], Erol EGRIOGLU[1], Ufuk YOLCU[2]
eren.bas@giresun.edu.tr, erole1977@yahoo.com, varyansx@hotmail.com

[1]*Giresun University, Department of Statistics, Forecast Research Laboratory, Giresun, Turkey*
[2]*Giresun University, Department of Econometrics, Forecast Research Laboratory, Giresun, Turkey*

In the literature, the training of multiplicative neuron model artificial neural networks (MNM-ANN) has been performed with some artificial intelligence optimization techniques such as genetic algorithm, particle swarm optimization, differential evolution algorithm and some derivative based algorithms. In this study, different from other studies in the literature, a new hybrid method for the training of MNM-ANN is proposed. In the proposed new hybrid method, artificial bat algorithm and back propagation learning algorithm is used together. Besides, the properties of an artificial intelligence optimization technique, bat algorithm, and a derivative based algorithm, back propagation learning algorithm, is used together by using the proposed method. The proposed method is applied to the Australian beer consumption (AUST) data time series data with 148 observations between the years 1956 and 1994. The last 16 observations of the time series were taken as test data. In addition to the proposed method, AUST data is analyzed by using seasonal autoregressive integrated moving average, Winter's multiplicative exponential smoothing, Multi-layer feed-forward neural network, Multilayer neural network based on particle swarm optimization, Back propagation algorithm based on MNM-ANN, MNM-ANN based on particle swarm optimization, MNM-ANN based on differential evolution algorithm, Radial basis artificial neural network, and Elman neural network methods. At the end of the analysis, it is seen clearly seen that the proposed method has the best performance compared with the methods given above according to root mean square and mean absolute percentage error criteria for AUST data.

Keywords: *multiplicative neuron model, artificial bat algorithm, back propagation, hybrid method.*

References

[1]Yadav R.N., Kalra P.K., John J. (2007), *Time series prediction with single multiplicative neuron model*, Applied Soft Computing, 7, 1157-1163.

[2]Rumelhart D.E., Hinton G.E., Williams R.J., (1986), *Learning represantations by back propagating errors*, Nature, 323, 533-536.

[3]Yang X.S. (2010), *A new metaheuristic bat-inspired algorithm*, Studies in Computational Intelligence, 284, 65–74, 2010.

Investigation of The Insurer's Optimal Strategy: An Application on Agricultural Insurance

Mustafa Asım Özalp[1], Uğur Karabey[1]
m.asim.ozalp@hacettepe.edu.tr,ukarabey@hacettepe.edu.tr

[1]*Hacettepe University, Ankara, Turkey*

We investigate an insurer's optimal investment and reinsurance ratio problem by maximizing the expected terminal wealth under exponential utility functions. It is assumed that there are 3 investment options for insurer and the insurer's risk process follows jump diffusion process. The problem is considered under the control theory and the closed- form solutions are obtained for the optimal investment strategy and reinsurance. In order to model the risk process of the insurer, the agricultural data from TARSİM were used.

Keywords: *Control theory, Optimal Investment, Jump-diffusion Process*

References

[1] Oksendal, B. and Sulem, A. (2004), *Applied Stochastic Control of Jump Diffusions*, Germany, Springer.

[2] ÖZALP, M. A. (2015), *Determination of The Optimal Investment and Liability For An Insurer with Dynamic Programming*, Hacettepe University, 11-17.

10TH ISC2017 INTERNATIONAL STATISTICS CONGRESS
December 6-8, 2017 ANKARA/TURKEY

Portfolio Selection Based on a Nonlinear Neural Network: An Application on the Istanbul Stock Exchange (ISE30)

Ilgım YAMAN[1], Türkan ERBAY DALKILIÇ[2]
ilgimyaman@gmail.com, tedalkilic@gmail.com

[1]*Giresun University, Giresun, TURKEY*
[2]*Karadeniz Technical University, Trabzon, TURKEY*

Portfolio selection problem is a very popular optimization problem in the optimization world. Hanry Markowitz [1] had proposed standard portfolio optimization in 1952. In the Portfolio optimization problem main goal is minimizing the risk, while maximizing the expected return of portfolio. Because of portfolio optimization problem is an NP-hard problem, many heuristic methods were used to solve portfolio optimization method such as particle swarm optimization, ant colony optimization etc. In fact these methods are not satisfied stock markets demands in financial world. In this study, in order to solve portfolio optimization problem, we prefer a nonlinear neural network. Since portfolio optimization problem is a quadratic programming (QP) problem, we use a new neural network which is represented in 2014 by Yan [2]. Proposed neural network is based on solving primal and dual problems simultaneously [3]. Istanbul stock exchange-30 data are used to solve nonlinear neural network which is adapted to solve portfolio optimization.

Keywords: *Portfolio optimization, Nonlinear neural network, ISE-30, Markowitz*

References

[1] Markowitz H., (1952), *Portfolio selection, The journal of finance,* 7(1):77-91

[2] Yan, Y. (2014), *A new nonlinear neural network for solving QP problems,* International Symposium on Neural Networks, Springer International Publishing, 347-357

[3] Nyugen, K.V., (2000), *A Nonlinear Network for Solving Linear Programming Problem s he title of proceeding,* International Symposium on Matematical Programming, ISMP 2000, Atlanta, GA, USA

10TH ISC2017 INTERNATIONAL STATISTICS CONGRESS
December 6-8, 2017 ANKARA/TURKEY

A Novel Approach for Modelling HIV-1 Protease Cleavage Site Preferability with Epistemic Game Theory

Bilge BAŞER[1], Metin YANGIN[1], Ayça ÇAKMAK PEHLİVANLI[1]
bilge.baser@msgsu.edu.tr, metin.yangin@msgsu.edu.tr, ayca.pehlivanli@msgsu.edu.tr

[1]*Mimar Sinan Fine Arts University, Statistics Department, Bomonti, İstanbul, Turkey*

HIV (human immunodeficiency virus) is a virus that attacks the immune system and making people much more vulnerable to infections and diseases. The HIV-1 protease is an important enzyme which is responsible of an imperative part in viral life cycle. The HIV-1 protease is a distinct target for the rational antiviral drug design because it is crucial for a successful viral replication. It cleaves the proteins to their component peptides and generates mature infectious particle. For this reason, HIV-1 protease enzyme inhibitor is one of the ways of struggling with HIV.

In recent works, it is observed that, HIV-1 protease prefers non-small and hydrophobic amino acids on both sides of the scissile bond [1]. Hsu, has also suggested for future research to focus on ways to inhibit the mutated cleavage sites. If cleavage site mutations are a rate limiting step in resistance development, simultaneous inhibition of cleavage site and protease could be very effective; HIV would have to mutate at both the protease and the cleavage site simultaneously to develop resistance [2].

In this study, it is the first time that combination of the game theory philosophy and HIV-1 protease cleavage site modelling. To address this approach, a two-player noncooperative game is designed with the players as HIV and inhibitor. The hydrophobicity values [3], the volumes [4], the relative mutabilities [5] of amino acids and the weighted frequencies of cleaved amino acids' combinations on both sides of the scissile bond in 1625 data set are used for generating the utility functions of both players. The choices of players are composed of all permutations of the two amino acids which are located on both sides of the scissile bond.

An epistemic model is constructed by using the utility function for each player and for each rational choice, that there is a type that expresses common belief in rationality and the types obtained are used for modelling HIV-1 protease preferability over the amino acids permutations.

Keywords: *HIV-1 protease cleavage sites, Epistemic Game Theory*

References

[1] You, L., Garwicz, D., Rögnvaldsson T. (2005), *Comprehensive Bioinformatic Analysis of the Specificity of Human Immunodeficiency Virus Type 1 Protease*, Journal of Virology, Vol.79, No.19, p.12477-12486.

[2] URL: https://web.stanford.edu/~siegelr/philhsu.htm *Accessed date*: 15/11/2017.

[3] URL: https://www.sigmaaldrich.com/life-science/metabolomics/learning-center/amino-acid-reference-chart.html *Accessed date*: 17/11/2017.

[4] Pommié C et al. (2004), *IMGT standardized criteria for statistical analysis of immunoglobulin V-REGION amino acid properties*, J Mol Recognit, Vol. Jan-Feb (17) 1, p.17-32.

[5] Pevsner, J. (2009), *Bioinformatics and Functional Genomics*, USA, Wiley-Blackwell, p.63.

Linear Mixed Effects Modelling for Non-Gaussian Repeated Measurement Data

Özgür Asar[1], David Bolin[2], Peter J Diggle[3], Jonas Wallin[4]
ozgur.asar@acibadem.edu.tr[1], david.bolin@chalmers.se[2], p.diggle@lancaster.ac.uk[3], jonas.wallin@stat.lu.se[4]

[1]Department of Biostatistics and Medical Informatics, Acıbadem Mehmet Ali Aydınlar University, Turkey
[2]Mathematical Sciences, Chalmers University of Technology and the University of Gothenburg, Gothenburg, Sweden
[3]CHICAS, Lancaster Medical School, Lancaster University, Lancaster, United Kingdom
[4]Department of Statistics, Lund University, Lund, Sweden

In this study, we consider linear mixed effects models with non-Gaussian random components for analysis of longitudinal data with large number of repeats [1]. The modelling framework postulates that observed outcomes can be de-composed into fixed effects, subject-specific random effects, a continuous-time stochastic process, and random noise [1, 2]. Likelihood-based inference is implemented by a computationally efficient stochastic gradient algorithm. Random components are predicted by either of filtering or smoothing distributions. The R package **ngme** provides functions to implement the methodology.

Keywords: *longitudinal data analysis, random-effects modelling, stochastic modelling*

References

[1] Asar Ö, Ritchie JP, Kalra PA and Diggle PJ (2016). *Short-term and long-term effects of acute kidney injury in chronic kidney disease patients: A longitudinal analysis*. Biometrical Journal, 58(6), 1552-1566.

[2] Diggle PJ, Sousa I and Asar Ö (2015). *Real-time monitoring of progression towards renal failure in primary care patients*. Biostatistics, 16(3), 522-536.

SESSION III

OPERATIONAL RESEARCH I

10TH ISC2017 INTERNATIONAL STATISTICS CONGRESS
December 6-8, 2017 ANKARA/TURKEY

A Robust Monte Carlo Approach for Interval-Valued Data Regression

Esra AKDENİZ[1], Ufuk BEYAZTAŞ[2], Beste BEYAZTAŞ[3]
ufuk,beyaztas@gmail.com, beste.sertdemir@medeniyet.edu.tr, esraakdeniz@gmail.com
[1]*Marmara University, Biostatistics Divison, İstanbul, Turkey*
[2]*Bartın University, Department of Statistics, Bartın, Turkey*
[3]*İstanbul Medeniyet University, Department of Statistics, İstanbul, Turkey*

Interval-valued data are observed with lower and upper bounds, representing uncertainty or variability. Interval-valued data often arise as a result of aggregation with the trend of big data. Regression methods for interval-valued data have been increasingly studied in recent years. The proposed procedures however are very sensitive to the presence of outliers, which might lead to poor fit of the data. This paper considers the robust estimation of the regression parameters for interval-valued data when there are outliers in the data set. We propose a new robust approach to fit a linear model combining the resampling idea and Hellinger-distance. The new procedure, called robust Monte Carlo Method (MCM) is compared with the method proposed by Ahn et al. (2012) by means of MSEs of regression coefficients, length of confidence intervals, coverage probabilities, lower and upper bound root mean-square errors demonstrating a better performance. An application is also demonstrated on a blood pressure data set to show the usefulness of the proposed method.

Keywords: *interval-valued data, robust regression, third key, Hellinger-distance*

References

[1] Ahn, J., Peng, M., Park, C., & Jeon, Y. (2012*). A resampling approach for interval- valued data regression.* Statistical Analysis and Data Mining: The ASA Data Science Journal, 5(4), 336-348.

[2] Billard, L., & Diday, E. (2000). *Regression analysis for interval-valued data. In Data Analysis,* Classification, and Related Methods (pp. 369-374). Springer, Berlin, Heidelberg.

[3] Sun, Y. (2016). *Linear regression with interval- valued data.* Wiley Interdisciplinary Reviews: Computational Statistics, 8(1), 54-60.

[4] Markatou, M. (1996). *Robust statistical inference: weighted likelihoods or usual m-estimation?.* Communications in Statistics--Theory and Methods, 25(11), 2597-2613.

ISC2017 INTERNATIONAL STATISTICS CONGRESS
December 6-8, 2017 ANKARA/TURKEY

sNBLDA: Sparse Negative Binomial Linear Discriminant Analysis

Dinçer GÖKSÜLÜK, Merve BAŞOL, Duygu AYDIN HAKLI
dincer.goksuluk@hacetttepe.edu.tr

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

In molecular biology, gene-expression based studies have great importance on examining the transcriptional activities in different tissue samples or cell populations [1]. With the recent advances, it is now feasible to examine the expression levels of thousands of genes at the same time. This leads researchers to focus on multiple analysis tasks: (i) clustering, (ii) differential expression and (iii) classification. Microarray and next-generation sequencing (NGS) technologies are the recent high-throughput technologies for quantifying gene expression. RNA sequencing (RNA-Seq), which is more recent technology than microarray, is the technique which uses the capabilities of NGS technology to characterize and quantify gene expression [2]. Microarray data consist of continuous values which are obtained from the log intensities of image spots. RNA-Seq, on the other hand, contains discrete count values which represent the RNA abundances with the number of sequence reads mapped to a reference genome or transcriptome. Hence, microarray-based algorithms are not directly applicable to RNA-Seq data since the underlying distribution of RNA-Seq data is totally different than microarrays. In a classification task, Poisson Linear Discriminant Analysis (PLDA) and Negative Binomial Linear Discriminant Analysis (NBLDA) are developed for RNA-Seq data [3, 4]. NBLDA should be preferred over PLDA when there is significant overdispersions. PLDA is a sparse method and able to select best subset of genes while fitting the model. However, NBLDA is not sparse and keeps all the genes (possibly thousands of genes) in the model even though most genes poorly contribute to discrimination function. In this study, we aim to develop sparse version of NBLDA by shrinking overdispersion parameter towards 1. With this improvement, insignificant genes can be removed from discriminant function. In addition, the complexity of the model is decreased in sparse models. The accuracy and sparsity of proposed model is compared to PLDA and NBLDA. Results showed that shrinking overdispersion towards 1 contributed to model simplicity by selecting a subset of genes. Although the accuracy of proposed model was similar (or better) with PLDA and NBLDA, the complexity of the model was lower.

Keywords: *classification, negative binomial distribution, RNA sequencing, gene expression*

References

[1]Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. (2015). limma Powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research*. 43(7):e47.

[2]Marioni JC, Mason CE, Mane SM, Stephens M, Gilad Y (2008). RNA-seq: An assessment of technical reproducibility and comparison with gene expression arrays. *Genome Research*. 18(9):1509–1517.

[3]Witten DM (2011). Classification and clustering of sequencing data using a Poisson model. *Annals of Applied Statistics*. 5:2493–2518.

[4]Dong K, Zhao H, Tong T, Wan X (2016). NBLDA: negative binomial linear discriminant analysis for RNA-seq data. *BMC Bioinformatics*. 17(1):369.

10TH ISC2017 INTERNATIONAL STATISTICS CONGRESS
December 6-8, 2017 ANKARA/TURKEY

Modelling Dependence Between Claim Frequency and Claim Severity: Copula Approach

Aslıhan ŞENTÜRK ACAR[1], Uğur KARABEY[1]
aslihans@hacettepe.edu.tr, ukarabey@hacettepe.edu.tr

[1]*Hacettepe University Department of Actuarial Science, Ankara, Turkey*

Claim frequency and claim severity are two main components used for premium estimation in non-life insurance. Frequency component represents the claim count while severity component represents the claim amount conditional on positive claim. Basic pricing approach relies on independence assumption between two components and the loss is obtained as the product of them. Independence assumption is restrictive and ignoring the dependence could lead to biased estimates. One of the possible ways to model dependence between claim severity and claim frequency is to jointly model the two components using copula approach ([1], [2], [3]).

In this study, dependence between claim severity and claim frequency is modelled with copula approach using a health insurance data set from a Turkish insurance company. Marginal distributions are specified using goodness of fit statistics and generalized linear models are fitted to both variables as margins. Mixed copula approach is used to obtain joint distribution of claim frequency and claim severity with different copulas. Results are compared.

Keywords: *copula, dependence, joint distribution, health insurance.*

References

[1] Czado, C., Kastenmeier, R., Brechmann, E. C., Min, A. (2012*). A mixed copula model for insurance claims and claim sizes*, Scandinavian Actuarial Journal, 2012(4), 278-305.

[2] Frees, E. W., Valdez, E. A., (2008), *Hierarchical insurance claims modeling*, Journal of the American Statistical Association, 103(484), 1457-1469.

[3] Krämer, N., Brechmann, E. C., Silvestrini, D., Czado, C. (2013). *Total loss estimation using copula-based regression models*, Insurance: Mathematics and Economics, 53(3), 829-839.

10TH ISC2017 INTERNATIONAL STATISTICS CONGRESS
December 6-8, 2017 ANKARA/TURKEY

Detection of Outliers Using Fourier Transform

Ekin Can ERKUŞ[1], Vilda PURUTÇUOĞLU[1, 2], Melih AĞRAZ[2]
ekincanerkus@hotmail.com, vpurutcu@metu.edu.tr, melih.agraz@yahoo.com

[1]*Department of Biomedical Engineering, Middle East Technical University, Ankara, Turkey*
[2]*Department of Statistics, Middle East Technical University, Ankara, Turkey*

The detection of outliers is one of the well-known challenges in data analyses since the outliers affect the outcomes of the analyses considerably. Therefore, it is typically used as the pre-processing step in advance of any modelling. Hereby, many parametric and non-parametric methods have been suggested to both detect the number of outliers and their locations in the dataset. Among many alternatives, the z-score test and the box-plot analysis can be though as two common parametric and non-parametric outlier detection method, respectively [1, 2].

Accordingly, in this study, we propose a novel non-parametric outlier detection method which is based on the Fourier transform [3] and is specifically used for time-series data, but can be also performed for no time-series data too. In our analyses, we implement this approach to find sparse and relatively high percentage of outliers under distinct number of observations. Furthermore, we consider that the outliers can be allocated periodically, e.g. putting in at every 5^{th} or 10^{th} observation, or aperiodically. As a result of the normally distributed datasets under various Monte Carlo scenarios, it is seen that our proposal method can more successfully detect both the number of outliers and their locations in the datasets than the findings of the z-score and box-plot approaches. Moreover, it is computationally less demanding that its competitors. Hence, we consider that our new method can be a promising alternative to find the outliers in the data under different conditions.

Keywords: *Fourier transform, outlier detection, Monte Carlo simulations*

References

[1] Ben-Gal, I. (2005), *Data mining and knowledge discovery handbook: a complete guide for practitioners and researchers,* Germany, Springer Science and Business Media, 117-130.

[2] Kutner, M. H., Nachtsheim, C. J., Neter, J. and Li, W. (2005), *Applied linear statistical models,* USA, McGraw-Hill, 390-400.

[2] Oppenheim, A. V., Willsky, A. S. and Withian, I. Y. (1983) *Signals and systems,* USA, Prentice-Hall International, 161-212.

10TH ISC2017 INTERNATIONAL STATISTICS CONGRESS
December 6-8, 2017 ANKARA/TURKEY

A perspective on analysis of loss ratio and Value at Risk under Aggregate Stop Loss Reinsurance

Başak Bulut Karageyik[1], Uğur Karabey[1]
basakbulut@hacettepe.edu.tr, ukarabey@hacettepe.edu.tr

[1]*Hacettepe University, Department of Actuarial Sciences, Beytepe, Ankara, Turkey,*

Reinsurance arrangement can be a prevalent risk management solution for the insurance companies. Aggregate stop-loss reinsurance is designed to protect an insurance company's overall losses among a specified loss ratio. Hence, the reinsurance company is obliged to cover the risks that exceed the pre-determined loss ratio.

Reinsurance agreements reduce the insurer's risk while increasing the insurance costs due to high reinsurance premiums. In most reinsurance studies Value at Risk is a widely used and effective risk management tool to ensure the optimal decision making.

In this work, we analyse the relevance of the confidence level of Value at Risk and the specified loss ratio under the aggregate stop-loss reinsurance arrangement. An application on Turkish agricultural insurance data is provided.

Keywords: *reinsurance; aggregate stop-loss reinsurance; Value at Risk*

References

[1] Dickson, D.C.M. (2005), *Insurance Risk and Ruin*, Cambridge University Press, Cambridge, 229p.

[2] Jorion, P. (1997), *Value at Risk: The New Benchmark for Controlling Market Risk*, Irwin Professional Pub, Chicago, 332p.

[3] Munich Reinsurance America-Munich RE (2010), *Reinsurance: A Basic Guide to Facultative and Treaty Reinsurance*, Princeton, 78p.

SESSION III

OPERATIONAL RESEARCH II

ISC2017 INTERNATIONAL STATISTICS CONGRESS
December 6-8, 2017 ANKARA/TURKEY

A Comparison of Goodness of Fit Tests of Rayleigh Distribution Against Nakagami Distribution

Deniz OZONUR[1], Hatice Tül Kübra AKDUR[1], Hülya BAYRAK[1]
denizozonur@gazi.edu.tr, hkubrasenol@gmail.com, hbayrak@gazi.edu.tr

[1]*Gazi University, Department of Statistics, Ankara, Turkey*

Nakagami distribution is one of the most common distributions used to model positive valued and right skewed data and widely used in a number of disciplines, especially in the analysis of the fading of radio and ultrasound signals. Recently, it has also been applied in other fields including hydrology and seismology. The distribution includes Rayleigh distribution as a special case. The purpose of the study is to apply tests of goodness of fit of Rayleigh distribution against Nakagami distribution. Specificially we applied likelihood ratio, $C(\alpha)$ and score tests. The goodness of fit tests are then compared in terms of empirical size and power using a simulation study.

Keywords: *Nakagami distribution, Rayleigh distribution, Likelihood Ratio,* $C(\alpha)$*, Score*

References

[1] Cheng, J., & Beaulieu, N. C. (2001). *Maximum-likelihood based estimation of the Nakagami m parameter*. IEEE Communications letters, 5(3), 101-103.

[2] Özonur, D., Gökpınar, F., Gökpınar, E., & Bayrak, H. (2016). *Goodness of fit tests for Nakagami distribution based on smooth tests*. Communications in Statistics-Theory and Methods, 45(7), 1876-1886.

[3] Schwartz, J., Godwin, R. T., & Giles, D. E. (2013). *Improved maximum-likelihood estimation of the shape parameter in the Nakagami distribution*. Journal of Statistical Computation and Simulation, 83(3), 434-445.

[4] Shankar, P. M., Piccoli, C. W., Reid, J. M., Forsberg, F., & Goldberg, B. B. (2005). *Application of the compound probability density function for characterization of breast masses in ultrasound B scans*. Physics in medicine and biology, 50(10), 2241.

10TH ISC2017 INTERNATIONAL STATISTICS CONGRESS
December 6-8, 2017 ANKARA/TURKEY

Generalized Entropy Optimization Methods on Leukemia Remission Times

Aladdin SHAMILOV[1], Sevda OZDEMIR[2], H. Eray CELIK[3]
asamilov@anadolu.edu.tr, sevdaozdemir@yyu.edu.tr, ecelik@yyu.edu.tr

[1]*Faculty of Science, Department of Statistics, Anadolu University, Eskişehir, Turkey*
[2]*Ozalp Vocational School, Accountancy and Tax Department, Yuzuncu Yil University, Van, Turkey*
[3]*Faculty of Science, Department of Statistics, Yuzuncu Yil University, Van, Turkey*

In this paper, survival data analysis is realized by applying Generalized Entropy Optimization Methods (GEOM). It is known that all statistical distributions can be obtained as $MaxEnt$ distribution by choosing corresponding moment functions. However, Generalized Entropy Optimization Distributions (GEOD) in the form of $MinMaxEnt$, $MaxMaxEnt$ distributions which are obtained on basis of Shannon measure and supplementary optimization with respect to characterizing moment functions, more exactly represents the given statistical data. In this research, the data for 21 leukemia patients is treated with 6-MP and the times to remission are examined (1983). The performances of GEOD are established by Chi-Square criteria, Root Mean Square Error (RMSE) criteria and Shannon entropy measure. Comparison of GEOD with each other in the difference senses shows that along of these distributions $(MinMaxEnt)_5$ is better in the senses of Shannon measure RMSE and Chi-Square criteria. Moreover, the distribution that the data set fits is computed by the method of survival data analysis with aid of the software R and in the sense of RMSE criteria, $(MinMaxEnt)_5$ distribution explains the data set better than survival distribution. For this reason, survival data analysis by GEOD acquire a new significance. The results are acquired by using statistical software MATLAB.

Keywords: *Generalized Entropy Optimization Methods, MaxEnt, MinMaxEnt Distributions, Survival Distribution*

References

[1] Deshpande & Purohit. (2005), *Life Time Data: Statistical Models and Methods.* India: Series on Quality, Reliability and Engineering Statistics.

[2] Shamilov (2007). *Generalized Entropy Optimization Problems And The Existence of Their Solutions*. Physica A: Statistical Mechanics and its Applications (382(2)) 465-472.

[3] Shamilov. (2009), *Entropy, Information and Entropy Optimization*, Eskisehir: T.C. Anadolu University Publisher, 54.

[4] Shamilov (2010). *Generalized entropy optimization problems with finite moment functions sets*. Journal of Statistics and Management Systems (Vol. 13, Issue 3) 595-603.

[5] Shamilov, Kalathilparmbil, Ozdemir (2017). *An Application of Generalized Entropy Optimization Methods in Survival Data Analysis*. Journel of Modern Phsics (8) 349-364.

10TH ISC2017 INTERNATIONAL STATISTICS CONGRESS
December 6-8, 2017 ANKARA/TURKEY

The Province on the Basis of Deposit and Credit Efficiency (2007 – 2016)

Mehmet ÖKSÜZKAYA[1], Murat ATAN[2], Sibel ATAN[2]
mehmetoksuzkaya@kku.edu.tr, atan@gazi.edu.tr, sduman@gazi.edu.tr

[1]*Kırıkkale University, Faculty of Economics and Administrative Sciences, Department of Econometrics, Kırıkkale / Turkey*
[2]*Gazi University, Faculty of Economics and Administrative Sciences, Department of Econometrics, Ankara / Turkey*

Banks face credit risk when they deposit their deposits from the market. Current account deficit of the country, debt stock, inflation, international credibility etc. are macro variables that create credit risk. On the other hand, asset and liability quality, liquidity position, credit quality and management quality, etc. variables are micro risk variables. The perceptions that arise from concepts such as uncertainties and regulations, market and country risks, especially in financial markets, negatively affect financial markets. In this case, the effects of deposits and loans on the banking sector are increased. In this study, it is aimed to calculate the relative efficiency values of the deposits and credit efficiency of the year 2007 - 2016 annual accounts using the total factor productivity of Malmquist by using the number of branches, number of bank employees, deposits and credit distributions as provincial branches of banks operating in the Turkish banking sector. The outcome of the study was assessed both in provincial and regional contexts. A mixed approach has been used in the efficiency measurement phase in the provincial banking sector. Accordingly, the number of branches and personnel were used as inputs and deposits and loans were used as outputs. Changes in technical efficiency, technological efficiency change, change in pure efficiency, change in scale efficiency and change in total factor productivity were calculated for provinces. As a result of the study, it was attempted to evaluate the increases in technological change index and technological change index and the increase in total factor productivity index in terms of banking inputs and outputs.

Keywords: *Banking Sector, Malmquist Total Factor Productivity Index (TFV), Efficiency, Provinces*

References

[1] Coelli, T. J., (1996). *A guide to DEAP Version 2.1: A Data Envelopment Analysis (Computer) Program*, CEPA Working Papers, 8/96, Department of Econometrics, University of New England, Australia, 1 - 49.

[2] Kılıçkaplan, S., Atan, M., Hayırsever, F., (2004), *Avrupa Birliği'nin Genişleme Sürecinde Türkiye Sigortacılık Sektöründe Hayat Dışı Alanda Faaliyet Gösteren Şirketlerin Verimliliklerinin Değerlendirilmesi*, Marmara Üniversitesi Bankacılık ve Sigortacılık Enstitüsü & Bankacılık ve Sigortacılık Yüksekokulu Geleneksel Finans Sempozyumu 2004, İMKB Konferans Salonu, 27 - 28 Mayıs, İstinye/İstanbul.

[3] Öksüzkaya, M., Atan, M., (2017), *Türk Bankacılık Sektörünün Etkinliğinin Bulanık Veri Zarflama Analizi ile Ölçülmesi*, Uluslararası İktisadi ve İdari İncelemeler Dergisi, Cilt 1, Sayı 18, Sayfa: 355 – 376.

[4] Akyüz, Yılmaz Yıldız, Feyyaz Kaya, Zübeyde, (2013), *Veri Zarflama Analizi (VZA) ve Malmquist Endeksi ile Toplam Faktör Verimlilik Ölçümü: Bist'te İşlem Gören Mevduat Bankaları Üzerine Bir Uygulama*, Atatürk Üniversitesi İktisadi ve İdari Bilimler Dergisi, Cilt: 27,Sayı:4, 110 – 130.

10TH ISC2017 INTERNATIONAL STATISTICS CONGRESS
December 6-8, 2017 ANKARA/TURKEY

On the WABL Ddefuzzification Operator for Discrete Fuzzy Numbers

Rahila ABDULLAYEVA[1], Resmiye NASIBOGLU[2]
rahila009@yahoo.com, resmiye.nasiboglu@deu.edu.tr

[1]*Department of Informatics, Sumgait State University, Sumgait, Azerbaijan*
[2]*Department of Computer Science, Dokuz Eylul University, Izmir, Turkey*

Let A - be fuzzy number given by means of LR-representation. The Weighted Averaging Based on Levels (WABL) operator for a fuzzy number A is calculated as below [1-3]:

$$WABL(A) = \int_0^1 (cR_A(\alpha) + (1-c)L_A(\alpha))p(\alpha)d\alpha, \quad (1)$$

where $c \in [0,1]$ is the "optimism" coefficient of the decision maker's strategy. The $p(\alpha)$ is a degree-importance function that is proposed as linear, quadratic etc. patterns up to value of the parameter k in [1, 2]:

$$p(\alpha) = (k+1)\alpha^k, \ k = 0, 1, 2, \dots \quad (2)$$

Based on this definition, a lot of methods can be constructed for obtaining these parameters (the degree-importance function and the optimism parameter). This allows the method to gain flexibility

The above formulations are valid for continuous universe and with continuous level interval [0, 1]. But in many situations, fuzzy information is operated for a given discrete universe $U = \{x_1, x_2, \dots, x_n | x_i \in R, i = 1, \dots, n\}$ and for a given discrete values of the membership degrees:

$$\Lambda = \{\alpha_0, \alpha_1, \dots, \alpha_t | \alpha_i \in [0,1]; \ \alpha_0 < \alpha_1 < \dots < \alpha_t\}. \quad (3)$$

Such fuzzy numbers are called discrete fuzzy numbers. In this case, the WABL value of the discrete fuzzy number can be formulated as follows:

$$WABL(A) = \sum_{\alpha\in\Lambda} p_\alpha (cR_\alpha + (1-c)L_\alpha), \quad \sum_{\alpha\in\Lambda} p_\alpha = 1, p_\alpha \geq 0, \ \forall\alpha \in \Lambda. \quad (4)$$

In our study, we investigate and prove analytical formulas to facilitate the calculation of WABL values for discrete trapezoidal fuzzy numbers $A = (l, m_l, m_r, r)$ with constant, linear and quadratic form degree importance functions of level weights.

Keywords: *fuzzy number, WABL operator, defuzzification.*

References

[1] Dubois D., Prade H. (1987), The Mean Value of a Fuzzy Number, *Fuzzy Sets and Systems*, 24, 279–300.

[2] Nasibov E.N. (2002), Certain Integral Characteristics of Fuzzy Numbers and a Visual Interactive Method for Choosing the Strategy of Their Calculation, *Journal of Comp. and System Sci. Int.*, 41, No.4, pp. 584-590.

[3] Nasibov E.N., Mert A. (2007), On Methods of Defuzzification of Parametrically Represented Fuzzy Numbers, *Automatic Control and Computer Sciences*, 41, No.5, pp. 265-273.

10TH ISC2017 INTERNATIONAL STATISTICS CONGRESS
December 6-8, 2017 ANKARA/TURKEY

Performance Comparison of the Distance Metrics in Fuzzy Clustering of Burn Images

Yeşim AKBAŞ[1], Tolga BERBER[1]
yesimyeginoglu@ktu.edu.tr, tolga.berber@fen.ktu.edu.tr

[1]*Faculty of Science, Department of Statistics and Computer Sciences, Karadeniz Technical University, Trabzon, TURKEY*

Statistical methods have been using in burn diagnosis, as well as in many medical fields. The fact that the annual number of deaths is determined as 180,000 by the World Health Organization in 2017, clearly reveals the importance of burn wound diagnosis. Percentage of burn is the one of the most important parameters which are needed to be determined in the planning of burn wound treatment. However, there is no accepted numerical approach available to calculate this parameter.

In this study, fuzzy clustering method [1, 3] have been used to determine the burn / normal skin [4] regions in order to calculate burn area percentage. We selected 10 sample images were from the burn wound image dataset of the patients who applied to the burn unit of the Karadeniz Technical University Faculty of Medicine Farabi Hospital. The information of each burn image is aggregated, then clustering is done for a single set of information (approximately 5 million data points). Although Euclidean distance is the most commonly used distance metric in image clustering methods, we examined the effects of different distance metrics on the clustering of burn wounds, in this study. We have evaluated the clustering performance of Euclidean, Cityblock (Manhattan), Jaccard, Cosine, Chebyshev, Minkowski distance metrics [2] to be used in FCM for all clusters C = [2, 20]. We measured the performance of the distance metrics in terms of PBMF validity measure which has proven success rates [5]. As a result, we found that the CityBlock distance metric gives the best result with 17 clusters.

Keywords: *Burn images, FCM, Distance Metrics*

References

[1] Badea, M.S., Felea, I.I., Florea, L.M., and Vertan, C., *The use of deep learning in image segmentation, classification and detection*. 2016.

[2] Deza, E. and Deza, M. M. (2009), Encyclopedia of Distances. Berlin, Heidelberg: Springer Berlin Heidelberg.

[3] Höppner, F., Klawonn, F., Kruse, R., and Runkler, T. (1999), Fuzzy Cluster Analysis: Methods for Classification, Data Analysis and Image Recognition. England: John Wiley & Sons Ltd.

[4] Suvarna, M., Sivakumar, and Niranjan, U.C.(2013), "Classification Methods of Skin Burn Images," *Int. J. Comput. Sci. Inf. Technol.*, vol. 5, no. 1, pp. 109–118.

[5] Wang, W. and Zhang, Y. (2007), "On fuzzy cluster validity indices," Fuzzy Sets Syst., vol. 158, no. 19, pp. 2095–2117.

SESSION IV

APPLIED STATISTICS V

Correspondence Analysis (CA) on Influence of Geographic Location to Children Health.

Pius Martin [1], *Peter Josephat* [2]
Martinpius01@gmail.com, mtakwimu@yahoo.com

[1]*Hacettepe University, Ankara, Turkey*
[2]*University of Dodoma (UDOM), Dodoma, Tanzania.*

This paper present a simple correspondence analysis (CA) of the primary data collected from 19 regions of Tanzania mainland (2013) with the general objective of identifying relationship between geographical location and health issues affecting children who are under 5 years old. Focus of the paper to children health was driven by the fact that according to various studies, child mortality rate is still high especially in sub-Saharan Africa Tanzania included.

For analysis, regions were further categorized into 5 zones namely Northern, Eastern, Southern, Western and Central zones. Meanwhile, at each zone various health problems affecting children were identified and categorized into 6 groups including Malaria, HIV-Aids, UTI/Fever, Physical/Skin problems, Stomach/Chest complications and Malnutrition/Obesity.

As an alternative to chi-square and a powerful multivariate technique in assessing relationship between two categorical variables at the level of category, CA was applied to our data by treating Zones as a row variable and Sickness as the column variable.

From our results we found that Chest/stomach complications is more connected to the Northern zone. Also a cluster of malaria and UTI/fever were more connected to Central and Eastern zones. Physical/skin problem is more connected with the Western zone. Apart from Southern zone HIV/Aids is not very far from either of the remaining zones. Southern was associated more with malaria. Finally, we have Malnutrition/Obesity located far from either of the zones which implies that although our variables were highly associated, not all categories will be related.
Therefore, holding other factors constant we can conclude that geographical location is associated with health problems facing under 5 population in Tanzania.

Keywords: *CA, HIV-AIDS, UTI.*

References

[1] Doey, L. and Kurta, J. (2011). Correspondence Analysis Applied to Psychological Research. *Tutorials in Quantitative Methods for Psychology*, Vol. 7(1): 5 – 14.

[2] Nagpaul PS. (1999). *Guide to advanced data analysis using IDAMS software.* New Delhi: United Nations Educational, Scientific and cultural Organization.Correspondence analysis

[3] Sourial N, Wolfson C, Zhu B, et al. (2010). *Correspondence analysis is a useful tool to uncover the relationships among categorical variables.* Journal of clinical epidemiology.63(6):638-646. doi:10.1016/j.jclinepi.2009.08.008.

[4] Sourial, N., Wolfson, C., Zhu, B., Quail, J., Fletcher, J., Karunananthan, S., Bandeen-Roche, K., Béland, F. Bergman, H. (2010). *Correspondence Analysis is a Useful Tool to Uncover the Relationships among Categorical Variables*. J Clin Epidemiol, 63(6): 638–646.

10TH ISC2017 INTERNATIONAL STATISTICS CONGRESS
December 6-8, 2017 ANKARA/TURKEY

Cluster Based Model Selection Method for Nested Logistic Regression Models

Özge GÜRER[1], Zeynep KALAYLIOGLU[2]
otanju@ankara.edu.tr, kzeynep@metu.edu.tr

[1]*Ankara University, Ankara, Turkey*
[2]*Middle East Technical University, Ankara, Turkey*

A parsimonious model explains the data with minimum number of covariates. Model selection methods are important to identify such models. Overfitting problem is one of the mostly encountered problems in model selection [1-2]. Especially in clinical, biological and social studies, researchers examine too many covariates. Therefore, tendency to overfit increases. The focus of this study is model selection in nested models with too many variables. We propose a new approach for logistic regression based on the distance between two cluster trees. We aim to overcome overfitting problems by use of a proper penalty term. This cluster tree based method is evaluated in an extensive simulation study. It is also compared with commonly used information based methods. Simulation scenarios include the cases when the true model is in the candidate set or not. Results reveal that this new method is highly promising. At the end, a real data analysis is also conducted to identify the risk factors of breast cancer.

Keywords: *model selection, overfitting, cluster tree, logistic regression, nested models*

References

[1] Babyak, M. A., (2004). *What you see may not be what you get: a brief, nontechinal introduction to overfitting in regression-type models.* Psychosom Med, 66, 411-21.

[2] Hawkins, D. M., (2004). *The problem of overfitting.* J Chem Inf Comput Sci, 44,1-12.

Dependence Analysis with Normally Distributed Aggregate Claims in Stop-Loss Insurance

Özenç Murat MERT[1], A. Sevtap SELÇUK-KESTEL[1]
mmert@metu.edu.tr, skestel@metu.edu.tr

[1]*Middle East Technical University, Institute of Applied Mathematics, Ankara, Turkey*

The reinsurance contracts in the insurance market have been playing an important role in the last couple of decades. One of the most important reinsurance contracts is the stop-loss reinsurance. It has an interesting property from the insurer point of view such that it is optimal if the criterion of minimizing the variance of the cost of the insurer is used. The word 'optimality' takes many researcher's attention so that optimal reinsurance contracts under different assumptions have been investigated for decades For instance, some researchers used utility functions to find the optimal contract while the others use aggregate claims with many different distributions such as gamma and translated gamma distributions [1],[2], [3].

This study aims to examine the stop-loss contracts with priority and maximum under the assumption the aggreate claims with normal distribution. The dependence between the cost of the insurer and the cost of the reinsurer is taken into account by implementing traditional dependence measures. Additional to these, the impact of tail dependence captured by copula approach is investigated. The deterministic retention is found when the correlation between the cost of the insurer and the cost of the reinsurer is maximum. Moreover, if the contract includes a maximum, the converging of the correlation of the parties is examined according to the distance between the maximum and the priority.

Keywords: *Stop-Loss reinsurance, reinsurance cost, priority, copula*

References

[1] Castañer, Anna and Claramunt Bielsa, M. Merce, *Optimal Stop-Loss Reinsurance: A Dependence Analysis* (April 10, 2014). XREAP2014-04.

[2] Guerra, M., & Centeno, M. D. L. (2008). *Optimal reinsurance policy: The adjustment coefficient and the expected utility criteria. Insurance: Mathematics and Economics*, 42(2), 529-539.

[3] Kaluszka, M., & Okolewski, A. (2008). *An extension of Arrow's result on optimal reinsurance contract. Journal of Risk and Insurance*, 75(2), 275-288.

ISC2017 INTERNATIONAL STATISTICS CONGRESS
December 6-8, 2017 ANKARA/TURKEY

Risk Measurement Using Extreme Value Theory: The Case of BIST100 Index

Bükre YILDIRIM KÜLEKCİ[1], A. Sevtap SELÇUK-KESTEL[1], Uğur KARABEY[2]
bukre@metu.edu.tr, skestel@metu.edu.tr, ukarabey@hacettepe.edu.tr

[1]*Middle East Technical University, Institute of Applied Mathematics, Ankara, Turkey*
[2]*Hacettepe University, Actuarial Sciences, Ankara, Turkey*

In recent decades increasing incidences on instabilities and shocks in financial markets are observed. This lead to search a risk management model which incorporates rare events (tail distributions) in the modeling of financial data [3]. In statistical modelling the events which are perceived as less likely are usually neglected. An alternative option to traditional statistical modeling that estimates the complete distribution, is the Extreme Value Theory (EVT) which is based on threshold exceedance methods and deal with the behavior specifically on the tail of a distribution [4].

EVT plays an important methodological role in risk management for insurance and finance as a method for modeling and measuring risk. Among common methods, we aim to implement Peaks Over Threshold (POT) method to model the exceedances over a given threshold with Generalized Pareto Distribution (GPD) whose distribution is as follows [1][2]:

$$G_{\varepsilon,\sigma} = \begin{cases} 1 - \left(1 + \frac{\varepsilon}{\sigma} y\right)^{-\frac{1}{\varepsilon}} & if\ \varepsilon \neq 0 \\ 1 - e^{-\frac{y}{\sigma}} & if\ \varepsilon = 0 \end{cases}$$

The aim of this study is to show the perfomance of proposed model in capturing the extereme tail behaviour of financial data and illustrate if high volatility as during the the subprime crise has impact on the proposed model. For this reason, we use daily returns of Turkish market index, BIST100, from 2001 to 2017. The popular risk measures such as VaR and ES as well as their confidence intervals are computed to implement the methodology. A comparison to traditional statistical modeling to extreme value distribution in the frame of financial crises will be done.

Keywords: *Extreme value theory, VaR, ES, confidence intervals, generalized pareto distribution, maximum likelihood estimation.*

References

[1] Embrechts, P., Resnick, S. I., & Samorodnitsky, G. (1999). *Extreme value theory as a risk management tool.* North American Actuarial Journal, 3(2), 30-41.

[2] Gilli, M. (2006). *An application of extreme value theory for measuring financial risk.* Computational Economics, 27(2-3), 207-228.

[3] Embrechts, P., Klüppelberg, C., and Mikosch, T. (1997). *Modelling extremal events, volume 33 of Applications of Mathematics*.

[4] Tancredi, A., Anderson, C., and O'Hagan, A. (2006). *Accounting for threshold uncertainty in extreme value estimation.* Extremes 9.2 : 87-106.

SESSION IV

APPLIED STATISTICS VI

10TH ISC2017 INTERNATIONAL STATISTICS CONGRESS
December 6-8, 2017 ANKARA/TURKEY

Examination of Malignant Neoplasms and Revealing Relationships with Cigarette Consumption

İrem ÜNAL[1], Özlem ŞENVAR
iremu96@gmail.com, *ozlemsenvar@gmail.com*

[1]*Marmara University, Department of Industrial Engineering, Istanbul, TURKEY*

In 2010s' Turkey, approximately 20% of deaths are caused by neoplasms and malignant neoplasms constitute almost all of this percentile. There are several main reasons of carcinoma such as biological, environmental, behavioural factors, and etc.

Tobacco smoking is overwhelmingly the most significant risk factor for cancer and across the board for chronic diseases. [1] Cigarette smoking is causally related to several cancers, particularly lung cancer, yet for some cancers there are inconsistent associations. [2]

In this study, malignant neoplasms of larynx and trachea/bronchus/lung, liver and the intrahepatic bile ducts and cervix uteri, other parts of uterus, ovary and prostate are examined according to their statistics of total death by gender. These three groups of data are obtained from Turkish Statistical Institute (TUIK) years between 2009-2016 and distribution of number of death that causes these three types of malignant neoplasms are compared between each other by gender. These three groups of malignant neoplasms are analysed with trend projection and simple linear regression analysis.

The aim of this study is to reveal the relationship between cigarette consumption and the number of deaths of malignant neoplasms and to perform forecasting for cigarette consumption. According to the predicted values of cigarette consumption, the numbers of deaths of malignant neoplasms are predicted.

Interpretations are provided based on the strength of these associations via correlation analysis.

Keywords: *Trend based forecasting, Correlation, Descriptive Statistics, Healthcare Data Analyses*

References

[1] Gelband, H., & Sloan, F. A. (Eds.). (2007). *Cancer control opportunities in low-and middle-income countries*. National Academies Press.

[2] Ray, G., Henson, D. E., & Schwartz, A. M. (2010). Cigarette smoking as a cause of cancers other than lung cancer: an exploratory study using the Surveillance, Epidemiology, and End Results Program. *CHEST Journal*, *138*(3), 491-499.

10TH ISC2017 INTERNATIONAL STATISTICS CONGRESS
December 6-8, 2017 ANKARA/TURKEY

Various Ranked Set Sampling designs to construct mean charts for monitoring the skewed normal process

Derya KARAGÖZ[1], Nursel KOYUNCU[1]
deryacal@hacettepe.edu.tr, nkoyuncu@hacettepe.edu.tr

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

In recent years, the statisticians tried to take the advantage of using various sampling designs to construct control chart limits. Ranked Set Sampling (RSS) is one of the most popular sampling and effective design. Most statisticians modify this design and proposed various ranked set sampling designs. They prefer to use these sampling designs since they give more efficient estimates compared to simple random sampling (SRS). In this study, we propose to use various ranked set sampling designs to construct the mean charts based on Shewhart, Weighted Variance and Skewness Correction methods that are applied to monitor the process variability under the skewed normal process. The performance of the mean charts based on various ranked set sampling designs are compared with simple random sampling by Monte Carlo simulation. Simulation results revealed that the mean charts based on various ranked set sampling perform much better than simple random sampling.

Keywords: *Skewed normal distribution, Ranked set sampling designs, Weighted variance method, Skewness correction method.*

References

[1] Karagöz D, Hamurkaroğlu C., (2012). *Control charts for skewed distributions: Weibull, Gamma, and Lognormal*, Metodoloski zvezki - Advances in Methodology and Statistics, 9:2, 95-106.

[2] Karagöz D.,(2016). *Robust $\bar{X}$Control Chart for Monitoring the Skewed and Contaminated Process*, Hacettepe Journal of Mathematics and Statistics, DOI: 10.15672/HJMS.201611815892.

[3] Koyuncu N., (2015). *Ratio estimation of the population mean in extreme ranked set and double robust extreme ranked set sampling*. International Journal of Agricultural and Statistical Sciences, 11:1, 21-28.

[4] Koyuncu N., (2016). *New difference-cum-ratio and exponential Type estimators in median ranked set sampling*. Hacettepe Journal of Mathematics and Statistics,45:1, 207-225.

[5] Koyuncu Nursel, Karagöz Derya (2017). *New mean charts for bivariate asymmetric distributions using different ranked set sampling designs*. Quality Technology and Quantitative Management. DOI: 10.1080/16843703.2017.1321220.

10TH ISC2017 INTERNATIONAL STATISTICS CONGRESS
December 6-8, 2017 ANKARA/TURKEY

Integrating Conjoint Measurement Data to ELECTRE II: Case of University Preference Problem

Tutku TUNCALI YAMAN[1]
tutkutuncali@gmail.com,

[1]*Marmara University, Istanbul, Turkey*

Conjoint analysis has a widespread usage in determination of consumer preferences with its different approaches after it was developed in early '60s [2]. A well-known approach in conjoint measurement is called Choice-Based Conjoint (CBC) and it revealed strong acceptance in marketing research after McFadden's 1986 study [3]. Lately, conjoint scores started to use as an input for Multi Dimensional Decision Making (MCDM) methods which run a ranking procedure, such as ELECTRE (Elimination Et (and) Choice Translating Reality) [1]. The technique has six different variations, namely ELECTRE I, ELECTRE II, ELECTRE III, ELECTRE IV, ELECTRE IS and ELECTRE TRI (B-C-nC). ELECTRE II developed by scholars Roy and Bertier [4] as a MCDM technique that provides rankings and superiorities of different alternatives according to their attributes' performance scores. Evaluation method of the technique is based on pairwise comparison of alternatives by concordance & nondiscordance principle. Main objective of this demonstrative study is presenting usage of conjoint data in ELECTRE II in the context of decision-making process. The purpose of the stated approach is gathering an objective ranking among substitute private universities. ELECTRE II procedure is based on the factors affecting the private (foundation) university's preference among candidates and marketing strategies of the school administrations. Preference data were collected by CBC method from 296 students who were in the preference process after 2016 university entrance exams. According to CBC results, some of the most important factors in preference process were appeared as, "presence of the field wishing to be studied", "academic reputation of university" and "campus facilities" respectively. Conjoint scores of these factors were used to develop payoff matrix (universities vs. factors array). In order to gain weights of each factor, in-phone interviews were realized with administrations or marketing professionals of selected private universities. Proportional distribution of marketing expenses for each factor in a 100-sum scale was gained from these interviews and the collected data accepted as the weighting vector. The results obtained from both CBC and weights were used as input in ELECTRE II in order to determine a complete and objective ranking of universities. As a result of this approach, which realized by empirical data, it could be seen how the rankings differ according to student preferences when marketing strategies of universities change the weights of factors. In addition to that, this approach also allowed us to describe the market situation in general thus each university could make a comparative assessment of its own.

Keywords: *conjoint measurement, ELECTRE II, multi attribute decision making*

References

[1] Govindan, K. and Jepsen, M. B. (2015), *ELECTRE: A comprehensive literature review on methodologies and applications,* European Journal of Operational Research, 250, 1-29.

[2] Luce, N. and Tukey, N. (1964), *Simultaneous conjoint measurement: A new type of fundamental measurement,* Journal of Mathematical Psychology, 1, 1-27.

[3] McFadden, D. (1986), *Estimating Household Value of Electric Service Reliability with Market Research Data,* Marketing Science 5, 4, 275-297.

[4] Roy, B. and Bertier, P. (1971), *La méthode ELECTRE II: Une méthode de classement en présence de critères multiples,* Paris, Sema (Metra-International) Direction Scientifique, 25.

Lmmpar: a package for parallel programming in linear mixed models

Fulya GOKALP YAVUZ[1], Barret SCHLOERKE[2]
fulyagokalp@gmail.com, bschloe@purdue.edu

[1]*Yildiz Technical University, Istanbul, Turkey*
[2]*Purdue University, West Lafayette, IN, USA*

The parameter estimation procedures of linear mixed models (LMM) include some iterative algorithms, such as Expectation Maximization (EM). The consecutive steps of the algorithm require multiple iterations and cause computational bottlenecks, especially for larger data sets. LMM packages, defined in R, are not feasible for larger data sets. Speedup strategies with parallel programming reduce the computation time by spreading workload between multiple cores simultaneously. The R package 'lmmpar' [1] is introduced in this study as one of the novel applications of parallel programming with a statistical focus. The implementation results for larger data sets with 'lmmpar' package are promising in terms of using less elapsed time than the classical approach with a single core.

Keywords: *mixed models, big data, parallel programming, speedup*

References

[1] Gokalp Yavuz, F. and B. Schloerke (2017), lmmpar: Parallel Linear Mixed Model, R package.

SESSION IV

APPLIED STATISTICS VII

10TH ISC2017 INTERNATIONAL STATISTICS CONGRESS
December 6-8, 2017 ANKARA/TURKEY

Structural Equation Modelling About the Perception of Citizens Living in Çankaya District of Ankara Province Towards the Syrian Immigrants

Ali Mertcan KÖSE[1], Eylem DENİZ HOWE[1]
alimertcankose@gmail.com, eylem.deniz@msgsu.edu.tr
[1]*Mimar Sinan Fine Arts University, İstanbul, Turkey*

As is well-known, Turkey's neighbour Syria experienced extensive protests and riots starting in 2011, which lead to an environment of confusion and a state of civil war. Not only did this condition affect Syrians, but the surrounding countries were affected, and especially Turkey, who is north of Syria. One of the major impacts on neighbouring countries has been through migration; Turkey has been disproportionately impacted, due to its open-door policy for refugees. As a result, Turkish citizens have come into a significant amount of contact with refugees & migrants fleeing war-torn Syria.

The aim of this study is to statistically examine the attitudes of Turkish citizens towards Syrian migrants, to determine if the situation has led to the development of prejudicial attitudes. We have performed a correlational study to measure citizens' empathy and social orientation, along *Empathy*, *Social Orientation,* and *Threat* scales. We used the *Threat* scale in two levels, to measure perceived *Socio-economic* and *Political* threat. We also used the *Social Orientation* scale in two levels: *Social Dominance* and *Social Egalitarianism* orientation. We gathered survey responses from 418 respondents living in the Cankaya district of the Ankara province. This data was analysed with structural equation modelling (SEM), which is one of the most important multivariate statistical methods used throughout the social sciences. SEM combines confirmatory factor analysis and path analysis to show – both visually and numerically – relationships between scales. Specifically, SEM allows us to express the level and degree of the relationship between scales.

For this study, we identified dependent latent variables as *Socio-economic Threat* (THDSE) and *Political Threat* (THTDP); independent latent variables were identified as *Social Dominance* (SBYD), *Social Egalitarianism* (SBYE), and *Empathy* (EMPT). With the surveyed data, we developed the two regression equations below to test the hypothesized relationships:

$$THDSE = 0.290SBYD - 0.173SBYE - 0.021EMPT$$
$$THTDP = 0.252SBYD - 0.185SBYE + 0.155EMPT$$

For these two equations, the standard goodness-of-fit metrics are: *RMSEA = 0.063*, *SRMR = 0.08*, *CFI=0.91*, *TLI =0.90*, which supports the hypothesized existence of prejudicial attitudes. We interpret these data and results to claim that the attitudes of Social Dominance, Social Egalitarianism, and Empathy can predict a respondents' perception of a political threat from the Syrian refugees. A prejudicial perception of a socio-economic threat, however, is only correlated with Social Dominance and Social Egalitarianism.

Keywords: *Structural Equation Modelling, Refugees, Migrants, Threat, Empathy, Social Orientation*

References

[1] Beaujean, A., A.,2014, *Latent Variable Modelling Using R A step by step Guide, Routledge,* New York.

[2] Mindrila, D., 2010, *Maximum likelihood(ML) and diagonaly weighted least squares(DWLS) estima tion prosedures: A comparison of estimation bias with ordinal and multivariate non- normal data,* Internation al Journal of Digital Society, 1 (1), 60-66.

10TH ISC2017 INTERNATIONAL STATISTICS CONGRESS
December 6-8, 2017 ANKARA/TURKEY

Compare Classification Accuracy of Support Vector Machines and Decision Tree for Hepatitis Disease

Ülkü ÜNSAL[1], Fatma Sevinç KURNAZ[2], Kemal TURHAN[1]
ulkunsal@gmail.com, fskurnaz@yildiz.edu.tr, kturhan@gmail.com

[1]*Karadeniz Technical University, Trabzon, Türkiye*
[2] *Yildiz Technical University, Istanbul, Türkiye*

Hepatitis is the medical term given to inflammatory liver diseases. There are five different types of hepatitis. It estimated 325 million people were living with chronic hepatitis infections (HBV or HCV) worldwide according to WHO (World Health Organization) report in 2015. Hepatitis disease kills more than 1.3 million people each year worldwide [3].

In this study, the dataset used (*hepatitis)* were obtained from KEEL (Knowledge Extraction Based on Evolutionary Learning) database, which is publicly available website. The dataset has no specific for any type of hepatitis disease and some values are not provided. [1].

In Biostatistics field, the machine learning methods has been used to classify of diseases. In this study, we compared classification accuracy of two methods which are SVM (Support Vector Machine) and DT (Decision Tree). The accuracy of classification between two methods was performed using R program [2]. Results show that accuracy of classification was 91,3% and 86,9% respectively SVM and DT. So, SVM has higher accuracy than DT. In conclusion, SVM method should be preferred over DT method for this type of dataset.

Keywords: *Hepatitis, Support Vector Machines, Decision Tree, Classification*

References

[1] http://sci2s.ugr.es/keel/dataset.php?cod=100, last access: April 2017
[2] Torti, E., Cividini, C., Gatti, A., et al., (2017), *Euromicro Conference on Digital System Design*, Austria, The publisher, 445-450.
[3] http://www.who.int/hepatitis/en/, last access: November 2017

10TH ISC2017 INTERNATIONAL STATISTICS CONGRESS
December 6-8, 2017 ANKARA/TURKEY

Effectiveness of Three Factors on Classification Accuracy

Duygu AYDIN HAKLI[1], Merve BASOL[1], Ebru OZTURK[1], Erdem KARABULUT
duygu.aydin@hacettepe.edu.tr, merve.basol@hacettepe.edu.tr, ebru.ozturk3@hacettepe.edu.tr,

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

We aimed to compare the accuracy of the classification methods in actual data sets, as well as in the simulation study using various correlation structures, number of variables and sample size in binary classification. We used simulated datasets and actual datasets. Three different factors are considered which may affect the classification performance in a simulation study. These are sample size, correlation structure and number of variables. Scenarios were created by considering these effects. 48 different scenarios including 4 different types of correlation structure (low, medium, high level correlation and similarity of the correlation structure created by using the real data set – medium-correlated), 4 different sample size (100, 250, 500, 1000) and 3 different number of variables (15, 25 and 50) were prepared and each scenarios was repeated 1000 times. CART (Classification and Regression Tree), SVM (Support Vector Machines), RF (Random Forest) and MLP (Multi-Layer Perceptron) methods have been used in the classification of data sets obtained from both simulation and actual data sets. Accuracy, specificity, sensitivity, balanced accuracy and F-measure were used as performance measures and 10-fold cross-validation was applied. The results were interpreted considering the F-measure. Data generation, classification methods and performance were obtained using R project. In our simulation work; as the sample size increased, the performance values increased. In the case of low correlated data, the performance values increased as the number of variables increased (15-25-50 variables), while the performance values decreased at other correlation levels. As the correlation level increases, it can be said that the performances increase. In the simulation data generated with both low and real data sets' correlation structure, the performance of SVM was found to be successful according to the performance of other classification methods. The MLP method is a preferred method when there is a nonlinearity. In our simulation study, MLP's performance results are lower than SVM because we derive linearly related data.

Keywords: *sample size, correlation structure, accuracy, classification methods*

References

[1] James A. Freeman, David M. Skapura. Neural Networks (1991), *Algorithms, Applications, and Programming Techniques*, Addison Wesley,1991.

[2] Burges, C. (1998), *A Tutorial on Support Vector Machines for Pattern Recognition*, Data Mining And Knowledge Discovery, 2, 121-167.

[3] Zhong, N. - Zhou, L., Springer Verlag (1999), *Methodologies for Knowledge Discovery and Data Mining*, Third Pacific-Asia Conference,13.

10TH ISC2017 INTERNATIONAL STATISTICS CONGRESS
December 6-8, 2017 ANKARA/TURKEY

Evaluation of the Life Index Based on Data Envelopment Analysis: Quality of Life Indexes of Turkey

Volkan Soner ÖZSOY[1], Emre KOÇAK[1]
volkansoner@gazi.edu.tr, emrekocak@gazi.edu.tr

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

Most of the governments and public authorities in the world are developed "Quality of Life Indexes" to measure the quality of life for all province or regions. It is created life indexes in the provinces by The Turkish Statistical Institute using objective and subjective indicators of the lives of the citizens. This index, which takes a value between zero and one, is calculated by taking 37 variables of life together with 9 dimensions of life such as housing, work life, income and wealth, health, education, environment, safety, access to infrastructure services and social life. However, the index does not allow to examine all aspects of life on provinces and to be improved. A new index based on linear programming is proposed to overcome this shortcoming in this study. Data envelopment analysis (DEA) based on linear programming has been widely used to evaluate the relative performance of decision making units (DMUs). Efficiency score of performance for each of DMUs as provinces in this study formed the index. The index, which takes between 0 and 1, indicates a better level of life as it approaches 1.

Keywords: *Quality of Life Indexes, linear programming, performance analysis, efficiency, data envelopment analysis*

References

[1] Banker, R. D., Charnes, A., & Cooper, W. W. (1984). *Some models for estimating technical and scale inefficiencies in data envelopment analysis.* Management science, *30*(9), 1078-1092.

[2] Charnes, A., Cooper, W. W., & Rhodes, E. (1978). *Measuring the efficiency of decision making units.* European journal of operational research, *2*(6), 429-444.

[3] Turkish Statistical Institute (TURKSTAT), Provincial Life Index, (2016)

ISC2017 INTERNATIONAL STATISTICS CONGRESS
December 6-8, 2017 ANKARA/TURKEY

Measurement Errors Models with Dummy Variables

Gökhan GÖK[1], Rukiye DAĞALP[1]
g.gok01@gmail.com, rdagalp@ankara.edu.tr

[1]*Ankara University, Ankara, Turkey*

Regression analysis, sometimes the explanatory variable, X, cannot be observed, either because it is too expensive, unavailable, or mismeasured. In this situation, a substitute variable W is observed instead of X, that is $W = X + U$, where U is *measurement error*. The substitution of W for X creates problems in the analysis of the data, generally referred to as measurement error problems. The statistical models used to analyze such data are called *measurement error models*. Measurement error problems occur in many areas such as environmental, agricultural or medical investigations. For example, the amount of air pollution in environmental studies, the glucose level of a diabetic or absorption of a drug in medical investigations cannot be measured accurately.

In regression analysis the dependent variable is frequently influenced not only by ratio scale variables but also by variables that are essentially qualitative, or nominal scale. Since such variables usually indicate the presence or absence of a "quality" or an attribute, such as male or female. One way we could quantify such attributes is by constructing artificial variables that take on values of 0 or 1, 1 indicating the presence (or possession) of that attribute and 0 indicating the absence of that attribute. Variables that assume such 0 and 1 values are called *Dummy variables*. Dummy variables can be incorporated in regression models just as easily as quantitative variables.

In this study, we introduced regression models with dummy variables and measurement error models for classical linear regression, and the parameters of regression models with dummy variables were obtained. In addition, the effect of measurement error on the parameter estimation for regression models with dummy variables was examined. The obtained results were supported by using the simulation study.

Keywords: *Measurement error models, Linear regression, Dummy variables, Error in variables*

References

[1] Gujarati, D. (2002), *Basic Econometrics*, 4th ed. New York: McGraw-Hill.

[2] Dağalp, R.E. (2001), *Estimators for generalized linear measurement error models with interaction terms*, Ph.D. Thesis, Department of Statistics, North Carolina State University, USA

[3] Stefanski, L.A. (1985), *The effects of measurement error on parameter estimation*, Biometrika 72, pp. 583-592.

[4] Carroll, R.J., Ruppert, D. & Stefanski, L.A. (1995), *Measurement Error in Nonlinear Models*, Chapman & Hall/CRC

SESSION IV

OTHER STATISTICAL METHODS II

10TH ISC2017 INTERNATIONAL STATISTICS CONGRESS
December 6-8, 2017 ANKARA/TURKEY

Sorting of Decision Making Units Using Mcdm Through the Weights Obtained with Dea

Emre KOÇAK[1], Zülal TÜZÜNER[1]

emrekocak@gazi.edu.tr , zulalturkoglu@gazi.edu.tr

[1] *Gazi University Department of Statistics, Ankara, Turkey*

Multi Criteria Decision Making (MCDM) is a procedure that consists in finding the best alternatives among a set of feasible decision making units (DMUs). Although there exist many different ranking methods for DMUs, these methods can be different ranking results due to different ranking algorithms or weighting methods. A Technique for Order Preference by Similarity to an Ideal Solution (TOPSIS), which is used to solve the ranking problem, is one of the most important MCDM methods. In this study, effecient DMUs in the analysis were ranked by using the TOPSIS method with the help of the weights of the effecient DMUs obtained by data envelopment analysis (DEA). The results obtained were compared with those obtained by different weighting methods and it was found that they had a high correlation value between them.

Keywords: *MCDM, TOPSIS, Data envelopment analysis*

References

[1] Charnes, A., Cooper, W.W., and Rhodes, E.L. (1978), *Measuring the efficiency of decision making unit*, European Journal of Operational Reserach, 2(6), 429-444.

[2] Paksoy, T., Pehlivan, N.Y., and Özceylan, E. (2013), *Bulanık Küme Teorisi*. Nobel Akademik Yayıncılık.

[3] Ramanathan, R. (2003), *An Introduction to Data Envelopment Analysis-A Tool for Performance Measurement*, New Delhi, Sage Publications.

10TH ISC2017 INTERNATIONAL STATISTICS CONGRESS
December 6-8, 2017 ANKARA/TURKEY

The Health Performances of the Turkey Cities by the Mixed Integer DEA Models

Zülal TÜZÜNER[1], H. Hasan ÖRKCÜ[1], Hasan BAL[1], Volkan Soner ÖZSOY[1], Emre KOÇAK[1]
zulalturkoglu@gazi.edu.tr, hhorkcu@gazi.edu.tr, hasanbal@gazi.edu.tr, volkansoner@gazi.edu.tr, emrekocak@gazi.edu.tr

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

Data envelopment analysis (DEA), developed by Charnes, Cooper and Rhodes [3] in 1978, is a method for assessing the efficiency of decision making units (DMUs) which use the same types of inputs to produce the same kinds of outputs. The lack of discrimination has been considered as an important problem in some applications of DEA. This discrimination is necessary to rank all DMUs and select the best DMU. In order to improve discrimination property of DEA, different approaches have been proposed in the literature [1, 2].The most popular of these are approaches to finding the most efficient DMU. In this study, using the health performance indicators such as the number of doctors, the number of hospitals, the number of inpatients and the number of surgeries, the health performances of the Turkey cities are examined by Wang and Jiang [5], and Toloo [4] DEA models.

Keywords: *DEA, ranking, most efficient DMU, health performance.*

References

[1] Aldamak, A. & Zolfaghari, S., (2017). *Review of efficiency ranking methods in data envelopment analysis.* Measurement 106, 161–172.

[2] Andersen, P. M., & Petersen, N. C. (1993). *A procedure for ranking efficient units in data envelopment analysis*. Management Science, 39, 1261–1264

[3] Charnes, A., Cooper, W. W., & Rhodes, E. (1978). *Measuring the efficiency of decision making units*. European Journal of Operational Research, 2, 429–444.

[4] Wang, Y.-M., & Jiang, P. (2012). *Alternative mixed integer linear programming models for identifying the most efficient decision making unit in data envelopment analysis*. Computers & Industrial Engineering, 62, 546–553.

[5] Toloo, M. (2015). *Alternative minimax model for finding the most efficient unit in data envelopment analysis*. Computers & Industrial Engineering, 81, 186–194.

Efficiency and Spatial Regression Analysis Related to Illiteracy Rate

Zülal TÜZÜNER[1], Emre KOÇAK[1]

zulalturkoglu@gazi.edu.tr , *emrekocak@gazi.edu.tr*

[1]*Gazi University Department of Statistics, Ankara, Turkey*

Data envelopment analysis (DEA), a nonparametric method based on Linear Programming model, has been a widely used method to measure efficiencies of decision making units (DMUs). This paper examines new combinations of DEA and spatial regression analysis that can be used to evaluate efficiency within a multiple-input, multiple-output framework and spatial interaction of DMUs in terms of illiteracy. A significant correlation was found between neighboring cities with efficiency measurement by DEA. Based on statistical analysis, the SEM (spatial error models) are more appropriate than the spatial lag models (SLM) and the results of the ordinary least squares (OLS) model was compared with the appropriate model in this study.

Keywords: *Spatial regression, Data envelopment analysis, Illiteracy*

References

[1] Anselin, L. (2005), *Exploring Spatial Data with GeoDa™ : A Workbook*, University of Illinois, Urbana-Champaign.

[2] Charnes, A., Cooper, W.W., and Rhodes, E.L. (1978), *Measuring the efficiency of decision making unit*, European Journal of Operational Reserach, 2(6), 429-444.

[3] Fischer, M.M. and Getis, A. (2009), *Handbook of Applied Spatial Analysis: Software Tools, Methods and Applications*, New York, Springer, 811p

[4] Ramanathan, R. (2003), *An Introduction to Data Envelopment Analysis-A Tool for Performance Measurement*, New Delhi, Sage Publications.

10TH ISC2017 INTERNATIONAL STATISTICS CONGRESS
December 6-8, 2017 ANKARA/TURKEY

Forecasting the Tourism in Tuscany with Google Trend

Ahmet KOYUNCU[1], Monica PRATESİ[1]
koyuncuahmet@gmail.com, monica.pratesi@unipi.it

[1]*University of Pisa, Pisa, Italy*

This study aims to forecast the number of tourists arrive in Tuscany with the help of the Google Trends dataset. In the first section, search queries dataset was collected from Google Trends and operated with weights derived from the nationality of tourist arrivals in Tuscany. Information about nationality of tourist arrivals was obtained from Tuscany Tourism Report in Regional Institute for Economic Planning of Tuscany. Moreover, tourist arrivals dataset was collected from Eurostat.

Then, linear regression was performed to the investigate the correlation levels with Google Trends data and Eurostat data. Result could indicate lag between the Google Trends dataset and Eurostat dataset were also examined in studies. In this study, correlation between city arrivals data and one month lagged Google Trend data is 0,8.

In the preliminary results, the tourist arrivals in 2016 was forecasted by using ARIMA model including tourist arrivals dataset from Eurostat and then, the tourist arrivals in 2016 was estimated by using Dynamic Regression Model including the search queries dataset from Google Trends and the tourist arrivals dataset in Eurostat. The actual numbers of tourist arrivals in 2016 was discussed and compared with estimated numbers with the ARIMA model and the dynamic regression model.

Keywords: *Forecasting, Google Trend, Time Series, Tourism*

References

[1] Hyndman, R.J. and Athanasopoulos, G. (2012), *Forecasting: Principles and Practice*, OTexts. https://www.otexts.org/fpp

[2] Box, George E P, Gwilym M Jenkins, Gregory C Reinsel, and Greta M Ljung. 2015. *Time Series Analysis: Forecasting and Control. 5th ed.* Hoboken, New Jersey: John Wiley & Sons.

[3] Brockwell, Peter J, and Richard A Davis. 2016. *Introduction to Time Series and Forecasting*. 3rd ed. New York: Springer.

[4] Pankratz, Alan E. 1991. *Forecasting with Dynamic Regression Models.* New York: John Wiley & Sons.

10TH ISC2017 INTERNATIONAL STATISTICS CONGRESS
December 6-8, 2017 ANKARA/TURKEY

A New Approach to Parameter Estimation in Nonlinear Regression Models in Case of Multicollinearity

Ali ERKOÇ[1] and M. Aydın ERAR[1]
ali.erkoc@msgsu.edu.tr, aydin.erar@msgsu.edu.tr

[1]*Mimar Sinan Fine Arts University, İstanbul, Turkey*

With the advancement of science and technology, the computer modeling of data and the development of future predictive methods have become popular. By modelling of obtained data, the estimation of the next step is gained importance, specifically in applied basic sciences such as physics, chemistry, engineering, medicine, space sciences.

Although these data sets can be modelled by using linear models, the generated models are often specified by nonlinear functions, since they are derived from solving systems of differential equations. For instance, the orbit of a spacecraft or a celestial body is generally determined by nonlinear regression models. Therefore, consistent estimation of the parameters is important for the accurate estimation of the orbit.

In regression analysis, the multicollinearity is a problem that prevents consistent and reliable estimation of parameters. In nonlinear regression, the estimation of reliable and consistent parameters is crucial to make consistent predictions of the model and to represent data as good as possible.

For this purpose, in this study, a new approach to parameter estimation is presented in case of multicollinearity in nonlinear regression models. The validity of the proposed approach was tested with the simulation study.

Keywords: *Nonlinear regression, multicollinearity, parameter estimation, iterative methods.*

References

[1] Bates, D. M. & Watts, D. G., (1988). *Nonlinear Regression Analysis and Its Applications.* New York: John Wiley & Sons.

[2] Belsley, D. A., (1991). *Conditioning Diagnostics: Collinearity and Weak Data in Regression*. New York: Wiley.

[3] Crouse, r. H., Jin, C. & Hanumara, R. C., (1995). Unbiased Ridge Estimation with Prior Information and Ridge Trace. *Communications in Statistics - Theory and Methods,* 24(9), pp. 2341-2354.

[4] Montgomery, D. C., Peck, E. A. & Vining, G. G., 2012. *Introduction to Linear Regression Analysis.* New Jersey: John Wiley & Sons.

[5] Swindel, B. F., (1976). Good Ridge Estimators Based on Prior Information. *Communications in Statistics - Theory and Methods,* 5(11), pp. 1065-1075.

SESSION IV

OPERATIONAL RESEARCH III

10TH ISC2017 INTERNATIONAL STATISTICS CONGRESS
December 6-8, 2017 ANKARA/TURKEY

Author Name Disambiguation Problem: A Machine Learning Approach

Cihan AKSOP[1]
cihan.aksop@gmail.com
[1]*The Scientific and Technological Research Council of Turkey, Science and Society Department, Ankara, Turkey*

Author name disambiguation problem is mostly encountered by scholarly digital libraries such as CrossRef[1], PubMed[2], DOAJ[3], DBLP[4], academic journal editors and various staffs that needs to assign experts to evaluate projects, studies etc. From perspective of digital libraries, this problem is classification of researches and from the perspective of editors, this problem is a part of referee or expert assignment problem. Hence author name disambiguation can be defined as the problem of the identification of an author from a given set of bibliographic source.

Author name disambiguation is a difficult problem since one has to classify the authors by using bibliographic sources in which "the same author may appear under distinct names, or distinct authors may have similar names." [1]. In deep, this problem is caused by bibliographic sources which consists of variability of academic writing rules, character encoding systems, typographic errors. Recently to overcome this problem, some unique identifiers like ORCID[5] and ResearcherID[6] are being used. However there is a limitation of this identifiers since most researchers do not have such ID's. Hence these ID's are inadequate to solve the author name disambiguation problem. In the literature, several approaches was developed to give a comprehensive solution of author name disambiguation problem [1-5]. In this paper, the author name disambiguation problem was investigated on a data received from a scholarly digital libraries on the field of computer science.

Keywords: *author name disambiguation, information retrieval, decision support system*

References

[1] Ferreira, A. A., Gonçalves, M. A., and Laender, A. H. F. (2012) *A Brief Survey of Automatic Methods for Author Name Disambiguation,* SIGMOD Record, 41 (2), 15-26.

[2] Torvik, V. I., Weeber, M., Swanson, D. R., and Smalheiser, N. R., (2005) *A Probabilistic Similarity Metric for Medline Records: A Model for Author Name Disambiguation*, Journal of the American Society for Information Science and Technology, 56 (2), 140-158.

[3] Protasiewicz, J., Pedrycz, W., Kolowski, M., Dadas, S., Stanislawek, T., Kopacz, A., Galezewska, M. (2016). *A Recommender System of Reviewers and Experts in Reviewing*, Knowledge-Based Systems, 106, 164-178.

[4] Wang, F., Shi, N., and Chen, B., (2010) *A Comprehensive Survey of the Reviewer Assignment Problem*, International Journal of Information Technology & Decision Making, 9 (4), 645-668.

[5] Liu, O., Wang, J., Ma, J. and Sun, Y. (2016) *An Intelligent Decision Support Approach for Reviewer Assignment in R&D Project Selection,* Computers in Industry, 76, 1-10.

[1] https://www.crossref.org/
[2] https://www.ncbi.nlm.nih.gov/pubmed/
[3] https://doaj.org/
[4] http://dblp.uni-trier.de/
[5] https://orcid.org/
[6] http://www.researcherid.com/

10TH ISC2017 INTERNATIONAL STATISTICS CONGRESS
December 6-8, 2017 ANKARA/TURKEY

Deep Learning Optimization Algorithms for Image Recognition

Derya SOYDANER
derya.soydaner@msgsu.edu.tr

Mimar Sinan University, Department of Statistics, Istanbul, Turkey

Deep learning is an active research area to solve many big data problems such as computer vision, speech recognition and natural language processing. In recent years, it has achieved several successful results in a broad area of applications. One of the main research areas of deep learning is image recognition that has become a part of our everyday lives from biometrics to self-driving cars. Image recognition is accepted as a true challenge of artificial intelligence because these types of tasks are easy to people to perform but hard to describe. Recognizing faces and objects are carried out by people intuitively. Recent studies have shown that convolutional networks are powerful models such computer vision tasks by means of their special structure and depth. However, deep neural networks are hard to optimize and it is quite common to invest days to months of time to train a deep neural network. Therefore, new optimization algorithms have been developed for training deep networks.

In this study, optimization algorithms with adaptive learning rates are used for training of convolutional networks. The effects of these algorithms are examined and their advantages are pointed out against basic optimization algorithms on a few benchmark image recognition datasets. Besides, the challenges of deep neural network optimization are emphasized in addition to importance of determining the structure of convolutional networks.

Keywords: *Deep Learning, Convolutional Networks, Optimization, Image Recognition*

References

[1] Duchi, J., Hazan, E. and Singer, Y. (2011), *Adaptive Subgradient Methods for Online Learning and Stochastic Optimization*, Journal of Machine Learning Research, 12, 2121-2159.

[2] Goodfellow, I., Bengio, Y. and Courville, A. (2016), *Deep Learning*, Cambridge, MIT Press.

[3] Kingma, D. and Ba, J. (2014). *Adam: A Method for Stochastic Optimization*. arXiv preprint arXiv:1412.6980

[4] LeCun, Y., Bengio, Y. and Hinton, G. (2015). *Deep Learning*, Nature, 521, 436-444.

Faster Computation of Successive Bounds on the Group Betweenness Centrality

Derya DİNLER[1], Mustafa Kemal TURAL[1]
dinler@metu.edu.tr, tural@metu.edu.tr

[1]*Department of Industrial Engineering, Middle East Technical University, Ankara, Turkey*

Numerous measures have been introduced in the literature for the identification of central nodes in a network, e.g., group degree centrality, group closeness centrality, and group betweenness centrality (GBC) [1]. The GBC of a group of vertices measures the influence the group has on communications between every pair of vertices in the network assuming that information flows through the shortest paths. Given a group size, the problem of finding a group of vertices with the highest GBC is a combinatorial problem. We propose a method that computes bounds on the GBC of groups of vertices of a network. Once certain quantities related to the network are computed in the preprocessing step taking time proportional to the cube of the number of vertices in the network, our method can compute bounds on the GBC of any number of groups of vertices successively, for each group requiring a running time proportional to the square of its size. Our method is an improvement of the method in [2] which has to be restarted for each group making it less efficient for the computation of the GBC of groups successively. In addition, the bounds used in our method are stronger and/or faster to compute in general. Our computational experiments on randomly generated and real-life networks show that in the search for a group of a certain size with the highest GBC value, our method reduces the number of candidate groups substantially and in some cases the optimal group can be found without exactly computing the GBC values which is computationally more demanding.

Keywords: *centrality, betweenness, social networks, probability bounds*

References

[1] Everett, M.G. and Borgatti, S.P. (1999), *The centrality of groups and classes*, The Journal of Mathematical Sociology, 23, 181-201.

[2] Kolaczyk, E.D., Chua, D.B. and Barthlemy, M. (2009), *Group betweenness and co-betweenness: Inter-related notions of coalition centrality*, Social Networks, 31, 190-203.

10TH ISC2017 INTERNATIONAL STATISTICS CONGRESS
December 6-8, 2017 ANKARA/TURKEY

Clustering of Tree-Structured Data Objects

Derya DİNLER[1], Mustafa Kemal TURAL[1], Nur Evin ÖZDEMİREL[1]
dinler@metu.edu.tr, tural@metu.edu.tr, nurevin@metu.edu.tr

[1]*Middle East Technical University, Industrial Engineering Department, Ankara, Turkey*

Traditional data mining techniques deal with data points, i.e., data objects which are represented by numerical vectors in the space. But improving technology and measurement capabilities, and need for deeper analyses result in collecting more complex datasets [4]. Such complex datasets may include images, shapes and graphs. Consider a dataset consisting of graphs. One may aim to partition those graphs into given number of clusters. Such graph clustering problems arise in many areas like biology, neuroscience, medical imaging, computer or social networks [1]. For example, assume that we have the retinal vascular image of a patient. Branching pattern of the vessels can be represented as a rooted tree. If we have set of retinal vascular images, i.e. rooted trees, of different patients, we can cluster those trees to see the difference between the retinopathy patients and normal patients [2].

In a graph clustering problem, data objects may be general graphs, rooted trees or binary trees. Edges in those graphs can be unweighted or weighted. When the edges are unweighted, only topology is considered. In the weighted case, graphs are clustered based on one or more attributes in addition to the topology. In this study, we consider a clustering problem in which the data objects are rooted trees with unweighted or weighted edges. For the solution of the problem, we use k-means algorithm [3]. The algorithm starts with initial centroids (trees) and repeats assignment and update steps until convergence. In the assignment step, each data object is assigned to the closest centroid. To measure the distance between two trees we utilize Vertex Edge Overlap (VEO) [5]. VEO is based on an idea that if two trees share many vertices and edges, they are similar. In the update step, each centroid is updated by considering the data objects assigned to it. For both of the cases (unweighted and weighted edges), we propose Mixed Integer Nonlinear Programing (MINLP) formulations to find the centroid of a given cluster which is the tree maximizing the sum of VEOs between trees in the cluster and the centroid itself. We tested our solution approaches on the randomly generated datasets and results are promising.

Keywords: *tree-structured data objects, clustering, heuristics, optimization*

References

[1] Aggarwal, C.C. and Wang, H. (2010), *A survey of clustering algorithms for graph data*, in *Managing and mining graph data*, US, Springer, 275–301.

[2] Lu, N. and Miao, H. (2016), *Clustering tree-structured data on manifold*, IEEE transactions on pattern analysis and machine intelligence, 38, 1956–1968.

[3] MacQueen, J. (1964), *Some methods for classification and analysis of multivariate observations*, in *Proceedings of 5th Berkeley symposium on mathematical statistics and probability*, 1, 281–297.

[4] Marron, J.S. and Alonso, A.M. (2014), *Overview of object oriented data analysis*, Biometrical Journal, 56, 732–753.

[5] Papadimitriou, P., Dasdan, A. and Garcia-Molina, H. (2010), *Web graph similarity for anomaly detection*, Journal of Internet Services and Applications, 1, 19-30.

SESSION IV

DATA MINING II

10TH ISC2017 INTERNATIONAL STATISTICS CONGRESS
December 6-8, 2017 ANKARA/TURKEY

The Effect of Estimation on Ewma-R Control Chart for Monitoring Linear Profiles under Non Normality

Özlem TÜRKER BAYRAK[1], Burcu AYTAÇOĞLU[2]
ozlemt@cankaya.edu.tr, burcu.aytacoglu@ege.edu.tr

[1]*Inter-Curricular Courses Department, Statistics Unit, Çankaya University, Ankara, Turkey*
[2]*Faculty of Science, Department of Statistics, Ege University, İzmir, Turkey*

In some industrial applications, the quality of a process or product is best described by a function, called a "profile". This function or a profile expresses a relation between a response variable and explanatory variable(s) and can be modeled via many models like simple/multiple, linear/nonlinear regression, nonparametric regression, mixed models, wavelet models. The aim is to detect any change in profile over time. This study focuses on simple linear profiles. One can find several methods proposed to monitor simple linear profiles (See for example [2] and [3]). The properties of the proposed methods are usually investigated when the in control parameter values are known in Phase II analysis and the error terms are normally distributed. However, these assumptions may be invalid in most of the real life applications. There are just few studies available for investigating estimation effect under normality [1],[4] and the effect of non-normality but with known parameter values [5]. Therefore there is a need to study the estimation effect under non-normality. One of the leading methods to monitor simple linear profiles is to examine residuals by using exponentially weighted moving average (EWMA) and range (R) charts jointly which is proposed by Kang and Albin [2]. In this method, j^{th} sample statistic for the EWMA chart is the weighted average of the j^{th} residual average and the previous residual averages. R chart is also used to monitor residuals in order to determine any unusual situation where the magnitudes of the residuals are large. In this study, the estimation effect on the performance of EWMA and R control charts combination under non-normality is investigated. For this purpose, average run length (ARL) and run length standard deviation (SDRL) values are obtained by simulation when the error terms are distributed as student's t with different degrees of freedom values. The results indicate that estimation of the parameters deteriorates the performance of the chart under t distribution. The performance of the known parameter case cannot be achieved even when the profile number used in phase I estimation is as high as 200. However, this profile number becomes sufficient as the degrees of freedom of the t distribution increases. Moreover, for some cases SDRL values are obtained to be very high which causes ARL values to be questionable and unreliable. The practitioners should be aware of this decline in the performance of the chart.

Keywords: *Control chart, Non-normality, Profile monitoring, Run length.*

References

[1] Aly, A. A., Mahmoud, M. A. & Woodall W. H. (2015), *A comparison of the performance of Phase II simple linear profile control charts when parameters are estimated*, Communications in Statistics – Simulation and Computation, 44, 1432-1140.

[2] Kang, L., & Albin, S. L. (2000), *On-line monitoring when the process yields a linear profile*, Journal of Quality Technology, 32(4), 418-426.

[3] Kim, K., Mahmoud, M. A., & Woodall, W. H. (2003), *On the monitoring of linear profiles*, Journal of Quality Technology, 35(3), 317-328.

[4] Mahmoud, M. A. (2012), *The performance of phase II simple linear profile approaches when parameters are estimated*, Communications in Statistics – Simulation and Computation, 41(10), 1816-1833.

[5] Noorossana, R., Vaghefi, A., Dorri, M. (2011), *Effect of non-normality on the monitoring of simple linear profiles*, Quality and Reliability Engineering International, 27, 425-436.

ISC2017 INTERNATIONAL STATISTICS CONGRESS
December 6-8, 2017 ANKARA/TURKEY

A Comparison of Different Ridge Parameters under Both Multicollinearity and Heteroscedasticity

Volkan SEVİNÇ [1], Atila GÖKTAŞ[1]
vsevinc@mu.edu.tr, *gatilla@mu.edu.tr*

[1]*Muğla Sıtkı Koçman University, Department of Statistics, Muğla, Turkey*

One of the major problems in fitting an appropriate linear regression model is multicollinearity which occurs when regressors are highly correlated. To overcome this problem, ridge regression estimator, which was first introduced by Hoerl and Kennard as an alternative method to the ordinary least squares (OLS) estimator, has been used. Heteroscedasticity, which violates the assumption of constant variances, is another major problem in regression estimation. To solve this violation problem, weighted least squares estimation is used to fit a more robust linear regression equation. However, when there are both multicollinearity and heteroscedasticity problem, weighted ridge regression estimation should be employed. Ridge regression depends on a value called ridge parameter which does not have an explicit form of calculation. There are plenty of ridge parameters proposed in the literature. To analyze the performances of these ridge parameters for both multicollinear and heteroscedastic data, we conduct a simulation study by generating heteroscedastic data sets of different sample sizes, having different number of regressors and different degrees of multicollinearity. Thereafter, a comparative study has been performed in terms of the mean square error values of the ridge parameters along with the two ones previously proposed by the authors. The study shows when severe amount of heteroscedasticity exists in highly multicollinear data, performances of the ridge parameters differs from the results that have been examined in a different study of Goktas and Sevinc (2016) for non-heteroscedastic data sets only having multicollinearity.

Keywords: *Multicollinearity, ridge parameter, heteroscedasticity, ridge regression, weighted ridge regression*

References

[1] Alkhamisi, M. A. and G. Shukur. 'A Monte Carlo Study Of Recent Ridge Parameters'. Communications in Statistics - Simulation and Computation 36.3 (2007).

[2] Dorugade, A. V. 'New Ridge Parameters For Ridge Regression'. Journal of the Association of Arab Universities for Basic and Applied Sciences 15 (2014).

[3] Hoerl, A. E., Kennard, R. and Baldwin, K. 'Ridge Regression: Some Simulations'. Comm. in Stats. - Simulation & Comp. 4.2 (1975).

[4] Hoerl, A. E. and Kennard, R. 'Ridge Regression: Biased Estimation For Nonorthogonal Problems'. Technometrics 12.1 (1970a).

[5] Hoerl, A.E. and Kennard, R. 'Ridge Regression: Applications To Nonorthogonal Problems'. Technometrics 12.1 (1970b).

[6] Kibria, G. 'Performance Of Some New Ridge Regression Estimators'. Communications in Statistics - Simulation and Computation 32.2 (2003).

10TH ISC2017 INTERNATIONAL STATISTICS CONGRESS
December 6-8, 2017 ANKARA/TURKEY

A Comparison of the Mostly Used Information Criteria for Different Degrees of Autoregressive Time Series Models

Atilla GÖKTAŞ [1], Aytaç PEKMEZCİ[1], Özge AKKUŞ[1]
gatilla@mu.edu.tr, aytac0803@mu.edu.tr, ozge.akkus@mu.edu.tr

[1] *Muğla Sıtkı Koçman University, Department of Statistics, Muğla, Turkey*

The purpose of this study is to compare the most well-known information criteria in stationary econometric time series modeling. As it is known researchers are confused of making the correct preference of such criteria for selecting the appropriate model in time series analysis. For this we generate data from AR(1) to AR(12) time series models allowing no constant, constant and constant with trend terms within the model for different varieties of sample sizes. Each generation type has been replicated for 10 000 times and the information criteria are calculated for each replication. It is found that as the sample size decreases the proportion of correct model selection in every type of information criteria tends to decrease. Since the log likelihood and MSE criteria seem to be failure in most of sample size types in most cases, we think that both are inapproriate to be used as model selector. For sample sizes that are less equal to 125, it is surprisingly found that the "Adjusted R Square" is best for selecting the correct model. For large sample sizes that are greater than 120 "Akaike Information Criterion" performs well. For very large sizes HQ and SIC criterion are best in selecting the appropriate fitted models. In conclusion we suggest SIC to be used for fairly large samples and FPE for small samples. Inclusion of constant or contstant with trend terms do not have any effect on the power of the information criteria.

Keywords: *Information Criteria, Time Series Data Generation, Model Selection*

References

[1] Akaike, H. (1981). *Likelihood of a model and information criteria*. J. Econometrics, 16, 3-14.

[2] Hannan, E. J., and B. G. Quinn (1979): "*The Determination of the Order of an Autoregression*, "Journal of the Royal Statistical Society, *B*, 41, 190-195.

[3] Schwarz, G. (1978*): "Estimating the Dimension of a Model,"* Annals of Statistics, 6, 461-464.

[4] Liew, V.K.S. (2004) *"Which Lag Length Selection Criterion Should We Employ?"* Economics Bulletin 3 (33), 1 – 9.

[5] Liew, Venus Khim−Sen and Terence Tai−leung Chong, (2005) *"Autoregressive Lag Length Selection Criteria in the Presence of ARCH Errors."* Economics Bulletin, Vol. 3, No. 19 pp. 1−5.

10TH ISC2017 INTERNATIONAL STATISTICS CONGRESS
December 6-8, 2017 ANKARA/TURKEY

Comparison of Partial Least Squares With Other Prediction Methods Via Generated Data

Atilla GÖKTAŞ [1],Özge AKKUŞ[1], İsmail BAĞCI[1]
gatilla@mu.edu.tr, ozge.akkus@mu.edu.tr, ibagci@mu.edu.tr

[1]*Muğla Sıtkı Koçman University, Department of Statistics, Muğla, Turkey*

When multicollinearity exists in linear regression model, using t test statistics for testing the coefficients of the independent variables becomes problematic. To overcome the problem there are great number of prediction methods used to fit an appropriate linear regression model. As a matter of fact that the purpose of our study is to compare Partial Least Squares Prediction method (PLS), Ridge Regression (RR) and Principal Components Regression (PCR), which are mostly used to fit regressors having severe multicollinearity against dependent variable. To realize this, a great number of different group of datasets are generated from standard normal distribution allowing the inclusion of different degree of collinearities for 10000 replications. For the design of the study, simulation work has been performed for five different degree of multicollinearity level (0.0, 0.3, 0.5, 0.7, 0.9) and five different sample sizes (30, 50, 100, 200 and 500). The proposed three different prediction regression methods were applied with the generated data. Thereafter the comparison has been made using the value of Mean Squares Error (MSE) of regression parameters. The smallest MSE was treated as determiner of which method was the most efficient and presented the best results under different circumstances. According to the findings obtained, an increase or a decrease in the sample size has definitely a vital effect on the predicting methods. It is found that there is no specific prediction method that can have a meaningful superiority to the others in any sample size or number of regressors. In the meantime each prediction method is affected by the size of the sample, number of independent variables or the degree of multicollinearity. However even in a super multicollinearity level, whatever the number of regressors is, in contrast to literature (say n<=200), it is observed that PCR method surprisingly had better results compared to the other two prediction methods.

Keywords: *Partial Least Squares, Ridge Regression, Principal Components Regression, Multicollinearity*

References

[1] Acharjee, A., Finkers, R., GF Visser, R. and Maliepaard, C. (2013), *Comparison of regularized regression methods for omics data*, Metabolomics, Vol:3 (3), 1-9.

[2] Firinguetti, L., Kibria, G. and Rodrigo, A. (2017), *Study of partial least squares and ridge regression methods*, Communications in Statistics-Simulation and Computation, Vol:0(0), 1-14.

[3] Mahesh, S., Jayas, D. S., Paliwal, J., and White, N. D. G. (2014) *Comparison of Partial Least Squares Regression and Principal Components Regression Methods for Protein and Hardness Predictions using the Near-Infrared Hyperspectral Images of Bulk Samples of Canadian Wheat*, Food and Bioprocess Technology, 8(1), 31–40

[4] Simeon, O., Timothy A.O., Thompson, O.O and Adebowale, O.A. (2014), *Comparison of classical least squares (CLS), Rigde and principal component methods of regression analysis using gynecological data*, IOSR Journal of Mathematics, Vol: 9 (6), 61-74.

[5]Yeniay, Ö. and Göktaş, A. (2002) *A comparison of partial least squares regression with other prediction methods*, Hacettepe Journal of Mathematics and Statistics, Vol: 31, 99-111.

SESSION V

FINANCE INSURANCE AND RISK MANAGEMENT

ISC2017 INTERNATIONAL STATISTICS CONGRESS
December 6-8, 2017 ANKARA/TURKEY

Maximum Loss and Maximum Gain of Spectrally Negative Levy Processes

Ceren Vardar Acar[1], Mine Çağlar[2]
cvardar@metu.edu.tr, mcaglar@ku.edu.tr

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

The maximum loss, or maximum drawdown of a process X is the supremum of X reflected at its running supremum. The motivation comes from mathematical finance as it is useful to quantify the risk associated with the performance of a stock.
The maximum loss at time t>0 is formally defined by

$$M_t^- := \sup_{0\le u\le v\le t} (X_u - X_v),$$

which is equivalent to $\sup_{0\le v\le t}(\sup_{0\le u\le v}(X_u - X_v))$ and $\sup_{0\le v\le t}(S_v - X_v)$, that is, the supremum of the reflected process $S - X$, or the so-called loss process, where S denotes the running supremum.
The loss process has been studied for Brownian motion (Salminen and Vallois 2007; Vardar-Acar et al. 2013), and some Le´vy processes (Mijatovic and Pistorius 2012). A spectrally negative Le´vy process X is a Le´vy process with no positive jumps, that is, its Le´vy measure is concentrated on $(-\infty, 0)$. Spectrally negative Levy process is a commonly used model for financial data.
In this study, the joint distribution of the maximum loss and the maximum gain is obtained for a spectrally negative Lévy process until the passage time of a given level. Their marginal distributions up to an independent exponential time are also provided. The existing formulas for Brownian motion with drift are recovered using the particular scale functions.

Keywords: *Maximum drawdown, spectrally negative, reflected process, fluctuation theory*

References

[1] Mijatovic, A., Pistorius, M.R. (2012): On the drawdown of completely asymmetric Lévy processes. Stoch. Proc. Appl. 122, 3812–3836

[2] Salminen, P., Vallois, P. (2007): On maximum increase and decrease of Brownian motion. Ann. I. H. Poincaré 43, 655–676

[3] Vardar-Acar C., Zirbel C. L., and Szekely G. J (2013), On the correlation of the supremum and the infimum and of maximum gain and maximum loss of Brownian motion with drift, Journal of Computational and Applied Mathematics, 248: 611775

10TH ISC2017 INTERNATIONAL STATISTICS CONGRESS
December 6-8, 2017 ANKARA/TURKEY

Price Level Effect in Istanbul Stock Exchange: Evidence from BIST30

Ayşegül İŞCANOĞLU ÇEKİÇ[1], Demet SEZER[2]
aysegulcekic@trakya.edu.tr, dsezer@selcuk.edu.tr

[1]*Trakya University, Edirne, Turkey*
[2]*Selcuk University, Konya, Turkey*

Volatility is a fundamental component of risk analysis and in general a good estimation of volatility increases the quality of the risk measurements. Therefore the factors which effect the volatility should be considered carefully. Low price effect is one of those factors which is an anomaly implying that the risk adjusted returns of low-priced shares outperform that of high-priced shares. The main reason behind is that low-priced assets show higher volatilities. In this study, we aim to investigate the existence of price effect on the assets trading in Istanbul Stock Exchange. In the analysis, we use 1761 daily observations of 28 stocks trading in BIST30 starting from 01/01/2011 to 01/10/2017. We divide stocks into four groups according to their price levels and we create four equally likely portfolios for each price level. Then we calculate the risk-adjusted returns (Sharpe ratio) where the risk measure is selected as Value at Risk (VaR) with time varying volatility. At this step the best volatility model is selected among various ARCH, GARCH and APARCH models according to AIC. Results show that the low price effect does not exist in Istanbul Stock Exchange. On the contrary, we detect a high price effect. These findings are also tested by using paired sample t-tests. In the study we also implement a risk analysis. For this purpose we estimate one-day VaR with the selected volatility model. Moreover, we try to improve the risk estimations by applying price correction methodology proposed by [2]. Finally, we demonstrate how the correction effects the quality of risk estimations.

Keywords: *Low price effect, Value-at-Risk, ARCH, GARCH, APARCH, Sharpe ratio*

References

[1] Muthoni, H. L., (2014), *Testing the existence of low price effect on stock returns at the Nairobi Securities exchange*,Unpublished Master Project, School of Business, University of Narobi.

[2] Siouris, G-J. and Karagrigoriou, A. (2017), *A Low Price Correction for Improved Volatility Estimation and Forecasting*, Risks, vol. 5, no. 45.

[3] Waelkens, K. and Ward, M. (1997), *The Low Price Effect on the Johannesburg Stock Exchange*, Investment Analysts Journal, 26:45, 35-48.

[4] ZarembaLow, A. and Zmudziński R. (2014), *Price Effect on the Polish Market*, Financial Internet Quarterly "e-Finanse", vol. 10, no.1, 69-85.

10TH ISC2017 INTERNATIONAL STATISTICS CONGRESS
December 6-8, 2017 ANKARA/TURKEY

Analysis of the Cross Correlations Between Turkish Stock Market and Developed Market Indices

Havva GÜLTEKİN[1], Ayşegül İŞCANOĞLU ÇEKİÇ[1]
havvagultekin@trakya.edu.tr, aysegulcekic@trakya.edu.tr

[1]*Trakya University Faculty of Economics & Administrative Sciences, Edirne, Turkey*

Linkage between financial markets has been a substantial problem after globalization. These linkages cause cross correlations among financial markets and affect the accuracy of risk predictions. Therefore, identifying and modelling of those linkages are important issues in the analysis of financial markets. Moreover, the cross correlations among financial markets exhibit a nonlinear behavior and thus, in general the well-known methods fail to predict such correlations. In this paper, we aim to show existence of nonlinear correlations between the financial markets of Turkey and developed countries. For this purpose, we use the Multifractal Detrending Moving-Average Cross-correlation Analysis (MF-XDMA) which is designed for detecting long-range nonlinear correlations. In the analysis we use the daily financial return series of Turkish stock market index BIST100 and developed market indexes which are S&P500, DAX30, FTSE100 for a 10 year period between 01/01/2007-01/01/2017. The results show the existence of nonlinear correlations.

Keywords: *Cross Correlations, MF-XDMA, BIST100, S&P500, DAX30, FTSE100*

References

[1] Cao, G., Han, Y. , Li, Q., Xu, W. (2017) *Asymmetric MF-DCCA method based on risk conduction and its application in the Chinese and foreign stock markets*, Physica A: Statistical Mechanics and its Applications, Volume 468, pp 119-130.

[2] Jiang, Z.-Q. and Zhou, W.-X. (2011) *Multifractal detrending moving-average cross-correlation analysis*, Phys. Rev. E, Volume 84, issue:1.

[3] Sun, X., Lu, X., Yue, G., Li, J. (2017) *Cross-correlations between the US monetary policy, US dollar index and crude oil market*, Physica A: Statistical Mechanics and its Applications, Volume 467, pp 326-344.

[4] Wang, G.-J. and Xie, C. (2013) *Cross-correlations between the CSI 300 spot and futures markets*, Nonlinear Dynamics, Volume 73, Issue 3, pp 1687–1696.

10TH ISC2017 INTERNATIONAL STATISTICS CONGRESS
December 6-8, 2017 ANKARA/TURKEY

Political Risk and Foreign Direct Investment in Tunisia: The Case of the Services Sector

Maroua Ben Ghoul[1], Md. Musa Khan[1]
benghoulmarwa@gmail.com, khanstatcu@gmail.com

[1]*Anadolu University, Faculty of Science Department of Statistics, Eskişehir, Turkey*

Political risk indicators have been considered as important factors which have impact on the Foreign Direct Investment (FDI). But, this relationship between the Political Risk and FDI still not highly covered as expected. In this context, it is crucial to point out the political risk factors' impact on the FDI especially for the Arab Spring countries which had embraced radical political change after the revolution in 2011. The aim of the paper is to investigate the relationship between political risk and the FDI in Tunisia for the case of service sectors. The research is based on aggregate variables that represent six pillars of Governance Indicators; Voice and Accountability, Political Stability and Absence of Violence/Terrorism, Government Effectiveness, Regulatory Quality, Rule of Law and Control of Corruption. The data was extracted from the Worldwide Governance Indicators and the Tunisian Central Bank, the data frequency is yearly from 2004 to 2016. The research confirms that the political factors especially the government effectiveness and voice and accountability have significant impact on the FDI and on the FDI in the services sector.

Keywords: *Political Risk, Tunisia, Foreign Direct Investment, Correlation, Regression model.*

References

[1] Campos, N.F., Nugent, N.B. 2002. "*Who is afraid of political instability?"* Journal of Development Economics 67(1): 157-172.

[2] Khan, M., & Ibne Akbar, M. (2013). *THE IMPACT OF POLITICAL RISK ON FOREIGN DIRECT INVESTMENT*. Munich Personal RePEc Archive.

[3] L. C. Osabutey, E., & Okoro, C. (2015). *Investment in Africa:The Case of the Nigerian Telecommunications Industry*. Wiley Periodicals.

[4] The Worldwide Governance Indicators (WGI). (s.d.). Consulté le October 2017, sur The Worldwide Governance Indicators (WGI): http://info.worldbank.org/governance/wgi/

10TH ISC2017 INTERNATIONAL STATISTICS CONGRESS
December 6-8, 2017 ANKARA/TURKEY

Bivariate Risk Aversion and Risk Premium Based on Various Utility Copula Functions

Kübra DURUKAN[1], Emel KIZILOK KARA[2], H.Hasan ÖRKCÜ[3]
kubraba@gmail.com, *emel.kizilok@gmail.com*, *hhorkcu@gazi.edu.tr*

[1]*Kirikkale University, Faculty of Arts and Sciences, Department of Statistics, Kırıkkale*
[2]*Kirikkale University, Faculty of Arts and Sciences, Department of Actuarial Science, Kırıkkale*
[3]*Gazi University, Faculty of Sciences, Department of Statistics, Ankara*

Copula functions, which have an important role in areas such as insurance, actuarial and risk, are often used to explain the dependency structure of random variables. The risk aversion coefficient is a decision-making parameter and insurance companies can calculate the risk premium associated with this parameter. In this study, it was aimed to calculate the risk aversion coefficient and the risk premium based on utility copula functions for dependent bivariate risk groups. For this, bivariate risk aversion coefficients based on various utility copula models were found. Then, bivariate risk premiums were calculated using these risk aversion coefficients. Numerical results are presented with some tables and graphs for various parameter values.

Keywords: *Dependence, utility function, utility copula, bivariate risk aversion, risk premium*

References

[1] Abbas, A. E. (2009), *Multiattribute utility copulas*, Operations Research, 57(6), 1367-1383.

[2] Denuit, M., Dhaene, J., Goovaerts, M., Kaas, R. (2005), *Actuarial Theory for Dependent Risks*, Measures, Orders and Models. John Wiley and Sons.

[3] Duncan, G. T. (1977), *A matrix measure of multivariate local risk aversion*, Econometrica: Journal of the Econometric Society, 895-903.

[4] Kettler, P. C. (2007), *Utility copulas*, Preprint series, Pure mathematics http://urn.nb. no/URN: NBN: no-8076.

[5] Nelsen, R.B. (2006), *An Introduction to Copulas*, 2nd edition, Springer, New York.

Linear and Nonlinear Market Model Specifications for Stock Markets

Serdar Neslihanoglu[1]
sneslihanoglu@ogu.edu.tr

[1]*Eskisehir Osmangazi University, Eskisehir, Turkey*

The aim of this research is to evaluate the modelling and forecasting performance of the newly defined nonlinear market model including higher moments (which is obtained by [2] and [4]). This model accounts for the systematic component of *co-skewness* and *co-kurtosis* by considering higher moments. Also, the analysis further expands a conditional (time-varying) market model by including time-varying *beta*, *co-skewness* and *co-kurtosis* in the form of the state-space model. Here, the weekly data from the several stock markets all over the world is obtained from the Datastream database provided by University of Glasgow, UK. The empirical findings overwhelmingly support the use of the time-varying market model approaches which perform better than linear model when modelling and forecasting the stock markets. In addition to the fact that higher moments are necessary for the data commonly involving structural changes.

Keywords: *Conditional Market Models, Higher-Moments, Nonlinear Market Model, Stock Markets Time-Varying Risk*

References

[1] Durbin, J. and Koopman, S. (2001). *Time Series Analysis by State Space Methods.*Oxford Statistical Science Series. Clarendon Press.

[2] Hwang, S. and Satchell, S. E. (1999). *Modelling emerging market risk premia using higher moments*. International Journal of Finance & Economics, 4(4), 271_296.

[3] Neslihanoglu, S. (2014). *Validating and Extending the Two-Moment Capital Asset Pricing Model for Financial Time Series*. PhD thesis, The School of Mathematics and Statistics, The University of Glasgow, Glasgow, UK.

[4] Neslihanoglu, S., Vasilios, S., McColl, J.H. and Lee, D. (2017), *Nonlinearities in the CAPM: Evidence from Developed and Emerging Markets*, Journal of Forecasting,36(8), pg. 867-897.

SESSION V

OTHER STATISTICAL METHODS III

10TH ISC2017 INTERNATIONAL STATISTICS CONGRESS
December 6-8, 2017 ANKARA/TURKEY

Small Area Estimation of Poverty Rate at Province Level In Turkey

Gülser Pınar YILMAZ EKŞİ[1], Rukiye DAĞALP[1]
gulpin81@gmail.com, Rukiye.Dagalp@ankara.edu.tr

[1]*Ankara University, Ankara, Turkey*

There are two main approaches for statistical inferences for sample surveys called such as model based and designed based. If determined sample size for survey is sufficient to produce reliable direct estimates , design based approach are taken. Small area or domain refers to determined sample size for survey is too small or insufficient in order to provide reliable estimate for interested area or domain. Interested small area can be geographical region or demographic group .This study is aimed to use model based methods combining information from other different reliable sources at interested area regarding to mixed model. Mixed models are classified into two groups such as area level models and unit level models. In this study, area level model such as Fay-Herriot model are taken into account and Empirical Best Linear Unbiased Prediction (EBLUP) and Hierarchical Bayes (HB) methods are exploited to estimate poverty rate relative to household expenditure at province level in Turkey by using Household Budget Survey micro level data and other related reliable auxilary data sources.

Keywords: *EBLUP, HB, Small Area Estimation, Poverty Rate*

References

[1] Fay R.E., Herriot R.A,1979. ,Estimates of income for small places: an application of James-Stein procedure to census data. Journal of the American Statistical Association, 74, pp. 269-277.

[2] Jiang, J. and Lahiri, P.2006b.,Mixed model prediction and small area estimation. Test,15:111–999.

[3] Henderson, C. R., 1975, Best Linear Unbiased Estimation and Prediction Under a Selection Model, *Biometrics*, 31, 423-447.

10TH ISC2017 INTERNATIONAL STATISTICS CONGRESS
December 6-8, 2017 ANKARA/TURKEY

Investigation of the CO_2 Emission Performances of G20 Countries due to the Energy Consumption with Data Envelopment Analysis

Esra ÖZKAN AKSU[1], Aslı ÇALIŞ BOYACI[2], Cevriye TEMEL GENCER[2]
ozkanesra@gazi.edu.tr, asli.calis@omu.edu.tr, ctemel@gazi.edu.tr

[1]*Gazi University, Ankara, Turkey*
[2]*Ondokuz Mayıs University, Samsun, Turkey*

In the 1980s, with the global climate change reaching appreciable dimensions, energy-economy-environment have started to be evaluated together. Within this context, at the conferences in Rio de Janeiro and Kyoto, some regulations and obligations have been introduced concerning emissions given to the atmosphere and environmental pollution. Also, in a consequence of economic development, CO_2 emission due to the energy consumption are gradually increasing. For this reasons, countries' efficiencies related to CO_2 emissions due to the energy consumption has become more of an issue. In this study, the Data Envelopment Analysis (DEA) method was used to evaluate inter-temporal energy efficiency based on fossil-fuel CO_2 emissions in G20 countries. Data used in the study were obtained from the World Bank website. For analysis, the data between 2007 and 2014 were used. Input variables of the model are land area, population and energy use; undesirable output variable of the model is fossil-fuel CO_2 emission and desirable output variable of the model is gross domestic product (GDP) per capita. These input and output variables are decided according to the information obtained literature and especially from [1] and [2] studies. EMS 1.3.0 package program was used for the calculation of efficiency scores of 20 countries according to these variables. Since CO_2 emission is an undesirable output transformation was applied to this variable. Efficiency scores were calculated separately for each year and it was aimed to observe the change in the energy efficiencies of the countries over the years. The computational results show that Argentina, Australia, Italy, South Korea, Turkey and United Kingdom are efficient for all years considered. In addition, France is efficient on 6 years except for 2007 and 2012; both Indonesia (in 2007, 2008 and 2014) and Saudi Arabia (in 2007, 2008 and 2012) is efficient on 3 years; Japan is efficient only in 2012. The remaining 10 countries (Brazil, China, Germany, India, Mexico, Russia, United States, South Africa, Canada and European Union) have not been efficient on any year, and comments have been made for these countries about what input and output variables they should change in order to be efficient. In the study, correlations were also examined using the SPSS Statistics 17.0 package program to see the relationships between inputs and outputs. As a result of this, it was seen that the correlation between CO_2 emission and population is relatively high to 0.770, and the correlation between GDP and energy use is high to 0,658. This situation indicates that during the research period, both energy use and population are important for countries' efficiencies. On the other hand, since the weights of input and output variables in the DEA vary with each decision-making unit, the weights of these important variables, which are the result of correlation calculations, may not have been considered for the countries that are inefficient. As a result, it may be advisable to include correlations between variables in the efficiency analysis to remove this disadvantage of the DEA.

Keywords: *data envelopment analysis, energy efficiency, CO_2 emission, G20 countries*

References

[1] Guo, X., Lu, C.C., Lee, J.H. and Chiu, Y.H. (2017), *Applying the dynamic DEA model to evaluate the energy efficiency of OECD countries and China*, Energy, 134, 392-399.

[2] Zhang, N. and Choi, Y. (2013), *Environmental energy efficiency of China's regional economies: A non-oriented slacks-based measure analysis*, The Social Science Journal, 50, 225-234.

10TH ISC2017 INTERNATIONAL STATISTICS CONGRESS
December 6-8, 2017 ANKARA/TURKEY

European Union Countries and Turkey's Waste Management Performance Analysis with Malmquist Total Factor Productivity Index

Ahmet KOCATÜRK[1], Seher BODUR[1], Hasan Hüseyin GÜL[1]
ahmetkocaturk@gazi.edu.tr, seherbodur@gazi.edu.tr, hasangul@gazi.edu.tr

[1]*Gazi University, Ankara, Turkey*

The global warming factor and waste is a very important environmental problem. The goal of solid waste management is develop the waste produced of collecting, transporting and final destruction in terms of economically and environmentally by the community after various processes. It is tried to determine the changes of performances of each country and position of Turkey in Europe Union Countries about solid waste management via comparing the scores which are calculated by years with the scores of previous year with Malmquist Total Factor Productivity Index.

Output oriented, constant returns to scale model is used. Waste management indicator data which belongs to the years 2006-2014,were taken from official European statistics site (Eurostat). The records are kept for 2 years. Inputs are waste, intensity and GDP per capita. Outputs are landfilling, deposit onto or into land, incinearation and recovery. Undesired output variable direction changed; its inverse received.

In this study, the performance of solid waste management in Europe Union Countries and Turkey is evaluated by using Malmquist Total Factor Productivity Index. Some suggestions and comments are made on the European Union countries and Turkey's waste management performance.

Keywords: *Data Envelopment Analysis, waste management performance, malmquist total factor productivity index.*

References

[1] Ball, E., Fare, R., Grosskop, S. and Zaim, O. (2005), *Accounting for externalities in the measurement of productivity growth: the Malmquist cost productivity measure*, Structural Change and Economic Dynamics, 16, 374–394.

[2] Banker, R. D. (1984), *Estimating Most Productive Scale Size Using Data Envelopment Analysis*, European Journal of Operational Research, 17, 35-44.

[3] Bjurek, H. (1996), *The Malmquist total factor productivity index*, Scandinavian Journal of Economics, 98 (2), 303–313.

10TH ISC2017 INTERNATIONAL STATISTICS CONGRESS
December 6-8, 2017 ANKARA/TURKEY

Evaluation of Statistical Regions According to Formal Education Statistics with AHP Based VIKOR Method

Aslı ÇALIŞ BOYACI[1], Esra ÖZKAN AKSU[2]
asli.calis@omu.edu.tr, ozkanesra@gazi.edu.tr

[1] *Ondokuz Mayıs University, Samsun, Turkey*
[2] *Gazi University, Ankara, Turkey*

Education raises the standards of life of individuals and societies. For this reason, a country should provide quality and healthy education to its individuals to grow and develop. Turkey is experiencing significant improvements in education compared to ten years ago. The schooling ratio increases at every level, and the number of students per teacher is gradually decreasing. However, this ratio is not evenly distributed among the regions. Education is divided into two, formal and informal education. Formal education is given in school and educational institutions. Formal education; includes pre primary, primary school, lower secondary school, upper secondary and tertiary educational institutions. Informal education does not have a systematic structure but it educates individuals about the environmental interactions during the lives of them, unplanned and unscheduled. In this study, it is aimed to rank the twelve regions in Turkey created by statistical factors such as population, geography and economy according to criteria which are net schooling ratio, the numbers of students per teacher and per classroom by using AHP based VIKOR method. AHP method was first brought forward by two researchers, Myers and Alpert, in 1968 and was developed as a model that can be used for solving the problems of decision-making by Professor Thomas Lorie Saaty in 1977. The VIKOR method was developed for multicriteria optimization of complex systems. It determines the compromise ranking-list, the compromise solution, and the weight stability intervals for preference stability of the compromise solution obtained with the initial weights. This method focuses on ranking and selecting from a set of alternatives in the presence of conflicting criteria. An analysis of the result obtained with these methods is presented in this paper.

Keywords: *Formal Education, AHP, VIKOR*

References

[1] Opricovic, S. and Tzeng, G.H. (2004), *Compromise solution by MCDM methods: A comparative analysis of VIKOR and TOPSIS*, European Journal of Operational Research, 156(2), 445-455.

[2] Opricovic, S. (2011), *Fuzzy VIKOR with an application to water resources planning*, Expert Systems with Applications, 38(10), 12983-12990.

[3] Thomas, S., (2008), *Decision making with the analytic hierarchy process*, International Journal of Services Sciences, 1(1), 85.

10TH ISC2017 INTERNATIONAL STATISTICS CONGRESS
December 6-8, 2017 ANKARA/TURKEY

On Sample Allocation Based on Coefficient of Variation and Nonlinear Cost Constraint in Stratified Random Sampling

Sinem Tuğba ŞAHİN TEKİN[1], Yaprak Arzu ÖZDEMİR[1],
Cenker METİN[2]
sinemsahin@gazi.edu.tr, yaprak@gazi.edu.tr, cenker.metin@gmail.com

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

Composite estimator is the weighted combination of two or more component estimators that are weighted with appropriate weights. This estimator has smaller mean square error than those of each component estimator. In practice, the aim of the sampling methods is to decrease the variance of the statistic that we are interested in under specific constraints. For a given cost constraint, decreasing the variance of a statistic in stratified random sampling is achieved by allocating the sample size to strata. Generally, this cost constraint used in allocation is linear. The allocation procedure makes use of composite estimators called as compromise allocation. In this study, a new compromise allocation method is proposed as an alternative to those of Bankier (1988), Costa et al. (2004) and Longford (2006) compromise allocation methods. Strata sample sizes were determined to minimize the composite estimator as shown in Eq.(1). The equation is obtained by weighting the both coefficient of variation of estimated population mean $CV(\bar{y}_{st})$ and coefficient of variation of strata means ($CV(\bar{y}_h)$).

$$\sum_{h=1}^{L} P_h CV^2(\bar{y}_h) + (GP_+)\, CV^2(\bar{y}_{st}) \qquad (1)$$

Where $P_h = N_h^q \bar{Y}_h^2$, $P_+ = \sum_{h=1}^{L} P_h$, $0 \le q \le 2$. The first component in Eq.(1) specifies relative importance, P_h, of each stratum *h*, while the second component attaches relative importance to $\bar{y}_{st}$ through the weight *G*. In this study, non-linear cost constraint was used to minimize the proposed estimator. The proposed allocation model was also interpreted by using the data from Statistics Canada's Monthly Retail Trade Survey [Choundry et al. (2012)].

Keywords: *Stratified Random Sampling, Composite Estimator, Compromise Allocation, Non-linear cost constraint.*

References

[1] Bankier J. (1989), *Sample allocation in multivariate surveys,* Survey Methodology, 15: 47-57.

[2] Choudhry G. H., Rao J.N.K., Hidiroglou M. A. (2012), *On sample allocation for efficient domain estimation*, Survey Methodology, 38(1):23-29.

[3] Costa A, Satorra A. and Venture E., (2004), *Using composite estimator to improve both domain and total area estimation,* Applied Statistics, 19, 273-278.

[4] Longford N. T., (2006), *Sample size calculation for small-area estimation*, Survey Methodology, 32, 87-96.

SESSION V

STATISTICS THEORY III

10TH ISC2017 INTERNATIONAL STATISTICS CONGRESS
December 6-8, 2017 ANKARA/TURKEY

Linear Bayesian Estimation in Linear Models

Fikri AKDENİZ[1] , İhsan ÜNVER[2], Fikri ÖZTÜRK[3]
akdeniz@gmail.coM, ozturkf@ankara.edu.tr, ihsanunver@gmail.com

[1]Çağ University, Tarsus, Turkey
[2]Avrasya University, Trabzon, Turkey
[3]Ankara University, Ankara, Turkey

Consider the classical linear model $y = Xb + e$, where $E(e) = 0, Cov(e) = s^2\mathbf{I}_n$. Let s^2 be a nuisance parameter, and we have $b: (0, s^2G^{-1})$ as a prior information. Under squared error loss the Bayes Estimator in the set of linear homogeneous estimators $\{\hat{b} : \hat{b} = Ay, A \in R^{p' n}\}$ is defined as

$$\hat{b}_{LB}(G) = \left(\arg\min_A MSE^B(A, s^2, G)\right)y$$

where,

$$\begin{aligned} MSE^B(A, s^2, G) &= E_b\left(E_y(Ay - b)(Ay - b)'\right) \\ &= E_b\left(s^2 Tr(AA') + Tr\left((AX - I)bb'(AX - I)'\right)\right) \\ &= s^2 Tr(AA') + Tr\left((AX - I)E_b(bb')(AX - I)'\right) \\ &= s^2 Tr(AA') + s^2 Tr\left((AX - I)G^{-1}(AX - I)'\right) \end{aligned}$$

[2]. $(X'X + G)^{-1}X' = \arg\min_A MSE^B(A, s^2, G)$ and

$$\hat{b}_{LB}(G) = (X'X + G)^{-1}X'y \qquad [1].$$

So, under the prior information $b: (0, s^2G^{-1})$ the Linear Bayes Estimator (LBE) is equal to General Ridge Estimator. Although being formally the same, these estimators are conceptionally different. A statistician employing the Bayes estimator uses the sample information with an extra prior information. A statistician employing the ridge estimator only uses the sample information and has to estimate the matrix G in order to make the estimator operational. The operational estimator is a nonlinear function of the sample.

The study discuss some statistical properties of the LBE in the context of shrinkage estimation.

Keywords: *Bayesian estimation, Ridge regression.*

References

[1] Gross, J. (2003), *Linear Regression, Berlin,* Springer,181-185.

[2] Rao, C.R. (1976), *Estimation of parameters in a linear model,* The Annals of Statistics, 4, 1023-1037.

10TH ISC2017 INTERNATIONAL STATISTICS CONGRESS
December 6-8, 2017 ANKARA/TURKEY

Alpha Logarithmic Weibull Distribution: Properties and Applications

Yunus AKDOĞAN[1], Fatih ŞAHİN[1], Kadir KARAKAYA[1]
yakdogan@selcuk.edu.tr, fatihsahin0644@gmail.com, kkarakaya@selcuk.edu.tr

[1]*Statistics Department, Science Faculty, Selcuk University, Konya, Turkey.*

In this study, a new distribution is introduced which is called alpha logarihtmic Weibull distribution(ALWD). Several properties of the proposed distribution including the moments, hazard rate function and etc. are obtained. Statistical inference on distribution parameters are also discussed. Simulation study is handled to observe the performance of the estimates. A real data example is provided.

Keywords: *Alpha logarihtmic family, Maximum likelihood estimation, Least square estimation, Weibull distribution,*

References

[1] Karakaya, K., Kinaci, I., Kus, C., Akdogan, Y. (2017). *A new family of distributions* Hacettepe Journal Of Mathematıcs And Statıstıcs. 46(2) 303-314.

[2] Mahdavi, A., Kundu, D. (2017). *A new method for generating distributions with an application to exponential distribution,* Commun. Stat. – Theory Methods. 46(13) 6543-6557.

10TH ISC2017 INTERNATIONAL STATISTICS CONGRESS
December 6-8, 2017 ANKARA/TURKEY

Binomial-Discrete Lindley Distribution

Coşkun KUŞ[1], Yunus AKDOĞAN[1], Akbar ASGHARZADEH[2], İsmail KINACI[1] , Kadir KARAKAYA[1]
coskun@selcuk.edu.tr, yakdogan@selcuk.edu.tr, a.asgharzadeh@umz.ac.ir, ikinaci@selcuk.edu.tr, kkarakaya@selcuk.edu.tr

[1]*Statistics Department, Science Faculty, Selcuk University, Konya, Turkey.*
[2]*Statistics Department, University of Mazandaran, Babolsar, Iran.*

In this study, a new discrete distribution called Binomial-Discrete Lindley (BDL) distribution is proposed by compounding the binomial and discrete Lindley distributions. Some properties of the distribution are obtained including the moment generating function, moments and hazard rate function. Estimation of distribution parameter is studied by methods of moments, proportions and maximum likelihood. A simulation study is performed to compare the performance of the different estimates in terms of bias and mean square error. Automobile claim data applications are also presented to see that new distribution is useful in modelling data.

Keywords: *Binom distribution, Dicrete Lindley distribution, Discrete distributions, Estimation*

References

[1] Hu, Y., Peng, X., Li, T. and Guo, H., On the Poisson approximation to photon distribution for faint lasers. Phys. Lett, (2007), 367, pp. 173-176.

[2] Akdoğan, Y., Kuş, C., Asgharzadeh, A., Kınacı I. and Sharafi, F., Uniform-geometric distribution. Journal of Statistical Computation and Simulation, (2016), 86(9), pp. 1754-1770.

10TH ISC2017 INTERNATIONAL STATISTICS CONGRESS
December 6-8, 2017 ANKARA/TURKEY

Asymptotic Properties of RALS-LM Cointegration Test Presence of Structural Breaks and G/ARCH innovations

Esin FİRUZAN[1], Berhan ÇOBAN[1]
esin.firuzan@deu.edu.tr, *berhan.coban@deu.edu.tr*

[1]*Department of Statistics, Faculty of Science, Dokuz Eylül University, Buca, IZMIR, Turkey*

Structural breaks and heteroscedastic error term in time series analysis such as unit root and cointegration tests have assumed great importance in both the theoretical and the applied time series literature. In the cointegration framework, especially, neglecting structural breaks and non-normal error term induces spurious rejection and the performances of conventional cointegration tests are affected. Former studies detected significant losses of power in the common cointegration tests when potential breaks and G/ARCH effect are ignored. Therefore, it would be meaningful to develop cointegration test establish multiple unknown structural breaks and non-normal cointegration error term.

Residual Augmented Least Squares–Lagrange multiplier (RALS-LM) test include a simple modification procedure to the least squares estimator designed to be robust to the presence of error terms which may exhibit non-normality and structural breaks. This approach utilizes information about the higher moments of the error terms for a construct of the test procedure. In this study, we investigate asymptotic properties of RALS-LM cointegration test that allows for aforementioned features in cointegration equation. Also, we extend and combine the works of Westerlund-Edgerton (2007) and Im et. al (2014).

The study presents the asymptotic behavior of RALS-LM cointegration test under structural break/s and non-normal and/or heteroscedastic innovations.

Keywords: *Cointegration, Residual Augmented Least Squares Estimators, Lagrange-Multiplier, Heteroscedasticity, Structural Breaks*

References

[1] Im, K. S., and P. Schmidt.(2008). More Efficient Estimation under Non-Normality when Higher Moments Do Not Depend on the Regressors, Using Residual-Augmented Least Squares. Journal of Econometrics 144, 219–233.

[2] Im, K. S., Lee, J., & Tieslau, M. (2014). More powerful unit root tests with non-normal errors. In R. C. Sickles & W. C. Horrace (Eds.), Festschrift in honor of Peter Schmidt: Econometric methods and applications (pp. 315–342). New York: Springer.

[3] Meng M., Lee J. and Payne J.E. (2016). RALS-LM unit root test with trend breaks and non- normal errors: application to the Prebisch-Singer hypothesis. Studies in Nonlinear Dynamics & Econometrics. Doi: 10.1515/snde-2016-0050

[4] Pierdzioch C., Risse M., Rohloff S. (2015) Cointegration of the prices of gold and silver: RALS-based evidence, Finance Research Letters, 15, 133-137

[5] Westerlund J. Edgerton D. L.(2007), *New Improved Tests for Cointegration with Structural Break*. Journal of Time Series Analysis. 28, 188-223.

10TH ISC2017 INTERNATIONAL STATISTICS CONGRESS
December 6-8, 2017 ANKARA/TURKEY

Transmuted Complementary Exponential Power Distribution

Buğra SARAÇOĞLU[1], *Caner TANIŞ*[1]
bugrasarac@selcuk.edu.tr, ctanis@selcuk.edu.tr

[1]*Selçuk University Department of Statistics, Konya, Turkey*

In this study, it has been introduced the transmuted complementary exponential power distribution by using quadratic rank transmutation map (QRTM) suggested by Shaw and Buckley [3], [4]. The some statistical properties of this distribution is provided. The unknown parameters of this model are estimated by the maximum likelihood (ML) method. The performances of ML estimator has been examined for unknown parameters of this new distribution via a monte-carlo simulation study according to bias and MSE.

Keywords: *Transmuted complementary exponential power distribution, maximum likelihood, monte-carlo simulation*

References

[1] Barriga, G. D., Louzada-Neto, F., & Cancho, V. G. (2011). *The complementary exponential power lifetime model*. Computational Statistics & Data Analysis, 55(3), 1250-1259.

[2] Saraçoğlu, B., 2017. *Transmuted Exponential Power Distribution and its Distributional Properties*, 6th International Eurasian Conference on Mathematical Sciences and Applications (IECMSA-2017), pg: 270.

[3] Shaw, W. T., & Buckley, I. R. (2007). *The alchemy of probability distributions: Beyond gram-charlier & cornish-fisher expansions, and skew-normal or kurtotic-normal distributions*. Submitted, Feb, 7, 64.

[4] Shaw, W. T., & Buckley, I. R. (2009). *The alchemy of probability distributions: beyond Gram-Charlier expansions, and a skew-kurtotic-normal distribution from a rank transmutation map.* arXiv preprint arXiv:0901.0434.

[5] Smith, R. M., & Bain, L. J. (1975). *An exponential power life-testing distribution.* Communications in Statistics-Theory and Methods, 4(5), 469-481.

SESSION V

MODELING AND SIMULATION II

ISC2017 INTERNATIONAL STATISTICS CONGRESS
December 6-8, 2017 ANKARA/TURKEY

The Determination of Optimal Production of Corn Bread Using Response Surface Method and Data Envelopment Analysis

Başak APAYDIN AVŞAR[1], Hülya BAYRAK[2], Meral EBEGİL[2], Duygu KILIÇ[2]
basak.apaydinavsar@sanayi.gov.tr, hbayrak@gazi.edu.tr, mdemirel@gazi.edu.tr, duygukilic4@gmail.com

[1]*The Ministry of Science, Industry and Technology, Ankara, Turkey*
[2] *Gazi University Department of Statistics 06500, Teknikokullar, Ankara, Turkey*

Optimization technology accelerates decision making processes and improves the quality of decision making in the solution of real-time problems [1]. In this study, the response surface methodology which optimizes the process with multiple responses was used combined with Data Envelopment Analysis (DEA). Response surface methodology is an empirical statistical approach for modelling problems in which several variables influence a response of interest [2]. Myers and Montgomery have described the response surface methodology as a method by which the statistical and mathematical techniques necessary for the development and optimization of processes that are used together [3]. On the other hand, a mathematical programming based approach, DEA, is a popular optimization technique used to determine the relative effectiveness of decision units responsible for transforming a set of inputs into a set of outputs. The response surface methodology allows to obtain a process through the regression equation without having to know the relation model between input and output. There are as many response equations as the number of responses, and so much surface and contour can be drawn. For this reason, the solution of the problem can become complex by increasing the number of responses. DEA method has the ability to hold the multiplicity of not only inputs but outputs and it is also an easy optimization technique to find the best alternatives. In the conventional response surface methodology, the combination of DEA and response surface method is quite advantageous in that it saves time by removing the difficulty of calculating each response individually. In this study, 81 loaves of corn bread were used and each of them was considered an experiment. The dataset consists of 4 inputs and 2 outputs. The inputs used for the analysis were wheat flour addition rate (%), yeast amount, oven temperature (^{0}C) and fermentation time (min). The amount of phytic acid (mg/100g) and loaf volume variables were used as the outputs. The desired parameter optimization is to have a uniformity that reduces the amount of phytic acid and increases the volume of the bread. The experimental responses were determined according to the measures mentioned in inputs and outputs. A central composite design was used to create the design of the experiment.

Keywords: *Optimization, Multiple Responses, Data Envelopment Analysis, Response Surface Method.*

References

[1] Winston, W. L. (2003), *Operations Research: Applications and Algorithms*, 4. Edition, International Thomson Publishing, Belmont, USA.

[2] Tsai, C. W., Tong, L. I. and Wang, C. H. (2010), *Optimization of Multiple Responses Using Data Envelopment Analysis and Response Surface Methodology*. Tamkang Journal of Science and Engineering, 13 (2), 197-203.

[3] Kılıç, D., Özkaya, B. and Bayrak, H. (2017), Response Surface Method in Food Agronomy and Application of Factorial Design, *XVIII. International Symposium on Econometrics Operations Research and Statistics*, Trabzon, Turkey.

ISC2017 INTERNATIONAL STATISTICS CONGRESS
December 6-8, 2017 ANKARA/TURKEY

A Classification and Regression Model for Air Passenger Flow Among Countries

Tuğba ORHAN[1], Betül KAN KILINÇ[2]
torhan4@gmail.com, bkan@anadolu.edu.tr

[1]*Turkish Airlines, Specialist, İstanbul, Turkey*
[2]*Department of Statistics Science Faculty Anadolu University, Eskişehir, Turkey*

Classification and regression tree (CART) is one of the widely used statistical techniques in dealing with classification and prediction problems. Classification tree is constructed when the dependent variable is categorical; on the other hand regression tree is developed. As CART does not assume any underlying relationship between the dependent variable and the predictors, the determinants of the demand of air transportation can be easily analysed and interpreted. In this paper, we build a regression tree model to examine air passenger flows among countries. This model considers the role of multiple factors as the independent variables such as income, distance, ... etc that can significantly influence the air passenger flows. The estimation results demonstrate that the regression tree model can serve as an alternative for analysing cross-country passenger flows.

Keywords: *air passenger flows, demand, regression and classification tree, airlines*

References

[1] Breiman, L., Friedman, J., Olshen, R., and Stone, C. (1984)., *Classification and Regression Trees*, Monterey, Calif., U.S.A., Wadsworth, Inc.

[2]R Development Core Team., *R: A Language And Environment For Statistical Computing*. Vienna (Austria): R Foundation for Statistical Computing. URL: http://www. R-project.org, 2013.

[3]Hastie, T., Tibshirani, R., Friedman, J. (2008), *The Elements of Statistical Learning*, Springer, Standford, California, Second Edition, 119, 308, 587

[4] Chang, Li-Yen. and Lin, Da-Jie. (2010), *Analysis of International Air Passenger Flows between Two Countries in the APEC Region Using Non-parametric Regression Tree Models*, Hong Kong, Vol I, 1-6.

10TH ISC2017 INTERNATIONAL STATISTICS CONGRESS
December 6-8, 2017 ANKARA/TURKEY

On Facility Location Interval Games

Mustafa EKİCİ[1], Osman PALANCI[2], Sırma Zeynep ALPARSLAN GÖK[3]
mustafa.ekici@usak.edu.tr, osmanpalanci@sdu.edu.tr, sirmagok@sdu.edu.tr

[1] *Usak University Faculty of Education Mathematics and Science Education, Usak, Turkey*
[2] *Suleyman Demirel University Faculty of Economics and Administrative Sciences, Isparta, Turkey*
[3]*Suleyman Demirel University Faculty of Arts and Sciences, Isparta, Turkey*

Facility location situations are a promising topic in the field of Operations Research (OR), which has many applications to real life. In a facility location situation, each facility is constructed to please the players [2]. Here, the problem is to minimize the total cost. This cost is composed of both the player distance and the construction of each facility. In the sequel, a facility location game is constructed from a facility location situation. In this study, we consider some classical results from facility location games and their Shapley value and Equal Surplus Sharing rules [3]. It is seen that these rules do not have population monotonic allocation schemes (PMAS). Further, we introduce facility location interval games and their properties [1].

Keywords: *facility location situations, cooperative games, cooperative interval games, Shapley value, Equal Surplus Sharing rules, uncertainty, PMAS.*

References

[1] Alparslan Gok, S.Z., Miquel, S. and Tijs, S. (2009), *Cooperation under interval uncertainty*, Mathematical Methods of Operations Research, 69, 99-109.

[2] Nisan, N., Roughgarden, T., Tardos, E. and Vazirani, V.V. (2007), *Algorithmic Game Theory*, Cambridge University Press, Cambridge.

[3] van den Brink, R. and Funaki, Y. (2009). *Axiomatizations of a class of equal surplus sharing solutions for TU-games,* Theory and Decision, 67, 303-340.

10TH ISC2017 INTERNATIONAL STATISTICS CONGRESS
December 6-8, 2017 ANKARA/TURKEY

Measurement System Capability for Quality Improvement by Gage R&R with an application

Ali Rıza FİRUZAN[1], Ümit KUVVETLİ[2]
Aliriza.firuzan@deu.edu.tr, umit.kuvvetli@gmail.com

[1]*Dokuz Eylul University, Izmir, Turkey*
[2]*ESHOT General Directorate, Izmir, Turkey*

Many manufacturers are using tools like statistical process control (SPC) and design of experiments (DoE) to monitor and improve product quality and process productivity. However, if the data collected are not accurate and precise, they do not represent the true characteristics of the part or product being measured, even if organizations are using the quality improvement tools correctly.

Therefore, it is very important to have a valid quality measurement study beforehand to ensure the part or product data collected are accurate and precise and the power of SPC and DoE are fully realized. Accuracy—in other words, no bias—is the function of calibration and is performed before a correct measurement study of the precisions of the gage and its operators.

In order to reduce the variations in a process, it is necessary to identify the sources of variation, quantify them and to have an understanding about the proper operation of the gage that is being used for collecting the measurements. In operating a gage, measurement error can be contributed to various sources like within-sample variation, measurement method, the gage/instrument used for measurement, operators, temperature, environment and other factors. Therefore, it is necessary to conduct a study on measurement system capability. This study is termed as Gage Repeatability and Reproducibility (GRR) study or gage capability analysis.

In this study, it was decided to examine measurement system although process is under control in a manufacturing company as a result of various problems about quality. Then measurement system was analysed and results obtained were shared.

Keywords: *quality improvement, gage R&R, process capability, measurement system analysis*

References

[1] Al-Refaie A. & Bata N. (2010). *Evaluating measurement and process capabilities by GR&R with four quality measures*, Measurement, 43 (6), 842-851.

[2] Box, G.E.P., Hunter, W.G., Hunter, J.S. (1978), *Statistics for Experimenters*. New York: Wiley.

[3] Van den Heuvel, E.R., Trip, A. (2003), *Evaluation of measurement systems with a small number of observers*. Quality Engineering, 15, 323 – 331.

[4] Karl D.M. & Richard W.A. (2002), *Evaluating measurements systems and manufacturing process using three quality measures*, Quality Engineering, 15(2), 243-251.

10TH ISC2017 INTERNATIONAL STATISTICS CONGRESS
December 6-8, 2017 ANKARA/TURKEY

Measuring Service Quality in Rubber-Wheeled Urban Public Transportation by Using Smart Card Boarding Data: A Case Study for Izmir

Ümit KUVVETLİ[1], Ali Rıza FİRUZAN[2]
umit.kuvvetli@gmail.com, aliriza.firuzan@deu.edu.tr

[1]*ESHOT General Directorate, Izmir, Turkey*
[2]*Dokuz Eylul University, Izmir, Turkey*

The quality of public transportation services is one of the most important performance indicators of modern urban policies for both planning and implementation aspects. Service performance of public transportation has direct impact on the future policies of local governments. Therefore, all the big cities, especially the metropolitan areas, have to directly deal with transportation issues and related public feedback. On the other hand, as in most service industries, it is very difficult to measure and assess the quality of service in public transportation, due to the intangible aspects of the service and the subjective methods used in quality measurement. Moreover, in the public transport sector where the potential problems associated with service quality should be determined and solved quickly, the current methods are insufficient to meet this need of public transport sector. In this project, it is aimed to fill this gap and a statistical model that measure service quality by using smart card boarding data and allows to measure service quality in detail such as route, time interval, passenger type and so on has been accordingly developed.

The main purpose of this project is to develop a model measuring quality of service for rubber-wheeled urban public transport firms have smart card systems. The model uses smart card data which is an objective data source as opposed to the subjective methods commonly used nowadays to measure service quality. The model measures service quality based on quality dimensions such as comfort, information, passenger density in the bus, type of bus stop etc. The weights of the dimensions in the model have been determined by statistical analysis of the data from passenger surveys. The results obtained from this model allow various detailed analyses for passenger types, routes and regions both on a general perspective with weighted criteria and on specific service dimensions requested. It is thought that the model results will guide the political decisions to provide the development of urban public transport systems, ensure standard service quality level and help to provide rapid intervention in problematic areas. Additionally, the project will contribute to the sector by measuring and monitoring service passenger satisfaction and comparing service quality offered by different cities.

Within the scope of the project, five routes with different passenger densities in Izmir/Turkey was selected as an example and the service quality for each passenger for a week (total 349.359 boarding) was measured and the results obtained were analyzed.

Keywords: *urban public transportation, service quality, smart card boarding data, servqual,*

References

[1] Cuthbert, P.F. (1996). *Managing service quality in HE: Is SERVQUAL the answer? Part 2*, Managing Service Quality, 6 (3), 31-35.

[2] Parasuraman, A., Zeithaml, V.A., & Berry L.L. (1985), *A Conceptual Model of Service Quality and its implications for Future Research,* Journal of Marketing , 49, 41-50.

SESSION V

STATISTICS THEORY IV

10TH ISC2017 INTERNATIONAL STATISTICS CONGRESS
December 6-8, 2017 ANKARA/TURKEY

Cubic Rank Transmuted Exponentiated Exponential Distribution

Caner TANIŞ[1], *Buğra SARAÇOĞLU*[1]
ctanis@selcuk.edu.tr, bugrasarac@selcuk.edu.tr

[1]*Selçuk University Department of Statistics, Konya, Turkey*

In this study, it is suggested a new distribution called "Cubic rank transmuted exponentiated exponential (CRTEE) distribution" using cubic rank transmutation map introduced by Granzotto et. al. [1]. The some statistical properties of this new distribution such as, hazard function and its graphics, moments, variance, moment generating function, order statistics are examined. The unknown parameters of this model are estimated by maximum likelihood method. Further, a simulation study performed in order to examine the performances of MLE according to MSE and bias.

Keywords: *cubic rank transmuted exponentiated exponential distribution, cubic rank transmutation map, maximum likelihood estimation, monte-carlo simulation*

References

[1] D. C. T. Granzotto, F. Louzada & N. Balakrishnan (2017) *Cubic rank transmuted distributions: inferential issues and applications*, Journal of Statistical Computation and Simulation, 87:14, 2760-2778, DOI: 10.1080/00949655.2017.1344239.

[2] Gupta, R. D., & Kundu, D. (2001). *Exponentiated exponential family: an alternative to gamma and Weibull distributions.* Biometrical journal, 43(1), 117-130.

[3] Merovci, F. (2013). T*ransmuted exponentiated exponential distribution*. Mathematical Sciences and Applications E-Notes, 1(2).

[4] Shaw, W. T., & Buckley, I. R. (2007). *The alchemy of probability distributions: Beyond gram-charlier & cornish-fisher expansions, and skew-normal or kurtotic-normal distributions.* Submitted, Feb, 7, 64.

[5] Shaw, W. T., & Buckley, I. R. (2009). *The alchemy of probability distributions: beyond Gram-Charlier expansions, and a skew-kurtotic-normal distribution from a rank transmutation map. arXiv preprint arXiv:0901.0434*.

Detecting Change Point via Precedence Type Test

Muslu Kazım KÖREZ[1], İsmail KINACI[1], Hon Keung Tony NG[2], Coşkun KUŞ[1]
kkorez@selcuk.edu.tr, ikinaci@selcuk.edu.tr, ngh@mail.smu.edu, coskun@selcuk.edu.tr

[1]*Department of Statistics, Selcuk University, Konya, Turkey*
[2]*Department of Statistical Science, Southern Methodist University, Dallas, Texas, USA*

The change point analysis interests whether there is a change in distribution of any process. In this study, the single change point problem is handled and the new algorithm is introduced based on precedence type test to detect the change point in single change point problem. It is also given some critical values and powers of the proposed test.

Keywords: *Change point, Nonparametric test, Precedence Test, Hyphothesis test*

References

[1] Balakrishnan, N. and Ng, H. K. T. (2006), *Precedence-Type Tests and Applications*, Hoboken, New Jersey, USA, A John Wiley & Sons, Inc., Publication, 2006, 31-34.

10TH ISC2017 INTERNATIONAL STATISTICS CONGRESS
December 6-8, 2017 ANKARA/TURKEY

Score Test for the Equality of Means for Several Log-Normal Distributions

Mehmet ÇAKMAK[1], Fikri GÖKPINAR[2], Esra GÖKPINAR[2]
mehmet.cakmak@tubitak.gov.tr, fikri@gazi.edu.tr, eyigit@gazi.edu.tr

[1]*The Scientific and Technological Research Council of Turkey, Ankara, Turkey*
[2] *Gazi University, Department of Statistics, Ankara, Turkey*

The lognormal distribution is one of the most extensively used distributions for modeling positive and highly skewed data. Therefore, it has wide areas of application such as geology and mining, medicine, environment, atmospheric sciences and aerobiology, social sciences and economics and etc. [1].

Let Y_{ij}, $j=1,\ldots,n_i$, $i=1,\ldots,k$ be random samples from the lognormal distributions which shape parameter is μ_i and scale parameter is ${\sigma_i}^2$, respectively, i.e., $Y_{ij} \sim \mathrm{LN}(\mu_i, {\sigma_i}^2)$. Then the mean of the i th population, M_i, is obtained as $M_i = \exp(\mu_i + {\sigma_i}^2/2)$. Our aim is to test H_0 hypothesis against H_1 hypothesis which are given below,

$$H_0 : M_1 = M_2 = \ldots = M_k, \qquad H_1 : M_i \neq M_{i'},\ \exists i \neq i'\ \ (i, i' = 1,\ldots,k).$$

In this paper, we propose a new test statistic for testing the equality of several lognormal means based on Score statistic. This test has an approximate chi-square distribution with *k*-1 degrees of freedom under the null hypothesis. In addition to traditional chi-square approximation, we also use a parametric bootstrap based method called computational approach test (CAT) to calculate the p-value of the test. This method does not require any sampling distribution and easy and fast to implement [2,3,4].

Keywords: *lognormal distribution, parametric bootstrap, score statistic, scale parameter.*

References

[1] Limpert, E., Stahel, W.A. and Abbt, M. (2001), *Log-normal Distributions across the Sciences: Keys and Clues*, BioScience, 51, 341-352.

[2] Pal, N., Lim, W. K. and Ling, C.H. (2007), *A computational approach to statistical inferences*, Journal of Applied Probability & Statistics, 2:13-35.

[3] Gökpınar, F. and Gökpınar, E. (2017), *Testing the equality of several log-normal means based on a computational approach*, Communications in Statistics-Simulation and Computation, 46(3): 1998-2010.

[4] Gökpınar, E. and Gökpınar, F. (2012), *A test based on computational approach for equality of means under unequal variance assumption*, Hacettepe Journal of Mathematics and Statistics, 41(4):605-613.

ISC2017 INTERNATIONAL STATISTICS CONGRESS
December 6-8, 2017 ANKARA/TURKEY

A New Class of Exponential Regression cum Ratio Estimator in Systematic Sampling and Application on Real Air Quality Data Set

Eda Gizem KOÇYİĞİT[1], Hülya ÇINGI[1]
eda.kocyigit@hacettepe.edu.tr, hcingi@hacettepe.edu.tr

[1]*Hacettepe University, Department of Statistics, Beytepe 06800, Ankara, Turkey*

Working with the sample saves researchers time, energy and money. In most cases, working on a well-defined small sample can yield better results than with a large batch. As a statistical sampling method, systematic sampling is simpler and more straightforward than random sampling.

In sample surveys, auxiliary information is commonly used in order to improve efficiency and precision of estimators while calculating sum, mean and variance of population estimations. Auxiliary information is used in ratio, product, regression and spread estimators due to its simplicity and precision. These estimators are preferable regarding correlation between auxiliary variable and study variable, and in some conditions, give results that have smaller variance, which means more precise, compared to estimators based on simple means.

In this paper, we propose a new class of exponential regression cum ratio estimator using the auxiliary variable for the estimation of the finite population mean under systematic sampling scheme. The Bias and Mean Square Error (MSE) equations of the proposed estimator are obtained and supported by a numerical example using original air quality data sets. We find the proposed estimator is more efficient than Swain's classical ratio estimators [5], Singh, H. P., Tailor, R., Jatwa, N. K modified ratio estimator [3], H. P. Tailor and R. S. Solanki efficient class of estimator [2], R. Singh and etc. improved estimator [4] and E. G. Kocyigit and Cingi's class of unbiased linear estimator [1] in systematic sampling.

Keywords: ***S****ampling theory, systematic sampling, estimators, MSE, air quality.*

References

[1] Kocyigit, E. G., Cingi, H. (2017), *A new class of unbiased linear estimators in systematic sampling*, Hacettepe Journal of Mathematics and Statistics, 46(2), 315-323.

[2] Singh, H. P., Solanki, R. S. (2012), *An efficient class of estimators for the population mean using auxiliary information in systematic sampling*, Journal of Statistical Theory and Practice, 6(2), 274-285.

[3] Singh, H. P., Tailor, R., Jatwa, N. K. (2011), *Modified ratio and product estimators for population mean in systematic sampling*, Journal of Modern Applied Statistical Methods, 10(2), 4.

[4] Singh, R., Malik, S., Singh, V. K. (2012), *An improved estimator in systematic sampling*, Journal of Scientific Research Banaras Hindu University, Varanasi, Vol. 56, 2012 : 177-182.

[5] Swain, A. K. P. C., (1964), *The use of systematic sampling ratio estimate*, J. Ind. Statist. Assoc., 2, 160–164.

10TH ISC2017 INTERNATIONAL STATISTICS CONGRESS
December 6-8, 2017 ANKARA/TURKEY

Alpha Power Chen Distribution and its Properties

Fatih ŞAHİN[1], Kadir KARAKAYA[1] and Yunus AKDOĞAN[1]
fatihsahin0644@gmail.com, kkarakaya@selcuk.edu.tr, yakdogan@selcuk.edu.tr.

[1]*Statistics Department, Science Faculty, Selcuk University, Konya, Turkey.*

Mahdavi and Kundu (2017) has been introduced a new family of distributions called APT-family. They considered a special case of this family with exponential distribution in details. In this paper, Chen distribution is considered as a baseline distribution for APT-family. Several properties of the APT-Chen distribution such as the moments, quantiles, moment generating function, order statistics etc. are derived. The maximum likelihood, moments and least square methods are discussed. Simulation study is also conducted to compare the estimation methods. An numerical example is provided to illustrate the capability of APT-Chen distribution for modelling real data.

Keywords: *Alpha power transformation, Chen distribution, Maximum likelihood estimation, Least square estimation.*

References

[1] Mahdavi, A., Kundu, D. (2017). *A new method for generating distributions with an application to exponential distribution*, Commun. Stat. – Theory Methods. 46(13) 6543-6557.

[2] Nassar, M., Alzaatreh, A., Mead, M., and Abo-Kasem, O. (2017). *Alpha power Weibull distribution: Properties and applications*, Commun. Stat. – Theory Methods. 46(20) 10236-10252.

SESSION VI

STATISTICS THEORY V

10TH ISC2017 INTERNATIONAL STATISTICS CONGRESS
December 6-8, 2017 ANKARA/TURKEY

Robust Mixture Multivariate Regression Model Based on Multivariate Skew Laplace Distribution

Y. Murat BULUT[1], Fatma Zehra DOĞRU[2], Olcay ARSLAN[3]
ymbulut@ogu.edu.tr , fatma.dogru@giresun.edu.tr, oarslan@ankara.edu.tr

[1]*Eskişehir Osmangazi University, Eskişehir, Turkey*
[2]*Giresun University, Giresun, Turkey*
[3]*Ankara University, Ankara, Turkey*

Mixture regression models were proposed by [4] and [5] as switching regression models. These models have been used many fields such as engineering, genetics, biology, econometrics and marketing to capture the relationship between variables coming from several unknown latent groups.

In literature, it is generally assumed that the error terms have the normal distribution. But the normality assumption is sensitive to the outliers and heavy tailed errors. Recently, [3] proposed robust estimation procedure for the mixture multivariate linear regression using multivariate Laplace distribution to cope with heavy tailedness. In the mixture model context, [2] proposed finite mixtures of multivariate skew Laplace distributions for modelling skewness and heavy tailedness in the heterogeneous data sets. In this study, we propose the mixture multivariate regression based on the multivariate skew Laplace distribution [1] to model both heavy tailedness and skewness simultaneously. Also, this mixture regression model will be an extension of the finite mixtures of multivariate skew Laplace distributions. We obtain the maximum likelihood (ML) estimators of the proposed mixture multivariate regression model using the expectation-maximization (EM) algorithm.

Keywords: *EM algorithm, mixture multivariate regression model, ML, multivariate skew Laplace distribution.*

References

[1] Arslan, O. (2010). *An alternative multivariate skew Laplace distribution: properties and estimation*. Statistical Papers, 51(4), 865-887.

[2] Doğru, F. Z., Bulut, Y. M., Arslan, O. (2017). *Finite Mixtures of Multivariate Skew Laplace Distribution.* arXiv:1702.00628.

[3] Li, X., Bai, X., Song, W. (2017). *Robust mixture multivariate linear regression by multivariate Laplace distribution.* Statistics and Probability Letters, 130, 32-39.

[4] Quandt, R. E. (1972). *A new approach to estimating switching regressions*. Journal of the American Statistical Association 67(338):306–310.

[5] Quandt, R. E., Ramsey, J. B. (1978). *Estimating mixtures of normal distributions and switching regressions.* Journal of the American Statistical Association 73(364):730–752.

10TH ISC2017 INTERNATIONAL STATISTICS CONGRESS
December 6-8, 2017 ANKARA/TURKEY

Robustness Properties for Maximum Likelihood Estimators of Parameters in Exponential Power and Generalized t Distributions

Mehmet Niyazi ÇANKAYA[1], Olcay ARSLAN[2]
mehmet.cankaya@usak.edu.tr, oarslan@ankara.edu.tr

[1]*Applied Sciences School, Department of International Trading, Uşak, Turkey*
[2]*Faculty of Sciences, Department of Statistics, Ankara, Turkey*

The normality assumption on data set is very restrictive approach for modelling. The generalized form of normal distribution, named as an exponential power (EP) distribution, and its scale mixture form have been considered extensively to overcome the problem for modelling non-normal data set since last decades. However, examining the robustness properties of maximum likelihood (ML) estimators of parameters in these distributions, such as the influence function and breakdown point has not been considered together. The well-known asymptotic properties of ML estimators of location, scale and added skewness parameters in EP and its scale mixture form distributions are studied and also these ML estimators for location, scale and scale variant (skewness) parameters can be represented as an iterative reweighting algorithm (IRA) to compute the estimates of these parameters simultaneously. The artificial data are generated to examine the performance of IRA for ML estimations of the parameters simultaneously. Real data examples are provided to illustrate the modelling capability of EP and its scale mixture form distributions.

Keywords: *Exponential power distributions; robustness; asymptotic; modelling.*

References

[1] Arslan, O., Genç, A.İ. (2009), *The skew generalized t distribution as the scale mixture of a skew exponential power distribution and its applications in robust estimation*, Statistics, 43(5), 481-498.

[2] Hampel, F.R., Ronchetti, E.M., Rousseeuw, P.J. and Stahel, W.A. (1986), *Robust Statistics: The Approach Based on Influence Functions*. Wiley Series in Probability and Statistics, 465.

ISC2017 INTERNATIONAL STATISTICS CONGRESS
December 6-8, 2017 ANKARA/TURKEY

Robust Inference with a Skew t Distribution

M. Qamarul ISLAM[1]
mislam@metu.edu.tr

[1]*Department of Statistics, Middle East Technical University, Ankara, Turkey*

There is a growing body of evidence that non-normal data is more prevalent in nature than the normal one. A number of examples can be quoted from areas of Economics, Finance, and Actuarial Sciences [1]. In this study a skew t distribution that can be used to model a data that exhibit inherent non-normal behavior is considered [3]. This distribution has tails fatter than a normal distribution and it also exhibits skewness. Although maximum likelihood estimators (MLE) can be obtained by solving iteratively the likelihood equations that are non-linear in form, this can be problematic in terms of convergence and in many other respects as well [4]. Therefore, we prefer to use the method of modified maximum likelihood (MML) in which the likelihood estimators are derived by expressing the intractable non-linear likelihood equations in terms of standardized ordered variates and replacing the intractable terms by their linear approximations obtained from the first two terms of a Taylor series expansion about the quantiles of the distribution [5]. These estimators, called modified maximum likelihood estimators (MMLE), are obtained in closed form and they are equivalent to the MLE, asymptotically. Even in small samples they are found to be approximately the same as MLE that are obtained iteratively. The MMLE are not only unbiased but substantially more efficient than the commonly used moment estimators (ME) that are obtained by applying the method of moments (MM). In conventional regression analysis, it is assumed that the error terms are distributed normally and, hence, the well-known least square (LS) method is considered to be most the suitable and preferred method for making the relevant statistical inferences. However, a number of empirical researches, particularly in the area of finance, have shown that non-normal errors are present as a rule and not an exception [2]. Even transforming and/or filtering techniques may not produce normally distributed residuals. So, we consider multiple linear regression models with random error having non-normal pattern; specifically, distributed as skew t distribution. Through an extensive simulation work it is shown that the MMLE of regression parameters are plausibly robust to the distributional assumptions and to various data anomalies as compared to the widely used least square estimators (LSE). Relevant tests of hypothesis are developed and explored for desirable properties in terms of their size and power. We also provide several applications where the use of such distribution is justified in terms of meaningful statistical hypotheses.

KeyWords: *Skew t distribution, Least square estimators, Maximum likelihood estimators, Modified maximum likelihood estimators, Linear regression*

References

[1] Adcock, C., Eling, M. and Loperfido, N. (2015), *Skewed distributions in finance and actuarial sciences: a review*, The European Journal of Finance, Volume 21(13), 1253-1281.

[2] Fama, E.E. (1965), *The behavior of stock market prices*, The Journal of Business, Volume 38(1), Pages 34-105.

[3] Fernandez, C. and Steel, M.F.J. (1998), *On Bayesian modeling of fat tails and skewness*, Journal of The American Statistical Association, Volume 93, Pages 359-371.

[4] Sazak, H.S., Tiku, M.L. and Islam, M.Q. (2006), *Regression analysis with a stochastic design variable*, International Statistical Review, Volume 74(1), Pages 77-88.

[5] Tiku, M.L. (1992), *A New method of estimation for location and scale parameters*, Journal of Statistical Planning and Inference, Volume 30(2), Pages 281-292.

ISC2017 INTERNATIONAL STATISTICS CONGRESS
December 6-8, 2017 ANKARA/TURKEY

Some Properties of Epsilon Skew Burr III Distribution

Mehmet Niyazi ÇANKAYA[1], Abdullah YALÇINKAYA[2], Ömer ALTINDAĞ, Olcay ARSLAN[2]
mehmet.cankaya@usak.edu.tr, ayalcinkaya@ankara.edu.tr, omeraltindag87@gmail.com, oarslan@ankara.edu.tr

[1]*Applied Sciences School, Department of International Trading, Uşak, Turkey*
[2]*Faculty of Sciences, Department of Statistics, Ankara, Turkey*

The Burr III distribution is used in a wide variety of fields of lifetime data analysis, reliability theory, and financial literature, etc. It is defined on the positive axis and has two shape parameters, say c and k. These shape parameters allow the distribution to be more flexible, compared to the distributions having only one shape parameter. They also determine the shape of tail of the distribution. Çankaya et al. [2] has extended the Burr III distribution to the real line via epsilon skew extension method which adds a skewness parameter, say ε, to the distribution. The extended version is called as epsilon-skew Burr III (ESBIII) distribution. When the parameters c and k have a relation such that $ck \approx 1$ or $ck < 1$, it is skewed unimodal. Otherwise, it is skewed bimodal with the same level of peaks on the negative and positive sides of real line. Thus, ESBIII distribution can capture fitting the various data sets even when the number of parameters are three. Location and scale form of this distribution can also be constructed. In this study, some distributional properties of the ESBIII distribution are given. The maximum likelihood (ML) estimation method for the parameters of ESBIII is considered. Robustness properties of the ML estimators are studied and tail behaviour of ESBIII distribution is also examined. The applications on real data are considered to illustrate the modelling capacity of this distribution in the class of unimodal and also bimodal distributions.

Keywords: *asymmetry; Burr III distribution; bimodality; epsilon skew; robustness.*

References

[1] Arslan, O., Genç, A.İ. (2009), *The skew generalized t distribution as the scale mixture of a skew exponential power distribution and its applications in robust estimation*, Statistics, 43(5), 481-498.

[2] Çankaya, M.N., Yalçınkaya, A., Altındağ, Ö., Arslan, O. (2017). On The Robustness of Epsilon Skew Extension for Burr III Distribution on Real Line, Computational Statistics, Revision.

[3] Hampel, F.R., Ronchetti, E.M., Rousseeuw, P.J. and Stahel, W.A. (1986), *Robust Statistics: The Approach Based on Influence Functions*. Wiley Series in Probability and Statistics, 465.

10TH ISC2017 INTERNATIONAL STATISTICS CONGRESS
December 6-8, 2017 ANKARA/TURKEY

Katugampola Fractional Integrals Within the Class of s-Convex Functions

Hatice YALDIZ[1]
yaldizhatice@gmail.com

[1]*Karamanoğlu Mehmetbey University, Department of Mathematics, Karaman, TURKEY*

The aim of this paper is to the Hermite-Hadamard and midpoint type inequalities for functions whose first derivatives in absolute value is $s-convex$ through the instrument of generalized Katugampola fractional integrals. Then, if we give the definition of Katugampola[4] fractional integrals,

Definition. Let f∈[a,b].

1.The left-sided Katugampola fractional integral ${}^{\rho}I_{a^+}^{\alpha} f$ of order α∈C, Re(α)>0 is defined by

$$ {}^{\rho}I_{a^+}^{\alpha} f(x) = \frac{\rho^{1-\alpha}}{\Gamma(\alpha)} \int_a^x \frac{t^{\alpha-1}}{\left(x^{\rho}-t^{\rho}\right)^{1-\alpha}} f(t)dt, x > a, $$

2.The right-sided Katugampola fractional integral ${}^{\rho}I_{b^-}^{\alpha} f$ of order α∈C, Re(α)>0 is defined by

$$ {}^{\rho}I_{b^-}^{\alpha} f(x) = \frac{\rho^{1-\alpha}}{\Gamma(\alpha)} \int_x^b \frac{t^{\alpha-1}}{\left(t^{\rho}-x^{\rho}\right)^{1-\alpha}} f(t)dt, x < b. $$

As a first application of this new concept, we state and prove Hermite-Hadamard type inequalities for the Katugampola fractional integrals by using s-convex functions. Second, we need to give a lemma for differentiable functions which will help us to prove our main theorems. Then, we present some theorems which are the generalization of those given in earlier works.

We close the paper by results presented in this study would provide generalizations of those given in earlier works. The findings of this study have a number of important implications for future practice.

Keywords: *$s-convex$ function, Hermite-Hadamard type inequality, Katugampola fractional integrals*

References

[1] Chen, H., Katugampola, U.N.(2017), *Hermite-Hadamard and Hermite-Hadamard-Fejer type inequalities for generalized fractional integrals*, J. Math. Anal. Appl., 446, 1274-1291.

[2] Dragomir, S.S.,Pearce, C.E.M. (2000), *Selected topics on Hermite--Hadamard inequalities and applications*, RGMIA Monographs, Victoria University.

[3] Gabriela, C. (2017), *Boundaries of Katugampola fractional integrals within the class of (h1; h2)-convex functions*, https://www.researchgate.net/publication/313161140.

[4] Katugampola, U.N. (2011), *New approach to a generalized fractional integrals*, Appl. Math. Comput., 218 (4), 860-865.

SESSION VI

APPLIED STATISTICS VIII

ISC2017 INTERNATIONAL STATISTICS CONGRESS
December 6-8, 2017 ANKARA/TURKEY

Intensity Estimation Methods for an Earthquake Point Pattern

Cenk İÇÖZ[1] and K. Özgür PEKER[1]
cicoz@anadolu.edu.tr, opeker@anadolu.edu.tr

[1] *Anadolu University, Eskişehir, Turkey*

A spatial point pattern is the set of points which is irregularly distributed within a region of space. Several examples for a spatial point pattern can be given as the locations of a certain tree type in a forest, crime areas in a neighbourhood and earthquakes occurred in a geographic region. These specific locations are defined as events to separate from arbitrary points in the domain. There are three fundamental pattern types for the spatial point patterns: clustered, regular and completely random patterns. Each of these patterns can be counted as the typical outcome of stochastic mechanisms called spatial point processes.

Intensity of a point pattern is the number of events occurred per unit area. For a spatial point process, intensity can be described as in the equation below [3].

$$\lambda(s) = \lim_{ds \to 0} \frac{E[N(ds)]}{ds}$$

Estimation of the intensity of a spatial point is the primary goal of the spatial point pattern analysis. It is an aid in determining risk and also determining the hot and cold spots. In addition, intensity is a one of the determinant for the pattern type. There are many estimation methods for intensity in point pattern literature. In this study, several estimation methods such as kernel density estimation with different bandwidths and adaptive smoothing for earthquake patterns are applied and the results are compared.

Keywords: *kernel density estimation, quadrat counts, adaptive smoothing, point processes, point patterns*

References

[1] Baddeley, A., Rubak, E. and Turner, R. (2015). *Spatial Point Patterns: Methodology and Applications with R*. London: Chapman and Hall/CRC Press

[2] Diggle, P. J. (2013) *Statistical Analysis of Spatial and Spatio-Temporal Point Patterns* Chapman and Hall/CRC Press

[3] Gatrell, A. C., Bailey, T. C., Diggle, P. J., & Rowlingson, B. S. (1996). Spatial point pattern analysis and its application in geographical epidemiology. *Transactions of the Institute of British geographers*, 256-274.

[4] Shabenberger, O., & Gotway, A. C. (2005). *Statistical Methods for Spatial Data Analysis.* Chapman & Hall/ CRC.

10TH ISC2017 INTERNATIONAL STATISTICS CONGRESS
December 6-8, 2017 ANKARA/TURKEY

Causality Test for Multiple Regression Models

Harun YONAR[1], Neslihan İYİT[1]
hyonar@selcuk.edu.tr, niyit@selcuk.edu.tr

[1] *Selcuk University, Science Faculty, Statistics Department, Konya, Turkey*

Regression analysis used in modeling the relationshis between variables involves a number of assumptions that can affect the model specification. The correct choice of variables is very important for testing the assumptions in multiple regression models. If the dependent or independent variables are not choiced correctly, explanations of the model will move away from its purpose. No matter how meaningful and strongly the statistical relationship between variables, this can not mean any causality relationship between these variables. When the time series is concerned, the relationship between variables is a sign of causality. In this study, multiple regression models are constituted in examining the economic development of countries and the results of causality analysis are taken into consideration to obtain the most suitable regression model among them. In this point, the effectiveness of causality test is investigated in the comparison of established regression models.

Keywords: *Causality test, multiple regression model, time series, economic development.*

References

[1] Kendall, M.G. and Stuart, A. (1961), *The advanced theory of statistics*, New York, Charles Griffin Publishers, p 279.

[2] Koop, G. (2000), *Analysis of economic data*, New York, John Wiley & Sons, p 175.

[3] Dobson, A.J. and Barnett, A. (1990), *An introduction to generalized linear models*, Chapman And Hall, p 59-89.

[4] McCullagh, P. and Nelder, J.,A. (1989), Generalized Linear Models, London, Second Edition, Chapman & Hall/CRC, p 21-48.

[5] Stock, .H. and Watson, M.W. (1989), *Interpreting the evidence on money-income causality*, North Holland, Journal of economics, p 161-181.

10TH ISC2017 INTERNATIONAL STATISTICS CONGRESS
December 6-8, 2017 ANKARA/TURKEY

Drought Forecasting with Time Series and Machine Learning Approaches

Ozan EVKAYA[1], Ceylan YOZGATLIGİL[2], A. Sevtap SELCUK-KESTEL[2]
ozanevkaya@gmail.com, ceylan.yozgatligil@gmail.com, skestel@metu.edu.tr

[1]*Atilim University, Ankara, Turkey*
[2]*Middle East Technical University, Ankara, Turkey*

As a main reason of undesired agricultural, economic and environmental damages, drought is one of the most important stochastic natural hazard having certain features. In order to manage the impacts of drought, more than 100 drought indices have been proposed for both monitoring and forecasting purposes [1], [3]. For different types of droughts, these indices have been used to understand the effects of dry periods including meteorological, agricultural and hydrological droughts in many distinct locations. In this respect, the future projections of drought indices allow the decision makers to assess certain risks of dry periods beforehand. In addition to the use of classical time series techniques for understanding the upcoming droughts, machine learning methods might be effective alternatives for forecasting the future events based on relevant drought index [2].

This study aims to identify the benefits of various methods for forecasting the future dry seasons with widely known drought indices. For that purpose, Standardized Precipitation Index (SPI), Standardized Precipitation Evapotranspiration Index (SPEI) and Reconnaissance Drought Index (RDI) have been considered over different time scales (3, 6, 9 months) to represent drought in Kulu weather station, Konya. The considered drought indices were used for forecasting the future period using both time series prediction tools and machine learning techniques. The forecast results of all methods with respect to different drought indices were examined with the data set of 1950-2010 for Kulu station. The potential benefits and limitations of various methods and drought indices were discussed in detail.

Keywords: *drought, drought index, forecast, machine learning*

References

[1] A. Askari, K.O. (2017), *A Review of Drought Indices*, Int. Journal of Constructive Research in Civil Engineering (IJCRCE), 3(4), 48-66.

[2] Belayneh, A. M., Adamowski, J. (2013), *Drought Forecasting using New Machine Learning Methods*, Journal of Water and Land Development, 18, 3-12.

[3] Zargar, A., Sadiq, R., Naser, B. and Khan, I. F. (2011), *A Review of Drought Indices*, Environ. Rev., 19, 333-349.

10TH ISC2017 INTERNATIONAL STATISTICS CONGRESS
December 6-8, 2017 ANKARA/TURKEY

Stochastic Multi Criteria Decision Making Methods for Supplier Selection in Green Supply Chain Management

Nimet YAPICI PEHLİVAN[1], Aynur ŞAHİN[1]
nimet@selcuk.edu.tr, aynursahin@selcuk.edu.tr

[1]*Selçuk University, Science Faculty, Department of Statstics, Konya, Türkiye*

Supplier selection is one of the most important problems in supply chain management (SCM) which considers multiple objectives and multiple criteria. Most of the earlier studies on supplier selection have focused on conventional criteria such as price, quality, production capacity, purchasing cost, technology and delivery time. But, more recent studies have dealt with the integration of environmental factors with supplier selection decisions. Green Supply Chain Management (GSCM) is defined as integrating environmental thinking into the SCM, including product design, material sourcing and selection, manufacturing processes, delivery of the final product to the consumers, as well as end-of-life management of the product after its useful life [2].
Several multi criteria decision making methods (MCDM) for supplier selection have been introduced such as AHP, ANP, TOPSIS, ELECTRE, GRA, etc. and their hybrids or fuzzy versions [2, 3, 4]. Stochastic analytical hierarchy process (SAHP) that can handle uncertain information and identify weights of criteria in the MCDM problem are proposed by [1] and [5]. In their studies, evaluations of the Decision Makers (DMs) including imprecise values are converted into crisp ones by utilizing the beta distribution to compute the weights.
In this study, we introduce stochastic multi criteria decision making methods to evaluate the supplier selection in green supply chain management which considers environmental criteria and sub-criteria, through a numerical example.

Keywords: *Stochastic multi criteria decision making, green supply chain, supplier selection*

References

[1] Çobuloğlu, H.I., Büyüktahtakın, İ.E. (2015), *A stochastic multi-criteria decision analysis for sustainable biomass crop selection,* Expert Systems with Applications, Volume 42, Issues 15–16, Pages 6065-6074.

[2] Hashemi, S.H., Karimi, A., Tavana, M., (2015), *An integrated green supplier selection approach with analytic network process and improved Grey relational analysis*, Int. J.Production Economics, Vol.159, Pages 178–191.

[3] Kannan, D., Khodaverdi, R., Olfat, L., Jafarian, A., Diabat, A. (2013), *Integrated fuzzy multi criteria decision making method and multiobjective programming approach for supplier selection and order allocation in a green supply chain*, Journal of Cleaner Production, Volume 47, Pages 355-367.

[4] Govindan, K., Rajendran, S., Sarkis, J. Murugesan, P.. (2015), *Multi criteria decision making approaches for green supplier evaluation and selection: a literature review*, Journal of Cleaner Production, Volume 98, Pages 66-83

[5] Jalao, E.R., Wu, T., Shunk, D. (2014), *A stochastic AHP decision making methodology for imprecise preferences*, Information Sciences, Volume 270, Pages 192-203.

10TH ISC2017 INTERNATIONAL STATISTICS CONGRESS
December 6-8, 2017 ANKARA/TURKEY

Parameter Estimation of Three-parameter Gamma Distribution using Particle Swarm Optimization

Aynur ŞAHİN[1], Nimet YAPICI PEHLİVAN[1]
aynursahin@selcuk.edu.tr, Nimet@selcuk.edu.tr

[1]*Selcuk University, Konya, Turkey*

Three-parameter (3-p) Gamma distribution is widely utilized for modelling skewed data in applications of hydrology, finance and reliability. The estimation of the parameters of this distribution is required in the most real applications. Maximum likelihood (ML) is the most popular method used in parameter estimation since ML estimators are unbiased and minimum variance. This method is based on finding the parameter values that make maximizing the likelihood function of a given distribution. Maximizing the likelihood function of a 3-p Gamma distribution is a quite difficult problem and this problem cannot be solved by using conventional optimization methods such as the gradient-based method. Thus, it is reasonable to use metaheuristic methods at this stage. Particle Swarm Optimization (PSO) is one the most popular population-based metaheuristic methods. In this paper, we proposed an approach to maximize the likelihood function of 3-p Gamma distribution by using PSO. Simulation results show that the PSO approach provides accurate estimates and it is satisfactory for the parameter estimation of the 3-p Gamma distribution.

Keywords: *Three parameter-Gamma distribution, Maximum Llikelihood Estimation, Particle Swarm Optimization.*

References

[1] Abbasi, B., Jahromi, A. H. E., Arkat, J. and Hosseinkouchack, M. (2006), *Estimating the parameters of Weibull distribution using simulated annealing algorithm*, Applied Mathematics and Computation, 85-93.

[2] Örkcü, H. H., Özsoy, V. S., Aksoy, E. and Dogan, M. I. (2015), *Estimating the parameters of 3-p Weibull distribution using particle swarm optimization: A comprehensive experimental comparison*, Applied Mathematics and Computation, 201-226.

[3] Vaidyanathan, V. and Lakshmi, R.V. (2015), *Parameter Estimation in Multivariate Gamma Distribution*, Statistics, Optimization & Information Computing, 147-159.

[4] Vani Lakshmi, R. and Vaidyanathan, V.S.N. (2016), *Three-parameter gamma distribution: Estimation using likelihood,spacings and least squares approach*, Journal of Statistics & Management Systems, 37-53.

[5] Zoraghi, N., Abbasi, B., Niaki, S. T. A. and Abdi, M. (2012), *Estimating the four parameters of the Burr III distribution using a hybrid method of variable neighborhood search and iterated local search algorithms*, Applied Mathematics and Computation, 9664-9675.

SESSION VI

OTHER STATISTICAL METHODS IV

Word Problem for the Schützenberger Product

Esra KIRMIZI ÇETİNALP[1], Eylem GÜZEL KARPUZ[1], Ahmet Sinan ÇEVİK[2]
esrakirmizi@kmu.edu.tr, eylem.guzel@kmu.edu.tr, sinan.cevik@selcuk.edu.tr

[1]*Karamanoğlu Mehmetbey University Department of Mathematics, Karaman, Turkey*
[2]*Selcuk University Department of Mathematics, Konya, Turkey*

Presentation of Schützenberger product play a crucial role in various subsections of mathematics such as automata theory, combinatorial group theory, semigroup theory . In this work, we consider monoid presentation of the Schützenberger product of n groups which is obtained by matrix theory [3]. We compute complete rewriting system for this monoid presentation. Thus, by this complete rewriting system we characterize the structure of elements of this product [2]. Therefore, we obtain solvability of the word problem [1].

Keywords: *Schützenberger Product, Rewriting Systems, Normal Form*

References

[1] Book, R. V. (1987), *Thue systems as rewriting systems*, J. Symbolic Computation, 3 (1-2), 39-68.

[2] Çetinalp, E. K. and Karpuz, E. G., Çevik, A. S. (2019) *Complete Rewriting System for Schützenberger Product of n Groups,* Asian-European Journal of Mathematics, 12(1).

[3] Gomes, G. M. S., Sezinando, H. and Pin, J. E. (2006), *Presentations of the Schützenberger product of n groups*, Communications in Algebra, 34(4) 1213-1235.

10TH ISC2017 INTERNATIONAL STATISTICS CONGRESS
December 6-8, 2017 ANKARA/TURKEY

Automata Theory and Automaticity for Some Semigroup Constructions

Eylem GÜZEL KARPUZ[1], Esra KIRMIZI ÇETİNALP[1], Ahmet Sinan ÇEVİK[2]
esrakirmizi@kmu.edu.tr, eylem.guzel@kmu.edu.tr, sinan.cevik@selcuk.edu.tr

[1] *Karamanoğlu Mehmetbey University Department of Mathematics, Karaman, Turkey*
[2]*Selcuk University Department of Mathematics, Konya, Turkey*

Automata theory is the study of abstract computing devices, or "machines". Before there were computers, in the 1930's, Alan Turing studied an abstract machine that had all the capabilities of today's computers. Turing's goal was to describe precisely the boundary between what a computing machine could do and what it could not do. Turing's conclusions apply not only to his abstract *Turing machines*, but to today's real machines [1].

In this talk, firstly, I will give some information about automata theory and automaticity. Then, I will present some results on automatic structure for some semigroup constructions; namely direct product of semigroups and generalized Bruck-Reilly *-extension of a monoid [2, 3].

Keywords: *automata, automatic structure; presentation; generalized Bruck-Reilly *-extension*

References

[1] Hopcroft, J. E., Motawa, R. and Ullman, J. D. (2000), *Introduction to Automata Theory, Languages, and Computation*, Pearson Educations, Inc.

[2] Karpuz, E. G., Çetinalp, E. and Çevik, A. S. *Automatic structure for generalized Bruck-Reilly *-extension* (preprint).

[3] Kocapınar, C., Karpuz, E. G., Ateş, F., Çevik, A. S. (2012), *Gröbner-Shirshov bases of the generalized Bruck-Reilly *-extension*, Algebra Colloquium, 19 (Spec 1), 813-820.

10TH ISC2017 INTERNATIONAL STATISTICS CONGRESS
December 6-8, 2017 ANKARA/TURKEY

The Structure of Hierarchical Linear Models and a Two-Level HLM Application

Yüksel Akay Ünvan [1], Hüseyin Tatlidil [2]
akay.unvan@gmail.com, tatlidil@hacettepe.edu.tr

[1]*Türk Eximbank, Ankara, Turkey*
[2]*Hacettepe University, Ankara, Turkey*

This study aims to describe the structure of Hierarchical Linear Models (HLM). The hierarchical linear models (HLM) structure, which is also known as "nested models", "multilevel linear models" (in sociological research), "mixed effect models" / "random effect models" (biostatistics), "random coefficient regression models" (in econometrics) or "covariance components models" (in the statistics), is used in the study in order to explain the structure of hierarchical data. The circumstances in which HLM is used and the basic points that HLM focuses are highlighted. The advantages of HLM, its mathematical theory, equalities and assumptions are also emphasized. Furthermore, previous studies on this subject are widely covered in the study. PISA 2012 application is the fifth of PISA assessments that began in 2000 and repeated every three years and PISA 2012 research has mainly focused on mathematics literacy skills. For this reason, some factors affecting the mathematical success of Turkish students participated in PISA 2012 were discussed in the study both in school and student level and the extent to which these factors explain the student's success scores were investigated by using the HLM method. A two-level HLM has been created that examines the effects of school and student-level characteristics on mathematical success. In the application part of the study, the HLM 6.0 software is used.

Keywords: *Hierarchical Data, Hierarchical Linear Models, PISA 2012*

References

[1] Abbott, M.L., Joireman, J. and Stroh, H.R. (2002), *The Influence of District Size, School Size and Socioeconomic Status on Student Achievement in Washington: A Replication Study Using Hierarchical Linear Modeling*, A Technical Report For The Washington School Research Center.

[2] Atar, H.Y. and Atar, B. (2012a), *Investigating the Multilevel Effects of Several Variables on Turkish Students' Science Achievements on TIMSS*, Journal of Baltic Science Education, 11.

[3] Erberber, E. (2010), *Analyzing Turkey's Data from TIMSS 2007 to Investigate Regional Disparities in Eighth Grade Science Achievement*, in Alexander W. Wiseman (ed.) The Impact of International Achievement Studies on National Education Policymaking (International Perspectives on Education and Society, Volume 13), Emerald Group Publishing Limited, pp. 119-142.

[4] Fullarton, S., Lokan, J., Lamb, S. and Ainley, J. (2003), *Lessons from the Third International Mathematics and Science Study*, TIMSS Australia Monograph No. 4. Melbourne: Australian Council for Educational Research.

[5] Heck, R.H. and Thomas, S. L., (2000), *An Introduction To Multilevel Modeling Techniques*, Lawrence Erlbaum Associates, London.

ISC2017 INTERNATIONAL STATISTICS CONGRESS
December 6-8, 2017 ANKARA/TURKEY

Credit Risk Measurement Methods and a Modelling on a Sample Bank

Yüksel Akay Ünvan [1], Hüseyin Tatlidil [2]
akay.unvan@gmail.com, tatlidil@hacettepe.edu.tr

[1]*Türk Eximbank, Ankara, Turkey*
[2]*Hacettepe University, Ankara, Turkey*

The accurate measurement of credit risk concept has kept the banking world busy for a long time. As a result of the crises experienced in Turkey, the banking sector has become more sensitive about the measurement and modelling of the credit risk. The credit risk measurement and modelling methods are applied within the framework of some international standards. The Basel II consensus takes place at this point. There is a need for banks to have sufficient equity to deal with the risks they encounter or may encounter during their operations. Effective and continuous control of this process by the authority is important. In this study, some of the credit risk calculation methods will be explained and an application will be made regarding the measurement and modelling of credit risk of an investment bank operating in Turkey.

Keywords: *Credit Risk, Basel II, Basel III, Equity, Capital Adequacy Ratio*

References

[1] Arunkumar, R., Kotreshwar, G. (2006), *Risk Management in Commercial Banks (A Case Study of Public and Private Sector Banks)*, Indian Institute of Capital Markets 9th Capital Markets Conference Paper, 1-22.

[2] Banking Regulation and Supervision Agency (BRSA) Report (2013), http://www.bddk.org.tr/websitesi/turkce/kurum_bilgileri/sss/10469basel6.pdf.

[3] Giesecke, K. (2004), *Credit Risk Modeling and Valuation: An Introduction*, Working Papers Series, 1-67. An abridged version of this article is published in Credit Risk: Models and Management, Vol. 2, D. Shimko (Editor), Riskbooks, London.

[4] Jacobson T, Lindé J., Roszbach K. (2005), *Credit risk versus capital requirements under Basel II: are SME loans and retail credit really different*, Journal of Financial Services Research, 28:1, 43, 75.

[5] Stephanou, C. , Mendoza, J. C. (2005), *Credit Risk Measurement Under Basel II: An Overview and Implementation Issues for Developing Countries*, World Bank Policy Research Working Paper No. 3556, 1-33.

10TH ISC2017 INTERNATIONAL STATISTICS CONGRESS
December 6-8, 2017 ANKARA/TURKEY

A Comparison on the Ranking of Decision Making Units of Data Envelopment and Linear Discriminant Analysis

Hatice ŞENER[1], Semra ERBAŞ[1], Ezgi NAZMAN[1]
haticesener42@gmail.com, serbas@gazi.edu.tr, ezgicabuk@gazi.edu.tr

[1]*Gazi University, Graduate School of Natural and Applied Sciences, Department of Statistics, Ankara, Turkey*

Data Envelopment Analysis (DEA) is a linear programming based on non-parametric method that is commonly used for ranking and classification of decision making units by utilizing certain inputs and outputs. Linear Discriminant Analysis (LDA) however, is a multivariate statistical method that is used to estimate group membership of units. The discriminant scores obtained using LDA can be used as an alternative to the DEA method for ranking of units. In this study, 9 variables representing the social development levels of 61 countries are employed. These countries are ranked separately according to the efficiency scores obtained by the DEA and the discriminant scores calculated by the LDA. The Spearman Rank Correlation Coefficient is examined in order to analyse the relationship between the rankings acquired by utilizing these two methods. Furthermore, in order to determine if there is a fit between DEA and LDA methods Mann- Whitney U ranking test - a non-parametric test - is used.

Keywords: *Data Envelopment Analysis, Linear Discriminant Analysis, Ranking units*

References

[1] Sinuany-Stern, Z. and Friedman, L. (1998), *DEA and the discriminant analysis of ratios for ranking units*, European Journal of Operational Research, 111, 470-478.

[2] Adler, N., Friedman, L. and Sinuany-Stern, Z.(2002),*Review of ranking methods in data envelopment analysis context*, European Journal of Operational Research, 140, 249-265

[3] Friedman, L. and Sinuany-Stern, Z. (1997), *Scaling units via the canonical correlation analysis in the DEA context*, European Journal of Operational Research, 100, 629-637.

[4] Bal, H. and Örkçü, H.H. (2005), *Combining the discriminant analysis and the data envelopment analysis in view of multiple criteria decision making:a new model*, Gazi University Journal of Science, 18(3), 355-364.

[5] Charnes, A.,Cooper, W.W. and Rhodes,E.(1978), *Measuring the efficiency of decision making units*, , European Journal of Operational Research, 2(6), 429-444.

SESSION VI

MODELING AND SIMULATION III

10TH ISC2017 INTERNATIONAL STATISTICS CONGRESS
December 6-8, 2017 ANKARA/TURKEY

Classifying of Pension Companies Operating in Turkey with Discriminant and Multidimensional Scaling Analysis

Murat KIRKAĞAÇ[1], Nilüfer DALKILIÇ[1]
murat.kirkagac@dpu.edu.tr, nilufer.dalkilic@dpu.edu.tr

[1]*Dumlupınar University, Kütahya, Turkey*

The Individual Pension System is a private retirement system that enables people to earn income that can maintain their standard of living in retirement periods by directing long-term investment in the savings they make during their active working life. The significance of the Individual Pension System in Turkey has increased considerably in recent years. As of the end of 2016, 7.789.431 contracts are in force. Besides, the number of participants increased by approximately 10% compared to the end of the previous year and reached 6.6 million in the system. Automatic enrolment in the Individual Pension System has also entered into force since January 1, 2017 [1].

The aim of this study is to classify fifteen pension companies operating in Turkey between 2012 and 2016, according to their financial performance. For this purpose, discriminant analysis and multidimensional scaling analysis, which are frequently used in statistical analyses, have been used. Discriminant analysis is a classification technique, where multiple clusters are known a priori and multiple new observations are classified into one of the known clusters based on the measured properties [2]. Multidimensional scaling analysis is a statistical method that reveals relationships between objects by making use of distances where distance between objects is not known but distances between them can be calculated [3].

The variables used in the analysis are the Individual Pension System basic indicators obtained from the Pension Monitoring Center [1] and main financial indicators obtained from the reports on insurance and private pension activities prepared by the Republic of Turkey Prime Ministry Undersecretariat of Treasury Insurance Auditing Board [4]. As a result of the study, the results obtained by both methods are examined and it is observed that the classification results obtained by these two methods are consistent with each other.

Keywords: *individual pension system, discriminant analysis, multidimensional scaling analysis.*

References

[1] Pension Monitoring Center, http://www.egm.org.tr/, (November,2017).
[2] Tatlıdil, H., (2002), *Uygulamalı çok değişkenli istatistiksel analiz*, Turkey, Akademi Matbaası, 256.
[3] Kalaycı, Ş., (2016), *Spss uygulamalı çok değişkenli istatistik teknikleri*, Turkey, Asil yayın dağıtım, 379.
[4] Undersecretariat of Treasury, https://www.hazine.gov.tr, (November,2017).

10TH ISC2017 INTERNATIONAL STATISTICS CONGRESS
December 6-8, 2017 ANKARA/TURKEY

A Bayesian Longitudinal Circular Model and Model Selection

Onur Camli[1], Zeynep Kalaylioglu[1]
camli@metu.edu.tr, kzeynep@metu.edu.tr

[1]*Department of Statistics, Middle East Technical University, Ankara, Türkiye*

The focus of the current study is the analysis and model selection for circular longitudinal data. Our research was motivated by a study conducted in Ankara University, Department of Gynecology that collects data on head angles of the fetus every 15 minutes of the last xx hour of the birth. There are a number of statistical methods to analyse longitudinal data in linear structure. However, the literature on statistical modeling of longitudinal circular response is limited and model selection methods in that context are not well addressed. We considered a Bayesian random intercept model on the circle to investigate relationships between univariate circular response variable and several linear covariates. This model enables simultaneous inference for all model parameters and prediction. For model selection purpose, we defined the predictive loss function in terms of angular distance between predicted and observed circular response variable and developed new criteria that are based on minimizing the total posterior predictive loss. Extensive Monte Carlo simulation studies controlled for the sample size and intraclass correlation were used to study the performances of the model and the model selection criteria under various realistic longitudinal circular settings. Relative bias and mean square error were used to evaluate the performance of the estimators under correctly specified models and robustness to model misspecification. Several quantities were used to evaluate the performances of the model selection criteria such as frequency of selecting the true model and a ratio that measures the strength of the particular selection. Simulations reveal a noticeable or equivalent gain in performance achieved by the proposed methods. A conventional longitudinal data set (sandhopper data) was used to further compare the Bayesian model selection methods for circular data. This research hopes to address and contribute to the model selection in circular data, a rather fertile area for methodological and theoretical development, while the demand increases with the circular complex data obtained through advancing technology in real life applications and studies.

Keywords: *Directional Statistics, Random Effects, Model Selection, Biology.*

References

[1] D'Elia, A. (2001), *A statistical model for orientation mechanism*, Statistical Methods and Applications, 10, 157–174.

[2] Fisher, N.I., and Lee A.J. (1992), *Regression models for angular response*, Biometrics, 48, 665–677.

[3] Nunez-Antonio, G., and Gutierrez-Pena E. (2014), *A Bayesian model for longitudinal circular data based on the projected normal distribution*, Computational Statistics and Data Analysis, 71, 506–519.

[4] Ravindran, P.K., and Ghosh, S.K. (2011), *Bayesian analysis of circular data using wrapped distributions*, Journal of Statistical Theory and Practice, 5, 547-561.

10TH ISC2017 INTERNATIONAL STATISTICS CONGRESS
December 6-8, 2017 ANKARA/TURKEY

A Computerized Adaptive Testing Platform: *Smart*CAT

Beyza Doğanay ERDOĞAN[1], Derya GÖKMEN[1], Atilla Halil ELHAN[1], Umut YILDIRIM[2], Alan TENNANT[3]
beyzadoganay@gmail.com, dgokmen2001@yahoo.com, ahelhan@yahoo.com, umutyil@gmail.com, alan.tennant@paraplegie.ch

[1]*Ankara University Faculty of Medicine Department of Biostatistics, Ankara, Turkey*
[2]*UMUTY Bilgisayar, Ankara, Turkey*
[3] *Rue Alberto Giacometti 13 Le Grand Saconnex, Geneva 1218, Switzer1and*

Computerized adaptive testing (CAT), which has also been called tailored testing, is a form of computer-based test that adapts to the examinee's ability level. In CAT, when a test is administered to a patient by using a program, the program estimates the patient's ability after each question, and then that ability estimate can be used in the selection of subsequent items. For each item, there is an item information function, and the next item chosen is usually that which maximises this information. The items are calibrated by their difficulty levels from the item bank. When a predefined stopping rule is satisfied, the assessment is completed [3].
In this study, a newly developed CAT software *Smart*CAT will be introduced. *Smart*CAT is a computer program for performing both simulated and real CAT, generating data for simulated CAT, creating item banks for real CAT with both dichotomous and polytomous items. Rasch family models (one-parameter, rating scale and partial credit models) [1,2] were supported by the program. The program provides different item selection methods (maximum Fisher information, maximum posterior weighted information, maximum likelihood weighted information), and theta estimation methods (maximum likelihood, expected a priori and maximum a pirori). The use of *Smart*CAT will be demonstrated by real and simulated data examples.

Keywords: *item bank, tailored test, computerized adaptive test, Rasch model*

References

[1] Doğanay Erdoğan B., Elhan A.H., Kaskatı O.T., Öztuna D., Küçükdeveci A.A., Kutlay Ş., Tennant A. (2017). *Integrating patient reported outcome measures and computerized adaptive test estimates on the same common metric: an example from the assessment of activities in rheumatoid arthritis.* Int J Rheum Dis.; 20(10):1413-1425.

[2] Elhan A.H., Öztuna D., Kutlay Ş., Küçükdeveci A.A., Tennant A. (2008). *An initial application of computerized adaptive testing (CAT) for measuring disability in patients with low back pain.* BMC Musculoskel Dis.; 9:166.

[3] Öztuna D., Elhan A.H., Küçükdeveci A.A., Kutlay Ş., Tennant A. (2010). *An application of computerised adaptive testing for measuring health status in patients with knee osteoarthritis.* Disabil Rehabil.; 32(23):1928-1938.

10TH ISC2017 INTERNATIONAL STATISTICS CONGRESS
December 6-8, 2017 ANKARA/TURKEY

Educational Use of Social Networking Sites in Higher Education: A Case Study on Anadolu University Open Education System

Md Musa KHAN[1], Zerrin AŞAN GREENACRE[1]
khanstatcu@gmail.com, zasan@anadolu.edu.tr

[1]*Anadolu University Department of Statistics, Eskisehir, Turkey*

The growth of the information communication technology, distance education as a primary means of instruction is expanding significantly at higher education. A growing number of higher education instructors are launching to link distance education delivery with "Social Networking Sites" (SNSs). In order to evaluate the largely unexplored educational benefits, importance and efficiency of SNSs in higher education a non-probability based web surveys was conducted on Open Education System's students in Anadolu University. This study explored how "Social Networking Sites" can be used to supplementary face-to-face courses instrument of enriching students' sense of community and, thus, to encourage classroom communities of practice in the background of higher education. Firstly, we use bivariate analysis for association among the selected variables and finally use logit regression on those variables which are significant in bivariate analysis. The results suggest that education based SNSs can be used most effectively in distance education courses as an information communication technological tool for betterment online communications among students for higher education.

Keywords: *Information communication technology, Distance education, Social networking sites (SNSs), Higher education, Open education system.*

References

[1] Anderson, T. (2005). *Distance learning—Social software's killer ap.* [Electronic version] Proceedings from Conference of the Open and Distance Learning Association of Australia (ODLAA). Adelaide, South Australia: University of South Australia.

[2] Correia, A., & Davis, N. (2008). *Intersecting communities of practice in distance education: The program team and the online course community.* Distance Education, 29(3), 289-306.

[3] Selwyn, N. (2000). *Creating a "connected" community? Teachers' use of an electronic discussion group.* Teachers College Record, 102, 750-778.

[4] Shea, P.J. 2006. A study of students' sense of learning community in an online learning environment. Journal of Asynchronous Learning Networks 10, no. 1: 35-44.

[5] Summers, J.J., and M.D. Svinicki. 2007. *Investigating classroom community in higher education.* Learning and Individual Differences 17, no. 1: 55-67.

ISC2017 INTERNATIONAL STATISTICS CONGRESS
December 6-8, 2017 ANKARA/TURKEY

An Improved New Exponential Ratio Estimators for Population Median Using Auxilary Information In Simple Random Sampling

Sibel AL[1], Hulya CINGI[2]
sibelal84@gmail.com, hcingi@hacettepe.edu.tr

[1]*General Director of Service Provision, Republic of Turkey Social Security Institution, Bakanlıklar, Ankara*
[2]*University of Hacettepe, Faculty of Science, Department of Statistics, Beytepe, Ankara, Turkey*

Median is often regarded as more appropriate measure of location than mean when variables have a highly skewed distribution, such as income, expenditure, production are studied in survey sampling. In literature, there have been many studies for estimating the population mean and population total but relatively less effort has been devoted to the development of efficient methods for estimating the population median.

In simple random sampling, Gross [2] defined sample median. Kuk and Mak [3] suggested a ratio estimator and obtained the MSE equation. Aladag and Cingi [1], made the first contribution in using exponential estimator for estimating the population median.

Following Singh et. al. [4], we define new exponential ratio estimators for population median and derive the minimum mean square error (MSE) equations of the proposed estimators for constrained and unconstrained choice of α_1 and α_2. We compare MSE equations and find theoretical conditions which make each proposed estimator more efficient than the others given in literature. These conditions are also supported by using numerical examples.

Keywords: *Auxiliary information, exponential estimator, median estimation, simple random sampling.*

References

[1] Aladag, S., Cingi, H., (2012), *A New Class of Exponential Ratio Estimators for Population Median in Simple Random Sampling*, 8th International Symposium of Statistics, 11-13 October, Eskisehir, Turkey.

[2] Gross, S. T., (1980), *Median estimation in sample surveys*, Proceedings of the Survey Research Methods Section, American Statistical Association, 181-184.

[3] Kuk, A. Y. C., Mak, T. K., (1989), *Median estimation in the presence of auxiliary information*, Journal of the Royal Statistical Society Series, B. 51(2), 261-269.

[4] Singh, R., Chauhan, P., Sawan, N., Smarandache, (2009), *Improvement in Estimating the Population Mean Using Exponential in Simple Random Sampling*, Bulletin of Statistics & Economics, F., 3 (A09), 13-19.

SESSION VI

OTHER STATISTICAL METHODS V

ISC2017 INTERNATIONAL STATISTICS CONGRESS
December 6-8, 2017 ANKARA/TURKEY

Demonstration of A Computerized Adaptive Testing Application Over A Simulated Data

Batuhan BAKIRARAR[1], İrem KAR[1], Derya GÖKMEN[1], Beyza DOĞANAY ERDOĞAN[1], Atilla Halil ELHAN[1]
batuhan_bakirarar@hotmail.com, iremkar.biostat@gmail.com, dgokmen2001@yahoo.com, beyzadoganay@gmail.com, ahelhan@yahoo.com

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

Computerized adaptive testing (CAT) is an algorithm which uses psychometric models to assess the examinees' abilities. Each of the examinees receives different items and number of items since CAT adapts the test to each examinee's ability level (θ). In CAT method, the answer given by the examinee to the first question plays key role in ordering the next questions [1]. The first question is generally at moderate strength in CAT. If the first question is answered correctly, the next one will become harder; if not, the next question will be easier. The logic behind this approach is that one cannot learn about the examined characteristic of the examinee from very easy or very hard questions, therefore the questions will be chosen from ones that will put forth individual's level of examined characteristic. A new estimate value ($\hat{\theta}$) is calculated based on the answers given to the items in this method. This process is repeated until the prespecified stopping criterion is met. Stopping criterion can be an indicator of certainty such as the number of applied items, change in the level of examined characteristic, the fact that questions to cover the target content have been applied, and standard error or a combination of these criteria [2]. The most advanced and efficient method used for measuring with questions bank is CAT. CAT applied with a suitable question bank is more effective than the classical method. While answering to all items of the scale in the classical method, in this method, examinees answer only the items in compliance with their level, which achieves estimation on the prespecified level of certainty with less number of items. Providing accurate results for examinees with all levels of skills and applying the evaluation whenever desired and achieving the results right away is the most distinctive advantage of CAT [2].
Use of CAT method for evaluation in health has been recently increasing, and studies on the subject indicate that results of evaluation through this method are successful and objectives are achieved. This study aims to provide general information about CAT and show that performance of CAT method is good when theta estimation is done with MLE. The study also tries to prove that information is achieved when all questions are answered with less questions. SmartCAT v0.9b for Windows was utilized for evaluation in the study.

Keywords: *computer adaptive testing, maximum likelihood estimation*

References

[1] Doğanay Erdoğan B., Elhan A.H., Kaskatı O.T., Öztuna D., Küçükdeveci A.A., Kutlay Ş., Tennant A. (2017), *Integrating patient reported outcome measures and computerized adaptive test estimates on the same common metric: an example from the assessment of activities in rheumatoid arthritis.* Int J Rheum Dis.; 20(10):1413-1425.

[2] Kaskatı O.T. (2011), *Rasch modelleri kullanarak romatoid artirit hastaları özürlülük değerlendirimi için bilgisayar uyarlamalı test yönteminin geliştirilmesi,* Ankara Üniversitesi, 100.

ISC2017 INTERNATIONAL STATISTICS CONGRESS
December 6-8, 2017 ANKARA/TURKEY

A Comparison of Maximum Likelihood and Expected A Posteriori Estimation in Computerized Adaptive Testing

İrem KAR[1], Batuhan BAKIRARAR[1], Beyza DOĞANAY ERDOĞAN[1], Derya GÖKMEN[1], Serdal Kenan KÖSE[1], Atilla Halil ELHAN[1]
iremkar.biostat@gmail.com, batuhan_bakirarar@hotmail.com, beyzadoganay@gmail.com, dgokmen2001@yahoo.com, s.kenan.kose@medicine.ankara.edu.tr, ahelhan@yahoo.com

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

A recent and probably most appealing new perspective offered by Item Response Theory (IRT) is the implementation of Computer Adaptive Testing (CAT) [2]. CAT algorithm uses psychometric models to assess the examinees' abilities. Each of the examinees receives different items and number of items since CAT adapts the test to each examinee's ability level (θ) [1]. The maximum likelihood estimator (MLE) and the expected a posteriori (EAP) estimator have been proposed for estimating a respondent's value of θ which are the two most frequently encountered in the literature. The MLE of θ is equal to the value of θ that maximizes the log-likelihood of the response pattern given fixed values of the item parameters. In contrast to the MLE, the EAP estimator yields usable estimates, regardless of the response pattern. The logic behind the EAP estimator is to obtain the expected value of θ given the response pattern of the individual [3].
The main purpose of this study is to compare MLE and EAP estimation in simulated data. In the simulated CAT application, the known item parameters simulate the responses of 1000 simulees having a uniform distribution between -2 and 2. All items are scaled using a 5-point Likert scale. The intraclass correlation coefficient and the Bland-Altman approach were used for evaluating the agreement between MLE (θ_{MLE}) and EAP (θ_{EAP}) estimates. The stopping rule allowed for the CAT was to stop once a reliability of 0.75 and 0.90 (i.e., its standard error equivalent) has been reached in both MLE (θ_{MLE}) and EAP (θ_{EAP}). Starting item was chosen as the item having median difficulty. SmartCAT v0.9b for Windows was utilized for evaluation in the study.

Keywords: *computer adaptive testing, expected a posteriori, maximum likelihood estimation*

References

[1] Doğanay Erdoğan B., Elhan A.H., Kaskatı O.T., Öztuna D., Küçükdeveci A.A., Kutlay Ş., Tennant A. (2017), *Integrating patient reported outcome measures and computerized adaptive test estimates on the same common metric: an example from the assessment of activities in rheumatoid arthritis.* Int J Rheum Dis.; 20(10):1413-1425.

[2] Forkmann, T., Kroehne, U., Wirtz, M., Norra, C., Baumeister, H., Gauggel, S., Elhan A.H., Tennant A., Boecker, M. (2013), *Adaptive screening for depression—Recalibration of an item bank for the assessment of depression in persons with mental and somatic diseases and evaluation in a simulated computer-adaptive test environment.* Journal of psychosomatic research, 75(5), 437-443.

[3] Penfield, R. D., Bergeron, J. M. (2005), *Applying a weighted maximum likelihood latent trait estimator to the generalized partial credit model.* Applied Psychological Measurement, 29(3), 218-233.

10TH ISC2017 INTERNATIONAL STATISTICS CONGRESS
December 6-8, 2017 ANKARA/TURKEY

Some Relations Between Curvature Tensors of a Riemannian Manifold

Gülhan AYAR[1], Pelin TEKİN[2], Nesip AKTAN[3]
gulhanayar@gmail.com, pelinpospos@gmail.com, nesipaktan@gmail.com

[1] *Karamanoğlu Mehmetbey University, Kamil Özdağ Science Faculty, Department of Mathematics, Karaman,Turkey,*
[2] *Trakya University, Science Faculty, Department of Mathematics, Edirne, Turkey*
[3]*Necmettin Erbakan University, Department of Mathematics-Computer Sciences, Konya,Turkey*

In this paper, properties of α-cosymplectic manifolds equipped with $M-$ projective curvature tensor are studied. First, we gave the basic definitions and curvature properties of $\alpha-$cosymplectic manifolds, then, we gave the definitions of Weyl projective curvature tensor W, con-circular curvature tensor C and conformal curvature tensor V and we obtain some relations between these curvature tensors of a Riemannian manifold. Also we proved that an $2n+1-$dimensional α-cosymplectic manifold M^{2n+1} is $M-$ projectively flat if and only if it is either locally isometric to the hyperbolic space $H_n(-1)$. And finally, we proved that the $M-$ projective curvature tensor in an $\alpha-$cosymplectic manifold M^{2n+1} is irrotational if and only if it is locally isometric to the hyperbolic space $H_n(-\alpha^2)$.

Keywords: *curvature tensor, manifold, cosymplectic manifold, Riemannian manifold*

References

[1] Ghosh A. , Koufogiorgos T. and Sharma R. (2001), *Conformally flat contact metric manifolds*, the country for pressing, J. Geom. , 70, 66-76.

[2] Chaubey S.K. and Ojha R.H. (2010), *On the m-projective curvature tensor of a Kenmotsu manifold, Differential Geometry - Dynamical Systems*, Geometry Balkan Press, 12, 2-60.

[3] Boothby M. and Wong R.C. (1958), *On contact manifolds*, Ann. Math. 68, 421-450.

[4] Sasaki S. and Hatakeyama Y. (1961), *On differentiable manofolds with certain structures which are closely related to almost contact structure*, Tohoku Math. J. 13, 281-294.

[5] Zengin F.Ö. (2012), *M-projectively flat Spacetimes*, Math. Reports 14(64), 4, 363-370.

10TH ISC2017 INTERNATIONAL STATISTICS CONGRESS
December 6-8, 2017 ANKARA/TURKEY

Comparisons of Some Importance Measures

Ahmet DEMİRALP[1], M. Şamil ŞIK[1]
ahmt.dmrlp@gmail.com, mhmd.sml85@gmail.com

[1] *Inonu University, Malatya, Turkey*

One of the system's efficiency measures is its survival probability as time goes by so called system reliability. In terms of system reliability some components are more important than the other components for the systems. Thus, several methods have been developed to measure the importance of components that affect system reliability. The importance measures are also used to rank the components in order to ensure that the system works efficiently or to improve its performance or design. The first method is Birnbaum reliability importance. Birnbaum Importance Measure of a component is independent of the reliability of the component itself. BIM is the rate of increase of the system reliability with respect to increase of the component reliability. Some of the other importance measures whose common properties are derived from Birnbaum are Structural Importance Measure, Bayesian Reliability Importance and Barlow-Proschan Importance. We obtained results for Birnbaum, Structural, Bayesian and Barlow-Proschan Importance from three different simulations with 100, 1000, 10000 repetition made for two different coherent systems.We observed that the components connected in serial with the system have the highest importance for the examined systems.

Keywords: *Birnbaum reliability importance, Structural importance, Bayesian reliability importance, Barlow-Proschan Importance.*

References

[1] Kuo, W. and Zuo, M. J. (2003), *Optimal Reliability Modeling: Principles and Applications*, USA, John Wiley & Sons.

[2] Kuo, W. and Zhu, X., (2012), *Importance Measures in Reliability, Risk and Optimization:Principles and Applications*, USA, John Wiley & Sons.

[3] Birnbaum, Z. W. (1969), *On the importance of different components in a multicomponent system, In Multivariate Analysis*, New York, Vol. 2, Academic Press.

ISC2017 INTERNATIONAL STATISTICS CONGRESS
December 6-8, 2017 ANKARA/TURKEY

Determining the Importance of Wind Turbine Components

M. Şamil ŞIK[1], Ahmet DEMİRALP[1]
mhmd.sml85@gmail.com, ahmt.dmrlp@gmail.com

[1]*Inonu University, Malatya, Turkey*

System analysts have been defined and derived various importance measures to determine the importance of a component in an engineered system. Wind turbines are widely preferred in recent years in the field of renewable energy due to their limited negative effect on the environment besides their high applicability in many terrains. In this study we aim to reduce maintenance and repair costs while improving performance of wind turbines in the structural design phase. The most known and used importance measure is Birnbaum component importance which is also defined as Marginal Reliability Importance (MRI). Derived from MRI, Joint Reliability Importance (JRI) measures two or more components contribution to the system reliability in a system. In this work we have obtained numerical results for 112 subsets of wind turbine components JRIs excluding null set, one component subsets, six components subsets and seven components subset. We calculated JRIs of some subsets of the wind turbine components by assuming all components with same p values and compared the results for $p = 0.1,0.2,0.3,0.4,0.5,0.6,0.7,0.8,0.9$ values. The reliability importance measure for a wind turbine improves as joint reliability of its components improves. This information extends understanding definition of the relevant components of a wind turbine system to improve its design. The numerical results show that {rotor, brake system, generator, yaw system, blade tip hydraulic} is the best components subset.

Keywords: *Wind Turbine, Joint Reliability Importance, System Reliability, Structural Design*

References

[1] Wu, S. (2005), *Joint importance of multistate systems,* Computers and Industrial Engineering 49(1), pp. 63-75.

[2] S. T. Sunder, S. T. and Kesevan, R. (2011),*Computation of Reliability and Birnbaum Importance of Components of a Wind Turbine at High Uncertain Wind*, International Journal of Computer App lications (0975 – 8887) Vol. 32– No.4.

[3] Kuo, W. and Zuo, M. (2003), *Optimal Reliability Modeling: Principles and Applications*, New Jersey, John Wiley&Sons, Inc., pp. 85-95.

[4] Gao, X., Cui, L. and Li, J. (2007), *Analysis for joint importance of components in a coherent system*, European Journal of Operational Research 182, pp. 282–299.

SESSION VI

APPLIED STATISTICS IX

10TH ISC2017 INTERNATIONAL STATISTICS CONGRESS
December 6-8, 2017 ANKARA/TURKEY

PLSR and PCR under Multicollinearity

Hatice ŞAMKAR[1] Gamze GÜVEN[1]
hfidan@ogu.edu.tr, gamzeguven@ogu.edu.tr

[1]*Eskisehir Osmangazi University, Eskisehir, Turkey*

The Least Squares (LS) estimator does not have minimum variance and may give poor results caused by multicollinearity problem [3]. Biased estimation techniques and dimension reduction techniques can be used to overcome this problem [5]. In literature, two of the most popular dimension reduction techniques are Partial Least Squares Regression (PLSR) and Principal Component Regression (PCR). These techniques construct new latent variables or components, which are linear combinations of available independent variables [4]. PCR and PLSR are based on a bilinear model that explains the existence of a relation between a set of p-dimensional independent variables and a set of q-dimensional response variables through k-dimensional scores t_i. with $k >> p$. The main difference between PCR and PLSR lies in the construction of the scores t_i. In PCR the scores are obtained by extracting the most relevant information present in the x-variables by performing a principal component analysis on the predictor variables and thus using a variance criterion. No information concerning the response variables is yet taken into account. In contrast, the PLSR scores are calculated by maximizing a covariance criterion between the x- and y-variables [1]. In this study, the mathematical models of PLSR and PCR were given and the properties of the techniques were briefly mentioned. In addition, a simulation study was conducted to compare predictive performances of PLSR and PCR techniques. For this aim, the optimal number of components and latent variables for PCR and PLSR, respectively, were considered. In the simulation study, correlated data was generated using the formula given in McDonald and Galarneau [2]. Besides different degrees of correlation, different numbers of variables and different numbers of observations were used in the simulation study. From the results of simulation study, it can be stated that generally PLSR become superior to PCR.

Keywords: *Multicollinearity, PLSR, PCR, dimension reduction techniques*

References

[1] Engelen, S., Hubert, M., Branden, K.V. and Verboven, S. (2004), *Robust PCR and robust PLSR: A comparative study.* In Theory and applications of recent robust methods, 105-117., Birkhäuser, Basel.

[2] McDonald, G.C. and Galarneau, D.I. (1975), *A Monte Carlo evaluation of some ridge-type estimators*, Journal of American Statistical Association, 70(350),407-416.

[3] Naes, T. and Martens, H. (1985), *Comparison of prediction methods for multicollinear data*, Communications in Statistics, 14(3), 545-576.

[4] Naik, P. and Tsai C.L. (2000), *Partial least squares estimator for single-index models*, Journal of the Royal Statistical Society: Series B, 62(4), 763-771.

[5] Rawlings, J.O., Pantula, S.G. and Dickey, D.A. (1998) *Applied regression analysis: a research tool*, Springer, New York.

ISC2017 INTERNATIONAL STATISTICS CONGRESS
December 6-8, 2017 ANKARA/TURKEY

On the Testing Homogeneity of Inverse Gaussian Scale Parameters

Gamze GÜVEN[1], Esra GÖKPINAR[2] , Fikri GÖKPINAR[2]
gamzeguven@ogu.edu.treyigit@gazi.edu.tr, fikti@gazi.edu tr

[1]*Eskisehir Osmangazi University, Ankara, Turkey*
[2]*Gazi University, Ankara, Turkey*

The Inverse Gaussian (IG) distribution is commonly used to model positive skewed data and it can accommodate a variety of shapes, from highly- skewed to almost normal. It is also noteworthy that the IG distribution is used in many applied sciences such as cardiology, finance, life tests. For applications to applied sciences and the comprehensive statistical properties of IG see refs. [1-5]. In practice, it is important to test equality of IG means. The classical method is applied under the assumption of homogeneity of the scale parameters. In the real world, this kind of assumption may or may not be true. One needs to check the validity of this assumption before applying the classical method. Furthermore, it can be said that it is very common problem in applied statistics to compare variances of several populations. The chief goal of this paper is to obtain a new test for the homogeneity of *k* IG scale parameters (λ's) and compare with the existing tests. For this reason, the hypotheses of interest:

$$H_0: \lambda_1 = \lambda_2 = \cdots = \lambda_k \text{ vs } H_1: at\ least\ one\ \lambda_i\ is\ different$$

The proposed test is based on simulation and numerical computations and uses the maximum likelihood estimates (MLEs) and restricted maximum likelihood estimates (RMLEs). In addition, it does not require the knowledge of any sampling distribution. In this paper we compare this test with the existing tests in terms of type I errors and powers using Monte Carlo simulation. Type I error rates and powers of the proposed test were computed based on 5,000 Monte Carlo runs for different values of the scale parameter λ, sample size *n*, and a number of groups *k*. For the range of parameters studied, the proposed test is very close to the nominal value of the significance level. Also, for all situations, the powers of the proposed performs well than the others, especially in cases of the sample sizes.

Keywords: *parametric bootstrap, computational approach test, Inverse Gaussian distribution.*

References

[1] Bardsley, W. E. (1980), *Note on the use of the Inverse Gaussian distribution for wind energy applications,* Journal of Applied Meteorology, 19, 1126-1130.

[2] Folks, J. L., Chhikara, R. S. (1978), *The Inverse Gaussian distribution and its statistical application-a review,* Journal of the Royal Statistical Society Series B (Methodological), 263-289.

[3] Seshadri, V. (1999), *The Inverse Gaussian distribution: statistical theory and applications*, Springer, New York.

[4] Takagi, K., Kumagai, S., Matsunaga, I., Kusaka, Y. (1997), *Application of Inverse Gaussian distribution to occupational exposure data,* The Annals of Occupational Hygiene, 41, 505-514.

[5] Tweedie, MC (1957), *Statistical Properties of Inverse Gaussian Distributions. I,* The Annals of Mathematical Statistics, 362-377.

10TH ISC2017 INTERNATIONAL STATISTICS CONGRESS
December 6-8, 2017 ANKARA/TURKEY

On an approach to ratio-dependent predator-prey system

Mustafa EKİCİ[1],Osman PALANCI[2]
mustafa.ekici@usak.edu.tr, osmanpalanci@sdu.edu.tr

[1]Usak University Faculty of Education Mathematics and Science Education, Usak, Turkey [2]Suleyman Demirel University Faculty of Economics and Administrative Sciences, Isparta, Turkey

The ratio-dependent predator-prey system is the main objective of the model, which is mutually resumed in two generations. In terms of view of human needs, the exploitation of biological resources and harvesting of populations are commonly practiced in forestry, fishery, and wildlife management. There is a wide range of interest in the use of bio-economic models to gain insight into the scientific management of renewable resources like fisheries and forestries concerning the optimal management of renewable resources. This paper presents an algorithm based on an improved differential transform method which is developed to approximate the solution of the ratio-dependent predator–prey system with constant effort harvesting. The divergence of the series is also eliminated by using the Padé approximation technique with this method. Some plots of the population of predator- prey versus time are presented to illustrate the performance and the accuracy of the method. The improved diferential transform method has the advantage of being more concise for numerical purposes. The advantage of method avoids the difficulties and massive computational work that usually arise from the parallel techniques and finite-difference method.

Keywords: *Differential transform method, predator-prey system, improved differential transform method, Padé approximation*

References

[1] Ekici, M. (2016). *Lineer Olmayan Bazı Matematiksel Modeller İçin Bir Yöntem*, Gazi University, 70-75.

[2] Tanner J. , T. (1975). *The Stability and The Intrinsic Growth Rates of Prey and Predator Populations*, Ecology, 56, 855-867.

[3] Berryman A. ,A. (1992). *The Origins and Evolution of Predator-Prey Theory*, Ecology, 73(5), 1530-1535

[4] Makinde O. , D. (2007). *Solving Ratio-Dependent Predator-Prey System With Constant Effort Harvesting Using Adomian Decomposition Method*, Applied Mathematics and Computation, 186, 17-22.

10TH ISC2017 INTERNATIONAL STATISTICS CONGRESS
December 6-8, 2017 ANKARA/TURKEY

Analysis of Transition Probabilities Between Parties of Voter Preferences with the Ecological Regression Method

Berrin GÜLTAY[1], Selahattin KAÇIRANLAR[2]
berringultay@comu.edu.tr, skacir@cu.edu.tr

[1]*Canakkale Onsekiz Mart University, Faculty of Art and Sciences, Department of Statistics, Canakkale, Turkey*
[2]*Cukurova University, Faculty of Art and Sciences, Department of Statistics, Adana, Turkey*

The ecological regression method is very useful in the analysis of election data aggregated concerning voters who voted for the same party as a result of two consecutive elections, in other words, voters who changed the party preference [1]. The aggregate electoral data for two consecutive elections can be expressed for two variables, X ; *party voted for the first election*, Y ; *party voted for the second election*. Expressed in multivariate multiple regression terminology, the explanatory variables are the proportions of the votes obtained in the first election, x_{ih} , for the part *i* and voting district *h*. As response variables, we use the proportions of votes obtained for party *j* in voting district *h* in the second election, y_{jh} . The system of *q* regression equations with *p* explanatory variables in each is of the form

$$\begin{aligned} y_{1h} &= \beta_{11}x_{1h} + \ldots + \beta_{p1}x_{ph} + e_{1h} \\ y_{2h} &= \beta_{12}x_{1h} + \ldots + \beta_{p2}x_{ph} + e_{2h} \\ &\vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \\ y_{qh} &= \beta_{1q}x_{1h} + \ldots + \beta_{pq}x_{ph} + e_{qh} . \end{aligned} \tag{1}$$

The parameter values β_{ij} are expected to be within the acceptable (0,1) range. Given information from *n* electoral districts, we can write the system of equations in matrix language as a multivariate linear regression model;

$$Y = XB + \varepsilon . \tag{2}$$

When the proportions are not stable enough, the estimates of transition parameters β using ordinary least squares (*OLS*) estimation might fall outside the acceptable range (0,1). Even though the equations in model (2) appear to be structurally unrelated, the fact that the disturbances are correlated across equations constitutes a ling among them. Such a behavior is reflected in a form

$$y = Z\beta + e \tag{3}$$

which is called *Seemingly Unrelated Regression Equations (SURE)* model considered by [3]. The aim of this study is to estimate the probabilities of the vote transitions in two consecutive special elections held on June 7 and November 1, 2015, using the *restricted modified generalized ridge estimator* which is used to estimate the Swedish elections (1988-1991) by [2].

Keywords: *Ecological Regression, Transitions probabilities, Shrinkage estimators, SURE Model*

References

[1] Gültay, B. (2009), *Multicollinearity and Ecological Regression*, MSc. Thesis, Cukurova University, Institute of Natural and Applied Sciences, University of Cukurova, Adana, 89.

[2] Fule, E. (1994), *Estimating Voter Transitions by Ecological Regression*, Electoral Studies, 13(4), 313-330.

[3] Zellner, A.(1962), *An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias*, Journal of the American Statistical Association, 57, 348-368.

10TH ISC2017 INTERNATIONAL STATISTICS CONGRESS
December 6-8, 2017 ANKARA/TURKEY

Variable Neighborhood – Simulated Annealing Algorithm for Single Machine Total Weighted Tardiness Problem

Sena AYDOĞAN[1]
senaaydogan@gazi.edu.tr

[1]*Gazi University Department of Industrial Engineering, Ankara, Turkey*

Scheduling problem is one of the important problems in the production area. Effective scheduling has become a necessity to survive in the modern competitive environment. Therefore, compliance with deadlines and avoidance of delay penalties are common goals of scheduling. The purpose of Single Machine Total Weighted Tardiness (SMTWT) Problem is to find the job sequence with the smallest total tardiness. It has been proved that the SMTWT problem is an NP-hard problem in terms of computational complexity. Exact methods such as dynamic programming and branch & bound algorithm are inadequate for solving the problem, especially when the number of jobs is over 50. For this reason, meta-heuristic methods have been developed to obtain near-optimal results at reasonable times. In this study, a variable neighborhood simulated annealing (V-SA) algorithm has been developed which can yield effective results for the SMTWT problem. The simulated annealing (SA) algorithm has been also developed and tested comparatively in different problem sizes. When the results are evaluated, it is seen that both algorithms give effective results in small, medium and large sized problems, but the V-SA algorithm which works to improve the solution with different neighborhood structures in terms of solution time worked at higher computation times as expected. Therefore, it is recommended that the V-SA algorithm can be preferred in cases where the solution quality is more important than the solution time, while the SA algorithm is preferred in cases where the solution time is more important than the solution quality.

Keywords: *Single machine total weighted tardiness problem, simulated annealing, variable neighborhood algorithm*

References

[1] Kirkpatrick, S. (1984). *Optimization by simulated annealing: Quantitative studies.* Journal of statistical physics, 34(5-6), 975-986.

[2] Lawler, E. L. (1964). *On scheduling problems with deferral costs*. Management Science, 11(2), 280-288.

[3] Mladenović, N. and Hansen, P. (1997). *Variable neighborhood search.* Computers & Operations Research, 24(11), 1097-1100.

POSTER PRESENTATION SESSIONS

10TH ISC2017 INTERNATIONAL STATISTICS CONGRESS
December 6-8, 2017 ANKARA/TURKEY

The Application of Zero Inflated Regression Models with the Number of Complaints in Service Sector

Aslı Gizem KARACA[1], Hülya OLMUŞ[1]
asligizem@gmail.com, hulya@gazi.edu.tr

[1]*Gazi University, Ankara, Turkey*

Count data are frequently used in biostatistics, econometrics, demography, educational sciences, sociology and actuarial sciences. These count data are frequently characterized by overdispersion and excess zeros. The distribution of the data set is skewed to the right when zero values are inflated and this does not lead to the assumption of the normal distribution required for the linear regression method. Applying conversion methods for zero values obtained in such cases, or ignoring zero values, results in biased and inefficient. Poisson Regression, Negative Binomial Regression, Zero Inflated Poisson Regression, and Zero Inflated Negative Binomial regression models are used in the model of counting data that has extreme zero and/or overdispersion. In this study, it is considered that gender, age, education and experience variables effect the number of complaints received from customers who work for any service sector. This count data was analyzed to evaluate zero inflated models using R program. In addition, the Akaike Information Criteria were used to evaluate regression models. In practice, it is determined which model is suitable for the last six months of 2016 (between July-December) and related to parameter estimates of these models comments were made. As a result, it has been determined that the Zero Inflated Poisson and Zero Inflated Negative Binomial Regression models are appropriate in the high zero inflated months; Poisson and Negative Binomial Regression models have been found to be more appropriate models for describing the data set in the months when the zero inflation is less.

Keywords: *count data, excess zeros, zero-inflated data, zero-inflated regression models*

References

[1] Akinpelu, K.,Yusuf B., Akpa M. and Gbolahan O. (2016), *Zero Inflated Regression Models with Application to Malaria Surveillance Data,* International Journal of Statistics and Applications, 6(4), 223-234.

[2] Hu M., Pavlicova M. and Nunes E. (2011), *Zero Inflated and Hurdle Models of Count Data with Extra Zeros: Examples from an HIV-Risk Reduction Intervention Trial*, American Journal of Drug & Alcohol Abuse, 37(5), 367-375.

[3] Kaya Y. and Yeşilova A. (2012), *E-Posta Trafiğinin Sıfır Değer Ağırlıklı Regresyon Yöntemleri Kullanılarak İncelenmesi*, Anadolu Üniversitesi Bilim ve Teknoloji Dergisi, 13(1), 51-63.

[4] Lambert, D. (1992), *Zero Inflated Poisson Regression, with an Application to Defects in Manufacturing*, Technometrics, 34(1), 1-14.

[5] Peng J. (2013), *Count Data Models for Injury Data from the National Health Interview Survey (NHIS)*, The Ohio State University, 60.

ISC2017 INTERNATIONAL STATISTICS CONGRESS
December 6-8, 2017 ANKARA/TURKEY

Burnout and Life Satisfaction of University Students

Kamile ŞANLI KULA[1], Ezgi ÇAĞATAY İN[1]
sanli2004@hotmail.com, ezgicagatayin@gmail.com

[1]*Ahi Evran University, KIRŞEHİR, TÜRKİYE*

The aim of this study is to determine whether burnout and life satisfaction of students who study at different faculties and junior college of Ahi Evran University differs according to variable of gender, date of birth, class, smoking, participation in social activities and weekly course schedule.

The population of this study is composed of all 3780 students who attended the 1st and 4th grades at different faculties/junior colleges in Ahi Evran University during the fall semester of 2016-2017.

In the study, it has been reached that girls were more burnout in the exhaustion and competence sub-dimension than the boys, whereas in the depersonalization sub-dimension, the boys were more burnout and the girls were higher in life satisfaction than the boys. It was determined that there was no difference in life satisfaction but there was difference burnout between the exhaustion, depersonalization sub-dimension of competence score statistically according to the date of birth. It was observed that there was a statistically significant difference burnout and depersonalization subscale scores of the students according to the grades and the life satisfaction of the first grade students was higher than the fourth grade students. It was determined that students who smoke had high burnout and low life satisfaction. It has been that students who participated in the social activities were exhausted in the burnout and depersonalization sub-dimension, in the dimension of competence, those who did not participate in the activities were more exhausted and the life satisfaction of the students participated in social activities was higher.

Keywords: *Burnout, life satisfaction, university students.*

This work was supported by the Scientific Research Projects Council of Ahi Evran University, Kırşehir, Turkey under Grant FEF.A3.16.036.

References

[1] Çapri, B., Gündüz, B., Gökçakan, Z. (2011). *Maslach Tükenmişlik Envanteri-Öğrenci Formu'nun (MTE-ÖF) Türkçe'ye Uyarlaması: Geçerlik ve Güvenirlik Çalışması,* Çukurova Üniversitesi Eğitim Fakültesi Dergisi, 01(40), 134-147.

[2] Diener, E., Emmons,R. A., Larsen, R. J. and Griffin, S. (1985), *The satisfaction with Life Scale*, Journal of Personality Assessment, 49(1), 71-75.

[3] Köker, S. (1991). *Normal ve Sorunlu Ergenlerin Yaşam Doyumu Düzeylerinin Karşılaştırılması*, Yayımlanmamış yüksek lisans tezi, Ankara Üniversitesi Sosyal Bilimler Enstitüsü, Ankara.

[4] Maslach, C., Schaufeli, W. B., and Leiter, M. P. (2001), *Job Burnout,* Annual Reviews of Psychology, 52, 397-422.

ISC2017 INTERNATIONAL STATISTICS CONGRESS
December 6-8, 2017 ANKARA/TURKEY

Examination of Job Satisfaction of Nurses

Kamile ŞANLI KULA[1], Mehmet YETİŞ[1], Aysu YETİŞ[2], Emrah GÜRLEK[1]
sanli2004@hotmail.com, drmehmetyetis@hotmail.com, draysuyetis@hotmail.com, e.gurlek@gmail.com

[1]*Ahi Evran University, Kırşehir, TÜRKİYE*
[2] *Ahi Evran University Education and Research Hospital, Kırşehir, TÜRKİYE*

In this study, Job satisfaction of nurses working at Ahi Evran University Training and Research Hospital will be examined in terms of various variables. The study will be conducted with the nurses working in Ahi Evran University Education and Research Hospital, volunteers who agree to participate in the research. For this purpose, Personal Information form developed by researchers as a data collection tool, Mineseto Job Satisfaction Scale will be used.

As a result of the research, it was determined that the average level of job satisfaction of the nurses was moderate. There was no difference between job satisfaction averages according to whether or not they made the choice of profession themselves. It was determined that there was a difference between the averages according to the idea of abandonment and that this difference was from all groups. According to wage satisfaction, those who are satisfied with their wage have higher external and general satisfaction averages. Job satisfaction of nurses who are satisfied with the working environment is higher in all dimensions. It has been achieved that nurses who enjoy doing his/her business have higher internal, external and general satisfactions.

Keywords: *Nurse, Job Satisfaction.*

This work was supported by the Scientific Research Projects Council of Ahi Evran University, Kırşehir, Turkey under Grant TIP.A3.17.005.

References

[1] Aras, A. (2014), *To research the job satisfaction and burnout and influential factors of doctors in primary health system in Erzurum*, Atatürk University, Medical School, Public Health, Erzurum.

[2] Çelebi, B. (2014), *Workers' burnout and job satisfaction: Alanya state hospital nurses sample*, Unpublished Master's Thesis, Beykent University Social Sciences Institute, Istanbul.

[3] Kurçer, M.A. (2005), *Job satisfaction and burnout levels of physicians working Harran University Faculty of Medicine in Şanlıurfa*, Harran Üniveritesi Tıp Fakültesi Dergisi, 2(3), 10-15.

[4] Sünter, A.T., Canbaz, S., Dabak, Ş., Öz, H., and Pekşen, Y. (2006), *The level of burnout, work-related strain and work satisfaction in general practitioners*, Genel Tıp Derg, 16(1), 9-14.

[5] Ünal, S., Karlıdağ, R., and Yoloğlu, S. (2001). *Relationships between burnout, job satisfaction and life satisfaction in physicians*, J. Clin Psy., 4(2) , 113-118.

10TH ISC2017 INTERNATIONAL STATISTICS CONGRESS
December 6-8, 2017 ANKARA/TURKEY

A Comparative Study for Fuzzification of the Replicated Response Measures: Standard Mean vs. Robust Median

Özlem TÜRKŞEN[1]
turksen@ankara.edu.tr

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

Classical regression analysis is a well-known probabilistic modelling tool in many researches. However, in some of the cases, the classical regression analysis cannot be proper to use, e.g. small sized data sets, unsatisfied probabilistic modelling assumptions, imprecision between the variables, uncertainty about the variables different than randomness. One of the example for the uncertainty on the response variable case is replicated response measured data set. In the replicated response measured data set, the response values cannot be identified exactly because of the uncertainty on the replications. In this case, fuzzy regression analysis can be considered as a modelling tool. In order to apply fuzzy regression, based on fuzzy least squares approach, it is needed to represent replicated measures as fuzzy numbers which is called fuzzification of the replicated measures. In this study, the replicated measures are presented as triangular type-1 fuzzy numbers (TT1FNs). Fuzzification is achieved according to the structure of replications from statistical perspective. For this purpose, mean and median are used to identify the center of TT1FN. The spreads from the center values are defined by using standard deviation and absolute deviation metrics which are calculated around the mean and the median statistics, respectively. A real data set from the literature is chosen to apply suggested robust fuzzification approach. It is seen from the fuzzy regression modelling results that median and median absolute deviation (MAD) should be preferred for fuzzification of the replicated response measures according to the root mean square error (RMSE) criteria.

Keywords: *Replicated response measured data set, triangular type-1 fuzzy numbers, fuzzy regression analysis, robust statistics.*

References

[1] Gladysz, B. and Kasperski, A. (2010), *Computing mean absolute deviation under uncertainty*, Applied Soft Computing, 10, 361-366.

[2] Leys, C., Ley, C., Klein, O., Bernard, P. and Licata, L. (2013), *Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median*, Journal of Experimental Social Psychology, 49, 764-766.

[3] Olive, D.J. (1998), *Applied Robust Statistics*, University of Minnesota, 517 pp.

[4] Rousseeuw, P.J. and Hubert, M. (2011), *Robust statistics for outlier detection*, WIREs Data Mining and Knowledge Discovery, 1, 73-79.

[5] Türkşen, Ö. and Güler, N. (2015), *Comparison of Fuzzy Logic Based Models for the Multi-Response Surface Problems with Replicated Response Measures*, Applied Soft Computing, 37, 887-896.

10TH ISC2017 INTERNATIONAL STATISTICS CONGRESS
December 6-8, 2017 ANKARA/TURKEY

Asymmetric Confidence Interval with Box-Cox Transformation in R

Osman DAĞ[1], Özlem İLK[2]
osman.dag@hacettepe.edu.tr, oilk@metu.edu.tr

[1] *Hacettepe University Department of Biostatistics, Ankara, Turkey*
[2] *Middle East Technical University Department of Statistics, Ankara, Turkey*

Normal distribution is important in statistical literature since most of the statistical methods are based on normal distribution such as t-test, analysis of variance and regression analysis. However, it is difficult to satisfy the normality assumption for real life datasets. Box–Cox power transformation is the most well-known and commonly utilized remedy [2]. The algorithm relies on a single transformation parameter. In the original article [2], maximum likelihood estimation was proposed for the estimation of transformation parameter. There are other algorithms to obtain transformation parameter. Some of them include the studies of [1], [3] and [4]. Box–Cox power transformation is given by

$$y_i^T = \begin{cases} \frac{y_i^\lambda - 1}{\lambda}, & if \quad \lambda \neq 0 \\ log\, y_i\,, & if \quad \lambda = 0 \end{cases}.$$

Here, λ is the power transformation parameter to be estimated, y_i's are the observed data, y_i^T's are transformed data.

In this study, we focus on obtaining the mean of data and a confidence interval for it when Box-Cox transformation is applied. Since the transformation is applied, the scale of the data has changed. Therefore, reporting the mean and confidence interval obtained from transformed data is not meaningful for the researchers. Besides, reporting mean and symmetric confidence interval obtained from original data becomes misleading for the researchers since the normality assumption is not satisfied. Therefore, it is pointed out that mean and asymmetric confidence interval obtained from back transformed data must be reported. We have written down a generic function to obtain the mean of data and a confidence interval for it when Box-Cox transformation is applied. It is released under R package **AID** with the name of "*confInt*" for implementation.

Keywords: *transformation, R package, asymmetric confidence interval*

References

[1] Asar, O., Ilk, O. and Dag, O. (2017), *Estimating Box-Cox power transformation parameter via goodness-of-fit tests*, Communications in Statistics - Simulation and Computation, 46(1), 91–105.

[2] Box, G. E. P. and Cox, D. R. (1964), *An analysis of transformations (with discussion)*, Journal of Royal Statistical Society Series B (Methodological), 26(2), 211–252.

[3] Rahman, M. (1999), *Estimating the Box-Cox transformation via Shapiro-Wilk W statistic*, Communications in Statistics–Simulation and Computation, 28(1), 223–241.

[4] Rahman, M. and Pearson, L. M. (2008), *Anderson-Darling statistic in estimating the Box-Cox transformation parameter*, Journal of Applied Probability and Statistics, 3(1), 45–57.

ISC2017 INTERNATIONAL STATISTICS CONGRESS
December 6-8, 2017 ANKARA/TURKEY

Visualizing Trends and Patterns in Cancer Mortality Among Cities of Turkey, 2009-2016

Ebru OZTURK[1], Duygu AYDIN HAKLI[1], Merve BASOL[1], Ergun KARAAGAOGLU[1]
ebru.ozturk3@hacetttepe.edu.tr

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

Cancer is the second leading cause of death in the Turkey (TURKSTAT, 2016) and world (GDB,2015). Moreover, mortality rate with respect to cancer has increased in Turkey over years (GDB, 2015). In this study, we focus on geographic differences in cancer mortality among cities of Turkey. The data at city level are significant and valuable since the public health policies planned and applied at the local level (Mokdad et al., 2017). Besides, local information might give benefits to health care professionals to understand the needs of community care and determining cancer hot spots.

According to Chamber et al. (1983) "There is no single statistical tool that is as powerful as a well-chosen graph". Therefore, we present cancer mortality by using statistical maps that are a method to represent geographic distribution of the data. In this study, we show statistical maps of cancer mortality based on gender between 2009 to 2016. In addition to these maps, we touch linked micromap which provide users to link statistical information to a series of small maps. We aim to show trends and patterns in cancer mortality among cities of Turkey by using these maps. We provide researchers and readers with an understanding of the distribution of cancer mortality that varies over the years. Moreover, we use R project in particular during this study to demonstrate drawing of statistical maps by using such free software. The data set on causes of death regards to usual residence (TURKSTAT, 2016) is provided by Turkish Statistical Institute (TURKSTAT).

Keywords: *cancer mortality, statistical maps, linked micromaps*

References

[1] Chambers, J. M., Cleveland, W. S., Kleiner, B., and Tukey, P. A. (1983*), Graphical Methods for Data Analysis*, London, UK: Chapman & Hall/CRC, 1.

[2] GBD 2015 Mortality and Causes of Death Collaborators. *Global, regional, and national life expectancy, all-cause mortality, and cause-specific mortality for 249 causes of death, 1980-2015: a systematic analysis for the Global Burden of Disease Study 2015*. Lancet. 2016;388(10053): 1459-1544.

[3] Mokdad AH, Dwyer-Lindgren L, Fitzmaurice C, et al: *Trends and patterns of disparities in cancer mortality among US counties*, 1980-2014. JAMA 317:388-406, 2017.

[4] TURKSTAT. (2017). Retrieved October 2017, Distribution of selected causes of death by usual residence with respect to gender, 2009-2016.

10TH ISC2017 INTERNATIONAL STATISTICS CONGRESS
December 6-8, 2017 ANKARA/TURKEY

A Comparison of Confidence Interval Methods for Proportion

Merve BASOL[1], Ebru OZTURK[1], Duygu AYDIN HAKLI[1], Ergun KARAAGAOGLU[1]
merve.basol@hacetttepe.edu.tr

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

Hypothesis tests and point/interval estimates for a population parameter are important parts of applied statistics when summarizing data. Although hypothesis tests have been reported using only p-values in most studies, it is suggested that hypothesis tests should be interpreted using both p-values and confidence intervals [2]. For a proportion, when sample size is large enough, one may estimate two-sided confidence intervals using traditional large sample theory, i.e Wald confidence interval as given $\hat{p} \pm z_{1-\alpha/2}\sqrt{\hat{p}(1-\hat{p})/n}$. However, two important problems arise from Wald confidence interval when sample size is small or proportion estimates are very close to 0 or 1; (i) the intervals that do not make sense, i.e. degenerate and (ii) the coverage probability that is quite different than nominal value $1-\alpha$. Hence, it is preferred to use alternative methods for estimating confidence intervals of population proportion in such cases [1,3]. In this study, we aimed to compare the performance of several confidence interval methods in terms of coverage probability and interval width under different conditions. The compared methods are simple asymptotic (Wald) with and without continuity correction, Wilson score with and without continuity correction, Clopper- Pearson (‘exact’ binomial), mid-p binomial tail areas, Agresti-Coull and bootstrap confidence method. For this purpose, we conducted a comprehensive simulation study which includes all the combinations of sample sizes (20, 50, 100 and 500) and population proportions (0.05, 0.10, 0.30 and 0.50). For each combination, 2000 datasets are generated and confidence intervals are estimated from each method. The analysis was made by using R 3.3.3 program with “DescTools” and “PropCIs” packages.
According to the results, when the sample size is small and proportion estimates are very close to 0 or 1, Wald method without continuity correction gives lower coverage probability. Wald with continuity correction, on the other hand, gives increased coverage probability and interval width. Clopper-Pearson method was very conservative since it is an exact method. In order to achieve coverage probability near nominal level, mid-p value is suggested rather than Clopper-Pearson.

Keywords: *confidence interval, proportion, Wald, simulation*

References

[1] Agresti, A. and Coull, B.A.(1998). *Approximate is better than “exact” for interval estimation of binomial proportions.* The American Statistician. 52(2), 119 – 126.

[2] Gardner, M. J., & Altman, D. G. (1986). *Confidence intervals rather than P values: estimation rather than hypothesis testing*. Br Med J (Clin Res Ed), 292(6522), 746-750.

[3] Newcombe, R. G. (1998). *Two-sided confidence intervals for the single proportion: comparison of seven methods.* Statistics in Medicine, 17(8), 857-872.

10TH ISC2017 INTERNATIONAL STATISTICS CONGRESS
December 6-8, 2017 ANKARA/TURKEY

Determining Unnecessary Test Orders in Biochemistry Laboratories: A Case Study for Thyroid Hormone Tests

Yeşim AKBAŞ[1], Serkan AKBAŞ[1], Tolga BERBER[1]
yesimyeginoglu@ktu.edu.tr, serkanakbas@ktu.edu.tr, tolga.berber@fen.ktu.edu.tr

[1]*Department of Statistics and Computer Sciences, Karadeniz Technical University, Trabzon, TURKEY*

Biochemistry laboratories, which perform many tests every day, have become one of the most important departments of hospitals, since they provide evidence to ease the disease identification process with the help of the tests they performed. Hence, doctors have begun to order biochemistry tests more often to make final decisions about diseases. According to the Ministry of Health, most of these tests consist of false or unnecessary tests for various reasons. These test orders cause considerable financial loss to hospitals and cause loss of time in terms of both laboratories and patients. The significant increase of health-care costs caused by unnecessary test orders could be reduced by identification of the tests that do not contribute to diagnosis and treatment of diseases.

In this study, we have examined all biochemistry test orders made by Emergency Unit of Farabi Hospital of Karadeniz Technical University between the dates between 01 January 2015 and 02 October 2017. We used association analysis approach to find out the most frequent test order co-occurrences and to assess necessity of them.

In this study we focused on TSH, FreeT3 and FreeT4 tests which are used to evaluate activity of thyroid hormones, since we identified them as the most frequent test orders which have been requested together from Emergency Unit. Moreover, these three tests have a procedural guideline, which suggests that order of the tests should be TSH, FreeT4 and FreeT3, respectively. According to the guideline, FreeT4 and FreeT3 tests should be performed when the value of the TSH test is out of the reference interval. We found that the number of co-occurrences of the three tests are close to one (TSH:2029, FreeT4:1967 and FreeT3:1526) which indicates that almost every order of TSH test include FreeT3 and FreeT4. As a result, necessary actions are being taken by Hospital Administration to prevent unnecessary test order requests.

This work is supported by KTU Scientific Research Projects Unit under project number FBB-2016-5521.

Keywords: *Unnecessary Test Order Identification; Association Analysis, Thyroid Hormone Tests*

References

[1] Demir, S., Zorbozan, N., and Basak, E. (2016), "*Unnecessary repeated total cholesterol tests in biochemistry laboratory*", Biochem. Medica, pp. 77–81.

[2] Divinagracia, R. M., Harkin, T. J., Bonk, S., and Schluger, N. W. (1998), "*Screening by Specialists to Reduce Unnecessary Test Ordering in Patients Evaluated for Tuberculosis"*, Chest, vol. 114, pp. 681–684.

[3] Mahmood, S., Shahbaz, M., and Guergachi, A. (2014), "*Negative and positive association rules mining from text using frequent and infrequent itemsets"*, Scientific World Journal.

[4] Tsay, Y.-J. and Chiang, J.-Y. (2005), *"CBAR: an efficient method for mining association rules",* Knowledge-Based Syst., vol. 18, pp. 99–105.

[5] Tiroid çalışma grubu (2015), "*Tiroid Hastalıkları Tanı Ve Tedavi Kılavuzu"*, Ankara, Türkiye Endokrinoloji ve Metabolizma Derneği.

10TH ISC2017 INTERNATIONAL STATISTICS CONGRESS
December 6-8, 2017 ANKARA/TURKEY

Box-Cox Transformation for Linear Models via Goodness-of-Fit Tests in R

Osman DAĞ[1], Özlem İLK[2]
osman.dag@hacettepe.edu.tr, oilk@metu.edu.tr

[1] *Hacettepe University Department of Biostatistics, Ankara, Turkey*
[2] *Middle East Technical University Department of Statistics, Ankara, Turkey*

Application of linear models requires the normality of the response and residuals for inferences, such as for hypothesis tests. However, normal distribution does not emerge so often in real life datasets. Box–Cox power transformation is a commonly used methodology to transform the distribution of the data into a normal one [2]. This methodology makes use of a single transformation parameter, which can be estimated from data generally via maximum likelihood (ML) method or ordinary least squares (OLS) method [3]. An alternative estimation technique is the use of goodness-of-fit tests [1].

In this study, we focus on estimating Box-Cox transformation parameter via goodness of fit tests for its use in linear regression models. In this context, Box–Cox power transformation is given by

$$y_i^T = \begin{cases} \frac{y_i^\lambda - 1}{\lambda} = \beta_0 + \beta_1 x_{1i} + \cdots + \beta_k x_{ki} + \varepsilon_i, & if \quad \lambda \neq 0 \\ log\, y_i = \beta_0 + \beta_1 x_{1i} + \cdots + \beta_k x_{ki} + \varepsilon_i, & if \quad \lambda = 0 \end{cases}.$$

Here, λ is the power transformation parameter to be estimated, y_i's are the observed response for the i^{th} subject, y_i^T's are transformed response, and $x_{1i}, \ldots x_{ki}$'s are the observed independent variables in the linear regression model. We employ seven popular goodness-of-fit tests for normality, namely Shapiro–Wilk, Anderson–Darling, Cramer-von Mises, Pearson Chi-square, Shapiro-Francia, Lilliefors and Jarque–Bera tests, together with ML and OLS estimation methods. We have written down an R function to perform Box-Cox transformation for linear models and to provide graphical analysis of residuals after transformation. It is released under R package **AID** with the name of "*boxcoxlm*" for implementation. The usage of the method is illustrated on a real data application.

Keywords: *transformation, R package, linear models*

References

[1] Asar, O., Ilk, O. and Dag, O. (2017), *Estimating Box-Cox power transformation parameter via goodness-of-fit tests*, Communications in Statistics - Simulation and Computation, 46(1), 91–105.

[2] Box, G. E. P. and Cox, D. R. (1964), *An analysis of transformations (with discussion)*, Journal of Royal Statistical Society Series B (Methodological), 26(2), 211–252.

[3] Kutner, M. H., Nachtsheim, C., Neter, J., Li, W. (2005). *Applied Linear Statistical Models. (5th ed.).* New York: McGraw-Hill Irwin, 132-134.

December 6-8, 2017 ANKARA/TURKEY

Semi-Parametric Accelerated Failure Time Mixture Cure Model

Pınar KARA[1], Nihal ATA TUTKUN[1],Uğur KARABEY[2]
pinarkara88@gmail.com, nihalata@hacettepe.edu.tr, ukarabey@hacettepe.edu.tr
[1]*Hacettepe University, Department of Statistics, Ankara, Turkey –*
[21]*Hacettepe University, Department of Actuarial Sciences, Ankara, Turkey*

The classical survival models used in cancer studies are based on the assumption that every patient in the study will eventually experience the event of interest. This assumption may not be appropriate when there are lots of patients in the study who never experienced the event of interest during the follow-up period. However, with advances of medical treatments, patients can be cured of some diseases, and researchers are interested in assessing effects of a treatment or other covariates on the cure rate of the disease and on the failure time distribution of uncured patients [5]. Therefore using mixture cure model which is firstly introduced by Boag (1949) and Berkson and Gage (1952) gains importance. Mixture cure models take into account both the cured and uncured parts in the population. Cox mixture cure model and accelerated failure time mixture cure models are the types of mixture cure models. In this study, semi-parametric accelerated failure time mixture cure model which is developed by Li and Taylor (2002) and Zhang and Peng (2007) is examined. The model is applied to a stomach cancer data to show the advantages and differences in interpretation of the results according to the classical survival models. The cured proportions are obtained for different scenarios.

Keywords: *censoring, cure models, accelerated failure time*

References

[1] Boag J.W. (1949), *Maximum likelihood estimates of the proportion of patients cured by cancer therapy*, Journal of the Royal Statistical Society, 11(1), 15-44.

[2] Berkson J. and Gage R.P. (1952), *Survival curve for cancer patients following treatment*, Journal of the American Statistical Association, 47(259), 501-515.

[3] Li C-S and Taylor J.M.G. (2002), *A semi-parametric accelerated failure time cure model*, Statist. Med., 21(21):3235–3247.

[4] Zhang J. and Peng Y. (2007), *A new estimation method for the semiparametric accelerated failure time mixture cure model*, Statist. Med., 26(16), 3157–3171.

[5] Zhang, J., Peng, Y. (2012), *Semiparametric estimation methods for the accelerated failure Time mixture cure model*, J Korean Stat Soc., 41(3), 415–422.

10TH ISC2017 INTERNATIONAL STATISTICS CONGRESS
December 6-8, 2017 ANKARA/TURKEY

The Conceptual and Statistical Considerations of Contextual Factors

Çağla ŞAFAK[1], Derya GÖKMEN[1], Atilla Halil ELHAN[1]
caglasafak01@gmail.com, dgokmen2001@yahoo.com, ahelhan@yahoo.com

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

The purpose of this paper is to introduce the conceptual variables (moderating, mediating and confounding variables) and their effects on the statistical analyses with examples. Moderator variable is a qualitative / quantitative variable that affects the direction and/or strength of the relation between an independent and a dependent variable [1]. In general, a given variable may be said to function as a mediator to the extent that it accounts for the relation between the independent and dependent variable [1]. Confounding variables or confounders are often defined as the variables correlate (positively or negatively) with both the dependent and independent variable [2]. In studies which contain conceptual variables, after defining the type of them, the effect of these variables should be considered by appropriate statistical analyses [3]. For example, when studying with a confounding variable, the analysis of covariance should be performed in order to determine the independent variable differences in terms of dependent variable. The path analysis should be used to determine the mediating effect of the variables under consideration. This study will show different analysis strategies when the study contains contextual variables.

Keywords: *conceptual variables, moderating, mediating, confounding*

References

[1] Baron RM, Kenny DA (1986). *The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations*. J Pers Soc Psychol;51:1173 – 1182.

[2] Pourhoseingholi MA, Baghestani AR, Vahedi M.(2012) *How to control confounding effects by statistical analysis.* Gastroenterol Hepatol Bed Bench. 5(2): 79–83.

[3] Wang PP, Badley EM, Gignac M. (2006). *Exploring the role of contextual factors in disability model.* Disability and Rehabilitation; 28(2): 135-140.

10TH ISC2017 INTERNATIONAL STATISTICS CONGRESS
December 6-8, 2017 ANKARA/TURKEY

GAP (Groups, Algorithms and Programming) and Rewriting System for Some Group Constructions

Eylem GÜZEL KARPUZ[1], Merve ŞİMŞEK[1]
eylem.guzel@kmu.edu.tr, mervenursmsk@hotmail.com

[1]*Department of Mathematics, Karamanoğlu Mehmetbey University, Karaman, Turkey*

GAP is a system for computational discrete algebra, with particular emphasis on Computational Group Theory. GAP provides a programming language, a library of thousands of functions implementing algebraic algorithms written in the GAP language as well as large data libraries of algebraic objects. GAP is used in research and teaching for studying groups and their representations, rings, vector spaces, algebras, combinatorial structures, and more [1].

In this work, firstly, we give some information about GAP and its applications. Then we present complete rewriting system and normal form structures for some group constructions with monoid presentations, namely direct product of finite cyclic groups and extended Hecke groups ([3]) by using a GAP package "IdRel: A package for identities among relators" written by A. Heyworth and C. Wensley [2].

Keywords: *group, algorithm, rewriting system, normal form, Hecke group.*

References

[1] https://www.gap-system.org/index.html
[2] https://www.gap-system.org/Packages/idrel.html
[3] Karpuz, E. G., Çevik, A. S. (2012), *Gröbner-Shirshov bases for extended modular, extended Hecke and Picard groups*, Mathematical Notes, 92 (5), 636-642.

Graph Theory and Semi-Direct Product Graphs

Eylem GÜZEL KARPUZ[1], Hasibe ALTUNBAŞ[1], Ahmet S. ÇEVİK[2]
eylem.guzel@kmu.edu.tr, hasbii_07@outlook.com, ahmetsinancevik@gmail.com

[1]*Department of Mathematics, Karamanoğlu Mehmetbey University, Karaman, Turkey*
[2]*Department of Mathematics, Faculty of Science, Selcuk University, Konya, Turkey*

Graph theory is a branch of mathematics which studies the structure of graphs and networks. The subject of graph theory had its beginnings in recreational mathematic problems, but it has grown into a significant area of mathematical research with applications in chemistry, operations research, social sciences and computer science. This theory started in 1736, when Euler solved the problem known the Konigsberg bridges problem [1].

In this work, firstly, we give some information about graph theory and its some applications to other science areas. Then, by considering a new graph based on semi-direct product of a free abelian monoid of rank n by a finite cyclic monoid [2], we present some graph properties on this new graph, namely diameter, maximum and minimum degrees, girth, degree sequence and irregularity index, domination number, chromatic number, clique number.

Keywords: *Graph theory, semi-direct product, presentation.*

References

[1] Bondy, J. A., Murty, U. S. R. (1978), *Graph Theory with Applications*, Macmillan press Ltd.
[2] Karpuz, E. G., Das, K. C., Cangül, İ. N. and Çevik, A. S. (2013), *A new graph based on the semi-direct product of some monoids*, J. Inequalities Appl., 118.

10TH ISC2017 INTERNATIONAL STATISTICS CONGRESS
December 6-8, 2017 ANKARA/TURKEY

An Application of Parameter Estimation with Genetic Algorithm for Replicated Response Measured Nonlinear Data Set: Modified Michaelis-Menten Model

Fikret AKGÜN[1], Özlem TÜRKŞEN[2]
[1]*fikretakgun1990@hotmail.com*, [2]*turksen@ankara.edu.tr*

[1] *Ankara University, Graduate School of Natural and Applied Science, Statistics Department Ankara, Turkey*
[1] *Republic of Turkey Energy Market Regularity Authority, Ankara, Turkey*
[2]*Ankara University, Faculty of Science, Statistics Department, Ankara, Turkey*

Many of the real life problems need an appropriate mathematical model. It is well known that the selection of an appropriate mathematical model is one of the main challenges in modelling part of statistical analysis. Nonlinear regression models can be preferred to apply to the nonlinear data sets for modelling stage considering the fact that many of the problems have nonlinear structure. And also, the nonlinear data sets can be composed of replicated response measures. In this case, it is possible to apply common used parameter estimation approach, minimization of the sum of square errors, for parameter estimation procedure. However, the minimization of the error function with derivative based optimization algorithms are difficult and time consuming due to the nonlinearity and complexity of the model structure. In this case, derivative free optimization algorithms should be used. One of the derivative free optimization algorithms is population based meta-heuristic algorithm. In this study, a replicated response measured data set is chosen from the literature. Modified Michaelis-Menten model is preferred to model for this data set since the data set is composed of replicated measures. Parameter estimation is achieved by minimizing the sum of square error function. Here, Genetic Algorithm, a well known population based meta-heuristic optimization algorithm, is preferred as a nonlinear optimization tool. The obtained results are compared with the presented results in the literature.

Keywords: *Replicated response measured nonlinear data set, nonlinear regression analysis, Modified Michaelis-Menten model, Genetic Algorithm.*

References

[1] Akapame, S.K. (2014), *Optimal and Robust Design Strategies for Nonlinear Models Using Genetic Algorithm*, Montana State University, 162.

[2] Bates, D.M. and Watts, D.G. (1988), *Nonlinear Regression Analysis and Its Applications*, U.S.A., John Wiley & Sons, 365.

[3] Heydari, A., Fattahi, M. and Khorasheh, F. (2015), *A New Nonlinear Optimization Method for Parameter Estimation in Enzyme Kinetics*, Energy Sources, Part A: Recovery, Utilization, and Environmental Effects, 37, 1275–1281.

[4] Mitchell, M. (1999), *An Introduction to Genetic Algorithms, England*, MIT Press, 158 pp.

[5] Türkşen,Ö. and Tez, M. (2016), *An Application of Nelder-Mead Heuristic-Based Hybrid Algorithms: Estimation of Compartment Model Parameters*, International Journal of Artificial Intelligence, 14(1), 112-129.

İSTKON 10

10th INTERNATIONAL STATISTICS CONGRESS

ISC-2017

CELEBRATING 25 Years Turkish Statistical Association

DECEMBER 6-8 , 2017
ANKARA UNIVERSITY

TURKISH STATISTICAL ASSOCIATION
www.turkistatistik.org

ANKARA UNIVERSITY

ANKARA UNIVERSITY
FACULTY OF SCIENCES

ANKARA UNIVERSITY
FACULTY OF APPLIED SCIENCES

www.istkon.net

Congress Mail : kayit@istkon.net
Association Mail : tid@turkistatistik.org