



የኢትዮጵያ ስታቲስቲክስ ባለሙያዎች ማኅበር
Ethiopian Statistical Association (ESA)

The 32nd Annual Conference of Ethiopian Statistical Association

“Statistical Methodology and Data Science for Population Mapping and Sustainable Development in Ethiopia”

Sponsors



United Nations
Economic Commission for Africa

14 – 15 June 2024

Addis Ababa, Ethiopia

Conference Program and Book of Abstracts

Venue: - Getfam Hotel, Addis Ababa

Program

ESA 2024

The 32nd Annual Conference of Ethiopian Statistical Association (ESA)

14 – 15 June 2024, Addis Ababa, Ethiopia

Venue: - Getfam Hotel, Addis Ababa

Sponsored by:

Ethiopian Statistical Service

African Center for Statistics-UNECA

Addis Ababa, Ethiopia

EXECUTIVE COMMITTEE MEMBERS

President	Zeytu Gashaw, PhD
Vice President	Denekew Bitew, PhD
Training Coordinator	Habte Tadesse, PhD
Research and Project Coordinator	Gezahagn Getahun
Members & Chapter Offices Affairs Coordinator	Eskedar Girma
Secretary-General	Haymanote Zeleke
Finance and Administrative Coordinator	Tadesse Kassahun, PhD

AUDITOR

Gezahegn Mekonnen, PhD

HEAD, SECRETARIAT

Samuel Hailu

EDITORIAL COMMITTEE

Editor-in-chief	Emmanuel GebreYohannes, PhD
Associate Editors	Mekonnen Tadesse, Assoc. Prof. Adem Kedir, PhD Bedilu Alamirie, PhD Tadesse Kassahun, PhD

ADVISORY BOARD MEMBERS

Addisu Fekadu	Ethiopia
Ayele Menbere	USA
Ayele Taye, PhD	Ethiopia
Ayenew Ejigou, Prof.	Ethiopia
Bute Gotu, PhD	Ethiopia
Dejen Tesfaw, PhD	Ethiopia
Demelash Megersa	Ethiopia
Emanuel Gabreyohannes, PhD	Ethiopia
Eshetu Wencheke, Prof.	Ethiopia
Fentaw Abegaz, PhD	The Netherlands

Genene Bizuneh	Ethiopia
Legesse Ayane	Ethiopia
Martha Sintayehu	Ethiopia
Mekonnen Tadesse, Assoc. Prof.	Ethiopia
Merga Belina, PhD	Ethiopia
Shibru Temesgen, PhD	Ethiopia
Sileshi Fanta, PhD	South Africa
Zerihun Tadesse	USA

LIST OF ABSTRACT REVIEWERS

- | | | |
|-------------------------------|----------------------------|---------------------------|
| 1. Dr. Abdisa Gurmessa | 10. Dr. Feysal Kemal | 19. Dr. Shiferaw Befikadu |
| 2. Dr. Akalu Banbata | 11. Dr. Gezahen Mekonnen | 20. Dr. Tadesse Kassahun |
| 3. Dr. Aragaw Eshete | 12. Dr. Girma Taye | 21. Dr. Tadelle Akaba |
| 4. Dr. Bedilu Alamirie | 13. Dr. Habte Tadesse | 22. Dr. Teshager Zerihun |
| 5. Dr. Cheru Astemegiorgis | 14. Dr. Haile Mekonnen | 23. Dr. Tilahun Ferede |
| 6. Dr. Denekew Bitew | 15. Dr. Jemal Ayalew | 24. Dr. Yonas Shuke |
| 7. Dr. Derebachew Asfaw | 16. Dr. Kassahun Tadele | 25. Dr. Zeytu Gashaw |
| 8. Dr. Dereje Danbe | 17. Dr. Markos Abiso | 26. Mr. Mekonnen Tadesse |
| 9. Dr. Emmanuel GebreYohannes | 18. Dr. Seyfemichael Amare | 27. Prof. Awoke Seyoum |

CONFERENCE ORGANIZING COMMITTEES

1. Scientific and Publication Committee

Tadesse Kassahun (PhD), Chair
 Haymanote Zeleke
 Samuel Hailu
 Zeytu Gashaw (PhD)

2. Logistics and Finance Committee

Gezahegn Getahun, Chair
 Eskedar Girma
 Gezahegn Mekonnen (PhD)
 Zeytu Gashaw (PhD)

Objectives of the Conference

The main objectives of the conference are to:

- promote statistics for addressing development sustainability challenges through presentations of scientific papers and panel discussions;
- discuss benefits of conducting regular population and housing census for sustainable development;
- discuss the integration of gender perspectives into the methodology of sample surveys and censuses;
- deliberate the cost of conflict and its consequences on health and social well-being of citizens;
- create opportunities for multi-stakeholder collaborations among statisticians, gender experts, data scientists, and demographers in the academia and other sectors;
- encourage early career professionals to involve in advanced statistical research for sustainable development;
- introduce the importance of machine learning and big data analytics for better decision making;
- initiate new project ideas together with different development partners;
- establish a forum of discussion on the challenges and opportunities of statistical advancement in Ethiopia and across Africa to the larger domain;
- foster dialogue among researchers, development partners, and policymakers.

Expected Outcomes

The 32nd annual conference of ESA has the following expected outputs.

- Promote the importance of statistics for sustainable development and better decision in order to make it an agenda so that the higher government officials and institutions pay attention.
- The integration of statistics and data science will be largely advertised.
- Ways to improve methods of data collection & analysis towards gender statistics in an easy to use format for policy makers and planners will be addressed.
- New platforms for research and development activities will be established among participants of the conference.
- Collaboration among ESA and local as well as international development partners will be strengthened.
- Share best practices and knowledge among the statistics community.
- The regularity of population and housing for economic growth and sustainable development along with challenges will be pinpointed.
- Establish sustainable partnership with development partners, data producers, development actors and MDAs

General Information

Conference Location

Getfam Hotel is located in Haya Hulet, Haile Geberesilassie Street, Addis Ababa, Ethiopia. Its location on Google map: 9°00'53.1"N 38°47'01.4"E 9.014737, 38.783731

Registration

The registration desk at the Hotel will be open from 08:00 AM on Friday 14th June 2024.

Internet Access

Free wireless Internet access will be provided to ESA 2024 conference participants during the days of the Conference at the Hotel. The information needed to access the wireless network will be provided at the registration desk of the conference venue.

Liability

The Conference fees DO NOT include provisions for the insurance of participants against personal injuries, sickness, and theft or property damage. This also applies to any event held during the Conference period. Neither the Conference Organizing Committee nor its sponsors nor committee members assume any responsibility for loss, injury or damage to persons or belongings, however, caused.

CONFERENCE PROGRAM

The 32nd Annual Conference of ESA, Ethiopia

Addis Ababa, Ethiopia,

14 –15 June 2024

Venue: Getfam Hotel, Addis Ababa

FRIDAY, 14 th June 2024	
08:00-9:00	Registration
SESSION 1: OPENING SESSION Main Hall	
9:00-09:45	<p>Master of Ceremony: Dr. Tadesse Kassahun, Ethiopian Statistical Association</p> <p>Welcoming Speech: Dr. Zeytu Gashaw, President of ESA</p> <p>Official Opening: Guest of honor</p> <p>Keynote Speech 1: H.E. Dr. Beker Shale, Director General, Ethiopian Statistical Service</p> <p>Keynote Speech 2: Mr. Oliver Chinganya, Director, African Center Statistics for Statistics - UNECA</p> <p>Keynote Speech 3: H.E Mr. Kora Tushune, State Minister, Ministry of Education</p>
	<p>Moderators:</p> <ol style="list-style-type: none">1) Mr. Molla Hunegnaw, African Center for Statistics UNECA2) Mrs. Yilfgn Abegaz, UnWOMEN.
09:45 – 10:30	<p>Speaker: Prof. Eshetu Gurm, Addis Ababa University, Ethiopia</p> <p>Topic: Uses of Statistical Methods and Procedures in Undertaking National Population Censuses: Opportunities and Challenges</p> <p>Speaker: Mrs. Desta Worash, Ethiopian Statistical Service</p> <p>Topic: Revolutionizing the Ethiopian Statistical Service: Harnessing Big Data as an Alternative Source</p> <p>Speaker: Mrs. Asnakech Tamene, Ethiopian Statistical Service</p> <p>Topic: Improved Socio-Economic Inclusive Gender Statistics for Sustainable National Development</p>

10:30-10:45	Networking and Refreshment Break
SESSION 2: PRESENTATIONS ON STATISTICS CURRICULA	
Main Hall	
Chair: Dr. Girma Taye, Addis Ababa University	
Rapporteur: Dr. Shibru Temesgen, Addis Ababa University	
10:45- 11:45	<p><i>15 minutes for each presenter</i></p> <p>Speaker: Dr. Bedilu Alamirie, Addis Ababa University, Ethiopia</p> <p>Topic: Overview of Statistics Education in Ethiopia: Opportunities and Challenge of BSc program in Statistics</p> <p>Speaker: Prof. Essey Kebde, Bahir Dar University, Ethiopia</p> <p>Topic: Statistics Postgraduate Education in Ethiopia: Challenges and Opportunities.</p> <p>Speaker: Dr. Getachew Tekle, Wachemo University, Ethiopia</p> <p>Topic: Development and challenges of statistics education in Ethiopian Universities</p> <p>Speaker: Prof. Yehenew G. Kifle, University of Maryland, Baltimore County, USA</p> <p>Topic: Graduate and Undergraduate Statistics Training at University of Maryland Baltimore County (UMBC), USA: Sharing Our Experiences</p>
11:45-13:00	Discussions
13:00-14:00	Lunch Break
SESSION 3: PARALLEL 1	
Room 1	
Chair: Prof. Awoke Seyoum, Bahir Dar University	
Rapporteur: Dr. Teshager Zerihun, Debre Tabor University	
14:00-14:45	<p><i>15 minutes for each presenter</i></p> <p>Presenter 1: Dr. Seyifemickael Amare</p> <p>Topic: <i>Estimating the Prevalence of Childhood Polio Vaccination in Ethiopian Zones Using a Hierarchical Bayes Small Area Estimation</i></p> <p>Presenter 2: Mr. Anteneh Bezabih</p> <p>Topic: <i>Survival Analysis of COVID-19 Patients Hospitalized in Hawassa University Comprehensive Specialized Hospital, Ethiopia: A Cohort Study</i></p> <p>Presenter 3: Mr. Yikeber Abebaw</p> <p>Topic: <i>A Joint Model of Longitudinal and Competing Risks Survival Data with Application to Chronic Heart Failure Patient's Data.</i></p>
SESSION 3: PARALLEL 2	
Room 2	
Chair: Dr. Merga Belina, Addis Ababa University	
Rapporteur: Dr. Jemal Ayalew,	

14:00-14:45	<p><i>15 minutes for each presenter</i></p> <p>Presenter 1: Dr. Getachew Tekle Topic: A New Alpha Power Type-1 Family of Distributions and Modelling the Overdispersed Count Outcome</p> <p>Presenter 2: Mr. Wudneh Ketema Topic: <i>Bayesian Propensity Score Approaches for Balancing Covariates Associated with Low Birth Weight in North Shoa Zone, Ethiopia</i></p> <p>Presenter 3: Mr. Ashagre Sharew Topic: <i>Identification of Confounders for the Association between Binary Exposures with Ordinal Outcome</i></p>
SESSION 3: PARALLEL 3 Room 3	
<p>Chair: Dr. Dejen Tesfaw, Addis Ababa University</p> <p>Rapporteur: Dr. Yonas Shuke, Hawassa University</p>	
14:00-14:45	<p><i>15 minutes for each presenter</i></p> <p>Presenter 1: Mr. Kedir Hussein Topic: <i>Artificial Intelligence-Driven Ensemble Model To Predict Mortality Due To Covid-19 In East Africa</i></p> <p>Presenter 2: Mr. Melaku Tadege Topic: <i>Cardiac patients' surgery outcome and associated factors in Ethiopia: application of Machine learning</i></p> <p>Presenter 3: Mr. Getahun Mulugeta Topic: <i>Developing Clinical Prognostic Models to Predict Graft Survival after Renal Transplantation: Comparison of Statistical and Machine Learning Models</i></p>
14:45 – 15: 30	Discussions
15:30 - 15:45	Networking and Refreshment Break
SESSION 3: PARALLEL 1 – CONTINUES Room 1	
<p>Chair: Dr. Haile Mekonnen, Bahir Dar University</p> <p>Rapporteur: Dr. Akalu Banbata, Jimma University</p>	

15:45 – 16:45	<p><i>15 minutes for each presenter</i></p> <p>Presenter 1: Mr. Maru Zewdu Topic: <i>Determinants of Blood Sugar Level and Time to First Remission for Type I diabetic patients under treatment; A comparison of Separate and Joint models</i></p> <p>Presenter 2: Mr. Melkamu Ayana Topic: <i>Spatio-temporal Distributions and Determinants of Diarrhea among under-five Children in Ethiopia</i></p> <p>Presenter 3: Dr. Yonas Shuke Topic: <i>Multivariate Spatio-temporal Modelling of Malaria count of Genus P. falciparum and P. vivax in Case of Southern Ethiopia</i></p> <p>Presenter 4: Dr. Haile Kebede Topic: <i>Spatiotemporal Models with Confounding Effects: Application on Under-Five Mortality across sub-Saharan African Countries</i></p>
<p>SESSION 3: PARALLEL 2 – CONTINUES Room 2</p>	
<p>Chair: Dr. Aragaw Eshete, University of Gondar Rapporteur: Dr. Feysal Kemal, Adama Science and Technology University</p>	
15:45 – 16:45	<p><i>15 minutes for each presenter</i></p> <p>Presenter 1: Dr. Shiferaw Befkadu Topic: <i>Model-free screening for variables with treatment interaction</i></p> <p>Presenter 2: Dr. Bisrat Misganaw Topic: <i>Linear regression model with doubly truncated generalized new symmetric error distribution</i></p> <p>Presenter 3: Dr. Melkamu Molla Topic: <i>Flexible Bayesian Semiparametric Mixed-Effects Model for Complex Longitudinal Data: A Simulation and Application Study</i></p> <p>Presenter 4: Dr. Denekew Bitew Topic: <i>Bayesian Multivariate joint modeling of longitudinal and time-to-event data using HIV/AIDS patients</i></p>
<p>SESSION 3: PARALLEL 3 – CONTINUES Room 3</p>	
<p>Chair: Dr. Dereje Danbe, Hawassa University Rapporteur: Dr. Tilahun Ferede, Ethiopian Public Health Institute</p>	

15:45 – 16:45	<p><i>15 minutes for each presenter</i></p> <p>Presenter 1: Mr. Daba Ketema Topic: <i>Multivariate Time Series Modeling of the Relationship Between Major Economic Indicators in Ethiopia</i></p> <p>Presenter 2: Mr. Yidnekachew Mare Topic: Additive Modeling of Zonal Level Crop Production in Ethiopia</p> <p>Presenter 3: Dr. Kindie Fentahun Topic: <i>Spatial and Survey Regression Analysis of Zero-Dose Immunization Coverage among 12-23 month-old Children in Ethiopia</i></p>
16:45 – 17:30	Discussions
SATURDAY, 15th June 2024	
SESSION 4: INVITED TALKS Main Hall	
Chair: Dr. Emmanuel GebreYohannes, Ethiopian Civil Service University	
Rapporteur: Dr. Derbachew Asfaw, Addis Ababa University	
09:00 – 10:30	<p>Speaker: Mr. Gezahagn Getahun, Ethiopian Statistical Association, Ethiopia Topic: Overview of the Post Measles Vaccination Coverage Survey of Ethiopia 2022</p> <p>Speaker: Prof. Akim Adekpedjou, Missouri University of Science & Technology, USA Topic:</p> <p>Speaker: Dr. Abera Wouhib, National Institute of Mental Health, USA Topic: Methodological and Computational Rigor to Facilitate Mental Health Research and Achieve National Institute of Mental Health (NIMH) Research Priorities: A Statistical Perspective</p>
10:30 – 10:45	Networking and Refreshment Break
10:45 – 11: 15	<p>Speaker: Prof. Yehenew G. Kifle Topic: Comparison of local powers of some exact tests for a common normal mean with unequal variances</p>
11:15 – 12:30	Discussions
12:30 – 14:00	Lunch Break
SESSION 5: BUSINESS SESSION Main Hall	
Chair: Dr. Butte Gotu and Mr. Mekonnon Tadesse, Addis Ababa University, Ethiopia	
Rapporteur: Mr. Genanew Timerga, Ethiopian Civil Service University, Ethiopia	

14:00: 15:30	<ul style="list-style-type: none"> • ESA-EC annual work report • Internal audit report • External audit report • Annual plan for 2023/24 • ESA Office
15:30-15:45	Networking and Refreshment Break
15:45: 17:30	<ul style="list-style-type: none"> • Discussion on the reports • Election of EC • Closing Speech
End of conference	

BOOK OF ABSTRACTS

Overview of Statistics Education in Ethiopia: Opportunities and Challenge of BSc program in Statistics

Bedilu Alamirie Ejigu, Department of Statistics, Addis Ababa University

The introduction of modern education system in Ethiopia by Emperor Menelik II in 1908 marked the beginning of higher education in the country, which officially commenced in 1950. Recognizing the need for empirical data collection, summarization, and analysis, the Imperial Government of Ethiopia embarked on a five-year development plan in the late 1950s. However, at that time, there was a dearth of professional statisticians to fulfill this need. To address this shortage, the first Statistical Training Center in Ethiopia was established in 1961 at the Faculty of Science of Haile Selassie I University with financial support from UN-ECA. The center initially offered a one-year program leading to a certificate (Junior Statistician) and drew participants from eight African countries, namely Ethiopia, Egypt, Libya, Sudan, Malawi, Zambia, Basutoland, and Somalia. In response to the growing demand for medium and high-level statisticians, two additional programs were introduced with support from UN - a three-year diploma program and a four-year degree program. Since 1966, the length of the BSc program varied depending on the regimen between three and four years. This presentation aims to provide an overview of statistics education in Ethiopia with a focus on BSc Statistics program's academic trajectory. It will also highlight opportunities and challenges within the program while engaging participants in discussion points aimed at enhancing its quality. In conclusion, it is evident that statistics education has evolved significantly over time in Ethiopia; however, there are still areas that require improvement for better quality outcomes. After a short presentation on the topic, conference participants will discuss on the BSc curriculum in statistics opportunities and challenges to address these issues while shedding light on potential opportunities within this field.

Development and challenges of statistics education in Ethiopian Universities

Getachew Tekle, Wachemo University

Abstract

Statistics education is the practice of teaching and learning of statistics, along with the associated scholarly research. Insignificant and informal statistical applications are given from K-12 grades in Ethiopia. These concepts are indirectly included in high school mathematics subjects mostly at the end of the textbooks which mathematics teachers may not teach or neglect due to lack of specialization. In Ethiopian universities, at most basic or introduction to statistics courses are given for specific fields of studies by non-statisticians. A recent case study conducted in Hadiya zone revealed that deficiencies in basic statistical knowledge, statistics courses are given as 'service teaching' with no link to subject area, lack of statistical literacy and inability of students in using statistics in everyday life, lack of theory-based work, including both construct issues with attitudes and lack of models to guide research, and narrow use of restricted research methods, including a heavy reliance on survey type measures are some of the challenges the finding reveals. The country's MOE gave less emphasis on statistics education though it recently launched a policy regarding quality of education. The policy unlikely recognizes other new subjects given for all freshman students in all higher institutions. While most of the universities have launched statistics departments, the issue of statistics course ownership has been a debating issue for years. The fact behind is that statistics education in the country must be owned by statisticians (who has at least bachelor degree in statistics). Thus, all statistics courses offered as service courses or major courses must be taught by statistics major educators. However, the country has no a standardized judging legislation regarding statistics courses given to non-statistics major students. Hence, there must be a national movement led by ESA to answer the professional, dignity, and national responsibility questions by influencing the country's education curriculum with the help of highly endured scientific facts and evidences. As the way forward, the common stand will be developed among the members, the basic game changing inputs/contextual scientific facts or research outputs will be collected, and the leading letter together with the developed persuasive document will be addressed to the HERQA and MOE. The progress will be evaluated until the improvement in the policy.

Improved Socio-Economic Inclusive Gender Statistics for Sustainable National Development

Asnakech Habtamu Tamene: CEO, Social and Inclusive Statistics, Ethiopian Statistical Service

Abstract

The purpose is to address the message to the academic community especially Statistics Professionals the importance of Gender Statistics which incorporates socioeconomic indicators. In doing so, we focuses on the current situation of the National Statistical System in data production and its usage, interaction between data producer and users, highlight about on the ongoing work and contributions of Ethiopian Statistical Service, and areas that needs capacity building. The overall approach is based on the available data source review what we have and show gaps and inform results as a story telling, challenges we face, possible solutions and finally recommendations.

Revolutionizing the Ethiopian Statistics Service: Harnessing Big Data as an Alternative Source

Abstract

In this presentation, we are at the frontier of a significant shift in how we handle data. Traditionally, we have relied on established sources for our statistical analysis, but now we are exploring the integration of big data into our processes. Big data offers vast and diverse datasets that can provide insights beyond what traditional sources offer. We will discuss the potential benefits of this integration, such as enhanced accuracy in our analyses due to the inclusion of a wider range of data points, increased efficiency through automation and faster processing, and improved timeliness of our outputs by tapping into real-time data streams. However, this transition will not be without its challenges. We will address issues like data quality, privacy concerns, and the need for robust infrastructure and analytical tools to handle the volume and variety of big data. Despite these challenges, we are excited about the transformative impact this shift can have on our work. By combining traditional and big data sources, we can gain a more comprehensive understanding of the phenomena we are studying and make better-informed decisions. As we are just beginning this journey, we will share our preliminary plans for using big data in our processes. We are eager to hear feedback and insights from our audience to help shape our approach as we move forward. This presentation is not just about sharing our vision but also about engaging in a dialogue to ensure that we are on the right track to harness the full potential of big data in statistics.

Methodological and Computational Rigor to Facilitate Mental Health Research and Achieve National Institute of Mental Health (NIMH) Research Priorities: A Statistical Perspective

Abera Wouhib

National Institute of Mental Health (NIMH), National Institutes of Health (NIH), USA

Abstract

Innovative and high-quality statistical methods are vital to scientific research for explicitly showing the importance of the research itself and to provide a strong foundation for future research based on current results. The fundamental goal of an extramural research program called Statistical Methods in Psychiatry, located at the National Institute of Mental Health (NIMH), is to foster novel statistical methods and analytical plans that identify and validate biomarkers and novel treatment targets corresponding to psychiatric disorders. In this talk, I will discuss potential pathways for advancement of novel statistical methods that are instrumental to achieving institutional research priorities and to facilitating innovative mental health research funded by the NIMH.

Comparison of Local Powers of Some Exact Tests for a Common Normal Mean With Unequal Variances

Yehenew G. Kifle

Department of Mathematics & Statistics, University of Maryland, Baltimore County, Maryland, USA

Abstract

The inferential problem of drawing inference about a common mean μ of several independent normal populations with unequal variances has drawn universal attention, and there are many exact and asymptotic tests for testing a null hypothesis $H_0: \mu = \mu_0$ against two-sided alternatives. In this talk, I will provide a review of some of these exact and asymptotic tests, and present theoretical expressions of local powers of the exact tests and a comparison. It turns out that, in the case of equal sample size, a uniform comparison and ordering of the exact tests based on their local power can be carried out even when the variances are unknown. At last, I will motivate a similar problem in the context of inference about a common mean vector from several independent multinormal populations with unequal and unknown dispersion matrices.

Estimating the Prevalence of Childhood Polio Vaccination in Ethiopian Zones Using a Hierarchical Bayes Small Area Estimation

Seyifemickael Amare Yilema^{1*}, Yegnanew A. Shiferaw², Haile Mekonnen Fenta³, Alebachew Taye Belay¹

¹Department of Statistics, College of Natural and Computational Science, Debre Tabor University

²Department of Statistics, University of Johannesburg, Johannesburg, South Africa

³Department of Statistics, College of Science, Bahir Dar University, Bahir Dar, Ethiopia

*Corresponding Author: Email: samarey1981@gmail.com

Abstract

Vaccination is one of the most effective, affordable, and life-saving medical therapies ever created. child polio vaccination is the cornerstone of creating a healthy and well-fare society, which is a crucial part of the 2063 African and 2030 global agenda. This study combines auxiliary variables from the 2007 population and housing census with data from the 2019 survey datasets to deploy a hierarchical Bayes (HB) small area estimation (SAE) approach for estimating the local-level child vaccination. The deviance information criterion (DIC) chose the logistic normal mixed model with known sampling variance over the other two models (Fay-Herriot model and log-normal mixed model). The results of this study show that model-based estimates generated from the HB-SAE approach are more precise and reliable than direct survey estimates. The root mean squared error (MSE) and coefficient of variation (CV) indicate that the proposed HB estimates improved survey-based estimates. In addition, this result contributes to the sustainable development goal for health (SDG3), which aims to ensure healthy lives and promote well-being for people around the globe.

Survival Analysis of COVID-19 Patients Hospitalized in Hawassa University Comprehensive Specialized Hospital, Ethiopia: A Cohort Study

Ali B. Anteneh¹, Marissa LeBlanc^{2,3}, Abebe A. Natnae⁴, Asfaw G. Zeytu⁵

¹ Department of Statistics, Hawassa University, Hawassa, Ethiopia, Email: antenehbezabih920@gmail.com

² Oslo Center for Biostatistics and Epidemiology, Oslo University Hospital, Oslo, Norway

³ Norwegian Institute of Public Health, Oslo University Hospital, Oslo, Norway

⁴ Hawassa University Comprehensive Specialized Hospital, Hawassa, Ethiopia

⁵ Department of Epidemiology and Biostatistics, School of Public Health, Addis Ababa University, Addis Ababa, Ethiopia.

Abstract

The COVID-19 pandemic is caused by the global spread of the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), and has resulted from 622,119,701 confirmed cases and 6,546,118 deaths to date. Most studies of the clinical features of hospitalized COVID-19 patients are from high-income countries, with little information available for developing countries such as Ethiopia. In this study, clinical and demographic characteristics, as well as risk factors for in-hospital mortality, are assessed in Hawassa, Ethiopia. The cohort includes 804 consecutive clinical diagnoses and/or radiological findings consistent with signs and symptoms of COVID-19 and/or laboratory-confirmed hospitalized COVID-19 patients from September 24, 2020, to November 26, 2021, at Hawassa University Comprehensive Specialized Hospital. The in-hospital mortality rate was predicted using Cox regression. The median age was 45 [min 0.006, max 100] years, with males making up 64.1% of the population. There were 173 (21.5%) deaths, with 125 (72.3%) deaths among males. Male patients died at a higher rate than females in each age category. Severe and critical cases represented 24% and 21% of the study population, respectively. Comorbidities were common, with 49.1% having at least one and 12.6% having several comorbidities. The most prevalent comorbidity was diabetes (15.9%), followed by hypertension (15.2%). The findings of univariate Cox regression in a large cohort of hospitalized COVID-19 patients in Ethiopia revealed that gender, age category, source of referral, severity status, various symptoms, and comorbidities were significantly associated with an increased risk of in-hospital death. The study indicated that age, gender, comorbidities, and symptoms at admission were all substantial risk factors for COVID-19 death, just as they were in high-income nations. However, in Ethiopia, unlike in wealthier countries, the average age of COVID-19 patients was remarkably young. Multivariate cox-proportional hazard analysis supports, patients suffering from COVID-19 and having certain factors such as disease severity, older age group, comorbidity, shortness of breath, sore throat, general body weakness, hypertension, diabetes, and those who visited health facilities before beginning had a higher risk of dying from the disease. Patients who had at least one symptom had a reduced risk of death. The study emphasizes the need to identify and address these patient factors during inpatient therapy to improve the chance of survival. As a result, the current study identifies COVID-19 patient factors that should be addressed throughout inpatient therapy.

A joint model of longitudinal and competing risks survival data; With Application to Chronic heart failure patient's data

Yikeber Abebaw*1, Mekuanent walle2

¹ Department of Statistics, Debre Tabor University, Debre Tabor, Ethiopia. Email: ykebera@gmail.com

² University of Messina, Italy.

Abstract

In recent years, there has been a significant increase in the methodological advancement and practical use of integrated models of longitudinal and time-to-event outcomes. In clinical and epidemiological studies, repeated measurements on a response are generally obtained along with time-to-event outcome, modeling them separately might lead to biased estimates. A joint modelling approach is required to obtain unbiased results and to evaluate their association. Most joint models for longitudinal and survival data have only one failure type for the time to event. However, in the joint model, a subject may be exposed to more than one type of failure event (competing risks) and considering the competing event as an independent censoring of the timeto-event process could lead to erroneous conclusions and a misinterpretation of an actual survival outcomes. Therefore, extend the methodology to allow for competing risks data is important. The main aim of this study is to model jointly competing risks failure time and longitudinal data by linking the cause-specific hazards model for competing risks analyses with a linear mixed model for longitudinal measurements through shared random effects model. We illustrated the proposed methods by analyzing real data on a chronic heart failure patients. For this study death, time to complication and mean longitudinal change of pulse rate was the event of interest, competing event and the longitudinal outcome respectively. The result revealed that the estimated value for the association parameter was positive and statistically significant. This indicates that there is significant evidence of an association between the longitudinal outcome, the risk of death ($\gamma_1 = 0.04$; $P = 0.0081$) and complication time of congestive heart failure patients ($\gamma_2 = 0.03$; $P = 0.0447$). Left ventricular ejection fraction, New York Heart Association class, gender, residence and age were statistically significant factors associated with outcome variables the in a competing risk context.

A New Alpha Power Type-1 Family of Distributions and Modelling the Overdispersed Count Outcome

Getachew Tekle¹, Rasool Roozegar^{*1}, Hamid Bidram²

¹Department of Statistics, Yazd University, Iran

²Department of Statistics, Isfahan University, Iran

Abstract

In this paper, we introduce a novel family of statistical models called a new 9 alpha power type-1 family of distributions. Three sub-cases of the family are discussed. Based on the novel family, a special model, explicitly, a new alpha power type-1-Weibull distribution is studied in depth. The new model has very interesting patterns of failure rates like increasing, decreasing, bathtub, and parabola-down. Hence, it is so flexible. Based on the comparison analysis, among five well-known models, it has an impact on the health data analysis. Furthermore, the count data models capable of handling overdispersion and zero-inflation are discussed and applied to the real health data. The zero-inflated negative binomial (ZINB) model in the frequentist approach has shown its popularity in handling both overdispersion and zero-inflation simultaneously, while the discrete Weibull (DW) model with the logit(q) link in the Bayesian approach outperformed its counterparts.

Bayesian Propensity Score Approaches for Balancing Covariates Associated with Low Birth Weight in North Shoa Zone, Ethiopia

Wudneh Ketema^{1,2*}, Awoke Seyoum Tegegne¹, Aweke A.Mitku^{1,5}, Solomon Hailemeskel³, Esubalew Tesfahun⁴

¹Department of Statistics, Bahir Dar University, Bahir Dar, Ethiopia. E-mail: wudnehketema@gmail.com

²Department of Statistics, Debre Berhan University, Debre Berhan, Ethiopia

³Department of Midwifery, Debre Berhan University, Debre Berhan, Ethiopia

⁴Department of Public Health, Debre Berhan University, Debre Berhan, Ethiopia

⁵School of Mathematics, Statistics and Computer Science, University of KwaZulu-Natal, Durban, South Africa

Abstract

The birth weight of an infant, typically measured within the first hours after birth, is crucial for assessing their health. Low birth weight (LBW) can result from intrauterine growth restriction, prematurity, or a combination of both factors. The study aimed to estimate the treatment effect using Bayesian causal inference and to assess covariates associated with LBW in the North Shoa Zone, Ethiopia. A prospective non-randomized (quasi-experimental study design) was employed from August 2019 to September 2020 in the North Shoa Zone, Amhara Regional State, Ethiopia. For the Bayesian causal inference approaches in estimating the average treatment effect (ATE), Markov Chain Monte Carlo algorithms were used. The performance of the standard nearest-neighbor propensity score and the weighting Bayesian propensity score matching procedure were compared to each other. In the Bayesian approach, the nearest neighbor matching method estimated an (ATE) of 2.065 (SE = 0.875), while the Frequentist approach yielded an estimate of 2.156 (SE = 0.86). Although bias differences were not statistically significant, Bayesian estimates exhibited less noise due to incorporating more information from a larger number of observations, resulting in higher accuracy. Both methods offer valuable insights, but the Bayesian approach offers advantages in noise reduction and accuracy to estimate ATE. The counterfactual estimate of the Bayesian method indicates that the true causal relationship was reliably estimated by the Bayesian credible interval effect of the LBW of newborn babies. More attention should be given to pregnant women with an abortion history, with high blood pressure, with more folic acid, with an emergency cesarean, women with premature babies, women with low nutritional status, and women with spontaneous vaginal delivery.

Identification of Confounders for the Association between Binary Exposures with Ordinal Outcome

Ashagre Sharew Iyassu^{1,4} Haile Mekonnen Fenta^{1,2} Zelalem G Dessie^{2,3} Temesgen T. Zewotir³

¹College of Science, Bahir Dar University, Bahir Dar, Ethiopia

² Department of Statistics, University of Pretoria, Pretoria, South Africa

³School of Mathematics, Statistics & Computer Science, University of KwaZulu Natal, South Africa

⁴DebreMarkos University, Debre Markos, Ethiopia

Abstract

In causal analyses, some third factor may distort the relationship between the exposure and the outcome variables under study, which gives spurious results. In this case treatment groups and control groups that receive and do not receive the exposure are different from one another in some other essential variables, called confounders. Place of birth and postnatal care as exposures and age-specific childhood vaccination status as outcome variables were used. Multiple logistic regression, log-binomial regression and modified (robust) Poisson regressions were used and compared to model the relationship between covariates and exposures. Ordinal logistic regression was used to model the relationship between outcome variable versus covariates and exposures variables. Significance testing at P value ≤ 0.2 and change in estimate methods were used and compared to identify confounders. Odds ratio estimated from logistic regression is greater than risk ratio, estimated from log-binomial and modified Poisson regression. Log-binomial regression performs better in postnatal care. Using the AIC and likelihood ratio test, disjunctive approach outshines the common cause approach. However, based on the percent change of exposures' effect, common cause approach outshines the disjunctive approach for place of delivery and the opposite is true for postnatal care. A 7% and 5% change in estimate of place of birth and postnatal care respectively give better results for confounder identification. Change in estimate method is stricter than the significance testing method. It is essential to use log-binomial or modified Poisson regression for binary outcome with more than 12% prevalence rate, which is consistent with literatures, 10%. It may be important to lower the cut point of change in estimate to 5% as compared to 10% cut-off point of change in estimate of exposure. It needs caution to choose significance testing or change in estimate method of confounder identification. It is important to identify and adjust confounders to know the real effect of an exposure on the outcome. Adjusting confounders, place of delivery and postnatal care increases the probability of getting age-specific full vaccination.

Artificial Intelligence-Driven Ensemble Model to Predict Mortality due to Covid-19 in East Africa

Kedir Hussein Abegaz^{1,2*} and İlker ETİKAN³

1. Department of Biostatistics, Faculty of Medicine, Near East University, Near East Avenue, Postcode: 99138, Nicosia, North Cyprus, Mersin 10, Turkey
2. Biostatistics and Health Informatics, Public Health Department, Madda Walabu University, Ethiopia Email: kedir6300@gmail.com
3. HOD Biostatistics, Faculty of Medicine, Near East University, Near East Avenue, Postcode: 99138, Nicosia, North Cyprus, Mersin 10, Turkey

Abstract

COVID-19 is the first pandemic that had a brutal impact and triggered a near-to-total shutdown of social and economic activity across the world. This pandemic killed more than six million people within less than three years and it demonstrated that the world is unprepared for such a rapid spread of the virus as a result of a devastating pandemic. However, it is believed that Artificial intelligence (AI) can understand how it happened and what will look like in the future. Hence, this study aimed to predict mortality due to COVID-19 using an AI-driven ensemble model in East Africa. The two years daily COVID-19 dataset was split into training and verification datasets. To predict COVID-19 mortality three steps were conducted. Firstly, non-linear sensitivity analysis is to select dominant inputs. Secondly, four AI-driven models (FFNN, ANFIS, SVM and MLR) to predict COVID-19 mortality independently. Finally, four ensemble approaches (NNE, ANFISE, SAE, and WAE) were developed. Then, the predicted COVID-19 mortality from the ensemble models was compared with results from each of AI-driven models. Four dominant inputs were selected to conduct single AI-driven models and the performance measures (DC and RMSE) were conducted. Hence, the DC of ANFIS with (4-6-1), FFNN with Gaussian, SVM with RBF, and MLR with (4-1) structures were 0.9273, 0.8586, 0.8490, and 0.7956 respectively. Among ensemble models, the DC of ANFISE was 0.9886 and improve the performance of single AI-drive models FFNN, ANFIS, SVM, and MLR by 13, 6.1, 13.9, and 19.3 percent, respectively. The non-linear ensemble approaches highly performed than the linear approaches in the current study. ANFISE was the most performing ensemble approach that boost the predicting performance of single AI-driven models in east Africa. This fact revealed the promising capability of ensemble models to predict the daily mortality of COVID-19 in other parts of the world.

Cardiac patients' surgery outcome and associated factors in Ethiopia: application of Machine learning

Melaku Tadege^{*1,2,3} Awoke Seyoum Tegegne¹, Zelalem G. Dessie^{1,4}

¹ College of Science, Bahir Dar University, Bahir Dar, Ethiopia E-mail: melakutadege@yahoo.com

² Department of Statistics, Injibara University, Injibara, Amhara, Ethiopia

³Regional Data Management Center for Health (RDMC), Amhara Public Health Institute (APHI), Bahir Dar, Ethiopia

⁴School of Mathematics, Statistics and Computer Science, University of KwaZulu-Natal, Durban, South Africa

Abstract

Cardiovascular diseases are a class of heart and blood vessel-related illnesses. In Sub-Saharan Africa, including Ethiopia, preventable heart disease continues to be a significant factor, contrasting with its presence in developed nations. Therefore, the objective of the study was to assess the prevalence of death due to cardiac disease and its risk factors of among heart patients in Ethiopia. The current investigation included all cardiac patients who had cardiac surgery in the country between 2012 and 2023. A total of 1520 individuals were participated in the study. Data collection took place between February 2022 and January 2023. The study design was a retrospective cohort since the study track back patients' chart since 2012. Machine learning algorithms were applied for data analysis. For machine learning algorithms comparison, lift and AUC was applied. From all possible algorithms, logistic algorithm at 90%/10% was the best fit since it produces the maximum AUC value. In addition, based on the lift value of 3.33, it can be concluded that the logistic regression algorithm was performing well and providing substantial improvement over random selection. From the logistic regression machine learning algorithms, age, saturated oxygen, ejection fraction, duration of cardiac center stays after surgery, waiting time to surgery, hemoglobin, and creatinine were significant predictors of death. Some of the predictors for the death of cardiac disease patients are identified as such special attention should be given to aged patients, for patients waiting for long periods of time to get surgery, lower saturated oxygen, higher creatinine value, lower ejection fraction and for patients with lower hemoglobin values.

Developing Clinical Prognostic Models to Predict Graft Survival after Renal Transplantation: Comparison of Statistical and Machine Learning Models

Getahun Mulugeta^{*1}, Temesgen Zewotir², Awoke Seyoum Tegegne³, Leja Hamza Juhar⁴, & Mahteme Bekele Muleta⁴

^{1*}Department of Statistics, Bahir Dar University, Ethiopia. Email: gech.marr@gmail.com

²School of Mathematics, Statistics & Computer Science, University of KwaZulu Natal, South Africa

³Department of Statistics, Bahir Dar University, Ethiopia

⁴St. Paul's Hospital Millennium Medical College, Ethiopia.

Abstract

Renal transplantation is a critical treatment that can save the lives of individuals who are suffering from end-stage renal disease (ESRD), but graft failure remains a significant concern. Accurate prediction of graft survival after renal transplantation is crucial as it enables clinicians to identify patients at higher risk of graft failure. This study aimed to develop clinical prognostic models for predicting graft survival after renal transplantation and compare the performance of various statistical and machine learning models. The study utilized data from a retrospective cohort of renal transplant recipients at the Ethiopian National Kidney Transplantation Center from September 2015 to February 2022. Various statistical and machine learning models were evaluated based on their discrimination, calibration, and interpretability. The comparison of models included standard Cox, Lasso-Cox, Ridge-Cox, Elastic net-Cox, Random Survival Forest, and Stochastic Gradient Boosting. The prognostic predictors of graft survival were selected based on the significance and relative importance of variables in different models. The study analyzed a total of 278 completed cases and observed the event of graft failure in 21 patients. The median graft survival time was 33 months, and the mean hazard of graft failure was 0.0755. The study found that the Random Survival Forest and Stochastic Gradient Boosting models demonstrated the best calibration and discrimination performance shown by an equal AUC of 0.97 and the overlapped calibration plots. On the other hand, the Cox proportional hazards model has the highest interpretability and established superior accuracy in estimating survival probabilities, as evidenced by its lowest Brier score of 0.000071. The current study indicates that an episode of chronic rejection, recipient residence, an episode of acute rejection, post-transplant urological complications, post-transplant nonadherence, blood urea nitrogen level, and number of post-transplant admissions were consistently identified as the top significant prognostic predictors of renal graft survival. The Random Survival Forest and Stochastic Gradient Boosting models demonstrated superior calibration and discrimination performance, while the Cox proportional hazards model offered accurate estimation of survival probabilities and interpretability. Clinicians should consider the trade-off between performance and interpretability when choosing a model. Incorporating these findings into clinical practice can improve risk stratification, enable early interventions, and inform personalized management strategies for kidney transplant recipients.

Determinants of Blood Sugar Level and Time to First Remission for Type I diabetic patients under treatment; A comparison of Separate and Joint models

Maru Zewdu Kassie¹

¹Department of statistics , Assosa University, Assosa , Ethiopia. E-mail: maruzewdu68@gmail.com

Abstract

This study aimed to investigate the association between longitudinal fasting blood sugar and time to first remission among T1DM patients at Debre Tabor General Hospital in North-west Ethiopia. A retrospective study was conducted on 217 randomly selected patients from January 2018 to January 2020. The linear mixed model and Cox proportional hazard model were used for the longitudinal and survival parts, respectively. The joint model analysis was employed to link these models. 68.1% of patients achieved first remission, with a significant negative association between fasting blood sugar and remission occurrence ($p < 0.0001$, $\alpha = -1.7914$). Thus, first remission was less likely to occur in patients with higher Fasting Blood Sugar and vice versa. The joint model outperformed separate models in variability, fit, and significance, making it the preferred choice for analyzing longitudinal and survival data.

Spatio-temporal Distributions and Determinants of Diarrhea among under-five Children in Ethiopia

Meskerem Tiku¹, Melkamu A. Zeru^{2*} and Denekew Bitew Belay²

¹Department of Statistics, Injibara University, Injibara, Ethiopia Email: melkamu.ayana@gmail.com

²Department of Statistics, Bahir Dar University, Bahir, Ethiopia

Abstract

The purpose of this study was to evaluate Ethiopia's childhood diarrhea disease spatio-temporal pattern and pinpoint its contributing factors. We conducted analyses on secondary data from four Ethiopian Demographic and Health Surveys conducted between 2000 and 2016. Moran's I was used to determine spatial dependence and spatial models were used to evaluate variables associated with diarrhea in under-five children at the zonal level. Childhood diarrhea showed spatial clustering in Ethiopia (Moran's I; $p < 0.05$). The spatial regression model revealed significant factors at the zonal level: children born at home ($e^{\theta}=1.355$, 95% CI: 1.052-1.544, $p < 0.001$), low birth weight ($e^{\theta}=1.18$, 95% CI: 1.017-1.691, $p < 0.05$), and unimproved source water ($e^{\theta}=0.8568$, 95% CI: 0.671-1.086, $p < 0.01$). The prevalence of diarrhea among under-five children varied over time by zone, with the Assosa, Hundene, and Dire Diwa zones having the highest rates. Home births and low birth weight contributed to the prevalence of childhood diarrhea. In high-risk zones of Ethiopia, reducing childhood diarrhea requires integrated child health interventions and raising awareness about the potential hazards associated with unimproved water sources.

Multivariate Spatio-temporal Modelling of Malaria count of Genus *P. falciparum* and *P. vivax* In Case of Southern Ethiopia

Yonas Kitawa*¹ and Zeytu Asfaw*^{2, 1}

¹Department of Statistics, Hawassa University, Hawassa, Ethiopia. Email: yonasshuke@hu.edu.et

²Department of Bio-statistics and Epidemiology, Addis Ababa University, Addis Ababa, Ethiopia

Abstract

*Although malaria incidence has fallen sharply over the past few years, the rate of decline varies by district, time, and malaria type. Despite this turn-down, malaria remains a major public health threat in various districts of Ethiopia. Consequently, the present study aims to develop a predictive model that helps to identify the spatio-temporal variation in malaria risk by multiple Plasmodium species. We propose a multivariate spatio-temporal Bayesian model to obtain a more coherent picture of the temporally varying spatial variation in disease risk. The spatial autocorrelation in such a data set is typically modelled by a set of random effects that assign a conditional autoregressive prior distribution. However, the autocorrelation considered in such cases depends on a binary neighbourhood matrix specified through the border-sharing rule. Over here, we propose a graph-based optimization algorithm for estimating the neighbourhood matrix that merely represents the spatial correlation by exploring the areal units as the vertices of a graph and the neighbour relations as the series of edges. Furthermore, we used aggregated malaria count in southern Ethiopia from August 2013 to May 2019. We recognised that precipitation, temperature, and humidity are positively associated with the malaria threat in the area. On the other hand, enhanced vegetation index, nighttime light (NTL), and distance from coastal areas are negatively associated. Moreover, nonlinear relationships were observed between malaria incidence and precipitation, temperature, and NTL. Additionally, lagged effects of temperature and humidity have a significant effect on malaria risk by either species. More elevated risk of *P.falciparum* was observed following the rainy season and unstable transmission of *P.vivax* was observed in the area. Finally, *P.vivax* risks are less sensitive to environmental factors than that of *P.falciparum*. The improved inference was gained by employing the proposed approach in comparison to the commonly used border-sharing rule. Additionally, different covariates are identified including delayed effects, and elevated risks of either of the cases were observed in districts found in the central and western regions. As malaria transmission operates spatially continuously, a spatially continuous model should be employed when it is computationally feasible.*

Spatiotemporal Models with Confounding Effects: Application on Under-Five Mortality across sub-Saharan African Countries

Haile Mekonnen Fenta^{*1,2} Ding-Geng Chen^{1,3}, Temesgen T Zewotir⁴, Najmeh Nakhaei Rad¹

¹Department of Statistics, University of Pretoria, Pretoria, South Africa. Email: hailemekonnen@gmail.com

²Department of Statistics, Bahir Dar University, Bahir Dar, Ethiopia

³College of Health Solutions, Arizona State University, Phoenix, Arizona, USA

⁴School of Mathematics, Statistics and Computer Science, University of KwaZulu-Natal, Durban, South Africa

Abstract

Different strategies have been developed to minimize under-five mortality (U5M) in sub-Saharan African (sSA) countries, but it is still a major child health concern in the region. Spatiotemporal modeling is important for areal data collected over time. However, when the number of time points and spatial areas is large, and the areas are disconnected, fitting the model becomes computationally complex due to the high number of required parameters to be estimated. Therefore, the main aim of this study is to adopt the spatiotemporal dynamic model which includes the confounding effects between time, space, and their interactions with fixed covariates, with special emphasis on U5M across disconnected sSA countries. We used the nationally publicly representative Demographic and Health Survey (DHS) data for the period from 2000 to 2020. Bayesian spatiotemporal hierarchical modelling with integrated nested Laplace approximation (INLA) program was used to model the spatiotemporal distribution of U5M among children across 37 districts located in four disconnected sSA regions, consists of Ethiopia, Nigerian, Zimbabwe and Ghana. A total of 170,356 under-five children from 37 districts were considered and 15,467 died before the age of five. The relative risk of U5M in the first DHS was 2.02 and it sharply decreased to 0.5 in the recent phase. The proportion improved access to water, sanitation, clean fuel use, urbanization and access to health facility in the district had a significant negative association with U5M: the higher the proportion of these covariate, the lower the prevalence of childhood mortality. This study revealed the evidence of strong spatial, temporal, and interaction effects influencing the under-five mortality risk across the districts. Improving the women's literacy index, access to improved water, the use of clean fuel, and the wealth index are associated with the improvement of risk of mortality among under-five children across the districts. Special attention should be paid to districts found in Nigeria and Ethiopia.

Model-free screening for variables with treatment interaction

Shiferaw Befkadu^{1*}, Jin Xu²

¹ Jimma University, Ethiopia

² East China Normal University, Shanghai 200062, China

Abstract

Precision medicine is a medical paradigm that focuses on making effective treatment decision based on individual patient characteristics. When there are a large amount of patient information, such as patient's genetic information, medical records and clinical measurements, available, it is of interest to select the covariates which have interactions with the treatment, for example, in determining the individualized treatment regime where only a subset of covariates with treatment interactions involves in decision making. We propose a marginal feature ranking and screening procedure for measuring interactions between the treatment and covariates. The method does not require imposing a specific model structure on the regression model and is applicable in a high dimensional setting. Theoretical properties in terms of consistency in ranking and selection are established. We demonstrate the finite sample performance of the proposed method by simulation and illustrate the applications with two real data examples from clinical trials.

Linear regression model with doubly truncated generalized new symmetric error distribution

Bisrat Misganaw Geremew ^{*1}, K.Srinivasa Rao²

¹Department of Epidemiology and Biostatistics, University of Gondar, Gondar, Ethiopia. Email: bisratuog@gmail.com

²Department of Statistics, Andhra University, Vishakhapatnam 530 003, India

Abstract

The linear model is one of the most popular models in statistical analysis. It has been received significant applications in many areas of Science, Engineering, Medicine and Econometrics. Most of the inferential procedures are based on the basic assumption of the linear regression model that the error terms have normal distribution. Regression analysis is one of the most commonly used statistical methodologies in many branches of science and engineering used for discovering functional relationships between variables. The regression model theories and applications are studied by many authors and usually to consider that the error term follows Gaussian distribution which is mesokurtic and having infinite range. Recently, there have been a several studies at the influences of non-Gaussian on linear regression analysis. In many data sets arising at places like biological experiments, agricultural experiments, financial analysis, industrial & chemical experiments, space experiments, etc., the error term may not have mesokurtic, rather distributed as symmetric and platykurtic, in nature and its range is finite. For such type of data using the regression model with Gaussian errors may badly fit. In this regard, Srinivasa Rao et al. (1997) have introduced a generalization of the normal distribution termed as the New Symmetric (NS) distribution. Then, the researcher developed and analysed linear models with NS distributed errors. But this distribution has infinite range $(-\infty, \infty)$. In practice the range of many random variables will have a finite range, non-normal errors and different shape of curves (γ). So, to incorporate the reality of such type of data sets we introduce a linear regression models with doubly truncated generalized new symmetric distributed (DTGNS) errors for the first time. The method of maximum likelihood (ML) and OLS were employed to estimate the model parameters. The various distributional properties like distribution function, the moment generating function, the characteristic function, mean, variance, skewness, kurtosis, and cumulants of this distribution are discussed. Different shapes of the frequency curves of the doubly truncated generalized new symmetric distribution were presented for various values of parameters. Simulation study is conducted to compare the efficiency of proposed model with that of new symmetric distributed errors and Gaussian errors. The Akaike information criterion (AIC) and Bayesian information criterion (BIC); and model diagnostics technique root mean square error (RMSE) are used for determining the best model for a given data. Through a Monte Carlo simulation study using the simulation study the properties of these estimators for the different sample sizes of $n=1000, 2000, 3000, 5000, \text{ and } 10,000$ are obtained and analyzed. It is observed that the proposed model, doubly truncated generalized new symmetric distributed errors, performs much better than linear model with new symmetric and normal distributed errors. It is also observed that the ML estimators are more efficient than OLS estimators when error terms follow doubly truncated generalized new symmetric distributions.

The models developed in this thesis are also useful for developing technological advancements and performance evaluations of several systems in engineering and business. The doubly truncated generalized new symmetric distribution which is used for characterizing the structure of error terms is much useful in analyzing several mesokurtic and platykurtic distributed variables. It is possible to develop many more regression models with plausible conditions in order to analyse the natural phenomena more close to reality. These models are more useful for analyzing data sets arising from agricultural experiments, biological experiments, financial analysis, space experiments, etc.

Flexible Bayesian Semiparametric Mixed-Effects Model for Complex Longitudinal Data: A Simulation and Application Study

Melkamu M. Ferede^{1*}, Workagegnehu H. Bilchut², Getachew A. Dagne³, Samuel M. Mwalili⁴

¹Department of Statistics, University of Gondar, Gondar, Ethiopia. Email: melkamum2m@gmail.com

²Department of Internal Medicine, University of Gondar, Gondar, Ethiopia.

³Department of Epidemiology and Biostatistics, University of South Florida, Tampa, FL 33612, USA.

⁴Department of Statistics and Actuarial Sciences, Jomo Kenyatta University of Agriculture and Technology (JKUAT), Nairobi, Kenya.

Abstract

In clinical trials and epidemiological research, mixed-effects models are commonly used to examine population-level and subject-specific trajectories of biomarkers over time. Despite their increasing popularity and application, the specification of these models necessitates a great deal of care when analyzing longitudinal data with nonlinear patterns and asymmetry. Fully parametric (linear) mixed-effect models may not capture these complexities flexibly and adequately. Additionally, assuming a Gaussian distribution for random effects and/or model errors may be overly restrictive, as it lacks robustness against deviations from symmetry. This paper presents a semiparametric mixed-effects model with flexible distributions for complex longitudinal data in the Bayesian paradigm. The nonlinear time effect on the longitudinal response was modelled using a spline approach. The multivariate skew-t distribution, which is a more flexible distribution, is utilized to relax the normality assumptions associated with both model errors and random-effects. Simulation studies were conducted to evaluate the performance of the proposed models. The models were then applied to chronic kidney disease data to assess covariates associated with the estimated glomerular filtration rate, measure of kidney function. Various model comparisons were performed by considering different functional and distributional specifications. First, we compared the proposed semiparametric partially linear mixed-effect (SPPLM) model with the fully parametric one (FPLM), and the results indicated that the SPPLM model outperformed the FPLM model. We then further compared four different SPPLM models, each assuming different distributions for the random effects and model errors. The proposed semi-parametric mixed-effect model with a skew-t distribution outperformed to the CKD data compared to the existing Gaussian models. The findings from the application revealed that hypertension, diabetes, and follow-up time had a significant association with estimates of declining kidney function. The simulation and application studies have demonstrated that our work has made a significant contribution towards a more robust and adaptable methodology for modeling complex longitudinal data.

Bayesian Multivariate joint modeling of longitudinal and time-to-event data using HIV/AIDS patients

Denekew Bitew Belay^{1,2*}, Ding-Geng Chen^{2,3}, Najmeh Nakhaei Rad²

¹Bahir Dar University, Department of Statistics, Bahir Dar, Ethiopia. Email: denekew.t.h@gmail.com

²University of Pretoria, Department of Statistics, Pretoria, South Africa

³Arizona State University, College of Health Solution, Phoenix, AZ, USA

Abstract

HIV/AIDS is still a global health burden and causes a significant number of morbidity and mortality. The application of joint modeling is a statistical approach that simultaneously analyses multiple longitudinal outcomes and time-to-event data which recently became crucial in HIV/AIDS research. This study aims to quantify the risk of time-dependent biomarkers and the time to death of HIV/AIDS patients using the Bayesian joint model estimation technique. A total of 571 patients' registration cards were accessed and their information is recorded from Felege Hiwot Comprehensive Referral Hospital, Ethiopia. From a total of 571 patients, there are 315 female and 256 male patient participants with 55.2% and 44.8% respectively. This study shows that the CD4 cell counts and weight of patients are significantly associated biomarkers with the time to death of patients. The estimated association parameters for CD4 counts and weight are significantly associated with time to death of the patients by considering different functional forms and association structures. The CD4 count has an estimated posterior mean of -0.14 which indicates that the hazard of death of a patient decreases by 13% as a unit decrease in squared root transformed of CD4 cell count. Similarly, a weight of the patients has an estimated posterior mean of -4.1 at its slope which indicates the hazard of a patient decreases by 98.3% as a unit increase of weight of a patient which shows an instantaneous effect on time to death of patients. The use of both longitudinal and time-to-event data jointly gives a more robust estimation and this will help the true patient-specific intervention. In this study we analyzed the two longitudinal biomarkers on time to death of HIV/AIDS patients which led to the new hypothesis that the patient's weight should be considered as important biomarker as CD4 cell count to determine the progress of the patient's health status.

Multivariate Time Series Modeling of the Relationship Between Major Economic Indicators in Ethiopia.

Daba Ketema^{1*}

¹Department of Statistics, Bule Hora University, Bule Hora, Ethiopia

Abstract

We model the relationship between major economic indicators in Ethiopia. We obtained 28 years data from National Bank of Ethiopia (NBE) for the period between 1991-2018, these data has been transformed into quarter data. The data exhibit up-ward trend and volatility clustering. Multivariate Generalized Autoregressive Conditional Heteroskedasticity (MGARCH) model were employed for volatility modeling with the specific interest in identifying; the trend of macroeconomic variables and the impact of macroeconomic variables on economic growth. The result from Dynamic Conditional Correlations (DCC(1,1))-GARCH showed that the sum of ARCH and GARCH effect is 0.672, which means volatility is neither permanent nor explosive. Meaning, a shock to volatility in one period will not lead to a greater volatility in the next period. This study recommends that; government and non-governmental organizations should increase agricultural products by improving infrastructure, provision of labor force, training to farmers as well as strategies like loan provision schemes with affordable interest rates and establishment of suitable markets for their products. NBE should be cautious in controlling money supply to control the problem of inflation, since the increase in money supply increases inflation, and inflation affects economic growth.

Additive Modeling of Zonal Level Crop Production in Ethiopia

Yidnekachew Mare^{1,2*}, Denekew Bitew Belay² and Temesgen Zewotir³

¹Department of Statistics, Arba Minch University, Arba Minch, Ethiopia. Email: yidnekachew.mare08@gmail.com

²Department of Statistics, Bahir Dar University, Bahir Dar, Ethiopia

³School of Mathematics, University of KwaZulu-Natal, Durban, South Africa

Abstract

Crop production is main source of food security and income for smallholder private farmers in Sub-Saharan countries, and to have a sustainable source of food security and economy, it is important to identify covariates affecting crop production linearly and nonlinearly to evaluate and change, if necessary, the current working policies. The main objective of this study is to identify the set of covariates that have linear and nonlinear effects on crop production and estimate their effects using an additive mixed effect model on zonal-level agricultural sample survey data for Meher seasons from 2012/13 to 2019/20. Based on the summary results, the minimum, mean, and maximum crop production by farmers across the country for the study period were 1.616, 8.693, and 147.843 quintals, respectively, and 50% of the farmers produced less than 6.95 quintals. The histogram, kernel density, and P-P plots suggested that log-transformed crop production is approximately normally distributed. From competing models' summary statistics, information criteria values and analysis of variance tests, relaxing linearity assumption and including random effect in the model has improved model performance, suggesting the additive mixed effects model best fits the data on hand. Gambella, SNNP, and Oromia regions have significantly different overall mean crop production than the reference in Dire Dawa town. Covariates like year, proportion of female farmers, household age and UREA fertilizer used have a significant nonlinear effect, while covariates like proportion of educated farmers, area used and proportion of farmers who received credit service have a significant linear effect on log crop production. The basic model assumptions are not violated, and as a recommendation, the final additive mixed effects model can be used for identifying the direction and underlying significant relationship between covariates and crop production and prediction and inference purposes. Based on the significant linear and non-linear relationships, stakeholders are strongly advised to take appropriate actions and change policies regarding participation of female and educated farmers, implementation of credit and advisory services and providing inputs.

Spatial and Survey Regression Analysis of Zero-Dose Immunization Coverage among 12-23 month-old Children in Ethiopia

Kindie Fentahun Muchie^{1*}

¹Department of Epidemiology and Biostatistics, Bahir Dar University, Bahir Dar, Ethiopia. Email: mkindief@gmail.com

Abstract

Vaccination is one of the successful and cost-effective public health interventions, saving millions of lives. Despite this, many children, especially those living in low and middle-income countries continue to miss out on life saving vaccines. Addressing zero-dose prevalence is critical to achieve the WHO's immunization agenda 2030 target, promises to leave no one behind. This study aimed to determine the prevalence, spatial distribution and associated factors of zero dose immunization among 12-23 children in Ethiopia. The EDHS 2019 data (survey data, GPS data and spatial covariate data) in Ethiopia was used. This study used children data of Ethiopia on immunization to assess zero dose immunization. Weighted percentages, mean and median were considered as descriptive measures. Spatial analysis and design weighted regression analysis were done. The overall weighted prevalence of zero dose immunization was 23.7% [95% CI: 18.7-28.5]. Significant clusters of high prevalence of zero dose immunization were identified. Rural residents [AOR=2.55; 95%CI: 1.05, 6.22], female children (AOR=1.78; 95%CI: 1.09, 2.91), children not delivered at the health institution (AOR=4.12; 95% CI: 2.39, 7.08), and children from a mother did not completed four or more ANC visits (AOR=2.55; 95% CI: 1.37, 4.75) were more likely to experience zero dose immunization as compared to their counter parts. There is high prevalence of zero dose immunization among 12-23 months old children in Ethiopia. However, interventions tailored on geographic areas, residence, sex of the child, four or more ANC visits and institutional delivery could help to improve immunization coverage among children in Ethiopia.