
MSC IN BIOSTATISTICS
PROGRAMME OUTLINE
2021

DIVISION OF EPIDEMIOLOGY AND BIOSTATISTICS
FACULTY OF MEDICINE AND HEALTH SCIENCES
STELLENBOSCH UNIVERSITY

06 CONTACT DETAILS


PROGRAMME CONVENOR

Prof Taryn Young

Division of Epidemiology and Biostatistics

Department of Global Health

Faculty of Medicine and Health Sciences, Stellenbosch University

 +27 21 938 9886

 tyoung@sun.ac.za


PROGRAMME ADMINISTRATOR

Mrs Liesel Floor

Division of Epidemiology and Biostatistics

Department of Global Health

Faculty of Medicine and Health Sciences, Stellenbosch University

 +27 21 938 9314

 iieselfloor@sun.ac.za

CONTENTS

Overview of course structure	2
Admission requirements	5
Structured modules: Objectives and contents	
Statistical Bridging Course	6
Mathematical Statistics	7
Linear Models	8
Fundamentals of Epidemiology	9
Data Management & Statistical Computing	10
Principles of Statistical Inference	11
Categorical Data and Generalized Linear Models	12
Analysis of Survival Data	13
Analysis of Observational Data	14
Longitudinal Data Analysis	15
Biostatistical Consulting	16
Bayesian Data Analysis	17
Clinical Biostatistics	18
Design and Analysis of Clinical Trials	19
Multivariate Statistics	20
Bioinformatics	21
Research assignment	22
Bursaries	23
Contact details	24

Information summarised in this document is subject to change

Biostatistics is a branch of statistics concerned with how we ought to make decisions when analyzing biomedical data. It is an evolving discipline concerned with formulating explicit rules to compensate both for the fallibility of human intuition in general and for biases in study design in particular.

MSc in Biostatistics courses offer rigorous training for those with a background or experience in quantitative or health-related discipline who wish to pursue a career in biostatistics. The programme would be of interest to potential biostatisticians who require practical and technical skills, and the application of principles of statistical reasoning to address public health problems and challenges.

01 OVERVIEW OF COURSE STRUCTURE

This is a 180 credit programme which consists of modules (120 credits) and a research assignment (60 credits).

Types of learning activities	Credits <i>1 credit = 10 hours</i>
Modules	120
Research assignment	60

Nature and duration of programme: The programme is offered on a full-time basis over a period of two years.

Language specification: English

Modules: Students need to do 12 modules, of which ten are compulsory and two elective. Student should have completed a statistical bridging course before enrolment into the programme.

PRE-COURSE

- Statistical bridging course: Calculus and Matrix algebra (0 credits)

COMPULSORY MODULES

- Mathematical Statistics (6 credits)
- Principles of Statistical Inference (12 credits)
- Linear models (6 credits)
- Fundamentals of Epidemiology (6 credits)
- Data Management and Statistical Computing (6 credits)
- Categorical Data Analysis & Generalized Linear Models (12 credits)
- Analysis of Survival Data (12 credits)
- Analysis of Observational Data: Causal Inference (12 credits)
- Longitudinal Data Analysis (12 credits)
- Biostatistical Consulting (12 credits)

OPTIONAL MODULES - SELECT 2

- Bayesian Data Analysis (12 credits)
- Clinical Biostatistics (12 credits)*
- Design and Analysis of Clinical Trials (12 credits)
- Multivariate Statistics (12 credits)*
- Bioinformatics (12 credits)*

Duration of a module: Semester (February – June or July – November) * not currently on offer

05 BURSARIES

The DELTAS Sub-Saharan Africa Consortium for Advanced Biostatistics Training (S2ACABT) is offering scholarships to develop and improve biostatistical skills among researchers, with an ultimate goal of creating research nodes of excellence to grow the discipline and a biostatistical network to nurture researchers with advanced skills and expertise. Bursaries are allocated on a competitive basis, based on academic merit and financial need. For more information on the application procedure contact the MSc Biostatistics programme coordinator/administrator.

The ACBE D43 Fogarty is offering scholarships to develop and improve biostatistical skills among researchers, with an ultimate goal of creating research nodes of excellence to grow the discipline and a biostatistical network to nurture researchers with advanced skills and expertise. Bursaries are allocated on a competitive basis, based on academic merit and financial need. For more information on the application procedure contact the MSc Biostatistics programme coordinator/administrator.

General bursaries: do look in the general University bursary booklet

04 RESEARCH ASSIGNMENT (60 CREDITS)

The aim of this module is that the student gains practical experience in a workplace setting, in the application of knowledge and skills learnt during the coursework of the MSc of Biostatistics program, under the supervision of an experienced biostatistician. The student will have to provide evidence having achieved this goal by producing and presenting two reports:

- Masters research assignment (contributes 70%)
- Summary of projects/tasks in a workplace environment (contributes 30%)

Progression to the research assignment module is subject to academic performance. The MSc in Biostatistics programme committee will consider the study results of first-year and second-year first semester, and the committee has the right to prevent a student from working on the research assignment module.

Please read in detail the *Practical guidelines for the research assignment*

SUBMISSION PROCESS

Two copies must be submitted, in temporary binding to the MSc Biostatistics Programme Coordinator. The submission deadline for December graduation is 1 September, and for March graduation it is 1 December.

NOTE: The programme coordinator must be informed 3 months in advance of the intention to hand in the research assignment.

We encourage MSc Biostatistics students to submit their research assignments for publication. In doing so, do be reminded that the Division of Epidemiology and Biostatistics, Stellenbosch University, should be stated as an affiliation. If you have more than one affiliation, the Division should be stated as a secondary affiliation.

Kindly also acknowledge MSc Biostatistics as follows – This research assignment has been conducted as part of the academic requirements of the MSc in Biostatistics, Stellenbosch University.

The Faculty offers incentives to those who publish (for further info, please contact Tashwell de Wet (tashwell@sun.ac.za)).

WHEN ARE CORE MODULES OFFERED?

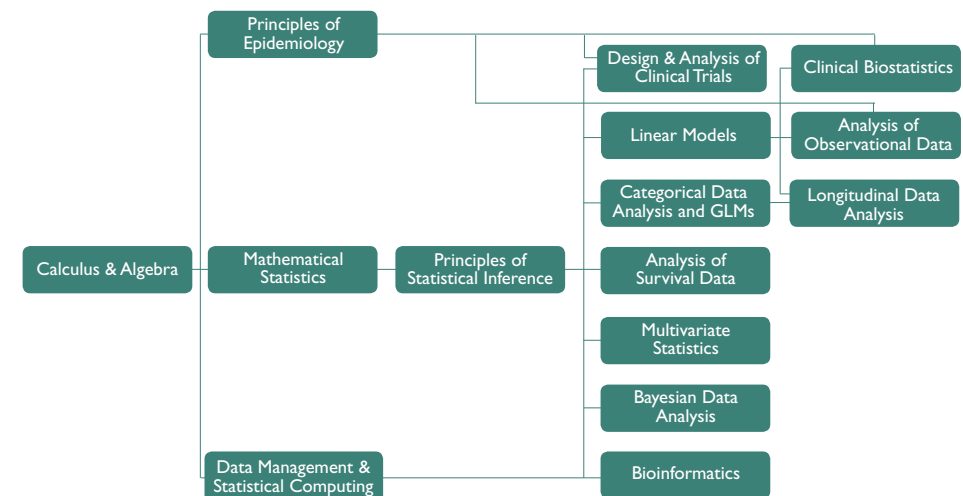
YEAR	Semester 1	Semester 2
1	Mathematical Statistics (6 credits)	Categorical Data Analysis & Generalized Linear Models (12 credits)
	Principles of Statistical Inference (12 credits)	Analysis of Survival Data (12 credits)
	Data Management & Statistical Computing (6 credits)	Analysis of Observational Data: Causal Inference (12 credits)
	Fundamentals of Epidemiology (6 credits)	
	Linear Models (6 credits)	
2	Longitudinal Data Analysis (12 credits)	
	Biostatistical Consulting (12 credits)	

WHEN ARE OPTIONAL MODULES OFFERED?

YEAR 2	
Design and Analysis of Clinical Trials	Clinical Biostatistics*
Bayesian Data Analysis	Multivariate Statistics*
Bioinformatics*	

* not currently on offer

SEQUENCE OF MODULES



A student may not register for a module if s/he has not passed the pre-requisite module(s). To register for the biostatistics consulting module, one should have completed course work in year 1 and passed the assessments.

MODULE STRUCTURE

Modules are offered using a combination of face-to-face teaching and e-learning using Sun Learn, Stellenbosch University's online learning environment. Typically a module (except Biostatistics consulting) consists of:

- 48 hours classroom time for 12 credit module or 24 hours for 6 credit module
- 72 hours (12 credit module) or 24 hours (6 credit module) self-study: reading / formal assignments / projects

MODULE ASSESSMENT

2-3 formative assessments and summative assessment. Continuous and summative assessment of modules will be conducted through written examinations, oral presentations, written assignments and participation in discussions. A pass mark of 50% is required for each module. The student will be required to participate successfully and to integrate knowledge in projects, reports and assignments. An external examiner is appointed for every module. A candidate who fails any module may be denied the right to reregister for the programme.

ATTENDANCE:

Students should inform the module convener if they are going to be absent for more than one session in a block or in the semester.

- Students missing sessions must make their own arrangements to obtain material they have missed.
- Students should ensure that the examination weeks are kept free of any competing engagements.
- Semester timetables should be consulted well in advance.

COMMUNICATION:

Students should ensure that the programme administrator has all their contact details, including any change in email address. Communication will take place using Sun Learn and email.

SHORT COURSES:

Some of the modules will be offered as short courses or on block release depending on availability of visiting professors/collaborators. Should a prospective full degree student have completed a short course offered by the programme, the student can apply for recognition of prior learning when entering the full degree programme.

RESEARCH ASSIGNMENT:

- The submission must be aligned to the prescribed format and will be assessed by both internal and external examiners.
- Students are expected to familiarise themselves with the University Ethics and research integrity guidelines and procedures.

GENERAL:

All students should consult Part I of the University Calendar for general information related to studying with Stellenbosch University such as the code of conduct guiding the relationship between the supervisor and student, responsibilities of students, supervisors, etc.

<http://www.sun.ac.za/university/yearbook/>

3.16 BIOINFORMATICS

Conveners: TBA

Requirements: Principles of Statistical Inference, Data Management & Statistical Computing

OBJECTIVES

- Understand basic concepts in bioinformatics
- Understand fundamental methods used in the discipline
- Be familiar with algorithms used in the analysis of genetic sequence
- Be capable of performing simple sequence analyses using existing tools
- Be familiar with key computational techniques needed
- Use key bioinformatics tools to analyse data from molecular biology
- Describe content and use of important bioinformatics databases and web portals

CONTENT

- Introduction to Bioinformatics
- Analyze DNA and protein sequences
- Multiple testing
- Linkage Analysis
- Association Analysis
- Gene Expression
- Pathway Enrichment

3.15 MULTIVARIATE STATISTICS

Conveners: TBA

Requirements: Principles of Statistical Inference

OBJECTIVES

After completion of the module the student will be able to:

- Explore and summarize multivariate data using graphical and numerical techniques
- Describe properties of multivariate distributions
- Demonstrate an understanding of limitations of some multivariate techniques
- Identify appropriate multivariate techniques to analyse multivariate data
- Perform multivariate analyses using R or Stata
- Explain and Interpret results from multivariate analyses

CONTENT

- Aspects of multivariate data (data organization & display, distance)
- Matrix algebra (positive definite matrices, mean vector and covariance matrix)
- Multivariate normal distribution
- Inference on mean vector
- MANOVA
- Multivariate Linear Regression
- Principal component analysis and PCA bi-plots
- Simple and Multiple Correspondence Analysis
- Multidimensional scaling
- Cluster Analysis
- Discriminant Analysis, Canonical Variate Analysis and Analysis of Distance

02 ADMISSION REQUIREMENTS

To be admitted into the MSc (Biostatistics) programme, the candidate should

1. Have completed the statistical bridging course module (Matrix algebra and Calculus) with a mark of at least 65%
2. Have average mark of at least 65% for the last qualification obtained.
3. Hold the National Senior Certificate with Mathematics (or equivalent) and one of the following:
 - an honours degree in Mathematics or Statistics
 - an honours degree in a biomedical sciences area
 - a medical degree (MBChB)
 - another qualification deemed adequate by Senate

Application by international students will be reviewed for equivalence of degree.

APPLICATION PROCEDURES:

Applications should include:

- Completed application form
- CV
- Letter of motivation
- Academic Record
- Matric certificate

03 STRUCTURED MODULES: OBJECTIVES & CONTENTS

3.1 STATISTICAL BRIDGING COURSE (CALCULUS AND MATRIX ALGEBRA)

Convener: TBA, Division of Epidemiology and Biostatistics
Requirements: Course entry requirements number 2 and 3

PART 1: CALCULUS

OBJECTIVES

After completion of the module the student will be able to:

- Apply the binomial theorem on any identified problem
- Find derivatives of functions using limits
- Find one-sided limits, limits, infinite limits, and limits at infinity
- Find derivatives by applying addition rule, product rule, quotient rule, chain rule, power rule, and rules for trigonometric functions
- Use implicit differentiation
- Demonstrate understanding of limits, continuity, derivative, anti-derivative, and definite integral
- Find anti-derivative of a function with one or two variables
- Use Taylor and MacLaurin series to represent functions

CONTENT

- Induction and binomial theorem
- Functions, limits and continuity
- Differentiation and applications
- Partial derivatives
- Integrals (single and multiple variable)
- Taylor series

PART 2: MATRIX ALGEBRA

OBJECTIVES

After completion of the module the student will be able to

- Perform basic matrix operations
- Compute determinant, Eigen values and Eigen vectors
- Solve matrix algebra problems

CONTENT

- Matrix operations
- Determinants
- Rank of a matrix
- Inverse and generalized inverses
- Eigen values and Eigen vectors
- Positive definite matrices
- QR factorization

3.14 DESIGN AND ANALYSIS OF CLINICAL TRIALS

Conveners: Carl Lombard, Division of Epidemiology and Biostatistics
Moleen Dzikiti, Division of Epidemiology and Biostatistics
Requirements: Principles of Statistical Inference, Principles of Epidemiology

OBJECTIVES

After completion of the module the student will be able to:

- Identify the benefits of randomisation as a mechanism for reduction and balancing bias, and implement a variety of randomisation schemes
- Demonstrate knowledge of the principles behind the common experimental designs and be able to implement, analyse and interpret data from a variety of randomised designs (superiority, non-inferiority, equivalence, cross-over, cluster, factorial, stepped wedge, parallel and quasi-random trials)
- Demonstrate an understanding of the principles underlying Phase I, II, III and IV studies (objectives, patients)
- appreciation of the scientific basis underlying issues in clinical studies including intention-to-treat, blinding, interim analyses, subgroup analyses and the reporting thereof
- Formulate an interim analysis plan for randomised trial and perform the corresponding specialized analyses
- Critically review reports/scientific articles on results of a trial

CONTENT

- Principles and methods of randomisation in controlled trials; treatment allocation, blocking, stratification and allocation concealment
- Parallel, factorial, cluster, stepped-wedge, equivalence, non-inferiority and crossover designs including n-of-1 studies
- Practical issues in sample size determination
- Intention-to-treat principle
- Phase I dose finding studies; phase II safety and efficacy studies and phase III and IV studies
- Interim analysis and early stopping
- Multiple outcomes/endpoints, multiple tests and subgroup analyses, including adjustment of significance levels and P-values
- Reporting trial results and use of the CONSORT statement

3.13 CLINICAL BIostatISTICS

Conveners: TBA

Requirements: Principles of Epidemiology, Linear Models

OBJECTIVES

After completion of the module the student will be able to:

- Demonstrate a broad understanding of statistical methods used in evidence-based health care
- Determine appropriate statistical methods of particular relevance to evidence-based health care
- Correctly apply these statistical methods
- Effectively communicate with clinicians on the application of statistical methods and interpretation of results

CONTENT

- Clinical agreement (kappa statistics, Bland-Altman agreement method, intraclass correlation)
- Diagnostic tests (sensitivity, specificity, predictive values, ROC curves, likelihood ratio, diagnostic odds ratio, Youden's index)
- Statistical process control (special and common causes of variation, Shewhart, CUSUM and EMWA charts)
- Systematic reviews (process, estimating treatment effect, assessing heterogeneity, publication bias, meta-analysis)

3.2 MATHEMATICAL STATISTICS

Convener: Merga Feyasa, Division of Epidemiology and

Biostatistics Requirements: Course entry requirements

OBJECTIVES

After completion of the module the student will be able to:

- Describe the concept of probability
- Understand and distinguish random variables, discrete and continuous distributions
- Apply calculus concepts to obtain expressions for parameters of distributions.
- Understand multivariate random distributions in the context of joint distribution function, marginal and conditional distributions, independence, expected values, and correlation and covariance
- Apply techniques for determining distributions of transformations of random variables
- Derive the moment generating functions of a random variable and use it to find mean and variance
- Understand the concept of the sampling distribution and standard error of an estimator of a parameter is presented, together with key properties of estimators

CONTENT

- Probability, random variables, discrete and continuous distributions, and the use of calculus to obtain expressions for parameters of these distributions such as the mean and variance (Bernoulli, binomial, Poisson, Geometric, Hypergeometric, Multinomial, negative binomial, Gaussian, Exponential, uniform, Gamma, Weibull, Chi-squared, Beta, Student's T, F-distribution)
- Joint distributions for multiple random variables are introduced together with the important concepts of independence, correlation and covariance, marginal and conditional distributions
- Techniques for determining distributions of transformations of random variables
- The concept of the sampling distribution and standard error of an estimator of a parameter is presented, together with key properties of estimators
- Large sample results concerning the properties of estimators are presented with emphasis on the central role of the normal distribution in these results
- Numerical simulation and graphing with Stata/R are used throughout to demonstrate concepts

3.3 LINEAR MODELS

Convener: Elphas Okango, Division of Epidemiology and Biostatistics
Requirements: Mathematical Statistics, Principles of Statistical Inference

OBJECTIVES

After completion of the module the student will be able to:

- Apply different linear models to biostatistical data analysis, with proper attention to underlying assumptions
- Perform model diagnosis and apply appropriate remedial measures
- Comprehend model building and validation process
- Interpret and communicate the results of linear models
- Use statistical software (R and Stata) for data analysis

CONTENT

- Simple linear regression
- Estimation of the regression parameters: least squares methods
- Regression models and related statistical inference
- Analysis of covariance to adjust for confounding
- Multiple regression with matrix algebra
- Model construction, validation and interpretation (use of dummy variables, parameterisation, interactions and transformations)
- Model checking, diagnostics and remedial measures
- Regression to the mean & handling of baseline values
- Flexible nonparametric regression
- The analysis of variance
- Multiple comparison techniques for cell means and contrasts
- Estimable functions
- Type I through IV sum of squares

3.12 BAYESIAN DATA ANALYSIS

Convener: Christel Faes, Hasselt University
Requirements: Principles of Statistical Inference

OBJECTIVES

After completion of the module the student will be able to:

- Demonstrate an understanding of Bayesian reasoning and inference
- Distinguish different prior knowledge and consequences of using each of them
- Illustrate the use of conjugate priors in Bayesian estimation
- Use WINBUGS and/or R2WINBUGS software for Bayesian analysis
- Use probability models to quantify uncertainty in statistical conclusions
- Perform practical Bayesian analysis relating to health research problems
- Summarise Bayesian analysis into a statistical report

CONTENT

- Sampling techniques to explore univariate posterior distributions
- Exploring prior distributions (conjugate prior distribution, non-informative priors, informative priors)
- One parameter models with conjugate prior distributions
- Multi-parameter inference (posterior distribution, Method of composition)
- The relationship between Bayesian methods and standard classical approaches to statistics (likelihood methods)
- Computational techniques (MCMC) for use in Bayesian analysis – use of simulation from posterior distributions using WinBUGS / R2WINBUGS
- Application of Bayesian methods to fitting hierarchical models to complex data structures
- Model building and Assessment in the Bayesian framework

3.1.1 BIOSTATISTICAL CONSULTING

*Conveners: Tonya Esterhuizen, Division of Epidemiology and Biostatistics
Lehana Thabane, McMaster University*

Requirements: All first year modules

OBJECTIVES

After completion of the module the student will be able to:

- Integrate statistical knowledge gained in the courses and apply them during biostatistical consulting sessions with researchers
- Communicate effectively with researchers to gather information required to make a link between research questions and statistical methods by asking relevant questions
- Provide guidance in study design and statistical analysis plans and communicate effectively with other researchers
- Research the statistical methodologies that are useful to researchers
- Communicate results in a clear manner in the form of a written report to researchers
- Provide leadership in research studies

CONTENT

- Communicate effectively about both statistical and biological concepts
- Develop data analysis plan
- Verbal and written skills will be developed through frequent oral presentations at meetings or journal clubs, seminars, drafting review papers on selected topics,
- Attending formal institutional workshops or courses on effective communication
- The workplace attachment will provide an opportunity for the student to develop skills learnt and also apply the theoretical knowledge gained during didactic classes

3.4 FUNDAMENTALS OF EPIDEMIOLOGY

Convener: Taryn Young, Division of Epidemiology and Biostatistics

Requirements: Course entry requirements

OBJECTIVES

After completion of the module the student will be able to:

- The history and development of clinical epidemiology
- How to frame research questions
- The principles, strengths and limitations of various study designs
- The different data sources
- How to make use of measures of disease occurrence, measures of effect and measures of public health impact
- The concepts of random error, bias, confounding and effect modification in epidemiological studies and know of strategies to deal with these challenges
- How to determine a causal link between exposure and outcome
- Epidemiological concepts related to infectious diseases, occupational health and chronic diseases

CONTENT

After completion of the module the student will be able to:

- History and contribution of epidemiology
- Development of clinical epidemiology
- Framing research questions
- Strengths and limitations of various research designs
- Cross sectional studies
- Case-control studies
- Cohort-studies
- Intervention studies
- Ecological studies
- Data sources (Disease notification systems, death and birth registration, census, health surveys)
- Early detection: screening
- Measuring disease occurrence (prevalence, incidence risk, incidence rates)
- Linking exposure and disease - measures of effect/association (risk difference, rate difference, risk ratio, rate ratio, odds ratio)
- Sensitivity, specificity, Positive Predictive Value (PPV) and Negative Predictive Value (NPV)
- The problem of error (random and systematic error)
- The challenge of confounding and effect modification
- External validity
- Association and causation
- Epidemiological concepts related to infectious diseases, occupational health and chronic diseases

3.5 DATA MANAGEMENT & STATISTICAL COMPUTING

Convener: Kamunhwala Gausi, Division of Epidemiology and Biostatistics
Requirements: Course entry requirements

OBJECTIVES

After completion of the module the student will be able to:

- Design case report forms
- Design relational database in MS Access
- Extract data from a relational database using SQL
- Manipulate and manage data using two major statistical software packages (Stata and R)
- Display and summarise data using statistical software
- Perform data verification and validation
- Link files through use of unique and non-unique identifiers
- Demonstrate fundamental programming skills for efficient use of software packages
- Understand key principles regarding confidentiality and privacy in data storage, management and analysis

CONTENT

- Case report forms: Principles for developing case report forms
- MS Access: Creating a relational database using data normalization rules, (creating tables, forms, reports and queries)
- Stata and R: The basics (importing and exporting data, recording data, formatting data, labelling variable names and data values; using dates; data display and summary presentation)
- Stata and R: graphs, data management and statistical quality assurance methods (including advanced graphics to produce publication-quality graphs)
- Data management using Stata and R (using functions to generate new variables, appending, merging, transposing longitudinal data; programming skills for efficient and reproducible use of these packages, including loops and arguments)
- Programming: Writing functions, macros and do files

3.10 LONGITUDINAL DATA ANALYSIS

Conveners: TBA, Division of Epidemiology and Biostatistics
Requirements: Linear Models, Categorical Data Analysis and Generalized Linear Models

OBJECTIVES

After completion of the module the student will be able to:

- Identify repeated/dependent/clustered data and understand the weakness of traditional methods when applied to these.
- Apply graphical techniques to explore repeated/dependent/clustered data
- Describe the statistical methods utilized to analyse continuous and non-continuous longitudinal data in a variety of settings
- Demonstrate an understanding of the theory behind modern methods for analysis of longitudinal/clustered data
- Analyse a scientific problem that requires repeated measurements, identify appropriate design, and identify the statistical methods required to analyse the data
- Identify and describe missing data and apply appropriate methods of analysis in the presence of missing data
- Utilize Stata and/or R procedures to perform analyses of longitudinal data
- Apply modern methods for the analysis of longitudinal data to a range of settings encountered in biomedical and public health research
- Prepare a scientific report describing methods used for analysis, results obtained and their interpretation
- Communicate methods used and the clinical/scientific meaning of the results from a longitudinal data analysis and defending their analysis

CONTENT

- The effect of non-independence on comparisons within and between clusters of observations
- Limitations of classical analysis methods
- Exploratory analysis of marginal distribution (average evolution, variance structure and correlation structure)
- Methods for continuous outcomes: Linear mixed effects (hierarchical or multilevel) models)
- Methods for discrete data: GEE and generalized linear mixed models (GLMM)
- Nonlinear Mixed Effect Models
- Missing Data (Missing completely at Random, Missing at Random, Not Missing at Random, Weighted GEE, Multiple Imputation)
- Joint Modelling of longitudinal and time-to-event data

3.9 ANALYSIS OF OBSERVATIONAL DATA

*Conveners: Rhoderick Machekano, Division of Epidemiology and Biostatistics
Moleen Dzikiti, Division of Epidemiology and Biostatistics*

Requirements: Principles of Epidemiology, Linear Models

OBJECTIVES

After completion of the module the student will be able to:

- Critically evaluate the pitfalls of observational studies
- Demonstrate in-depth understanding of bias and able to identify potential sources of bias
- Understand of the role and potential of different methodological approaches to overcome bias
- Use analytical tools and apply statistical models to estimate causal effects of interventions, treatments or exposures of interest from observational studies
- Critically review research that claims to estimate causal effects with non-experimental data

CONTENT

- Fundamental problem of causal inference – potential outcomes
- Randomization and estimation of causal effects
- Sources of bias in observational studies
- Graphical representation of causal effects
- Approaches to causal inference
 - Subclassification, matching
 - Standardization
 - Propensity scores (matching, inverse weighting)
 - Instrumental variables
 - Marginal structural models
 - G-estimation

3.6 PRINCIPLES OF STATISTICAL INFERENCE

Convener: Innocent Maposa, Divisions of Epidemiology and Biostatistics

Requirements: Mathematical Statistics

OBJECTIVES

After completion of the module the student will be able to:

- Explain fundamental concepts in statistical inference and their practical interpretation and importance in biostatistical contexts
- Apply the theoretical basis for frequentists and Bayesian approaches to statistical inference
- Develop and apply parametric methods of inference, with particular reference to problems of relevance in biostatistical contexts
- Demonstrate an understanding of the theoretical basis to justify more complex statistical procedures
- Explain the basic alternatives to standard likelihood-based methods, and be able to identify situations in which these methods are useful
- Describe general approaches to obtaining estimators of parameters
- Demonstrate an understating of the theory of sufficient statistics
- Set up confidence interval for parameters and derive sample size
- Understanding Neyman-Pearson theory for hypothesis testing and use it to derive tests based on the generalized likelihood ratio statistic

CONTENT

- Core concepts of statistical inference, beginning with estimators, confidence intervals, type I and II errors and p-values
- The emphasis will be on the practical interpretation of these concepts in biostatistical contexts, including an emphasis on the difference between statistical and practical significance
- Classical estimation theory, bias and efficiency
- Likelihood function, likelihood based methodology, maximum likelihood estimation and inference based on likelihood ratio, Wald and score test procedures
- Bayesian approach to statistical inference vs classical frequentist approach
- Nonparametric procedures, exact inference and resampling based methodology (Jackknife and bootstrap)

3.7 CATEGORICAL DATA AND GENERALIZED LINEAR MODELS

Conveners: Jim Todd, London School of Hygiene and Tropical Medicine

Requirements: Principles of Statistical Inference

OBJECTIVES

After completion of the module the student will be able to:

- Be familiar with descriptive and inferential methods for contingency tables
- Use generalized linear models (GLMs) and other methods to analyse categorical data with proper attention to the underlying assumptions.
- Present results clearly and accurately in a structured report, such as might form the basis of a report by a statistical consultant
- Interpret and communicate results to researchers

CONTENT

- Conventional methods for contingency tables especially in epidemiology (odds ratios and relative risks, chi-squared tests for independence, Mantel-Haenszel methods for stratified tables, and methods for paired data)
- Generalized linear models (GLMs) (Components, parameter estimation, logistic model, logit model and probit model)
- Hypothesis testing using GLMs
- Simple, multiple and exact logistic regression models
- Model building in multiple logistic regression
- Methods for assessing model adequacy (residual analysis, likelihood ratio statistic, score statistic, Wald statistic, deviance statistic, Pearson chi-square statistic, Pearson residuals and deviance residuals, over-dispersion)
- Nominal and ordinal logistic regression for categorical response variables with more than two categories (Baseline-category logit model, Cumulative logits model, Proportional odds model, Adjacent-category logit model)
- Models for count data (Poisson regression and log-linear models)

3.8 ANALYSIS OF SURVIVAL DATA

Conveners: TBA

Requirements: Principles of Statistical Inference

OBJECTIVES

After completion of the module the student will be able to:

- Know principles and methodologies of time to event analysis with proper attention to underlying assumptions
- Apply non-parametric methods commonly used in survival analysis
- Formulate parametric and semi-parametric models for survival outcomes
- Interpret estimates and perform hypothesis tests obtained from regression models for survival outcomes and communicate the results.
- Perform goodness of fit analyses
- Analyse data with time-varying covariates

CONTENT

- Censoring mechanisms (Type 1 right censoring, type 2 right censoring, interval censoring and left censoring)
- Survival likelihood
- Kaplan-Meier method and its variance
- Confidence interval for survival function
- Hypothesis testing in nonparametric setting (log-rank test, test for trend)
- Cox Proportional Hazards model (assumptions, model building, diagnostic techniques, checking proportional odds assumption)
- Parametric survival models in the proportional hazards metric (Exponential, Weibull)
- Parametric survival models in the accelerated failure time metric
- The extended Cox model, interactions and time-varying covariates
- Joint modelling/Multivariate/Clustered survival data
- Frailty models
- Sample size calculations for survival studies