

2<sup>nd</sup> International Conference on Applied Statistics (ICAS 2019)  
December 27-29, 2019

*Emerging challenges in a data-centric world*

PROGRAMME AND ABSTRACT BOOK

Institute of Statistical Research and Training (ISRT)  
University of Dhaka, Dhaka-1000, Bangladesh

## Contents

<b>Programme Book</b>	<b>1</b>
Directions to Conference Rooms . . . . .	3
Programme Summary . . . . .	5
Day 1 (December 27, 2019) . . . . .	5
Day 2 (December 28, 2019) . . . . .	6
Day 3 (December 29, 2019) . . . . .	7
Programme Details . . . . .	8
Day 1 (December 27, 2019) . . . . .	8
Day 2 (December 28, 2019) . . . . .	15
Day 3 (December 29, 2019) . . . . .	27
<b>Abstract Book</b>	<b>41</b>
Keynote Papers . . . . .	43
Markov neighborhood regression for high-dimensional inference . . . . .	43
Ethics for applied statisticians . . . . .	44
Plenary Papers . . . . .	45
Developments in response surface designs for comparative experiments and related topics	45
Big data and the precision medicine - A methodological study . . . . .	46
Online Grocery Recommendation using Bandit Models . . . . .	47
Clustering, Optimization and Applications . . . . .	48
Invited Papers . . . . .	49
Graduation of mortality rates of a national pension scheme for pricing annuity products	50
Autoregressive integrated distributed optimal Lagged (ARIDOL) model . . . . .	50
On efficiency of split-plot response surface designs when some observations are missing .	51
Real-time visual subject tracking and classification by combining motion signal analysis and tridimensional-shape feature classifiers with group-induction boosting algo- rithms . . . . .	53
Geographically-dependent individual-level models for infectious diseases transmission . .	55
Time series analysis of nigeria stock returns using garch models with structural breaks .	55
Parameter estimation for the mixtures of normal distributions using discriminant analysis	56

Simulated annealing-backpropagation algorithm on parallel trained maxout networks (SABPMAX) in detecting credit card fraud . . . . .	56
On extended negative binomial distribution for count data with inflated frequencies . .	57
Investigating the role of ICT determinants of private commercial banks with stochastic frontier models in Bangladesh . . . . .	58
Diagnostic analysis of statistics in biomedical publication - A cautionary note . . . . .	59
A modified approach for testing independence of error and covariates in nonparametric regression . . . . .	60
Integrating pathway information with machine learning approach to construct a sufficient predictor gene set for phenotype classification . . . . .	61
Nonparametric performance hypothesis testing with the information ratio . . . . .	61
Assessment of nutritional status among adult Ho Tribal population of Debra Block of Paschim Medinipur district, West Bengal, India . . . . .	62
Second and Forth Digit Finger Lengths are Good Predictors of Height among Adult Muslim Men of Murshidabad District, West Bengal, India . . . . .	63
Assessment of nutritional status among adult Ho Tribal population of Debra Block of Paschim Medinipur district, West Bengal, India . . . . .	63
Efficient estimators using characteristics of Poisson distribution . . . . .	64
Dealing with treatment-confounder feedback and sparse follow-up in longitudinal studies - An application of the marginal structural model with a multiple sclerosis cohort	64
Assessing the livelihood status of fishermen at Sunamganj district in Bangladesh . . . .	65
ARIMA with intervention with GJR-GARCH in DSEX Index . . . . .	66
A three-parameter poisson-weighted lindley distribution with properties and applications	66
Prevalence of malnutrition and anaemia among tribal women in India: A study on NFHS-4 Data . . . . .	67
A new approach to determine the coefficient of skewness and an alternative form of boxplot	68
Inequalities in farm sector labour force in India . . . . .	68
Low birth weight is the consequence of maternal psychiatry, nutrition and gestational age: Results from a marginal model approach . . . . .	70
Violence against women in West Bengal: Its extent and causes . . . . .	71
Modeling the stability and determinants of household food insecurity: A pair copula construction approach . . . . .	72
Divide and recombine approach for big data analysis and prediction of trajectory probabilities from repeated measures data using machine learning algorithms . . . . .	72

Analyzing the factor structure and sleep quality of Pittsburgh sleep quality index in indian information technology sector . . . . .	73
Comparing means of two correlated variables using pooled data of paired and independent samples . . . . .	74
Impact of Differential Sectoral Output Growth on Environment: A Panel Data Analysis of Bangladesh . . . . .	74
Relative efficiency of higher normed estimators over the least squares estimator . . . . .	75
Improvement of the Performance of Clustering Methods Using Outlier Modification Rule	76
Spatial optimal design for habitat condition models . . . . .	78
Regression-type Estimators in the Presence of Non-response . . . . .	78
Predictive Comparison of Vine Copula Models . . . . .	80
Tobacco use among Bangladeshi secondary school students: A comparison between 2007 and 2013 GYTS surveys . . . . .	80
Machine learning application for geospatial big data: Case studies for wheat area mapping and brick kilns identification from satellite images. . . . .	82
Farmers role and challenges to adaptation hinders sustainability in agricultural sector of Bangladesh: A study based on early flash flood areas in Bangladesh . . . . .	82
Internal consistency and structure of the Statistical Anxiety Rating Scale (STARS): A case study of SUST, Sylhet. . . . .	84
A consultant's care book: Interesting cases from 17 years of statistical collaboration in Australia . . . . .	84
Testing the Mediation effects using bootstrapping in CB-SEM:The why and how? . . . .	86
Correlates of childhood overweight/obesity in Bangladesh: A community-level cross-sectional study . . . . .	87
The efficiency of bivariate Fay-Herriot small area estimators . . . . .	87
Impact of pre-disclosed inducements on response rate and quality of primary data . . . .	88
Divide and Recombine Approach for Fitting Logistic Regression Model for Big Data . . .	89
Empirical profiling of banks in distress in Bangladesh using logit model with panel data	90
A Marginal Conditional Model for Covariate Effects on Depression Data . . . . .	91
Breeding design of important economic crops by Statistical analysis of NGS Data . . . .	93
Evaluation of sample size requirements for developing a risk prediction model . . . . .	94
Sample size calculation for external validation of risk models . . . . .	95
Bayesian analysis of accelerating stress-strength models . . . . .	95

Inequalities in Farm sector labour force in India . . . . .	96
Factors affecting food grain production and their forecast in Bangladesh 2050 . . . . .	97
Factors affecting educational performance of academically non at-risk primary students: A multilevel modelling approach . . . . .	97
Application of deep learning in bioinformatics . . . . .	99
Integration analysis of multi-omics data using hierarchical structural component models	100
Data visualization & machine learning: Impact in drug development . . . . .	100
Bivariate Chi-Square Distribution and Some of its Applications . . . . .	102
Response-Based Multiple Imputation Method for Minimizing the Impact of Covariate Detection Limit in Logistic Regression . . . . .	103
Nexus between Bank-Based Financial Inclusion and Economic Growth in Asia: Do Size of Economy, Governance, Financial Crisis and Regulations Matter? . . . . .	103
CPI compilation from big data . . . . .	104
Measure and assessment of state statistical quality in Vietnam Statistics . . . . .	104
Current status of sustainable development goals: 2016-2030 A “one world” approach to the global development agenda . . . . .	105
Joint model for Longitudinal data subject to nonlinear mixed-effects model and time-to- event data: A two stage approach . . . . .	106
Recent trends in fertility behavior in Bangladesh . . . . .	107
Contributed Oral Papers . . . . .	108
Monitoring small variation of a non-normal process . . . . .	108
Multiple regression model to examine the incidence of government expenditure on neona- tal mortality in Nigeria . . . . .	108
Bias-reduced and separation-proof poisson regression model with small or sparse count data . . . . .	109
Statistical prediction of chemical components in groundwater of Bangladesh . . . . .	110
Association of biomass fuel smoke with respiratory symptoms among children under five in urban areas: Results from Bangladesh urban health survey, 2013 . . . . .	110
On the improvement of likelihood estimation for the shape parameter of generalized extreme value lifetime distribution . . . . .	111
Knowledge and practices on exclusive breastfeeding among mothers in rural area of Ra- jshahi district in Bangladesh . . . . .	112
The penalized likelihood for AFT models with small or rare event survival data . . . . .	112

Dealing with separation or near-to-separation in the multinomial logit models with application to childhood health seeking behavior data . . . . .	113
Influencing factors of maternal health service utilization in Bangladesh: A statistical analysis of the BDHS data 2014 . . . . .	114
Using fractional polynomials for examining the association of age with anemia and malnutrition of children under five years . . . . .	115
Determinants of unmet need for modern contraceptive in Bangladesh: A mixed model analysis . . . . .	115
Determinants of coital frequency among currently married women in Bangladesh . . . .	116
Variable selection for censored data using modified correlation adjusted correlation (MCAR) scores . . . . .	117
Analysis of rainfall occurrence in consecutive days using Markov models with covariate dependence in selected regions of Bangladesh . . . . .	117
Spatial patterns of child growth indicators in Bangladesh: A hierarchical bayesian spatial model . . . . .	118
Quality of corporate environmental performance and financial performance: Evidence from select Asian countries . . . . .	119
Piecewise linear multiple change point model: An application to survival analysis . . . .	120
Construction of windows for pharmacokinetic sampling in phase I clinical trial . . . . .	120
Predictive performance of the logistic regression model under separation problem and minimal EPV . . . . .	121
Simulation study: An analysis of the incorporation of mixed variables into principal component approach . . . . .	122
A non parametric likelihood ratio test for comparison of several count data model and its application to GATS data . . . . .	122
Variable selection for censored data with greedy algorithm based adaptive quantile regression models . . . . .	123
Patient-specific dose finding in seamless phase I/II clinical trials . . . . .	124
Structural equation model: Biosocial factors affecting adolescent's reproductive health knowledge in rural Bangladesh . . . . .	124
Using machine learning algorithm: "Analysis and prediction of diabetes" . . . . .	125
Using penalized methods for propensity score models with rare exposure in observational study with survival outcome . . . . .	126
Estimating heterogeneous causal effects of size at birth on children's nutritional status in Bangladesh . . . . .	126

Performance of some model selection criterions to identify the correct parametric survival model . . . . .	127
Geostatistical exploratory analysis of groundwater level of Barisal . . . . .	128
Increasing burden of multi-drug resistant tuberculosis in HIV infected individuals: a systematic review and meta-analysis . . . . .	128
Early stopping in seamless phase I/II clinical trials . . . . .	129
Classification of elderly health status in Sylhet district, Bangladesh: A frailty index approach . . . . .	130
Early childbearing: more women in Bangladesh are having their first child early in life .	131
A spatial multilevel model using conditional autoregressive processes . . . . .	131
Tree-based ensemble methods: A comparison using simulation and application to hypertension data . . . . .	132
Determining the impact of employment status on contraceptive use in Bangladesh . . .	133
The risk of plasmodium vivax parasitaemia after plasmodium falciparum infection: an individual patient data meta-analysis . . . . .	133
Availability of essential and priority medicines for treating sick children among health facilities in Bangladesh . . . . .	134
An analysis of high dimensional logistic regression for variable selection with lasso technique	135
The impact of wage remittances on the economy of SAARC region: An econometric approach . . . . .	136
Relationship of contraception prevalence rate and total fertility rate in the context of Bangladesh . . . . .	136
Performance of CAR and SAR model in smoothing spatial data under different neighborhood assumptions . . . . .	137
Generalized linear model for multivariate polytomous response . . . . .	138
Survival probabilities of stomach and colon cancer patients in Bangladesh . . . . .	138
Deep neural network for predicting neonatal mortality in Bangladesh . . . . .	139
Exploring the influencing factors for contraceptive use among married women: A special mixed method study of Bangladesh and other twenty low and middle income countries based on DHS data . . . . .	140
A cause and effect study on road accident in Bangladesh . . . . .	141
Estimation of coverage of antenatal and postnatal care at district level in Bangladesh: An application of small area estimation method . . . . .	141

Regional and country-level variation of prevalence of malnutrition under five children in South Asian countries: An evidence from 2014-2017 demographic and health surveys . . . . .	142
Determinants of hemoglobin level among children (6-59 months) in South Asian countries: A multilevel modeling approach from recent South Asian DHS surveys . . . . .	143
Shifting the patterns of overall chronological age into the biological age among senior citizen . . . . .	143
Lack of knowledge and misperceptions about thalassaemia among college students in Bangladesh: A community-level cross-sectional study . . . . .	144
Comparison of classification performance between decision tree and random forest algorithms: a case study on women empowerment issue in Bangladesh . . . . .	145
Divide and Recombine Approach for Fitting Logistic Regression Model for Big Data . .	146
Spatio-temporal change of green space dynamics and rapid urbanization in Bangladesh .	146
Correlates of stunting among under-five children in Bangladesh: A multilevel approach .	147
Patterning prevalence and correlates of stunting among under-five children in Bangladesh based on BDHS (2007-2014) data . . . . .	148
Prevalence and associated risk factors of hypertension among adults in Bangladesh: A case study in south western Bangladesh . . . . .	149
Effect of maternal factors on nutritional status of under-five children in Bangladesh: A statistical investigation . . . . .	149
Reference based genome assembling to recover full genome sequence from NGS dataset .	150
Identification of biomarker genes influencing colorectal cancer: Insights from the systems biology approach . . . . .	151
In-silico identification and characterization of RNAi genes in sweet orange . . . . .	152
Microbial clustering based on 16S rRNA sequence: A comparative study . . . . .	153
Regression based fast multi-trait QTL analysis . . . . .	153
Statistical phenomics in crop growth and development research . . . . .	154
De-novo assembling to recover whole genome from NGS dataset . . . . .	155
Detection of toxic chemical compounds and their associated toxicogenomic biomarkers using robust co-clustering approach . . . . .	156
Genome wide association studies to detect important SNPs: A comparative study . . .	156
Gene expression data analysis using robust model based clustering . . . . .	157
Computational identification of RNAi genes in wheat genome . . . . .	158
Diagnosis of competitiveness among listed companies in Bangladesh . . . . .	158



Morbidity and associated factors of nutrition among under five children of garment workers in Dhaka city . . . . .	159
Comparison of Statistical Models for Cross-over design . . . . .	160
Monitoring small variation of a non-normal process . . . . .	160
Contributed Posters . . . . .	162
Intimate partner violence (IPV) with miscarriages, stillbirths and abortions: Identifying vulnerable households for women in Bangladesh . . . . .	162
Assessing vulnerability to poverty using downside mean semi-deviation approach: An empirical analysis on a balanced panel from rural Bangladesh . . . . .	162
Investigating causal effects of low birth weight on malnutrition in children of age under three years: evidence from Bangladesh demographic and health survey 2014 . . . . .	163
Impact of sample sizes on the accuracy of estimates for a two-level logistic regression model	164
Determinants of depression among university students . . . . .	164
Socio-economic and demographic factors influencing double burden of anemia among mothers and their under-five children in Bangladesh: Logistic regression analysis	165
Factors influencing early initiation of breastfeeding among Bangladeshi mothers: Two level logistic regression model . . . . .	166
Comparison of multiple logistic regression and artificial neural network models in predicting acute respiratory infections of under-five children in Bangladesh . . . . .	167
Rapid consumption methodology: A dynamic approach for household expenditure survey and poverty measures . . . . .	167
Systematic review and meta-analysis of injecting drug use as a risk factor of HIV/AIDS in Bangladesh . . . . .	168
Socio-demographic factors associated with women empowerment in Bangladesh: the role of interactions . . . . .	169
Diagnosis and prediction of heart disease using machine learning and data mining techniques	169
Continuously observed multi-state model applied to london hospital neonatal care data . . . . .	170
Association between the type of fuel use with acute respiratory infections among children under-five in Bangladesh . . . . .	171
Comparison of individual ARIMA model and wavelet-ARIMA model for forecasting financial time series data . . . . .	171
Missing continuous outcomes in stepped wedge cluster randomised trials . . . . .	172
Breastfeeding and childhood diseases: Evidence from nationally representative survey . . . . .	173
Predicting malnutrition status of under-five children using tree based models . . . . .	173

On the performance of Newton-Raphson, modified Newton-Raphson, Fisher scoring and EM algorithm techniques under maximum likelihood estimation . . . . .	174
Meta-analysis on child malnutrition and it's determinants . . . . .	175
LogitLDA: A logistic transformed linear bayes classifier for gene expression data classification . . . . .	175
Influence of kernel for disease status prediction through support vector machine . . . . .	176
The impact of international oil price on Dhaka stock exchange . . . . .	177
A systematic review of causes and solutions of water logging in Dhaka city . . . . .	177
Performance evaluation of machine learning algorithms for MRI brain image classification	178
Conditional inference for discrete longitudinal data in the presence of excess zeros . . . . .	179
Effect and control of automotive emission in Dhaka city . . . . .	179
A zero-inflated negative binomial longitudinal model for count data: Generalized quasi-likelihood approach for inference . . . . .	180
First-order integer valued AR process for data with excess zeros and its inference . . . . .	181
Blockwise robust singular value decomposition for MRI brain image denoising . . . . .	182
Comparing different measures of model selection for forecasting very short time series: ARIMA vs. exponential smoothing . . . . .	182
Prediction of child malnutrition in Bangladesh: A comparison between predictive ability of logistic regression and classification trees . . . . .	183
Comparing rainfall prediction of north-eastern side of Bangladesh: SARIMA, ANN and simple exponential smoothing approach . . . . .	184
Identifying the associated factors of infant and under-five child mortality in Bangladesh: A semi-parametric approach . . . . .	185
Identifying and predicting factors associated with contraceptive use among women in Bangladesh: A machine learning approach . . . . .	185
ECG beat analysis using discrete wavelet coefficient . . . . .	186
A satisfactory threshold of body mass index to identify elderly nutritional risk: A cross-sectional study . . . . .	187
Spatial distribution of infant mortality rate at district level in Bangladesh using small area estimation method . . . . .	187
Finding the spatial auto correlation and multiple linear regression analysis of literacy rate, 2011 of Bangladesh . . . . .	188
Determinants of stunting, wasting and underweight in under five children: An analysis of BDHS data from 2004, 2007, 2011 and 2014. . . . .	189

Relation between demographic data and photoplethysmography (PPG) signal features in estimating human health metrics . . . . .	190
Spousal violence among married women in the slum area: A study in Khulna city of Bangladesh . . . . .	190
Rejection sampling scheme for simulating from multivariate normal density . . . . .	191
Bayesian accelerated failure time survival regression model: An application to under-five mortality data in Bangladesh . . . . .	192
Nested frailty model for two-level clustered survival data: Analyzing infant mortality in Bangladesh . . . . .	192
Zero inflated negative binomial mixed model for analyzing antenatal care visits in rural Bangladesh: A comparative study . . . . .	193
Poisson regression with observation-level random effects: A comparative study . . . . .	194
Consumer buying behavior and preferences of online shopping: A study on Dhaka city, Bangladesh . . . . .	194
<b>Author Index</b>	<b>196</b>

# Programme Book

~

## Directions to Conference Rooms

The venue of the conference “the Nabab Nawab Ali Senate Bhaban” is a three-storied building. Room locations and directions to the rooms are given below.

### Room A (Main Senate Hall)

This room is located at the second floor of the building. Directions to the “Room A”.

1. Go straight up on the grand staircase
2. Go up to the 2nd floor using the elevators/stairs on either side ( left or right)
3. The entrance to Room A is just up ahead (to the left or right, depending on the elevators taken)

NB. Please note there are six entrances/exits from “Room A”. The directions will be placed adjacent to the pathways.

### Room B

This is the special conference room (seminar room) of the senate bhaban and is located at the first floor of the building. Directions to “Room B”.

1. Go straight up on the grand staircase
2. Move straight ahead, then enter through the main entrance and you will find “Room B” (the special conference room)

### Room C

This is the conference room of the senate bhaban and is located at the second floor of the building. Directions to “Room C”.

1. Go straight up on the grand staircase
2. Go up to the 2nd floor using the elevator/staircase on the left
3. Move ahead a few steps and you will find “Room C” (named the conference room)

### Room D

This room is located in the ground floor of the senate bhaban and is also known as Rajanigandha Members’ Lounge of the Dhaka University Alumni Association (DUAA). Directions to “Room D”.

1. Move along the pathway to the right from the Senate Building grand staircase
2. Turn left at the corner
3. Keep going until you see the entrance on the left marked Room D (Rajanigandha Member’s Lounge)

**Dining Hall**

The dining hall of the senate bhavan is located in the first floor of the building. Directions to “Dining Hall”

1. Go straight up on the grand staircase
2. Move straight ahead and enter through the main entrance. Turn left and you will find the cafeteria

## Programme Summary

### Day 1 (December 27, 2019)

Time	Details	Room
08:30-10:00	Registration	Main Lobby (First floor)
10:00-10:30	Tea break	Dining Hall
10:30-11:30	Inaugural session	A
12:00-13:00	Keynote-1	A
13:00-14:00	Lunch	Dining Hall
14:00-16:00	Invited session-1	A
	Invited session-2	B
	Invited session-3	C
16:00-16:20	Tea break	Dining Hall
16:20-17:50	Invited session-4	A
	Invited session-5	B
	Invited session-6	C



**Day 2 (December 28, 2019)**

Time	Details	Room
09:00-10:30	Plenary session-1	A
10:30-11:00	Tea break	Dining Hall
11:00-12:00	Keynote-2	A
12:30-13:15	Poster session-1	North Lobby (Second floor)
13:00-14:00	Lunch	Dining Hall
14:00-16:00	Invited session-7	A
	Invited session-8	B
	Invited session-9	C
	Contributed session-1	D
15:30-16:15	Poster session-2	North Lobby (Second floor)
16:00-16:20	Tea break	Dining Hall
16:20-17:20	Contributed session-2	A
	Contributed session-3	B
	Contributed session-4	C
	Contributed session-5	D

**Day 3 (December 29, 2019)**

Time	Details	Room
09:00-10:30	Plenary session-2	A
10:30-11:00	Tea break	Dining Hall
11:00-13:00	Invited session-10	A
	Invited session-11	B
	Invited session-12	C
	Contributed session-6	D
12:30-13:15	Poster session-3	North Lobby (Second floor)
13:00-14:00	Lunch	Dining Hall
14:00-16:00	Invited session-13	A
	Invited session-14	B
	Invited session-15	C
	Contributed session-7	D
16:00-16:20	Tea break	Dining Hall
16:20-17:20	Contributed session-8	A
	Contributed session-9	B
	Contributed session-10	C
	Contributed session-11	D

## Programme Details

### Day 1 (December 27, 2019)

---

December 27, 2019                      Room A (12:00 - 13:00)  
Keynote Session - 1  
Session chair: Shahariar Huda, Kuwait University, Kuwait

---

---

Paper Id	Presenter	Paper title
P-232	Martinez, W	Ethics for Applied Statisticians

---

---

December 27, 2019                      Room C (14:00 - 16:00)

Invited Session - 1: Applied statistics

Session chair: Ehsan Karim, University of British Columbia, Canada

---

---

Paper Id	Presenter	Paper title
P-193	Arman, MHR	Empirical Profiling of Banks in Distress in Bangladesh using Logit Model with Panel Data
P-195	Sultana, P	Probability models for identifying best among the novice candidates
P-214	Monir, MM	Breeding design of important economic crops by Statistical analysis of NGS Data
P-218	Sen, R	Bayesian Analysis of Accelerating Stress Strength Models
P-244	Dau, TTQ	Measure and assessment of state statistical quality in Vietnam Statistics

---

---

December 27, 2019                      Room B (14:00-16:00)

Invited Session - 2: Mathematical statistics

Session chair: Nizam Uddin, University of Central Florida, Orlando, FL, USA

---

Paper Id	Presenter	Paper title
P-63	Rahman, UH	A three-parameter Poisson-weighted Lindley distribution with properties and applications
P-100	Uddin, N	Comparing means of two correlated variables using pooled data of paired and independent samples
P-111	Khan, S	Effect-size measures for binary outcome variables and their estimation methods in meta-analysis for public health studies
P-132	Roy, DC	Regression-type Estimators in the Presence of Non-response
P-198	Gulshan, J	A Marginal Conditional Model for Covariate Effects on Depression Data
P-240	Joarder, AH	Bivariate Chi-Square Distribution and Some of its Applications

---

---

December 27, 2019                      Room A (14:00-16:00)  
 Invited Session - 3: Machine learning and mathematical statistics  
 Session chair: Mezbahur Rahman, Minnesota State University, Mankato, USA

---

Paper Id	Presenter	Paper title
P-12	Roitman, L	Real-time visual subject tracking and classification by combining motion signal analysis and tridimensional-shape feature classifiers with group-induction boosting algorithms.
P-20	Rahman, M	Parameter Estimation for the Mixtures of Normal Distributions using Discriminant Analysis
P-22	Golingay, SMD	Simulated Annealing-Backpropagation Algorithm on Parallel Trained Maxout Networks (SABPMAX) in Detecting Credit Card Fraud
P-39	Jahan, N	Integrating Pathway Information with Machine Learning Approach to Construct a Sufficient Predictor Gene Set for Phenotype Classification
P-47	Sharma, P	Efficient Estimators Using Characteristics of Poisson Distribution
P-109	Das, S	Relative Efficiency of Higher Normed Estimators Over the Least Squares Estimator

---

---

December 27, 2019                      Room A (16:20-17:50)

Invited Session - 4: Bioinformatics

Session chair: Md. Nurul Haque Mollah, University of Rajshahi, Bangladesh

---

---

Paper Id	Presenter	Paper title
P-116	Rashid, MM	Deterministic and Stochastic Modeling of the Drosophila Circadian Clock
P-204	Mollah, MNH	Global analysis of drug-target interactions via meta-predictor with convolution on protein sequences
P-208	Rahaman, MM	Statistical Phenomics in Crop Growth and Development Research
P-210	Hasan, MN	Detection of Toxic Chemical Compounds and their Associated Toxicogenomic Biomarkers using Robust Co-clustering Approach
P-233	Shahjaman, M	Identification of Potential Biomarkers from Transcriptomics Data using Integrated Bioinformatics Approaches

---

---

December 27, 2019                      Room C (16:20-17:50)

Invited Session - 5: Healthcare analysis

Session chair: Utpal Kumar De, North Eastern Hill University, Shillong, India

---

Paper Id	Presenter	Paper title
P-9	Pal, M	Unequal Effect of Parents' Education on Child Health and Health Care
P-11	Bose, K	Age Trends in Anthropometric Characteristics and Nutritional Status among Adult Mahali Females of Bankura District, West Bengal, India
P-13	Bharati, P	Prevalence of over nutrition and anaemia among adult Indians: A study from NFHS-4
P-64	Bharati, S	Prevalence of malnutrition and anaemia among tribal women in India: A study on NFHS-4 Data
P-82	De, UK	Does Road Connectivity Improve Access to Healthcare? - A Study in Rural Meghalaya

---



---

December 27, 2019                      Room B (16:20-17:50)  
 Invited Session - 6: Epidemiology and environmental statistics  
 Session chair: Mohammad Abdul Hoque, University of Portsmouth, UK

---

Paper Id	Presenter	Paper title
P-24	Mitra, A	Challenges of Valuation of Environmental Goods-A Case Study of Ecotourism in the Eastern Himalayas (India)
P-87	Ahmed, SS	Determining the best modelling approach for forecasting height yield of hybrid spruce in the tree improvement program
P-134	Kopecky, J	iArsenic - Instant arsenic screening of hand pump tubewells in Bangladesh using a statistical model embedded in a Web application
P-161	Ullah, S	How far do people travel to dialysis? Time and distance travelled for metropolitan haemodialysis patients
P-189	Feist, SE	Spatio-temporal controls on salinization within the Ganges Delta, southwest Bangladesh

---

**Day 2 (December 28, 2019)**

---

December 28, 2019                      Room A (09:00-10:30)

Plenary Session - 1

Session chair: Faming Liang, Purdue University, USA

---

---

Paper Id	Presenter	Paper title
P-106	Das, K	Big Data and the Precision Medicine - A methodological study
P-245	Bandyopadhyay, S	Clustering, Optimization and Applications

---

---

December 28, 2019                      Room A (11:00 - 12:00)

Keynote Session - 2

Session chair: Sanghamitra Bandyopadhyay, Indian Statistical Institute, Kolkata, India

---

---

Paper Id	Presenter	Paper title
P-226	Liang, F	Markov Neighborhood Regression for High-Dimensional Inference

---

---

 December 28, 2019 North Lobby (12:30-13:15)

Contributed Poster Session - 1

Session chair: Shahadut Hossain, UAE University, UAE

---

Paper Id	Presenter	Paper title
P-50	Sanim, MAI	Investigating causal effects of low birth weight on malnutrition in children of age under three years: Evidence from Bangladesh demographic and health survey 2014
P-74	Hossain, S	Comparison of Multiple Logistic Regression and Artificial Neural Network Models in Predicting Acute Respiratory Infections of Under-five Children in Bangladesh
P-89	Chowdhury, S	Systematic Review and Meta-analysis of Injecting Drug Use as a Risk Factor of HIV/AIDS in Bangladesh
P-108	Khurshid, S	Diagnosis and prediction of heart disease using machine learning and data mining techniques
P-129	Shihab, S	Meta-Analysis on Child Malnutrition and it's Determinants
P-140	Bosu, P	LogitLDA: A Logistic transformed Linear Bayes Classifier for Gene Expression Data Classification
P-141	Paul, S	Influence of Kernel for Disease Status Prediction through Support Vector Machine
P-149	Arefin, AB	A Systematic Review of Causes and Solutions of Water Logging in Dhaka City
P-151	Rashid, M	Performance Evaluation of Machine Learning Algorithms for MRI Brain Image Classification
P-152	Ferdous, Z	Conditional inference for discrete longitudinal data in the presence of excess zeros
P-162	Akter, R	First-order integer valued AR process for data with excess zeros and its inference
P-167	Hasan, MN	Blockwise Robust Singular Value Decomposition for MRI Brain Image Denoising
P-179	Adib, A	ECG beat analysis using discrete wavelet coefficient
P-219	Shuzan, MNI	Relation between demographic data and Photoplethysmography (PPG) signal features in estimating human health metrics
P-228	Rayhan, MAI	Bayesian Accelerated Failure Time Survival Regression Model: An Application to Under-Five Mortality Data in Bangladesh
P-229	Pal, B	Nested Frailty Model for Two-Level Clustered Survival Data: Analyzing Infant Mortality in Bangladesh

---

---

December 28, 2019                      Room A (14:00-16:00)

Invited Session - 7: Biostatistics and data science

Session chair: Gareth Ambler, University college London, UK

---

---

Paper Id	Presenter	Paper title
P-48	Karim, ME	Dealing with treatment-confounder feedback and sparse follow-up in longitudinal studies - An application of the marginal structural model with a multiple sclerosis cohort.
P-178	Nesa, MK	The Efficiency of Bivariate Fay-Herriot Small Area Estimators
P-216	Ambler, G	Evaluation of Sample Size Requirements for Developing a Risk Prediction Model
P-217	Pavlou, M	Sample size calculation for external validation of risk models
P-237	Nandy, K	Data Visualization and Machine Learning: Impact in Drug Development
P-241	Hossain, S	Response-Based Multiple Imputation Method for Minimizing the Impact of Covariate Detection Limit in Logistic Regression

---

---

December 28, 2019                      Room B (14:00-16:00)

Invited Session - 8: Time series analysis

Session chair: TBA

---

Paper Id	Presenter	Paper title
P-6	Amaefula, CG	Autoregressive integrated distributed optimal Lagged (ARIDOL) model
P-18	Musa, Y	Time Series Analysis of Nigeria Stock Returns Using GARCH Models with Structural Breaks
P-61	Huq, MM	ARIMA with Intervention with GJR-GARCH in DSEX Index
P-99	Chatterjee, A	Analyzing the Factor Structure and Sleep Quality of Pittsburgh Sleep Quality Index in Indian Information Technology Sector
P-101	Khan, ZS	Impact of Differential Sectoral Output Growth on Environment: A Panel Data Analysis of Bangladesh
P-135	Hoque, ME	Predictive Comparison of Vine Copula Models

---

---

December 28, 2019                      Room C (14:00-16:00)  
 Invited Session - 9: Survival analysis  
 Session chair: Kalyan Das, Indian Institute of Technology, Bombay, India

---

Paper Id	Presenter	Paper title
P-36	Al-Mutairi, DK	On a generalization of lifetime distribution under unknown environments
P-130	Roy, SS	Estimating the hazard functions of two alternating recurrent events in the presence of covariates
P-145	Bardhan, A	Health Scenarios of Tribal women in West Bengal
P-158	Das, U	Analysis of Interval Censored Competing Risks Data Under Missing Causes
P-238	Lee, S	Kaplan-Meier median survival time based multifactor dimensionality reduction method for identifying gene-gene interactions associated with the survival phenotype
P-247	Chakraborty, A	Joint model for Longitudinal data subject to nonlinear mixed-effects model and time-to-event data: A two stage approach

---

---

December 28, 2019                      Room D (14:00-16:00)  
 Contributed Session - 1: Survival analysis  
 Session chair: Wasimul Bari, University of Dhaka, Bangladesh

---

Paper Id	Presenter	Paper title
P-30	Islam, MM	On the Improvement of Likelihood Estimation for the Shape Parameter of Generalized Extreme Value Lifetime Distribution
P-49	Alam, TF	The penalized likelihood for AFT models with small or rare event survival data
P-62	Mimi, MA	Variable Selection for Censored Data Using Modified Correlation Adjusted coRelation (MCAR) Scores
P-69	Chakraborty, S	Piecewise Linear Multiple Change point Model: An Application To Survival Analysis
P-81	Nishat, MNS	Variable Selection for Censored Data with Greedy Algorithm based Adaptive Quantile Regression Models
P-95	Hossain, M	Using machine learning algorithm: Analysis and prediction of diabetes
P-105	Chowdhury, SH	Performance of Some Model Selection Criteria to Identify the Correct Parametric Survival Model
P-142	Ghosh, S	An analysis of high dimensional logistic regression for variable selection with lasso technique
P-154	Ahammed, T	Survival Probabilities of Stomach and Colon Cancer Patients in Bangladesh
P-190	Das, A	Comparison of Classification Performance between Decision Tree and Random Forest Algorithms: A Case Study on Women Empowerment Issue in Bangladesh

---



---

 December 28, 2019 North Lobby (15:30-16:15)

Contributed Poster Session - 2

Session chair: Shahid Ullah, Flinders University, Australia

---

Paper Id	Presenter	Paper title
P-29	Afiaz, A	Intimate Partner Violence (IPV) with miscarriages, stillbirths and abortions: Identifying vulnerable households for women in Bangladesh
P-46	Basit, A	Assessing Vulnerability to Poverty Using Downside Mean Semi-deviation Approach: An Empirical Analysis on a Balanced Panel from Rural Bangladesh
P-51	Biswas, MR	Impact of sample sizes on the accuracy of estimates for a two-level logistic regression model
P-56	Sorkar, MNU	Determinants of Depression among University Students
P-58	Limon, MA	Socio-economic and demographic factors influencing double burden of anemia among mothers and their under-five children in Bangladesh: Logistic regression analysis
P-59	Ayesha, U	Factors influencing early initiation of breastfeeding among Bangladeshi mothers: Two level logistic regression model
P-83	Mazumder, SR	Rapid Consumption Methodology: A Dynamic Approach for Household Expenditure Survey and Poverty Measures
P-92	Chowdhury, MA	Socio-Demographic Factors Associated with Women Empowerment in Bangladesh: The Role of Interactions
P-110	Akter, MT	Continuously Observed Multi-state Model Applied to London Hospital Neonatal Care Data
P-114	Anjum, A	Association between the type of fuel use with acute respiratory infections among children under-five in Bangladesh
P-117	Nazra, KU	Comparison of Individual ARIMA Model and Wavelet-ARIMA Model for Forecasting Financial Time Series Data
P-119	Hossain, S	Missing Continuous Outcomes in Stepped Wedge Cluster Randomised Trials
P-122	Hasan, MN	Breastfeeding and Childhood Diseases: Evidence from Nationally Representative Survey
P-125	Uddin, B	On the Performance of Newton-Raphson, Modified Newton-Raphson, Fisher Scoring and EM Algorithm Techniques under Maximum Likelihood Estimation.
P-143	Hisham, AM	The Impact of International Oil Price on Dhaka Stock Exchange
P-157	Sultana, S	A zero-inflated negative binomial longitudinal model for count data: Generalized quasi-likelihood approach for inference

---

---

December 28, 2019                      Room A (16:20-17:20)  
 Contributed Session - 2: Applied statistics  
 Session chair: Ohidul Islam Siddiqui, University of Dhaka, Bangladesh

---

Paper Id	Presenter	Paper title
P-65	Mahmud, S	Analysis of Rainfall Occurrence in Consecutive Days Using Markov Models with Covariate Dependence in Selected Regions of Bangladesh
P-68	Maji, SG	Quality of Corporate Environmental Performance and Financial Performance: Evidence from Select Asian Countries
P-128	Tusty, MA	Tree-Based Ensemble Methods: A Comparison Using Simulation and Application to Hypertension Data
P-165	Das, KR	A Cause and Effect Study on Road Accident in Bangladesh
P-215	Sharmin, S	Diagnosis of Competitiveness among Listed Companies in Bangladesh

---

---

December 28, 2019                      Room B (16:20-17:20)  
 Contributed Session - 3: Public health  
 Session chair: Papia Sultana, University of Rajshahi, Bangladesh

---

Paper Id	Presenter	Paper title
P-86	Methun, MIH	Structural Equation Model: Biosocial Factors Affecting Adolescent's Reproductive Health Knowledge in Rural Bangladesh
P-123	Hossain, MK	Classification of Elderly Health Status in Sylhet District, Bangladesh: A Frailty Index Approach
P-138	Hakim, S	Availability of essential and priority medicines for treating sick children among health facilities in Bangladesh
P-170	Rahman, MA	Estimation of Coverage of Antenatal and Postnatal Care at District Level in Bangladesh: An Application of Small Area Estimation Method
P-177	Islam, MA	Regional and country-level variation of prevalence of malnutrition under five children in South Asian countries: an evidence from 2014-2017 Demographic and Health Surveys
P-221	Rumana, AS	Morbidity and Associated Factors Of Nutrition Among Under Five Children of Garment Workers in Dhaka City

---

---

December 28, 2019                      Room C (16:20-17:20)  
Contributed Session - 4: Environmental statistics  
Session chair: Ohid Ullah, Shahjalal University of Science and Technology, Sylhet, Bangladesh

---

---

Paper Id	Presenter	Paper title
P-25	Siddiquee, R	Statistical Prediction of Chemical Components in Groundwater of Bangladesh
P-66	Lipi, N	Spatial patterns of child growth indicators in Bangladesh: A Hierarchical Bayesian spatial model
P-112	Islam, H	Geostatistical Exploratory Analysis of Groundwater Level of Barisal
P-127	Islam, MM	A spatial multilevel model using conditional autoregressive processes.
P-148	Tamanna, RJ	Performance of CAR and SAR Model in Smoothing Spatial Data under Different Neighborhood Assumptions

---

---

December 28, 2019                      Room D (16:20-17:20)  
 Contributed Session - 5: Epidemiology  
 Session chair: M. Rezaul Karim, University of Rajshahi, Bangladesh

---

Paper Id	Presenter	Paper title
P-26	Hasan, M	Association of biomass fuel smoke with respiratory symptoms among children under five in urban areas: Results from Bangladesh Urban Health Survey, 2013
P-85	Shanto, SI	Patient-specific dose finding in seamless phase I/II clinical trials
P-113	Sultana, ZZ	Increasing Burden of Multi-drug Resistant Tuberculosis in HIV Infected Individuals: A Systematic Review and Meta-analysis
P-136	Hossain, MS	The risk of plasmodium vivax parasitaemia after plasmodium falciparum infection: An individual patient data meta-analysis
P-199	Muyeed, A	Prevalence and Associated Risk Factors of Hypertension among Adults in Bangladesh: A Case Study in South Western Bangladesh

---

**Day 3 (December 29, 2019)**

---

December 29, 2019                      Room A (09:00-10:30)  
Plenary Session - 2  
Session chair: Nizam Uddin, University of Central Florida, USA

---

---

Paper Id	Presenter	Paper title
P-34	Huda, S	Developments in Response Surface Designs for Comparative Experiments and Related Topics
P-220	Kumar, UD	Online Grocery Recommendation using Bandit Models

---

---

December 29, 2019                      Room A (11:00-13:00)  
 Invited Session - 10: Big data  
 Session chair: Tariqul Hasan, University of New Brunswick, Canada

---

Paper Id	Presenter	Paper title
P-96	Chowdhury, RI	Divide and recombine approach for big data analysis and prediction of trajectory probabilities from repeated measures data using machine learning algorithms.
P-115	Auwul, MR	Improvement of the Performance of Clustering Methods Using Outlier Modification Rule
P-147	Matin, MA	Machine learning application for geospatial big data: Case studies for wheat area mapping and brick kilns identification from satellite images
P-234	Hasan, MAM	Application of Deep Learning in Bioinformatics
P-235	Park, T	Integration analysis of multi-omics data using hierarchical structural component models
P-243	Doan, NV	CPI compilation from big data

---

---

December 29, 2019                      Room B (11:00-13:00)  
 Invited Session - 11: Public health  
 Session chair: Shahid Ullah, Flinders University, Australia

---

Paper Id	Presenter	Paper title
P-16	Mahsin, M	Geographically-dependent individual-level models for infectious diseases transmission
P-42	Bhandari, S	Assessment of Nutritional Status among Adult Ho Tribal Population of Debra Block of Paschim Medinipur district, West Bengal, India
P-88	Begum, MR	Low birth weight is the consequence of maternal psychiatry, nutrition and gestational age: Results from a marginal model approach
P-137	Ahammed, T	Tobacco use among Bangladeshi secondary school students: a comparison between 2007 and 2013 GYTS surveys
P-168	Hossain, MS	Correlates of childhood overweight/obesity in Bangladesh: A community-level cross-sectional study
P-184	Siddiquee, MH	Lack of knowledge and misperceptions about thalassaemia among college students in Bangladesh: A community-level cross-sectional study

---



---

December 29, 2019                      Room C (11:00-13:00)  
 Invited Session - 12: Design of experiment and agriculture statistics  
 Session chair: TBA

---

Paper Id	Presenter	Paper title
P-3	Appiah, SK	Graduation of mortality rates of a national pension scheme for pricing annuity products
P-8	Yakubu, Y	On efficiency of split-plot response surface designs when some observations are missing
P-91	Yimam, JA	Modeling the Stability and Determinants of Household Food Insecurity: A Pair Copula Construction Approach
P-156	Ferdushi, KF	Farmers role and challenges to adaptation hinders sustainability in agricultural sector of Bangladesh: A study based on early flash flood areas in Bangladesh
P-222	Chakraborty, S	Inequalities in Farm sector labour force in India
P-223	Matin, KA	Factors affecting food grain production and their forecast in Bangladesh 2050.

---

---

December 29, 2019                      Room D (11:00-13:00)  
Contributed Session - 6: Bioinformatics  
Session chair: Taesung Park, Seoul National University, Korea

---

---

Paper Id	Presenter	Paper title
P-155	Chowdhury, S	Deep Neural Network for Predicting Neonatal Mortality in Bangladesh
P-201	Hasan, MM	Reference Based Genome Assembling to Recover Full Genome Sequence from NGS Dataset
P-202	Tasmia, SA	Identification of Biomarker Genes Influencing Colorectal Cancer: Insights from the Systems Biology Approach
P-203	Mosharaf, MP	In-Silico Identification and Characterization of RNAi Genes in Sweet Orange
P-205	Islam, MM	Microbial Clustering Based on 16S rRNA Sequence: A Comparative Study
P-206	Alam, MJ	Regression Based Fast Multi-trait QTL Analysis
P-209	Kibria, KM	De-Novo Assembling to Recover Whole Genome from NGS Dataset
P-211	Firoj, MG	Genome Wide Association Studies to Detect Important SNPs: A comparative Study
P-212	Ahmed, FF	Gene Expression Data Analysis Using Robust Model based Clustering
P-213	Akond, Z	Computational Identification of RNAi Genes in Wheat Genome

---

---

 December 29, 2019 North Lobby (12:30-13:15)

Contributed Poster Session - 3

Session chair: TBA

---

Paper Id	Presenter	Paper title
P-124	Hemo, SA	Predicting Malnutrition Status of Under-Five Children using Tree Based Models
P-153	Islam, J	Effect and Control of Automotive Emission in Dhaka City
P-169	Zaman, T	Comparing Different Measures of Model Selection for Forecasting Very Short Time Series: ARIMA vs. Exponential Smoothing
P-171	Chowdhury, AA	Prediction of Child Malnutrition in Bangladesh: A Comparison Between Predictive Ability of Logistic Regression and Classification Trees
P-172	Islam, SN	Comparing Rainfall Prediction of North-Eastern Side of Bangladesh: A SARIMA, ANN and Simple Exponential Smoothing Approach
P-173	Rahman, MM	Identifying the Associated Factors of Infant and Under-five Child Mortality in Bangladesh : A Semi-parametric Approach
P-174	Ara, T	Identifying and Predicting Factors Associated with Contraceptive Use Among Women in Bangladesh : A Machine Learning Approach
P-181	Farhana, Z	A Satisfactory Threshold of Body Mass Index to Identify Elderly Nutritional Risk: A Cross-sectional Study
P-182	Shah, MRT	Spatial Distribution of Infant Mortality Rate at District Level in Bangladesh Using Small Area Estimation Method
P-185	Yousuf, MPH	Finding the Spatial Auto Correlation and Multiple Linear Regression Analysis of Literacy Rate, 2011 of Bangladesh (for poster presentation)
P-186	Dhar, A	Determinants of Stunting, Wasting and Underweight in under five children:an analysis of BDHS data from 2004, 2007, 2011 and 2014.
P-224	Pranti, MR	Spousal Violence among Married Women in the Slum Area: A Study in Khulna City of Bangladesh
P-227	Akter, SF	Rejection Sampling Scheme for Simulating from Multivariate Normal Density
P-230	Mahi, MH	Zero In ated Negative Binomial Mixed Model for Analyzing Antenatal Care Visits in Rural Bangladesh: A Comparative Study
P-231	Maria, M	Poisson Regression with Observation-Level Random Effects: A Comparative Study
P-248	Tanvir, K	Consumer buying behavior and preferences of online shopping : a study on Dhaka city, Bangladesh

---

---

December 29, 2019                      Room A (14:00-16:00)

Invited Session - 13: Demography

Session chair: Mizanur Rahman, University of North Carolina, Chapel Hill, USA

---

Paper Id	Presenter	Paper title
P-53	Ullah, MO	Assessing the livelihood status of fishermen at Sunamganj district in Bangladesh
P-72	Chakraborty, S	Inequalities in Farm sector labour force in India
P-90	Biswas, CS	Violence against Women in West Bengal: Its Extent and Causes
P-225	Majumder, UK	Factors Affecting Educational Performance of Academically non At-Risk Primary Students: A Multilevel Modelling Approach
P-246	Moslehuddin, M	Current status of sustainable development goals: 2016 -2030 a “one world” approach to the global development agenda
P-249	Rahman, M	Recent fertility trends in Bangladesh

---

---

December 29, 2019                      Room B (14:00-16:00)

Invited Session - 14: Miscellaneous

Session chair: TBA

---

Paper Id	Presenter	Paper title
P-27	Serra, IJA	On extended negative binomial distribution for count data with inflated frequencies
P-43	Ghosh, M	Second and Forth Digit Finger Lengths are Good Predictors of Height among Adult Muslim Men of Murshidabad District, West Bengal, India
P-159	Kawsar, LA	Internal consistency and structure of the Statistical Anxiety Rating Scale (STARS): A case study of SUST, Sylhet
P-188	Islam, S	Impact of pre-disclosed inducements on response rate and quality of primary data
P-239	Mannan, HR	Gains in life expectancy in the Australian population due to reductions in smoking: Comparisons between targeting the population versus a specific high risk group
P-242	Banerjee, PK	Nexus between Bank-Based Financial Inclusion and Economic Growth in Asia: Do Size of Economy, Governance, Financial Crisis and Regulations Matter?

---

---

December 29, 2019                      Room C (14:00-16:00)

Invited Session - 15: Non-parametric statistics

Session chair: Shahadut Hossain, UAE University, UAE

---

Paper Id	Presenter	Paper title
P-2	Abdelrahman, MM	A novel combination of Biochanin-A/ 5-Fluorouracil: Impact on wnt/beta-catenin signaling
P-28	Baten, MA	Investigating the Role of ICT Determinants of Private Commercial Banks with Stochastic Frontier models in Bangladesh
P-35	Maiti, SI	A Modified Approach for Testing Independence of Error and Covariates in Nonparametric Regression
P-41	Aboy, JBC	Nonparametric Performance Hypothesis Testing with the Information Ratio
P-67	Hossain, MF	A New Approach to Determine the Coefficient of Skewness and An Alternative Form of Boxplot
P-164	Rahman, DMM	Testing the Mediation effects using bootstrapping in CB-SEM: The why and how?

---

---

December 29, 2019                      Room D (14:00-16:00)  
 Contributed Session - 7: Biostatistics and clinical trial  
 Session chair: Menelaos Pavlou, University College London, UK

---

Paper Id	Presenter	Paper title
P-19	Mondol, MH	Bias-reduced and separation-proof Poisson regression model with small or sparse count data
P-37	Rana, MM	Knowledge and Practices on Exclusive Breastfeeding among Mothers in Rural Area of Rajshahi District in Bangladesh
P-52	Nusrat, N	Dealing with separation or near-to-separation in the multinomial logit models with application to childhood health seeking behavior data
P-71	Sultana, N	Construction of windows for pharmacokinetic sampling in Phase I clinical trial
P-75	Tabassum, KF	Predictive Performance of the Logistic Regression Model under Separation Problem and Minimal EPV
P-103	Akter, E	Using Penalized Methods for Propensity Score Models with Rare Exposure in Observational Study with Survival Outcome
P-104	Taskin, F	Estimating Heterogeneous Causal Effects of Size at Birth on Children's Nutritional Status in Bangladesh
P-121	Khan, NM	Early stopping in seamless phase I/II clinical trials
P-180	Amin, R	Determinants of Hemoglobin Level among children (6-59 months) in South Asian Countries: A multilevel modeling approach from recent South Asian DHS Surveys
P-200	Naima, US	Effect of maternal factors on nutritional status of under- five children in Bangladesh: A statistical investigation

---

---

December 29, 2019                      Room A (16:20-17:20)  
 Contributed Session - 8: Demography and social statistics  
 Session chair: Pk. M. Matiur Rahman, University of Dhaka, Bangladesh

---

Paper Id	Presenter	Paper title
P-10	Usman, A	Multiple Regression Model to examine the incidence of Government expenditure on Neonatal Mortality in Nigeria
P-57	Islam, MK	Determinants of Unmet Need for Modern Contraceptive in Bangladesh: A Mixed Model Analysis
P-60	Bashar, MI	Determinants of coital frequency among currently married women in Bangladesh
P-196	Rahman, MM	Correlates of Stunting Among Under-five Children in Bangladesh: A Multilevel Approach
P-197	Rahman, MT	Patterning prevalence and correlates of stunting among under-five children in Bangladesh based on BDHS (2007-2014) Data

---



---

December 29, 2019                      Room B (16:20-17:20)  
 Contributed Session - 9: Demography  
 Session chair: Khan Abdul Matin, University of Dhaka, Bangladesh

---

Paper Id	Presenter	Paper title
P-126	Farhana, Z	Early Childbearing: More Women in Bangladesh are having their First Child Early in Life
P-133	Ahmed, A	Determining the impact of employment status on contraceptive use in Bangladesh
P-146	Hema, PS	Relationship of Contraception Prevalence Rate and Total Fertility Rate in the context of Bangladesh
P-163	Islam, MA	Exploring the Influencing Factors for Contraceptive Use among Married Women: A Special Mixed Method Study of Bangladesh and Other Twenty Low and Middle Income Countries Based on DHS Data
P-183	Rahman, MM	Shifting the Patterns of Overall Chronological Age into the Biological Age among Senior Citizen
P-250	Haider, MM	Explaining the discrepancies in data on the use of long-acting reversible contraceptives and permanent methods in Bangladesh

---

---

December 29, 2019                      Room C (16:20-17:20)  
Contributed Session - 10: Applied statistics  
Session chair: Enayetur Raheem, Senior Data Scientist at Cigna, USA

---

---

Paper Id	Presenter	Paper title
P-5	Mukta, MMR	Monitoring Small Variation of a Non-Normal Process
P-54	Islam, MA	Influencing Factors of Maternal Health Service Utilization in Bangladesh: A Statistical Analysis of the BDHS Data 2014
P-55	Afrose, M	Using fractional polynomials for examining the association of age with anemia and malnutrition of children under five years
P-191	Islam, MK	Divide and Recombine Approach for Fitting Logistic Regression Model for Big Data
P-192	Islam, MM	Spatio-Temporal Change of Green Space Dynamics and Rapid Urbanization in Bangladesh

---

---

December 29, 2019                      Room D (16:20-17:20)  
 Contributed Session - 11: Miscellaneous  
 Session chair: Md. Abdul Baten, Shahjalal University of Science and Technology, Bangladesh

---

Paper Id	Presenter	Paper title
P-76	Dewan, MN	Simulation Study: An Analysis of the Incorporation of Mixed Variables into Principal Component Approach
P-78	Rudra, S	A nonparametric likelihood ratio test for comparison of several count data model and its application to GATS data
P-144	Islam, MM	The Impact of Wage Remittances on the Economy of SAARC Region: An Econometric Approach
P-150	Sultana, N	Generalized linear model for multivariate polytomous response
P-236	Kamruzzaman, M	Comparison of Statistical Models for Cross-over design

---

# Abstract Book

~

## Keynote Papers

---

[P-232-K-01]

### Markov neighborhood regression for high-dimensional inference

*Faming Liang\**, Purdue University, USA

#### Abstract

We propose an innovative method for constructing confidence intervals and assessing p-values in high-dimensional linear and generalized linear models. The proposed method has successfully reduced the high-dimensional inference problem to a series of low-dimensional inference problems: For each regression coefficient  $\beta_i$ , the confidence interval and p-value can be calculated by regressing on a subset of variables selected according to the conditional independence relations between the corresponding variable  $X_i$  and other variables. Since the subset of variables forms a Markov neighborhood of  $X_i$  in the Markov network formed by all the variables  $X_1, X_2, \dots, X_p$ , the proposed method is coined as Markov neighborhood regression. We test the proposed method on high-dimensional linear, logistic and Cox regression. The numerical results indicate that the proposed method significantly outperforms the existing ones. Based on the Markov neighborhood regression, a method for learning causal structures for high-dimensional linear and generalized linear models is proposed and applied to the problems of identification of drug sensitive genes and cancer driver genes. The idea of using conditional independence relations for dimension reduction is general and potentially can be extended to other high-dimensional or big data problems as well.

*Keywords:*

#### Biography

Professor Liang is working in the Department of Statistics at Purdue University where he leads the Statistical Computing Research Group. His research interests are in data science from the perspective of statistical computing. He has published several books and over 100 research papers in top ranked journals in the areas of machine learning, Markov Chain Monte Carlo and high performance computing with a focus on big-data and high-dimensional statistical problems encountered in biomedical science, engineering, and social science. Professor Liang is a fellow of the ASA and IMS and an elected member of the ISI. He has served as Associate Editor of many reputed statistical journals including JASA, Technometrics and Biometrics. He has received several prestigious awards for his outstanding contributions to research.

[P-232-K-02]

**Ethics for applied statisticians**

*Wendy Martinez\**, President, The American Statistical Association and the Mathematical Statistics Research Center of the U.S. Bureau of Labor Statistics

**Abstract**

I had the honor of moderating a panel session on ethics at the 2019 American Statistical Association (ASA) Conference on Statistical Practice. Talks from the panel members and remarks from the audience reinforced for me that as statisticians and data scientists we must be intentional in our application of the Ethical Guidelines for Statistical Practice. Throughout our careers, we must confront issues such as company privileged information, biases in algorithms and models, data ownership, and more. Our analyses always start with data, so I will focus more on data ethics in this presentation. I will begin by offering my definition of data ethics and will then provide some real-world examples where ethical concerns arose. I will facilitate our discussion by exploring case studies, which will help all of us think through some of the issues and prepare us to be ethical data scientists.

*Keywords:* data ethics, algorithm bias, case studies.

**Biography**

Dr. Wendy Martinez is the president of the American Statistical Association for the year 2020 and directs the Mathematical Statistics Research Center of the U.S. Bureau of Labor Statistics. She is also the coordinating editor of the journal *Statistics Surveys* (jointly sponsored by four major statistical societies). Her research interest includes computational statistics, exploratory data analysis, data visualization, outlier detection, statistical pattern recognition, spatial statistics, and the analysis of unstructured text. She published several books and over 30 publications in the field of computational statistics and data visualization and analysis. Dr. Martinez was elected a Fellow of the American Statistical Association in 2006, and a member of the International Statistical Institute in 2007. She has received many awards including the Founders Award (2017) of the American Statistical Association, for outstanding leadership and support of statistical and multidisciplinary research and BLS Commissioner's Award (2018) for management excellence.

## Plenary Papers

---

[P-034-PL-1]

### Developments in response surface designs for comparative experiments and related topics

*Shahariar Huda\**, Kuwait University, Kuwait

#### Abstract

“Response Surface Designs (RSDs)” are considered for the situation where the experimenter is primarily interested in comparing the responses at different points rather than obtaining the response at individual locations in the factor space. A closely related situation is that in which the experimenter wishes to obtain the slope(s) of a response surface. Designs for “differences” and “slopes” play an important role in “sensitivity analysis” and in “product optimization”. This article provides a comprehensive survey of developments in these fields and discusses the implications of some recent results.

*Keywords:* A-minimax, D-minimax, Optimal design, Rotatability, Slope-rotatability.



---

[P-106-PL-2]

### **Big data and the precision medicine - A methodological study**

*Kalyan Das\** and *Siuli Mukhopadhyay*, Indian Institute of Technology, Bombay, India

#### **Abstract**

Trendy research in biomedical sciences juxtaposes the term ‘precision’ to medicine and public health with companion words like big data, data science, and deep learning. The creation of powerful systems for the effective use of biomedical Big Data in Personalized Medicine (Precision Medicine) will require significant scientific and technical developments. One of the important challenges posed to biomedical research by our increasing ability to tackle large datasets is to frame up a possible model after fulfilling the need for standardization of data content, format, and clinical definitions. The present work focuses on analyzing both the technical and societal hurdles related to the development of prediction models of health risks, diagnoses and outcomes from integrated biomedical databases. In the context of longitudinal studies which involve high heterogeneity of treatment effects particularly while working with big data, there is a necessity to develop model for personalized treatment. As different individuals react to the same treatment very differently, subject specific information play a big role here. We estimate unobserved subject-specific treatment effects and apply the random forest algorithm to allocate effective treatments for individuals. This article focuses on the potential impact of big data analysis to improve health, prevent and detect disease at an earlier stage taking into consideration the personalize interventions.

*Keywords:*

[P-220-PL-3]

**Online Grocery Recommendation using Bandit Models***Nandini Seth and U Dinesh Kumar\**, Indian Institute of Management, Bangalore, India**Abstract**

Due to higher frequency of purchase and customer stickiness, grocery e-tailing has emerged as an attractive and profitable online business. Similar to other e-retail businesses, online grocery retailers also employ personalised recommender systems in order to support sales and learn about customer preferences. Most existing recommender systems are designed to suggest one-time purchases but are not designed to efficiently make recommendations for grocery purchases which are recurrent in nature. Therefore, the mechanism for grocery recommender system is different due to the frequency and recurrence of purchase, absence of explicit customer rating and difficulty in accurately predicting item pairs bought together. While most recommender systems are targeted at improving immediate sales, a better understanding of a customer's preferences is also necessary in grocery context since the customer engagement period is longer. This paper aims at identifying a better iterative recommendation mechanism for online groceries based on the explore-exploit dilemma of the Multi-Armed Bandit models. The paper proposes a two-stage mechanism to optimally choose grocery recommendations for a customer while balancing the economic goal of accumulating sales and sequentially learning about evolving customer preferences. The first stage uses frequency of purchase to reveal the offline propensity of choice for each customer-item pair. This output is used in stage two to calculate reward priors for Bayesian bandit model which is employed to make the final online recommendations. The authors have used the customer basket data from the biggest grocery e-tailer in India to show the usability of this mechanism. The work presented in this paper can be used as a starting point for designing recommender systems for other items characterised by repeat and frequent purchases.

*Keywords:* Page Ranking Algorithms for recommender systems, Multi-arm Bandit Models, Recommender Systems, Reinforcement learning algorithms

---

[P-245-PL-4]

### Clustering, Optimization and Applications

*Sanghamitra Bandyopadhyay\**, Indian Statistical Institute, Kolkata, India

#### Abstract

Clustering is an unsupervised exploratory data analysis tool which groups the data on the basis of some similarity/dissimilarity metric such that a predefined criterion is optimized. The problem of clustering is therefore essentially one of optimization. The use of metaheuristics like genetic algorithms has been made successfully in the past for clustering a data set. It is to be noted that the clustering problem admits a number of criteria or cluster validity indices that have to be simultaneously optimized for obtaining improved results. Hence in recent times the problem has been posed in a multiobjective optimization (MOO) framework and popular metaheuristics for multiobjective optimization have been applied. In this talk, we will first briefly discuss about the fuzzy *c*-means algorithm and the basic principles of MOO. Subsequently it will be shown how a popular multiobjective optimization algorithm may be used for solving the clustering problem. Since such algorithms provide a number of solutions, a way of combining the multiple clustering solutions so obtained into a single one using supervised learning will be explained. The talk will conclude by demonstrating an application of the multiobjective clustering technique on gene expression data sets.

*Keywords:*

## Invited Papers

---

[P-002-I-01]

### A novel combination of Biochanin-A/5-Fluorouracil: Impact on wnt/beta-catenin signaling

*Mohamed Mahmoud\**, Ain Shams University, Abbassia, Cairo, Egypt, *Mohamed El-Sesy*, National Cancer Institute, Giza, Egypt, *Mai F Tolba*, Chapman University, Irvine, CA, USA, and *Dalal A. Abou El Ella*, Ain Shams University, Abbassia, Cairo, Egypt

#### Abstract

Colon cancer is one of the most common types of cancer worldwide. It is considered the second leading cause of cancer deaths in both women and men combined. Although 5-fluorouracil (5FU) is the first line of therapy for colon cancer, its use is significantly hampered by drug toxicity and development of tumor resistance. Therefore, it is important to investigate strategies to overcome such stumbling block. Biochanin-A (Bio-A) is a promising natural isoflavone that has selective cytotoxicity against cancer cells. This study aimed at investigating the modulatory impact of Bio-A on 5FU cytotoxicity in colon cancer cells and exploring the possible underlying mechanisms with emphasis on Wnt/beta catenin signals. Cytotoxicity and Calcosyn synergy analyses were conducted to assess IC50 and synergism between Bio-A and 5FU in both Caco-2 and HCT-116 cell lines. The effect of the different treatments on Wnt/ $\beta$ -catenin signals was assessed using phospho-tracer enzyme linked immunoassays.

Bio-A alone showed cytotoxic activity against Caco-2 and HCT-116 colon cancer cells with IC50 values of 34.34 and 34.41  $\mu$ M respectively using sulforhodamine-B cytotoxicity assay. Combined treatment of 5FU/ Bio-A at 1:2 and 1:4 ratios reduced the IC50 of 5FU significantly by 26.7 and 57% in Caco-2 cell line. The same treatments produced reduction of 5FU IC50 by 13.2 and 39.6% in HCT-116. Calcosyn® analysis indicated that the combinations were synergistic in both cell lines with combination index CI < 1. Since Wnt/  $\beta$ -catenin signaling plays an important role in the pathogenesis of colon cancer, the study was substantiated to investigate the effects of Bio-A and Bio-A/5FU combination on some key signals in addition to cell cycle analysis and assessment of cyclin D1 expression. Bio-A significantly suppressed Akt-phosphorylation and hence reduced GSK3 $\beta$  phosphorylation and preserved it in the active form. Hence, treatment with Bio-A boosted the ratio of phosphorylated  $\beta$ -catenin(p-S45)/total  $\beta$ -catenin indicating an increased degradation of  $\beta$ -catenin and reduced signaling. This study results underscore the potential merit of Bio-A as an adjuvant to 5FU for the management of colon cancer. Further tumor xenograft studies are encouraged in order to verify this effect in vivo.

*Keywords:* Colon cancer, biochanin, 5-FU, wnt/beta-catenin

---

[P-003-I-02]

### Graduation of mortality rates of a national pension scheme for pricing annuity products

*Derrick Asamoah Owusu\**, *Simon K. Appiah*, and *Awura Amma A. Danso*, Kwame Nkrumah University of Science and Technology, Kumasi, Ghana

#### Abstract

Mortality improvements continue to pose a severe challenge for the management of public pension schemes due to increasingly life expectancy. An understanding of mortality patterns of insured lives of in a given population is key in determining the pricing and reserving in life insurance or pension products such as annuities. Annuities provide a guarantee income payout for the annuitant, which makes most pensioners resort to purchasing annuity products with their lump sum. Unfortunately, annuity markets do not exist in developing countries such as Ghana. In this study a non-parametric method of graduation was applied to crude mortality rates of a national pension scheme. Graduated mortality rates were produced at various smoothing ( $h$ ) and fitness ( $z$ ) parameter values. The optimal set of graduated mortality rates were obtained at smoothing and fitness values of  $h = 100$  and  $z = 3$ , respectively, which provided a good fit at both young and old ages. The optimal graduated mortality rates were then used to compute life expectancy of the scheme members for the annuity tables to be developed for pricing the annuity products.

*Keywords:* Crude mortality rates, Graduation, Whittaker-Henderson, Pension scheme, Annuity pricing

---

[P-006-I-03]

### Autoregressive integrated distributed optimal Lagged (ARIDOL) model

*Chibuzo Gabriel Amaefula*, Federal University Otuoke Bayelsa, Nigeria

#### Abstract

The paper proposed a multiple regression with autoregressive lagged ordered identity called autoregressive integrated distributed optimal lagged (ARIDOL) model for long-run equilibrium and adjustment to short-run equilibrium relationships. The proposed model infuses some pre-tests as regularity conditions. The demonstration of the model applicability via empirical data sets indicates that at the long-run, previous month CPI has positive effect on current CPI and this is significant under 1% level. Current UKEXR has positive influence on current CPI and it is significant under 10% level. The result also indicates that adjustment to short-run equilibrium relationship is negative and significant at 1% level as expected. And some model diagnostic tests helped to conclude that the proposed model is adequate and

can be recommended for modelling long- run equilibrium relationship where adjustment to short-run equilibrium relationship may be required.

*Keywords:* Autoregression, cointegration, PACF, short-run adjustment, integration order

[P-008-I-04]

### On efficiency of split-plot response surface designs when some observations are missing

*Yisa Yakubu\**, Federal University of Technology, Minna, Niger State, Nigeria, *Angela Unna Chukwu*, University of Ibadan, Ibadan, Oyo state, Nigeria, and *Usman Abubakar*, Federal University of Technology, Minna, Niger State, Nigeria

#### Abstract

In most experimental situations, some observations are lost due to unforeseen circumstances. This leads to destruction of desirable design properties like independence, normality of errors and optimality. Efficiency of completely randomized response surface designs when some observations are missing has been extensively studied in literature. However, complete randomization of experimental runs is often unrealistic in most industrial experimental situations where some factors are difficult to change. Therefore such designs are often conducted within a split-plot structure and their performance thus depends on relative magnitude ( $d$ ) of model's whole-plot and subplot variance components. Split-plot central composite designs (CCD) consist of factorial ( $f$ ), whole-plot axial ( $\alpha$ ), subplot axial ( $\beta$ ), and center ( $c$ ) points. This study examines the effect of missing pairs of observations of these points on efficiency of split-plot central composite designs in terms of  $trace(A)$ , maximum prediction variance ( $G$ ), and integrated average prediction variance ( $V$ ) optimality criteria, under various values of  $d$ . Efficiency functions were formulated in terms of these criteria and efficiency of reduced designs (due to missing observations), relative to the corresponding full designs, were examined. Maximum  $A$ -efficiency losses of 19.1, 10.6, and 15.7% due to missing pairs of observations,  $ff$ ,  $\beta\beta$ , and  $f\beta$ , respectively, were observed at  $d = 0.5$ ; maximum  $G$ - and  $V$ -efficiency losses of 10.1, 0.1, 16.1, 0.1% and 0.1, 0.1, 1.1, 0.2% were observed, respectively when the pairs  $ff$ ,  $\alpha\alpha$ ,  $\beta\beta$ , and  $cc$ , were missing.  $A$ -efficiency was observed to be robust to missing  $cc$ ,  $\alpha\alpha$ ,  $\alpha c$ ,  $fc$ ,  $f\alpha$  observations while  $G$ - and  $V$ -efficiencies were each observed to be robust to missing  $\alpha\alpha$ . The study revealed that as  $d$  increases, the observed losses in efficiency of these designs become insignificant.

*Keywords:* Efficiency, Split-plot central composite design, Missing observations, Optimality criteria

[P-009-I-05]

**Unequal effect of parents' education on child health and health care**

*Manoranjan Pal\**, *Utpal Kumar De*, *Premananda Bharati*, and *Susmita Bharati*, Indian Statistical Institute, 203 BT Road, Kolkata 700108, India

**Abstract**

Parents have tremendous influence on growth and development of children. At the initial period of life, it is believed that mother is more important because there are some physical needs of children e.g., breastfeeding, which can only be done by mother. Both the parents should have emotional touch with the child. Apart from breastfeeding, it is generally believed that mother plays more vital role than father on the life of her child. Are there scientific evidences to support the belief that mother has more influence on child? We have not yet come across any such investigations other than taking opinion surveys. While we do not go against these opinions, we emphasize on the need of judging it scientifically. The paper carries out a systematic way of finding the effects of education of parents and their relative importance on the life of a child so far as Birth Weight, BCG, DPT, Polio and Measles Vaccination and Status of Anemia are concerned using unit level data of the latest round of National Family Health Survey (NFHS IV, 2015-16) of India. State wise secondary data are also used to see the effect of literacy of parents on the Sex Ratio at Birth of Children. Apart from preparing two-way tables, statistical tools like correlation, partial correlation and regression techniques are used for this purpose. Results of our analysis suggest that there is really unequal effect of mother and father and it is the mother, who counts more.

*Keywords:* Educational Levels of Parents, NFHS, Birth Weight, Sex Ratio at Birth, Vaccination, Status of Anemia

[P-011-I-06]

---

**Age trends in anthropometric characteristics and nutritional status among adult mahali females of Bankura District, West Bengal, India**

*Kaushik Bose\**, *Shilpita Bhandari*, *Mihir Ghosh*, *Binoy Kumar Kuiti*, and *Soma Pal*,  
Vidyasagar University, West Bengal, India

**Abstract**

The present study was undertaken to assess age trends in anthropometric measures and nutritional status among adult Mahali females. It was a community-based cross-sectional study, carried out in selected four villages of Bankura district, West Bengal, India. A total of 118 Mahali tribals, aged over 18 years were included in our study. The subjects were further classified into three age groups:  $\leq 30$  years, 31-49 years,  $\geq 50$  years. Anthropometric variables included height, weight, sitting height, knee height, mid-upper arm, medial calf circumferences and body mass index (BMI). In general, an inverse age trend was observed in all these anthropometric variables. This age trend was statistically significant ( $p < 0.05$ ) in case of height, weight, sitting height and medial calf circumference. In nutritional assessment, 75 individuals were found to have chronic energy deficiency (CED). This could have serious health implications.

*Keywords:* Age, Body mass Index, Nutritional Status

[P-012-I-07]

---

**Real-time visual subject tracking and classification by combining motion signal analysis and tridimensional-shape feature classifiers with group-induction boosting algorithms**

*Lucas Agudiez Roitman\**, Stanford University, Stanford, CA, USA

**Abstract**

This paper provides a novel and unprecedented approach for integrating motion features in the detection and classification of moving subjects in a static environment. More specifically, we measure the impact of the use of trajectory history, rotation history, blob orientation, motion frequency in the three axes, motion acceleration, segmentation errors, and flickering scores, and how they can influence classification of moving people, pets, and other objects. We apply our method to data captured by a combined color and depth camera sensor. We find that, while some motion descriptors slightly improve accuracy, the use of them in conjunction outperforms previous approaches in the classification and tracking of real-world moving subjects in real-time.



*Keywords:* real-time tracking; moving subjects; classification; motion signal; motion statistics; accelerometer; orientation; rgb camera; depth images; computer vision; machine learning; classifiers; boosting; artificial intelligence

---

[P-013-I-08]

### Prevalence of over nutrition and anaemia among adult Indians – A study from NFHS-4

*Premananda Bharati\**, *Manoranjan Pal*, and *Susmita Bharati*, Indian Statistical Institute, Kolkata, India

#### Abstract

Overweight or obesity is a big problem among Indians like other Asian countries, though its intensity varies due to interstate disparity in socio-economic and socio-cultural differences. Besides over nutrition, there is a great disparity of prevalence of anaemia. Thus, it is worth to study statewise variation of overnutrition and anemia among Indians, by gender also by socio-economic variables. For this study, we have used the part of published study from NFHS-4 adult Indians, (15-49 years) age groups. For women, the sample sizes are 647150 and 679415 respectively for over nutrition and anemia and for men, the corresponding sample sizes are 99167 and 98721 respectively. The socio-economic variables are age-groups, type of residence, marital status, educational status, religion and wealth index of the family. Over nutrition has been assessed through Body Mass Index (BMI), Classification of BMI is according to WHO (1988).

The study reveals that in India, percentage of over nutrition among women is 20.7 and among men it is 18.9. The gender difference is only 1.8%. Percentage of anemia among adult women is 53.0 and among men is 22.7, the gender difference being 30.3%, i.e females are much more affected than males. Over-nutrition concentrated zones in India are North, West and South whereas high anaemic affected zone is the East zone. If we compare our results of the present study with those of NFHS-3 data, it is seen that over nutrition has increased by 8.1% for women and 9.6% for men, whereas anaemia has decreased by 2.3% for women and 1.5% for men over the last ten years. The possible socio-economic factors responsible for over nutrition and anaemia have been identified and it is discussed in detail in the analysis section of the paper.

---

[P-016-I-09]

### Geographically-dependent individual-level models for infectious diseases transmission

*Md Mahsin\** and *Rob Deardon*, University of Calgary, Canada, and *Patrick Brown*, University of Toronto, Canada

#### Abstract

Modelling of infectious diseases has been increasingly used to evaluate the potential impact of different control measures and to guide public health policy decisions. In recent years, individual-level models (ILMs) have been effectively used to model infectious disease transmission. These models are well developed but assume the probability of disease transmission between two individuals depends only on their spatial (or network based) separation. In this study, we extend ILMs to geographically-dependent ILMs (GD-ILMs) that allow the evaluation of the effect of spatially varying risk factors (e.g., education, social deprivation), environmental factors (such as temperature, air quality, rainfall, and humidity), as well as unobserved spatial structure, upon the transmission of infectious disease. We consider a conditional autoregressive (CAR) model to capture the effects of unobserved spatially structured latent covariates or measurement error. We show how GD-ILMs can be fitted to data on both simulation and Alberta in influenza outbreaks epidemic within a Bayesian statistical framework using Markov chain Monte Carlo (MCMC) methods.

*Keywords:*

---

[P-018-I-10]

### Time series analysis of nigeria stock returns using garch models with structural breaks

*Yakubu Musa\** and *Ibrahim Adamu*, Usmanu Danfodiyo University, Sokoto, Nigeria

#### Abstract

This study examines the stock returns series using Symmetric and Asymmetric GARCH models with structural breaks in the presence of varying distribution assumptions. Volatility models of Symmetric GARCH(1,1), Asymmetric Exponential GARCH(1,1), Power GARCH(1,1) and GJR-GARCH(1,1) models were considered in estimating and measuring shock persistence, leverage effects and mean reversion rate. The empirical findings showed the high persistence of shock in returns series for the estimated models. However, when structural breaks were incorporated in the estimated models by including dummy variable in the conditional variance equations of all the models, there was significant reduction of shock persistence parameter and mean reversion rate. The study found the GJR-GARCH(1,1) with skewed student-t distribution to best fit the series. The volatility was forecasted

using GJR-GARCH(1,1) model and the values are compared with the actual values and the results indicates a continuous increase in unconditional variance.

*Keywords:* Volatility, GARCH, Structural Breaks

---

[P-020-I-11]

### Parameter estimation for the mixtures of normal distributions using discriminant analysis

*Mezbahur Rahman\**, Minnesota State University, Mankato, USA

#### Abstract

Parameter estimation procedure for the mixtures of normal distributions using discriminant analysis is introduced. Parameter estimation procedure for the mixtures of normal distributions using EM algorithm and using maximum likelihood are reviewed. A simulation study is conducted in comparing five variant methods involving maximum likelihood, EM algorithm, and discriminant analysis.

*Keywords:* Discriminant analysis; EM algorithm; Grid search; Maximum likelihood estimate; Newton-Raphson algorithm; Normal probability density function.

---

[P-022-I-13]

### Simulated annealing-backpropagation algorithm on parallel trained maxout networks (SABPMAX) in detecting credit card fraud

*S. D. Golingay\**, University of the Philippines-Diliman, Philippines

#### Abstract

Based on the Backpropagation (BP) artificial neural network algorithm, this study introduces the idea of combining Simulated Annealing (SA), a global searching algorithm and then proposes a new neural network algorithm: Simulated Annealing-Backpropagation Algorithm on Parallel Trained Maxout Networks (SABPMAX) algorithm. The proposed algorithm can improve the numerical stability and evaluation measures in detecting credit card fraud. It makes use of the global searching capability of SA and the precise local searching element of the backpropagation algorithm to improve the initial weights of the network towards improving detection of credit card fraud. Several models were made and tested using different fraud distributions. Furthermore, separate applications of BP algorithm and SABPMAX algorithm were compared. Numerical results show higher accuracy rate, higher sensitivity, shorter computing time, and overall better performance of the SABPMAX algorithm.

*Keywords:* Neural Network, Fraud Detection; Maxout Network, Class Imbalance, Credit Card Fraud

---

[P-024-I-14]

### **Challenges of valuation of environmental goods - A case study of ecotourism in the Eastern Himalayas (India)**

*Amitava Mitra\**, Rajiv Gandhi University, Itanagar, Arunachal Pradesh, India

#### **Abstract**

Ecotourism is a way to travel to natural areas that takes responsibility to conserve the environment and sustain well-being of local communities. Ecotourism came into prominence in the late eighties as a strategy for reconciling conservation with development in ecologically rich areas. Our case study is the Eastern Himalayas i.e. Arunachal Pradesh which is one of the 'Hot spot' of biodiversity in the world. The rich forest resources and biodiversity have to be conserved since the biodiversity does not correspond to manmade boundaries and hence there is an urgent need for collaboration between India and her neighbours in south Asia. An alternative measure has to be adopted which can act as an economic incentive for preservation of forest and biodiversity in the State. In this respect the promotion of ecotourism or nature-based tourism appears to be one of the best ways of generating revenue for the Government, as well as the generation of economic resources for the local community provided it is properly managed. It is expected that if local people reap direct benefits from protecting their surroundings, hopefully they will go on to protect nature and biodiversity. However, there needs to be economic valuation of nature-based tourism, which can help the policy makers to take unbiased decision on its management and sustainable utilization. Hence, the present paper will deal with the economic valuation of eco-tourism in Arunachal Pradesh, which can help to preserve the rich forests and biodiversity based on Travel Cost Methodology (TCM).

*Keywords:* Ecotourism, Valuation, Conservation, Travel Cost Method

---

[P-027-I-15]

### **On extended negative binomial distribution for count data with inflated frequencies**

*I. J. Serra\**, University of the Philippines Cebu, Philippines, *D. L. Lim-Polestico*, Mindanao State University, Marawi City, Lanao del Sur, Philippines

**Abstract**

This study extends the existing zero-inflated distributions through the flexibility of peaks in the data with excessive counts other than zeros and handles an over dispersion in the data. Moreover, this study formulates a proposed zero and  $k$  inflated negative binomial (ZkINB) distribution which is a mixture of a multinomial logistic and negative binomial distribution. The multinomial logistic component captures the occurrence of excessive counts, zero and  $k \geq 1$ , while the negative binomial component captures the counts that are assumed to follow a negative binomial distribution. Furthermore, this study derived the moment generating function of the distribution in order to solve some structural properties of the ZkINB, including the mean, variance and the skewness and kurtosis. Two real datasets were used to analyze the characteristics of the ZkINB. In both examples we used zero and  $k$  inflated negative binomial (ZkINB) distribution and compare it to the zero and  $k$  inflated Poisson (ZkIP) and zero-inflated count distributions. The first example illustrated a ZkINB distribution with inflations at 0 and  $k = 3$ , while the second example has inflations at 0 and  $k = 1$ . As a result, the zero and  $k$  inflated negative binomial distribution seems to exhibit a better fit than the inflated NB and POI count data distributions.

*Keywords:* Count data, inflated frequencies, zero inflated, over dispersion

---

[P-028-I-16]

**Investigating the role of ICT determinants of private commercial banks with stochastic frontier models in Bangladesh**

*Sakera Begum and Md. Azizul Baten\**, Shahjalal University of Science and Technology, Sylhet, Bangladesh

**Abstract**

ICT has now become the engine block of banking institutions worldwide and the Bangladesh banking industry is no exception. This study measures the impact of ICT determinants on 17 private commercial banks (PCBs) in Bangladesh employing Stochastic Cobb-Douglas and Trans-log cost and profit efficiency models considering a panel data set during 2007-2018. The average cost efficiency score was 65.8% and the average profit efficiency score was 50.5% for PCBs for Cobb-Douglas cost and profit frontier analysis models while in Translog stochastic cost and profit frontier analysis models, the average cost efficiency score was 66.3% and the average profit efficiency score was 53.9% for PCBs. Translog Stochastic Frontier model was more preferable than Cobb-Douglas Stochastic Frontier model. The IT personnel expenses (0.0018) and Credit Card Transaction (0.0013) were positively significant for profit efficiency of Stochastic Cobb-Douglas profit frontier model while the IT personnel expenses (0.0006) and credit card transaction (0.000006) were positively significant for PCBs in cost efficiency of Stochastic Translog frontier Model. The IT factors had a positive impact on private commercial banks so it may conclude that the Private commercial banking system was technologically more advanced.

*Keywords:* ICT, Private Commercial Banks, Stochastic frontier model, Tobit regression model, Likelihood-Ratio test, Bangladesh

[P-031-I-16]

**Diagnostic analysis of statistics in biomedical publication - A cautionary note***Ravinder Kumar Soni\**, Dayanand Medical College and Hospital, Ludhiana, India**Abstract**

Statistics is usually seen as a theoretical subject by majority of scientists. With emerging advances in science, this branch has also grown from its infancy era. Statistics is widely used in all branches of Medical Sciences. Right from the descriptive statistics, plenty of different statistical tools and techniques are used. Appropriate usage of statistical technique(s) for analysis of biomedical data is of utmost importance in order to make rationale and reliable inferences. The study explores critically the usage of various statistical method(s) in biomedical publications.

The data for the present study was collected from MedIND; A national Database of Indian Medical journals. We scrutinized all original research publications in the latest volume of available journals on MedIND website. Editorials, letters, case reports and review articles were excluded. There were 137 available journals having 1191 research publications. All these publications were scrutinized for their statistical content. Out of 1191 publications, 39.4% did not mention the design/ type of study and 36.2% did not specify place of study. Sampling techniques was mentioned/described in 65.9% publications. In methodology section of publication 733(61.5%) described the statistical methods used. Statistical software applied for analysis was mentioned only in 54.9% of studies. 68.1% of publications described the results with proper statistical explanation. Surprisingly 98.8% of publications did not specify/acknowledge the statistical consultation.

There were 85.7% publications in which one of statistical method was applied. Of the 550(46.2%) publications which applied Chi square test, 64(11.6%) did not check the assumption of test. ANOVA was used in 185(15.4%) studies, none of which followed the assumption of ANOVA. Despite the use of statistical software, exact p value was not mentioned in 170(14.3%) of publications. Despite wide use of statistical method(s) in biomedical research, there seems to be significant issue for scientific misconduct for its usage with applicable assumptions.

*Keywords:* Statistical technique, Biomedical publications

---

[P-035-I-18]

**A modified approach for testing independence of error and covariates in nonparametric regression**

*Saran Ishika Maiti\** and *Sthitadhi Das*, Visva-Bharati University, West Bengal, India

**Abstract**

Under nonparametric regression model  $Y = m(X) + \varepsilon$ , where  $m$  being an unknown regression function,  $Y$  being a real-valued response variable,  $X$  being the real covariate and  $\varepsilon$  the error term, testing the independence between error and covariate has always been paramount interest. Quite a few measures of association on distance-based approach have already existed in literature. Apart from distance-based tests, association-measures based tests too emerge in latest literature. The stepping stone of this approach lays on Kendall's  $\tau$ . Bergsma et al. (2014) proposed a modification of  $\tau$ , named  $\tau^*$ , which is constructed on the second order differences of neighbouring triplets of responses. Dhar et al. (2018) studied the asymptotic distribution of  $\tau^*$  and power of the test of independence under contiguous alternative. Mirroring the same way, we developed the test of independence based on third order differences of response variables. Further we unravel the local power of the proposed tests using Le Cam's contiguous alternative followed by a simulated example on a number of bootstrap samples. Finally, a fascinating biological data is handled where effect of an environmental stress, say light bandwidth, on some physiological parameters such as plant height, root length and relative water content is analyzed through our proposed test procedure.

*Keywords:* Kendall's  $\tau$ , Measures of Association, Asymptotic power, Contiguous alternative, Nonparametric regression model

---

[P-036-I-19]

**On a generalization of lifetime distribution under unknown environments**

*Dhaif K. Al-Mutairi\**, Kuwait University, Kuwait

**Abstract**

A generalized lifetime distribution for a unit that operates under unknown environment, where the environment follows an extended Lindley distribution, is developed. Several properties of this lifetime distribution are investigated. An application is given to show the usefulness of this distribution.

*Keywords:* Statistical modeling, stochastic models, survival models

---

[P-039-I-20]

**Integrating pathway information with machine learning approach to construct a sufficient predictor gene set for phenotype classification**

*Nusrat Jahan\**, James Madison University, USA

**Abstract**

We propose a 2-step phenotype classification method based on functional genomic data. In the first step, pathways relevant to specific phenotype conditions are identified. Then a sufficient dimension reduction technique selects the most relevant sufficient sets of predictors from each pathway. In the second step, a machine learning approach is used to detect the most significantly differential sets of sufficient pathway-based predictors. We demonstrate our method using an RNA-sequencing dataset from a study of pediatric acute myeloid leukemia (AML).

---

[P-041-I-21]

**Nonparametric performance hypothesis testing with the information ratio**

*J. B. I. Aboy\**, University of the Philippines Cebu, Philippine, and *J. Magadia*, University of the Philippines Diliman, Philippine

**Abstract**

This study proposes a nonparametric bootstrap-based test to compare performances between two portfolios in terms of their Information ratio. This serves as an extension to the literature that tests performance between two portfolio investment strategies that uses Sharpe ratio. Monte Carlo experiments show that the test has appropriate sizes and is powerful to most of the scenarios. Moreover, the test does not perform well in highly correlated portfolio returns, but is better when the mean of portfolio return is modeled using an autocorrelated process.

*Keywords:* Nonparametric bootstrap test, portfolio investments, information ratio.



[P-042-I-22]

**Assessment of nutritional status among adult Ho Tribal population of Debra Block of Paschim Medinipur district, West Bengal, India***Shilpita Bhandari\** and *Kaushik Bose*, Vidyasagar University, West Bengal, India**Abstract**

The tribal population in India, as per 2011 census is 104.3 million, constituting 8.6% of the total population. Overall, 89.97% of them live in rural areas. Nutrition has been a major health issue in India for centuries. Chronic hunger and under-nutrition is the worst tribulation of the poverty that still plagues millions of tribal households in India. Health and nutritional status of an individual depends on the food he/she eats. Humans need a wide range of nutrients to lead a healthy and active life and these are derived through the diet they consume daily. Good nutrition is a basic component of health. The objective of this study was to assess the anthropometric characteristics and nutritional status of the population.

The present cross-sectional study was carried out at 5 villages (Duan, Baguan, Bankakul, Amra Gerya and Kismat Duan Roypukur) under Debra block of Paschim Medinipur district, West Bengal, India, which is approximately 100 km away from Kolkata, the state capital of West Bengal. A total 161 adult (Female= 92 and Male= 69) Ho tribal people were included in this study. The mean (SD) value of height (cm), weight (kg), mid-upper arm circumference (cm), body mass index (BMI;  $kg/m^2$ ) was higher among males 160.35 (5.75), 56.32 (9.02), 25.37 (2.94) and 21.88 (3.20) respectively, than females 148.71 (5.52), 47.91 (8.42), 22.86 (2.73) and 21.62 (3.32). These value were statistically significant ( $p < 0.001$ ) except BMI. Based on MUAC, the prevalence of under nutrition was higher among females (43.5%) than males (21.7%). This was statistically significant ( $\chi^2 = 8.28$ ,  $p < 0.01$ ). However, based on BMI, the prevalence of under nutrition was similar in both sexes (males = 18.8%; females = 17.4%). According to WHO classification of chronic energy deficiency (CED), the rate of under nutrition indicated a poor situation. Thus, to reduce the nutritional stress among this ethnic group, appropriate nutritional intervention programmes are needed to be initiated.

*Keywords:* Body Mass Index, Mid Upper Arm Circumference, Nutritional Status

---

[P-043-I-23]

**Second and Forth Digit Finger Lengths are Good Predictors of Height among Adult Muslim Men of Murshidabad District, West Bengal, India**

*Mihir Ghosh\** and *Kaushik Bose*, Vidyasagar University, West Bengal, India

**Abstract**

Adult height is an important indicator of personal identification. There are numerous ways to establish adult height and their significance lies in the simplicity of measurement, applicability and accuracy in prediction. The present study was conducted on 388 adult Bengalee Muslim male brick-kiln workers of Murshidabad District of West Bengal, India. Measurements were taken following the standard techniques. Data were analyzed statistically to establish the relationship between finger lengths (both hands, second and forth digit fingers) and height. We found that both fingers (second and forth digits) of both hands had significant association with height and can be an important tool for height estimation. Significant differences in measurements and formulae for second and forth digit fingers in both hands were observed. The differences between the predicated values of height and actual values (direct measurement) were statistically not significant. Thus, we can conclude that it may be possible to accurately predict adult height from finger length measurements. This may have immense applications in anatomy and forensic sciences.

*Keywords:* Anthropometry, Finger length, Height estimation, Adult Male.

---

[P-044-I-24]

**Assessment of nutritional status among adult Ho Tribal population of Debra Block of Paschim Medinipur district, West Bengal, India**

*Rabindra Nath Das\**, The University of Burdwan, Burdwan, West Bengal, India, and  
*Youngjo Lee*, Seoul National University, Seoul, Korea

**Abstract**

The industrial practitioners try to find the operating condition that achieves the target mean value, and simultaneously reduces the process variability. Many techniques such as Taguchi's signal-to-noise ratios process, dual response surface methodology, joint generalized linear models, hierarchical generalized linear models are commonly used to achieve this goal. However, a source of process variability would be due to an extra variation in the settings of the experimental factors. In this paper extra-setting variation is allowed via random effects not only in the mean but also in the various sources of variability

of quality characteristic of a product. An illustrative example shows that random effects are necessary to reflect such phenomena. Meanings of these random effects in the models are discussed herein.

*Keywords:* Dual response surface methodology; Double hierarchical generalized linear models (DHGLMs); Joint generalized linear models (JGLMs); Signal-to-noise ratios; Structured dispersion.

---

[P-047-I-25]

### Efficient estimators using characteristics of Poisson distribution

*Prayas Sharma\**, University petroleum and Energy Studies, Dehradun, India, and *Rajesh Singh*, Banaras Hindu University, Varanasi, India

#### Abstract

This paper proposed a generalized class of estimators, exponential class of estimators based on adaption of Sharma and Singh (2015) and Solanki and Singh (2013) and simple difference estimator for estimating unknown population mean in case of Poisson distributed population in simple random sampling without replacement. The expressions for mean square errors of the proposed class of estimators are derived to the first order of approximation. It is shown that the adapted version of Solanki and Singh (2013), exponential class of estimator, is always more efficient than usual estimator, ratio, product, exponential ratio and exponential product type estimators and equal efficient to simple difference estimator. Moreover, the adapted version of Sharma and Singh (2015) estimator are always more efficient than all the estimators available in literature. In addition, theoretical findings are supported by an empirical study to show the superiority of the constructed estimators over others with earthquake data of turkey.

*Keywords:* Auxiliary attribute, point bi-serial, mean square error, simple random sampling, Poisson distribution.

---

[P-048-I-26]

### Dealing with treatment-confounder feedback and sparse follow-up in longitudinal studies - An application of the marginal structural model with a multiple sclerosis cohort

*M. Ehsan. Karim\**, *Helen Tremlett*, *Feng Zhu*, and *Elaine Kingwell*, The University of British Columbia, Vancouver, Canada, and the MINT Study Group

**Abstract**

The interferon-betas (IFN $\beta$ ) are widely prescribed platform therapies for patients with multiple sclerosis (MS). Very few studies have investigated the potential survival advantage associated with IFN $\beta$  exposure among patients with MS, and none have been able to take potential treatment-confounder feedback into account. The patient's or physician's decision to initiate, continue or stop these treatments may be affected by the complex dynamics of the patient's disease progression and comorbidity status, and vice versa. Longitudinal observational data are required to assess this association adequately. These data provide the opportunity to account for potential treatment-confounder feedback between comorbidities, MS disease progression and IFN $\beta$  exposure. In real-world applications, sparse follow-up due to irregular patient visits provides one of the biggest challenges in longitudinal analyses. We accessed a cohort of patients with relapsing onset MS from British Columbia, Canada (1995-2013) to examine the potential survival advantage of IFN $\beta$  exposure using a marginal structural model, and considered several imputation approaches appropriate for longitudinal data to deal with sparse follow-up.

*Keywords:*

---

[P-053-I-27]

**Assessing the livelihood status of fishermen at Sunamganj district in Bangladesh**

*Mohammed Amran Hossain*, Deputy Commissioner Office, Sunamganj, Bangladesh, *Shahnaj Sultana Sathi*, *Md. Sabbir Hossain*, and *Mohammad Ohid Ullah\**, Shahjalal University of Science and Technology, Sylhet, Bangladesh.

**Abstract**

Bangladesh is a riverine country. A large number of people settled in the river banks and their livelihood depends on fishing. About 17 million people (1.4 million women) depending on fisheries sector for their livelihoods through fishing, farming, fish handling and processing. Most of the lands in Sunamganj are haors and cannals and therefore many people in this area are involved in fishing. It's a rising sector for income generation and employment and its contribution of GDP is 4.42%. But fisherman does not get proper benefit and fishery sector does not getting proper attention though government enacted Jalmahal Nitimala, 2009. In this study we aimed to assess the livelihood status of fishermen at Sunamganj. For this, we randomly collected data based on a questionnaire from 425 fishermen during April 2018. Though 96.4% was registered fishermen, most of them took loan (40.6%) from the local businessmen. The financial condition of fishers was observed very poor as the land owned by them was decreasing day by day. We found that earning members had influence ( $OR = 1.77$ ,  $CI : 0.965 - 3.272$ ) on taking loan, indicates that households with only one earning member are 1.77 times as likely to take loan than household with more than one earning members. About 89.9% fishermen were afraid about their future earnings due to the flood. Taken together, we may conclude that overall situation of the livelihood status is not so good because of more illiteracy, more loan and natural disasters like flood. So government and non-government organization should play role to improve their economic status by

providing well education to their children, give more incentives so that they don't need to take loan and building embankment or dam for protecting flood.

*Keywords:* Livelihood status, fishermen, Sunamganj

---

[P-061-I-28]

### **ARIMA with intervention with GJR-GARCH in DSEX Index**

*Md. Monimul Huq\** and *Md. Ayub Ali*, University of Rajshahi, Bangladesh

#### **Abstract**

Intervention analysis is the modeling to control the natural or man-induced interventions on the mean level of a time series. The mathematics of intervention analysis is presented followed by a systematic method of analysis that is used to analyze in various field but intervention in asymmetric GARCH model is rare in previous study. Here, we used Dhaka Stock Exchange Broad Index (DSEX) data from January 27, 2013 to June 30, 2019. Dhaka Stock Exchange Broad Index includes all stocks must have a float-adjusted market capitalization above 100 million BDT. Additionally, if a current index constituent falls below the 100 million BDT threshold, but is no less than 70 million BDT, then the stock remains in the index provided it also meets the other inclusion criteria. Based on AIC, BIC and HQ comparison, ARIMA with intervention with GJR-GARCH model was performed better than other GARCH family models. Testing for the robustness of this outcome, we showed that the results were highly time dependent. The forecasting analysis divulged the usefulness of the proposed model for clarifying the values of the DSEX. The present proposed model should be widely used in modeling and forecasting of financial volatile data.

*Keywords:* Intervention, ARIMA, GJR-GARCH model, DSEX.

---

[P-063-I-29]

### **A three-parameter poisson-weighted lindley distribution with properties and applications**

*U. H. Rahman\**, Assam University, Assam, India, *R. Shanker*, *Devendra Mishra*, Institute of Statistics, Patna, India, and *T. D. Roy*, Assam University, Assam, India

**Abstract**

A three-parameter Poisson-Weighted Lindley distribution has been proposed which includes several one parameter and two-parameter discrete distributions. A general expression for the  $r$ th factorial moment has been derived and hence moments about the origin and moments about the mean have been obtained. Moments based measures including coefficient of variation, skewness, kurtosis and index of dispersion have been studied. The estimation of parameters has been discussed using maximum likelihood estimation. Some examples of discrete datasets have been presented to test the goodness of fit of the proposed distribution over other one and two-parameter discrete distributions.

*Keywords:* Poisson-Weighted Lindley distribution, Poisson-Lindley distribution, Moments, Maximum Likelihood Estimation, Goodness of fit.

---

[P-064-I-30]

**Prevalence of malnutrition and anaemia among tribal women in India: A study on NFHS-4 Data**

*Susmita Bharati\**, *Manoranjan Pal*, and *Premananda Bharati*, Indian Statistical Institute, Kolkata, India

**Abstract**

In India, Scheduled Tribe (ST) population consists of 8.2 per cent and they are mostly isolated from the general population. Very recently, national level (NFHS-4) data, collected during 2015-16 shows that malnutrition (underweight and overweight together) consists of 32.9 per cent. Again, 49.5 per cent, among adult tribal women, are either malnourished or anaemic. The sample size in this paper is 96100 of 15-49 year adult non-pregnant women. For this study, the dependent variables are Body Mass Index (BMI) for assessing the nutritional status and haemoglobin level for anaemia. To see the differential effects on BMI and anaemia, the independent variables are taken as type of residence education, religion, food habits and wealth index of the family. The result shows that, in India, the prevalence of under nutrition (CED) and over nutrition are 20.8 and 12.1 percent respectively. Middle belt of India is the mostly affected region of CED. And high overweight or obesity zones are concentrated in South zone and then followed by North-east zone. In case of anaemia, the highest concentrated zone is East zone and then followed by Central zone while the lowest anemic zone is North-east zone. Subsequent analysis shows that, at present, under nutrition among tribal women is more or less similar to that of the all India women, but percentage of obese are 6 to 7 unit less than that of all India women. In case of anaemia, it is a bit less than the corresponding all India level. Beside socio-economic factors; low cost calorie based food in semi-urban areas, different cost-free health facilities from the government and efforts of NGOs may have significant influence in reducing malnutrition among the tribal women in India.

*Keywords:*

---

[P-067-I-31]

**A new approach to determine the coefficient of skewness and an alternative form of boxplot**

*Ummay Salma Shorna* and *Md. Forhad Hossain\**, Jahangirnagar University, Savar, Dhaka, Bangladesh

**Abstract**

To solve the problems in measuring coefficient of skewness related to extreme value, irregular distance from the middle point and distance between two consecutive numbers, "Rank skewness" a new measure of the coefficient of skewness has been proposed in this paper. Comparing with other measures of the coefficient of skewness, proposed measure of the coefficient of skewness performs better specially for skewed distribution. An alternative of five point summary boxplot, a four point summary graph has also been proposed which is simpler than the traditional boxplot. It is based on all observation and give better result than the five point summary.

*Keywords:* Skewness, Rank Skewness, Boxplot, Four Point Summary Graph.

---

[P-072-I-32]

**Inequalities in farm sector labour force in India**

*Sonali Chakraborty\**, Indian Statistical Institute, Kolkata, India

**Abstract**

The back bone of rural economy in India is the farm sector or agricultural sector. It also contributes an overall economic and social development towards the nation. Farm activities include agriculture (crop production), plantation, animal husbandry (milk, meat, egg etc), forestry and logging and fishing. In India the lion share of labour force are engaged in farm sector. Though the female labour force participation is typically low in our country their share in agriculture is commendable and not decreasing as has been noticed in all other sectors in the recent past. But whether this participation in true sense is in gain full economic activities is to be looked into. Wage discrimination, stereotyping of works, working hours are some inequalities that females are facing at work place. In this analysis an attempt has been made to see the inequality in male female participation in farm sectors. Their wage differentials in rural urban sectors, in different states has been computed following NIC (National Industry Classification) and NCO (National Classification of occupation ) using unit level data of NSS(68th round 2011-12 and Periodic labour force survey 2017-18). From preliminary observations it is noted that there is an increase of rural female as market oriented crops producers or animal producers and related workers from 2011-12 to 2017-18. The NIC gives a more vivid picture, where they mainly engaged in mixed farming, plant

propagation, post-harvest activities along with perennial and non-perennial crop production. Reduction in rural male in all these industries are noted as an opposite criterion. In both rural and urban sector gender inequalities persists in wages in almost all NIC activities. Some exception like raising of cattle, supporting activities for crop production, fresh water aquaculture, growing of bush fruits and nuts where average earnings of females are little bit higher than male workers working as either casual labour or regular waged salaried person. The female to male wage ratio is minimum in case of MNREGA workers.

*Keywords:* Wage inequality, gender stereotyping, Agriculture labour, activity status.

---

[P-082-I-33]

### **Does road connectivity improve access to healthcare? - A study in rural Meghalaya**

*Utpal Kumar De\**, North Eastern Hill University, Shillong, India

#### **Abstract**

This paper examined the impact of rural road connectivity on the accessibility to healthcare services by the people of remote and isolated villages in Meghalaya. Primary data have been collected from the rural areas of Meghalaya where scattered villagers face varied connectivity to different desired destinations, particularly of health and education centres. We then constructed a Road Development Index, which has been related to various healthcare indicators across the villages. The findings confirmed that households in villages with good road connectivity have better access to various healthcare centres like public health centres, government hospital and private hospitals for delivery of baby and other related post-natal cares. Apart from their socio-economic conditions, reduced transportation costs help in accessing more medical help during sickness and better implementation of related immunization programmes like (BCG, DPT, Polio, measles), post-delivery care and female sterilization significantly for the households with good road connectivity. Further, it also indicates that with better road connectivity, incidence of various sicknesses like malaria, typhoid, tuberculosis and others are reduced significantly.

*Keywords:* Road Connectivity, Road Development Index, Access to Healthcare, Meghalaya

---

[P-087-I-34]

### **Determining the best modelling approach for forecasting height yield of hybrid spruce in the tree improvement program**

*Suborna Shekhor Ahmed\**, University of British Columbia, Canada



**Abstract**

The Canadian forest industry is facing increased competition for forest land from urban expansion and other resource sectors, including agriculture, oil, and gas, resulting in a diminishing fiber supply. At the same time, climate changes are impacting forests via changes in forest growth rates and natural disturbances. Using improved stock can ameliorate timber supply shortages and facilitate a more competitive forest sector while continuing to provide a broad range of ecosystem services by supporting sustainable forest management. In tree improvement programs, provenances are planted at several sites, and the growth and yield of trees are measured over time to specify the best performing trees. To assess the performance of provenances on a planting site, an accurate forecast to harvest age would be beneficial. In this research, an NLMM (nonlinear mixed-effects model) for the height yields of hybrid spruce (*Picea engelmannii* Parry ex Engelmann  $\times$  *Picea glauca* (Moench) Voss) provenances was developed. Then, alternative methods to forecast yields were compared, specifically using: (i) population-averaged forecasts based on the fixed-effects part of the NLMM model; (ii) subject-specific (i.e., a provenances within sites) forecasts using the random effects with the available repeated measures; and (iii) forecasts using the population-averaged model along with prior repeated measures in the autocorrelation model. Forecast accuracies were negatively impacted by having a less accurate fixed-effects part of the NLMM model, indicating the importance of careful selection and testing of the fixed-effects part of the model. However, the subject-specific forecasts were comparatively much less affected when a less accurate fixed-effects part of the model was used. The subject-specific forecasts were performing the best with accuracies similar to the measurement precision using standard height measurement devices given five or more prior measurements.

---

[P-088-I-35]

**Low birth weight is the consequence of maternal psychiatry, nutrition and gestational age: Results from a marginal model approach**

*Musammet Rasheda Begum\** and *Soma Chowdhury Biswas*, University of Chittagong, Bangladesh

**Abstract**

Low birth weight (LBW) babies occur commonly worldwide and face greater risks in later stages of life. This study aimed to know the prevalence and potential risk factors of LBW considering the clustering effects. A longitudinal study carried out throughout pregnancy and postpartum for child birth weight during September 2015 to August 2017. Generalized estimating equation (GEE) with independence, unstructured, exchangeable and auto regressive first order was applied to explore clustering effects. The prevalence of LBW (LBW < 2500 gm) was 21.2% (95% CI: 18.69 to 23.82). Smaller QIC value in GEE with independence working correlation structure indicated that the observations within cluster were not correlated. Anxiety in first trimester, hemoglobin and gestational age were associated with LBW. A non informative cluster size confirmed the response among observations in a cluster was independent. Since

clustered observations violated the assumption, a multiple logistic regression was employed to explore more potential risk factors. The model showed that mother with anxiety in third trimester were more likely to give birth of LBW babies (OR: 6.458, CI: 2.073-20.118) than non-anxious ones. Underweight mothers were more expected (OR: 2.54, CI: 1.786-3.589) to have LBW newborn than normal weighted women. The probability of LBW was lower (OR: 0.042, CI: 0.006-0.292) for term birth ( $> 37$  weeks) than preterm birth ( $\leq 37$  weeks). Mothers' psychiatry, nutrition and gestational age were associated with LBW of babies. Proper intervention emphasized on reduction of maternal mental stress, proper nutrition and maintaining the full gestational age should be taken to reduce the rates of LBW neonates.

*Keywords:* LBW, factors, GEE.

---

[P-090-I-36]

### **Violence against women in West Bengal: Its extent and causes**

*Chaiti Sharma Biswas\** and *Manoranjan Pal*, Indian Statistical Institute, India

#### **Abstract**

Violence against women is a common phenomenon in orthodox Indian society. Here, since their childhood, women are trained up to tolerate violence against them. Though it is an age old problem, but it come to the spotlight after Beijing world conference. Since then various countries take numerous steps to eradicate this crime from the society. Indian Government and NGOs also take several measures to abolish this crime, but it continues to persist in the society still now. In this background, this study attempts to locate the deep rooted factors responsible for such type of crime against women. Violence related variables found from the project funded by ICSSR are clubbed into six groups namely verbal abuse, threats, control and isolation, economic exploitation, physical assault and sexual assault in this study. It is found that women mainly experience "control and isolation" in their married life, which is followed by "verbal" and "physical" abuse. More than 30 percent married women face threat and economic exploitation. Percentage of women facing sexual assault is found to be very low. It is possibly due to the lack of disclosure. Rural women are tortured more often than that of the urban women. Also, violence against women decreases with age, though this rate of decrease is small. Relative to educated women illiterate women face more violence. It is surprising to note that working women experience more violence than non-working women. Non-payment or part payment of dowry is also an important factor of violence. Increasing the level of education and decision making power of women seem to be the main two solution for reducing violence against women. Moreover, girls and women should be trained to learn self-defense.

*Keywords:* Violence, Women, India

---

[P-091-I-37]

**Modeling the stability and determinants of household food insecurity: A pair copula construction approach**

*Jemal Ayalew Yimam\**, Wollo University, Dessie, Ethiopia, and *John O. Olaomi*, University of South Africa, Florida, South Africa

**Abstract**

Assessment of household food security status as well as determinate predictors using cross sectional data does not reflect the entire situation of it, rather, the stability over time is the key component to assure household food security. The aim of this study was to assess the stability of household food security over time and its determinant factors using longitudinal data through pair copula construction approach. This longitudinal study uses household food-security status, environmental and climate change related data collected three times at six months interval from June 2014 to 2015. The study participants were 646 farmer households randomly selected from selected Weredas of the South Wollo Zone of Amhara regional state, Ethiopia. The status of household food security determined using both the quartile and composite food security index for each rounds. Pair copula construction cumulative logit model assessed the stability of household food security status over time and the determinants of it simultaneously. For parameter estimation of the parameter and standard error full maximum likelihood method was used. The copula parameter revealed that individual household food security status is not stable over time. Moreover, the marginal parameter indicated that presence of crop disease; market price increase and medium weathering condition were the significant recurrent factors for households to be chronically to mildly food in-secured throughout the study period. One time cultivation per year was the temporal significant factor for household food insecurity. The results of this model suggest that household food security is not stable over time, so great attention is required for granting households to be food secured taking valuable intervention for the identified recurrent determinant factors. Moreover, the model provided easily interpretable and understandable outputs, thus, we suggest the model for any longitudinal discrete data analysis.

*Keywords:* Chronically Food In-secured, Longitudinal Ordinal Outcome, Cumulative Logit, Pair Copula Construction, Full Maximum Likelihood

---

[P-096-I-38]

**Divide and recombine approach for big data analysis and prediction of trajectory probabilities from repeated measures data using machine learning algorithms**

*R. I. Chowdhury\**, *M. Ataharul Islam*, University of Dhaka, Bangladesh, and *T. Hasan*, University of New Brunswick, Canada

**Abstract**

Due to the advancement of electronic data capturing system, the amount of data being collected and stored is ever increasing. This huge amount of data is complex and poses great statistical challenges in methodology, computation and theory. One approach to analyzing such data is to divide and recombine where large data set is broken into pieces in some sort of meaningful way and analyze each subset separately. Then results are recombined somehow in a manner that produces a statistically valid output. We proposed a simpler approach to analyze large datasets generated from repeated measures. In the proposed approach follow-up time is a natural conditioning variable that allows large data to divide into small subsets. Then using the relationship between joint, marginal and conditional probabilities we can recombine the results into a statistically valid way. A recombined result allows answering various questions that are required by researchers while analyzing repeated measures data.

*Keywords:*

---

[P-099-I-39]

**Analyzing the factor structure and sleep quality of Pittsburgh sleep quality index in indian information technology sector**

*Arijit Dutta\** and *Arindam Chatterjee*, Heritage Institute of Technology, Kolkata, India, and  
*Rimu Chaudhuri*, Heritage Business School, Kolkata, India

**Abstract**

The Pittsburgh Sleep Quality Index (PSQI) has gained widespread acceptance as a useful tool to measure sleep quality. In order to formulate the diagnosis process, it is essential that we understand the factor structure inherent in the PSQI data. In this work, we seek to estimate such a structure with a focus on the Indian Information Technology (IT) workers. We have used Confirmatory Factor Analysis (CFA) and the Exploratory Factor Analysis (EFA) for this purpose. Instead of using random imputation to handle the frequently occurring missing values, we have used more sophisticated techniques thereby improving the accuracy of our results. We have also used the Multi layer perceptron based method to see how we can classify the sleep quality of the sampled population. We have discovered that contrary to the general perception most Indian IT employees have sleep quality belonging good and very good classes.

*Keywords:* Exploratory and Confirmatory Factor Analysis, Pittsburg Sleep Quality Index, Goodness of Fit measure, Global Sleep Quality. Multi layer perceptron (MLP), Missing Value Handling.

---

[P-100-I-40]

**Comparing means of two correlated variables using pooled data of paired and independent samples**

*Nizam Uddin\**, University of Central Florida, Orlando, FL, USA

**Abstract**

A test statistic for testing equality of means of two correlated random variables  $X$  and  $Y$  is presented. The proposed test statistic is based on  $(n_1 + n_2 + 2n)$  pooled data of a sample of  $n$  paired  $(X, Y)$  observations, and independent samples of  $n_1$  observations on  $X$  and  $n_2$  observations on  $Y$ . Assuming that the joint distribution of  $X$  and  $Y$  is bivariate normal, the likelihood of  $(n_1 + n_2 + 2n)$  paired and unpaired sample observations is utilized to obtain the maximum likelihood estimate of all parameters involved. Properties of maximum likelihood method are utilized to approximate the null distribution of the proposed test statistic by scaled t-distribution. The proposed test statistic is evaluated empirically with respect to type I error and power against some other alternative methods available in the literature.

*Keywords:* Bivariate normal distribution, paired and independent samples, power, simulation, scaled t-distribution, test of equality of means, type I error.

---

[P-101-I-41]

**Impact of Differential Sectoral Output Growth on Environment: A Panel Data Analysis of Bangladesh**

*M. Zakir Saadullah Khan\**, Comilla University, Bangladesh

**Abstract**

Natural environment is central to economic activity and growth, which provide the resources economy needs to produce goods and services. The act of the laws of thermodynamics has implications for the economic systems - excessive use of natural resources, mostly in the form of energy, affects environmental quality; and poor environmental quality in turn affects economic growth by lowering the quantity and quality of resources. Energy used in economic activities results in emissions of carbon dioxide ( $CO_2$ ) that changes the natural environment. Different production sectors use varying volume of energy as inputs and in different intensity. The scale and composition of the economy, the change in share of different sectors in GDP, may have the potential to reduce the environmental impacts. This paper investigates the impacts of differential sectoral output growth of Bangladesh on  $CO_2$  emissions using panel data of major five sectors over the period 1990 to 2017. Analyses applied second generation panel data econometric techniques by using Augmented Mean Group analysis developed by Eberhardt and

Bond (2009) which takes into account parameter heterogeneity and cross-sectional dependence. Cross-sectional dependence is tested by CD-Bias test of Pesaran (2008), while Cross-Sectionally Augmented Dickey-Fuller test proposed by Pesaran (2007) is employed for testing for the presence of a unit root and cointegration among the variables is tested by using the methodology of Westerlund (2008). Result shows that distribution of energy use between sectors is the prime catalyst of  $CO_2$  emission. Uneven growth across sectors contributes significantly to lower  $CO_2$  emission. The role of manufacturing, power and transport sector GDP are more responsible for  $CO_2$  emission than agriculture and service GDP. The growing trend of economic activities should be planned in such a way that increases the energy efficiency of the production process, which can substantially reduce the level of  $CO_2$  emissions.

*Keywords:* Sectoral Output Growth,  $CO_2$  Emission, Panel Data Analysis.

---

[P-109-I-42]

### Relative efficiency of higher normed estimators over the least squares estimator

*Gopal K. Basak, Samarjit Das\**, *Arijit De*, and *Atanu Biswas*, Indian Statistical Institute, Kolkata, India

#### Abstract

In this article, we study the performance of the estimator that minimizes  $L_{2k}$ - order loss function (for  $k \geq 2$ ) against the estimators which minimizes the  $L_2$ - order loss function (or the least squares estimator). Commonly occurring examples illustrate the differences in efficiency between  $L_{2k}$  and  $L_2$ -based estimators. We derive an empirically testable condition under which the  $L_{2k}$  estimator is more efficient than the least squares estimator. We construct a simple decision rule to choose between  $L_{2k}$  and  $L_2$  estimator. Special emphasize is provided to study L4 estimator. A detailed simulation study verifies the effectiveness of this decision rule. Also, the superiority of the  $L_{2k}$  estimator is demonstrated in a real life data set.

*Keywords:*

---

[P-111-I-43]

### Effect-size measures for binary outcome variables and their estimation methods in meta-analysis for public health studies

*Shahjahan Khan\** and *Muhammed Ashraf Memon*, University of Southern Queensland, Toowoomba, Australia

**Abstract**

The effect-size is the name given to a family of indices that measure the magnitude of effectiveness of a treatment or intervention on the subjects. Depending on the type of study there are various measures that can be used to determine the effect-size for the intervention of interest. Statistical meta-analysis has been extensively used in public health to synthesise published summary data on specific effect-size from a number of independent studies in order to make credible and scientifically valid conclusions. Selection of measures of effect-size is the key in performing meta-analysis. The effect-size measure broadly depends on the type of outcome variables involved. For binary or categorical outcome variables effect-size is measured by proportion, relative risk or risk ratio and odds ratio. This presentation introduces various effect-size measures for binary outcome variables, discusses their estimators, sampling distribution of the estimators and confidence intervals. Some examples from recent literature on randomised control trials in medical studies are used to illustrate the estimation methods of different effect-sizes and their applications in public health.

*Keywords:* Meta-analysis, effect-size, binary outcomes, estimation, public health.

---

[P-115-I-46]

**Improvement of the Performance of Clustering Methods Using Outlier Modification Rule**

*Md. Rabiul Auwal\** and *Chongqi Zhang*, Guangzhou University, China, and *Md. Shahjaman*, Begum Rokeya University, Bangladesh

**Abstract**

Today's world is sinking by big amount of data. Data mining algorithms is a vital process to solve these problems. The Clustering methods are the greatest commonly used data mining models that are widely used to extract valuable knowledge from unsupervised data. There are some popular clustering methods in the literature for extracting valuable knowledge from big amount of data. Nevertheless, most of them often produce misrepresentative results in presence of outliers. Therefore, in this study, we announce a robust approach to overawe the problems of classical methods. We use median and median absolute deviation (MAD) for our robust procedure. In this procedure, a gene was considered as outlying gene if at least one of the expressions of this gene does not belong to a certain interval of the proposed outlier detection rule. Otherwise, this gene was considered as a non-outlying gene. We investigate the performance of the proposed method in a comparison of the traditional method using both simulated and real gene expression data analysis. From a real colon cancer gene expression data analysis, the proposed method gives the highest accuracy. For simulated data, the proposed method gives approximately the same result as with original data in presence of outliers. The simulation as well as real cancer gene expression data sets results show better performance with our proposed procedure.

*Keywords:* Gene expression data, Clustering, Outlier detection and modification, DE gene, MAD and Robustness.

---

[P-116-I-45]

### Deterministic and Stochastic Modeling of the *Drosophila* Circadian Clock

Md. Mamunur Rashid\*, Hiroyuki Kurata, Kyushu Institute of Technology, Japan

#### Abstract

Circadian rhythms are omnipresent in cellular organisms, generated by interlocked transcriptional-translational feedback loops. It produces phase shift, with the robustness of oscillation period, amplitude and phase, against external and internal variations. However, it is not clear exactly how these oscillation properties are achieved simultaneously. What is the underlying design principle of such an interlocked network? To address this question, in this study, we present two distinct coupling models of *Drosophila* circadian clocks; activator coupled oscillators (ACO) and repressor coupled oscillators (RCO). By examining the phase, period, and amplitude, we identified two key parameters: **coupling dissociation constant** and **coupling time-delay** which play important roles in the robustness of the oscillatory properties. In particular, ACO can produce anti-phase oscillators with applied subtle asymmetry kinetics, whereas RCO provides in-phase oscillators. However, compared to RCO, ACO can generate greater fluctuation in period and amplitude of the coupled oscillator depending on different sets of dissociation coupling constant along with coupling delay. Also, ACO shows fragility in phase, period, and amplitude to external noise compared to RCO. In addition, the ACO model shows significant variance in the intrinsic period and amplitude as dissociation coupling constant changes in the stochastic environment, whereas RCO ensures an almost robust period and amplitude. Moreover, we have shown that dissociation coupling constant significantly affects the entrainability of the ACO to the day-night master clock, whereas RCO produces almost constant entrainment irrespective of the coupling strength. The ACO generates an anti-phase oscillator, while reducing the robustness of period and amplitude and the entrainability to the day-night cycle.

*Keywords:* Circadian rhythms, Interlocked feedback loop, Phase difference, Coupling strength, Coupling delay, Robustness

---

[P-130-I-46]

### Estimating the hazard functions of two alternating recurrent events in the presence of covariates

Sugata Sen Roy\*, University of Calcutta, India



**Abstract**

The motivation for this paper is a cystic fibrosis data which records a patient's times to relapse and times to cure under several recurrences of the disease. The idea is to study the impact of covariates on the hazard rates of two alternately occurring events. The dependence between the times to the two events over the different cycles is modeled through an autoregressive type set up. The partial likelihood function is then derived and the estimators obtained. The estimators are shown to be consistent and asymptotically normal.

---

[P-131-I-47]

**Spatial optimal design for habitat condition models**

K. S. Bakar\*, CSIRO, Australia, A. H. M. MAhbul Latif, University of Dhaka, Bangladesh

**Abstract**

This paper develops a method for optimal spatial design from remote sensing data to inform policy decisions related to habitat condition assessment system (HCAS) in Australia. One of the key issues of the HCAS output is to determine the reference pixels of the remote sensing data, which are assumed to be in a good condition and are distributed based on homogeneous clusters. Some clusters, however are very large and contain thousands of reference pixels, whereas, some are very small with as minimum as 20 pixels. Thus, a habitat condition output may be biased towards the clusters with large reference pixels. Hence, the proposed spatial optimal design for the large homogeneous cluster aids to solve the problem and initiates a better understanding of the habitat conditions. We explore the proposed method with simulations and with a HCAS case-study region in Australia. The method can be applicable to any similar situations, where a spatial optimal design is required.

*Keywords:* Optimal design, remote sensing, habitat condition.

---

[P-132-I-48]

**Regression-type Estimators in the Presence of Non-response**

Dulal Chandra Roy\*, University of Rajshahi, Bangladesh

**Abstract**

In this paper, we have constructed, employing a reasonable predict format, certain imputation-based regression-type estimators of the population mean in the presence of non-response. These estimators are structurally explicit and meaningful. We have compared them with the existing ones due to Rao (1990) and Hansen Hurwitz (1946), and shown that, compared with four estimators due to Rao (1990), the two of the proposed estimators are superior while the other two are identical with Rao's. An example has been given to highlight the viability of the proposed estimators. We have also find the variance estimator in respect of each of the four proposed estimators.

---

[P-134-I-49]

**iArsenic - Instant arsenic screening of hand pump tubewells in Bangladesh using a statistical model embedded in a Web application**

Jacek Kopecky\*, Dillon O'Shea, Fionn Plant, Lewis Newton, Mohammad Hoque, Jim Smith, Shubhankar Dam, University of Portsmouth, UK, M. Atikul Islam, Kazi Matin Ahmed, University of Dhaka, Bangladesh, Adrian Butler, Simon Parker, Imperial College London, UK

**Abstract**

Groundwater is an attractive source of drinking water particularly in developing countries, as it can be supplied untreated. To avoid waterborne pathogenic diseases, governments and NGOs encouraged installation of millions of tubewells, low-cost mechanical hand-pump wells, in Bangladesh. This has inadvertently exposed ca. 40% consumers to arsenic, which is naturally present in the groundwater. Using over 10 million tubewells, over 20 million Bangladeshi people are exposed to this insidious toxin. Switching to a safe well is a major mitigation option for rural inhabitants, but 6 million wells are yet to be tested. Screening of these untested wells requires trained personnel with a chemical test kit that are outside reach of most rural inhabitants, and testing kits typically produce toxic by-products. Building on existing research, we use computer modelling and digital technology to solve this issue. We used the existing dataset of ca. 4 million tubewells in a web based application to estimate arsenic levels in untested (or disremembered) wells, based on a geochemical indicator (staining of the tubewell platform) coupled with a 3D location indicator (produced using the well depth and street address) input by the user. Provided with the above-mentioned indicators, our model gives the user a high-confidence estimate of arsenic risk in their tubewell. Communicating the research directly with the user, particularly in an actionable format, is a key factor overlooked so far in arsenic mitigation efforts. The widespread internet coverage and increasing use of smartphones in Bangladesh provide a powerful opportunity both for gathering data and for disseminating information on arsenic pollution directly to the users. The app will not only allow users to make an informed decision about continuation of the use of the tubewell for drinking (saving people from ingesting arsenic); it will also help the government save screening costs, and adopt risk-based mitigation strategies.

*Keywords:* Arsenic, web application, groundwater, tubewell, Bangladesh.

---

[P-135-I-50]

### **Predictive Comparison of Vine Copula Models**

*Md Erfanul Hoque\** and *Elif F. Acar*, University of Manitoba, Canada

#### **Abstract**

Vine copulas are a popular tool for flexible and tractable specification of high dimensional joint densities for representing multivariate data. There are various vine constructions such as D-Vines, C-Vines and more general R-Vines. In recent years, these models have been considered in regression contexts to predict conditional mean and conditional quantiles of a variable of interest given the other variables. In this work, we compare the predictive performance of various vine constructions models and evaluate the predictive utility of vine copula regression over classical regression models through simulation studies and real data applications.

*Keywords:* Copula regression, prediction error, vine copulas.

---

[P-137-I-51]

### **Tobacco use among Bangladeshi secondary school students: A comparison between 2007 and 2013 GYTS surveys**

*M. J. Uddin* and *T. Ahammed\**, Shahjalal University of Science and Technology, Sylhet, Bangladesh

#### **Abstract**

Using tobacco, usually begins in adolescence, is burning healthcare and economical issue, especially in the developing countries. We aimed to examine the changes in the pattern of smoking over time and whether these changes are uniform for socio-demographic factors. We used two nationally representative surveys, named the Global Youth Tobacco Survey (GYTS) conducted in Bangladesh (the year 2007 and 2013). GYTS was a school-based survey of students targeting adolescents age 13-15 years developed by the World Health Organization and the Centers for Disease Control and Prevention. Data were analyzed using logistic regression models and chi-square tests. Our study showed that overall tobacco usage declined from 8.43% in 2007 to 7% in 2013. Between 2007 and 2013, the prevalence fell from 41.09 to 24.3% among grade nine or ten students and from 37.15 to 14.87% among female students. However,

the prevalence among males, 13 to 14 years old, studied in grade seven or eight was increased. Awareness about second-hand smoking among participants (2007: 90.54%; 2013: 90.61%), class discussion against usages of tobacco (2007: 59.95%; 2013: 67.84%) were significantly increasing and the rate of offering free tobacco (2007: 8.70%; 2013: 6.15%), opinion on banning smoking in the public place (2007: 73.89%; 2013: 64.69%) receiving anti-tobacco messages (2007: 97.65%; 2013: 84.95%) and tobacco advertisement (2007: 93.4%; 2013: 88.88%) were significantly decreased with time. Overall male students (OR: 2.078, 95% CI: 1.37-3.15, p-value = 0.0006), students older than 14 years (OR: 2.094, 95% CI: 1.044-2.539, p-value = 0.0121), studying in higher grade, against banning smoke (OR: 1.628, 95% CI: 1.18-3.73, p-value = 0.0315) and exposed to second-hand tobacco using significantly associated and had higher odds of using tobacco products. A combination of social, school and home-based antismoking interventions may be effective in controlling adolescent smoking in Bangladesh.

*Keywords:* Adolescents, School, Smoking prevalence, Tobacco use.

---

[P-145-I-52]

### Health Scenarios of Tribal women in West Bengal

Aditi Bardhan\*, Indian Statistical Institute, Kolkata, India, Poulami Bharati, Clubtown Residency, Kolkata, India, Susmita Bharati, Manoranjan Pal, Premananda1 Bharati, Indian Statistical Institute, Kolkata, India

#### Abstract

It is well-known that the health status of the tribal population is very poor. Here, national level (NFHS-4) data, collected during 2015-16 has been considered. Our objectives are (i) to study the prevalence and inter-division and inter district variations of under nutrition, anaemia and non-communicable diseases among the adult tribal women in West Bengal; (ii) to see the extent of deviations of these results from Indian tribal women and (iii) to show the relationship of under nutrition, anaemia and non-communicable diseases with some socio-economic variables. Here, the sample size is 658 of (15-49)-year-old non-pregnant tribal women. Dependent variables are under nutrition, anaemia and certain non-communicable diseases like Diabetes, Asthma, Thyroid, Heart-diseases and cancer. The independent variables are type of residence, education of women and also of their partners, religion and wealth index. The result shows that in WB, percentage of under nutrition and anaemia are 30.5 and 69.4 respectively. The prevalence of Diabetes, Asthma, Thyroid and Heart diseases are 1.7%, 2.6%, 1.1%, 2.0% respectively. None of the selected tribal has been seen to suffer from cancer. Division-wise, the highest prevalence of undernutrition is found in Medinipur division and the highest anaemic prone division is Malda followed by Burdwan and then Medinipur. When compared with the all India tribal women, it is seen that tribal women of WB suffer more by 10 and 20 percentage points respectively for under nutrition and anaemia. The reasons can be traced in the place of residence, literacy status, religion and wealth index. In case of Diabetes, Asthma, Thyroid and heart diseases; place of residence and wealth index are found to be the two most important factors. And these results are statistically significant.

---

[P-147-I-53]

**Machine learning application for geospatial big data: Case studies for wheat area mapping and brick kilns identification from satellite images.**

*Mir A Matin\**, *Varun Tiwari*, and *Suman Ghimire*, International center for Integrated Mountain Development

**Abstract**

The amount of data collected by environmental research has been grown tremendously. A large part of these is geospatial data that refers to remote sensing satellite images, in-situ data collected by various sensors and location-based data. The processing of these data is challenging in terms of proper algorithms and computing power. Machine learning has evolved as key tools for handling the large amount of Geospatial data. Machine learning is a branch of artificial intelligence that uses computers to learn from big data using statistical methods and adapt them without explicitly programmed. The machine learning algorithms are based on different learning models including regression, clustering, Bayesian, decision tree, artificial neural network (ANN) and support vector machine (SVM). In this paper, we present two case studies of using machine learning with remote sensing images. In the first case, we applied a random forest decision tree model to a large amount of time series optical and RADAR remote sensing images to estimate wheat cultivated areas in Afghanistan. In the second case, we applied deep learning models of convolution neural network with high-resolution satellite images to automatically detect brick kilns in the Hindukush Himalayan region. For the first case, we have achieved an overall accuracy of 83.8% and 89.0% for irrigated and rain-fed wheat area respectively. In the second case, the highest classification accuracy for detecting brick kilns was 93%.

---

[P-156-I-54]

**Farmers role and challenges to adaptation hinders sustainability in agricultural sector of Bangladesh: A study based on early flash flood areas in Bangladesh**

*Kanis Fatama Ferdushi\**, Shahjalal University of Science and Technology, Bangladesh, *Anton Abdulbasah Kamil*, Istanbul Gelisim University, Turkey, and *Mohammad Nayeem Hasan*, Shahjalal University of Science and Technology, Bangladesh

**Abstract**

Bangladesh remains one of the most vulnerable countries in the world to the effects of climate change. Given the reliance of a large segment of the population on the agricultural sector for both their livelihoods as well as national food security, climate change adaptation in the agricultural sector is crucial

for continued national food security and economic growth. Using household data from lowland rice farmers of selected haor areas in Sylhet, the current work presents an analysis of the determinants behind the implementation of different climate change adaptation strategies by lowland rice farmers. The first objective was to find out those items which were making hurdle for adaptation. To achieve that goal, respondents were interviewed with respect to climate change related barriers circumstances they faced in their daily lives. The item for measuring degree of challenges were calculated through principal confrontation matrix. The results show that unpredictable weather is ranked one as well as high cost of farm inputs is ranked as two. There were a lot of respondents (294) who had given their judgment as those variable was highly challenging for adaptation. The second objective of this study was to explore the extent of awareness of climate change within this population as well as the type of effective decisions whether to willingness of adaptation held by lowland rice farmers with respect to climate change. To serve this purpose, ordered logit model (LM) was subsequently employed to evaluate the attitudes of 378 farmers with respect to climate change vulnerability. This results also state that lack of access subsidies, change crop variety and crop to livestock were have positive impact on the behavioral intension (BI) that mean adaptation choices of the farmers with climate change but those were not significant. On the other hand, unpredictable weather and crop to livestock were had negative impact on the BI of the farmers. We had also found that the limited access to agri-market of the household head is significant at threshold level ( $p=0.05$ ) and limited access to agri-extension is also significant at ( $p < 0.05$ ). The results also implies that farmers with negative attitudes towards adaptation are less likely to adapt. This study will help to implement of new strategies and policy in the haor areas which are more vulnerable to climate change.

*Keywords:*

---

[P-158-I-55]

### **Analysis of interval censored competing risks data under missing causes**

*Debanjan Mitra, Ujjwal Das\**, Indian Institute of Management, Rajasthan, India, and *Kalyan Das*, Indian Institute of Technology, Bombay, India

#### **Abstract**

In this work, we discuss analysis of interval censored competing risks data when some of the causes of failure are missing. This problem has not been investigated much, albeit in lifetime studies such data occur frequently. The vertical modeling approach has been proposed here. This approach utilizes the data to extract information to the maximum possible extent especially when some causes of failure are missing. The maximum likelihood estimates of the model parameters are obtained. The asymptotic confidence intervals for the model parameters are constructed using approaches based on observed Fisher information matrix, and parametric bootstrap. A simulation study is considered in detail to assess the performance of the point and interval estimators. It is observed that the proposed analysis performs better than the complete case analysis. This establishes the fact that our methodology is an useful

technique for interval censored competing risks data when some of the causes of failure are missing. Such analysis seems to be quite useful for smaller sample sizes where complete case analysis may have a significant impact on the inferential procedures. Through Monte Carlo simulations, the effect of a possible model mis-specification is also assessed on the basis of the cumulative incidence function. For illustration purposes, a real data set is analyzed and the conclusion appears to be quite realistic.

*Keywords:* Interval censoring; Competing risks; Cumulative incidence function; Gompertz model; Maximum likelihood estimates; Confidence intervals.

---

[P-159-I-56]

**Internal consistency and structure of the Statistical Anxiety Rating Scale (STARS): A case study of SUST, Sylhet.**

*Farha Nusrat Zahan, Md Atiqul Islam, and Luthful Alahi Kawsar\**, Shahjalal University of Science and Technology, Bangladesh.

**Abstract**

The need for basic statistics knowledge is increasing day by day. However, many students find it difficult to figure out statistical concepts and have become one of the greatest challenges among students at all levels of study. Few studies related to statistics anxiety has been done in the regions other than English-speaking populations. No study has yet been done to examine the internal consistency and structure of Statistics Anxiety Rating Scale (STARS) in Bangladesh. This study aimed to examine the internal consistency and structure of the STARS at Shahjalal University of Science and Technology (SUST), Sylhet, Bangladesh. Data were collected from the undergraduate students who enrolled Statistics as a non-major course from January 2019 to June 2019 of SUST using the modified version of STARS questionnaire. The STARS was identified as a reliable tool for measuring level of statistics anxiety in Bangladesh. The internal consistency reliabilities, ranging from 0.70 to 0.86, were found to be satisfactory. Findings of the present study would be helpful for the students, teachers, even policymakers to design student-friendly curriculum.

*Keywords:* Statistics anxiety, STARS, Internal Consistency, Mental Health.

---

[P-160-I-57]

**A consultant's care book: Interesting cases from 17 years of statistical collaboration in Australia**

*Shahid Ullah\**, Flinders University, Australia

**Abstract**

The use of statistical methods in clinical investigations and public health settings has been increasing in the recent years. However, the accurate and robust statistical techniques are limited. The book brings together experts working in clinical and public health settings to present robust statistical methods and their applications, mostly in kidney health research. We outlined the methodological issues related to traditional matrices and consideration in robust statistical methods with care. We addressed some “new frontiers”, including the competing risk framework, the use of time varying exposures, the concept of joint modelling approach for longitudinal exposures and survival outcomes, and a series of parametric models along with semiparametric cox regression model. We highlighted the approach to improved risk communication and the value of graphical displays including spider map, radar plot and hot spots using ArcGIS mapping system. De-identified Australia and New Zealand Dialysis and Transplant Registry data were used for model building in health data. Stata codes are available from the author on request. Through collaboration with clinicians and health professionals in medicine and public health areas, this book contributes to a high-quality service to the patients, carers, health professionals, health organisations, policymakers, funders and other interested parties through application of appropriate statistical methodologies in health settings.

*Keywords:* Consultant book, robust statistical methods, medicine and public health, collaboration

---

[P-161-I-65]

**How far do people travel to dialysis? Time and distance travelled for metropolitan haemodialysis patients**

*Stephen McDonald, Shahid Ullah\*, Emily Duncanson, Kathryn Dansie, Aarti Gulyani, Chris Davies, and Shilpa Jesudason, South Australian Health and Medical Reserach Institute, Adelaide, Australia*

**Abstract**

Travel to haemodialysis (HD) treatment has been identified as a significant burden for HD patients, but data on Australian-specific travel distance and time to HD is lacking. To characterise the burden of travel to HD in metropolitan Australia. Using data from ANZDATA, distance and travel time from the population weighted centroid of residential postcode to treatment centre was estimated utilising Google Maps Directions API for 6,043 Australian patients receiving facility HD (hospital or satellite) in major cities at 31 December 2017. Median two-way travel distance to patients’ actual treatment facility was 18.3 kilometres, with substantial variation between states. Median estimated two-way travel time was 32.6 minutes. Times and distances were greater for those in the first year of treatment and substantial different between states. Half of the patients (50.0%) received treatment at their closest facility. Of those not treated at the closest centre, a substantial proportion were receiving treatment in a hospital-facility when the closer facility was a satellite unit For those treated in a satellite facility that was not the closest facility, only a minority of these were attributable to the closer facility being a hospital-based



one. Through application of geospatial technology to the renal replacement population, this study provides insight into a previously suspected but unmeasured aspect of patient experience. Facility HD patients in Australia have a substantial travel burden - the time estimates equate to a median of 4 days per year spent travelling to treatment in addition to the time required for treatment. The increased distance in the first year presumably reflects a combination of local capacity limitations and the need to access hospital- rather than satellite-based facilities. For patients (often aged and frail), the actual burden is likely to be greater than that estimated; many people will require support from family and friends with transport. The fact many people do not access treatment in their closest centre suggest potential for substantial reduction in this burden.

*Keywords:* Travel burden; time and distance; hospital-facility haemodialysis, Google API

---

[P-164-I-59]

### Testing the Mediation effects using bootstrapping in CB-SEM: The why and how?

*Mohammad Mizanur Rahman\**, *Md. Saidur Rahaman*, Metropolitan University Sylhet, Bangladesh, and *Samuel Babatunji Adedeji*, Universiti Putra Malaysia, Malaysia

#### Abstract

This research paper has tried to provide answers to some of the challenging issues relating to mediation analysis using bootstrapping in Covariance Based-Structural Equation Modeling (CBSEM). Particularly, an emphasis is given on four key burning aspects with reference to up-to-date literature on mediation analysis using bootstrapping with a practical example. Researchers have the benefits of updating their knowledge base from the current citations provided to carry out proper analysis based on informed decisions. This study has created awareness in relation to the drawbacks and applications of the CB-SEM as a statistical technique for mediation analysis in the behavioural and social sciences and its relevance in the future. In addition to the justification as a method for research analysis, the study derives its relevance from the need to address the challenges faced by researchers in the application of the technique and the required ways out of them by setting a practical example.

*Keywords:* Bootstrapping; CB-SEM; Mediation effects and Social science

---

[P-168-I-610]

**Correlates of childhood overweight/obesity in Bangladesh: A community-level cross-sectional study**

*Mohammad Sorowar Hossain\**, Independent University, Bangladesh, *Enayetur Raheem, Shameema Ferdous*, Biomedical Research Foundation, Dhaka, Bangladesh, *Mahbubul H. Siddiquee*, BRAC University, Bangladesh, and *Anthony D. Okely*, University of Wollongong, Australia

**Abstract**

Many low- and middle-income countries are facing a rising trend of childhood obesity in the past few decades. However, a little is known about the factors responsible for this increase in these settings including Bangladesh. This study aimed to determine the prevalence of overweight/obesity among children (4-7 years) in a district town, and to identify correlates associated with socio-demographic characteristics and adherence to physical activity guidelines. A cross-sectional study with stratified sampling was conducted using a self- or interviewer-administered questionnaire (N=585). Weight status and Body Mass Index of children were assessed according to CDC criteria. Binary logistic regression models were used to calculate crude and adjusted odds ratios to identify the correlates. The overweight or obesity was observed among 14% of children. Caesarean mode of delivery (AOR: 1.82; 95% CI: 1.02 to 3.34) was positively associated with overweight/obesity while children meeting the combined 24-hour movement guidelines were less likely to be overweight/obese (AOR: 0.20; 95% CI: 0.03 to 0.70). This is the first study to examine the factors responsible for childhood overweight/obesity at a district town level in Bangladesh. Caesarean delivery and not meeting the 24-hour movement guidelines were correlated with childhood overweight/obesity. These factors should be targeted in interventions and public health policies to reduce the increasing level of obesity in Bangladesh.

---

[P-178-I-61]

**The efficiency of bivariate Fay-Herriot small area estimators**

*Mossamet Kamrun Nesa\**, Shahjalal University of Science and Technology, Bangladesh, *Robert Clark*, Australian National University, Australia, and *Carole Birrell*, University of Wollongong, Australia

**Abstract**

Fay-Herriot estimators are widely used to produce small area statistics when only area-level aggregate data are available. This paper investigates the conditions under which bivariate Fay-Herriot models

give useful reductions in approximate prediction mean squared error (APMSE) compared to separate univariate models. The APMSE is shown to be equal under these two approaches if the sampling errors and the area-level random effects have proportional variance-covariance matrices, even if there is high correlation between the two variables of interest. The ratio of APMSEs is calculated numerically for a range of settings, and this numerical study is summarized using a novel regression tree approach. Univariate and bivariate estimators are compared using data on 30 indicators from the 2011-2012 New Zealand Health Survey, with MSEs estimated by a parametric bootstrap approach. The results suggest that bivariate modelling can be worthwhile, but only for a minority of cases.

*Keywords:* Small Area Estimation; Bivariate Fay-Herriot Model; Univariate Fay-Herriot Model.

---

[P-188-I-62]

### **Impact of pre-disclosed inducements on response rate and quality of primary data**

*Shahidul Islam\**, MacEwan University, Canada

#### **Abstract**

Inducements of different forms and at different stages are used for motivating people to participate in human subject research. Although it is accepted that inducements, in general, play a positive role in increasing participation rate, there are exceptions. Inducements may contaminate the quality of data and as such the research findings. Although inducements are used for increasing response rate, these can have opposite effects depending on the respondents. If the intention of data collection is purely for public goods and that message is clearly conveyed to the prospective respondents, a material inducement may not be needed to acquire appropriate data. Under such a situation, peoples' altruistic behavior takes precedence over psychological egoism.

*Keywords:* Primary data collection, human subject, inducement.

---

[P-189-I-63]

### **Spatio-temporal controls on salinization within the Ganges Delta, southwest Bangladesh**

*S. E. Feist\**, *M. A. Hoque*, *M. C. Bloor*, University of Portsmouth, UK, *K. M. Ahmed*, *M. A. Islam*, University of Dhaka, Bangladesh, and *A. M. Dewan*, Curtin University, Australia

**Abstract**

Human and ecosystem health within Asian deltas are at risk from rising salinity linked to climate change and anthropogenic activities. Southwest Bangladesh is identified as a region at particular risk due to reduced upstream discharge, low elevation and relative sea-level rise. However, a comprehensive and integrated understanding of the spatio-temporal controls on salinity within the delta is currently lacking and provides the focus for this presentation. A large amount of in-situ measurement data from a network of salinity ( $n = 54$ ) and water level ( $n = 19$ ) monitoring sites, maintained by Bangladesh Water Development Board, were collated and analyzed. Multiple regression modelling and Seasonal Trend Decomposition of hydro-morphological parameters including elevation, distance from the primary river source, and water levels have been used to constrain temporal trends and determine relative influence on salinity. Results show high salinity ( $> 5000 \text{ mg/L}$  of Total Dissolved Solids) is prevalent in the southwest of Bangladesh, associated with a region of low elevation ( $< 3m$  above mean sea level). Lower salinity ( $< 1000 \text{ mg/L}$ ) is observed where elevation exceeds  $3m$ , and is consistently  $< 500 \text{ mg/L}$  at elevations  $> 6m$ . Inter-annual variation within the southwest is high, with salinity  $> 12500 \text{ mg/L}$  extending up to 120 km inland. Multiple linear regression models show a strong relative influence of increased distance from the primary river source with increased salinity, which also correlates with lower elevation. Tidal water level trends also show a strong spatial distribution pattern, corresponding with three hydro-morphological regions of the delta. Spatial variation in salinity results from complex interactions between geomorphological and hydrological factors. Regional trends in salinization and tidal water levels are identified which are also frequently subject to localized modification of landform/land-use. The identification and quantification of regionally specific primary drivers will help guide adaption strategies and focus mitigation efforts on vulnerable areas.

*Keywords:* Salinization, tidal water levels, time series, hydro-morphology, salinity.

---

[P-191-I-64]

**Divide and Recombine Approach for Fitting Logistic Regression Model for Big Data**

*Md. Kamrul Islam, Md. Razanmiah, and Md. Rezaul Karim\**, University of Rajshahi, Bangladesh

**Abstract**

Divide and recombine (D&R) is a new statistical approach to the analysis of big data. In the D&R approach, the data are divided into manageable subsets, an analytic method is applied independently to each subset, and the outputs are recombined. This paper applies the D&R approach for fitting the logistic regression model for big data. As an example, it considers the Bangladesh Demographic and Health Survey (BDHS) 2014 data set for modeling Bangladeshi women's empowerment status at the household level. It is shown that the D&R approach has great potential in analyzing big data.

*Keywords:* Salinization, tidal water levels, time series, hydro-morphology, salinity.

---

[P-193-I-65]

### Empirical profiling of banks in distress in Bangladesh using logit model with panel data

*Shirin Sharmin*, LR Global Bangladesh Asset Management Company Ltd, Bangladesh, *Mohammed Azizur Rahman Shuman*, Sohar International Bank SAOG, Oman, and *Mohammad Arman\**, North South University, Bangladesh

#### Abstract

This paper studies the application of Altman Z-score bankruptcy model in predicting bank distress esp. in the context of Bangladesh. The modified model that is applicable for non-manufacturing companies in emerging economies was applied using market capitalization data of 30 listed commercial banks at Dhaka Stock Exchange (DSE). Using financial data from 2013-2018 the Z-score was calculated for all the thirty banks and the lowest Z-score was found to be  $-5.73$  while the highest was  $8.10$  with a median value of  $2.61$ , which suggests around 50% of the observations fell under **gray** or, **distress** zone. Later, an ordered logit model was developed using the panel data with the Z-score being the dependent variable, while type of bank, age since incorporation, number of branches, total loans and deposits being the explanatory variables. Acknowledging the fact that Z-score model, which was developed in the context of developed nations, is not readily applicable for banks in Bangladesh, the capital adequacy ratios (CAR) of these banks were calculated for year 2013-2018 and setting a minimum benchmark of 10%, the banks were classified as compliant or, not. Accordingly, a binary logit model was run on the panel data with explanatory variables, earlier set. Results are expected to help the regulators, such as the central bank, profile a commercial bank in distress with relative ease and take corrective measures accordingly.

*Keywords:* Altman Z-score, banks, Dhaka Stock Exchange (DSE), Capital adequacy ratio (CAR), logit model for panel data

---

[P-195-I-66]

### Probability models for identifying best among the novice candidates

*Papia Sultana\**, University of Rajshahi, Bangladesh

#### Abstract

In biomedical sciences, education, psychology, sociology and also in other fields, firms or institutions may need to select best candidates among the novice. These candidates can be persons, test methods or any other proper tool depending on the problem. In this paper, I will describe a probability model for

which will allow us to get agreement index between the candidates for classifying subjects into two or more mutually exclusive categories. In addition, the agreement index will allow us to identify the best one among the candidates. Other statistical inference, like estimation, distribution and test procedure will be described, too.

*Keywords:* Probability model, agreement index, estimation, distribution, test.

---

[P-198-I-67]

### **A Marginal Conditional Model for Covariate Effects on Depression Data**

*Jahida Gulshan\**, *Azmeri Khan*, and *M Ataharul Islam*, University of Dhaka, Bangladesh

#### **Abstract**

Elderly population suffer from repeated spells of depression that may change over time and result in other health problems and chronic illness. The literature on depression among elderly helped understanding the factors associated with depression and also the outcome of depression. In this study we analyzed Depression data from Health and Retirement Study conducted by University of Michigan using a marginal conditional model where effects of difficulties in activities of daily living and other covariates on depression status at different waves are of interest. The covariate effects on depression status disaggregated at different waves given the depression status at previous waves are expected to be helpful in understanding of depression and associated factors in a more explicit manner.

*Keywords:* Marginal Conditional Model, Depression, activities in daily living.

---

[P-204-I-68]

### **Global analysis of drug-target interactions via meta-predictor with convolution on protein sequences**

*Md. Mehedi Hasan*, Kyushu Institute of Technology, Fukuoka, Japan, *Md. Nurul Haque Mollah\**, University of Rajshahi, Bangladesh

**Abstract**

Drug-target interactions (DTIs) identification is one of the major challenges for drug discovery. The labor-intensive and high-cost nature of *in vivo* and *in vitro* experiments have highlighted the importance of *in silico*-based DTI prediction approaches. In several *in silico* models, conservative protein descriptors have been shown to not be adequately informative to predict accurate DTIs. Thus, in this study, we propose a meta-predictor based DTI prediction model capturing local residue patterns of proteins participating in DTIs. When we employ meta-predictor on raw protein sequences, we perform convolution on various lengths of amino acid subsequences to capture local residue patterns of generalized protein classes. We train our model with large-scale DTI information and validate the performance of the proposed model using an independent dataset that is not seen during the training phase. As a result, the proposed model performs better than previous protein descriptor-based models. It also outperformed the recently developed deep learning models for the massive prediction of DTIs. By examining pooled intricacy results, we assured that our model can perceive binding sites of proteins for DTIs.

*Keywords:* Drug-target interactions, Protein sequence, Meta-predictor

---

[P-208-I-69]

**Statistical phenomics in crop growth and development research**

*Md. Matiur Rahman\**, Bangabondhu Sheikh Mujibur Rahman Science and Technology University, Gopalganj, Bangladesh

**Abstract**

In coming decades, global cereal production must need to increase more than 70%-80% from current decades due to the consumption of food, feed, and fuel by a larger, more affluent population. The rapid progression of genomic technology has improved plant research. Recently, available components of “OMICS” approaches offered and integrated suite of new technologies, and we are in a path to improve the description of complex phenotype through high-throughput phenotype technologies. High-throughput plant phenotyping technologies have been accelerated by advanced image sensing and analyzing techniques. Development of this concept termed as plant Phenomics. The main goal of plant statistical phenomics is to bridge the genotype–phenotype gap. Bridging the genotype–phenotype gap also demands that large-scale biological (OMICS) data and associated bioinformatics resources. The advanced statistical algorithms are constantly supporting to bind together such genetic and phenotypic data. Researchers are facing a big challenge during the handling these type of dataset. However, using advanced data-mining (DM) and machine learning (ML) algorithms, it is possible to overcome analytical bottleneck of the Phenomics dataset. DM and ML are an inherently multidisciplinary approach to data analysis that draws inspiration, and borrows heavily, from statistics, probability theory, decision theory, optimization, and visualization and also perform well when **BigData** problem arise. Several image-based studies have used and evaluated DM and ML methods performance in biology, and images obtained in

highthroughput screening through a real-time phenotypic platforms and the generated dataset represent a BigData problem. However, we are interested to discuss the contributions of the statistical learning methods and algorithms in Phenomics study for crop growth and development research.

*Keywords:* Genotype, Phenotype, Image sensing, Phenomics, Data Mining, Machine Learning.

---

[P-210-I-70]

### Detection of toxic chemical compounds and their associated toxicogenomic biomarkers using robust co-clustering approach

*M. Nazmol Hasan\**, Bangabandhu Sheikh Mujibur Rahman Agricultural University, Bangladesh, *M. Nurul Haque Mollah*, University of Rajshahi, Bangladesh

#### Abstract

Detection of toxic chemical compounds (CCs) or drugs and their associated toxicogenomic biomarkers is one of the most important tasks in the toxicogenomic studies as well as in pre-clinical phase of drug development pipeline. There are few statistical methods for the detection of toxic CCs and their associated toxicogenomic biomarkers. For example, probabilistic hidden variable model (PHVM) and logistic probabilistic hidden variable model (LPHVM). These are the EM (expectation-maximization) based iterative approaches, which produce co-clusters between CCs and genes in absence and presence of outliers in the dataset to detect toxic CCs and their associated biomarkers. However, these approaches consume more time for computation and produces equal number of clusters for CCs and genes though there may be more clusters in the genes than the CCs. Because, a CCs cluster may up-regulate a set of genes and down-regulate another set of genes. Therefore, a CCs cluster makes more than a single co-cluster with the gene clusters. These limitations can be overcome applying co-clustering algorithm (CCA) on the hierarchical clustering (HC) approach. Nonetheless, the HC is also sensitive to outlying observations. On the other hand, toxicogenomic dataset often contains outlying observations. In this regard, the robust hierarchical clustering (RHC) together with CCA can be applied for the detection of toxic CCs and their associated toxicogenomic biomarkers in absence and presence of outlying observations in the dataset.

*Keywords:* Chemical compounds, Toxicogenomic biomarkers, Logistic probabilistic hidden variable model, Co-clustering algorithm, Robust hierarchical clustering.

---

[P-214-I-71]

### Breeding design of important economic crops by Statistical analysis of NGS Data

*Md. Mamun Monir\**, Zhejiang University, China, and *Md. Nurul Haque Mollah*, University of Rajshahi, Bangladesh



**Abstract**

Optimizing crop yields through artificial genetic selection to improve economically important traits is the ultimate goal of crop research. With the advance of next-generation sequencing (NGS) technologies, the identification of the genes associated with complex traits is a huge challenge due to different types of genetic and epigenetic factors. Single nucleotide polymorphisms (SNP), insertion and deletion of the short sequence identified from whole-genome sequence data, are important genetic variants that could associate with many complex traits of living organisms. Interactions between the genetic variants and environmental factors could also largely control the complex traits. Same genotype of a genetic variant could behave differently in different environments or geographical locations due to their interaction with environments or locations. Gene-environment interaction analysis is crucial for crop trait analysis in developing environmental specific high productive crops. This may also help to design superior hybrid crops. The computational complexity in analysis increases with the increasing number of these genetic variants and epigenetic factors. Appropriate statistical genetic models and advanced computational technology techniques are needed to identify important genetic and epigenetic factors. Finding the best genotype combination for multiple loci can improve the quality of the trait or production, which in turn brings another complexity. In this paper, we review statistical models suitable for analyzing crop complex traits, as well as methods for finding good genotypes for multiple loci to obtain optimal crop yields.

*Keywords:* Crops, Breeding design, Next generation sequencing, Complex trait and Single nucleotide polymorphisms.

---

[P-216-I-72]

**Evaluation of sample size requirements for developing a risk prediction model**

*G. Ambler\**, *Q. Chen*, *M. Pavlou*, and *R. Z. Omar*, University College London, UK

**Abstract**

Clinical prediction models are routinely used in practice for prognosis or diagnosis. Prediction models are often developed using regression models and it is important that the development dataset is of an appropriate size. In particular, if the development dataset is too small, the resulting model may fit the development data too well (overfitting) and predict poorly in validation data.

Therefore, there is a need for clear sample size guidelines in this area. Until recently, the main guideline was the ‘rule of ten’ that suggests that at least 10 events per predictor variable (EPV) are required. However, recent work by van Smeden (2018) and Riley (2018) suggest that other factors, in addition to EPV, need to be taken into consideration.

In this talk, we review the current research in this area, then present the results of simulation studies based on real data to investigate the effects of various factors including EPV, model discrimination and outcome prevalence on risk model performance. In contrast to much of the research in this area, our

focus is on both average model performance over simulations and the variability around that average, rather than just the former. Specifically, in each simulation scenario, we seek to find the sample size required to achieve good risk model performance 90% of the time.

*Keywords:* Prediction, model development, sample size.

---

[P-217-I-73]

### Sample size calculation for external validation of risk models

*M. Pavlou\**, *C. Qu*, *R. Omar*, and *G. Ambler*, University College London, UK

#### Abstract

Risk prediction models for health outcomes are used in the clinical management of patients and assessment of institutional performance. Given their importance in health care, it is essential that the predictive performance of a risk model is assessed adequately in an external validation study. However, there has been limited research into the sample size requirements for validation studies and hence we aim to provide guidance to researchers through this work. We propose precision methods to calculate the minimum sample size required for an external validation study.

The performance of a risk model is often assessed using the calibration slope and c-statistic, hence it is important to estimate these measures with acceptable precision. We derive analytical formulae for the variance of the C-statistic and the calibration slope that can be used when calculating the sample size for prospective external validation studies. We need to make few assumptions regarding the features of the validation dataset and require information that is easily obtained from previous studies or expert clinical opinion (the outcome prevalence and the anticipated value of the C-statistic). We assess the performance of our estimators using simulation studies in several prevalence and model strength scenarios.

The simulation results show that the proposed expressions provide good approximations to the standard errors. That is, the expressions may be used to estimate the precision of the c-statistic and calibration slope for a given sample size, and hence can be used to determine a suitable sample size for a prospective external validation study.

*Keywords:* risk models, validation, sample size

---

[P-217-I-74]

### Bayesian analysis of accelerating stress-strength models

*Rijji Sen\**, Behala College, Kolkata, India, *S. K. Upadhyay*, Banaras Hindu University, Varanasi, India.

**Abstract**

A body placed under the influence of repeatedly increasing stress, fails, once the cumulative effect of the resulting damage exceeds its intrinsic strength. Statistically speaking, there are two random variables  $X$  and  $Y$ , one representing the stress and the other the strength. The resulting survival Bayesian probabilities are talked in terms of  $P(X < Y)$ . Considering a Gaussian assumption for the strength of the system and a large sample approximation for the stepwise accumulating damage, we consider a set of models emerging from it. An extensive Bayesian analysis of these models have been carried out using some sample based approaches. The plausibility of the models are verified using graphical representation of Rao-Blackwellized predictive density estimates. Finally, the models are compared using predictive loss and a conclusive decision is made regarding which model best explains the considered failure data.

*Keywords:* Accelerating stress-strength models, Metropolis algorithm, Rao-Blackwellized Predictive density estimates, Predictive Loss Criterion

---

[P-222-I-75]

**Inequalities in Farm sector labour force in India**

*Sonali Chakraborty\**, Indian Statistical Institute, India

**Abstract**

The back bone of rural economy in India is the farm sector or agricultural sector. It also contributes an overall economic and social development towards the nation. Farm activities include agriculture (crop production), plantation, animal husbandry (milk, meat, egg etc), forestry & logging and fishing. In India the lion share of labour force are engaged in farm sector. Though the female labour force participation is typically low in our country their share in agriculture is commendable and not decreasing as has been noticed in all other sectors in the recent past. But whether this participation in true sense is in gain full economic activities is to be looked into. Wage discrimination, stereotyping of works, working hours are some inequalities that females are facing at work place. In this analysis an attempt has been made to see the inequality in male female participation in farm sectors. Their wage differentials in rural urban sectors, in different states has been computed following NIC (National Industry Classification) and NCO (National Classification of occupation ) using unit level data of NSS (68<sup>th</sup> round 2011-12 and Periodic labour force survey 2017-18). From preliminary observations it is noted that there is an increase of rural female as market oriented crops producers or animal producers and related workers from 2011-12 to 2017-18. The NIC gives a more vivid picture, where they mainly engaged in mixed farming, plant propagation, post-harvest activities along with perennial and non-perennial crop production. Reduction in rural male in all these industries are noted as an opposite criterion. In both rural and urban sector gender inequalities persists in wages in almost all NIC activities. Some exception like raising of cattle, supporting activities for crop production, fresh water aquaculture, growing of bush fruits and nuts where average earnings of females are little bit higher than male workers working as either casual labour or regular waged salaried person. The female to male wage ratio is minimum in case of MNREGA workers.

*Keywords:* Wage inequality, gender stereotyping, Agriculture labour, activity status.

---

[P-223-I-76]

### Factors affecting food grain production and their forecast in Bangladesh 2050

*Khan A. Matin\**, University of Dhaka, Bangladesh

#### Abstract

The study aims at focusing on the factors of food grain production and forecasting them up to 2050. Time series data on production of food grain (rice and wheat), and various inputs for the period 1982-2016 were obtained from various government publications like the Bangladesh Economic Review, Agricultural Statistics Yearbook and Statistical Year book of Bangladesh. The application of Cobb-Douglas production function revealed that the area under food grain production (ACREAGE), distribution of improved variety of seed (SEED), and area irrigated (IRRIG) all had positive but statistically insignificant output elasticity. The output elasticity of fertilizer consumption (FERTCON) was negative but statistically highly significant ( $P < .01$ ). The output elasticity of disbursement of credit (CREDIT) and cropping intensity (INTENSITY) was positive and statistically highly significant ( $P < .01$ ). The fitted multiple linear regression model under the Cobb-Douglas production function explained 97% of the variation ( $R^2 = 0.97$ ) in total production of food grain (TFGP) and had the desirable statistical properties ( $F_{7,27} = 126.31$ ,  $P = .000$ ,  $D - W = 1.07$ ). The ADF test suggested that all the variables were stationary at 1st difference. We also had increasing forecast value for TFGP, SEED, FERTCON, IRRIG, CREDIT and INTENSITY. The Forecast for total food grain production (TFGP) was found to be 55.6 million M. tons in 2050 under Random Walk model with drift=0.576. The forecast for SEED was found to be 180.9 Million M. tons. The value of FERTCON exhibited an increase to 521.9 Kg per hectare. Irrigation increased to 61.46%. The CREDIT disbursement showed an increase to Tk 3.25 Trillion. Cropping INTENSITY was likely to increase to 236%. But the forecast for ACREAGE and AGLF exhibited no sign of increase during the forecast period under consideration.

*Keywords:* Food grain. Inputs. Cob-Douglas production function, ADF test, ARIMA model, Random Walk model, Forecasting.

---

[P-225-I-77]

### Factors affecting educational performance of academically non at-risk primary students: A multilevel modelling approach

*L. N. Roy*, Dinajpur Laboratory School and College, Bangladesh, *U. K. Majumder\**, Khulna University, Khulna, Bangladesh, and, *D. N. R. Paul*, Uttara University, Dhaka, Bangladesh

**Abstract**

This research study is to identify the significant predictors of primary level educational performance for academically non at-risk primary students and also to test the significance of the school effect on the performance scores applying Multilevel Linear Regression Model. Using two-stage cluster sampling technique, data were collected from 1230 5<sup>th</sup>-grade students of 60 different categories of primary schools under different locations. These schools were randomly selected from 10 Upazilas in Rangpur Division, Bangladesh with probability proportional to size (PPS). For this study, 63% of 5th-grade student's, i.e. 775 students were found to be academically non at-risk considering the risk factors. The results indicated that due to the differences between schools, more than half of the total variation in the Primary Education Completion (PEC) achievement scores was found for academically non at-risk students. The student-level variables can predict the variation both at student-level (89%) and at school-level (99%). The effect of school is found to be significant only due to the student-level factor. Performance score was affected by three student-level factors: i) study factor (prior achievement in class-4, time spent in learning, time needed to complete the homework), ii) family background factor (academic atmosphere at home), and iii) behavioral attitude factor (student's self-educational expectation and student's self-educational confidence); one school-level factor (academic atmosphere at school perceived by the students) and two cross-level interactions (prior achievement in class-4 & academic atmosphere at school perceived by the students; and time spent in learning & academic atmosphere at school perceived by the students). For better academic performance, this study suggests that students should pay more emphasis on their study factor & behavioral attitude factor and both parents & school authority should improve their respective academic atmospheres.

*Keywords:* Educational performance, education, 5<sup>th</sup>-grade, multilevel model and non-at-risk.

---

[P-233-I-78]

**Identification of potential biomarkers from transcriptomics data using integrated bioinformatics approaches**

*Md. Shahjaman\**, Begum Rokeya University, Rangpur, Bangladesh, and *Md. Nurul Haque Mollah*, University of Rajshahi, Rajshahi, Bangladesh

**Abstract**

Transcriptomics data analysis (TDA) is one of the most important OMICS research wings for bioinformatics. Transcriptomics is divided into two fields: one is microarray and other is RNA-Seq. Difference of these two technologies is that microarray produces continuous measurement and RNA-Seq produces discrete measurement of thousands of genes from small number of samples. Transcriptomics datasets is the high-dimensional big datasets ( $p > n$ ). Thus analyzing of these types of datasets has become complicated and challenging for the researchers. The main objective of TDA is to identify the biomarker genes/transcripts whose expressions differ between two or more experimental conditions. So far a number of statistical methods have been developed to identify the biomarker genes from both types of

transcriptomics data. But they produce lower accuracies and higher false discoveries in presence of outlying expressions. Significance analysis of microarrays (SAM) is the popular and widely used methods to identify the biomarker genes for both small-and large-sample cases in the microarray platform. However, it is not robust against outliers. Therefore, in this study, we robustify the SAM approach using the minimum  $\beta$ -divergence method for both unpaired and paired samples. From a breast cancer (unpaired) and head-and-neck cancer (paired) dataset the proposed method identified respectively, three and two genes that were not detected by the other competing methods. Using pathway analysis, disease association study, protein-protein interactions and survival analysis we found that our proposed identified genes might be involved in the critical pathways of cancer. Furthermore, we retrieved 12 significant drugs using these genes that might be therapeutic target in cancer. We then apply the proposed method in RNASeq Datasets to identify the potential biomarker genes. The proposed method identified five and two genes from a real mouse and Nigerian datasets, respectively. Using integrated bioinformatics analysis including gene ontology and KEGG pathway we explored that identified genes are significantly enriched in some important pathways and biological processes.

*Keywords:* Transcriptomics, SAM, Outlier,  $\beta$ -weight function, Robust SAM.

---

[P-234-I-79]

### Application of deep learning in bioinformatics

*Md. Al Mehedi Hasan\**, Rajshahi University of Engineering and Technology, Rajshahi, Bangladesh

#### Abstract

Transformation of biological and biomedical big data into valuable knowledge has become one of the most important challenges in bioinformatics. To extract knowledge from big data in bioinformatics, machine learning has been a widely used and successful methodology. Recently, deep neural network architectures have become increasingly popular as machine learning tools for analyzing dig data. The availability of bigger computational resources, easy to use libraries for implementation and new algorithms for training deep neural architecture are the main drivers for its popularity. Moreover, deep learning has exhibited unprecedented performance in various applications. In this reason, application of deep learning in bioinformatics to gain insight from data has been emphasized in both academia and industry. In this work, we reviewed deep learning in bioinformatics by analyzing various examples from current research. We gave attention both in the bioinformatics domain (i.e., omics, biomedical imaging, biomedical signal processing) and deep learning architecture domain (i.e., deep neural networks, convolutional neural networks, recurrent neural networks, emergent architectures) to provide a useful and comprehensive perspective in each domain. Additionally, we explained some theoretical and practical issues of deep learning in bioinformatics and suggest future research directions. We expect that this study will provide valuable insights and serve as a meaningful starting point for researchers to apply deep learning methods in bioinformatics studies.

*Keywords:*

---

[P-235-I-80]

### **Integration analysis of multi-omics data using hierarchical structural component models**

*Taesung Park\**, Seoul National University, Seoul, Korea

#### **Abstract**

In personalized medicine era, different omics data are commonly generated from each patient. Identification of multi-markers from these multi-omics data is one of the most challenging issues. Although many methods have been developed to identify candidate markers for each type of omics data, few methods can facilitate multi-marker identification from multi-omics data. In this talk, we will introduce a hierarchical structural component model (HisCoM) for integrative analysis of multi-omics data. We apply the proposed HisCoM to integrate two types of omics data. In the first application, we will integrate microRNA and mRNA data and in the second application we will integrate SNP data and metabolomic data. Through this application real multi-omics data, our proposed HisCoM was shown to effectively identify integrated two types of omics data well, providing a much broader biological interpretation.

*Keywords:*

---

[P-237-I-81]

### **Data visualization & machine learning: Impact in drug development**

*Kaustav Nandy\**, Novartis Healthcare Private Limited, Hyderabad, India

#### **Abstract**

With progress in modern computing system, machine learning (ML) is becoming increasingly popular in every industry, pharma is no exception. Once a trail is completed, identifying subgroups for which a certain drug has better/worse performance has tremendous clinical interest. In pharma industry diverse data from wide range of therapeutic area like Oncology, Ophthalmology, Neurology, Endocrinology etc. is explored. Classically, in pharmaceutical industry, such analyses are often driven by intuition and with predefined input from Medical. In this talk, we shall walk through some examples where modern data visualization techniques as well as data driven analysis using machine learning helped to answer these important clinical questions. Additionally, we shall glimpse through some examples where we

used ML tools developed by using R Shiny App to build predictive modeling and analyze the clinical data dynamically.

*Keywords:*

---

[P-238-I-82]

**Kaplan-Meier median survival time based multifactor dimensionality reduction method for identifying gene-gene interactions associated with the survival phenotype**

*Seung Yeoun Lee\**, Mira Park, Jung Wun Lee, and Taesung Park, Sejong University, Seoul, Korea

**Abstract**

In order to identify gene-gene interaction associated with the survival time, we propose a Kaplan-Meier median survival time based multifactor dimensionality reduction method, which referred to KM-MDR. We use a median survival time to classify multi-level genotypes into high and low risk groups and identify the best single nucleotide polymorphism (SNP) pairs associated with the survival time. Through intensive simulation studies, we compare the power of KM-MDR with that of Surv-MDR, Cox-MDR and AFT-MDR. It is found that KM-MDR has the similar power with Surv-MDR with less computing time and has comparable power with Cox-MDR and AFT-MDR even when there is covariate effect. We apply KM-MDR to a real dataset of the ovarian cancer patients from the Cancer Genome Atlas (TCGA).

*Keywords:*

---

[P-239-I-83]

**Gains in life expectancy in the Australian population due to reductions in smoking: Comparisons between targeting the population versus a specific high risk group**

*Haider R Mannan\**, Western Sydney University, Campbelltown, NSW, Australia



**Abstract**

Due to prolonged use of population-based tobacco control strategies smoking prevalence has declined substantially in Australia. However, smoking prevalence is still double in socially disadvantaged groups compared to those who are not disadvantaged despite the narrowing down in this disparity. But not all tobacco control strategies effectively used in the general population is effective in specific high-risk population groups. Therefore, in addition to population-based strategies, another way for effectively reducing smoking in high risk population groups may be to target these groups to more effectively identify and support those smoking to quit. In this backdrop we examined whether the population-based tobacco control intervention is more effective in increasing life expectancy among Australians compared to a high risk tobacco control intervention or a combination of the two when smoking prevalence is reduced to 10% and 0% respectively.

Using the risk percentiles approach, analyses were performed separately for men and women using simulated data from various sources such as the Australian health survey and population census. We examined the key high risk population subgroups such as the socioeconomically disadvantaged, indigenous, homeless, prisoners and those having mental health problems. The results showed that the combined approach of reducing smoking is most effective for improving life expectancy of Australians particularly for the socially disadvantaged and prisoners both of which have a high fraction of smokers in the population. Targeting high-risk population groups having large fraction of smokers in the population for smoking reductions can strongly complement the existing population-based smoking reduction strategies. As both population and high risk approaches for smoking reductions are important, the national prevention policies should make judicious use of both to maximize health gain.

*Keywords:*

---

[P-240-I-84]

**Bivariate Chi-Square Distribution and Some of its Applications**

*Anwar H. Joarder\**, Sultan Qaboos University Muscat, Oman

**Abstract**

The bivariate chi-square distribution has been popular in wireless communication systems. We will discuss product moments and in particular product moments of integer order. Marginal and conditional distributions, conditional moments, coefficient of skewness and kurtosis of conditional distribution will also be discussed. Results match with the independent case when of the variables are uncorrelated. We also show an application in estimating common variance of a bivariate normal population. Some open problems of further investigation will be highlighted.

*Keywords:*

---

[P-241-I-85]

**Response-Based Multiple Imputation Method for Minimizing the Impact of Covariate Detection Limit in Logistic Regression**

*Shahadut Hossain\**, United Arab Emirates University, UAE

**Abstract**

Presence of detection limit (DL) in covariates causes inflated bias and inaccurate mean squared error to the estimators of the regression parameters. This paper suggests a multiple imputation method to correct the deleterious impact introduced by the covariate DL in the estimators of the parameters in logistic regression model. The proposed method exploits the inherent structural relationship between the response and the covariate constrained by the DL in logistic regression. The performance of the method has been thoroughly investigated, and found to outperform the existing competing methods. The proposed method is computationally simple and easily implementable by using three existing R libraries. Moreover, the method is robust to the violation of the distributional assumption for the covariate of interest, which is constrained by the detection limit.

*Keywords:*

---

[P-241-I-85]

**Nexus between Bank-Based Financial Inclusion and Economic Growth in Asia: Do Size of Economy, Governance, Financial Crisis and Regulations Matter?**

*Prashanta Kumar Banerjee\**, Bangladesh Institute of Bank Management (BIBM), Dhaka, Bangladesh, and *Md. Zakir Hossain*, The University of Western Australia, Australia

**Abstract**

This paper investigates whether a bank-based financial inclusion is significant to promote economic growth in Asian economies. To this end, we have incorporated loans, deposits and number of banks as a measure of financial inclusion into our model. We document access to financial services, particularly loans and deposits, is the key driver of economic growth in most Asian economies. Our results reveal when a bank branch is established, it takes about 4 years for this to have a positive impact on GDP. We provide empirical evidence that the effect of financial inclusion is more noticeable in relatively developed economies. However, a higher level of corruption negatively influences the association between financial inclusion and GDP growth in Asia. In effect, the effectiveness of government policy formulation and implementation is weak in most of these countries which drives away the potential benefits of financial inclusion. Furthermore, our analysis suggests a financial crisis does not have a direct bearing on the

underlying structure of Asian economies. Finally, we find Basel regulations do not impede the extent of financial inclusion; rather higher capital ratios positively interact with the extent of financial inclusion to increase GDP.

*Keywords:* Bank, Financial Inclusion, Economic Growth, Asian Economies

---

[P-243-I-87]

### **CPI compilation from big data**

*Nguyen Van Doan\**, *Vu Thi Van Anh*, *Nguyen Van Thuy*, *Nguyen Cong Hoan*, *Nguyen Thi Minh Anh*, *Vu Hai Bang*, and *Pham Thi Hanhand*, Institute of Statistical Science, Vietnam

#### **Abstract**

Currently, in most countries, the resources for state statistical activities are declining while the needs for state statistics are increasing in both quantity and level of detail. Using the traditional method for data collection (data collection by conducting statistical surveys), only a small part of the current needs for statistics is met. This is a huge challenge that requires Vietnam Statistics to research new data sources for replacing traditional ones. The Institute of Statistical Science under the General Statistics Office of Vietnam has researched and exploited big data to produce state statistics. This article aims to share experience of data mining on retail prices of goods and services on websites for compiling CPI in Vietnam.

*Keywords:* E-commerce market, state statistics, traditional data sources, big data, software programs, retail goods lists, CPI compilation.

---

[P-244-I-88]

### **Measure and assessment of state statistical quality in Vietnam Statistics**

*Nguyen Van Doan* and *Dau Thi Quynh Trang\**, , Institute of Statistical Science, Vietnam

#### **Abstract**

Nowadays, statistics play an increasingly important role in policy making, management and administration in all areas of social life. Accordingly, in order to improve the effectiveness of statistics-based decision making, the quality of statistics need be assured. This is one of the major challenges need be solved in a data-centric world. For statistical agencies, to produce high-quality statistics is their first mission. Many statistical agencies in the world determine that the quality of statistics is a vital issue

of their entire statistical system. All the factors such as institution, organizational structure, budget management, legal regulations and other resources of the statistical system are aimed at producing high-quality statistics. Also, Vietnam Statistics is paying more attention to statistical quality management and has made the first efforts in this field. The project to strengthen state management of statistical quality by 2030 approved by the Prime Minister in May 2017 has outlined a long-term quality management roadmap. Accordingly, the set of state statistical quality criteria by 2030 of Vietnam (VSQF) was promulgated by the Prime Minister in January 2019. Currently, Vietnam Statistics is implementing the measure and assessment of state statistical quality with the VSQF, namely developing the methods, processes and tools for measuring and assessing statistical quality of statistical agencies in the state statistical system of Vietnam, which is described in detail in this paper.

*Keywords:* official statistics, statistical quality management, set of state statistical quality criteria

---

[P-244-I-88]

### **Current status of sustainable development goals: 2016-2030 A “one world” approach to the global development agenda**

*M. Mosleh-Uddin\**, Ivytech Community College, USA

#### **Abstract**

In September 2015, 193 Global leaders agreed on the 17 Sustainable Development Goals (SDGs), 169 Targets and 231 Indicators. The past three decades have seen unprecedented progress when it comes to reducing extreme poverty around the world – but there’s still an awful lot more to be done. More than 1.3 billion people live in extreme poverty ( less than \$1.25/day). The reduction in extreme poverty was mainly driven by East Asia and the Pacific and South Asia ( China, Indonesia and India).Bangladesh has been widely acclaimed as one of the forerunners of Millennium Development Goals (MDGs) implementation. It made outstanding progress in the areas of poverty alleviation, ensuring food security, primary school enrolment, and gender parity in primary and secondary level education, lowering infant and under-five mortality rate and maternal mortality ratio, improving immunization coverage and reducing the incidence of communicable diseases. In the 2015+ period , we must move away from poverty reduction to eliminating extreme poverty of all forms. We must ensure that none should be left out because of race, religion, caste, geographical locations, disability, gender or any other reason from their fundamental rights and minimum basic needs. Access to good schools, health care, electricity, safe water and other critical services remains elusive for many people , often determined by socio-economic status, gender, ethnicity and geography. Moreover, for those who have been able to move out of poverty, progress is often temporary: economic shocks, food insecurity and climate change threaten to rob them of their hard earned gains and force them back into poverty. It will be critical to find ways to tackle these issues as we make progress towards 2030. The most challenge will be on data to measure the achievement of the goals.

*Keywords:*

[P-247-I-90]

**Joint model for Longitudinal data subject to nonlinear mixed-effects model and time-to-event data: A two stage approach***Arindom Chakraborty\** and *Srimanti Dutta*, Visva-Bharati University, India**Abstract**

Medical and epidemiological studies often involve instances where we have to analyze repeated evaluations of outcomes of a particular characteristic of a patient along with event history data which may involve death, dropout or progress of a disease. These data needs to be analyzed using a joint modelling approach where we take into account simultaneously two processes i.e., the longitudinal and the time-to-event. However, classical approach suffers from some serious drawbacks as current maximum likelihood methods for joint modeling are only advantageous when the dimension of random-effects is not high. Further, for multivariate linear mixed-effects models, the model fitting and analysis becomes intractable in cases where dimension of variables are greater than four due to the increase of parameters in the covariance matrix for random-effects. Keeping in mind these difficulties in estimation, we have adopted a two-stage approach to our proposed joint model with  $n + 1$  components. Here, the first  $n$  components describe the longitudinal process which is viewed in nonlinear framework. The  $(n + 1)th$  component describes the time-to-event and our main objective is to examine how the time-to-event depends on the longitudinal process. The association between the two processes has been defined through Bartlett decomposition of the covariance matrix. We had considered an accelerated failure time model under selection model framework. We consider a linear mixed-effects model in the first stage for the longitudinal process and use the information obtained from that stage to estimate the parameters in the survival process. This has been achieved by adopting gradient descent algorithm with adaptive learning rate. A simulation study has been done to study the performance of the proposed methodology. Finally a data set on AIDS was analysed using this two-step method.

*Keywords:*

[P-248-I-91]

**Recent trends in fertility behavior in Bangladesh***Mizanur Rahman\**, University of North Carolina, Chapel Hill, USA**Abstract**

Bangladesh made remarkable improvements in health, education, and economic growth in recent decades. For example, total fertility rate (TFR) declined from over 6 births per woman in the early 1980s to 2.3 in the 2010s. Since then TFR seemed to have stalled, contraceptive prevalence rate (CPR) stalled at 62%, even slightly declined between 2014 and 2017-18. We ask several questions in this paper: Has fertility stalled? Is there any possibility of even reversal of fertility? How some regions achieved replacement-level fertility with lower CPR than the national targeted CPR of 75%? Is the stall or potential reversal of fertility associated with economic improvement? Is the family planning job done? Why the poor have higher TFR than the rich despite former's higher and more effective contraceptive method use? We attempt to address these questions by using BDHS data.

There is an indication of stalling of fertility as age specific fertility rates in prime ages 20-34 remained constant or even slight increase over recent years. Desire for having no more children increased steadily since the early years of fertility transition but it began to decline since 2011 which apparently coincides with economic improvement. Fertility is likely to decline after some years associated with increased women's education and employment and parents' aspiration of quality children in terms of better education and health. The western region of Bangladesh has achieved a replacement-level fertility (or below) with a CPR of 70% or below probably because of their high practice of menstrual regulation (MR) or medical MR. The family planning job is not yet finished at all; there is still an unmet need for family planning of 12%. The program needs reinforced and better organized services in the areas of behavior change communication, quality of care, and range of services. The higher fertility of the poor despite their higher contraceptive use than the rich is associated with the former's lack of information and resources for MR services.

*Keywords:* Total fertility rate, desire for children, menstrual regulation, family planning

## Contributed Oral Papers

---

[P-005-CO-01]

### Monitoring small variation of a non-normal process

*Md. Mizanur Rahman Mukta\**, Bangabandhu Sheikh Mujibur Rahman Science Technology University, Bangladesh and *Md. Pear Hossain*, City University of Hong Kong, Hong Kong

#### Abstract

Control charts are the most attractive tool among the magnificent seven tools of SPC (Statistical Process Control) and they have been readily used as a sophisticated device to monitor a statistical process. Most of the control charts are basically designed under normality assumptions. But in countless pragmatic situations in real life this normality assumptions may be violated. One such non-normal situation is to monitor the process variability from a non-normal (skewed) parent distribution where we propose the use of a VREWMA control chart which is developed under Rayleigh distribution. We modified the exponentially weighted moving average (EWMA) control chart under Rayleigh distribution for detecting small shifts for process monitoring. We introduces pivotal quantity for the scale parameter of the Rayleigh distribution which follows a gamma distribution. Probability limits and L-sigma limits are contemplated alongside with the performance measure based on different RL (run length) attributes like, ARL (average run length),SDRL (Standard deviation of run length), MDRL (median of run length). Power curve, RL curve and percentiles are also used to measure the performance. We also compared our proportionate control chart with classical EWMA to illustrate the superiority our proposed chart. To avoid the complexity of future calculations for practitioners, factors for constructing control chart for monitoring the Rayleigh parameter are given for different sample sizes and different false alarm rate. We also provide simulated data to illustrate and judge the behavior of VREWMA control chart for monitoring small variations of non-normal process. We show that the proposed VREWMA control chart can detect the small variations for the non-normal process for different smoothing constants. Finally, couple of real life examples (including latest wind speed data of Bangladesh) has been provided to show the significance of such a control chart.

*Keywords:* Rayleigh distribution, gamma distribution, process monitoring, EWMA control chart, VREWMA, average run length, smoothing constant, wind speed energy.

---

[P-010-CO-02]

### Multiple regression model to examine the incidence of government expenditure on neonatal mortality in Nigeria

*Abubakar Usman\**, Federal University of Technology, Minna, Nigeria and *Sulaiman Muhamad*, Federal University of Technology, Minna, Nigeria

**Abstract**

The study examined the incidence and accessed the impact of governments' expenditure on the incidence of Neonatal Mortality in Nigeria. This was tested with use of multiple regression analysis of Neonatal Mortality against the total recurrent expenditure ( $X_1$ ) and total capital expenditure ( $X_2$ ) of the federal government of Nigeria allocated to the Health sector taking into consideration the available data set from 1990 to 2017. The Multiple regression model derived from the data set is  $Y = 50.9000 - 0.06023X_1 - 0.0535X_2$ . The null hypothesis test was rejected, which was an indication that the study has statistically significantly identified that both the Recurrent Expenditure and Capital Expenditure contributes greatly to the decline in Neonatal Mortality in Nigeria

*Keywords:* Regression, under-five, recurrent, endogenous, antenatal, and neonatal.

---

[P-019-CO-03]

**Bias-reduced and separation-proof poisson regression model with small or sparse count data**

*Momenul Haque Mondol\**, University of Barishal, Bangladesh, *Mohammad Shafiqur Rahman*, University of Dhaka, Bangladesh, and *Wasimul Bari*, University of Dhaka, Bangladesh

**Abstract**

Separation or monotone likelihood can be observed in the fitting process of Poisson regression using maximum likelihood estimation (MLE) technique when one or more parameters diverge to infinity. The separation is very common in count data when sample size is small or there is huge number of zero count or there is sufficient number of strong predictors or mixture of two or more such conditions. The study investigates the consequence of separation in the standard Poisson models and provides a solution by incorporating Firth's (1993) type penalty term, which was originally proposed for bias reduction in MLE, in likelihood score equation that we may call penalized Poisson. The modified score equation guaranteed an achievement of convergence and finite estimate of the regression coefficient. An extensive simulation study was conducted to assess the performance of penalized Poisson model over standard Poisson and Zero-inflated Poisson in the presence of separation. Several simulation scenarios were considered for creating complete or quasi-complete or near-to-quasi-complete separation by varying the sample size, proportion of event in the binary predictor which make separation, and the magnitude of regression coefficient, log odds ratio  $\beta$ , relating the binary predictor to the response. The results revealed that the penalized Poisson having profile penalized likelihood based confidence interval, performed better than the standard maximum likelihood based model, standard Poisson and Zero-inflated Poisson, in terms of bias, MSE and length of 95% confidence interval (precision) in all simulation scenarios. An illustration using real data also supported the simulation findings.

*Keywords:* Separation problem, bias reduction, sparse count data, Poisson model, zero-inflated Poisson.



---

[P-025-CO-04]

### Statistical prediction of chemical components in groundwater of Bangladesh

*Rifat-E-Jahan Siddiquee\**, a2i ICT Division, Bangladesh, and *Paritosh K Roy*, University of Dhaka, Bangladesh

#### Abstract

Imbalanced presence of various chemical components like arsenic, iron, calcium etc. in the groundwater of an area is a cause of serious concern in the field of public health as it may significantly affect the disease prevalence in that area. Therefore, the prediction of these chemical components and mapping is of paramount importance to the policy planner for designing and implementing interventions to improve the human health and well-being. In this paper, two distinct procedures for the construction of predictive maps with regards to the spatial prediction of chemicals are proposed; machine learning procedures and model based approaches. For machine learning methods, we are considering regression tree, random forest and stochastic gradient boosting. In the model based approaches, geostatistical linear regression model, simple kriging are contemplated. This study also executes a comparative analysis of both methods in the hopes of finding out which method performs better. The study was conducted throughout the 64 districts of Bangladesh consisting of 3534 tube wells. Observations on a total of 23 variables were obtained from these locations. In comparison with model based approaches, it's seen that machine learning methods perform quite better both in terms of predictive accuracy and precision. In conclusion, machine learning approaches like random forest and stochastic gradient boosting are very effective methods for the purpose of explaining the spatial variability among the chemical components in various locations and also for the construction of spatial predictive maps.

*Keywords:* Machine learning, spatial prediction, mapping.

---

[P-026-CO-05]

### Association of biomass fuel smoke with respiratory symptoms among children under five in urban areas: Results from Bangladesh urban health survey, 2013

*Md. Hasan\**, Bangabandhu Sheikh Mujib Medical University (BSMMU), Bangladesh

#### Abstract

A little is known regarding the effect of exposure to biomass fuel smoke inhalation on respiratory symptoms in Bangladeshi population which is a major health hazard in most of the developing countries. This study aims to explore the association between respiratory symptoms and biomass fuel smoke exposure among children under five years of age. Data were extracted from the Bangladesh Urban

Health Survey conducted in 2013. A total of 10575 mothers with least one surviving children were selected. Respiratory symptoms among children under five were considered as the primary outcome. Sequential multiple logistic regression models were used to observe the association between respiratory symptoms and biomass fuel smoke exposure adjusting the effect of residential factors, mother and child characteristics. Around 40% of the mothers exclusively used biomass fuel irrespective of the kitchen location and 54% of them were habituated in indoor cooking. Prevalence of respiratory symptoms of under-five children among in-house and outdoor biomass fuel users were 23.0% and 21.9%, respectively. Results of fitted multiple logistic regression models showed that the odds of having respiratory symptoms among children under five years of age were increased due to in-house biomass fuel use [OR=1.18; 95% CI, 1.04 - 1.36]. An increased risk of respiratory symptoms was also significantly associated with mother's birth complication [OR=1.51; 95% CI, 1.36 - 1.67]; NGO membership of mothers [OR=1.32; 95% CI, 1.16 - 1.51]; age of the child (6-23m) [OR=1.29; 95% CI, 1.10 - 1.52], and nutritional status (stunting) [OR=1.18; 95% CI, 1.06 - 1.31]. This study found the use of in-house biomass fuel as a significant risk factor associated with respiratory symptoms of children under five. More longitudinal studies should be designed to establish a causal relationship between HAP (household air pollution) and respiratory symptoms among children with more direct measures of HAP and clinical procedure.

*Keywords:* Biomass fuel smokes, respiratory symptoms, children under five, urban health, Bangladesh

---

[P-030-CO-06]

### **On the improvement of likelihood estimation for the shape parameter of generalized extreme value lifetime distribution**

*Md. Mazharul Islam\**, Bangladesh Institute of Governance and Management, Bangladesh,  
and *Md Hasinur Rahaman Khan*, University of Dhaka, Bangladesh

#### **Abstract**

In presence of nuisance parameters, profile likelihood inference is often unreliable and biased, particularly in small sample scenario. Over past decades several adjustments have been proposed to modify profile likelihood function in literature including a modified profile likelihood estimation technique introduced in Barndorff-Nielsen. In this study, adjustment of profile likelihood function of parameter of interest in the presence of nuisance parameter is investigated deeply. We mainly focus to implement the Barndorff-Nielsen's technique on generalized extreme value (GEV) distribution for estimating its shape parameter. The accelerated failure time models are used for lifetimes having GEV distribution. Monte-Carlo simulation study is used to demonstrate the performance of the profile and modified profile approaches. Simulation results confirm the expected superiority of the modified profile likelihood estimates over the profile likelihood estimates for the parameters of interest. It is found that the modifications can improve considerably the overall performance of the estimators through reducing their biases and standard errors. Real-data example is provided to demonstrate the performance of the methods. The real data analysis also suggests similar conclusions as found in the simulation study.

*Keywords:*

---

[P-037-CO-07]

**Knowledge and practices on exclusive breastfeeding among mothers in rural area of Rajshahi district in Bangladesh**

*Md. Masud Rana\**, *Md. Rafiqul Islam*, and *Md. Golam Hossain*, University of Rajshahi, Rajshahi, Bangladesh

**Abstract**

World Health Organization (WHO) suggests that exclusive breastfeeding (EBF) is the best nutrition for the neonate. Still, it remains a big challenge to establish EBF in Bangladesh. The aim at the study was to determine the level of knowledge and practices on EBF in Rajshahi District, Bangladesh. A total 513 number of mothers were selected who having at least one child age (6-12) months from 32 different community clinic in the rural area of Rajshahi District, Bangladesh during September to December 2015. A composite index, chi-square test, and binary logistic regression model were utilized in this study. The prevalence of good EBF knowledge and practices was 32.0% and 27.9% among mothers having at least one child age (6-12) months. From analyses, mothers age from =31 age in years were (AOR=0.040, 95% CI:0.021-0.079;  $p<0.05$ ) and (AOR=0.084, 95%CI:0.050-0.143;  $p<0.05$ ) times less likely to had a good EBF knowledge and practices compared to mother's age =30 year's. Housewife mothers was (AOR=21.352, 95% CI: 5.170-88.174,  $p<0.05$ ) and (AOR=9.992, 95% CI: 4.485-22.260,  $p<0.05$ ) times more likely to have a good knowledge and practices than their service holder mothers. Those mothers monthly family income = 10,000 BDT were (AOR=0.197, 95% CI; 0.088-0.442;  $p<0.05$ ) and (AOR=0.092, 95% CI: 0.050-0.168,  $p<0.05$ ) times less likely to had a good knowledge and practices respectively compared with their counterpart. This study found huge gap about knowledge and practices on EBF among mothers having at least one child age (6-12) months. This study suggested that education and EBF related intervention could play an important role to increase good knowledge and practices on EBF among mothers having at least one child age (6-12) months. Malnutrition will be decreased if EBF was widely established in Bangladesh.

*Keywords:* Exclusive breastfeeding, knowledge and practice, composite index, chi-square test, binary logistic regression.

---

[P-049-CO-08]

**The penalized likelihood for AFT models with small or rare event survival data**

*Tasneem Fatima Alam\** and *Mohammad Shafiqur Rahman*, University of Dhaka, Bangladesh

**Abstract**

The Accelerated failure time (AFT) model is widely used in medical science and reliability engineering for its intuitive interpretation. The model parameters are generally estimated by maximum likelihood estimation (MLE) which reports unbiased and consistent estimates when sample size is large and/or rate of censoring is low; however, its small sample performance is unknown. This paper investigated the properties of MLEs of the regression parameters of the AFT models for small sample or rare event (high rate of censoring) situation and introduced a penalized likelihood approach to address the problems. The penalized likelihood function and the corresponding score equation were derived by adding a penalty term, equivalent to the Jeffreys invariant prior, to the existing likelihood function, which was originally proposed by Firth (Biometrika, 1993) for reducing the first order bias in MLEs of the regression parameters of the exponential family models. The penalized method was illustrated for the most commonly applied log-location-scale family models such as Weibull, Lognormal and Loglogistic distributions. The illustration showed that the Jeffreys-prior based penalized likelihood succeeds to achieve convergence, providing finite estimates of the regression coefficients and solves the problem of separation or monotone likelihood created by a covariate, which are not often possible by the MLE. Extensive simulation studies conducted separately for each of the log-location-scale models demonstrated the penalized approach to have a substantial improvement over MLE by providing smaller amount of bias, mean squared error (MSE) and narrower confidence interval. An application of the methods using data with small sample and rare event supports the findings from the simulation. The penalized likelihood approach is therefore recommended to use in both cases of small ( $N < 50$ ) and large sample with high censoring rate in practice.

*Keywords:* Bias reduction, monotone likelihood, Jeffreys prior, log-location-scale family.

---

[P-052-CO-09]

**Dealing with separation or near-to-separation in the multinomial logit models with application to childhood health seeking behavior data**

*Nowrin Nusrat\** and *Mohammad Shafiqur Rahman*, University of Dhaka, Bangladesh

**Abstract**

Logistic regression models are commonly applied to analyze multinomial data frequently arising in many areas of research including medicine and social sciences. However, separation leading to monotone-likelihood created by a covariate often exists in multinomial logistic model when the sample size is relatively small, the certain response categories is rare, and if there is sufficient number of influential covariates. In the presence of separation, the maximum likelihood (MLE) may fail to achieve convergence or provide biased or infinite estimate of at least one regression coefficients of the model, particularly for the coefficient associated with the covariate responsible for creating separation. This study addressed the problems of separation by applying a penalized likelihood (PMLE) approach, which was originally proposed by Kosmidis and Firth (2011) to remove the first-order bias in the MLEs of the multinomial

logit model via Poisson log-linear model. The penalized likelihood is shown to achieve convergence and provide finite estimate of the regression coefficient in the presence of separation. We investigated the performance of both MLE and PMLE using an extensive simulation study against scenarios with complete (e.g., more than one empty cell) or quasi-complete-separation (one empty cell) and a new form termed as “near-to-separation” (non-zero cell but with few observations-less than 15% of total sample), which is more common in practice than the other two forms. The simulation study showed that the MLE failed to achieve convergence and/or provided infinitely large estimate of the regression-coefficient in the presence of complete or quasi-complete-separation, whereas the PMLE showed some improvements by achieving convergence and providing finite estimate. In the presence of near-to-separation, the PMLE also outperform the MLE by providing smaller amount of bias and MSE and better coverage. An application of the method is provided to analyze childhood health seeking behavior data consisting of different forms of separation.

*Keywords:* Monotone likelihood, Poisson log-linear model, penalty function.

---

[P-054-CO-10]

### **Influencing factors of maternal health service utilization in Bangladesh: A statistical analysis of the BDHS data 2014**

*Md. Akhtarul Islam, Shadman Shakib Ahmed, and Henry Ratul Halder\**, Khulna University, Bangladesh

#### **Abstract**

Antenatal care (ANC), use of skilled birth attendants (SBAs) and postnatal care(PNC) services are crucial for maternal health; these can remarkably minimize maternal mortality. The factors that affect these service utilization can help to intrigue a suitable approach and policies for the achievement of the targeted maternal mortality ratio (MMR) of the Sustainable Development Goals (SDGs). In our study, we have analyzed BDHS 2014 data using the Binary logistic regression model. Education of women, place of residence, wealth index, and husband’s education has a significant association with the use of maternal health services. Women who completed higher education are more likely to use ANC(AOR=3.76, CI=2.07-6.52), SBAs (AOR=2.82, CI=1.96-1.4.08) and PNC (AOR=1.87, CI=1.31-2.68). Women from urban areas have a higher chance of using ANC (AOR = 1.30, CI = 1.05-1.60), SBAs (AOR=1.64, 9 CI=1.40-1.92)and PNC (AOR=1.45, CI =1.23-1.70) more than the women from rural areas.Women from the richest households tend to use more ANC (AOR=2.30, CI=1.80-2.95), SBAs (AOR = 2.23, CI = 1.83-2.71) and PNC (AOR = 1.78, CI = 1.47-2.16) than those who are from the poorest households. There is a high inconsistency in these service utilization among women with dissimilarities in education, household wealth and residence. ANC is an essential entry point for ensuring the application of institutional delivery and PNC services. Schemes that seek to enhance maternal health services utilization should target the improvement of education, economic status, and empowerment of women.

*Keywords:* Maternal health service, antenatal care, postnatal care, Bangladesh.

---

[P-055-CO-11]

**Using fractional polynomials for examining the association of age with anemia and malnutrition of children under five years**

*Mahin Afrose\** and *M. Shafiqur Rahman*, University of Dhaka, Bangladesh

**Abstract**

Assessing association of continuous covariates with the binary outcome in a regression model is challenging, because the functional form for the effect of a covariate is often unknown. Incorrect specification of the continuous covariate in modeled may lead to misleading inference. When the true association is non-linear, using categorization or imposing a linear association lead to large reductions in power. Fractional polynomial correctly specifies the nonlinear association of continuous covariate with the outcome and hence it is widely considered as an alternative to the traditional approaches for analyzing continuous covariates. However, its use in public health research is very limited. For example, several studies have been conducted to assess the association of age with anemia and nutrition, none of them have reported the correct functional form of the association, and most of them provided solution by categorizing age, which leads to misleading inference. However, identifying the correct association of age on anemia or malnutrition might be useful for the policy makers to identify the target group for designing appropriate policies. This study aims at examining the association of age with anemia and malnutrition among children of age under five years using fractional polynomial. Data from the 2011 Bangladesh demographic and health survey were used in this study. The results revealed that a non-linear association of age with both anemia and malnutrition was found. Fractional polynomial regression model identified that the second order polynomial of power (-2, -1) for anemia, power (0.5, 1) for stunting and (1,1)for underweight is significant with lowest deviance. The association was further investigated after controlling for the effect of some known confounders such as mother's education, socio-economic status etc. Based on the findings, some recommendations are discussed.

*Keywords:* Malnutrition, anemia, non-linear relationship, fractional polynomial, DHS data.

---

[P-057-CO-12]

**Determinants of unmet need for modern contraceptive in Bangladesh: A mixed model analysis**

*Md Kawsarul Islam\**, Bangladesh University of Textiles, Bangladesh, *M. Ershadul Haque*, University of Dhaka, Bangladesh, and *Zillur R. Shabuz*, University of Dhaka, Bangladesh

**Abstract**

Unmet need for contraceptive is the key demographic indicator for any country. This study examines the determinants of unmet need for modern contraceptive among ever-married women of reproductive age in Bangladesh. The data were extracted from the Bangladesh Demographic and Health Survey (BDHS) 2014, which is a nationally representative survey implemented using a stratified two-stage cluster sample design. We considered 11932 ever-married women of age 15-49 for analysis in this study. A two-level random intercept binary logistic regression model as well as single-level binary logistic regression model have been used to identify the determinants of unmet need. The findings suggest that place of residence, division, age of the women, working status, number of living children, visited by FP workers and decides on healthcare have significant effects on unmet need. The study reveals that the odds ratio from the single-level logistic regression underestimates estimated parameters.

*Keywords:* Unmet need, modern contraceptive, family planning.

---

[P-060-CO-13]

**Determinants of coital frequency among currently married women in Bangladesh**

*Mamun Ibn Bashar\** and *Md. Moinuddin Haider*, icddr, Bangladesh

**Abstract**

Coital frequency is an important determinant of fertility. The study explores the socio-demographic differentials of coital frequency among currently married Bangladeshi women. The study used a nationally representative survey, Bangladesh Demographic and Health Survey 2011, to study the association of coital frequency in the last four weeks with variables like women age, education, religion, contraceptive use, years of cohabitation, asset index, urban-rural residence, husband's occupation, etc. Multiple logistic regressions were used to measure association. Average number of sexual intercourse among women aged less than 20 years is 7.7 and 3.9 among women aged above 35 years. Women with below secondary level of education are 13 percent more likely to have intercourse in a week than who never attended school and secondary completed women are 23 percent more likely to have it. Couples who are living together for less than three years are 20 percent less likely to have it in a week and 21 percent higher to have it among who are cohabiting for 10-19 years than who are cohabiting for 3-9 years. Any contraception users have it more frequently than nonusers. Religion is also an important factor that affects intercourse: non-Muslim couples are 18 percent less likely to have it in a week compared to Muslim couples. No significant association was found between coital frequency and the independent variables TV watching, body mass index or asset quintiles. The possible explanations for the differentials in coital frequency in Bangladesh context are explored in this paper.

*Keywords:* Coitous, socio-demographic determinants, logistic regression.

---

[P-062-CO-14]

**Variable selection for censored data using modified correlation adjusted correlation (MCAR) scores**

*Mst. Afsana Mimi\** and *Md Hasinur Rahaman Khan*, University of Dhaka, Bangladesh

**Abstract**

Dealing with modeling for high-dimensional censored data is very challenging because of the complexities in data structure. This paper focuses on developing a variable selection procedure for censored high-dimensional data with the AFT models using modified Correlation Adjusted coRrelation (MCAR) scores method which is developed based on CAR scores method (Zuber, Strimmer, et al., 2011) that provides a canonical ordering that encourages grouping of correlated predictors and down-weights antagonistic variables. The proposed MCAR scores method is developed as an extension of the CAR scores method through using NOVEL integration of the sample and threshold estimator of the correlation matrix as suggested by Huang and Frylewicz (N. Huang & Fryzlewicz, 2015). The proposed MCAR exhibits computationally more efficient estimates under model sparsity and can provide a canonical ordering among the predictors. The MCAR method is a greedy method that is also easy to understand and can perform estimation and variable selection simultaneously. Performances of variable selection of the MCAR method have been compared with other existing regularized techniques in literature- the lasso, elastic net and with an machine learning technique- the boosting and with the the censored CAR by a number of simulation studies and a real microarray data set called Diffuse large-B-cell lymphoma. It reveals that when correlation exists among the covariates, the MCAR method outperforms all five techniques while for uncorrelated data, the MCAR performs quite similar to the CAR method but clearly outperforms other three methods. The empirical study further reveals that the MCAR method exhibits the best predictive performance among the methods.

*Keywords:* Censoring, AFT model, high-dimensional data, MCAR score.

---

[P-065-CO-15]

**Analysis of rainfall occurrence in consecutive days using Markov models with covariate dependence in selected regions of Bangladesh**

*Sultan Mahmud\**, Innovation for Poverty Action, Bangladesh, *M. Ataharul Islam*, University of Dhaka, Bangladesh, and *Syed Shahadat Hossain*, University of Dhaka, Bangladesh



**Abstract**

The objective of this study is: to reveal the regional variation and identify the responsible factors for making the transition of rainfall status in two consecutive days in Bangladesh at three selected regions. The covariate dependent first-order Markov models have been used to find out how the change of rainfall status (rain or no rain) is affecting by the selected factors. Among the considered factors, we found the relative humidity, sunshine hour, maximum temperature and cloud cover show positive association with the transition type no rain to rain for all three stations however the minimum temperature, maximum temperature, and sea level pressure have negative effect on this transition type in all selected stations with exception that the maximum temperature in Rajshahi and minimum temperature in Rajshahi and Sylhet do not have any significant effect. On the other hand, with the exception, the maximum temperature in Rajshahi, and sunshine hour in Dhaka and Sylhet the relative humidity, maximum temperature sunshine hour and cloud cover show positive association with the transition type rain to rain and the minimum temperature and sea level pressure show negative association with this transition type. The regional variations and responsible factors for changing the rainfall status from disclosed as well among three selected regions over four seasons such as pre-monsoon, monsoon, post-monsoon as well as winter. In all four seasons, the transition of the rainfall status from experienced statistically significant positive impact for the exploratory variables relative humidity, sunshine hour and cloud cover and negative impact for maximum temperature, minimum temperature as well as sea level pressure with some exceptions. However, the humidity and cloud cover play the most important role in making the transition the rainfall status between two consecutive days.

*Keywords:* Rainfall analysis, Markov model, statistical method, longitudinal data analysis.

---

[P-066-CO-16]

**Spatial patterns of child growth indicators in Bangladesh: A hierarchical bayesian spatial model**

*Nasrin Lipi\** and *Syed Shahadat Hossain*, University of Dhaka, Bangladesh

**Abstract**

Anthropometric indicators such as height adjusted for age (HAZ), weight-adjusted for height (WHZ) and weight adjusted for age (WAZ) are usually used to monitor child growth. Studying geographical patterns of child growth and their determining factors should improve the understanding of the nutritional situation in any country and thus help in making health policies and designing interventions at sub national level. We used a Hierarchical Bayesian spatial models based on conditional autoregressive (CAR) priors to model the spatial patterns in HAZ, WHZ and WAZ scores and their relation to different socio-demographic characteristics such as child age, mother's education, wealth index, place of residence, etc. The data for our illustration comes from the Multiple Indicator Cluster Survey in 2012-2013, a nationwide sample survey conducted by Bangladesh Bureau of Statistics (BBS). Posterior medians of the spatial effects are shown in choropleth map. Spatial analyses for each of the anthropometric growth

indicators showed geographical patterns of child growth and areas with high or low level of the indicators were identified.

*Keywords:*

---

[P-068-CO-17]

**Quality of corporate environmental performance and financial performance: Evidence from select Asian countries**

*Santi Gopal Maji\** , and *Utpal Kumar De*, North Eastern Hill University, India

**Abstract**

After more than three decades of theoretical and empirical research on the association between environmental performance and financial performance of corporate firms, the results still remained inconclusive. The purpose of this study is to investigate the association between quality of environmental performance and financial performance of major firms belong to three Asian countries: Japan, South Korea and India. Among these three selected nations, Japan and South Korea are industrially advanced and also in the publication of the sustainability reports, while India is in emerging stage on both counts. The sample for the study consists of all listed non-financial firms who have been publishing sustainability reports in their respective website in English language continuously over the years since 2009. Based on the above criteria, 93 companies have been included in the sample of which 36 from Japan, 29 from South Korea and 28 from India respectively. The study period is 10 years, from 2008-09 to 2017-18, which is considered on the basis of the beginning of sustainability reporting in India (2009 onward) and the latest period of data availability.

Since the extant literature denotes that environmental performance is more associated with the market value of firms, we have used market to book ratio (MBR) as the proxy for corporate financial performance. Also, content analysis technique has been used to compute the disclosure score of environmental performance of the firms. Using GMM model of regression it is found that environmental disclosure and financial performance are strongly and positively associated in Japan and South Korea, while the relation has been found insignificant in case of India. But profitability is found to encourage the firms to adopt better environmental management as reflected through their disclosure and sustainability measure, which is very strong in case of industrially advanced Japan and Korea and India has to go a long way in this direction.

*Keywords:* Environmental performance, financial performance, profitability, content analysis

---

[P-069-CO-18]

### **Piecewise linear multiple change point model: An application to survival analysis**

*Sukanta Chakraborty\**, Bangladesh Bank, Chittagong, Bangladesh

#### **Abstract**

The hazard function is a frequently used function for modeling and evaluating the time related events. In the Cox-proportional hazard model this hazard function assumes to be constant. Earlier many statisticians investigated this assumption and found that this function is not constant over time and it has some change point in survival time of patient. Many authors Matthews and Farewell (1982), Nguyen et al. (1984), Matthews et al. (1985), Yao (1986), Worsley (1988), Henderson (1990), Loader (1991), Pons (2002), Gijbels and Gurler (2003), and Zhao et al. (2009) investigated about one change point in hazard function through piecewise constant hazard model. Moreover, Goodman et al. (2006, 2007), Qian and Zhang (2014) investigated about multiple change point in hazard function through piecewise linear multiple change point model by Nelder mead simplex method. Rois et al. (2017) investigated about multiple change point in hazard function through piecewise linear multiple change point model by Cross entropy algorithm. In our paper we calculate the likelihood function for piecewise linear multiple change point model considering three change point. We found the likelihood function non-linear. Since this likelihood function is non-linear so we use non-linear minimization method to estimate the parameters of the model. Later in this paper we make comparison among the parameters estimated by non-linear method with those estimated by Nelder-mead simplex method & Cross entropy algorithm in a simulation study. Simulation study shows that the standard error of the estimates estimated by Non-linear minimization method is lowest among the three method. A Wald type test statistics shows that the estimated parameters are significant. Power analysis is also conducted to test the adequacy of the sample size. So, change point estimation in hazard function by non-linear minimization method can be considered as an improvement in change point analysis.

*Keywords:* Hazard function, change point.

---

[P-071-CO-19]

### **Construction of windows for pharmacokinetic sampling in phase I clinical trial**

*Nigar Sultana\**, University of Dhaka, Bangladesh

**Abstract**

This paper describes a method for the construction of pharmacokinetic sampling windows in a way so that they are around the D-optimum time points. Here we consider the situation where a pharmacokinetic (PK) study is accompanied by a dose-finding study in phase I. The D-optimal criterion is often used to determine the optimal time for collecting blood samples so that they provide maximum information regarding the population PK parameters. However, blood samples collection at the D-optimal time points are often difficult. Instead, sampling time point chosen from a suitable time interval can ease the process. The proposed method is conceptually simple and considers the average value and standard deviation of D-optimal time points up to create sampling windows. Random time points can be chosen then from these windows to collect blood samples from the next cohort. The nonlinear random effects model has been used to model the PK data. Also, we employ the continual reassessment method for dose allocation to the patients. Comparisons of the accuracy and precision for the PK parameter estimates obtained at the D-optimal and random time points are also presented.

*Keywords:*

---

[P-075-CO-20]

**Predictive performance of the logistic regression model under separation problem and minimal EPV**

*Kazy Farhat Tabassum\** and *Md Hasinur Rahaman Khan*, University of Dhaka, Bangladesh

**Abstract**

The predictive performance of the logistic regression model is generally influenced by the problems of low events per variable (EPV) and separation when the number of variables is considerably high. Firth's correction and existing approaches for simultaneous identifying and handling these two problems under models with many variables may lead to substantially different predictive performances. This study aims to examine the reasons of different predictive performances obtained under various settings of EPV and separation problems and to propose several alternative techniques that will improve the model prediction accuracy. Several simulation studies have been performed to examine the basic rule of 10 EPV along with the effects of separation. Results reveal that the traditional Firth's correction method and validation methods - apparent performance, repeated split sample, bootstrap optimism as proposed in Austin and Steyerberg (2017) suffer from criticisms while predictive performance is evaluated using c-statistic, AUC, brier score and error rates. Simulation results also demonstrate that among three regularized machine learning techniques (support vector machine, random forest, neural network) penalized SVM to handle separation problem and a novel repeated splitting technique to handle low EPV problems outperform the traditional techniques. We also notice that the 10 EPV rule for binary logistic regression is weak under the traditional methods than these proposed methods. The predictive performance of the methods is also measured by implementing a real dataset. The regularized machine learning techniques offer to use penalty terms to obtain control over over fitting

naturally. These proposed modifications in predictive measures through machine learning methods will be the potential to provide more accurate predictions of binary outcomes.

*Keywords:* EPV, logistic regression, machine learning, separation.

---

[P-076-CO-21]

**Simulation study: An analysis of the incorporation of mixed variables into principal component approach**

*Md. Nayem Dewan\**, icddr,b, Bangladesh

**Abstract**

Publications that are concerned to principal component analysis (PCA) and wealth index is a very popular topic in economics for last several years. This paper also aims to explore these two things. So far many authors established that the use of traditional PCA approach is not suitable for wealth index construction with different types of data. It is a more serious problem when the data set contains mixed variables. Researchers are trying to solve this issue for many years. This study is dedicated to explore the two ways of using mixed data into PCA. Here we define mixed data as data with a mixture of numerical and categorical variables. Two methods that are discussed and compared are namely nonlinear PCA using homogeneity analysis and PCAmix. Properties of these two methods have also been discussed here with proper references. A large simulation study is carried out to find the best approach for dealing with mixed data. We have used a confirmatory factor analysis model in our simulation study where we consider 30 to 70 percent of discrete variable and rest is continuous. Data are generated from normal, log-normal, uniform and bimodal distribution under two conditions such as randomly selected distribution with random sample size and randomly selected distribution with fixed sample size. From our simulation study, under first condition we have found that PCAmix explains 2.38% more variation, 4 times greater correlation with true scores and gives 28% less misclassification than nonlinear PCA. Almost similar result has been found under second condition. Overall, PCAmix showed better result than nonlinear PCA in terms of proportion of explained variance, misclassification rate and rank correlation. Statistical analysis and simulation were done by using programming language R.

*Keywords:* Principal component analysis, nonlinear PCA, PCAmix, simulation, wealth index

---

[P-078-CO-22]

**A non parametric likelihood ratio test for comparison of several count data model and its application to GATS data**

*Sujan Rudra\** and *Soma Chowdhury Biswas*, University of Chittagong, Bangladesh

**Abstract**

A comparison of count data models is obligatory to evaluate the best performance model. Lack of discrepancy of competing models; it is requisite to know which model performs the best based on the given set of observations. This study proposes a non-parametric likelihood ratio test for the comparison of parametric likelihood multiple models. The proposed multiple ( $m > 2$ ) comparison test particularly will be useful for over-dispersed, mis-specified, nested, non-nested or overlapping count data models. The proposed test is developed on the basis of the Kullback-Leibler Information Criterion (KLIC) and Young (1989) test for comparing two parametric models, comparing two moment-based models Kitamura (2000), comparing parametric and moment-based models Chen (2007). We evaluated the performance of the proposed tests in a Monte Carlo study, and with an example from the GATS (Global Adult Tobacco Survey) data ( $n = 2038$ ). Results show that the effect of competing count data models is not the same propinquity; However, Hurdle Negative Binomial Regression Model is the best-fitted model for the data set.

*Keywords:* KLIC, extension of Young test, count Data models, model selection tests.

---

[P-081-CO-23]

**Variable selection for censored data with greedy algorithm based adaptive quantile regression models**

*Md Nasim Saba Nishat\** and *Md Hasinur Rahaman Khan*, University of Dhaka, Bangladesh

**Abstract**

The sparsity associated with high dimensional censored data becomes more complex, because of the incompleteness due to censoring. In such case, quantile regression provides an attractive and flexible alternative tool to analysis of censored data particularly with accelerated failure time (AFT) model, because the conditional quantile functions are often of direct interest in regression analysis. This study focuses on extending the quantile regression for high-dimensional censored data through AFT models implemented with a set of greedy algorithms' sure independence screening (SIS), tilting and PC-simple that are known as Quant-SIS, Quant-Tilting and Quant-PC respectively. The techniques are proved to be easily adapted efficient and consistent variable selection algorithms for high-dimensional data sets because of sure independence, tilting correlation and partial faithfulness properties. Performance of the proposed methods are conducted with a set of simulation studies conducted under a variety of settings including different collinearity level among the covariates, level of censoring and quantiles. A real microarray data example on Diffuse Large B-cell Lymphoma (DLBCL) patients is used to demonstrate the performance of the proposed methods for identifying correct genes that are related with the survival time of the DLBCL patients. Both analyses suggest that the modified quantile regression techniques for censored data considerably outperform the existing methods for most of the cases and at their least performing as parallel to the traditional methods.

*Keywords:* Accelerated failure time models, high dimensional censored data, greedy algorithms, variable selection.

---

[P-085-CO-24]

### **Patient-specific dose finding in seamless phase I/II clinical trials**

*Shantonu Islam Shanto\** and *M. Iftakhar Alam*, University of Dhaka, Bangladesh

#### **Abstract**

This study proposes an extension of seamless phase I/II design which will consider patients' diversity as covariate to determine the optimum dose. In this study, a simple design is proposed that will take binary values as covariate effect. Trinomial response is assumed here that categories the dose-response into three outcomes and the continuation ratio model is used to model the response variable. The parameters are estimated by Bayesian approach. Four plausible scenarios are discussed that show that the proposed design is likely to identify the true optimum dose more accurately than the case that does not consider any covariate. This study also show the sample size has an effect on the design.

*Keywords:* Dose finding studies, seamless phase I/II clinical trials, patients heterogeneity, optimum dose, dose-response model.

---

[P-086-CO-25]

### **Structural equation model: Biosocial factors affecting adolescent's reproductive health knowledge in rural Bangladesh**

*Md Injamul Haq Methun\**, Tejgaon College, Dhaka, Bangladesh, *Sheikh Giash Uddin*, Jagannath University, Dhaka, Bangladesh, *Md. Iqramul Haq Shimul*, Sher-e-Bangla Agricultural University, Dhaka, Bangladesh, *A. A. S. Saleheen*, Jagannath University, Dhaka, Bangladesh, and *Md. Ismail Hossain*, Jagannath University, Dhaka, Bangladesh

#### **Abstract**

This study aims to explore how social context beyond individuals may influence adolescent's sexual and reproductive health knowledge in the rural area of Bangladesh. The adolescence period is a historical landmark in human life; this period orchestrates the visible transformation of a "child" into an adult. Adolescent's sexual and reproductive health is of the utmost importance for a smooth transition from childhood into adulthood. Unluckily, most of the adolescents of our country are deprived of healthy

sexual and reproductive life. The social contexts of rural Bangladesh accelerate this deprivation by limiting the adolescents' knowledge about sexual and reproductive health. A total of 1211 sample adolescents were interviewed for the study. The structural equation model (SEM) is used to see how social contexts like family, peers, school, and community affect adolescent's sexual and reproductive health knowledge. Whereas, confirmatory factor analysis has been employed to measure the validity and reliability of the latent construct and multi-group confirmatory factor analysis was additionally used for model identification purposes. A latent variable, adolescent's reproductive health knowledge, was constructed based on the perception of contraceptive methods, menstruation, pregnancy, maternal health, and sexually transmitted disease. The results revealed age, mass media exposure, maternal education had a positive impact on reproductive health knowledge. Parents by practicing monitoring, behavioral and psychological control on their adolescent childrens limited their knowledge on reproductive health. Whereas, Teacher's concern for the student and adolescent's peer connection showed positive direct as well as mediating effect on the knowledge level. Besides, the Community of being more disorganized increased the adolescent's reproductive health knowledge level. However, the economic condition of the family and parental working status exposed no significant effect on knowledge. Eventually, it is obvious from the study that essential social contexts such as family and community are still impeding adolescent's reproductive health knowledge.

*Keywords:* Structural equation model (SEM), confirmatory factor analysis (CFA), reproductive health knowledge.

---

[P-095-CO-26]

### Using machine learning algorithm: "Analysis and prediction of diabetes"

*Md. Murad Hossain\** and *Md. Rana Ahmed*, Bangabandhu Sheikh Mujibur Rahman Science and Technology University, Gopalganj, Bangladesh

#### Abstract

Diabetes is a cardiovascular disease. It is not only an epidemic in Bangladesh but also in the whole world that is increasing rapidly. Machine learning techniques (MLT) are used to predict the diabetes datasets at an early stage of safe human life. In our research paper we use Pimadiabetes dataset from Kaggle UCI machine learning data repository. Machine learning techniques (MLT) are cost-effective and time saving for diabetic patients and doctors. Currently in above datasets we apply KNN, Naive Bayes, Bayes Nets, Random forest, Support vector machine, Simple logistic and J48. By using these algorithms with WEKA software we create an ensemble (Vote) hybrid model by combining individual methods which provides the highest performance and accuracy.

*Keywords:*



---

[P-103-CO-27]

**Using penalized methods for propensity score models with rare exposure in observational study with survival outcome**

*Ema Akter\** and *Mohammad Shafiqur Rahman*, University of Dhaka, Bangladesh

**Abstract**

Propensity score methods are increasingly being used for estimating the marginal effects of exposure or treatment controlling for the effect of some baseline confounders in observational studies. The propensity score (PS) is an estimated probability of being exposed/treated obtained by fitting logit or probit model of exposure with the observed confounders. Then the propensity score is used in matching, weighting, and covariate-adjustment for estimating the marginal effect of the exposure/treatment on the outcome. All these procedures perform well when the prevalence of exposure is relatively high. However, there is much doubt about the performance of different PS methods when the prevalence of exposure/treatment is low or rare in practice. This is because the PS model estimated using maximum likelihood approach may provide biased estimate of the predicted probability (i.e., PS) of being exposed or treated. Therefore, the incorrect estimate of the PS may, in turn, influences the estimate of the marginal effect of the exposure. In this research, we proposed penalized maximum likelihood methods for estimating PS model with rare exposure/treatment for achieving the correct estimate of the propensity score. We explored two penalized methods: one is based on Jeffreys invariant prior (Firth, 1993; Puhr et al. 2017) and the other is based on log  $F(1,1)$ -prior (Greenland and Mansournia, 2015). The PS estimated from penalized logit models were then used in matching, weighting, and covariate-adjustment to examine if the penalized PS model improves the performance over the standard ML-based PS model. Simulation results revealed that all the penalized methods performed well, in terms of bias and MSE in the marginal estimate, in comparison with the standard ML-based logit model. Of the penalized methods, the log  $F(1,1)$ -prior based penalized method performed better than the Jeffreys-prior based penalized method. The methods were illustrated using a real dataset having rare exposure.

*Keywords:* Confounding, marginal estimate, bias reduction, Jeffreys prior, log- $F(1,1)$  prior.

---

[P-104-CO-28]

**Estimating heterogeneous causal effects of size at birth on children's nutritional status in Bangladesh**

*Fariha Taskin\** and *Mohaimen Mansur*, University of Dhaka, Bangladesh

**Abstract**

Size at birth has been widely established as a potential risk factor for child malnutrition in public health literature. Existing studies that investigated such relationships are mostly correlational in nature. This paper examines potential causal effect of children's size at birth on their nutritional status using Bangladesh Demographic and Health Survey (BDHS) data. Quantifying causal effect of a treatment variable from survey data is of growing interest, but it is particularly challenging because of observational (non-experimental) nature of data and the possibility of the presence of many confounding variables. We exploit the recent advancements in machine learning algorithms that are specifically developed to address causal inference. State-of-art methods including Causal Tree, Causal Forest, Causal Boosting and Bayesian Additive Regression Tree (BART) have been compared against each other and also against widely used propensity adjusted logistic regression method through a simulation exercise based on real data involving many confounding covariates. Assuming that birth size may have differential effects for different groups of subjects defined by covariates (e.g., poor versus rich or urban versus rural children) we estimate heterogeneous causal effect. Simulation results show superior performance of machine learning methods, especially BART and Causal Forest in providing a more accurate estimate of causal effect of size at birth on level of under nutrition. These findings highlight the importance of systematic investigations of determinants of birth-weights, and also implementing early interventions ensuring increased nutrition and maternity care for pregnant mothers.

*Keywords:* Causal inference, causal tree, causal forest, causal boosting, Bayesian additive regression tree (BART).

---

[P-105-CO-29]

**Performance of some model selection criteria to identify the correct parametric survival model**

*Sancharee Hom Chowdhury\** , *Tahmina Akter*, and *Mohammad Shafiqur Rahman*, University of Dhaka, Bangladesh

**Abstract**

Parametric survival models are commonly used to analyze right censored survival data collected in both health studies and reliability engineering because of their intuitive interpretation and flexibility in prediction. However, the most challenging issue on fitting parametric-model is to identify the correct distribution of baseline survival function in addition to the usual variable selection. Moreover, censoring makes the process more complicated. Mis-specification of the model leads to biased estimate of the model parameters and invalid inference. Several model- selection-criterion proposed in the literature including Akaike Information Criteria (AIC), Bayesian Information Criteria (BIC) and their different versions such as conditional AICc, and BICc, Delta-AIC and BIC, weighted-AIC and BIC. However, it is unknown which criterion perform uniformly to identify the correct model. This study investigates the performance of these criterion using an extensive simulation study with aim to provide some practical

recommendations. We evaluated the criterion firstly in identifying the correct baseline distribution and secondly in selecting correct variables. Several simulation scenarios were considered by varying the sample size and censoring percentage and changing the distribution of true model. The results revealed that all these criterion performed equally when sample size is relatively large and percentage of censoring is low. The performance of all criterion decreased markedly with the increasing degrees of censoring and decreasing sample size, except for the delta AIC and BIC. Both AIC and BIC performed equally in all simulation scenarios. We further illustrated the methods using two case-studies of real data and discussed some practical recommendations.

*Keywords:* Censoring, AIC, BIC, AICc, BICc, delta-AIC and BIC, weighted-AIC and BIC, Cox-Snell residual.

---

[P-112-CO-30]

### Geostatistical exploratory analysis of groundwater level of Barisal

*Hasina Islam\** and *Syed Shahadat Hossain*, University of Dhaka, Bangladesh

#### Abstract

Groundwater is the main source of drinking and irrigation water supplies. Good management of water resources is crucial to sustainable development for agriculture. This study used groundwater level data collected by Bangladesh Water Development Board (BWDB) from 39 tube wells at different locations of Barisal. The geostatistical analysis of groundwater level was conducted by using R-language. At the preliminary step, exploratory spatial data analysis is done in which histogram, trend of data, semivariogram cloud and cross-covariance cloud of the raw data is observed. Spatial interpolation is also carried out using geostatistical kriging model throughout modeling semivariogram and covariance. Mapping piezometric surface was conducted from groundwater level raster.

*Keywords:*

---

[P-113-CO-31]

### Increasing burden of multi-drug resistant tuberculosis in HIV infected individuals: a systematic review and meta-analysis

*Zeeba Zahra Sultana\** , *Farhana Ul Hoque*, *Mohammad Hayatun Nabi*, *Mohammad Delwer Hossain Hawlader*, *Gias U. Ahsan*, *Shakil Ahmed*, and *Ahmed Hossain*, North South University, Bangladesh

**Abstract**

The increasing number of multidrug-resistant tuberculosis (MDR-TB) and the progressive HIV epidemic are serious threats to the global tuberculosis control programs. Contradictory findings on the association were observed in different studies. The aim was to summarize available evidences on the association, and to provide a pooled estimate of risks. We searched PubMed/MEDLINE, Google Scholar and Science Direct databases to select eligible observational studies, published between January 2010 and July 2019 with language restricted to English. Random-effects model was used to obtain the unadjusted odds ratio (OR) of the crude association with 95% CI. Publication bias was affirmed by funnel plot symmetry and confirmed by Egger's test. Heterogeneity was determined by Cochran's Q statistic and quantified by I<sup>2</sup>. The protocol was registered with PROSPERO, number CRD42019132752. We identified 1496 records through database search and after subsequent elimination, included 47 articles (55 101 enrolled TB patients) in our systematic review and meta-analysis based on eligibility criteria. Overall, the effect estimate was 1.47 (95% CI 1.19-1.81, I<sup>2</sup>=80.49%) with substantial heterogeneity and evidence of publication bias (p=0.13). Subgroup analysis revealed that the estimated pooled OR for European countries (OR=2.31, 95% CI 1.80-2.96, I<sup>2</sup>=32.26%) was higher than the other three continents. Further analysis showed that pooled OR was higher for primary MDR-TB (OR=3.13, 95% CI 1.59-6.13, I<sup>2</sup>=13.13%) with no heterogeneity among the studies. Additionally, the pooled OR for MDR-TB and HIV increased with age, higher in female predominant studies, significant in countries with high income and high burden for TB/HIV and/or MDR-TB. The meta-analysis summarizes that the risk of the association increased significantly in the last decade. As such, it has become imperative to improve the detection and overall management of MDR-TB in HIV-endemic settings to achieve the aims of WHO End TB Strategy.

*Keywords:* Multidrug-resistant tuberculosis, HIV, MDR-TB, systematic review, meta-analysis.

---

[P-121-CO-32]

**Early stopping in seamless phase I/II clinical trials**

Noor M Khan\*, Mawlana Bhashani Science and Technology University, Bangladesh, and M. Iftakhar Alam, University of Dhaka, Bangladesh

**Abstract**

In recent years, seamless phase I/II clinical trials have drawn attention for their convenience in finding an optimal dose. Engaging an appropriate number of patients in a trial is always a challenging task. A dynamic stopping rule for seamless phase I/II clinical trials is proposed here so that resources can be saved. That is, the stopping rule aims to save patients from unnecessary toxic or subtherapeutic doses. Particularly speaking, we allow a trial to stop early when width of the confidence intervals for the dose-response parameters become narrower or when the sample size is equal to a predefined size, whichever comes first. Our simulation study of dose-response scenarios in various settings demonstrates that the

proposed stopping rule can engage an appropriate number of patients and therefore, it is suggested to be used in clinical trials.

*Keywords:* Seamless phase I/II trial, early stopping, optimum dose, continuation ratio model.

---

[P-123-CO-33]

### **Classification of elderly health status in Sylhet district, Bangladesh: A frailty index approach**

*Mohammad Kamal Hossain\**, Bangabandhu Sheikh Mujibur Rahman Science and Technology University, Gopalganj, Bangladesh, *Md. Nazrul Islam*, and *Mohammed Taj Uddin*, Shahjalal University of Science and Technology, Bangladesh

#### **Abstract**

Aging, a leading public health concern is a sure event of a person who has a death at the ending part of human natural lifespan. An important issue in public health sector is to assess the elderly health for making sustainable aged health policies. Therefore, our aim is to classify the health status among the aged people in Sylhet district, Bangladesh. A random sample of size 944 elderly persons, aged 55 or more were collected from the Sylhet district in Bangladesh. A frailty index having 33 indicators was computed for each elderly. The health status of the aged person was classified into four groups, e.g., No Frail, Vulnerable, Frail and Most Frail. Appropriate statistical tests of significance were done.

The average  $\pm$  SD age of the elderly is  $62.66 \pm 8.95$  years. About two-third of the older people are frail, 17 percent are vulnerable and 16 are most frail. Results found that age, gender, education, marital status, occupation, religion, type of family and locality of the elderly are significantly associated with the frailty. The old age elderly, female, illiterate and Muslim elderly are more frail than their counterparts. The elderly aged 70 or more are 5 times more likely to be frail than the pre-elderly. It was also observed that female elderly is 1.55 times more likely to be frail than male. Additionally, housewife or others elderly are also have higher risk of frailty than those who are employed in govt. or private services. Frailty has become a major public health condition related to aging. Our study confirms that the health status of early elderly people is at risk. This classification of elderly people in different frailty groups and associated socio-demographic variables will be helpful for public health policymakers in Bangladesh.

*Keywords:* Elderly, health complexity, frailty, public health, Bangladesh.

---

[P-126-CO-34]

**Early childbearing: more women in Bangladesh are having their first child early in life**

*Zaki Farhana\** , *Mohammad Meshbahur Rahman*, and *Kanis Fatama Ferdushi*, Shahjalal University of Science and Technology, Bangladesh

**Abstract**

The age when women become a mother is known as age at first birth and it is a transition marks to a woman into motherhood. But early pregnancy may adversely affect the health of mothers and infant, economic welfare of the family and it caused many health complexities such as bad parental health care and vaccination behavior, fetal death, lower birth weights, etc. To address this issue, this study aims answering the relationships between age at first birth and socio-demographic factors of 17,569 married women of Bangladesh. For determining the associated factors of age at first birth, the Weibull regression model had been used. During the time of teenage, most of the women having experience about their first birth in Bangladesh. Approximately 72.8% of women have given their first birth before age 20 years and the rural women were most suffering by giving early birth than urban women. The significant factors of age at first birth were found Age at marriage, residence, education, husband's education. Higher educated women give 1.326 times late birth as compared to illiterate women and the women whose husband were higher educated gives 0.95 times less early births than the women of illiterate husband. No use of contraception and the involvement of media is also significantly responsible for giving early birth of women. Age at marriage, less use of contraception, husband and women's literacy are the most influential factors explaining variations in age at first birth. Although Bangladesh government have taken some steps regarding this issue, it is still now an alarming problem and can be controlled by improving such problems. It is time to rethink our government and policymakers about the fact and take the necessary steps. A healthy mother can lead a healthy family as well as a healthy nation.

*Keywords:* Early childbearing, socio-demographic factors, frailty model, Weibull regression model, Bangladesh.

---

[P-127-CO-35]

**A spatial multilevel model using conditional autoregressive processes**

*Md. Mynul Islam\** , *Syed Shahadat Hossain*, and *Mohammad Samsul Alam*, University of Dhaka, Bangladesh

**Abstract**

Multilevel Modeling (MLM), where the lower-level units are nested into higher-level ones based on a geographical hierarchy, helps us to understand individual and contextual effects of a complex survey data. However, the classical MLM does not involve spatial effect. It assumes that lower-level units are correlated with the same correlation belonging to a higher level unit and takes into account the higher-level unit as independent of each other. It does not consider directly the distance between them. As a consequence, spatial interaction effects remain unmodeled. To deal with this, the concept of multilevel modeling is extended to propose a spatial multilevel model to incorporate spatial interactions at both lower and higher levels in this study. In the proposed model, the outcome variable is modeled through a spatially dependent random effect at both level and conditional autoregressive process (CAR) is used at both higher and lower level random effects. To assess, the performance of the proposed model, a series of Monte Carlo simulations are conducted. The results show that proposed models perform better than classical MLM to retrieve true model parameters and the true scenario of the process. When the response variable is not itself an autoregressive process but spatially correlated, estimation bias for the regression coefficient for classical MLM is similar to the proposed spatial MLM, but RMSE for the regression coefficient is large for classical MLM. When the response variable is itself a spatial autoregressive process, not only the estimation bias for the regression coefficient but also the RMSE for regression coefficients provided by the classical MLM is very large compared to proposed spatial MLM. Also, it is seen that the MLM produces biased estimates for the variance parameters. As a demonstration of the proposed model, we applied it to HYVboro rice production data in Bangladesh.

*Keywords:* Spatial multilevel model, spatial interaction, CAR, MLM, higher level effect.

---

[P-128-CO-36]

**Tree-based ensemble methods: A comparison using simulation and application to hypertension data**

*Mohsina Akanda Tusty\** and *Mohaimen Mansur*, University of Dhaka, Bangladesh

**Abstract**

In Bangladesh today, hypertension stands as one of the most prevalent non-communicable diseases and it is a proven risk factor for various fatal health consequences including deaths through heart failures and other cardiovascular diseases. Since the prevalence of hypertension in the country has increased substantially and consistently over years, need for a better understanding of social, demographic, economic and geographical factors and how they interact to contribute to the risk of hypertension has become ever more important. Related empirical studies often miss the important interplay of these factors. Possible reasons for this negligence are over-reliance of parametric regression type models which suffer from dimensionality problem arising from interactions of different orders and the difficulty in identifying potentially important factors in advance. In this paper we exploit the classification tree, a non-parametric statistical learning approach, which is well-suited to capture complex non-linear interactions among

variables. The method is applied to the 2011 Bangladesh Demographic Health Survey Data to identify socio-demographic groups at high risk of hypertension. Tree-based interactions which are simple rules involving predictors are then incorporated in a logistic regression to quantify their effects and related statistical significance. Among women, elderly (aged 45 years or above) individuals with relatively higher BMI and who are diabetic record the highest prevalence of hypertension (almost 70%). Among men, higher-BMI elderly individuals who belong to rich wealth class constitute the highest risk group. These findings have potential implications for targeted and multi-factor focused interventions aiming reduction of hypertension.

*Keywords:* Machine learning, tree-based ensemble methods, hypertension.

---

[P-133-CO-37]

### **Determining the impact of employment status on contraceptive use in Bangladesh**

*Arzo Ahmed\** , *G K Paul*, and *P Deb*, Mawlana Bhashani Science and Technology University, Bangladesh

#### **Abstract**

Population explosion is one of the alarming issues in all around the world that's why is used to prevent for contraceptives method. Contraceptive use plays a significant role in controlling fertility, particularly in reaching the replacement level of fertility. Using Bangladesh Demographic and Health Survey (BDHS)-2014 data this study attempts to examine the current contraceptive use status of married women. For doing this information of 16,830 married women were extracted. The analysis showed that the contraceptive use was found higher among employed women than that of unemployed women. Women's age, region, place of residence, religion, number of living children, and child preference were found to be significantly associated with current use of contraception among employed women. On the other hand, women's age, education, region, place of residence, religion, number of living children, ever heard about family planning and child preference were identified as the significant predictors of contraceptive use among unemployed women. The study also shows that sex preference has a significant effect on the use of contraceptive methods. It was also observed that female education has a positive effect on contraceptive use among almost all regions.

*Keywords:* Contraceptive, employed, BDHS, education level.

---

[P-136-CO-38]

### **The risk of plasmodium vivax parasitaemia after plasmodium falciparum infection: an individual patient data meta-analysis**

*Mohammad S Hossain\** , icddr,b, Bangladesh



**Abstract**

A recent literature review highlighted a high risk of *P. vivax* parasitaemia following treatment of *P. falciparum* infections. We evaluated the risk of recurrence of *P. vivax* parasitaemia and associated risk factors following treatment of *P. falciparum* in co-endemic locations. A systematic review identified efficacy studies of uncomplicated falciparum malaria treated with artemisinin-based combination therapy, undertaken in regions co-endemic for *P. vivax*. Investigators were invited to share individual patient data which were pooled using standardised methodology. The risk of vivax parasitaemia at day 42 and 63 and associated risk factors were investigated by multivariable Cox regression analyses.

A total of 33 studies, enrolling 13,427 patients from 10 countries were included in the analysis. Recurrent parasitaemia between day 7 and 42 was recorded in 1,722 (12.8%) patients, of which 899 (52.2%) were due to *P. vivax* mono-infection, 98 (5.7%) mixed *P. vivax* and *P. falciparum*, and 725 (42.1%) *P. falciparum* mono-infection. At day 42, the cumulative risk of vivax parasitaemia following treatment of *P. falciparum* with artemether-lumefantrine (AL) was 33.0% (95%CI 30.6-35.6), compared to 16.1% (12.2-21.1) after artesunate-amodiaquine (AA), 7.4% (6.7-8.2) after artesunate-mefloquine (AM) and 4.1% (3.4-5.0) after dihydroartemisinin-piperaquine (DP);  $p < 0.001$ . By day 63 the risks had risen to 41.0% (37.7-44.5), 43.5% (35.8-52.2), 23.0% (21.4-24.6) and 14.3% (12.5-16.3), respectively. In multivariable analysis, the risk of vivax parasitaemia was greatest in patients recruited from areas of short relapse periodicity, or presenting with *P. falciparum* gametocytes, mixed infections, high parasitaemia ( $> 50,000$  parasites/uL) or anaemia ( $Hb < 10$  g/dL). The risk of vivax malaria after falciparum infection is high after all ACTs. In the setting of clinical trials, universal radical cure with primaquine would likely be most effective in south-east Asia. Further studies are needed to confirm these benefits in areas of low malaria transmission.

*Keywords:*

---

[P-138-CO-39]

**Availability of essential and priority medicines for treating sick children among health facilities in Bangladesh**

*Shariful Hakim\**, Chander Hat Degree College, Bangladesh and *Md Jamal Uddin*, Shahjalal University of Science Technology, Bangladesh

**Abstract**

Better availability and accessibility of medicines is a prerequisite for elevating health outcomes in children. We aimed to assess the availability of essential and priority medicines for treating sick children and to identify health facility characteristics associated with the availability of those medicines. The 2014 Bangladesh Health Facility Survey (BHFS) data was used in this study. A total of 1,437 facilities that provide curative care services for sick children was used in the analyses. The outcome variables, essential medicines for sick child care was calculated as a count score of ranging from 0 to 7 and priority

medicine was also calculated as count score of ranging from 0 to 4, by which higher scores indicated the greater availability of medicines. Data we reanalyzed using Poisson regression model. Of the selected facilities, 58 (4%) facilities had no essential medicine and 1243 (86.5%) facilities had no priority medicine for treating sick children-among facilities that offer curative care service for sick children, 85.6% had paracetamol syrup/suspension available on the day of the survey visit. Only 5.7% eligible facilities had ceftriaxone powder for injection available on the day of the survey. The chance of availability of essential medicine was 1.20 times higher in public facilities compared to private-for-profit facilities (Relative Risk (RR): 1.20; Confidence Interval (CI): 0.97-1.47). The chance of availability of priority medicine was 5.08 times higher in those facilities situated in urban areas compared to those situated in rural areas (RR: 5.08; CI: 3.42-7.56). The chance of availability of priority medicine was 1.47 times higher in facilities that had guidelines compared to facilities that had no guidelines (RR: 1.47; CI: 1.16-1.86). Since, priority medicines are in worse supply than essential medicines in Bangladesh health facilities, there is a want for ameliorating the availability of priority medicines for children.

*Keywords:* Essential medicine, priority medicine, BHFS, curative care service, sick children.

---

[P-142-CO-40]

### **An analysis of high dimensional logistic regression for variable selection with lasso technique**

*Sanjib Ghosh\** and *Soma Chowdhury Biswas*, Chittagong University, Bangladesh

#### **Abstract**

High dimensional variable selection is one of the formidable challenges in biomedical statistics. It is also well recognized in many disciplines, and has been the focus of much research. In regression analysis, it is extremely important when we meet situations where a huge number of variables are available. Several variable selection techniques have been developed for linear regression models. Some of the most interesting penalization regression methods are the LASSO, Least Angle Regression (LARS), Elastic Net and Adaptive LASSO. Logistic regression model proved very useful in many contexts, though high dimensionality causes difficulties for model fitting and model selection. This study extended the idea of LASSO to least squares regression and logistic regression with a set of diabetes data to identify the risk factors related to the change in disease status and to give an idea of path diagram using the concept of likelihood which describes the order that variables enter the model. This study, found that the LASSO has not only correctly given much smaller standard errors to the signal predictors, but also the minimum cross-validation error at the point corresponding to a set of coefficient estimates for which only the signal variables are non-zero. Hence cross-validation together with the LASSO has correctly identified the signal variables in the model. It can also be concluded that the ability of the LASSO to shrink and select the variable simultaneously and performance is superior compared to ordinary least squares regression.

*Keywords:* High dimensionality, LASSO, LARS, cross validation.

---

[P-144-CO-41]

**The impact of wage remittances on the economy of SAARC region: An econometric approach**

*Mohammad Mohidul Islam\** and *Muhammad Amir Hossain*, Bangladesh Bank, Bangladesh

**Abstract**

Developing countries having stable population growth or experiencing population dividend have been earning foreign currencies by exporting both skilled and unskilled manpower abroad. The inflow of foreign currencies from the expatriates, known as remittance, plays an important role in the economic development and is considered as a major source of international reserves in comparison with export earnings from goods and FDI inflows respectively. Considering the importance of wage remittances in socioeconomic development in the developing economies, this study tries to investigate the impact of wage remittances on the economy of the SAARC countries (Afghanistan, Bangladesh, Bhutan, India, Maldives, Nepal, Pakistan and Sri Lanka) using panel data for the period 2008 to 2017. The empirical analysis is performed employing the econometric techniques of panel data regression approach in order to examine whether the relationship between remittances and economic growth among these countries are significant or not. The findings of the study reveal that wage remittances have significant positive influence on economic growth in the SAARC region, but the impact is not substantial because the remittances are mostly spent for bearing the household expenditure rather than savings and domestic investments.

*Keywords:* Wage remittances, economy, panel data, and SAARC region.

---

[P-146-CO-42]

**Relationship of contraception prevalence rate and total fertility rate in the context of Bangladesh**

*Prianka Sultana Hema\**, , BRAC University, Bangladesh

**Abstract**

The negative relationship between contraception prevalence rate and total fertility rate is well documented in global context. At the same time some of the study suggested that socio-economic factors playing major role for the dramatic decline of fertility in the context of Bangladesh. However, this study aims to explore whether the observed national TFR of Bangladesh is as expected as the national CPR or not. By using the empirical model of the relationship between TFR and CPR developed by Bongaarts and Potter ( $TFR = 7.3 - 0.063 * CPR$ ), this study found that over the last four decades from

1975 to 2014, the observed TFR is lower than the estimated TFR considering the national CPR. Though the contraceptive prevalence rate is increasing without no exception, over time there was an evidence of stagnant TFR for some particular years (1993-97 and 2011-14). This may further emphasize the fact that socio-economic factors may influence the total fertility rate as the social change takes a long time and so the stagnation of TFR can be a usual scenario. This study will be helpful to give a new insight on demographic perspectives in Bangladesh.

*Keywords:* Contraception prevalence rate, total fertility rate, socio-economic factors, Bangladesh.

---

[P-148-CO-43]

### **Performance of CAR and SAR model in smoothing spatial data under different neighborhood assumptions**

*Rownak Jahan Tamanna\** and *Syed Shahadat Hossain*, University of Dhaka, Bangladesh

#### **Abstract**

Areal data is one of the major fields of spatial statistics. There are several models for smoothing areal data. Among those, conditional autoregressive (CAR) model and simultaneous autoregressive (SAR) model are frequently used. These models capture spatial dependence in the covariance structure as a function of a neighborhood matrix. There are several approaches for defining neighbors but there is no strategy to decide on which approach is appropriate one. However the widely used approach is adjacent based neighbor. In our study, we explore the role of choice of neighborhood structure in estimation and prediction by two real life examples. From these two examples we observed that under different neighboring assumption, the performance of the models vary both in terms of estimation and prediction. For examining such variation in the performance of the methods, we conduct a simulation study for both CAR and SAR model under different neighboring assumption. Among many we have chosen eight neighborhood structure for our study. They are adjacent based, Euclidean distance based with cut point 0.6 and 0.8, Fixed point (k-point) distance based for  $k=2$  and  $k=4$ , triangular graph based, Gabriel graph based and relative graph based neighbor. We compare these neighborhood structure in terms of which neighbor gives more robust estimated variance parameter. Also we compared their prediction error. In the simulation study we observed that Euclidean distance based with cut point 0.6 neighbor (M2) performs better both in terms of estimation and prediction irrespective of the true neighborhood scenario of both CAR and SAR models. It gives the smallest root square error of estimated variance parameter for each scenario means it gives more robust estimated variance parameter. Also it performs better in prediction as it gives smaller value of RMSPE in most of the scenarios.

*Keywords:* CAR model, SAR model, neighborhood structure.

---

[P-150-CO-44]

### Generalized linear model for multivariate polytomous response

*Naznin Sultana\**, *A. H. M. Mahbub Latif*, and *M Ataharul Islam*, University of Dhaka, Bangladesh

#### Abstract

Multivariate polytomous outcome, which are often observed in public health research, can be aroused from both longitudinal (e.g. by measuring a variable repeatedly over time from the same subject) and observational studies (e.g. more than one categorical variables are used to define the outcome). Instead of using the commonly used models for analyzing correlated discrete response, such as the marginal models (e.g. generalized estimation equations methods) and the conditional models (e.g. generalized linear mixed effects models), a joint modeling approach is considered in this study. Marginal models occasionally suffer from inappropriate specification of dependence in responses and on the other hand, conditional models require additional assumptions regarding the distribution of random effects. In the proposed approach, joint probabilities corresponding to multivariate response are obtained by the product of marginal and conditional probabilities, which are defined by the Markovian structure of a specific order. Using such modeling allows considering different covariates for different marginal and conditional models. As an example, the data extracted from the Bangladesh health survey (BDHS 2014) are used to show an application of the proposed model. The response of interest birthweight (measured categorically) is assumed to depend on antenatal care during pregnancy, and antenatal care is then assumed to depend on women autonomy. The joint probabilities of these three responses can be defined by the product of the marginal probability of women autonomy, and conditional probabilities of antenatal care services given women autonomy, and birthweight given other two responses. Important factors for different marginal and conditional models are identified. Prediction for birthweight is made from the joint probabilities. The area under the ROC curve shows that the proposed model has the better predictive ability in comparison with the other models.

*Keywords:* Categorical outcomes, marginal probabilities, conditional probabilities, joint probabilities, prediction, ROC.

---

[P-154-CO-45]

### Survival probabilities of stomach and colon cancer patients in Bangladesh

*Tanvir Ahammed\**, *Afroza Jannat Suchana*, *Anisa Ahmed Chowdhury*, *Md. Rafil Tazir Shah*, Shahjalal University of Science and Technology Sylhet, *M. S. Rahman*, Begum Rokeya University, Rangpur, Bangladesh, *A. Akter*, M. C. College, Sylhet, National University, Bangladesh, and *Mohammad Ohid Ullah*, Shahjalal University of Science and Technology, Sylhet, Bangladesh

**Abstract**

Cancer is predicted to be an increasingly important cause of morbidity and mortality in the next few decades, in all regions of the world. In Bangladesh, about 150781 new cancer cases occur in 2018 and caused about 108137 deaths. Stomach and colon cancer have 6.3% and 2% of the total cancer deaths in Bangladesh. However, to the best of knowledge, no study has been conducted in Bangladesh on the survival probabilities of stomach and colon cancer patients. Therefore, we aimed to conduct a study on the survival probabilities of stomach and colon cancer patients with respect to sex, age, and body mass index. In this study, we observed only sixty-five patients with stomach or colon cancer from the eight divisional cities in Bangladesh. A non-parametric, Kaplan-Meier product limit estimate for survival probabilities was used in this study. We found less median survival time for female patients (22 months) than for males (31 months). A similar pattern was observed for patients older than 45 years compared to younger patients. We did not observe any differences in survival probabilities of over weight and healthy weight. Taken together, we conclude that male and younger patients; survival probabilities are higher than female and older patients respectively. Further study may be needed including more samples.

*Keywords:* Cancer, stomach, colon, K-M estimates, survival probabilities

---

[P-155-CO-46]

**Deep neural network for predicting neonatal mortality in Bangladesh**

*Srizan Chowdhury\**, BRAC University, Bangladesh, *Tamanna Howlader*, University of Dhaka, Bangladesh, and *Nahian Ibn Hasan*, BRAC University, Bangladesh

**Abstract**

Recent studies have compared the performance of deep neural networks (DNNs) and Cox PH models in survival prediction, mostly using clinical survival data e.g.- cancer patients' data. However, DNNs have never been applied to large scale survey data, such as mortality data extracted from demographic health surveys. In such data, precise clinical measurements are unavailable and it is difficult to map indirect and convoluted relationships from a set of socio-demographic features to a survival outcome in the presence of large number of covariates and heavy censoring. In this study, simulations were designed to compare predictive performances of Cox PH and DNN using different C-indices under multiple settings obtained by varying sample size, number of covariates, shape of baseline hazard, censoring rate and by considering linear and non-linear specifications of covariates in the risk function. Parameters of Cox PH model were estimated from training data and applied to test data. Modern DNN optimizing hyperparameters, such as size and depth of network, optimizer, activation function, l1 and l2 regularizers were used. Comparisons were also made on neonatal mortality data extracted from Bangladesh Demographic and Health Surveys 2014. Simulation results indicate that for linear risk function, DNN performs in parallel with Cox PH and does slightly better for heavy censoring and small samples. For non-linear risk function, DNN distinctly outperforms Cox PH in all settings. In the

real data, it was found that choice of predictor set had an influence on predictive performance of DNN relative to Cox PH. For example, DNN performed significantly better than Cox PH models with LASSO, ridge and elastic net regularizations when the predictor set contained a mixture of socio-demographic, antenatal care and delivery related variables. The study concludes that DNN holds promise in predicting survival times derived from large scale survey data.

*Keywords:* Survival prediction, deep neural network, socio-demographic, survey data.

---

[P-163-CO-47]

**Exploring the influencing factors for contraceptive use among married women: A special mixed method study of Bangladesh and other twenty low and middle income countries based on DHS data**

*Md. Akhtarul Islam\** , *Hasin Raihan*, and *Md Nafzul Alam Khan*, Khulna University, Bangladesh

**Abstract**

Contraceptive is a method to control pregnancy. High Contraceptive uses prevail in Countries with a high growth rate. Different Social, economic factors are related to high contraceptive use. More than one in ten married women worldwide need family planning; that is then what to stop or delay childbearing. Finding factors influencing the use of contraceptives among married women from Bangladesh and other low and middle-income countries, determine which countries use more contraceptive methods, to raise awareness among women for the proper use of contraceptive, to see whether married women from Bangladesh uses more Contraceptive than other countries.

Individual data on contraceptive use along with other information and partner's information were extracted From DHS data set, we use Binary Logistic Regression for BDHS 2014, then apply meta-analysis approach to 21 DHS data to find out the consistency of the significant variables among lower and middle-class countries. We use the random-effects model and forest plot for determining the effects and relationship of the variables We have found that, women's age (.654), women's education level OR (1.225), media exposure OR (1.10), wealth index OR (.885), Husband/partner's Education OR (1.052), women's working status OR (1.44), to be significant variable for using contraceptive. For meta-analysis we found Respondent Education (OR: 1.34), Husband's Education (OR: 1.52), Respondent Currently Working (OR: 1.38), Respondent Media Exposure (OR: 1.42) are consistent factors across the countries. Thus we have found that Media Exposure, Women's education, Husbands education, Women's working status are the key factors behind more contraceptive use among women in Bangladesh and other low and middle-income countries.

*Keywords:* Logistic regression, meta-analysis, DHS, contraceptive use.

---

[P-165-CO-48]

### **A cause and effect study on road accident in Bangladesh**

*Keya Rani Das\**, Bangabandhu Sheikh Mujibur Rahman Agricultural University, Bangladesh

#### **Abstract**

In Bangladesh, road traffic accident is an alarming issue which creates injuries and sometimes damage of lives. To reduce road traffic accident we need to study about the significant causes which may lead on this issue. In this study, the main purpose is to find out the significant causes and its effect on road traffic accident in Bangladesh. In this regard, 100 data was collected from each division in Bangladesh which constitute a total of 800 observation checklist through interview method from the drivers, passengers and the pedestrians. This study reveals most significant cause is unskilled drivers for the road traffic accident in Bangladesh. According to the Pareto diagram the Primary data also shows that six main causes are more liable for almost 80% road accident in Bangladesh. This study will recommend some curative measure and suggestions to reduce road traffic accident in Bangladesh.

*Keywords:* Road traffic accident, cause and effect, fishbone diagram, Pareto diagram, Bangladesh.

---

[P-170-CO-49]

### **Estimation of coverage of antenatal and postnatal care at district level in Bangladesh: An application of small area estimation method**

*Sabina Islam, Md. Ahadur Rahman\**, S and *Mossamet Kamrun Nesa*, Shahjalal University of Science and Technology, Sylhet, Bangladesh

#### **Abstract**

At present, there has been a raising demand for district level estimates in Bangladesh due to existence of decentralized approach to government and policymakers. National level prevalence of maternal health care indicators can be obtained from a nationwide survey, but prevalence of micro-level administrative units is not possible due to lack of data. In such case small area estimation (SAE) techniques are applied by combining a survey data with a contemporaneous census data. As maternal health care indicators, ANC and PNC play most vital role on mothers and infants health condition during and after pregnancy. In this research, we consider Fay-Herriot model as a SAE technique which borrow strength across spatial regions. The resulting Fay-Herriot estimates of ANC and PNC are compared with the corresponding direct estimates. Direct area-level estimates of maternal health indicators are obtained from 2011 Bangladesh Demographic Health Survey. A Fay-Herriot model for predicting small area estimates with



their mean squared errors is fitted using explanatory variables from the 2011 Population Census. For ANC and PNC in Chittagong Division, Khagrachhari, Bandarban and Rangamati districts and in Sylhet Division, Sunamganj, Maulvibazar, and Habiganj districts are found highly vulnerable.

*Keywords:* Small area estimation, Fay-Herriot model, mean squared error.

---

[P-177-CO-50]

**Regional and country-level variation of prevalence of malnutrition under five children in South Asian countries: An evidence from 2014-2017 demographic and health surveys**

*Md Atiqul Islam\**, *Ruhul Amin*, *Md Najmul Islam*, and *Md Sium Hossain*, Shahjalal University of Science and Technology, Sylhet, Bangladesh

**Abstract**

**Background:** Childhood malnutrition is one of the major global public health issues, especially in developing countries. The aim of the study is to identify the major risk factors of stunting, underweight and wasting among children under five years of South Asian countries taking into account the random effects at the hierarchical geographical structure of the data (Bangladesh, India, Maldives, Myanmar, Nepal and Pakistan). **Methods:** The study used data from the recent six countries Demographic Health Surveys (DHS) conducted from 2014 to 2017. A total of 2,44,966 children aged under five years were included in the analysis. Generalized Linear Mixed Effect Modeling technique was applied to investigate the influence of various risk factors on stunting, underweight and wasting among children under five. **Result:** This study observed that overall 37.66%, 33.39% and 19.36% of the children were stunted, underweight and wasted respectively. The highest prevalence of stunted, underweight and wasted children was found in India and the lowest in the Maldives. Results showed that number of under-five children in the household, number of living children in the household, toilet facility, access to piped water, residence, wealth index, maternal educational status, maternal age, maternal nutritional status, currently child breastfeeding, diarrhea and fever in the two week prior to the survey, children birth order number and total child ever born in the household were significantly associated with the child malnutrition. The study also determined that the overall situation between country level variability was about 13% than the between regional level variability (4-9%) than the cluster level variability (14%). **Conclusion:** This study suggests to improve women's education as educated mothers have a better knowledge of nutrition that will improve childhood malnutrition.

*Keywords:* Malnutrition, South-Asia, stunting, underweight, wasting.

[P-180-CO-51]

**Determinants of hemoglobin level among children (6-59 months) in South Asian countries: A multilevel modeling approach from recent South Asian DHS surveys**

*Ruhul Amin\**, *Md Atiqul Islam*, *Monjur Ahmed*, *Md. Ahadur Rahman*, and *Md Shamiyol Islam Shakil*, Shahjalal University of Science and Technology, Sylhet, Bangladesh

**Abstract**

Hemoglobin concentration (Hb) is used clinically to determine the presence of anemia or other blood disorder. Anemia is one of the crucial global public health headaches, especially in developing countries. Challenge of health care systems, low hemoglobin level impacts are high in children (6-59 months) South Asian countries. The main goal of the study is to identify the risk factors of low Hb concentration of South Asian countries (Bangladesh, India, Nepal, Maldives, and Myanmar) children (6-59 months) considering the multilevel cluster structures. The study used data from the recent five countries Demographic Health Surveys (DHS) (2011 Bangladesh, 2015-16 India, 2016 Nepal, 2016-17 Maldives, and 2015-16 Myanmar) conducted from 2011 to 2016. A total of 2,09,533 children aged from 6 to 59 months were included in the analysis. Linear Mixed Effect Modeling technique was applied to investigate the influence of various risk factors on the hemoglobin levels of children aged 6 to 59 months considering the hierarchical geographical structure of the data. The highest mean of Hb level was found in Maldives 109.92 29 g/L and the lowest in the India 106.39 04 g/L. The mean  $\pm$  SD of Hb level was also observed by children's nutritional status, where stunted, underweight and wasted had the lowest. Result from the linear mixed model showed that number of under-five child in the household, toilet facilities, residence, wealth index, maternal educational status, maternal anemia, maternal age, maternal nutritional status, currently child breastfeeding, diarrhea and fever in the two week prior to the survey, child nutritional status such as stunting, wasting and underweight and total child ever born in the household had significant effect on the child hemoglobin level. Additionally, it was also found that a four-level random intercept model is better compared to the other model in fitting the data well.

*Keywords:* Hemoglobin, hierarchical, mixed model, South-Asia.

[P-183-CO-52]

**Shifting the patterns of overall chronological age into the biological age among senior citizen**

*Mohammad Meshbahur Rahman\** and *Md. Nazrul Islam*, Shahjalal University of Science and Technology, Sylhet, Bangladesh

**Abstract**

There are two age of an individual's, one is chronological age (CA) and another is biological age (BA). BA is an important indicator to assess elderly health and numerous genetic and non-genetic factors contribute to biological aging. The current study focuses to facilitate the elderly health status by assessing the BA among the urban aged people in north-eastern Bangladesh. A random sample of 400 elderly person aged 55 or more has been used and an index known as frailty index (FI) has been computed for each elderly using 30 health indicators. Then using regression and inverse regression approach of the square root of age on the square root of the FI and sex, biological age has been computed for each elderly. Different statistical analyses has been carried out to identify the gaps between chronological age and biological age. The observed average CA and BA of the elderly are found 63.6 ( $\pm 8.73$ ) and 63.61 ( $\pm 3.81$ ) years respectively. To identify the age shifting, we classified both CA and BA into four categories: pre-elderly (aged 55-59), young elderly (aged 60-64 years), old elderly (aged 65-69 years) and oldest elderly (aged 70 and above). BA are increased with the increase of CA until the age group 60-64 years and after that it declined. The distribution of the BA is ranged from 54.29 to 72.40 years, which is closer than the range of CA's (55 to 90 years) and as a consequence, it strongly predicted longer life of the elderly. Overall, the gap between CA and BA found insignificant ( $p > 0.05$ ) but it showed statistically significant ( $p < 0.05$ ) for subgroup (gender and religion basis comparison) analysis. Elderly shifting probability to the upper age groups are showed higher among the pre-elderly age group whereas the old and oldest elderly showed the higher probability of shifting to the lower age groups. Shifting to the higher biological age group from chronological age indicates comparatively poor health status of elderly and vice-versa. From this shifting behavior, it is observed that pre-elderly health is more vulnerable than the others. Biological aging has become a major health condition related to frailty as well as health deficits. This assessment of biological aging will be helpful for public health policymakers in Bangladesh.

*Keywords:* Pattern shifting, biological age, chronological age, senior citizen.

---

[P-184-CO-53]

**Lack of knowledge and misperceptions about thalassaemia among college students in Bangladesh: A community-level cross-sectional study**

*Mahbubul H. Siddiquee\**, *Mahbub Hasan*, *Enayetur Raheem*, and *Mohammad Sorowar Hos-sain*, Biomedical Research Foundation, Bangladesh

**Abstract**

Thalassaemia is a life-threatening but potentially preventable inherited hemoglobin disorder. It is highly prevalent in South Asian countries, including Bangladesh. Raising public awareness is an important component of a thalassaemia prevention strategy, and requires an understanding of the local socio-cultural context. The aim of this study is to assess knowledge and perceptions about thalassaemia among college students in Bangladesh. A cross-sectional survey was undertaken in 1578 randomly

selected college students using a self-administered questionnaire. Over two-thirds (67%) of the students had never heard of thalassaemia. The proportion of students who had heard of thalassaemia was nearly twice as high in urban colleges as in semi-urban or rural setting (46.4% vs. 25.8% respectively). Students from a science background had the highest knowledge scores, while the lowest were reported among students of humanities. Textbooks were found to be the main source of thalassaemia-related knowledge. Nearly 40% of the students were not sure or did not want to be a friend to a thalassaemia patient and 39% either declined or remained hesitant about helping thalassaemia patients by donating blood. Most of the respondents (88%) showed a positive attitude towards “premarital” screening to prevent thalassaemia. This study has identified critical knowledge gaps and societal misperceptions. A better understanding of these issues will be important for disseminating thalassaemia related information in college student. These findings could contribute to developing effective intervention strategies in resource-limited countries like Bangladesh.

*Keywords:*

---

[P-190-CO-54]

### **Comparison of classification performance between decision tree and random forest algorithms: a case study on women empowerment issue in Bangladesh**

*Antora Das\** and *Md. Rezaul Karim*, University of Rajshahi, Bangladesh

#### **Abstract**

Among various supervised classification algorithms, decision tree and random forest both are important and significant classifiers for approaching model and prediction. This paper compares the performances of these two classification algorithms based on a data set relating to women empowerment issue, which mainly focuses on gender discrimination issues on Bangladeshi women. The variable representing the decision making on using the contraceptive method given in the Bangladesh Demographic and Health Survey (BDHS) 2014 data set is considered as an indicator of empowerment of women. The decision tree represents a collection of decision nodes or branches to classify the model, pruning reduces the size of decision trees and improves predictive accuracy. Whereas the random forest includes constructions of decision trees of the given training data, approaching model, matching the test data with these approaching boosting, bagging, out-of-bag error and thus reduces bias and determines prediction accuracy overall the model. This research might be useful to create expert decision making systems and a solution for women rights and decision making policy to construct an optimal prediction model for women empowerment.

*Keywords:* Decision tree, random forest, women empowerment, classification, accuracy, prediction.

---

[P-191-I-64]

### **Divide and Recombine Approach for Fitting Logistic Regression Model for Big Data**

*Md. Kamrul Islam\**, *Md. Razanmiah*, and *Md. Rezaul Karim*, University of Rajshahi, Bangladesh

#### **Abstract**

Divide and recombine (D&R) is a new statistical approach to the analysis of big data. In the D&R approach, the data are divided into manageable subsets, an analytic method is applied independently to each subset, and the outputs are recombined. This paper applies the D&R approach for fitting the logistic regression model for big data. As an example, it considers the Bangladesh Demographic and Health Survey (BDHS) 2014 data set for modeling Bangladeshi women's empowerment status at the household level. It is shown that the D&R approach has great potential in analyzing big data.

*Keywords:* Salinization, tidal water levels, time series, hydro-morphology, salinity.

---

[P-192-CO-55]

### **Spatio-temporal change of green space dynamics and rapid urbanization in Bangladesh**

*Md Mozahidul Islam\**, Biomedical Research Foundation, Bangladesh, *Meshbahur Rahman*, Shahjalal University of Science and Technology, Bangladesh, *Mahbubul H Siddiquee*, and *Mohammad Sorowar Hossain*, Biomedical Research Foundation, Bangladesh

#### **Abstract**

Urbanization has increased globally over the last several decades and it caused many environmental impacts associated with the reduction of inland water body, river area, dense vegetation, light vegetation as well as the reduction of green space dynamics. Insights into how urban green space dynamics changes in response to urban expansion, greening and water policies are very crucial for guiding sustainable urban development. This study employed integrated GIS and statistical approaches to characterize the changing patterns and effects of green space dynamics in Bangladesh from 1999 to 2019. A random sample of 13 major changing districts and its subdistricts (towns) has taken under consideration for the remote sensing-based satellite image processing of the year 1999, 2009 and 2019, and the green space dynamics- built up area, dense vegetation, inland water body, light vegetation and river area (in hectore) are examined. Normalized Difference Vegetation Index (NDVI) has been incorporated which has been optimized and random sampling driven accuracy assessment to match with the field level. Minimum 50 points/per class random accuracy assessment points was generated in ArcGIS 10.5 environment and cross-matched with field throw Google earth pro 7.1 with an overall accuracy per class of 83-90%.

After getting satisfactory accuracy rate, we found that the highest increased rate of buildup area in the year 2009-2019 rather than 1999-2009. In 1999, the average buildup area, dense vegetation, inland water body, light vegetation and river area were 8225.50, 9299.59, 4053.47, 5482.84 and 1575.65 hectares respectively while 13868.16, 3127.47, 1354.64, 9476.47 and 868.16 hectares respectively in 2019. Paired t-test for all green space dynamics has been performed to examine the significant changes between time 1999 to 2009, and 2009 to 2019. From this analysis, built up area and light vegetation in 2009 were found significant ( $p < 0.0001$ ) change of increase from the past to present. The dense vegetation and inland water body are found significant ( $p < 0.05$ ) change of decrease in 2009 and 2019 than their counter year. The river area are also found an increasing but not in a significant rate. The study demonstrated the usefulness of the concentric and directional landscape analyses in characterizing the spatial-temporal variations of urban green space dynamics in Bangladesh with a concentric development form.

*Keywords:* Urbanization, green space dynamics, satellite image processing, Bangladesh.

---

[P-196-CO-56]

### Correlates of stunting among under-five children in Bangladesh: A multilevel approach

*Papia Sultana*, University of Rajshahi, Bangladesh, *Md. Mahfuzur Rahman\**, and *Jahanara Akter*, Global Public Health Research Foundation, Bangladesh

#### Abstract

Child malnutrition still remains a major cause of childhood morbidity and mortality in Bangladesh. This study aims to determine the prevalence and identify the associated risk factors of child malnutrition in Bangladesh using multilevel logistic regression model on data from the Bangladesh Demographic and Health Survey (BDHS), 2014. A total sample of 6965 children aged 0-59 months was extracted from BDHS 2014. We performed descriptive analysis and multilevel generalized linear regression analysis with clustered data structure.

Our findings show that among children the prevalence of moderate and severe values was respectively: 25% and 12% for stunting; 11% and 3.1% for wasting; 25% and 7.9% for underweight. The probability of stunting increased with age, with highest rate among children aged 36-47 months, which was significantly higher than children aged less than 6 months ( $OR=6.71$ , 95% $CI=4.46, 10.10$ ). Female children are found to be 11% less likely to be stunted than male children ( $OR=0.89$ , 95%  $CI=0.78, 1.02$ ). Children with birth interval less than 24 months were significantly more likely to be stunted than children of first birth by 36% ( $OR=1.36$ , 95%  $CI=1.11, 1.67$ ). Mothers with a normal BMI were 16% less likely to have children with stunting compared to mothers who are underweight ( $OR=0.84$ , 95%  $CI=0.76, 0.93$ ). Other factors which were associated with a higher risk of stunting included parents with lower educational levels, children from the poorest wealth index, and mothers aged less than 20 years as first birth. Government and non-government organization should generate effective program to aware women of reproductive age about adverse effect of short birth interval, and to aware parents about standard height and weight according to age and gender of children. Overall, necessary steps may be taken

to make people educated and to reduce household wealth inequality to improve nutritional status of children.

*Keywords:* Child malnutrition, stunting, prevalence, multilevel generalized linear regression, odds Ratio, Bangladesh

---

[P-197-CO-57]

**Patterning prevalence and correlates of stunting among under-five children in Bangladesh based on BDHS (2007-2014) data**

*Md. Tahidur Rahman\**, Islamic University, Bangladesh, *Noyon Ahmed, Md. Jahangir Alam, Dulal Chandra Roy, and Papia Sultana*, University of Rajshahi, Bangladesh

**Abstract**

Malnutrition is referred to as the greatest single threat to the world's public health, especially for developing countries. Rates of malnutrition in Bangladesh are among the highest in the world, with six million children estimated to be chronically undernourished (DHS, 2011). The nutritional status of under-five children is the most sensitive indicator of a society, a country as well as the world's public health status. This study investigated the differential impact of some demographic, socio-economic, environmental and health related factors on nutritional status. The objectives of this study are to find the prevalence and patterns of stunting, wasting and underweight among under-five children in Bangladesh from 2007 to 2014.

Secondary data of sample size 5300, 7647 and 6965 are extracted from Bangladesh Demographic and Health Survey (BDHS- 2007, 2011 & 2014) respectively. Indicator of malnutrition stunted (Height-for-age-z-score < -2) was considered as the outcome variable. Along with descriptive analysis and factor analysis had been used to find patterns of malnutrition among under-five children in Bangladesh and its correlates. Prevalence of stunted among the under-five children were 41.70%, 40.67% and 36.57 in 2007, 2011 and 2014 respectively. From the factor analysis, it had been found that top most five correlates for stunting in 2014 were father's education, household wealth index, mother's education, age of child and duration of breast feeding, in 2011 those were household wealth index, mother's education, father's education, father's occupation and frequency of antenatal visits during pregnancy; and in 2007 those were duration of breast feeding, age of child, father's education, mother's education and household wealth index respectively. This study revealed that father's education level, wealth index, mother education, age of children and duration of breastfeeding are the most significant factors and these factors are directly influenced on the stunting among under-five children Bangladesh. For improving the nutritional status of children government or policy maker should take necessary steps for controlling these factors by modifying some of the existed intervention strategy.

*Keywords:* Child malnutrition, stunted, factor analysis.

[P-199-CO-58]

**Prevalence and associated risk factors of hypertension among adults in Bangladesh: A case study in south western Bangladesh***Abdul Muyeed\**, Jatiya Kabi Kazi Nazrul Islam University, Bangladesh**Abstract**

Hypertension is a major non-communicable disease (NCD) around the world, which affects cardiovascular diseases (CVDs) like stroke, kidney failure and heart attack etc. This study aimed to reveal the prevalence of hypertension and risk factors associated with it among the adults in Bangladesh. An observational study on 200 people were used to collect the anthropometrics and socioeconomic data from M.R.S industries limited, Kushtia; Municipal market, Kushtia and different village areas of Koya union. The Chi-square test and multinomial logistic regression were used to assess the degree of association and net effect of risk factors on hypertension using odd ratios (ORs). Almost 72.5% male and 27.5% female participants were suffering from hypertension and among them 75% were working people. The Chi-square test revealed that the gender (0.017\*), residence area (0.005\*\*), working status (0.049\*), diabetes history (0.002\*\*), smoking habit (0.003\*\*), food habit (0.002\*\*), sleeping status (0.002\*\*), education (0.034\*), family record (0.006\*\*), age (0.000\*\*) and wealth index (0.020\*) were significantly associated with hypertension. The male patients were less likely (OR=0.418) to have normal pressure than female patients but more likely (OR=2.707) to have low pressure. Diabetic patients were less likely (OR=0.362) to have normal pressure than non-diabetic patients but more likely (OR=1.212) to have low pressure. Smoker participants were less likely to have normal and low pressure (OR=0.112 and OR=0.063) than nonsmoker. Those who had no problem in sleeping were more likely (OR=0.362) to have normal pressure than who had suffered from insomnia and also more likely (OR=13.329) to have low pressure. Wealthier and old aged people were less likely to have normal and low pressure compared to poorer and younger people. The policy makers of health system in Bangladesh need to develop effective strategies including early diagnosis, creating awareness via mass media and should undertake long run policy about this concern.

*Keywords:* Hypertension, risk factors, Chi-Square test, multinomial logistic regression, odds ratio.

[P-200-CO-59]

**Effect of maternal factors on nutritional status of under-five children in Bangladesh: A statistical investigation***Papia Sultana* and *Umma Shakera Naima\**, Rajshahi University, Bangladesh



**Abstract**

Nutritional status of children is an indicator of nutritional profile of the entire community. Poor nutritional status of children is a major public health problem throughout the developing world. Maternal factors may significantly affect child's nutritional status. Therefore, aim of this study was to investigate the effect of maternal factors on under-five child malnutrition in Bangladesh using logistic regression analysis. This study was based on secondary data extracted from BDHS-2014 collected under the authority of NIPORT. Information on 6965 woman's children was considered for measuring nutritional status. This study classified the nutritional status of under-five children on the basis of Z-scores and standard deviations. The anthropometric method for measuring the nutritional status included three widely used indicators to assess the growth of children: height-for-age (stunted), weight-for-height (wasted), and weight-for-age (underweight). Logistic regression analysis was performed to evaluate the association of nutritional status with various factors along with descriptive analysis using Stata11.

From the bivariate analysis, it was found that 12.20% were severely malnourished, 24.91% were moderately malnourished and 62.89% were normal (well-nourished) for height for age. 8.34% were severely malnourished, 24.69% were moderately malnourished and 66.96% were normal (well-nourished) for weight for age. 3.19% were severely malnourished, 11.74% were moderately malnourished and 85.07% were normal (well-nourished) for weight for height. Logistic regression analysis showed that mother's age, mother's education, mother's BMI, mother's age at 1st birth were significant to child nutrition. Good nutritional child carried healthy mother considered a precondition for the long-term socioeconomic progress of a society. Children are the most vulnerable members of society and to allow their development to be affected by poor nutrition is to perpetuate the vicious cycle of poverty, malnutrition and to waste human potential. Maternal factors significantly affect a child's nutritional status. Government of Bangladesh would be unable to accelerate economic development over the long term until their children are assured of optimal nourished.

*Keywords:* Child malnutrition, maternal factors, stunted, wasted, underweight, logistic regression, odds ratio.

---

[P-201-CO-60]

**Reference based genome assembling to recover full genome sequence from NGS dataset**

*Md Mehedi Hasan\**, *Md. Kaderi Kibria*, *Md. Golam Firoj*, *Md. Md. Mazharul Islam*, *Md. Mamun Monir*, and *Md. Nurul Haque Mollah*, Rajshahi University, Bangladesh

**Abstract**

Next-generation sequencing (NGS) is a high-throughput methodology that enables rapid sequencing of the base pairs in DNA or RNA samples. It requires sequence assembling which refers to aligning and merging sequence fragments from a longer DNA sequence in order to reconstruct the original sequence, since there is no sequencer yet to produce full nucleotide sequence of a genome (DNA/RNA). A reference

genome is the initial sequence to which all subsequent sequences are ultimately compared. Assembling the billions of nucleotide bases that make up the conifer genome is a huge challenge, but important because the reference genome leads to the identification of all or most of the genes in an organism, and reveals features of the genome structure, such as the amount and order of repetitive elements and the nature of regulatory elements. Re-sequencing, or sequencing, other individuals within the same species is vastly less time consuming and costly once a reference genome exists. There are several reference based genome assemblers. In this work, an attempt is made to investigate the performance of different algorithms for reference based assembling (Fastp, FastQC, Velvet, Ragout, BWA, Bowtie2, VarScan, SAMtools and BCFtools, etc.) to recover whole genome from NGS datasets highlighting the statistical importance in this area. To perform comparative study of different online reference based assemblers, the necessary NGS datasets are available in the online databases (e.g. SEED, NCBI SRA, MG-RAST, RDP-SILVA, etc.). The output of this work may contribute to the discovery of new drugs/vaccine in the health sectors and high yielding varieties in the agricultural sectors by reducing the cost and time.

*Keywords:* Whole genome sequence, next generation sequencing (NGS), short sequences, reference sequence, assembling algorithm & online software.

---

[P-202-CO-61]

### **Identification of biomarker genes influencing colorectal cancer: Insights from the systems biology approach**

*Samme Amena Tasmia\** and *Md. Nurul Haque Mollah*, Rajshahi University, Bangladesh

#### **Abstract**

Colorectal cancer (CRC) is the 2nd most cause of cancer related death in the world, but early diagnosis ameliorates the survival of CRC. This report directed to identify molecular biomarker signatures in CRC. In this study, systems-based approaches have been considered to explore the potential key biomarker genes and reporter biomolecules (i.e., TFs and miRNAs). We analyzed GSE8671 human CRC microarray gene expression datasets from NCBI-GEO. In total, 64 specimens (32 CRC specimens and 32 normal samples) were analyzed. We used four popular existing methods, such as the classical parametric approach ANOVA(F-test), the non-parametric approaches SAM and KW (Kruskal-Wallis test) and empirical Bayes (EB) approach LIMMA to identify common genes. Differentially expressed genes were identified from these common genes. We performed functional over representation, pathway enrichment, protein-protein interaction (PPI), reporter biomolecules, and survival analyses were done on common DEGs. In this study, we identified 2300 common genes, which are highly significant ( $p\text{-value} < 0.05$ ) using ANOVA, KW, SAM and LIMMA. Total 197 up-regulated using adj.P.Val (adjusted  $p\text{-value}$ )  $< 0.01$  &  $\log\text{FC}$  (log fold change)  $> 2$  and 61 down-regulated using adj.P.Val (adjusted  $p\text{-value}$ )  $< 0.01$  &  $\log\text{FC}$  (log fold change)  $< -2$ . DEGs were detected from these common genes. The 8 hub proteins (RPL22L1, NR3C1, CD44, AXIN2, SRCAP, TNS1, INHBA, MMP1) were selected as proteomic signatures from PPI network using the criteria: nodes degree = 11. Analyses revealed reporter transcription factors and reporter microRNAs as regulatory component. The prognostic power

analysis revealed that hub proteins and reporter biomolecules related with worse survival of patients in CRC. We employed the systems biology analyses where transcriptome datasets were incorporated with molecular networks to reveal molecular biomarker signatures at RNA and protein levels, which showed significant potentiality in CRC.

*Keywords:* Colorectal cancer, differentially expressed genes, biomarkers, protein-protein interaction, reporter biomolecules.

---

[P-203-CO-62]

### **In-silico identification and characterization of RNAi genes in sweet orange**

*Md. Parvez Mosharaf\**, University of Rajshahi, Rajshahi, Bangladesh, *Fee Faysal Ahmed*, Jashore University of Science and Technology, Bangladesh, *Md. Mazharul Islam*, University of Rajshahi, Rajshahi, Bangladesh, *Zobaer Akond*, BARI, Gazipur, Bangladesh, and *Md. Nurul Haque Mollah*, Rajshahi University, Bangladesh

#### **Abstract**

RNA interference (RNAi) plays key roles in both transcriptional and post-transcriptional levels as well as regulates various eukaryotic gene expressions which involved in stress responses, development and maintenance of genome integrity during various developmental stages. The whole mechanism of RNAi pathway is directly involved with the functional interaction activities of Dicer-Like (DCL), Argonaute (AGO) and RNA-dependent RNA polymerase (RDR) gene families. However, the genes of these three families are largely unknown in sweet orange (*Citrus sinensis*), which is an economically important fruit plant. Hence, focusing the significance, a comprehensive genome-wide identification, characterization and diversity analysis of RNAi genes in *C. sinensis* was conducted. Through the various bioinformatics analysis, 4 CsDCL, 8CsAGO and 4 CsRDR as RNAi genes was identified and characterized. Phylogenetic analysis scattered the proposed CsDCLs, CsAGOs and CsRDRs genes into four, six and four subgroups respectively along with the relevant genes of *Arabidopsis thaliana*. The functional domain and motif composition analysis and the gene structure showed that these inevitable regions are conserved into the RNAi genes as like as *Arabidopsis*. The GO enrichment analysis results clearly indicated that the predicted genes have direct involvement into the RNAi process as expected in *C. sinensis*. The sub cellular location of the *C. sinensis* RNAi proteins revealed that the proteins are largely allocated into cytoplasm and nucleus. This study will be more helpful and provide a basis to explore the expression analysis of the proposed RNAi genes in *C. sinensis* for biotechnological application.

*Keywords:* DCL, AGO, RDR, domain and motif, *in silico* approach, sweet orange.

---

[P-205-CO-63]

### Microbial clustering based on 16S rRNA sequence: A comparative study

*Md. Mazharul Islam\**, *Md. Kaderi Kibria*, *Md. Mehedi Hasan*, *Md. Mamun Monir*, and *Md. Nurul Haque Mollah*, Rajshahi University, Bangladesh

#### Abstract

Unsupervised microbial clustering is now playing the vital role to identify the diseases related virus and bacteria. The 16S rRNA sequence is commonly used for identification, classification and quantitation of microbes in complex biological mixtures such as environmental samples like water, soil or air, and microbiome samples. Phylogenetic tree approach is one of the most popular approaches for microbial clustering. The 16S rRNA gene is also used for phylogenetic studies as it is highly conserved between different species of microbes. Molecular phylogenetic analysis is the use of macro molecular sequences to reconstruct the evolutionary relationships between organisms. Phylogenetic trees represent the evolutionary relationships of sequences or species. An optimal alignment of the primary structures and a careful data selection are prerequisites for reliable phylogenetic conclusions. The 16S rRNA based phylogenetic trees can be reconstructed and the significance of their topologies evaluated by applying Neighbor-Joining, UPGMA, Maximum Parsimony and Maximum Likelihood methods of phylogeny inference in comparison, and by fortuitous or directed resampling of the data set. Up to date, nobody compare the phylogenetic tree methods yet. Only few papers was using one of the phylogenetic tree methods among the several methods of phylogenetic tree. In this study, an attempt is made to explore better phylogenetic tree approach for microbial clustering based on 16s RNA sequence. The analyzing results of this study apprise that for microbial clustering, using phylogenetic tree based on sequence dataset, the maximum likelihood method is comparatively better than the other three methods.

*Keywords:* Microbial clustering, 16s rRNA sequence, phylogenetic tree, maximum likelihood approach, diseases related microbes.

---

[P-206-CO-64]

### Regression based fast multi-trait QTL analysis

*Md. Jahangir Alam\**, *Md. Ripter Hossain*, *S.M. Shahinul Islam*, and *Md. Nurul Haque Mollah*, Rajshahi University, Bangladesh

**Abstract**

In many line crossing experiments of genome-wide QTL mapping studies, measurements are taken on multiple traits along with the marker genotypes. Very often, such traits are correlated and there are common chromosome regions that affect multiple traits. Although single-trait simple interval mapping methods can be applied to each trait one-by-one, such approaches do not take into account the pleiotropic effects. The joint analyses of multiple traits, which include all quantitative traits together in a single model, can increase the power of QTL identification and improve the QTL localization accuracy when multiple traits are correlated genetically in the population. Multivariate maximum likelihood estimation (MMLE) and multivariate least squares estimation (MLSE) based interval mapping are two widely used approaches for multi-trait QTL analysis. The recent advancements of technologies facilitate the generation of high-dimensional genotypic marker data along with phenotypic data on multiple traits with a large number of individuals in genome-wide QTL experiments. When the data is high-dimensional, then computational time is a matter of consideration. In this study, we have proposed a new and fast approach for multi-trait QTL analysis based on the assumption that the phenotypes and the conditional probability of putative QTL genotype given the marker genotypes follow a multivariate normal distribution. In this method, the calculation of the LOD statistic is very straight forward because it is calculated only based on the sample variance-covariance matrix of the phenotypes and the conditional probability of QTL genotype given the marker genotypes. Our proposed method is able to identify the same QTL positions as identified by the other two existing methods (MMLE and MLSE). Also, power analysis shows that the proposed method has almost the same power as the existing methods. Moreover, the proposed method takes comparatively less computation time than the existing methods.

*Keywords:* Genome-wide QTL mapping, multiple traits, fast multi-trait QTL analysis, backcross population, multivariate normal distribution.

---

[P-208-CO-65]

**Statistical phenomics in crop growth and development research**

*Md. Matiur Rahaman\**, Bangabandhu Sheikh Mujibur Rahman Science and Technology University, Bangladesh

**Abstract**

In coming decades, global cereal production must need to increase more than 70%-80% from current decades due to the consumption of food, feed, and fuel by a larger, more affluent population. The rapid progression of genomic technology has improved plant research. Recently, available components of "OMICS" approaches offered and integrated suite of new technologies, and we are in a path to improve the description of complex phenotype through high-throughput phenotype technologies. High-throughput plant phenotyping technologies have been accelerated by advanced image sensing and analyzing techniques. Development of this concept is termed as plant Phenomics. The main goal of plant

statistical phenomics is to bridge the genotype-phenotype gap. Bridging the genotype-phenotype gap also demands that large-scale biological ('OMICS') data and associated bioinformatics resources. The advanced statistical algorithms are constantly supporting to bind together such genetic and phenotypic data. Researchers are facing a big challenge during the handling these type of dataset. However, using advanced data-mining (DM) and machine learning (ML) algorithms, it is possible to overcome analytical bottleneck of the Phenomics dataset. DM and ML are an inherently multidisciplinary approach to data analysis that draws inspiration, and borrows heavily, from statistics, probability theory, decision theory, optimization, and visualization and also perform well when "Big Data" problems arise. Several image-based studies have used and evaluated DM and ML methods performance in biology, and images obtained in high-throughput screening through a real-time phenotypic platforms and the generated dataset represent a "Big Data" problem. However, we are interested to discuss the contributions of the statistical learning methods and algorithms in Phenomics study for crop growth and development research.

*Keywords:* Genotype, phenotype, image sensing, phenomics, data mining, machine learning.

---

[P-209-CO-66]

### De-novo assembling to recover whole genome from NGS dataset

*Kaderi Md Kibria\**, *Md. Mehedi Hasan*, *Md. Golam Firoj*, *Md. Mazharul Islam*, *Md. Mamun Monir*, and *Md. Nurul Haque Mollah*, Rajshahi University, Bangladesh

#### Abstract

Next generation sequencing (NGS) allows massively parallel sequencing with thousands to millions of short DNA/RNA sequence fragments in one experiment at considerably low cost compared to other sequencing technologies. It requires sequence assembling which refers to aligning and merging sequence fragments from a longer DNA sequence in order to reconstruct the original sequence, since there is no sequencer yet to produce full nucleotide sequence of a genome (DNA/RNA). De-novo sequence assemblers are a type of algorithm that assembles short nucleotide sequences into longer ones without the use of a reference genome. Different assemblers are designed for different type of read technologies. Reads from second generation technologies (called short read technologies) like Illumina are typically short (with lengths of the order of 50-200 base pairs). However reads from third generation technologies like PacBio and fourth generation technologies like Oxford Nanopore (called long read technologies) are longer with read lengths typically in the thousands or tens of thousands. This necessitates different algorithms for assembly from short and long read technologies. In this work, an attempt is made to investigate the performance of different de-novo assemblers (SPAdes, Ray, ABySS, ALLPATHS-LG, Trinity, HGAP, Falcon, Canu, Hinge, Velvet, etc.) to recover whole genome from NGS datasets highlighting the statistical contributions. To perform comparative study of different online assemblers, the necessary NGS datasets are available in the online databases (e.g. SEED, EBI meta genomics and HMP, MG-RAST, COG, etc.). The output of this work may contribute to the discovery of new drugs/vaccine

in the health sectors and high yielding varieties in the agricultural sectors by reducing the cost and time.

*Keywords:* Whole genome sequence, next generation sequencing (NGS), short sequences, de-novo assembling, computational algorithm and online software.

---

[P-210-CO-67]

### **Detection of toxic chemical compounds and their associated toxicogenomic biomarkers using robust co-clustering approach**

*Mohammad Nazmol Hasan\**, Bangabandhu Sheikh Muzibur Rahman Agricultural University, Bangladesh, and *M. Nurul Haque Mollah*, University of Rajshahi, Bangladesh

#### **Abstract**

Detection of toxic chemical compounds (CCs) or drugs and their associated toxicogenomic biomarkers is one of the most important tasks in the toxicogenomic studies as well as in pre-clinical phase of drug development pipeline. There are few statistical methods for the detection of toxic CCs and their associated toxicogenomic biomarkers. For example, probabilistic hidden variable model (PHVM) and logistic probabilistic hidden variable model (LPHVM). These are the EM (expectation-maximization) based iterative approaches, which produce co-clusters between CCs and genes in absence and presence of outliers in the dataset to detect toxic CCs and their associated biomarkers. However, these approaches consume more time for computation and produces equal number of clusters for CCs and genes though there may be more clusters in the genes than the CCs. Because, a CCs cluster may up-regulate a set of genes and down-regulate another set of genes. Therefore, a CCs cluster makes more than a single co-cluster with the gene clusters. These limitations can be overcome applying co-clustering algorithm (CCA) on the hierarchical clustering (HC) approach. Nonetheless, the HC is also sensitive to outlying observations. On the other hand, toxicogenomic dataset often contains outlying observations. In this regard, the robust hierarchical clustering (RHC) together with CCA can be applied for the detection of toxic CCs and their associated toxicogenomic biomarkers in absence and presence of outlying observations in the dataset.

*Keywords:* Chemical compounds, toxicogenomic biomarkers, logistic probabilistic hidden variable model, co-clustering algorithm, robust hierarchical clustering.

---

[P-211-CO-68]

### **Genome wide association studies to detect important SNPs: A comparative study**

*Md. Golam Firoj\**, *Md. Kaderi Kibria*, *Md. Mehedi Hasan*, *Md. Mamun Monir*, and *Md. Nurul Haque Mollah*, Rajshahi University, Bangladesh

**Abstract**

A single-nucleotide polymorphism (SNP) is a substitution of a single nucleotide that occurs at a specific position in the genome, where each variation is present at a level of more than 1% in the population. SNPs make up over 90% of all human genetic variations, risk to certain diseases and response to drugs. SNPs underline differences in our susceptibility to a wide range of diseases (e.g. Cancer, Sickle-cell anemia, thalassemia, Cystic fibrosis, etc.). There are more than 1000000 SNPs for the human population. Among these huge candidates, few SNPs say 1-20 SNPs are usually significantly associated with the occurrence of a certain diseases or any other phenotypic variations. Identification of these few SNPs responsible for a certain diseases or any other phenotypic variation out of huge number of candidate SNPs is a challenging task. To explore better statistical algorithm for identification of disease related important SNPs, we used both simulated and real SNP datasets, where the real datasets were downloaded from the online databases (Swiss-Port, NCBI, dbSNP, dsSNP, GDB, EMBL). In this study, we considered contingency table, linear regression, ridge regression, linear mixed model, etc. Also some performance measures were used to assess the performance of different methods, such as FDR and power curve, etc. We considered some computer software (EMMA, plink, GenABEL, etc) during the comparative study. We observed that mixed model approach outperform over the other approaches to detect important SNPs.

*Keywords:*

---

[P-212-CO-69]

**Gene expression data analysis using robust model based clustering**

*Fee Faysal Ahmed\**, Jashore University of Science and Technology, Bangladesh, *Md. Hadiul Kabir*, University of Rajshahi, Bangladesh, and *Md. Nurul Haque Mollah*, Rajshahi University, Bangladesh

**Abstract**

The goal of cluster is to clustering given objects into one of the several groups. But most of the existing clusters are very much sensitive to outliers. To overcome this problem, an attempt is made to propose a highly robust Gaussian Mixture Model (GMM) based Clustering by the minimum  $\beta$ -divergence method. The performance of the proposed method depends on gene/feature selection, the value of tuning parameter  $\beta$ , number of clusters, initialization of Gaussian parameters, initialization and breakdown points and detection of outlier dataset. We have discussed some techniques in this paper to improve the performance of the proposed method by tackling these issues. The proposed cluster reduces to the MLE-based Gaussian Mixture Model (GMM) based clustering when  $\beta \rightarrow 0$ . The performance of the proposed method is investigated using both synthetic and real datasets. It is observed that the proposed method improves the performance over the classical GMM clusters in presence of outliers. Otherwise, it keeps equal performance.



*Keywords:* Cluster, Gaussian mixture model, minimum  $\beta$ -divergence estimators, outlier detection, robustness.

---

[P-213-CO-70]

### Computational identification of RNAi genes in wheat genome

*Zobaer Akond\**, BARI, Bangladesh, *Md. Parvez Mosharaf*, and *Md. Nurul Haque Mollah*, Rajshahi University, Bangladesh

#### Abstract

Gene expression and anti-stress phenomenon in plants are maintained by RNA interference (RNAi) genes. This involves three genes such as Argonaute (AGO), Dicer-like (DCL), and RNA-dependent RNA polymerase (RDR). These prospective genes in one of the major cereal crops wheat (*Triticum aestivum* L.) have not been identified and characterized yet. Conserved domain assembly, genetic structure, the phylogenetic relationship, identification of cis-acting components of seven DCLs, gene ontology (GO), and subcellular locations of total 62 RNA silencing identified genes were analyzed in this article. Overall, seven Dicer-like (DCL), 39 Argonaute (AGO) and 16 RNA-dependent RNA polymerase (RDR) genes were found. Phylogenetic analysis with *Arabidopsis* and rice showed that TaDCL, TaAGO and TaRDR genes clustered into four, eight and four subgroups respectively. A number of cis-acting components related to various environmental stressors presented in the 1.5 kb promoter areas of the TaDCL genes implied that TaDCLs are probable to help plants react against different environmental factors. GO classification revealed that a number of biological and molecular pathways are significantly linked to RNAi mechanism in wheat. Most of the RNA silencing genes were found in Cytosol probable to involve in different metabolic pathways in wheat. The results of this study would however help researchers to get necessary clues for more functional genomic investigation of Dicer-like, Argonaute and RDR gene families in wheat against important biotic and a biotic stressors.

*Keywords:* Wheat, RNA Silencing, RNA interference (RNAi), Dicer-like (DCL), Argonaute (AGO), RNA-dependent RNA Polymerase (RNA).

---

[P-215-CO-71]

### Diagnosis of competitiveness among listed companies in Bangladesh

*Shirin Sharmin\**, LR Global Bangladesh AMCL, Bangladesh, and *Mohammad Arman*, North South University, Bangladesh

**Abstract**

Abstract: Measuring the level of competitiveness in an industry should help the decision makers and regulators formulate policies to facilitate proper growth of the industry. Popular measures such as Herfindahl-Hirschman Index (HHI), 3-firm concentration ratio, and entropy concentration index were measured using market capitalization data of companies across different sectors such as banks, non-banking financial institutions (NBFI), insurance companies, IT, hospitality & tourism, pharmaceuticals, textiles, cement, food & allied, fuel & power, and industrial. A total of about 400 companies listed at the Dhaka Stock Exchange (DSE) were considered in this study and daily trading data from last ten years starting from January 2009 were analyzed. Herfindahl index for sectors such as banks, general insurance, textile and industrial sectors were found to be competitive compared to the rest. Here, the average HHI was much less than 1,000 and showing signs of improvement over the last 11 years. Sectors such as fuel & power, pharmaceuticals, NBFI, IT, and life insurance etc. were found to have HHI hovering around 2000 that is indicative of moderate level of market concentration. Food & allied, tourism & leisure, mutual funds, and cement industries were found to have concentrated market with average HHI values exceeding 3,500 that indicates lack of competition. The other measures are also reported and changes in market concentration were compared across the different industries and with respective sectors return.

*Keywords:* Dhaka stock exchange (DSE), competitive market, Herfindahl-Hirschman index, entropy concentration index

---

[P-221-CO-72]

**Morbidity and associated factors of nutrition among under five children of garment workers in Dhaka city**

*Afrin Sadia Rumana\**, Bangladesh University of Professionals, Bangladesh, and *Md. Mazharul Islam*, Bangladesh Institute of Governance and Management, Bangladesh

**Abstract**

The aim of the current study is to examine the relationship between different demographic and socio-economic factor associated with Child Nutrition and Care of Under 5 Children of Garment Workers in Dhaka City. The study also assesses the current status of breastfeeding practice, dietary diversity practices and meal frequency at different age group of respective children. Cross sectional research design was considered where questionnaire survey method was used. The research will be mainly based on informal interview of the female garment worker having children under 5 years old. Convenience sample technique, a part of non-probability sample design was employed to collect the data for conducting analysis. A total of 188 usable questionnaires were found with a response rate of 50.13%. The result of the analysis showed that most of the children of the garment workers did not get proper nutrition facility. Because of the economic condition of the family, children of the garments' worker do not get proper breastmilk from their mother. Furthermore, almost 30% of the female workers only completed

their primary education. In this situation, they did not have proper knowledge about dietary diversity practices and meal frequency for their children. Thus, from the literature support and findings of the analysis, authority should take necessary steps. So that female workers can get proper working environment and careful about the nutrition of their child.

*Keywords:* Morbidity, z-score, sociodemographic factors.

---

[P-236-CO-73]

### Comparison of Statistical Models for Cross-over design

*Md. Kamruzzaman\** and *Yonggab Kim*, Seoul National University, Republic of Korea, *Yeni Lim* and *Oran Kwon*, Ewha Womans University, Republic of Korea, and *Taesung Park*, Seoul National University, Republic of Korea

#### Abstract

Cross-over designs have been widely used in clinical trials to investigate the efficacy of new treatments. In cross-over design, each subject is treated subsequently with different treatments. Many methods such as linear mixed models (LMMs) and generalized estimating equation (GEE) models have been used to analyze the repeated measurements from cross-over design. When we consider repeated measured response variables, estimation of random components for LMMs is not always easy. In this article, we applied the GEE method to cross-over design to overcome the limitation of LMMs. To apply the GEE model to the data from the cross-over designs, we need to switch the role of variables in LMM such a way that the independent variable in LMMs is considered as a response variable in GEE model and vice versa. The purpose of this study is to compare the performance of these GEE models and LMMs for cross-over designs. Through simulation studies, we checked the type I errors and compared power to evaluate the performance of the proposed GEE model and LMMs.

*Keywords:* Correlated data, Cross-over design, Mixed effect model, Generalized estimating equation model.

---

[P-250-CO-74]

### Monitoring small variation of a non-normal process

*M Moinuddin Haider\**, icddr,b, *Mizanur Rahman*, University of North Carolina, Chapel Hill, USA, *Sharad Barkataki*, *Nurul Alam*, and *Quamrun Nahar*, icddr,b

**Abstract**

Bangladesh Demographic and Health Survey 2014 showed that the use rate of long-acting reversible contraceptives (LARC–IUD and implant) and permanent methods (PM–tubectomy and NSV) was 8%, which was less than half of government’s family planning MIS’s (FPMIS) estimate (17%) in the year. This study aims to examine the accuracy of government FPMIS’s estimate of LARC/PM use rate. The study was conducted in 16 purposively selected villages covered by Matlab HDSS and FPMIS. Data collectors of both of the systems visit women on every two months and record their FP use status. Women were matched between the two systems by their names, husband’s name, and village for the reference period of May–June 2017. Independent data collectors visited the villages to identify the unmatched cases and interviewed the unmatched women, and a sample of matched women to collect contraceptive information using a structured questionnaire. Of the 3,057 women in the FPMIS and 2,802 women in the HDSS in May–June 2017, 1,805 were matched between the two sources. The LARC/PM use was 13.2% based on FPMIS records, 8.2% based on HDSS records and 8.2 from the survey. The primary reason for higher LARC/PM use rate in the FPMIS records is the selectively keeping women, in the records, with higher use of contraceptives, including LARC/PM. Many non-users were not found in the FPMIS records and, in contrast, the HDSS records were more complete regardless of method use.

*Keywords:*

## Contributed Posters

---

[P-029-CP-01]

### **Intimate partner violence (IPV) with miscarriages, stillbirths and abortions: Identifying vulnerable households for women in Bangladesh**

*Awan Afiaz\**, University of Dhaka, Bangladesh, *R. K. Biswas*, University of New South Wales, Australia, *R. Shamma*, University of Dhaka, Bangladesh, and *N. Ananna*, Ibrahim Medical College, Dhaka, Bangladesh

#### **Abstract**

Intimate partner violence (IPV) is a persistent problem in Bangladesh and it poses a serious threat to the well-being of women and maternal healthcare. The current study investigated the probable sociodemographic factors that are likely to contribute to IPV in Bangladesh and its overall association with reproductive healthcare of women, particularly in cases of miscarriages, stillbirths and induced abortions (MSA). The study hypothesized that IPV could be associated with MSA and targeted to determine the vulnerable cohorts at risk. A generalized linear model was fitted to the Bangladesh Demographic Health Survey 2007 data consisting of 3,920 women as the sample by adjusting for survey weights, and cluster and strata-wise differences. Results showed that approximately 1 out of every 4 women who were subject to IPV reported having one or more MSA(s) prior to the study. Furthermore, IPV and MSA were significantly associated with the age of women, residence, age of women at their first birth, sex of household head and household financial state. Moreover, IPV was estimated to have significantly increased the odds of experiencing MSA(s) by 35%. There appeared to be a need to address both paradigms, particularly for the poor rural women in Bangladeshi patriarchal society. The findings emphasized that a united intervention effort is required in the vulnerable cohorts, if Bangladesh intends to continue down the path of meeting the Sustainable Development Goals (SDG), particularly 3.1 and 5.2, proposed by the UN.

*Keywords:* Domestic violence, gender discrimination, patriarchy, sociodemographic factors, wealth index, sustainable development goals, BDHS.

---

[P-046-CP-02]

### **Assessing vulnerability to poverty using downside mean semi-deviation approach: An empirical analysis on a balanced panel from rural Bangladesh**

*Md. Israt Rayhan* and *Abdul Basit\**, University of Dhaka, Bangladesh

**Abstract**

This study estimates vulnerability to poverty in rural Bangladesh using the downside mean semi-deviation approach whereby vulnerability is decomposed into severe vulnerability (vulnerability by expected poverty) and moderate vulnerability (vulnerability by downside risk) components. Two rounds of the Bangladesh Integrated household Survey (2011/12 and 2015) designed and conducted by International Food Policy Research Institute (IFPRI) is used to analyze poverty situations in both vulnerability and poverty dimensions. Predictive performance of the vulnerability to poverty index is assessed using accuracy rate, sensitivity, specificity and area under the Receiver Operating Characteristics (ROC) curve based on a simulation study. A majority of the poor and about one-tenth of non-poor households under study are found to be vulnerable to future poverty in 2015. The incidence of both poverty and vulnerability to poverty has reduced throughout the period under study. The incidence of vulnerability to poverty is found to be majorly determined by its mean component, i.e., severe vulnerability. Vulnerability and poverty maps of rural Bangladesh (district level) are constructed to gain insights about the vulnerability and poverty situation at a more disaggregated level.

*Keywords:* Vulnerability to poverty, downside mean semi-deviation, multilevel models.

---

[P-050-CP-03]

**Investigating causal effects of low birth weight on malnutrition in children of age under three years: evidence from Bangladesh demographic and health survey 2014**

*Md. Ariful Islam Sanim\** and *Mohammad Shafiqur Rahman*, University of Dhaka, Bangladesh

**Abstract**

Bangladesh has showed rapid progress in many development indicators by achieving most of the millennium development goals, except for child malnutrition. Therefore, this become a new challenge for government to achieve its target under sustainable development goals by 2030. Given high prevalence of malnutrition among the children in Bangladesh, several studies have been conducted to identify the important risk factors of malnutrition. However, a few of them investigated the causal effects of low birth weight (LBW) on malnutrition, despite its high prevalence. As there are some potential confounders such as child's age, sex, maternal education, socio-economic status etc, associated with both the LBW and malnutrition, it is difficult to estimate the casual effect of LBW on malnutrition in the presence of such confounders. This study aims at estimating the casual effect of LBW on malnutrition after controlling for the confounders using data from the Bangladesh Demographic and Health Survey (BDHS) 2014. Propensity Score (PS) methods such as matching, stratification, covariate adjustment and inverse probability weighting, have been applied to estimate the causal effect. The results revealed that there is significant marginal effect of LBW on stunting, underweight and wasting with odd ratio estimated under PS matching 1.70, 2.13 and 1.68 respectively. Similar results can be observed for the other PS

methods: stratification, weighting, and covariate adjustment. Based on findings some recommendations and policy implications are discussed.

*Keywords:* Malnutrition, causal effect, propensity score.

---

[P-051-CP-04]

### **Impact of sample sizes on the accuracy of estimates for a two-level logistic regression model**

*Md Rasel Biswas\** and *Noor Jahan Akter*, University of Dhaka, Bangladesh

#### **Abstract**

In a multilevel framework an important problem is determining an adequate sample size that produces accurate estimates. While designing a research study, applied practitioners are frequently concerned with the sufficient sample sizes required to estimate unbiased fixed effects and variance components, and produce accurate confidence intervals around the point estimates. In this study, we investigate the sample behavior of estimates in a binary response multilevel model. Particularly, we study the influence of different factors on the accuracy of the estimates and their profile-likelihood confidence intervals, for a 2-level logistic regression model, through a Monte Carlo simulation study. We investigate the condition of: (a) different level-1 sample sizes; (b) different level-2 sample sizes; and (c) different intra-class correlation coefficients. We analyze the bias of the estimates by relative bias and, the accuracy of the confidence interval through a non-coverage indicator. In all cases we have examined, the fixed effect parameter estimates are unbiased (even though the sample size is very small), while the variance components are always underestimated. Despite some exceptions for the variance components, the profile-likelihood confidence interval performs very well in all simulated conditions.

*Keywords:* Multilevel logistic models, simulation, estimation, sample size, profile-likelihood interval.

---

[P-056-CP-05]

### **Determinants of depression among university students**

*Md Nesar Uddin Sorkar\** , Shahjalal University of Science and Technology, Sylhet, Bangladesh, *Md. Taj Uddin*, Shahjalal University of Science and Technology, Sylhet, Bangladesh, and *Mohammad Ziaul Islam Chowdhury*, University of Calgary, Calgary, Canada

**Abstract**

Depression is a state of low mood and displeasure for activity. It can disturb a person's thoughts, motivation, behavior, feelings, and sense of well-being. Depression has severe consequences and can lead to mental disorder and even committing suicide. Depression is quite common in all walks of life. This study attempts to find the factors that are associated with depression among the university students. A total of four hundred randomly selected students from Shahjalal University of Science and Technology was interviewed through a structured questionnaire to identify the potential factors for depression. Basic statistical tools such as Chi-square test and logistic regression were used for the data analysis. From Chi-square test, several factors including feeling withdrawn when around others, amount anything, feeling emotionally empty, feeling worthless, feeling sad, spending more time alone, study disrupted by distracting thoughts, more tired than used to be, rested after sleeping, daily tasks take longer than used to, overwhelmed with challenges in studies, trouble completing study tasks, finding interest in study, finding interest in department subject, taking medicine regularly, feeling insecure in life, and hypertension variables were identified significantly associated with depression. Similarly, the logistic regression analysis finds the factors feeling withdrawn when around others, feeling emotionally empty, division, taking medicine regularly, aim in life, and hypertension are potential risk factors for depression. Depression is highly prevalent among the university students. It is extremely important to find the root cause for depression and take necessary preventive measures. Further research on how to mitigate this problem is also warranted.

*Keywords:* Determinant, University, Student, Depression

---

[P-058-CP-06]

**Socio-economic and demographic factors influencing double burden of anemia among mothers and their under-five children in Bangladesh: Logistic regression analysis**

*Md. Akhtaruzzaman Limon\**, *Md. Abdul Wadood*, *Md. Nurul Islam*, *Md. Ripter Hossain*,  
*ASMA Mamun* and *Md. Golam Hossain*, University of Rajshahi, Rajshahi, Bangladesh

**Abstract**

People of all ages and sexes may develop anemia, but especially mothers and under-five children are the main sufferers. High prevalence of anemia in both women and children create risk of its simultaneous presence in both the mothers and their children in the same households that ultimately would make a major public health concern for the country. These issues are either poorly documented, or not studied yet. Considering the prevailing situation, this study was designed to determine the prevalence and investigate the associated factors of anemia among Bangladeshi mothers of reproductive age (15-49 years) and their under-five (6-59 months) children pair. Data for this study was extracted from the Bangladesh Demographic and Health Survey 2011 (BDHS-2011). Anemic mothers and their anemic under five children were considered as dual burden of anemia. A total of 1,281 pairs of mothers of reproductive age (15-49 years) and their under-five children (6-59 months) were considered as samples in



the current study. Binary logistic regression model was utilized for investigating the associated factors. The dual burden (prevalence) of anemia among the defined mother-children pairs was found to be 45.8%, out of which 54.0% pairs had mild, 44.3% moderate and 1.7% severe anemia. Logistic regression model demonstrated that religion, household wealth index, mothers' nutritional status, number of total ever born children, children's age and current breastfeeding status were the most influential factors (95% confidence interval) associated with anemia among the population. The government and non-government health authorities and other stakeholders should take measures in this context to reduce this cause of maternal and child morbidity and mortality, the two major components of Sustainable Development Goals to be achieved by 2030.

*Keywords:* Anemia, household study, mothers and children pair, public health, bangladesh

---

[P-059-CP-07]

### **Factors influencing early initiation of breastfeeding among Bangladeshi mothers: Two level logistic regression model**

*Ummay Ayesha\**, *Md. Ariful Islam*, *Md. Abdul Wadood*, *ASMA Mamun*, *Md. Ripter Hossain*, *Md. Nurul Islam*, and *Md. Golam Hossain*, Rajshahi University, Bangladesh

#### **Abstract**

Early initiation of breastfeeding (EIBF) is associated with better health of the mothers and reduced risk of neonatal mortality. EIBF increases skin-to-skin contact between mother and infant, and helps to prevent hypothermia of the baby, establishes bonding between mother and child, and increases the potential for exclusive breastfeeding practices. The objective of this study was to determine the prevalence and factors associated with EIBF among Bangladeshi mothers after removing the clustering effects. A total number of 4,092 married non-pregnant Bangladeshi mothers who had at least one child aged under 2 were included in this study. Data was extracted from Bangladesh demographic and health survey (BDHS), 2014. BDHS collected data using two-stage stratified cluster sampling. The observations were derived from several levels of hierarchy, therefore it was possible to obtain a clustering effect in outcome variable. To remove the clustering effect, two levels of multiple logistic regression analysis were used to detect the impact of socioeconomic, demographic and nutritional factors on EIBF among Bangladeshi mothers. The prevalence of early initiation of breastfeeding among Bangladeshi mothers was 51.4%. Two-level logistic regression model demonstrated that mothers living in

*Keywords:* Early initiation of breastfeeding, Reproductive age, Bangladesh, Associated factors, Two-level logistic regression

---

[P-074-CP-08]

**Comparison of multiple logistic regression and artificial neural network models in predicting acute respiratory infections of under-five children in Bangladesh**

*Sabbir Hossain\** and *Nayeem Hasan*, Shahjalal University of Science and Technology Sylhet, Bangladesh

**Abstract**

Acute respiratory infections (ARI) are one of the most important global cause of morbidity and mortality of under-five children and come up with significant health complications for developing countries like Bangladesh. In this study, we compared the performance of multiple logistic regression (MLR) and artificial neural network (ANN) models in predicting ARI of under-five children in Bangladesh. The cross-sectional data were collected from the 2014 Bangladesh Demographic and Health Survey (BDHS), which designed to provide information on demographic indicators, maternal health, and children's health and nutritional status. After excluding missing values and variables that are not related to child ARI, a total of 966 children under-five years were acceptable for our study. Children with a cough coincide with chest related short, rapid breathing in the 2 weeks before the study was assumed having an ARI. ANN and MLR models were used to predict ARI in this study. The performance of the two models were compared by the area under the receiver operating characteristic (AUROC) curve, Kappa statistics and accuracy rate. By using the accuracy rate and Kappa statistic among the 10 ANNs, the ANN with 30 neurons in the hidden layer was considered the best ANN model based on the highest accuracy rate and the kappa statistic. The results showed that the accuracy rate and AUROC curve for ANN (30) model was 96.38% and 93.9% while for the MLR model it was 87.06% and 78.2%. In comparison with the conventional MLR model, the ANN model in this study appeared to be more accurate in predicting ARI of under-five year children.

*Keywords:* Acute respiratory infections, multiple logistic regression, artificial neural network, under-five children, Bangladesh.

---

[P-083-CP-09]

**Rapid consumption methodology: A dynamic approach for household expenditure survey and poverty measures**

*Saifur Rahman Mazumder\** and *Md. Israt Rayhan*, University of Dhaka, Bangladesh

**Abstract**

In the household expenditure survey, the questionnaire records consumption and expenditure for a comprehensive list of food and non-food items. As a result, the administering time often exceeds about 2-3 hours. In shock-prone countries, the face to face interview time is limited for security concerns. So, The lengthy administering time hampers the survey very often. Also because of the cost issue, the poverty estimate is not very frequent in developing country. This study discussed a new approach to do a household expenditure survey and measure the poverty rate which is very cost and time efficient. The list of items are partitioned into one core module and more than one optional modules here. Then every household will be questioned about the core module items and one randomly assigned optional module items. So, the other optional module consumption will be missing for every household. After the survey, the missing module consumption data will be estimated based on the household characteristics and core module consumption. The multiple imputations used over single imputation in estimating the missing values here because single imputation ignores uncertainty and almost always underestimates the variance. And because of the skewed data, Predictive Mean Matching (PMM) method is applied under Multiple Imputation using Chained Equation (MICE) which works efficiently for skewed data. IFPRI's BIHS-2015 data is used in this study and BIHS-2011 is used for cross validation. After estimating the missing data, it is compared with the full consumption data (since in BIHS-2015 and BIHS-2011 full consumption data are available). In BIHS-2015 data, MICE works best with FGT0= 27.85% (from the full consumption data FGT0= 28.01%) and standard error 0.81. Also similar results have been seen with BIHS-2011 data. The face to face interview time will be reduced and poverty can be estimated efficiently by using this methodology.

*Keywords:* Household expenditure survey, Lengthy administering time, Cost and time efficient, Core module, Optional modules, Multiple Imputation, PMM, Mice.

---

[P-089-CP-10]

**Systematic review and meta-analysis of injecting drug use as a risk factor of HIV/AIDS in Bangladesh**

*Noor Jahan Akter and Swarna Chowdhury\**, University of Dhaka, Bangladesh

**Abstract**

HIV/AIDS is a new, emerging problem, especially, among injecting drug users (IDU) in Bangladesh. Although HIV/AIDS prevalence in Bangladesh compared to the surrounding countries is still low, ongoing high-risk behaviors among injecting drug users could facilitate more extensive transmission of HIV/AIDS to the general population. This study aims to assess the relationship between injecting drug use and HIV/AIDS. We included 19 studies which met the inclusion/exclusion criteria. Pooled prevalence, odds ratio, 95% CI and measures of heterogeneity were calculated by random-effects model. Publication bias was examined by funnel plots and Egger's test. The overall pooled prevalence was 3.65% (95% CI: 2.10-5.56%) which indicates that Bangladesh is at the brink of concentrated epidemic.

The female pooled prevalence was low (0.26%; 95% CI: 0.00-1.19%), as expected. But in male IDUs it was much higher (2.44%; 95% CI: 1.51-3.57%). IDU was significantly associated with HIV status. IDUs were 6.085 times more likely to be infected with HIV than non-IDUs (pooled OR: 6.085; 95% CI: 4.654-7.956). Injecting drug use continues to fuel the HIV epidemics spreading throughout Bangladesh. Many pragmatic strategies are being implemented but need further evaluation.

*Keywords:* Meta-analysis, HIV/AIDS, injecting drug users.

---

[P-092-CP-11]

### **Socio-demographic factors associated with women empowerment in Bangladesh: the role of interactions**

*Mehzabin Ahmed Chowdhury\** and *Mohaimen Mansur*, University of Dhaka, Bangladesh

#### **Abstract**

Empowerment of women is one of the most accentuated issues of the present world. The study aimed at investigating the role of interactions among socio-demographic factors in predicting women empowerment. The data used in this survey is extracted from Bangladesh Demographic Health Survey (BDHS) 2014. A total of 17,863 ever married women were considered for the analysis. Classification tree and then logistic regression were implemented to the data for assessing the variable interactions in predicting women's role in various decision making. The exploration shows that only 9.3% of the women take decision on how to spend her earnings herself. Then 13.3% of women take decision on health care alone. Again, only 7.8% of women take decision on large household purchase and 9.3% of them take decision on their visits to relatives alone. These percentages reveal that majority of women does not have decision making power in Bangladesh. The analysis also reports several explanatory variables to be significant as sex of the household head, division, residence, highest education level of women, religion, wealth, current working status. The interactions in some cases are also found predictively significant. This project report highlights the importance of interactions among predictor variables in assessing women empowerment in Bangladesh in addition to individual predictors.

*Keywords:* Interactions, women empowerment, classification tree, logistic regression.

---

[P-108-CP-12]

### **Diagnosis and prediction of heart disease using machine learning and data mining techniques**

*Salman Khurshid\** and *Md. Murad Hossain*, Bangabandhu Sheikh Mujibur Rahman Science and Technology University Gopalganj, Bangladesh

**Abstract**

In medical sciences prediction of Heart disease is one of the most difficult task. Nowadays heart disease is a major cause of morbidity and mortality in modern society. Heart disease is a term that assigns to a large number of medical conditions related to heart. Medical diagnosis is extremely important but complicated task that should be performed accurately and efficiently. Although significant progress has been invented in the diagnosis and treatment of heart disease, further investigation is still needed. The availability of huge amounts of medical data leads to the need for powerful data analysis tools to extract useful knowledge. Heart disease diagnosis is one of the application where data mining and machine learning tools have shown successful results. In our research paper we use Heart disease UCI data set from kaggle machine learning data repository. Machine learning techniques (MLT) are cost effective and time saving for heart disease patients and doctors. In our study we apply the machine learning algorithms KNN, Naive Bayes, Random forest, Logistic regression, Support vector machine, J48 and Decision tree by WEKA to identify which method provides maximum performance and accuracy.

*Keywords:* Heart disease, Data mining, Machine learning, KNN, Naive Bayes, Random forest, Support vector machine, J48, Decision tree

---

[P-110-CP-13]

**Continuously observed multi-state model applied to london hospital neonatal care data**

*Mst. Tanzila Akter\** and *Md Hasinur Rahaman Khan*, University of Dhaka, Bangladesh

**Abstract**

In longitudinal medical studies, it is usually observed that patients are followed over time and covariate information is collected at several occasions and patients move several finite states. A multi-state model is a natural choice to implement to such data. We use a continuously observed multi-state model, a time stochastic process allowing patients to move among a finite number of states to a neonatal care data obtained from the University College London Hospital in 2007. Parametric survival models including the Cox regression model are used for fitting survival data to know the effects of several covariates for the patients. We observed that three transitions, NICU-SCBU, SCBU-Discharge and NICU-Discharge were found significant relative to others. On the other hand, SCBU-HDU, HDU-NICU and HDU-SCBU transitions are found as less important while gestation had been found as significantly associated with transition rate. Birth-weight of newborn and hospital length of stay by mothers were found significant.

*Keywords:* Multi-state model, Cox regression, stochastic process.

[P-114-CP-14]

---

**Association between the type of fuel use with acute respiratory infections among children under-five in Bangladesh**

*Aniqua Anjum\**, *Tanvir Ahammed*, and *Mohammad Nayeem Hasan*, Shahjalal University of Science Technology Sylhet, Bangladesh

**Abstract**

Acute respiratory infections (ARI) can be a leading explanation for mortality in children beneath 5 years in Bangladesh. Low-income families typically rely on fuels like wood, coal, and animal dung for cookery. As it's not clear whether or not the employment of different fuel has a health advantage over biomass fuels, therefore, we aim to conduct a study to discover the consequences of fuel usage on ARI in children. Using Demographic & Health Survey (BDHS) 2014 data, we estimated the effect of fuel use on ARI by constructing logistic regression models. A multiple logistic regression model was applied to assess the risk factors of ARI due to fuel use. From the analysis, we found that crude (only variety of fuel within the model) odds ratios (OR) for the ARI were 1.63 (95% confidence interval (CI): 1.16-2.36) which means children in households using polluting fuels were 63% more likely to have an ARI event than children from households using fossil fuels. Once adjusting for child's sex, age, and place of residence; mother's education, and household wealth index, the effect of biomass fuels is also more acute as adjusted odds ratio (AOR) for the ARI were 1.19 (95% CI: 0.78-1.87) compared with the effect of fossil fuel. Children belonging to cohort 0-11 months were significantly more likely to have ARI (AOR = 1.92; 95% CI: 1.44-2.53) as compared to children belonging to 24-59 age groups. The mother's education level is another major factor in this field. There's an association between fuel usage and ARI in children. This study also identified various socio-demographic, nutritional and environmental risk factors for ARI which may be tackled by effective education of the community and applicable initiatives taken by the government.

*Keywords:* Acute respiratory infections, childhood diseases, public awareness, type of fuel.

[P-117-CP-15]

---

**Comparison of individual ARIMA model and wavelet-ARIMA model for forecasting financial time series data**

*Kazi Umama Nazra\**, *Mohammad Nayeem Hasan*, and *P. Bhattacharja*, Shahjalal University of Science and Technology, Sylhet, Bangladesh

**Abstract**

Wavelet transform (WT) has received much attention in many fields and applications such as physics, engineering, signal processing, applied mathematics and statistics. Recently, it has also received attention among financial time series and financial researchers. In this study, we compared wavelet autoregressive integrated moving average (WA) model with the individual autoregressive integrated moving average (ARIMA) model. The study was performed by daily return data from 2013 to 2018 of Dhaka Stock Exchange (Bangladesh). This paper suggests a novel technique for forecasting the financial time series data, based on WA and individual ARIMA model. A test for constant variability and selection of appropriate variants was performed using the Bartlett's test, and the stability of the transformed series was tested with the Augmented Dickey-Fuller Unit Root Test (ADF) for both the WA and ARIMA models. To find out the most effective model, the two competing models were subjected to comparative analysis tests. Goodness of fit of the model was calculated using the accuracy data criterion (AIC, BIC, and AICc), while the predictive performance was evaluated with RMSE, MAE and MAPE. The results of the analysis show that the wavelet transformation model (WAVELET-ARIMA (0,1,0)) is adequately better fitted from the individual ARIMA (ARIMA (0, 1, 0)) model. In terms of model precision (goodness-of-fit), the model for the WA model is found to have better fit for DSE data (AICc = 953.30) from individual ARIMA model (AICc = 968.84). In forecast performance, the model for wavelet transformed series outperformed (RMSE = 194.99) the model for individual ARIMA (RMSE = 217.54). Therefore, the result of the study shows that the WA model is more accurate than the individual ARIMA model.

*Keywords:* ARIMA model, wavelet transform, time series analysis, Dhaka stock exchange.

---

[P-119-CP-16]

**Missing continuous outcomes in stepped wedge cluster randomised trials**

*Sorif Hossain\** and *Anower Hossain*, University of Dhaka, Bangladesh

**Abstract**

In stepped wedge cluster randomized trials (SW-CRTs), clusters are sequentially randomized until the point at which all clusters are exposed to the intervention groups. Two commonly used analysis approaches for SW-CRTs are cluster level analysis and individual level analysis using linear mixed models (LMM). Missing outcomes are a commonly occurring problem in SW-CRTs which can lead to biased and inefficient inferences if ignored or handled inappropriately. It is plausible to have missing outcomes in SW-CRTs depends on baseline covariates. In this thesis, we will restrict only on continuous outcomes and missingness only in outcomes depends on baseline covariates. We investigated analytically and through simulations the validity of cluster level analysis and linear mixed models using complete record analysis (CRA) and multiple imputed data sets. Cluster level analysis using CRA gives biased estimate unless the missingness mechanism is the same between the two intervention groups. LMM using CRA

gives valid estimate regardless of the the missingness mechanisms. In case of LMM, multiple imputed procedures achieve no gain over CRA.

*Keywords:* Stepped wedge cluster randomized trials, control treatment, missing values, complete record analysis, multiple imputation.

---

[P-122-CP-17]

### **Breastfeeding and childhood diseases: Evidence from nationally representative survey**

*Mohammad Nayeem Hasan\**, *N. N. Nisha*, and *Md Jamal Uddin*, Shahjalal University of Science and Technology, Sylhet, Bangladesh

#### **Abstract**

Breastfeeding is a normal feeding method for infants and young children that ensures optimal growth and development. Moreover, it protects against common childhood illnesses and can also have long-term health benefits for mother and baby. We aimed to determine the association between exclusive breastfeeding (EBF) and common childhood diseases. We used the 2014 Bangladesh demographic and health survey data and considered 763 children aged less than seven months, in which 376 (52%) were non-exclusively breastfed and 387 (48%) were exclusively breastfed. We analyzed data using the Poisson regression model. In adjusted model (important confounding added to the model), infants who were not exclusively breastfed (Risk Ratio (RR) 1.33, 95% CI 1.14-1.55,  $p = 0.000$ ) were found to be a greater risk of having childhood diseases rather than who were exclusively breastfed. The risk of a child getting affected by diseases was more acute if the mothers' age lies between 15-19 years (RR 1.12, 95% CI: 0.78-1.62). Infants belonging to mothers who have secondary education were (RR 0.80, 95% CI: 0.64-1.00) less likely to be affected by diseases than non-educated mothers. Children delivered via normal delivery were (RR 0.77, 95% CI: 0.61-0.97) less likely to be getting diseases compared to the children with C-section delivery. We also found some other factors, such as division, father occupation, C-section, and child's sex, were significantly associated with childhood disease. In conclusion, we found a significant association between EBF and childhood diseases in Bangladesh. So, the necessary steps should be taken to improve the existing situation of exclusive breastfeeding practice by the Ministry of Health and Family Welfare in Bangladesh.

*Keywords:* Exclusive breastfeeding, childhood diseases, Poisson regression.

---

[P-124-CP-18]

### **Predicting malnutrition status of under-five children using tree based models**

*Sabbir Ahmed Hemo\** and *Md. Israt Rayhan*, University of Dhaka, Bangladesh



**Abstract**

Malnutrition is one of the leading causes of morbidity and mortality in children under the age of five in most developing countries like Bangladesh. The main objective of this study is to design a model that predicts the nutritional status of under-five children using tree based model and classical approach. This study used secondary data from Bangladesh Demographic and Health Survey 2014 for 7886 children. Decision tree based model like classification tree, random forest and classical model like multiple binary logistic regression model are fitted to assess the association of malnutrition of children with potential socioeconomic and demographic factors. In this particular study, predictive model is developed using random forest having an accuracy of 70.1% & 72.4% and area under receiver operating characteristic curve of 69.8% and 70% for stunting and underweight respectively. The prevalence of stunting and underweight are found 36.5% and 33% respectively among children aged less than 60 months and higher in rural setting than in urban areas. Similarly, wealth index, exposure of mother to the mass media, age of child, size of child at birth, and parents' education are significantly associated with stunting and underweight of children.

*Keywords:* Predictive modeling, data mining, random forest, classification tree.

---

[P-125-CP-19]

**On the performance of Newton-Raphson, modified Newton-Raphson, Fisher scoring and EM algorithm techniques under maximum likelihood estimation**

*Bushra Uddin\** , *Md. Faisal Ahamad Seyam*, and *Md Hasinur Rahaman Khan*, University of Dhaka, Bangladesh

**Abstract**

The term parameter estimation refers to the process of using sample data to estimate the parameters of the selected distributions. Several parameter estimation methods are available in literature. One of them is the Maximum Likelihood Estimation (MLE) that attempts to find the parameter values which maximize the likelihood function, given the observations. We estimate the parameter value of parametric distributions under the Maximum Likelihood method. Several iterative algorithms including the Newton-Raphson (NR), modified Newton-Raphson, Fisher Scoring and Expectation Maximization (EM) Algorithm will be used in estimating parameters under MLE. The purpose of this study is to examine and compare the performance of traditional and alternative methods of parameter estimation and to suggest which method is good under which situation with several performance criteria - number of iteration, system time and bias. A number of simulation studies has been conducted to investigate and compare the performances of three iterative algorithms and their alternative methods for the estimation of parameters associated with three probability distributions: Poisson, Exponential and Normal. Among the methods, different method is preferable indifferent situation.

*Keywords:* Parameter, estimation, mle, likelihood.

---

[P-129-CP-20]

### Meta-analysis on child malnutrition and it's determinants

Noor Jahan Akter and Safwan Shihab\*, University of Dhaka, Bangladesh

#### Abstract

The intensity of child malnutrition in developing countries is severe especially the ones in South Asia. Bangladesh, India two developing countries of South Asia still suffer from high child malnutrition rates though the rates have declined compared to previous ones. But the rates of these countries are still above the threshold of very high prevalence. This study demonstrates the prevalence of stunting, wasting and underweight for under-five year children of Bangladesh and India and identifies the socio-economic factors responsible for child malnutrition. A meta-analysis has been conducted using 15 articles for estimating the prevalence of stunting and 13 articles for estimating the prevalence of wasting and underweight. The analysis indicates that the prevalence of stunting, wasting and underweight for Bangladesh is lower compared to India. Several risk factors are also found significant which are responsible for high prevalence rates of child malnutrition in these two countries.

*Keywords:* Meta-analysis, child malnutrition.

---

[P-140-CP-21]

### LogitLDA: A logistic transformed linear bayes classifier for gene expression data classification

Priyanka Bosu\*, Bandhan Sarker, Suman Khan, and Md. Matiur Rahaman, Bangabandhu Sheikh Mujibur Rahman Science and Technology University, Gopalganj, Bangladesh

#### Abstract

Gene expression are one of the advanced breakthroughs in molecular biology, which allow monitoring of gene expression for tens of thousands of genes in parallel and produce huge amounts of valuable data. Analysis and handling of such data is becoming one of the major challenges in the life science research. Gene expression data analysis plays the significant role in identifying biomarkers (genes)/patients of the species because of the advanced generation of genomic technologies. Classification and clustering methods are essential to classify/predict the genes/patients that are highly correlated for prognosis diseases, imposing classes of the diagnosed diseases, as well as analyzing and explaining molecular level for detecting potential biomarkers that are linked to causing specific diseases in organism. Statistical approaches employed for the classification of gene-expression data are based on either a continuous scale (e.g., microarray data) or need to fulfill the normality assumptions. However, gene-expression data are

often contaminated by outliers (noise), which affects data analysis results. Therefore, these statistical algorithms can be applied with proper modifications. In my study, we proposed a robust approach for data modification and classification based on some popular data transformation method and suggested a Linear Bayes classifier method named as LogitLDA. To evaluate the performance of our proposed method simulated and a real gene expression data analysis were carried out. The proposed method improves classification results than the classical method when data sets are contaminated. Otherwise, it keeps equal performance.

*Keywords:* Gene expression, microarray, logitLDA, outliers, transformation.

---

[P-141-CP-22]

### **Influence of kernel for disease status prediction through support vector machine**

*Shethika Paul\** , *Md. Shuzon Ali*, and *Nishith Kumar*, Bangabandhu Sheikh Mujibur Rahman Science and Technology University, Gopalganj, Bangladesh

#### **Abstract**

Support Vector Machine (SVM) is a sophisticated state of the art classification technique. The SVM classifier is extensively used in bioinformatics, biostatistics, public health, as well as other disciplines due to its high precision, capability to deal with high-dimensional data and also flexibility in modelling of various types of data. SVM is a supervised multivariate classification method that treats each feature vector as a point in a high dimensional space. Although, there are several kernel functions in the literature, however, the performance of SVM is significantly affected by the choice of a kernel function which is the hyper parameter of SVM among other factors. Since biological data sets may contain noise/outliers and, almost all the existing kernels used in SVM are more or less affected by outliers. In this research, for disease status prediction through SVM, different biological data sets are taken as part of the study. Here, we measure the influence of 9 kernel functions for disease status prediction through support vector machine. The results show that the Laplace kernel function produce higher accuracy and lower misclassification error rate compared to the other kernel function for different data sets. Therefore, our recommendation is to use the Laplace kernel function in SVM for disease status prediction and we believe that this research will be helpful to doctors to predict disease status.

*Keywords:* Support Vector Machine (SVM), kernel function, Laplace kernel function, classification and accuracy.

---

[P-143-CP-23]

### The impact of international oil price on Dhaka stock exchange

*Mohammad Nayeem Hasan*, Shahjalal University of Science and Technology, Sylhet, Bangladesh, *Abdullah Md. Hisham\**, North South University, Dhaka, Bangladesh, and *Kazi Umama Nazra*, Shahjalal University of Science and Technology, Sylhet, Bangladesh

#### Abstract

Stock prices of Dhaka Stock Exchange are highly affected by International Oil Price. Most of the manufacturing company's main raw material is crude oil. If the oil price goes up the cost of production will be high as a result the overall market get affected. So, our aim is to find the actual impact of international oil price on Dhaka stock exchange (DSEX). DSEX is an index which covers almost 97% companies of the Dhaka Stock Exchange. By studying the DSEX price, international oil price and Exchange Rate from February, 2013 to July, 2019 we have identified negative relationship between oil price and DSEX. Our data sources were macro trends.com for oil price, Dhaka Stock Exchange for DSEX and Bangladesh Bank for Exchange rate. We have distributed the data normally first then used Multiple Linear Regression model to analyze the relationship between variables. We also assumed other factors are constant which can affect the DSEX. The result we have revealed, that international oil price has some negative effect on DSEX. The slope coefficient indicates that one unit increase in oil price can reduce the DSEX price by 8.07 units. P-value for oil price is close to zero means our data is highly significant. From R Square we can ensure that, approximately 58% of the observed variation can be explained by this model's input. So, while making policies which affect economy and stock market, policy makers should take the movement of crude prices.

*Keywords:* Dhaka stock exchange, international oil price, slope coefficient, multiple linear regression.

---

[P-149-CP-24]

### A systematic review of causes and solutions of water logging in Dhaka city

*Ahsan Bin Arefin\**, *Ahmed Fahim*, *Haripodo Shill*, *Maisha Maliha*, *Zahid Hasan Mahmood*, *Noor Jahan Akter*, and *Md Hasinur Rahaman Khan*, University of Dhaka, Bangladesh

#### Abstract

Heavy rainfall causes water logging in many places within the Dhaka city, creating manifold problems for its citizen. The rapid urbanization in the recent years has further worsened the water logging conditions of Dhaka city. This study attempts to use the systematic review techniques as an alternative to traditional secondary data review for the study of water logging of Dhaka city. Two dimensions

of analysis are proposed- qualitative analysis and meta analysis. Rigorous protocols are maintained throughout the data collection procedures. Causes and solutions of water logging in the city are primarily evaluated by qualitative analysis. The effect sizes of the causes and the solutions were collected to do the meta-analysis. Results presented a detailed insight into all the areas indicated in the traditional reviews done earlier (Tawhid, 2004; Mowla, & Islam, 2013). The qualitative analyses pointed out the lower gradient, siltation, insufficient drainage, increase of impervious lands, reduction of wetlands, incompatible river water level, river embankment and management limitations as the key factors of water logging in the city. The solutions included rainwater harvesting, supplementary drainage, leaky wells, redesigning of the hydrologic system, pump stations, integrated management and others. The solutions showed considerable improvement in quantity also. From all the observations, it is clear that we must either harmonize all development plans or take supplementary steps to maintain the natural drainage system of the city. To state some, the naturally lower gradient must be compensated by water retention, natural siltation by supplementary drainage, infiltration by direct infiltration (e.g. leaky well) and river water level by water rerouting. It is evident that developing artificial drainage without considering the natural system, like embankment, served short-term purposes for specific areas only.

*Keywords:* Systematic review, traditional review, water logging, Dhaka city.

---

[P-151-CP-25]

### Performance evaluation of machine learning algorithms for MRI brain image classification

*Md. Mamunar Rashid\**, *Md. Najibul Hasan*, and *Nishith Kumar*, Bangabandhu Sheikh Mujibur Rahman Science and Technology University, Gopalganj, Bangladesh

#### Abstract

Magnetic resonance imaging (MRI) is one of the sophisticated medical imaging techniques for diagnosing some particular diseases specially brain tumor, brain cancer etc. Sometimes, Cancer diagnosis is highly invasive, time consuming and expensive. However, one of the main advantages of MRI brain image is that it is a non-invasive and safer approach which saves time as well as cost for detecting brain tumor / abnormality. The traditional approach of brain tumor/ abnormality detection from MRI brain image is manual inspection, which may be inappropriate for large volume of images. Therefore, automated brain tumor / abnormality detection can solve the above problems. In the literature, there are several machine learning techniques including support vector machine, neural networks, k-nearest neighbor, naive Bayes, linear discriminant analysis, random forests etc. for MRI brain image classification. However, the impact/influence of machine learning techniques for MRI image classification are different. Therefore, it is important to find out the suitable machine learning technique for automated brain abnormality detection. In this paper, we have made a comparative study of machine learning algorithms for MRI brain image classification. This study has been based on three stages namely feature extension, dimensionality reduction and classification. We have also measured the performance of different machine learning techniques using classification accuracy, sensitivity and specificity etc. This

analysis will help to choose the suitable machine learning technique for detecting brain abnormality automatically.

*Keywords:* Magnetic resonance imaging, machine learning, classification, feature extraction, dimension reduction.

---

[P-152-CP-26]

### Conditional inference for discrete longitudinal data in the presence of excess zeros

*Zannatul Ferdous\**, *J. F. Antu*, and *T. S. Mallick*, University of Dhaka, Bangladesh

#### Abstract

In medical science, pharmaceutical studies, public health and socio-economic researches, we often encounter the situation of excess zeros in count data. In count data literature, there exist modified count models, such as hurdle and zero-inflated models, to accommodate excess zeros in count data. The extensions of these modified count models to the longitudinal data with excess zeros have been widely used in the literature, where the correlations among repeated responses are modeled through certain random effects. In this research, we have proposed a new regression model for longitudinal count data that incorporates excess zeros as well as first-order autoregressive correlation structure among repeated count responses. A conditional generalized quasi-likelihood (CGQL) estimating equation is derived for estimating the regression parameters; while moment method is used for estimating other nuisance parameters of the model such as correlation and inflation parameters. Extensive simulation studies have been conducted to verify the performances of the proposed estimators. It was found that the CGQL estimates has less biases compared to the existing random effect model based maximum likelihood (ML) estimates, irrespective of the degree of correlation or zero inflation presents in the data set. As an illustration, a pharmaceutical data have been reanalyzed using our proposed approach.

*Keywords:* Longitudinal data, zero inflation, quasi-likelihood, generalized quasi-likelihood, random effect.

---

[P-153-CP-27]

### Effect and control of automotive emission in Dhaka city

*Johirol Islam\**, *Ahsan Hafiz*, *Asif Kabir*, Bangladesh University of Engineering Technology, Dhaka, Bangladesh, *Foysal Mahmud Faraby*, Ahsanullah University of Science Technology, Dhaka, Bangladesh, and *Sajib Aninda Dhar*, Bangladesh University of Engineering and Technology, Dhaka, Bangladesh

**Abstract**

Due to the increase in living standards, using automotive vehicles is expanding day by day. These automotive vehicles transmit many negative substances, for example, Carbon Monoxide (CO), Nitrogen Oxide (NOx), Suspended Particulate issue (SPM), and so on because of deficient burning which was seen as at risky levels in various territories of Dhaka city. This causes numerous medical problems, for example, eye bothering, skin malignant growth, cardiovascular infections, queasiness, asthma, and sickness. In this viewpoint, emission control is significant for lessening air contamination for the occupants of a city. To decrease this air pollution because of outflows from petroleum and diesel motors numerous principles and guidelines have been given in numerous nations alongside the USA. An extensive analysis of energy usage and pollution shows that alternative power systems are still a long way behind the conventional ones. The monthly report of May 2018 shows that the average density of PM2.5 & PM10 were ranged from  $33.8 g/m^3 \sim 59.3 g/m^3$  and  $63.6 g/m^3 \sim 110 g/m^3$ , respectively. These ranges were  $38.2 g/m^3 \sim 69.6 g/m^3$  and  $98.7 \mu g/m^3 \sim 158 \mu g/m^3$ , respectively for PM2.5 and PM10 in the month of May, 2019. The average density of SO<sub>2</sub>, NO<sub>2</sub>, CO and O<sub>3</sub> was ranging from 7.33ppb, 5.57~17ppb, 0.28 ~ 2.65ppm and 2.28 ~ 5.77ppb, respectively in May 2018 in different areas of Dhaka city. This scenario is 1.87 ~ 15.8ppb, 12.6 ~ 20.1ppb, 0.54 ~ 1.09ppm, 5.34 ~ 12.0ppb, respectively for SO<sub>2</sub>, NO<sub>2</sub>, CO and O<sub>3</sub> for the month of May, 2019. For controlling the emission control NAQS (National Air Quality Standard) suggested making Auto Rickshaw (AR) restricted in Dhaka city. Furthermore, different types of technical steps like-Engine Modifications, Combustion Cylinder Alternations, Exhaust Gas Recirculation (EGR), Selective Catalytic Reduction, Non-Selective Catalytic Reduction and Three-Way Catalysts, etc. can be taken in consideration to reduce the emission of pollutants from automotive vehicles.

*Keywords:* Emission control, IC engine, air quality, carbon monoxide.

---

[P-157-CP-28]

**A zero-inflated negative binomial longitudinal model for count data: Generalized quasi-likelihood approach for inference**

*Sabrina Sultana\** and *T. S. Mallick*, University of Dhaka, Dhaka, Bangladesh

**Abstract**

Excess zeros in count data are common problem in practice, where standard count data models, such as Poisson or negative binomial fail to provide valid inference for the parameters associated with the model. As a remedy, zero-inflated models under mixture distribution framework have been used, where a separate binary distribution is assumed to account for the excess zeros in the data. In this context, zero-inflated negative binomial (ZINB) is more flexible than zero-inflated Poisson (ZIP), since ZINB can also accommodate overdispersion present in the data. Analysis of zero-inflated count data in longitudinal studies has become of interest in recent years. To accommodate correlation among zero-inflated count responses, the use of random effects in conjunction with the standard ZIP or ZINB is very popular

and widely used. In this research, a new ZINB model for longitudinal count data with excess zeros has been proposed which accommodates autoregressive of order-1 correlation structure, zero-inflation and overdispersion among repeated count responses. The basic properties of the model have also been investigated. The generalized quasi-likelihood (GQL) estimating equation for the estimation of the regression parameters and moment equation for correlation parameter have been derived. Maximum likelihood (ML) equation for the overdispersion parameter and moment equation for the inflation parameter associated with the proposed model have been used based on responses at the first time point. Extensive simulation study reveals that the proposed inference technique provides estimates with smaller biases as compared to the existing random effect model based ML estimates when the zero-inflated count data are stochastically correlated. A pharmaceutical data have been reanalyzed using our proposed model and inference technique.

*Keywords:* Longitudinal data, zero-inflation, generalized quasi-likelihood, random effect, negative binomial regression.

---

[P-162-CP-29]

### **First-order integer valued AR process for data with excess zeros and its inference**

*Romana Akter\** , *N. M. Ahmed*, and *T. S. Mallick*, University of Dhaka, Bangladesh

#### **Abstract**

Many researchers in medical, social, psychology and behavioral sciences have the interest to examine the effects of a set of covariates on the count response. Poisson regression model is commonly used to analyze this kind of data. In modeling count data, it is common to encounter a large number of zeros than expected which is referred to as zero inflation. In such cases, traditional Poisson model may not be pertinent for inference. Hurdle model and zero inflated Poisson model are most commonly used to adjust for excess zeros in count data. In many public health surveys, medical trials and insurance longitudinal zero inflated count data often occurs. In modeling longitudinal zero inflated count data, an important issue is to incorporate correlation among repeated responses along with zero inflation. Analysis of such longitudinal zero inflated count data is not adequately addressed in the literature. In this research, a new stationary first order integer valued autoregressive process has been introduced which is suitable for data with excess zeros. The basic properties of the model have been investigated. A generalized quasi-likelihood (GQL) approach is proposed for the estimation of the regression parameters of the newly proposed model for longitudinal zero inflated count data, whereas correlation and inflation parameters are proposed to be estimated by method of moments. An extensive simulation study has been conducted and the performance of the proposed estimation approach was verified. It was found that the proposed estimation technique gives unbiased estimates for the parameters of the proposed model for zero inflated longitudinal count data. As an illustration, a real life longitudinal data with excess zeros have been reanalyzed using our proposed model and inference technique.



*Keywords:* Longitudinal data, integer valued process, zero inflation, generalized quasi-likelihood, hurdle model.

---

[P-167-CP-30]

### **Blockwise robust singular value decomposition for MRI brain image denoising**

*Md. Najibul Hasan\**, *Md. Mamunur Rashid*, and *Nishith Kumar*, Bangabandhu Sheikh Mujibur Rahman Science and Technology University, Gopalganj, Bangladesh

#### **Abstract**

Magnetic Resonance Imaging (MRI) is one of the high-dimensional high-throughput technologies that are playing a major role for diagnosing brain tumor. With the help of MRI images, brain tumor is diagnosed at advanced stages. Correct identification of brain tumor or abnormality, Image processing in MRI of brain is highly essential that can reduce the chance of fatal stage. Occasionally, these MR images are introduced with noise during acquisition which reduces the image quality and limits the accuracy in diagnosis. Therefore, preliminary diagnosis of MRI brain images from the hospital may not be always reliable for further analysis because of the presence of noise. Reduction/Elimination of noise in medical images is an important task in preprocessing which is one of the previous crucial parts for further image analysis. Although, several denoising techniques are available in the literature including median filter, wiener filter, wavelet filter, wavelet based wiener, non-local mean (NLM), wavelet based NLM, wavelet based weighted median filter etc. In this paper, we have measured the influence of the above methods for denoising. We have also proposed a new blockwise robust singular value decomposition technique for denoising image. The results of our analysis showed that blockwise robust singular value decomposition technique gives the better performance compared to the other methods. Therefore, our recommendation is to use the blockwise robust singular value decomposition technique for brain image denoising.

*Keywords:* Magnetic resonance imaging (MRI), preprocessing, denoising, image quality metrics, wavelet transformation.

---

[P-169-CP-31]

### **Comparing different measures of model selection for forecasting very short time series: ARIMA vs. exponential smoothing**

*Tasnuva Zaman\**, Bangladesh University of Engineering and Technology, Dhaka, Bangladesh, and *Md. T. Ahmed*, Chittagong University of Engineering and Technology, Bangladesh

**Abstract**

Auto Regressive Integrated Moving Average (ARIMA) and Exponential Smoothing (ES) models are very popular for forecasting non-seasonal linear time series. Though it is easy to forecast with large sample data using the above models, they can also predict with small samples given the model parameters are reduced to suit the sample size. The aim of this paper is to check which statistical measure performs best to select the model with lower error in prediction of short time series. For this, 10 non-seasonal short time series data (n= 36) randomly selected from secondary sources have been analyzed. Each series can be fit to an ES model and its equivalent ARIMA model. As the sample is small, very simple versions of the models with one or two parameters have been generated using SPSS. Validation sampling has been done to test the best forecast performance by each model and the result has been cross checked by root mean squared error (RMSE), R square, mean average percentage error (MAPE), normalized BIC and the log of the accuracy ratio (lnQ2) measures, to test the performance of the selected measures. Different measures suggested using different models. The best fit model suggested by most measures were not the best predicting model for nine cases. However, the log of accuracy ratio performed the best to identify the model whose predictions are closest to the actual values.

*Keywords:* Short time series, model selection measures, ARIMA, exponential smoothing.

---

[P-171-CP-32]

**Prediction of child malnutrition in Bangladesh: A comparison between predictive ability of logistic regression and classification trees**

*Md. Ahadur Rahman, Md Atiqul Islam, Md. Azizul Baten, Md Shamiyol Islam Shakil, Ruhul Amin, and Anisa Ahmed Chowdhury\**, Shahjalal University of Science and Technology, Sylhet Bangladesh

**Abstract**

Child Malnutrition refers to under nutrition where a child is not getting enough calories, protein, or micronutrients for his physical and mental growth. Detecting malnourished children is important as their health and growth seriously affected by it and they become vulnerable to various diseases. Malnourished child can be detected based on its three indicators, stunting, underweight and wasting. The aim of the study is to predict the malnourished child in Bangladesh using the determinants of malnutrition. Both logistic regression and machine learning classification trees were applied to predict the malnourished child. We have used Bangladesh Demographic and Health Survey (BDHS) 2011 data for training purpose and test those two models on BDHS 2014 data to see which model perform better while predicting in future timeline. In terms of stunting, logistic regression detects 67.3% correctly while classification tree detects 67.9% correctly. For underweight, logistic regression detects 70.1% and classification tree detects 69.7% correctly. For wasting, true positive detection rate in logistic regression has found 0%, while classification trees identify 25.1%. Result concludes that based on the classification

trees, past information on several determinants of malnutrition can easily predict malnourished child in future generation.

*Keywords:* Child malnutrition, logistic regression, classification trees, predictive capability.

---

[P-172-CP-33]

### **Comparing rainfall prediction of north-eastern side of Bangladesh: SARIMA, ANN and simple exponential smoothing approach**

*Shabab Noor Islam\**, *Nahid Sultana*, *Aniqua Anjum*, and *Hridoy Ahmed*, Shahjalal University of Science Technology, Sylhet, Bangladesh

#### **Abstract**

Forecasting rainfall is one of the most significant and challenging tasks in the modern world. It is important to precisely decide to predict the rainfall for compelling utilization of water resources, crop profitability and pre-planning of water structures. This study aims to utilize Simple Exponential Smoothing and Seasonal Autoregressive Integrated Moving Averages (SARIMA) modeling technique together with Artificial Neural Networks (ANNs) methodology in order to forecast Historical Rainfall data of Bangladesh provided by Bangladesh Meteorological Department: only choosing the north-eastern side of Bangladesh's rainfall information. To forecast rainfall, Simple Exponential Smoothing method was used as well SARIMA (1,0,0)(1,1,0) was found to be the most parsimonious model to predict future precipitation. ANN was also applied which is a good alternative to the Box-Jenkins approach, especially if data shows non-linearity. For the ANN method, a multi-layer perception (MLP) with a supervised learning technique called back propagation was applied. Prediction performances of the models were reevaluated using two types of error measurement; mean absolute error (MAE) and root mean squared error (RMSE). The model with the smallest value of MAE and RMSE was considered best in predicting rainfall. Results indicated that the proposed ANN method predicts the next month's rainfall value with a low RMSE (116.63) than SARIMA (213.07) and Simple Exponential Smoothing (288.78). The MAE for ANN, SARIMA, Simple Exponential Smoothing were 79.15, 138.46, 202.74 respectively. Comparing RMSE, MAE, it can be concluded that ANN relatively performs better than other models in this study. These models are compared with each other and comparison has been conducted through studying their efficiency in modeling and forecasting the monthly data. It is expected that this long-term forecast will help policymakers inefficient scheduling of flood prediction, rainwater harvesting, and crop management.

*Keywords:* Artificial neural network, exponential smoothing, forecast.

[P-173-CP-34]

---

**Identifying the associated factors of infant and under-five child mortality in Bangladesh: A semi-parametric approach***Md. Mahabubur Rahman\** and *Tasnim Ara*, University of Dhaka, Bangladesh**Abstract**

In the past few decades, the world made remarkable progress in child survival and millions of children have better survival chances now. Data from the 2014 BDHS shows that in Bangladesh infant and under-five mortality rate were 38 and 46 deaths per 1,000 live births respectively. Mortality of children is one of the main indicators of a country's socio-economic development and quality of life so it should be at a minimal level. The main objective of this study is to identify the factors associated with infant and under-five child mortality using BDHS-2014 dataset. Two time (age at death or age at the end of the study) variables are considered separately for infant and under-five children. Cox PH model is used to calculate the hazard rate that evaluates the effect of the socio-economic and demographic factors on infant and under-five child mortality. Finally, model assumptions are tested with the goodness of fit. Division, number of household members, NGO membership, level of educational attainment of mother, birth order, total number of children ever born, number of living children, births in last five years, husband's occupation, age difference of parents are found to have a significant effect on hazard rate for both infant and under-five children. For both infant and under-five children, only in Khulna division hazard rate (1.57, 1.74) is higher compared to Dhaka division. Among the NGO members, hazard rate is higher compared to non-NGO members. Increasing mother's educational level lower the risk of mortality. Hazard rate is lower for the second child where three and above birth order produces a high risk of mortality. If the father is employed hazard rate decreases by 77% and 78% respectively for infant and under-five children compared to unemployed father. Findings indicate that in order to reduce infant and under-five child mortality, policy should be made to increase mother's educational level and eradicate the unemployment problem.

*Keywords:* Bangladesh Demographic and Health survey (BDHS), Non-governmental organization (NGO), Cox Proportional Hazard (Cox PH).

[P-174-CP-35]

---

**Identifying and predicting factors associated with contraceptive use among women in Bangladesh: A machine learning approach***Tasnim Ara\**, *Md. Hasinur Rahaman Khan*, and *Elizabeth Bianca Sarker*, University of Dhaka, Bangladesh

**Abstract**

Contraceptive use is viewed as a safe and affordable way to confined rapid population growth, reduce maternal, and infant mortality. High contraceptive rate is always expected for controlling births in countries with high population growth like Bangladesh. The aim of this study is to identify the factors associated with contraception use along with the prediction of factors under both usage of traditional and modern methods among the women of child-bearing age for combined data of Bangladesh Demographic and Health Survey (BDHS) for years 2004, 2007, 2011 and 2014. Apart from initial exploratory analysis, a mixed effect logistic regression has been implemented to combined BDHS data for identifying important factors. Traditional machine learning techniques including random forest, bagging, boosting have been used for predicting the factors of contraceptive use. Analysis of mixed random effect model reveals that there is an increasing pattern in modern method usage from 2004 to 2014 but contraception usage slightly decreased in 2007 and increase again in 2011 and 2014. A number of factors including division, place of residence, religion, sex of household head, number of household members, woman's age, occupation, BMI, current breastfeeding status, husband's age difference with wife, husband's and wife's combined educational level, living status with wife, recent sexual activity, desire for children, current amenorrheic and abstaining status, number of living children, number of child born in last five years, total children ever died, frequency of reading newspaper or magazine and frequency of watching television were significantly associated with contraception use in Bangladesh. Analysis of machine learning techniques also support these findings along with clear pathway of predicting the chance of being contraceptive user.

*Keywords:* Bangladesh Demographic and Health Survey (BDHS), Divisions, Generalized estimating equation(GEE), Mixed effect, Random Forest, Boosting.

---

[P-179-CP-36]

**ECG beat analysis using discrete wavelet coefficient**

*Azfar Adib\**, Concordia University, Canada

**Abstract**

This work aimed to develop a new approach to identify different types of electrocardiogram (ECG) beats using discrete wavelet transform (DWT) coefficients. Five types of cardiac phenomena were considered and for each of those some particular records from MIT-BIH arrhythmia database were selected. For these records DWT coefficients up to level 4 were calculated in the Matlab environment using different types of mother wavelets. A comparison was made between the performances of different types of mother wavelets to select the mother wavelet providing the best result.

*Keywords:*

[P-181-CP-37]

**A satisfactory threshold of body mass index to identify elderly nutritional risk: A cross-sectional study**

*Zaki Farhana\**, *Meshbahur Rahman*, and *Mohammed Taj Uddin*, Shahjalal University of Science and Technology, Sylhet, Bangladesh

**Abstract**

Malnutrition is one of the key prevalent concerns for frail health condition among elderly not only in Bangladesh but also all over the world. It is noticeable that diseases, impaired physical and mental condition are common for malnourished elderly. The main objective of this study is to assess the nutrition status and defines cut-off value of Body Mass Index (BMI) for indicating Underweight condition among the rural elderly. This cross-sectional study covers the rural site of Sunamgonj district which is located at the north-eastern part of Bangladesh. A total 400 elderly who are aged 50 or more considered in this study by simple random sampling. The nutritional profile is assessed with Short Form Mini Nutritional Assessment (MNA-SF) and Body Mass Index (BMI), according to WHO criteria. Receiver Operating Characteristic (ROC) Curve is used to define the predictive cut-off value of BMI for underweight condition in relation to MNA-SF. It is found that only 0.5% elderly has adequate nutritional status, 40.0% are at risk for malnutrition and 60.0% are considered malnourished. ROC curve indicated that a BMI threshold as high as 23.04 kg/m<sup>2</sup> is needed to identify 89% of these ?at risk for malnutrition? elderly according to the MNA-SF (area under the curve [AUC]: 0.828, P<0.001). A BMI cut-off value of 22.26 kg/m<sup>2</sup> has a sensitivity of 67.9% and a specificity of 81.2%. Early detection of nutrition disorder paves the way for an early nutritional approach and creates scope to prevent undesirable outcomes with respect to the health of those individuals. Higher BMI values, up to 22.26 kg/m<sup>2</sup>, should be considered as threshold for better detecting elderly malnutrition. The current BMI cut-off value (< 18.5kg/m<sup>2</sup>) may not be appropriate for the North-eastern rural elderly of Bangladesh.

*Keywords:* Body mass index (BMI), mini nutritional assessment (MNA-SF), receiver operating characteristic (ROC) curve, elderly, Bangladesh

[P-182-CP-38]

**Spatial distribution of infant mortality rate at district level in Bangladesh using small area estimation method**

*Md. Ahadur Rahman*, *Mossamet Kamrun Nesa*, *Md. Rafil Tazir Shah\**, *Anisa Ahmed Chowdhury*, and *Afroza Jannat Suchana*, Shahjalal University of Science and Technology, Sylhet, Bangladesh

**Abstract**

Infant Mortality Rate (IMR) is considered as an important indicator of population health and directly related to mothers and children health and population growth of a nation. National level estimates of often hide the real scenario across the country. In order to construct a map of IMR, accurate estimates of the IMR are required at very small spatial scales, typically the administrative units of a country. Although comprehensive data on maternal health indicators are collected in national surveys, the small-scale estimates cannot be calculated using the standard estimation methods in national surveys since such methods are designed to produce national level estimates and assume large samples. Area level small area estimation methods (e.g., Fay-Herriot method) are widely used to calculate such micro-level small-scale estimates. In this research, Fay-Herriot model has been used to calculate district wise estimates of IMR in Bangladesh with mean squared error in two different survey timeline and compare them to find district wise improvement over time. The Bangladesh Demographic Health Survey (BDHS) 2011 and 2014 data along with Population Census 2011 are utilized for this study. It has been found that in 2011, Joypurhat, Sherpur, Jhalokati and Sirajganj were most vulnerable in terms of IMR. But over time they improved well and in 2014 Sunamganj, Habiganj, Gazipur and Cox'sbazar took their place and identified as most vulnerable.

*Keywords:* IMR, small area estimation, spatial distribution

---

[P-185-CP-39]

**Finding the spatial auto correlation and multiple linear regression analysis of literacy rate, 2011 of Bangladesh**

*Md. Parvez Hasan Yousuf\**, *Jinat Ara Nasreen*, and *Anikah Tasnim Raka*, Bangladesh University of Professionals (BUP), Bangladesh

**Abstract**

The spatial auto correlation, spatial interpolation, spatial regression etc. are the common methods used to understand the presence and absence of the spatial relationship among the variables with distance. The research objectives of this paper are to find the spatial auto correlation of the literacy rate, 2011 of Bangladesh. And to explain and predict the relationship between this literacy rate and other independent variables through multiple linear regression analysis. It is a quantitative study. Data were collected from secondary sources mostly from the census and statistical yearbooks published by Bangladesh Bureau of Statistics (BBS). Then analysis was done using Microsoft Excel and SPSS software. Maps were produced using ArcGIS software (version-10.3). Moran's  $I$  and Geary's  $C$ , two widely used spatial statistic for detecting global and local spatial patterns were calculated to find out spatial autocorrelation. In order to define the neighbors of districts, contiguity spatial weights matrices were created and used in excel by observing the districts' boundary of Bangladesh map. The results show that there is spatial autocorrelation of the literacy rate of the 64 districts of Bangladesh. The value of Moran's  $I$  is positive and that of Geary's  $C$  is significantly lower than 1 which demonstrates spatial

influence and positive spatial autocorrelation respectively. The maps produced visually represents the literacy rate and its spatial dependence. The multiple regression analysis results also reveal that the independent variables namely Spatial Dependence (X<sub>Lag</sub>), Population of farmers (%), Population of 2011 and Projected Population of 2016 are the significant predictors of literacy rate, 2011. Most of the p-values of independent variables are less than 0.05 so, these predictors are significant. Thus, various tools and methods used for spatial analysis seems to be fruitful enough to represent and predict the demographic data.

*Keywords:* Spatial auto correlation, contiguity, literacy rate, Moran's I, Geary's C, multiple regression analysis.

---

[P-186-CP-40]

**Determinants of stunting, wasting and underweight in under five children: An analysis of BDHS data from 2004, 2007, 2011 and 2014.**

*Argha Dhar\**, University of Dhaka, Bangladesh

**Abstract**

Despite improvements in child malnutrition, poor nutritional status remain as one of the most concerned thing in public health sector of Bangladesh. The children are remain as the most vulnerable cases for poor nutritional status. This study investigated the impact of some socioeconomic, demographic and health and community factors on stunting, wasting and underweight in Bangladeshi children aged less than 5 years. This study has used Bangladesh Demographic and Health Survey data of 2004, 2007, 2011 and 2014. Bivariate analysis was performed to find out the differentials in prevalence and binary logistic regression was performed to assess the association of stunting, wasting underweight with potential risk factors. This analysis also revealed the overall stunting, wasting and underweight trend. The main contributing factors revealed from our study for stunting, wasting and underweight are found to be mother's education, mother's body mass index, family wealth index, husband's education, husband's occupation, division, mother's age at first child. Trends for stunting, wasting and underweight are found to be decreasing across the year.

*Keywords:*



---

[P-219-CP-41]

**Relation between demographic data and photoplethysmography (PPG) signal features in estimating human health metrics**

*Moajjem Hossain Chowdhury, Md. Nazmul Islam Shuzan\**, North South University, Bangladesh, *Muhammad E. H. Chowdhury*, Qatar University, Qatar, *Zaid B Mahbub*, and *M. Monir Uddin*, North South University, Bangladesh

**Abstract**

Photoplethysmography (PPG) technique utilizes infrared light to measure the blood flow to the skin. The PPG signals are regularly used in clinical practice to measure heart rate, oxygen saturation etc. however can be used for other clinical applications like non-invasive blood pressure, total hemoglobin concentration measurement. In this work, a short PPG data set recorded in China was used to study the correlation between the signal features of a PPG signal with the demographic information of a healthy subject. Several unique features from PPG signals were extracted using signal processing techniques in MATLAB. These features were used to correlate between the Signalling information with the demographic data. It was observed that the age, weight, height and body mass index (BMI) of a person affect the key features (i.e., Systolic Peak, Diastolic Peak, Noticeboard etc.) of a PPG signal. Therefore, demographics cannot be ignored when estimating different metrics of health using PPG signal. Demographic features should be incorporated along with signal features in future studies of PPG signal to predict human health metrics.

*Keywords:* Telephotography (PPG), demographic information, body mass index (bmi), signal features, correlation.

---

[P-224-CP-42]

**Spousal violence among married women in the slum area: A study in Khulna city of Bangladesh**

*Maliha Raihan Pranti\** and *U. K. Majumder*, Khulna University, Bangladesh

**Abstract**

Despite having many socio-economic advances throughout the last couple of decades, violence against women (VAW) is still a glaring social problem in Bangladesh as well as all over the world. In the study area (slum area of Khulna City Corporation, Bangladesh), the surprising acceptance of VAW in day to day life even from woman's perspective proves the sad reality of having a patriarchal society where women are brought up as submissive or lesser part of the society. In addition, the sheer ignorance of

being victim of domestic violence or spousal violence especially in terms of sexual and psychological abuse has also been a frustrating truth to accept. By performing a cross-sectional study where the data were collected during July to September 2019 from 400 married women in the study area using face to face interview technique, the aims of the study are to find out the magnitude of spousal violence in the slum area among the married women where the socio-economic status is considered the lowest in the urban life and detect any factors which are responsible. Here, the sample was selected following the cluster sampling method, and the further statistical analysis was done using chi-square analysis, point-interrelationship and logistic regression analysis. The findings of the study are as follows: the overall prevalence rate of spousal violence is 69% where the percentage of moderately, seriously and dangerously abused women are 60%, 7% and 2% respectively. Woman's poorer economic background, working status, and husband's controlling behavior emerged as strong predictors for domestic violence in the slum area of Khulna City Corporation. Elimination of structural inequalities inherent in the society, caste, and the traditional male hierarchy in society could prevent domestic violence.

*Keywords:* Spousal violence, married women, slum, Khulna

---

[P-227-CP-43]

### Rejection sampling scheme for simulating from multivariate normal density

*Syeda Fateha Akter\**, University of Dhaka, Bangladesh

#### Abstract

The multivariate normal (MVN) distribution is often considered to be the underlying distribution of many observed samples for modelling purposes, and hence simulation from this distribution is required to verify the fitted model. The decomposition based approach is currently being used to simulate sample from MVN distribution whose building block is Cholesky or Eigen decomposition. Although the decomposition approach is routinely used to generate MVN in almost all statistical package (R, SAS, ST AT A), but this approach may have a numerical issue (Ripley, 1987). Unfortunately, there is no other alternative of this approach to generate MVN variates. Motivated by this problem, we develop an alternative method to generate MVN variates whose building block is rejection sampling. We have empirically investigated the performance (normality and randomness) of our proposed method and found that it can be used as an alternative method for generating MVN variates.

*Keywords:* Efficient decomposition methods, decomposition based MVN generation, Cholesky and Eigen decomposition, gibbs sampling, ratio-of-uniforms, rejection sampling.

---

[P-228-CP-44]

**Bayesian accelerated failure time survival regression model: An application to under-five mortality data in Bangladesh**

*Md. Azharul Islam Rayhan\** and *Wasimul Bari*, University of Dhaka, Bangladesh

**Abstract**

Although point estimate provides a single value to guess the unknown parameter of interest based on the sample information, it may lead to a misleading inference. To overcome this problem, one may use interval estimate using point estimate and margin of error. Frequentist approach of inference provides confidence interval, which is considered as a random variable and either includes the parameter with probability 1 or not with probability 0. Alternatively, there exists Bayesian approach of inference, which considers credible interval as fixed and parameter as random variable allowing a direct probabilistic interpretation of credible intervals to state directly the probability that a parameter say, lies between upper and lower limits of the interval, whereas confidence interval can not provide this type of interpretation. Besides, Bayesian inference incorporates prior information along with information contained in data, to provide marginal posterior distribution for each parameter which covers the whole range of inferential solutions rather than a point estimate and a confidence interval as in frequentist approach. In this thesis, both frequentist and Bayesian inference approaches have been studied for accelerated failure time (AFT) survival regression models. An extensive simulation study has been conducted to examine the performance of frequentist and Bayesian approaches under AFT model. Finally, under-5 mortality data extracted from Bangladesh Demographic and Health Survey (BDHS), 2014 have been analyzed using both frequentist and Bayesian approaches to find out potential determinants of under-5 mortality. A number of recommendations has been suggested based on results obtained from AFT model using Bayesian approach to reduce the under-5 mortality in Bangladesh to a great extent.

*Keywords:* Bayesian analysis, survival regression, accelerated failure time, under-5 mortality

---

[P-229-CP-45]

**Nested frailty model for two-level clustered survival data: Analyzing infant mortality in Bangladesh**

*Bikash Pal\** and *Wasimul Bari*, University of Dhaka, Bangladesh

**Abstract**

Cox proportional hazard (PH) model has been extensively used in analyzing survival data to examine the effect of a set of covariates on time to occur an event of interest assuming that observed survival times are independent. In clustered survival data, the observations from the same cluster share certain unobserved characteristics and hence the observations within a cluster are likely to be correlated. It is also assumed that heterogeneity exists among the observations of different clusters. Frailty models consider random effect terms to model this correlation and/or heterogeneity in survival analysis. To obtain consistent and efficient estimates of parameters of interest, it is required to take this correlation into account. For this purpose, the Cox PH shared frailty model is widely used in practice, where baseline hazard function is left unspecified and frailty term acts multiplicatively on the hazard function. One parameter gamma distribution is commonly used as the frailty distribution. In practice, it may happen that data may arise from a hierarchical structure i.e., a cluster is nested within another cluster. In this case, nested frailty models are appropriate to analyze survival data to obtain optimal estimates of parameters of interest. In this study, an extensive simulation study has been conducted to examine the performance of Cox PH, shared, and nested frailty models. Misspecification of models is also examined through simulation study. To identify significant determinants of infant mortality in Bangladesh, survival data were extracted from Bangladesh Demographic and Health Survey (BDHS), 2014, which involves two-stage stratified cluster sampling. Because of the presence of two-level clustering in data, nested frailty model is recommended for the analysis purpose. Recommendations have been suggested based on the results obtained from appropriate survival model to reduce the infant mortality to a great extent in Bangladesh.

*Keywords:* Cox PH model, clustered survival data, heterogeneity, random effect terms, shared and nested frailty model.

---

[P-230-CP-46]

**Zero inflated negative binomial mixed model for analyzing antenatal care visits in rural Bangladesh: A comparative study**

*Mahfuza Haque Mahi\** and *Wasimul Bari*, University of Dhaka, Bangladesh

**Abstract**

Though the Poisson regression model is a common choice to analyze the count response to examine the effects of different covariates, in the presence of overdispersion, this model may provide misleading inference. To overcome this problem, a Negative Binomial (NB) model is commonly used to explain variability in the data. In practice, it may happen that count response may arise through two-stage sampling. Hence, count responses within a cluster are likely to be correlated. In that case, the NB mixed regression model is more appropriate for analyzing the overdispersed count data which is correlated. Count data often contain excessive number of zero outcomes than expected in NB distribution, and then, the Zero Inflated Negative Binomial (ZINB) regression model has been fitted for zero inflated

data where overdispersion also exists. If the observations in the count data are overdispersed as well as zero inflated and simultaneously correlated due to the sampling design, then ZINB mixed regression model is appropriate for analyzing the data. In this talk, NB and NB mixed regression models, ZINB and ZINB mixed regression models are applied to number of antenatal care (ANC) visits of women during pregnancy in rural Bangladesh, where data are extracted from Bangladesh Demographic and Health Survey (BDHS), 2014 and a comparison is made among the results obtained from these regression models. Antenatal care service is a significant intervention to improve maternal and child health and to prevent maternal and infant deaths. However, this service is poorly developed in Bangladesh, particularly in rural areas. This study mainly focuses to identify potential factors associated with ANC visits of women in rural Bangladesh.

*Keywords:* Overdispersion, Zero inflation, Negative Binomial, Mixed model, Antenatal care.

---

[P-231-CP-47]

### **Poisson regression with observation-level random effects: A comparative study**

*Maria Mukta\** and *Zakir Hossain*, University of Dhaka, Bangladesh

#### **Abstract**

Poisson regression is commonly used as the base or standard model for analyzing count data with the restrictive property that the mean and variance of the responses are equal. However, overdispersion, i.e. extra-Poisson variation is very common in many practical fields of health sciences. In such cases, Poisson regression produces misleading inferences and hence give incorrect interpretations of the results. Poisson regression with observation-level random effects (OLRE), where each data point receives a unique level of random effect that models the overdispersion present in the data, is found to be the best choice (smallest AIC) than Poisson, quasi-Poisson and negative binomial for modelling and analysis of overdispersed antenatal care count data of women in Bangladesh.

*Keywords:* Overdispersion, Pearson chi-square, generalized linear models (glms), generalized linear mixed models (glmms).

---

[P-248-CP-48]

### **Consumer buying behavior and preferences of online shopping: A study on Dhaka city, Bangladesh**

*KM Tanvir\** and *Farhia Azrin Oishe*, University of Dhaka, Bangladesh

**Abstract**

Business world is changing and internet has played a significant role. Many people are thinking of starting online business. Before starting they should understand the consumer behavior towards online shopping. The purpose of the study is to know the answer of “Should I start online business and if I do what should be my strategies?” Study on online consumer buying behavior is important because it helps to understand consumers demand. We need to identify the factors that motivate customers to decide whether to do online shopping or not. As part of my primary data collection, I had conducted a survey and in total 500 people responded and 40 of them were taken for the in-depth interviews to ensure the collection of the data was reliable. The survey reveals that consumers shop online to save time and energy and sometimes discount and offers motivate them. As well as majority of the people don't prefer online shopping because they can't trust online merchant and they are aware of the quality of the product. We also have found that the people from the age group 30 and above are less likely to do online shopping. Those who want to start online business, they need to gain trust from the consumers first. They need to sell the products in a reasonable price and give discount and offers. They need to have positive reviews as many consumers are motivated to buy online for the positive reviews.

## Author Index

- Aboy JBI, 61  
Adib A, 186  
Afiaz A, 162  
Afrose M, 115  
Ahammed T, 80, 138  
Ahmed A, 133  
Ahmed FF, 157  
Ahmed SS, 69  
Akond Z, 158  
Akter E, 126  
Akter MT, 170  
Akter R, 181  
Akter SF, 191  
Al-Mutairi DK, 60  
Alam MJ, 153  
Alam TF, 112  
Amaefula CG, 50  
Ambler G, 94  
Amin R, 143  
Anjum A, 171  
Ara T, 185  
Arefin AB, 177  
Arman M, 90  
Auwul MR, 76  
Ayesha U, 166
- Bandyopadhyay S, 48  
Banerjee, PK, 103  
Bashar MI, 116  
Basit A, 162  
Baten MA, 58  
Begum MR, 70  
Bhandari S, 62  
Bharati P, 54  
Bharati S, 67  
Biswas CS, 71  
Biswas MR, 164  
Bose K, 53  
Bosu P, 175
- Chakraborty A, 106  
Chakraborty S, 68, 96, 120  
Chowdhury MA, 169
- Chowdhury RI, 72  
Chowdhury S, 139, 168  
Chowdhury SH, 127
- Das A, 145  
Das K, 46  
Das KR, 141  
Das RN, 63  
Das S, 75  
Das U, 83  
De UK, 69  
Dewan MN, 122  
Dhar A, 189  
Doan NV, 104  
Dutta A, 73
- Farhana Z, 131, 187  
Feist SE, 88  
Ferdous Z, 179  
Ferdushi KF, 82  
Firoj MG, 156
- Ghosh M, 63  
Ghosh S, 135  
Golingay SD, 56  
Gulshan J, 91
- Hakim S, 134  
Halder HR, 114  
Hasan M, 110  
Hasan MAM, 99  
Hasan MM, 150  
Hasan MN, 93, 156, 173, 182  
Hema PS, 136  
Hemo SA, 173  
Hisham AM, 177  
Hoque ME, 80  
Hossain M, 125  
Hossain MF, 68  
Hossain MK, 130  
Hossain MS, 87, 133  
Hossain S, 103, 167, 172  
Huda S, 45
- Huq MM, 66
- Islam MA, 140  
Islam H, 128  
Islam J, 179  
Islam MA, 142  
Islam MK, 115, 146  
Islam MM, 111, 131, 136, 146, 153  
Islam S, 88  
Islam SN, 184  
Islam SR, 141
- Jahan N, 61  
Joarder AH, 102
- Kamruzzaman M, 160  
Karim E, 64  
Karim MR, 89  
Kawsar LA, 84  
Khan MZS, 74  
Khan NM, 129  
Khan S, 75  
Khurshid S, 169  
Kibria KM, 155  
Kumar D, 47
- Lee SY, 101  
Liang F, 43  
Limon MA, 165  
Lipi N, 118
- Mahi MH, 193  
Mahmoud M, 49  
Mahmud S, 117  
Mahsin M, 55  
Maiti SI, 60  
Maji SG, 119  
Majumder UK, 97  
Mannan HR, 101  
Maria M, 194  
Martinez W, 44  
Matin KA, 97

Matin MA, 82  
 Mazumder SR, 167  
 Methun MIH, 124  
 Mimi MA, 117  
 Mitra A, 57  
 Mollah NH, 91  
 Mondol MH, 109  
 Monir MM, 93  
 Mosharaf MP, 152  
 Mosleh-Uddin M, 105  
 Mukta MMR, 108  
 Musa Y, 55  
 Muyeed A, 149  
  
 Naima US, 149  
 Nandy K, 100  
 Nazra KU, 171  
 Nesa MK, 87  
 Nishat MNS, 123  
 Nusrat N, 113  
  
 Owusu DA, 50  
  
 Pal B, 192  
 Pal M, 52  
 Park T, 100  
 Paul S, 176  
 Pavlou M, 95

Pranti MR, 190  
  
 Rahaman MM, 154  
 Rahman M, 56, 107  
 Rahman MAI, 183  
 Rahman MAN, 187  
 Rahman MM, 86, 92, 143,  
     147, 185  
 Rahman MT, 148  
 Rahman UH, 66  
 Rana MM, 112  
 Rashid M, 178  
 Rayhan MAI, 192  
 Roitman LA, 53  
 Roy DC, 78  
 Rudra S, 122  
 Rumana AS, 159  
  
 Sanim MAI, 163  
 Sen R, 95  
 Serra IJ, 57  
 Shahjaman M, 98  
 Shanto SI, 124  
 Sharma P, 64  
 Sharmin S, 158  
 Shihab S, 175  
 Shuzan MNI, 190

Siddiquee MH, 144  
 Siddiquee R, 110  
 Soni RK, 59  
 Sorkar MNU, 164  
 Sultana N, 120, 138  
 Sultana P, 90  
 Sultana S, 180  
 Sultana ZZ, 128  
  
 Tabassum KF, 121  
 Tamanna RJ, 137  
 Tanvir KM, 194  
 Taskin F, 126  
 Tasmia SA, 151  
 Trang DTQ, 104  
 Tusty MA, 132  
  
 Uddin B, 174  
 Uddin N, 74  
 Ullah MO, 65  
 Ullah S, 84, 85  
 Usman A, 108  
  
 Yakubu Y, 51  
 Yiman JA, 72  
 Yousuf MPH, 188  
  
 Zaman T, 182