YENEPOYA

(DEEMED TO BE UNIVERSITY)
Recognized under Sec 3(A) of the UGC Act 1956
Accredited by NAAC with 'A' Grade

YENEPOYA (DEEMED TO BE UNIVERSITY)

Deralakatte, Mangaluru -575018

REGULATIONS AND CURRICULUM GOVERNING

POSTGRADUATE PROGRAM

MASTER OF SCIENCE BIOSTATISTICS

(CURRICULUM - EFFECTIVE FROM 2020-21)

ATTESTED

Dr.Gangadhara Somayaji K.S.
Registrar
Yenepoya(Deemed to be University)
University Road, Deralakatte
Mangalore-575 018, Karnataka

YENEPOYA
(DEEMED TO BE UNIVERSITY)
Recognized under Sec 3(A) of the UGC Act 1956
Accredited by NAAC with 'A' Grade

Office of the Registrar
University Road
Deralakatte
Mangalore - 575018
Ph: 0824-2204667/68/69/71
Fax: 0824-2203943

Ref: No. Y/REG/ACA/38-ACM/2020

14.05.2020

NOTIFICATION – 38-ACM/11/2020 dtd.06.05.2020

Sub: Proposal to start M.Sc. Statistics and M.Sc. Biostatistics courses

Ref: Resolution of the Academic Council at its 38th meeting held on 27.04.2020, vide agenda -22

******

The Academic Council at its 38th meeting held on 27.04.2020 & subsequently the Board of Management at its 49th meeting held on 30.04.2020 have resolved to approve the proposal to start 2 year courses of M.Sc. Statistics and M.Sc. Biostatistics.

The curriculum of M.Sc. Statistics and M.Sc. Biostatistics courses as drafted and finalized by the Board of Studies, Statistics has been approved.

This notification is issued for implementation with effect from the academic year 2020-2021.

REGISTRAR

To:

Dr. Ismail B., Professor of Statistics

Copy to:

1. Controller of Examinations
2. File copy

REGULATIONS, SCHEME AND SYLLABUS UNDER
CHOICE BASED CREDIT SYSTEM (CBCS)

Post Graduate Program: **M.Sc., in Biostatistics**

Department: **Department of Postgraduate Studies and Research in Statistics**

1. **Preamble:**

The goal of this program is to train students in statistical methods, as applied to the problems in the biomedical and health sciences. During the program, students learn the basic principles of biostatistics, data analysis and specific computing applications. The teaching and practical training on various courses under M.Sc. Biostatistics program enable students to acquire knowledge in the subject and develop practical skills to contribute effectively in the fields of data analysis, academics and research.

The role of Statistics as the technology of data analysis and decision making under uncertainty has expanded vastly in the past few decades. Statistical expertise has been central to many areas of research and health service activities including not only traditional areas such as clinical trials and epidemiological studies, but also the rapidly emerging fields of statistical genetics, bioinformatics and computational biology. It's descriptive and inferential roles not only formulate the basis of growth of almost all the disciplines of the contemporary world but also provide an array of employment avenues in industry, medical and pharmaceutical research, academia , government and R&D organizations. Almost all organisations including medical, agricultural, Pharmaceutical industries and Computer software companies are looking for Data Scientists particularly Statisticians who can effectively carry out the scientific analysis of their data.

2. Scope & need for starting the Course

There has been a clearly increasing recognition of the contribution of Bio statistical expertise to an ever expanding range of areas in medical, genetic, and health science research. By developing new quantitative methods, characterizing underlying theory, and making innovative applications to scientific problems, biostatisticians play a leading role in the advancement of biomedical, pharmaceutical and public health research. Demand for expertise in biostatistics continues to be high in academia and in public and private sectors emphasizing health science research. There is a widely recognised shortage of graduates in biostatistics relative to what appears to be a continually increasing demand for expertise in this area.

The theoretical foundations, design guidelines and computational skills associated with different fields of Biostatistics help the students to pursue research in advanced areas of **Biostatistics**.

In view of the present trend and demand for Bio-Statistics graduates, YENEPOYA (Deemed to be University) decided to start the Department of Statistics at YENEPOYA (Deemed to be University) and offer postgraduate programme in **Biostatistics.** The Department of Statistics shall be an independent Post Graduate Department in the University involving teaching, research and consultancy services.

3. For whom is it meant

Biostatistics is meant for under graduate candidates who are interested to work as Data Scientists in Pharmaceutical Industries, Medical and Agricultural Research Institutions and State and Central Government organizations. Biostatistics deals with the application of statistical techniques to scientific research in health-related fields, including medicine, dentistry, genetics, nursing and public health, and development of new tools to study these areas.

4. Eligibility

Eligibility for admission:

The following candidates are eligible for M.Sc. Biostatistics program:

1. B.Sc. Statistics or B.Sc. Mathematics with Statistics as subsidiary subject of any University recognized by UGC and have secured a minimum of 50% marks in aggregate.
2. BE or B. Tech.
3. B.Sc. with Computer Science as one of the subject and have secured a minimum of 50% marks in aggregate.

Selection of candidates shall be on merit basis (Overall percentage of marks in the qualifying examination excluding languages).

5. Objectives of the Course/Program:

Main objective: To produce quality human resource in Biostatistics and evolve as Centre for Advanced studies in Biostatistics.

The goal of this program is to train students in statistical methods, as applied to the problems in the biomedical and health sciences.

- Specific learning objectives and Learning outcome are given for each Course in Appendix I.

Students who successfully complete this program will be:

- Exposed to broad base of practical techniques to work effectively in biomedical industry and academic settings.
- Effectively use biostatistician methods in biomedical and health science problems for data collection related to variable/variables, data management, designs, analysis and interpretation.
- Identify distributions relating to the variable/variables under study.
- Provide consultancy, carryout research and resource building in Biostatistics.
- Acquire skills in data display, summary presentation and pattern recognition using statistical packages like SPSS, STATA and R.
- The theoretical foundations, design guidelines and computational skills associated with different fields of Biostatistics help the students to pursue research in advanced areas of Biostatistics.

6. Duration of the Course: including hours of theory/practical/clinical/hands-on/community visits

Duration: Two years with four semesters.

{3 Semester regular classes and one semester exclusively for project/internship}

Course pattern and Credits: Choice based credit System (CBCS)

Total credits:	94 credits
Hard Core courses:	55 credits
Soft Core/ Discipline specific electives:	09 Credits
Open elective papers-	06 Credits
Project with Internship	24 Credits

1 credit=1 hour of lecture per week/ 2 hours of Laboratory or practical

Types of Courses

i. **Core course**: The programme of study will have "Core" and "Elective" courses. The Core course will further consist of "Hard" and "Soft" Core courses. Hard core courses will have 4 credits while soft courses will have 3 credits. Hard core is a course that should compulsorily be studied by a candidate whereas there can be choice between Soft Core courses.

ii. **Open Elective**: Generally a course which can be chosen from a pool of courses and which may be very specific or specialized or advanced or supportive to the discipline/ subject of study or which provides an extended scope or which enables an exposure to some other discipline or subject or domain or nurtures the candidates proficiency skill.

- Two open elective courses of 3 credits each shall be offered in the second and third semesters only. Open elective will be chosen from other program within the faculty or across the faculty.

- The list of open elective courses offered shall be displayed in the website.
- A student shall not take the courses offered by the department in which she/he is enrolled.
- Registration for the open elective courses shall be at least one week prior to the commencement of the course with the CBCS coordinator.

Table: 1. Course, Scheme of instruction and examination for Semester-wise:

Course code	Type of Course	Course name	Hrs/ Week L:T:P	Exam (hours)	IA Marks	End Semester Marks	Max marks	Credits
First Semester								
FS02BT-1C1	Hard Core-Theory	Mathematical Foundation	3:1:0	3	40	60	100	4
FS02BT-1C2	Hard Core-Theory	Probability and Probability Distributions	3:1:0	3	40	60	100	4
FS02BT-1C3	Hard Core-Theory	Statistical Inference	4:0:3	3	40	60	100	4
FS02BT-1C4	Hard Core-Theory	Nonparametric methods	4:0:3	3	40	60	100	4
FS02BT-1P1	Hard Core-Practical	Practical I: Data Management and Statistical Computing -1	1:0:5	3	40	60	100	3
FS02BT-1P2	Hard Core-Practical	Practical II: Based on FS02BT-1C3 & FS02BT-1C4	0:0:6	3	40	60	100	3
Total Credits								22
Second Semester								
FS02BT-2E1	Open Elective	Statistical Methods	3:0:0	3	40	60	100	3
FS02BT-2C1	Hard Core-Theory	Sampling methods and Design of Experiments	4:0:4	3	40	60	100	4
FS02BT-2C2	Hard Core-Theory	Epidemiology	4:0:2	3	40	60	100	4
FS02BT-2C3	Hard Core-Theory	Applied Regression Analysis	4:0:4	3	40	60	100	4
FS02BT-2C4. 1	Soft Core Theory	a. Categorical Data Analysis and Bayesian Methods	3:0:2	3	40	60	100	3
FS02BT-2C4.2		b. Demography and Vital Statistics						

FS02BT-2P1	Hard Core-Practical l	Practical III: FS02BT-2C2 & FSO2BT-2C3	0:0:6	3	40	60	100	3
FS02BT-2P2	Hard Core-Practical l	Practical IV: Data management and Statistical Computing II (based on FS02BT-2C1 & FS02BT-2C4.1/FS02BT-2C4.2)	0:0:6	3	40	60	100	3
Total Credits								24
Third Semester								
FS02BT-3E1	Open Elective	Basic Data Analysis Techniques.	3:0:0	3	40	60	100	3
FS02BT-3C1	Hard Core-Theory	Survival Analysis	4:0:2	4	40	60	100	4
FS02BT-3C2	Hard Core-Theory	Methods in Clinical Trials	4:0:0	4	40	60	100	4
FS02BT-3C3	Hard Core Theory	Hierarchical linear and Hierarchical Generalized linear Models.	4:0:2	4	40	60	100	4
FS02BT-3C4.1	Soft Core Theory	a. Genetic epidemiology and Bioinformatics	3:0:1	3	40	60	100	3
FS02BT-3C4.2		b. Psychometrics						
FS02BT-3C5.1	Soft Core-Theory	a. Time Series Analysis	3:0:1	3	40	60	100	3
FS02BT-3C5.2		b. Nonparametric regression						
FS02BT-3P1	Hard Core-Practical	Practical V: Data management and Statistical Computing III (based on all core theory papers)	0:0:6	3	40	60	100	3
Total Credits								24

Fourth Semester

Subject Code	Subject	Maximum Marks				Credits
		IA Marks	Viva	End Semester Marks	Total marks	
FS02BT-4C1	Project	150	50	200	400	24
Total					400	24

TOTAL CREDITS for the M.Sc. Program	94

7. Curriculum including syllabus/course contents Preamble for the syllabus:

The program consists of theoretical components on Mathematical foundation, Probability theory and Probability Distributions, Sampling theory and Design of Experiments, Statistical inference, Epidemiology, Regression Analysis, Categorical Data Analysis and Bayesian Methods/ Demography and Vital Statistics and Statistical Software I (Excel, SPSS and R) for the first year. The second year curriculum include study of Survival Analysis, Methods in Clinical Trials, Hierarchical linear and Hierarchical Generalized linear Models, Genetic epidemiology and Bioinformatics/ Psychometrics, Time Series Analysis/ Nonparametric regression, Statistical software II (R and STATA software) and Project work.

M. Sc. Bio-Statistics program is of minimum 94 credits (15 Hard Core Courses of 55 credits + 03 Soft Core Courses of 09 credits , two open elective Courses with 6 credits and Project work with 24 credits) spread over four semesters. The Choice Based Credit System to be implemented through this curriculum would allow students to develop a strong footing in the fundamentals and specialize in the disciplines of their liking and abilities.

The program emphasizes theory, methods and applications of statistics and it is structured to provide knowledge and skills in depth, necessary for the employability of students in biomedical industry, other organizations, as well as in academics. The program has some unique features such as good theoretical foundation in the first year, Full semester project work and computer training for statistical computations including standard software packages like SPSS, STATA and R.

The syllabus of the first year (two semesters) covers most of the Hard Core courses. In the second year, there are **Three** Hard core courses, **two** Soft Core courses, one practical and Project work with Internship. There are two open electives courses- one in 2nd Semester and One in 3rd Semester. The syllabus has been framed to have a good balance of theory, methods and applications of Biostatistics. It is possible for the students to study basic courses from other disciplines such as Public Health, Social work, Biosciences , Hospital Administration and Master of Physiotherapy under open elective courses.

The Fourth semester is only for Project Work with Internship and constitutes 24 credits which carry 400 marks. During this period the students will be posted in different health care research centers and pharmaceutical industries. Department facilitate the students through its contact with various organizations in identifying the suitable place for their project work. The candidate will undergo Internship for 8 weeks and carryout the Project work in the fourth semester. All the students need to submit a proposal, the thesis or dissertation work will be done under the

guidance of a faculty member or professional, must be submitted before the notified dates.

The aim of this project is to make the students understand various practical problems in data collection, data management, designs, analysis and interpretation.

The detailed syllabus on various courses in M.Sc. Biostatistics are given in Appendix I

8. Teaching/Learning modalities including online instructions, contact programmes etc.,

- The program consists of two year (Four Semesters) of full time study which includes successful completion of written exams and the Master's thesis. The thesis is based on the project work done in the fourth semester of the program. The course consists of combination of lectures, seminars and practical exercises. In practical sessions students are trained to analyse real life problems from bio medical, bioinformatics, Psychometrics and health science research. The first three semesters are regular classes and the Fourth Semester is exclusively for project work and Internship.
- Also students are trained to work directly using the Statistical software like SPSS/STATA & R
- Each theory and practical papers are 56 hours and 84 hours of teaching respectively.
- In the third semester students will be posted for statistical consultation.
- There will be regular journal presentation.

9. Evaluation:

Evaluation of the course shall be done on a continuous basis with continuous internal assessments (CIA) followed by university examination at the end of the semester (SEE) for each course. The weights for CIA shall be 40% and SEE shall be 60%.

There will be two sessional examinations followed by university exam (at the end of the semester) for each theory and practical papers. The University examination is for 60 marks of 3 hours duration. Each theory and practical course carry an internal assessment of 40 marks which is based on two sessional exams, one seminar/ one Journal paper, Case studies/Data Analysis and assignments.

a) Internal Evaluation:

Internal Assessment**:** Marks for internal assessment shall be awarded on the basis of seminars, journal paper presentations, tests, assignments etc. The assessment gives importance to continuous and comprehensive evaluation. Each theory and practical course carry an internal assessment of 40 marks which is based on two sessional exams, one seminar/ one Journal paper, Case studies/Data Analysis and assignments.

The split up is as follows:

Components of CIA	Details	Distribution of Marks
Sessional Tests	Average of the two tests	20
Seminar/ other related activities	One Seminar / One Journal paper relevant to the core courses	07
Creativity /Skill enhancing Exercise	Case studies / Data Analysis /outreach activities/other creative activities	07
Assignments	Discipline specific as required by the course	06

CIA for practical's (40 marks) shall be based on the performance in laboratory practical sessions class room problem solving during the semester (20 marks) and average score of two practical tests conducted (20 Marks).
The internal assessment marks shall be notified to the students before it is communicated to the Controller of Examinations, before the commencement of the University examinations

b) University Examinations:

There shall be Semester End Examination **(SEE)** at the end of each semester ordinarily during December/January for odd semesters and during June/July for even semesters. The SEE duration shall be three hours. The University examination is for 60 marks of 3 hours duration.

Pattern of question paper for semester end examination is as follows:

Sl. No	Key Criteria	No. Of questions to be answered and Marks	Max marks
1	Short answer type questions of 2 marks each (Answer 6 out of 8 questions)	6 X 2 marks	12
2	Long answer type questions of 12 marks each (Answer 4 out of 6 questions)	4x12 marks	48
TOTAL			**60**

Pattern for practical examination:

Sl. No	Key Criteria	No. Of questions and Marks	Max marks
1	Practical Record		**10**
2	Viva		**10**
2	Four questions from Six questions without leaving any core papers	**4x10**	**40**
TOTAL			**60**

Valuation of answer scripts:

a. Each theory examination shall be evaluated by one internal and one external examiner. Thereshall be a third evaluation if the difference is more than 15%.
b. Practical examination shall be jointly conducted and evaluated by one internal examiner and one external examiner.

Project evaluation:

Project: In-semester assessment is with a weight of 150 marks, and supervisor/s is/are responsible for the continuous assessment of candidate's performance regarding progress during internship, protocol presentation and execution of project. At the end of the fourth semester the students need to submit two copies of their project report in the format prescribed by the department. Also, the students have to submit a copy of their internship report along with the project report to the department.

End-semester examination with a weight of 250 marks is based on evaluation of report, presentation of report and viva-voce shall be jointly conducted and evaluated by one internal examiner and one external examiner.

A student with 'F' grade must re-register for fourth semester of the program.

c) Criteria for eligibility to appear for the University Exam

Candidates having minimum 80% attendance in each of the courses can only qualify to appear for the semester end examinations. The candidates shall register for all the courses of a semester when he/she appears for the examination of that semester for the first time

Student who has less than 80% shall not be permitted to appear for the Semester end examination in the course in which the short fall exists. The HOD/course coordinator through the Dean of Faculties shall announce the names of the students who will not be eligible to take the Semester End- Examinations in the various courses and send a copy of the same to the Controller of Examination's (CEO) office. Registrations of such students for those courses shall be treated as cancelled.

Re-Entry after Break of the study

a. Students admitted to a program abstaining for more than 3 months must seek readmission into the appropriate semester.
b. The student shall follow the syllabus in vogue (currently approved/is being followed) for the program
c. All re admissions of students are subject to the approval of the University.

Maximum period for completion of the Programme

A candidate shall complete the four semesters (two years) programme within five years from the date of admission

d) Criteria for Pass /I Class/Distinction:

i. A candidate shall be declared to have passed the PG program if he/she secures at least a CGPA of 5.0 (Course letter Grade P) in the aggregate of both internal assessment and semester end

examination marks.

ii. For each course the total of 100% is determined from the CIA evaluation and the SEE and aggregate of CIA and SEE at 50% as minimum for pass.

iii. The candidates who pass all the semester examinations in the first attempt in two years are eligible for ranks provided they secure at least a CGPA of 6.0 (at least letter Grade A).

iv. The results of the candidates who have passed the fourth semester examination but not passed the lower semester examinations shall be declared as NCL (Not Completed Lower semester examinations). Such candidates shall be eligible for the degree only after completion of all the lower semester examinations.

v. A candidate who passes the semester examinations in parts is eligible for only CGPA and letter Grade but not for ranking.

vi. Carry over shall be allowed for candidate who failed in not more than two courses in a semester.

vi. Candidate who fails in any of the Course /project work shall reappear in that Course/project work and pass the examination subsequently

Letter Grades

The results of successful candidates at the end of each semester shall be declared in terms of Grade Point Average (GPA) and alpha sign grade. The results at the end of the fourth semester shall be classified on the basis of the Cumulative Grade Point Average (CGPA) obtained in all the four semesters and the corresponding overall alpha-sign grade. The letter grade as described below shall be adopted.

CGPA Range	Programme alpha-sign grade
9.0-10.0	O+(Outstanding)
8.0 – 8.99	O(Excellent)
7.0 - 7.99	A+ (Very Good)
6.0 - 6.99	A(Good)
5.5 - 5.99	B^+ (Average)
5.0 – 5.49	B (Pass)
<5.0	F (Fail)

Calculation of Cumulative Grade Point Average (CGPA): It is a measure of overall cumulative performance of a student over all semester. The CGPA is the ratio of total credit points secured by a student in various courses in all semesters and the sum of all the total credits of all courses in all the semesters. It is expressed up to two decimals.

$$CGPA = \frac{\sum_{i=1}^{4} C_i S_i}{\sum_{i=1}^{4} C_i},$$

where S_i is the SGPA of the i^{th} Semester and C_i is the total number of credits in that semester.
The following is the sample illustration of computing semester grade point averages (GPA), cumulative grade point average (CGPA) and the letter grades assigned.

	Semester I	Semester II	Semester III	Semester IV
Credit (C_i)	20	18	20	24
SGPC (S_i)	8.23	7.31	6.95	8.2

CGPA= (20x8.23+ 18x7.31+ 20x6.95+
24x8.2)/82
= 631.98/82
= 7.71

Letter Grade: **A^+**

10. Attendance requirements

Each course (theory, practical, etc.) shall be treated as an independent unit for the purpose of attendance. A student shall attend a minimum of 80% of the total instruction hours in a course including tutorials and seminars in each semester.

11. Additional Faculty Requirement including internal & external faculty:

A. Faculty : 04

i. Professor : ----

ii. Associate Professor : 01

iii. Assistant Professor : 03

Qualification for Teaching Staff:

The qualification for the above teaching posts shall be as prescribed in the regulations issued by UGC/ Government, from time to time.

Knowledge of working on Statistical software and R-programming is desirable.

B. <u>Nonteaching Staff</u>

i. Tutor 01

ii. Typist –cum- Clerk 01

iii. Attender 01

Qualification:

i) Tutor: B.Sc. graduates with Statistics as one of course and knowledge of Statistical software suchas SPSS/STATA and R.

ii) Typist-cum- clerk: Any degree from a recognised University with the knowledge of computers.

12. Furniture and Equipments required for Department of Statistics

Sl. No.	Office Tables with drawers	Total: 09 01- Prof/HOD 02- Two Assoc. Prof. 05-Asst. Prof. 01-Office Clerk
1	Tables without drawers	Total: 05 02-Two Class Rooms 01-Seminar Room 01-Practical Lab 01-Attender
2	Chairs for Staff	Total: 08 01 (Executive Chair): Prof. & HOD 05 Cushion Chair: (02: Assoc. Prof. & 03: Assist. Prof.) 03 Normal Chair (01 Lab Assistant, 01: Clerk and 01: Attender)
3	Chairs	Total: 08 02-Cushion Chair (for HOD Room) 05-Cushion Chair (for each Faculty)
4	Cub-board	Total:10 01- for HOD Room 01-for Office Clerk 05-for each faculty 02- Glass-Door cub-board for Library 01-Computer Lab
5	Computer Table	Total: 02 01-for HOD Room 01-for Office Clerk 01-Large or Long Common Table for Computer Lab
6	Computer	Total: 22 PCs 01-for HOD Room 01-for Office Purpose 20-for Computer Lab (With basic configuration)
7	Laptop	Total: 01 For Teaching and Presentation Purpose (Common)
8	Statistical Software	SPSS and STATA
9	Xerox cum Printer Machine	Total: 01 01-for Office Purpose
10	Intercom	Total: 02 01: HOD Room 01: Office Room
11	White & Black Board	Total: 04 02:Marker White Board 02: Black Board
12	Projector & Screen	Total: 02 Each for Two Class room

13. Additional Infrastructure Required

Details of Infrastructural facility for the department of Statistics:

Place: Fourth Floor, East Wing , Soorya Buiding, Yenepoya (Deemed to be University).

Sl. No	Details	Sq. feet	Number of rooms	Total area Square feet
1	HOD Room	14ft x14 ft	01	196
2	Supporting staff room (SDA & office)	14 ft x14 ft	01	196
3	Faculty room	12ft x12ft	04	576
4	Board room	13ftx13ft	01	169
5	Classroom	25 ftx30 ft	02	1500
6	Classroom	25 ft x 20ft	0 2	1000
7	Computer Lab	25 ft x 35 ft		875
8	General for Attenders	8ft x 8ft	01	64
9	Library/Research Room	20 ft x 30 ft	01	600
Total				5176 Sq.ft.

14. Requirement of Books/Journals if not already available. : **List Enclosed**

15. Budget requirement :

i. For initial establishment : **Rs.44,48,000**/-
ii. Recurring annual budget : **Rs 33,98,000/-**

Cost Estimate for Furniture and Equipments, Books & journals to the Department of P.G. Studies and Research in Statistics

(a) Books and Journals: Rs 1,50,000/=

(b) Building: For 5,400 Square feet area as per Engineer estimate

(c) Faculty and Nonteaching Staff salary :

i. Existing Faculty: Rs. 24,48,000/-

ii. Guest Faculty and Non- teaching Staff : Rs. 8,00,000/-

iii. Furniture, Equipment and consumables : Rs. 10,50,000/-

16. **Inspection of the Department by team of Experts**.(with minimum of one External Subject Expert in the team):

i) Meeting of Adhoc committee dated 27.6.2019

ii) Meeting of Expert committee dated

07.8.2019 iii)Meeting of BOS in

Statistics dated23.01.2020

17. **Fees Structure:** Rs.75,000/- (Rs. Seventy five thousand) per year.
Intake : **12 seats**

18. **Stipend**, if any : -

19. **Academic Calendar** including date of commencement, hours of instructions in class room teaching, practical, seminars, tutorials etc. Time table for transfer of curriculum, examination(Internal) and University Exam.

Date of Commencement: August 2020*

Semesters: An academic year shall consist of two semesters;

Odd Semester 1st& 3rd	July/August to December/January*
Even semester 2nd& 4th	January/February to June/July*

Hours of instructions in class room Teaching Theory, Practical, Project work etc.

Per week Teaching theory/practice Hours					
Semester	Theory	Practical	Project/Field Work	Total Hours	Total Credits
I	16	12	-	28	22
II	18	12	-	30	24
III	21	6	-	27	24
IV	-	-	24	24	24

Course Scheme of instruction and examination semester wise is given in Table 1.

Academic Calendar*

	I Semester	II Semester	III Semester	IV Semester
Date of commencement	July/August 2020	January 2021	July/August 2021	January 2022
Date of First Internal	8 - 10 Weeks after the date of commencement of the regular classes	8 - 10 Weeks after the date of commencement of the regular classes	8 - 10 Weeks after the date of commencement of the regular classes	8 - 10 Weeks after the date of commencement of the regular classes
Date of second Internal	16 – 18 Weeks after the date of commencement of the regular classes	16 – 18 Weeks after the date of commencement of the regular classes	16 – 18 Weeks after the date of commencement of the regular classes	16 – 18 Weeks after the date of commencement of the regular classes
Date of Main Exam	December-2020/ January- 2021	June -2021	December-2021/ January-2022	June -2022

The date will be finalized/applicable as per the Academic Calendar decided by of the Yenepoya (Deemed to Be University).

20. Award of Certificate/Diploma : M.Sc. in Biostatistics

21. If the course needs extra hours other than routine curriculum how will you manage?

: Arrange special lectures

Annexure - I: Syllabus of M.Sc. Biostatistics

First Semester M.Sc.

Course Code	Type of Course	Course name	Hrs/Week	Credits
FS02BT-1C1	Hard Core-Theory	Mathematical Foundation	3:1:0	4
FS02BT-1C2	Hard Core-Theory	Probability and Probability distributions	3:1:0	4
FS02BT-1C3	Hard Core-Theory	Statistical Inference	4:0:3	4
FS02BT-1C4	Hard Core-Theory	Non-Parametric methods	4:0:3	4
FS02BT-1P1	Hard Core-Practical	Practical I: Data Management and Computing -1	1:0:5	3
FS02BT-1P2	Hard Core-Practical	Practical II**:** Based on APBST 103 & APBST 104	0:0:6	3
Total Credits				22

Hard Core	FS02BT-1C1:Mathematical Foundation	No. of hrs./week: 4 (42L +14T)

Learning Objectives: To acquaint the students with

- Necessary knowledge on theoretical foundation on Calculus and their applications in different branches of Biostatistics.
- Necessary theoretical foundations on Matrix algebra and highlight their applications in different statistical computations.

Learning Outcome: At the end of the Course the student will be able to

- Concentrate on Statistical concepts in subsequent Courses and not be distracted by the mathematics involved in the study.
- Follow the mathematics used in Biostatistics at master's degree level.
- Understand the mathematics behind statistical methods introduced at that level.
- Work out matrix computations, determine different characteristics of a matrix and demonstrate the various decomposition of matrices and their applications.

Unit I: Set Theory: Definition, basics of set operations, counting principles. Sequences of sets and their limits. Set function; Measure, Lebesgue measure, measurable functions.

(08 hours)

Unit-II: Numerical sequences and series: sequences and their convergence; sub sequences and Cauchy sequences. Series- convergence and divergence, tests for convergence, addition and multiplication of series.

(10 hours)

Unit-III: Continuity and differentiation of functions. Differentiation of vector valued functions. Special functions; Power series, series expansion and approximations. The Riemann Stieltjes' integration, properties. Integration by parts, change of variables, step functions as integrators.

(10 hours)

Unit IV: Basics of linear Algebra: Vector spaces, linear dependence and independence; basis and dimension of a vector space. Ortho normal basis and orthogonal projections. Gram-Schmidts Orthogonalization Process. Matrix Operations: Elementary Row operations; Row

reduced Echelon Form, Rank and inverse of a matrix. Partitioned matrices; Generalized inverse of matrices. Linear equation-homogenous and non-homogenous systems, solution spaces.

(14 hours)

Unit V: Characteristic roots and vectors, Cayley-Hamilton theorem, Determinant, rank and trace of a matrix in terms of characteristic roots. Diagonalisation; LU Decomposition; Cholesky Decomposition, Spectral Decomposition, Singular value decomposition. Positive definite matrices. Real quadratic forms, classification of quadratic forms, reduction of quadratic forms.

(14 hours)

References:

1. R B Bapat, Linear Algebra and Linear Models. 3rd edition. Hindustan Book Agency; 2011
2. Apostol, T.M,Mathematical Analysis, Narosa Publshing House, New Delhi. Second Edition; 1996
3. Khuri, A.T , Advanced Calculus with Applications in Statistics. John Wiley & Sons. Inc; 1993
4. Lucy D C, Lay SR and Mc Donald JJ.Linear Algebra and Its Applications. 5th Ed. Pearson,Boston; 2016
5. Malik S.C, and Savitha Arora, Mathematical Analysis. Wiley Eastern;1993
6. Rao, C.R, Linear Statistical Inference and its Applications. 2nd edition, wiley;1995
7. Searle S.R, Matrix Algebra Useful for Statistics. John Wiley& Sons; 1982

Hard Core	FS02BT-1C2**: Probability and Probability Distributions**	**No. credits:4** **(42L+14T)**

Learning objective: To provide necessary foundation in probability and probability distributions so as to apply it to the study of inferential problems studied under Biostatistics.

Learning Outcomes: At the end of the Course the student will be able to

- This course along with other units in the statistical inference provide the basics of mathematical statistics and the students will be able to use it in the study of health science problems. Application of the limit theorems in the study of inferential problems in Biostatistics.
- Identify distribution form relating to the variable/variables related to real life problems in biomedical and health science problems.

Unit I: Random Experiment; sample point; sample space; events, frequency and classical definitions of probability. Axiomatic definition of probability; addition and multiplication theorems; conditional probability and independence; Bayes' theorem (thrust is on numerical problems & applications)

(13 hours)

Unit II: Random variable, cumulative distribution function, probability mass function, probability density function. Joint distributions, marginal distributions, conditional distribution, expectation of a random variable and its properties, conditional expectation, addition and multiplication theorem of conditional expectation, moments and moment generating function, characteristics function, and probability generating function.

(14 hours)

Unit III: Discrete distributions: Binomial, Poisson distribution, Uniform distribution, Negative Binomial distribution, Hyper geometric distribution, Geometric distribution and Multinomial distributions (Definition , moments, properties, moment generating function, Characteristic function and application) .

(10 hours)

Unit IV: Continuous distributions: Normal distribution, Student t distribution, F distribution, Chi- square distribution, Exponential distribution, Weibull distribution, Beta distribution, gamma distribution and lognormal distribution (Definition, moments, properties, characteristic function and applications). Concept of non-central distributions Chi-square, t and F.

(13 hours)

Unit V: Sampling Distributions:. Distribution of sample mean from a normal population; chi-square distribution, F and t statistics, distribution and their applications. The Concept and applications of Law of large numbers, central limit theorem (CLT) and applications. CLT for iid case (statement and examples only). Evaluation of probabilities from the binomial and Poisson distributions using Central limit theorem.

(06 hours)

References:

1. Bernard Rosner, Fundamentals of Biostatistics. 8th Ed. CENGAGE Learning;2016
2. Bhat B.R, Modern Probabilty theory.3rd addition. New Age publishers;1999
3. John N L, Kotz S and Balakrishnan N Continuous Univariate Distributions I & II. John Wiley & Sons; 1991
4. John N L, Kotz S and Kemp AW Univariate Discrete Distributions. John Wiley &Sons; 1992
5. Kai Lai Chung , A Course in Probability Theory. Academic Press; 2001
6. Laha R.G, and Rohatgi V.K. Probability theory. John Wiley; 1979
7. Parimal Mukhopadhyay, Mathematical Statistics. Books and Allied (P) Ltd. Calcutta; 2015
8. Rohatgi V.K. and A.K.M.E Saleh, Introduction to Probability and Statistics. John Wiley and Sons; 2001
9. Sheldon M. Ross, Introduction to Probability models. 9th Ed. Elsevier Inc; 2007

Hard Core	FS02BT-1C3**: Statistical inference**	**No. of hrs./week: 4 (56 L+42P)**

Learning objectives:

- To provide a basic foundation in the concepts and methods of statistical inference, with a strong emphasis on practical aspects of the interpretation of statistically based conclusions in health science problems.

Learning outcome: Students will be able to

- choose an appropriate estimation method and testing procedure in real life situation
- Apply hypothesis testing via some of the statistical distributions.
- Determine the sample size for various situations in biomedical and health science problems.

Unit I: Elementary Concepts in statistics: Concepts of population and sample from a population; qualitative and quantitative data; nominal, ordinal, ratio, interval data; cross sectional and time series data; discrete and continuous data. Collection and scrutiny of data; primary data; designing a questionnaire and schedule; secondary data and its sources. Presentation of data, diagrammatic and graphical representation of data; histogram, stem and leaf charts, ogives, boxplot, Lorenz curve, skewness and kurtosis

(07 hours)

Unit II: Point Estimation: Basic concepts and properties of estimators: Parametric models, parameters, random sample and its likelihood, statistic and its sampling distribution, problem of inference, estimator and estimate, mean square error (MSE), properties of estimators- unbiasedness, consistency, efficiency and sufficiency, Cramer-Rao lower bound, Minimum variance unbiased estimator, relative efficiency of an estimator.

(12 hours)

Unit III: Methods of estimation: Fisher information, complete and sufficient statistic, Rao-Blackwell theorem, UMVUE, method of moments, method of MLE, properties of MLE (statements only), method of minimum chi-square, linkage estimation (Examples from Genetics). Interval estimation: concepts of confidence interval, confidence coefficient, confidence interval for the parameters of univariate normal, proportion, mean, difference of means. Small sample and large sample confidence intervals. Large sample confidence intervals for binomial and Poisson parameters, boot strap methods.

(16 hours)

Unit IV: Hypothesis testing: Introduction, types of errors, critical region, level of significance, one & two tailed, power, p value- interpretation, statistical & clinical significance, test for normality, NP lemma (Statement only) and its application in testing of hypothesis. Maximum likelihood ratio test statistic, hypothesis testing for – a single population mean, difference between population means, Levene's test for equality of variance, ratio of two variances, paired comparisons, single population proportion, difference between two population proportions, a single population variance. Introduction to Score statistic or Wald statistic. One Way ANOVA & post hoc tests, power calculation.

(14hours)

Unit V: Sample size calculation: Importance of Sample size estimation (SSE), basic principle in sample size determination, SSE for Mean, proportion. SSE for comparison of two independent group means, proportion, comparison for more than two group means (Bonferroni's correction), testing mean difference in related samples.

(7 hours)

References:

1. Bernard Rosner, Fundamentals of Biostatistics. 8th Ed. CENGAGE Learning;2016
2. CampbellM J , Statistics at squareone. Ninth Edition, BMJ Publishing Group (Online Book) ;1997
3. Casella G and Berger RL (2002) , Statistical Inference. 2nd Edition. Thompson Duxbury press;2002
4. DouglasG. Altman, Practical Statistics for Medical Research Chapman & Hall; 1991
5. Lehmann E.L. and Casella G, Theory of Point Estimation. Springer, New York ;1998
6. Wayne W. Daniel, Biostatistics. John Wiley &Sons;2009

Hard Core	**FS02BT-1C4: Non-Parametric Methods**	**No. of hrs./week: 4 (56L +42P)**

Learning Objective: To acquaint students with

- Basics of nonparametric methods and its application in biostatistics.
- The methods like bootstrapping and Jackknife methods and their application.
- Analysis and interpretation of results using various nonparametric methods.
- Strength of association using contingency tables.

Learning Outcome: At the end of the course students will be able to

- Identify appropriate nonparametric tests for a real life situation and carryout the statistical analysis and interpret.
- Execute bootstrapping and Jackknife methods.

Unit I: Introduction to Non-parametric statistics:

Data & variable type for nonparametric analysis.

Robustness and distribution free, advantages and disadvantages of nonparametric methods. Introduction to bootstrapping, bootstrap standard error and confidence interval, Jackknife method. Analysis of 2x2 tables-assumptions and limitations, chi-square test and Fisher exact test, extension to RxC tables.

(12 hours)

Unit II: **One sample methods and Test of goodness of fit**

A non parametric test and confidence interval for median-binomial test. Inference for percentiles, Tolerance limits, sign test. Chi square test for goodness of fit. Kolmogorov Smirnov one sample statistic, test for normality. Visual analysis of goodness of fit.

(12 hours)

Unit III: Two independent sample comparisons and its extension:

Two sample permutation test, sign test, Willcoxon rank sum test, Mann Whitney U test, Equivalence of Mann Whitney U test and Willcoxon rank sum test, selection among Willcoxon rank sum and t-test, relative efficiency. Test for equality of scale parameters- Siegel Tukey and Ansari-Bradley test. Kruskal Wallis one way ANOVA and multiple comparison.

(16 hours)

Unit IV: Two related sample comparison & its extension

Willcoxon sign rank test, McNemar test and paired comparison permutation test, sample size for McNemar test, Friedman ANOVA and multiple comparison. Cochran Q test and Page test for ordered alternatives.

(10 hours)

Unit V: Measures of association and its significance: Cramer coefficient, Phi coefficient, lambda, gamma, kendall's tau, spearmen's rank correlation.

(06 hours)

References:

1. Gibbons J K , Practical Nonparametric Statistics. 3rd edition. Wiley publication; 1998
2. Gibbons and Chakraborthy, Nonparametric statistical inference .CRC press; 2010
3. James J. Higgins, Introduction to Nonparametric Statistics: Duxbury advanced Series; 2004
4. Sidney Siegel & Castellen Jhon, Nonparametric Statistics for behavioural Sciences. 2nd Edition. McGraw Hill. International Edition ;1998

Hard Core	**FS02BT-1P1: Practical I - Data Management and Statistical Computing -I**	**No. of credits: 03** **(14 L + 70P)**

Learning Objectives: To acquaint the student with

- Essential concepts and tools required for the management and analysis of data using packages like Excel, R-software and SPSS/STATA.
- Data manipulation, descriptive analysis and interpretation using statistical software

Learning outcome: Students will be able to

- Acquire skills in data display, summary presentation and pattern recognition using the statistical software tools.
- Develop data management skill in Excel, R –software and SPSS.
- Interpret the statistical computations to the real life problem under study.

Content

- Module 1 – Data Management Concepts
- Module 2 – Introduction to R, SPSS and Excel.
- Module 3 – Data Management Using Excel, R and SPSS.

Examples from Biostatistics Context (Biomedical research, Bioinformatics and Psychometrics)

References:

1. Alain F Zuur et.al, A Beginer's Guide to R. Use R Series. Springer; 2009
2. Andy Field , Discovering Statistics Using SPSS. 3rd Edition. Sage Publication Ltd;2009 ISBN 978-1-84787-9056
3. Phil Spector, Data Manipulation with R. Use R Series. Springer; 2008
4. Sabine Landau and Brian S, Everitt: A handbook of Statistical Analyses using SPSS. Chapman & Hall/ CRC, ISBN 1-58488-369-3;2004

Hard Core	**FS02BT-1P2: Practical II (Based on the theory papers on** FS02BT-1C3 & FS02BT-1C4.**)**	**No. of Credits : 03** **(84P)**

Learning Objective: To acquaint the student with

- The analysis and interpretation of descriptive statistics
- Different estimation techniques and their application.
- Basics of testing of hypothesis , Various parametric tests and their application in health Science research.
- Application of various nonparametric tests and their applications in real life problems.
- Estimation of sample size under different situations.

Learning Outcome: Students will be able to

- Understand and verify the basic concepts of parametric and nonparametric tests.
- Estimate appropriate sample size for different statistical tests in research problems.
- Analyse the health science related problems using appropriate parametric tests and interpret the results.
- Apply suitable nonparametric tests and interpret the statistical computations for the real life situation. This course includes practical problems using data from Bio statistical context (Biomedical Research, Bioinformatics and Psychometrics) based on the statistical methods studied in the courses APBST 103 and APBST 104.

1. Descriptive statistics and their interpretation
2. Methods of Estimation

- Interval estimation

3. Exercise on Parametric test

- Size and power calculation. Inference based on p- value and confidence interval.
- Test for normality
- Independent sample t- test
- One sample t- test
- Paired t- test
- ANOVA, Post hoc test and power calculation
- Levene's test for equality of variances

4. Determination of sample size
5. Exercise on Non-Parametric test statistics

- Chi-square test
- Fisher's exact test
- Mann Whitney U test
- Willcoxon sign rank test
- McNemar's test
- Kruskal-Wallis test
- Friedman's test
- Spearman's rank correlation coefficient

Second Semester M.Sc. Biostatistics

Course Code	Type of Course	Course name	Hrs/Week	Credits
FS02BT-2E1	Open Elective Theory	Statistical Methods	3 :0:0	3
FS02BT-2C1	Hard Core- Theory	Sampling methods and Design of Experiments	4:0:0	4
FS02BT-2C2	Hard Core- Theory	Epidemiology	4:0:2	4
FS02BT-2C3	Hard Core- Theory	Applied Regression Analysis	4:0:4	4
FS02BT-2C4.1	Soft Core -Theory	a. Categorical Data Analysis and Bayesian Methods	3:0:2	3
FS02BT-2C4.2		b. Demography and Vital Statistics		
FS02BT-2P1	Hard Core-Practical	Practical III: Applied Regression Analysis and Epidemiology	0:0:6	3
FS02BT-2P2	Hard Core-Practical	Practical IV: Data management and Statistical Computing II (based on FS02BT-2C1&FS02BT-2C4.1/ FS02BT-2C4.2)	0:0:6	3
Total Credits				24

Open Elective	FS02BT-2E1**:** **STATISTICAL METHODS**	**No. of credits: 3** **(42L)**

Learning Objective: To acquaint the students with

- Basic knowledge on Statistics for non -statisticians
- An exposure to various concepts in statistics and their interpretation.
- Training the students on the choice of basic statistical measures

Learning Outcome: At the end of the Course the student will be able to

- Use suitable statistical measures for specific problems and interpret the results.
- Demonstrate an application of the basic statistical measures and their interpretations with real life problems.

Unit I: Statistics: meaning and role as a decision making science, Data-types and scales of measurement. Presentation-tables, diagrammatic and graphical methods. Descriptive Statistics - measures of central tendency, positional averages, measures of dispersion, skewness and kurtosis - Definition and properties. Exploratory Data Analysis using descriptive measures and graphical tools.

(11hours.)

Unit II: Probability theory: random experiment, simple events, sample space - types of events, probability of an event, rules of probability, conditional probability, Bayes' theorem. Probability distributions: random variables - discrete and continuous type, Bernoulli, Binomial, Poisson and normal distributions - applications.

(11 hours.)

Unit III: Sampling methods: population and sample, parameter and statistic, concept of a random sample, simple random sampling, stratified sampling, systematic sampling, Non probability sampling.

(10 hours.)

Unit IV: Correlation: bivariate data, correlation, scatterplot, correlation coefficient and its properties, testing for correlation coefficient, rank correlation.

Regression: linear relationship, linear regression model, simple linear regression, fitting the regression model, coefficient of determination, standard error of the estimated model.

(10 hours.)

References:

1. R.C. Campbell, Statistics for Biologists. Cambridge University Press; 1974
2. Christopher Chatfield , Statistics for Technology. Chapman and Hall; 1981
3. Douglas A. Lind, William C. Marchal, Samuel A. Wathen. "Basic Statistics for Business & Economics" McGraw-Hill Education; 2012
4. Harry Frank and Steven C. Athoen: Statistics, Concepts & Applications. Cambridge University Press; 1997
5. J.Medhi, Statistical Methods. An Introductory Text. Wiley Eastern Limited; 1992
6. Ross,S.M, Introductory Statistics. Academic Press; 2017
7. Rice,J.A, Mathematical Statistics and Data Analysis Dux; 2006

Hard Core	FS02BT-2C1**: Sampling Methods and Design of Experiment**	**No. of credits: 4 (56L+56P)**

Learning objective: Students will be able to

- Understand the basic principles and concepts in design of experiment.
- To provide foundations on types of experimental designs.
- Understand different Experimental designs and their application in real life situation.
- Understand the basic concepts and theory related to different sampling methods.
- Application of sampling methods in biostatistics.

Learning Outcome: Students will be able to

- Identify appropriate sampling method for health surveys, carry out statistical computation and interpret based on the sample drawn.
- Identify suitable experimental design for the research problem and to carry out Statistical Analysis.

Part A: Sampling Theory:

Unit I: Concepts of population and sample, need for sampling, census and sample surveys, basic concepts on sampling , sampling and non-sampling errors, Random sampling , finite population sampling techniques-SRSWR, SRSWOR, estimation of mean or total in each case and their variances. Systematic sampling, stratified sampling, allocation problems in stratified sampling, PPS sampling, estimation of mean or total in each case and their variances. Examples from health sciences.

Nonrandom sampling: Quota sampling, convenience sampling, Snowball sampling-methodology, merits and demerits, their application.

(13 hours)

Unit II: Cluster and Multistage Sampling: cluster sampling, two stage and multi-stage sampling, Auxiliary information in sample surveys, Ratio and regression estimators based on SRSWOR method of sampling, double sampling. Randomized response technique: Warner's model-related and unrelated questionnaire methods. Examples based on health science.

(12 hours)

Part B: Designs of Experiments:

Unit III: Introduction to design of experiments: Need for design of experiments, fundamental principle of design of experiments, basic design-CRD, RBD and LSD, Repeated measure design, layouts and analyses, fixed, random and mixed effects model-Multiple comparison tests-Tukey, Newman- Keul, Scheffe's test, estimation of missing plots and analyses for RBD.

(13 hours)

Unit IV: Factorial designs: 2^n designs, illustration, main effects, interaction effects, and analysis. 3^2 design, illustrations, main effects, interaction effects and analyses. Complete and Partial confounding in 2^n designs in RBD and analyses.

(12 hours)

Unit V: Fractional factorial designs, orthogonal and balanced arrays and their connection with confounded and fractional factorial. Split plot and split block experiments.
Response surface experiments, first order design and orthogonal design. Experiments with mixture models, design and analyses.

(06 hours)

References:

1. Cochran W G and Cox GM , Experimental designs. John Wiley; 1992
2. Cochran W G, Sampling Technique. 3rd Edition. Wiley Eastern;1989
3. Das, M N and Giri N C , Design and Analysis of Experiments. New Delhi :Wiley Eastern Limited; 1979
4. Daniel WW , Biostatistics. A foundation for analysis in the health sciences. John wiley; 2005
5. Montgomery D C , Design and Analysis of Experiments. Wiley India 5th Edition;2006
6. Khuri A and Cornell M, Response Surface methodology. Marcel Dekker; 1996
7. Mukhopadhyay P, Theory and Methods of Survey Sampling. Prentice-Hall of India; 1998
8. Sukhatme et.al, Sampling Theory of Surveys with Applications. Indian Society of Agricultural Statistics. New Delhi; 1984
9. Swain AKPC, Finite Population sampling- Theory and Methods. New Delhi: South Asian Publishers; 2003
10. Ranjan K Som, Practical Sampling Techniques. 2nd edition. CRC publication. ISBN: 9780824796761;1995
11. Zar J H , Biostatistical Analyses. Pearson; 2010

Hard Core	FS02BT-2C2 **: Epidemiology**	**No. of credits: 4 (56L+28P)**

Learning Objectives: To acquaint the students with

- Foundation on basics epidemiological study designs.
- Different study designs, selection of appropriate research design.
- Analysis and interpretation of results from Epidemiological studies.

Learning outcome: At the end of the Course the student will be able

- Understand and interpret various concepts under Epidemiological study.
- Translate the research objectives into clear, testable statistical hypothesis.
- Identify appropriate Epidemiological study design for the research problem and carryout the Statistical Analysis.

Unit-I: Research methodology: Definition, meaning and objectives of research, need and importance of epidemiological research in public health. Types of research. Good research criteria, formulation of research question and its key considerations, formulation of hypothesis, need and importance of review of literature, steps in formulating a good research proposal.

Ethics in Research: Orientation to Institutional Ethics committee, importance of ethics in research, ICMR ethical guidelines, Ethical and professional issues related to data security and publication.

Fundamentals of Intellectual Property Rights (IPR): What are patents and inventions? Key differences - Invention and discovery - Understanding inventorship, ownership and rights in the context of collaborative research- Novelty, inventive step and industrial applicability - Introduction to patent specifications - Content of provisional and complete specifications**.** - Conduct search (novelty and inventive steps) - Conduct prior art with keywords - Booleans operators - Use of parenthesis - Patent description writing - Patent claims writing - Precautions of writing the specifications. Patent specifications and filing procedure

(14 hours)

Unit II: Measures in Epidemiology: Definition of Epidemiology, Historical developments in epidemiology, John snow's study and Doll Hill study, epidemiological triad, emergence of modern epidemiology, role of epidemiology in health science, classification of diseases, measures of disease frequency, rate ratio and proportion, prevalence and incident rate, rate risk and ratio, point & period prevalence, relation between prevalence and incidence, measures of exposures and outcomes, types of exposure and outcome, sources of exposures, measures of morbidity and mortality rate, Cumulative rate and risk, person years, age specific incidence rates, case fatality rate, measures of a disease and exposure, standardization of rates: Concept , direct and indirect methods and introduction to confounding.

Validity and reliability of measures of exposure and outcome, basic concepts of surveillance and levels of prevention –Primary, secondary and tertiary. Screening- definition and requirements, biases, sensitivity and specificity, positive predictive, negative predictive test results, likelihood ratio positive and negative, ROC analysis.

(15 hours)

Unit III: Observational studies: Concept and objectives of observational studies, explorative, descriptive and analytical designs, study design and their importance, advantages and disadvantages; descriptive study designs. Case report, Case study, case series, cross sectional study, Strengths and limitations of different study designs.

(07 hours)

Unit IV: Analytical study designs: Case Control: Definition and selection of cases and controls, measuring exposure, Odds ratio (OR), Confidence interval for OR, Attributable Risk and Population attributable Risk, Interpretation of results, advantages and disadvantages of case control studies, advantages and disadvantages of population based case control studies over hospital based. Nested case control study. Matched case control studies, analysis, advantages and disadvantages of matching, Ecological study.

Cohort studies : Choice of study population, definition of cohort, choice of comparison group, measurement of exposure, outcomes, Relative Risk(RR), rate difference, confidence interval for RR, interpretation of RR, Case cohort studies and historical cohort studies. Advantages and disadvantages of prospective, historical cohort study.

(14 hours)

Unit V: Dealing with Confounding variables: various methods of dealing with confounding, concept of matching, propensity score matching, restricted sampling. Introduction to stratification, Mantel Haenszel summary measures, MH Odds ratio, MH risk ratio, MH confidence interval for OR and RR. Interaction, relative and absolute measures, attributable risk, attribute risk percentage, population excess risk. Relative risk (Risk ratio, rate ratio, odds ratio).

(06 hours)

References:

1. Altman DG, Practical statistics for medical research. CRC press; 1990
2. Bonita R, Beaglehole R, Kjellström T. Basic epidemiology. World Health Organization; 2006
3. Clayton and Hills, Statistical Models in Epidemiology. Oxford : Oxford University Press; 1993
4. Ethical Guidelines for Statistical Practice prepared by a committee on Professional Ethics of the American Statistical Association, Approved by the ASA Board in April; 2018
4. Gordis L, Epidemiology. Saunders-Elsevier Philadelphia; 2013
5. Fleiss JL, Levin B, Paik MC . Statistical methods for rates and proportions. John Wiley & Sons; 2013
6. Kenneth J Rothman, Causal inference in Medicine. Epidemiology Resources. ISBN-13 978- 0917227035; 1988
7. Last JM, Abramson JH, Freidman GD. editors. A dictionary of epidemiology. New York: Oxford university press; 2001
8. Olinda Timms, Biomedical Ethics. 2nd Ed. New Delhi, Elsevier; 2019
9. Park K, Preventive and Social Medicine. Jabalpur: M/s Banarsidas Bhanot Publishers; 2011
10. Rothman KJ, Greenland S, Lash TL. Modern epidemiology. Lippincott Williams & Wilkins 2008
11. Schlesselman JJ, Case-control studies: design, conduct, analysis (Monographs in Epidemiology and Biostatistics) Oxford University Press; 1982

https://www.amstat.org/ASA/Your-Career/Ethical-guidelines for Statistical Practice.aspx

Hard Core	FS02BT-2C3**:** **Applied Regression Analysis**	**No. of credits: 4** **(56L+56P)**

Learning Objectives: To acquaint the students with

- Necessary tools for the analysis of problems in Biomedical and health science research using regression methods.
- Tools used to detect the problems such as multicollinearity, heteroscedasticity, autocorrelation etc in biomedical and health science research. Statistical analysis of such data using appropriate regression methods.
- Take up Case studies and carryout regression analysis.

Learning Outcome: At the end of the Course the student will be able to

- Apply regression analysis to identify significant factors influencing the response variables in real life situations.
- Use appropriate statistical tool for the Analysis of problems in Biomedical research, Health science research and Bioinformatics.

Unit-I: Simple linear regression: Introduction to regression analysis, Assumptions and least square estimation of model parameters, Standard error of the estimators, testing of hypothesis on the significance of slope and intercept. Coefficient of determination, residual analysis, ANOVA.

(10 hours)

Unit II: Multiple linear regression: Description of the model. Estimation of regression parameters and error variance, properties, test of hypothesis of regression parameters, Significance of regression (ANOVA, R^2, Adjusted $R^{2)}$, dummy variable regression-general concepts and uses. Residual analysis.

(12 hours)

Unit III: General linear models: Introduction- Gauss Markov set up, assumptions-homoscedasticity and heteroscedasticity, testing for heteroscedasticity, generalised least squares estimator, multicollinearity, its detection and methods to deal with multicollinearity. Autocorrelation, Durbin- Watson test, estimation of model parameters. Variance stabilising transformations to linearize the model. Analytical methods for selecting a transform.

(16 hours)

Unit IV: Variable selection:

Selection of variables, forward selection, backward elimination and stepwise regression (Algorithms only), Information Criteria, Akaike information criteria. Residual analysis

(06 hours)

Unit V: Introduction to non-linear regression.

Nonlinear regression- Transformation to a linear model, Usefulness of the nonlinear regression method, limitations of the nonlinear regression method, use of re-sampling procedures in regression. **Logistic Regression**: Introduction, logistic regression models- Dichotomous, Polytomous and Continuous independent variables. Fitting logistic regression model, Conditional Logistic Regression, Interpreting coefficients. Testing the significance of model. Adjustment for Confounding and interaction factors.

(12 hours)

References:

1. Draper N.R. and Smith H, Applied Regression Analysis. 3rd Ed. John Wiley and Sons. Inc; 1998
2. Hosmer D W and Lemeshow S, Applied Logistic Regression. Wiley; 1989
3. Johnston J, Econometric Methods. 3rd Ed. McGraw Hill; 1984
4. Johnston J. and Dinardo J, Econometric Methods. 4th Ed. McGraw-Hill Companies; 1997
5. Montgomery, D. C., Peck, Introduction to linear regression analysis. John Wiley and Sons. 3rd Ed; 2003
6. Weisberg S, Applied Linear Regression Wiley; 2005
7. Zar J H , Bio statistical Analysis.Pearson Education. 4th edition;2006

Soft Core	**FS02BT-2C4.1a:** **Categorical Data Analysis and Bayesian Methods**	**No. of credits: 3 (42L+28P)**

Learning Objectives: To acquaint the students with

- Various basic concepts and tests used in categorical data analysis and Bayesian methods.
- An emphasis on the practical interpretation and communication of results to colleagues and clients who may not be statisticians.
- Application of these techniques to problems in biomedical and health science.

Leaning Outcome: At the end of the Course the student will be able to

- Carry out analysis of two way and higher order contingency tables, its interpretation in terms of measures of strength of association.
- To have good understanding of Bayes estimate and its credible intervals, choice of priors in obtaining it and testing of a statistical hypothesis.

A: Categorical data analysis

UNIT I: Contingency tables analysis: Introduction to contingency table. Probability distribution used for the analysis of categorical data. Sampling schemes for the construction of contingency table given the marginal totals-Poisson, multinomial, binomial sampling schemes. (Discussion confined to hypothesis, deriving expected frequency and the corresponding chi square test). Exact test of independence for RxC contingency table using multiple hyper geometric distribution, Likelihood ratio test procedure for a contingency table. The analysis of ordered tables by assessing scores to categories. Cochran Armitage trend test.

(09 hours)

UNIT II: Test of association and measure of agreement

Testing for independence, in RxCxK tables- Complete independence, partial independence and conditional independence. Testing for Conditional independence in 2x2xK tables and RxCxK tables by Chochran Mantel Haenszel test.

Measures for paired data – Stuart Maxwell Mc- Nemar test for marginal homogeneity for related square table, Chochran Mentel Haenszel correlation statistics for testing marginal homogeneity in a RxCxK table. Measuring agreement between observers- kappa measure of agreement. Analysis of categorical outcomes in a crossover design or Latin square design and clustered randomized trials (Chi square test adjusted for design effect).

(13 hours)

B: Bayesian Inference

UNIT III: Basic concepts of Bayesian inference

Basics of Probability. Conditional probability, theorem of total probability. Bayes's theorem. Likelihood function, prior distribution, posterior distribution. Methods of construction of priors and computation of the posterior distribution. Application of Bayes theorem for positive predictive value, negative predictive value, Likelihood ratio and posterior odds of a diagnostic test.

Basic concept of Bayesian inference; Posterior distribution, Likelihood function and prior distribution . Difference between Bayesian and Frequentist approach. Posterior distribution summaries; mean, median, percentiles and variance. Choice of priors- Informative priors, conjugate priors and improper priors. Introduction to 95% Credible intervals and highest probability density (HPD) credible intervals. Difference between credible and confidence interval.

(10 hours)

UNIT IV: Inference for proportion & mean and Bayesian hypothesis testing

Choice of prior distribution; Beta distribution, Uniform (0,1) priors and informative priors. Posterior distribution of proportion under prior distribution specified above. Posterior distribution summaries; Mean , median, mode and variance. Bias and means square error of Bayesian estimation of proportion

Precise and imprecise null hypothesis. Hypothesis testing using posterior distribution (For precise hypothesis), posterior odds, Bayes factor and interpretation. Hypothesis test using 95% HPD credible interval.

(10 hours)

References :

1. Agresti A, Categorical data analysis. John Wiley & Sons; 2013
2. Agresti, A, An introduction to categorical data analysis. Second Edition. John Wiley & SonsInc. Publication; 1996
3. Broemeling, L.D Bayesian biostatistics and diagnostic medicine. CRC Press; 2007
4. Everitt, B S , The analysis of contingency tables. CRC Press; 1992
5. Wiliam M. Bolstad, Introduction to Bayesian Statistics. John Wiley & Sons; 2004

Soft Core	FS02BT-2C4.2 b : Demography & Vital Statistics	No. of credits: 3 (42L+28P)

Learning Objectives: To acquaint the students with

- Various basic concepts on demography and vital statistics
- Necessary knowledge on the practical application of demography and vital statistics.
- Acquire knowledge on life table, its construction, application and interpretation.
- Population projection and statistical models for measurement of population growth.

Learning Outcome: At the end of the Course the student will be able to

- Describe the various concepts and results on Demography and vital statistics and work on the theoretical advancements on these topics.
- Construct life table and give interpretations.
- Calculate various measures of mortality and fertility.
- Demonstrate suitable growth models for population projection with real life data.

Unit I: Concept of demography and vital statistics, scope and importance of Demography and vital statistics. Data sources and their limitations: Population Census, Vital Registration, Population Register, Demographic Health Surveys. Vital registration system in India- civil registration system (CRS), Sample registration system (SRS) and cause of death reporting system in India.

(8 hours)

Unit II: Measures of fertility: Crude and specific rates of fertility, uses of fertility indicators, data sources, various measures of fertility: CBR, GFR, GMFR, TFR, TMFR, Age specific fertility rate, GRR, NRR. Distribution of time to first birth, inter live birth intervals and number of births (for both homogeneous and non homogeneous groups of women), estimation of parameters, estimation of parity regression ratios from open birth interval data.

(12 hours)

Unit III: Measures of Mortality: Crude and specific rates of mortality, various measures of mortality CDR, ASDR, IMR, CSMR, PMR, NMR, Postnatal mortality, infant mortality rate); Mortality- levels, trends and differentials. Life tables, elements of life table, construction of complete and abridged life tables. Expectation of life and application of life tables.

(11 hours)

Unit IV: Migration: Migration Rates and Ratios; Indirect measures of net-internal migration. National growth rate method. Estimation of Measures of Mortality. Measurement of Population growth. Linear, Geometric, exponential Gompertz, Logistic Population Growth models. Method of population projection, use of Leslie matrix. Stochastic models for population growth.

(11 hours)

References:

1. Altman Douglas G, Practical Statistics for Medical Research. London:Chapman &Hall; 2000
2. Breslow and Day, Statistical Methods in Cancer Research Vol.1 . The Analysis of Case Control Studies. WHO; 1980
3. Gordis Leon , Epidemiology. Elsevier. Philadelphia; 1996
4. Park K,Park's Text Book of Preventive and Social Medicine. Little Brown &Co. Boston; 2009
5. Pathak KB & F. Ram, Techniques of Demographic Analysis. 2nd Edition. Himalaya Publication House; 2016
6. Nathan Keyfitz and John A Beckman, Demography through problems. Problem book in Mathematics Springer: ISBN 10: 1441928170 ;2010
7. Roger Detels , Robert Beaglehole, Mary Ann Lansang and Martin Gulliford. Oxford Text Book of Public Health. 5th edition. Edited volume, Oxford: Oxford University Press; 2009
8. Swanson DA, Siegel JS, Stockwell EG, Shryock HS .The Methods and Materials of Demography second edition; 1973

Hard Core	**FS02BT-2C2&FS02BT-2C3: Practical III: based on Applied Regression Analysis and Epidemiology**	**No. of credits: 3 84P**

Learning Objective: To acquaint the students with

- Hands on working knowledge of regression analysis using R language.
- Basic analysis of data from various study designs in Epidemiology.

Learning Outcome: At the end of the Course the student will be able to

- Carry out Regression Analysis using software and interpret the results.
- Identify the various problems related to regression model, choose appropriate method of estimation and testing. Carryout diagnostic checking through residual analysis.
- Analyse the data related to different study designs and ratios under Epidemiology.

List of Practical on Applied Regression Analysis and Epidemiology using statistical packages with examples from biomedical and health science.

1. Simple linear regression.
2. Multiple linear regression., Residual Analysis
3. Testing for autocorrelation and fitting auto-correlated model.
4. Testing Heteroscedasticity and fitting multiple linear regression model.
5. Multicollinearity, Ridge regression.
6. Best subset selection based on MSE, R^2 and Mallows C_p-criterion.
7. Logistic Regression
8. Analysis of data from different study designs
9. Mantel Haenszel summary measures.
10. Calculation of Relative risks and ROC analysis
11. Sensitivity and specificity analysis

Hard Core	**FS02BT-2P2: Practical IV:** **Data Management and Statistical computing II** **(on** FS02BT-2C1 & FS02BT-2C4.1/ FS02BT-2C4.2**)**	**No. of credits:3** **84P**

Learning Objectives: To acquaint the students with

- Use of SPSS and R- software for solving real life problems.
- Random sampling methods and obtaining the random sample. Application and limitations of these methods in biomedical and health science applications.
- Non probability sampling methods and its application in health science study.
- Various Designs of experiments and their applications in biomedical study.

Learning outcome: At the end of the Course the student will be able to

- Choose appropriate sampling design for a real life problem and determine the sample size.
- Draw random sample, carryout statistical analysis and report the findings.
- Identify appropriate Experimental design in biomedical and health science applications and carryout ANOVA.

List of practical: Practical on APBST202 and APBST 205 with examples from biomedical and health science using statistical packages.

Sampling Methods:

1. Method of drawing a sample: i) Random sampling and ii) non probability sampling
2. SRSWR and SRSWOR sampling
3. Cluster sampling with clusters of equal size.
4. Cluster sampling with clusters of unequal size.
5. Two stage sampling with SRSWOR at both the stages.
6. Ratio Method of estimation. And Regression method of estimation.
7. Non probability sampling

Designs of Experiments:

8. ANOVA in CRD , RBD and LSD
9. Repeated measure design
10. Multiple comparison
11. 2^n factorial design – Main effects, interactions, their best estimates and ANOVA
12. 3^n factorial design – Main effects, interactions, their best estimates and ANOVA
13. 2^n Complete and partial confounding
14. Fractional Factorial designs.

FS02BT-2C4.1: Categorical Data Analysis and Bayesian methods

1. Test of independence for RxC contingency table
2. Testing for Complete independence, Partial independence and conditional independence in RxCxK contingency table.
3. Measuring agreement between observers- Kappa measure of agreement.
4. Analysis of Categorical Outcomes in a cross over design or Latin square design.
5. Bayes estimation under conjugate family, hyper parameters of the conjugate family and mixturesof conjugate families.
6. Bayesian credible interval, HPD credible interval from exponential family.
7. Estimation of posterior density, HPD credible intervals using importance sampling from exponential family.
8. Posterior density estimation, HPD credible intervals using Gibbs sampler

FS02BT-2C4.2: Demography and Vital statistics

1. Various measures of fertility
2. Construction of life table and interpretation
3. Estimation of measure of fertility
4. Estimation of measure of mortality
5. Population growth models and population projection

Third Semester M.Sc. Biostatistics

Course Code	Type of Course	Course name	Hrs/Week	Credits
FS02BT-3E1	Open Elective Theory	Basic data analysis techniques	3 :0:0	3
FS02BT-3C1	Hard Core- Theory	Survival Analysis	4:0:2	4
FS02BT-3C2	Hard Core- Theory	Design and Analysis of Clinical Trials	4:0:0	4
FS02BT-3C3	Hard Core- Theory	Hierarchical Linear and Hierarchical Generalized Linear models	4:0:2	4
FS02BT-3C4.1 FS02BT-3C4.2	Soft Core -Theory	a. Genetic epidemiology and Bioinformatics b. Psychometrics	3:0:1	3
FS02BT-3C5.1 FS02BT-3C5.2	Soft Core Theory	a. Time Series Analysis b. Nonparametric regression	3:0:1	3
FS02BT-3P1	Hard Core-Practical	Practical IV: Data management and Statistical Computing III (Based on all theory papers)	0:0:6	3
Total Credits				24

Third Semester: M.Sc. Biostatistics

Open Elective	FS02BT-3E1: Basic data analysis techniques	No. of credits: 3 (42 L)

Learning Objectives: To acquaint the students with

- Various statistical tests for testing and their application.
- The situation where the statistical tests are used.
- The assumptions related to the tests and their verification.

- **Learning outcome**: At the end of the Course the student will be able to
- Verify the conditions and choose appropriate statistical test for testing of hypothesis related to specific real life problem.
- Interpret the computational results.

Unit 1

Population and sample, parameter, statistic, estimator, statistical properties of estimators. Basic concepts concerning testing of hypotheses, procedure for hypothesis testing. Null hypothesis, alternate hypothesis, statistical test procedures, test statistic, two types of errors, level of significance, p-value, size and power of the test. One sided and two sided test procedures. Parametric and nonparametric tests.

(12 hours)

Unit 2

Assumptions, test procedures and examples - One sample Z test, hypothesis testing of means, hypothesis testing for differences between means under equal variance and unequal variances, paired t- test, tests for proportions. Sample size and its determination.

Hypothesis testing for comparing a variance to some hypothesized population variance, testing the equality of variances of two normal populations, hypothesis testing of correlation coefficients and regression coefficients, Confidence intervals.

(14 hours)

Unit 3

Non-parametric tests, sign test, Wilcoxon signed rank test, Wilcoxon rank sum test-Mann-Whitney test, Contingency tables - Chi-square test for independence of attributes,

Principles of design of experiments, basic principle of ANOVA, CRD, RBD, LSD. Tukey multiple comparison test with equal sample sizes, Tukey-Kramer test with unequal sample sizes.

(16 hours.)

References:

1. Douglas A. Lind, William C. Marchal, Samuel A. Wathen. Basic Statistics for Business & Economics. McGraw-Hill Education; 2012
2. J. Medhi , Statistical Methods . An Introductory Text. Wiley Eastern Limited; 1992

Hard Core	**FS02BT-3C1: SURVIVAL ANALYSIS**	**No. of credits: 4** **(56L+28P)**

Learning Objectives: To Acquaint the student with

- Basics Knowledge in survival analysis and its application.
- Analysis, interpretation and reporting of parametric and non-parametric methods used in survival analysis.
- Regression models used in survival function.

Learning Outcome: At the end of the Course the student will be able to

- Use of different models and Probability Distributions for the study of survival function.
- Select appropriate Statistical model and method of estimation of survival function for real life situation.

Unit-I: Why survival analysis? Censored Data , types of censoring. Functions of Survival time-definitions, Relationships of the survival functions. Life distributions -Exponential, Gamma, Weibull, Lognormal, Pareto, Proportional Hazards family. Product-Limit estimates of survivorship function, Nelson-Aalen estimator of survival function. Estimation of parameters for exponential and gamma distributions under various censoring situations.

(12 hours)

Unit-II: Life tables: Life table analysis, Relative, Five Year and Corrected Survival Rates. Standard methods for uncensored and censored data; Asymptotic properties of estimates under a random censorship model. Failure rate, mean residual life and their elementary properties. Estimation of survival function - Kaplan Meier Estimator, The Mantel-Haenszel test, Greenwood's formula. Other life table estimators. Case study. Test for Comparison of Survival Functions (Log-Rank, Breslow's, Taronware, Cox F test)

(12 hours)

Unit-III:

Parametric methods for regression model fitting

Preliminary examination of data, General structure of Parametric regression model. Exponential Regression model, Weibull Regression model, Lognormal Regression model, Generalised Gamma Regression model, log logistic regression model. Case study.

Semi-parametric regression for failure rate – Cox's proportional hazards model with one and several covariates, partial likelihood function, Wald, score and likelihood ratio tests, Breslow estimator, stratification, estimation of the Survivorship function with covariates, Adequacy Assessment of the Proportional Hazards Model. Cox regression model diagnostics: residuals, functional forms, outlying and influential cases, model validation, Residual Analyses (Cox-Snell, Deviance, Martingale, Schoenfeld).

(14 hours)

Unit IV: Identification of Prognostic Factors related Survival Time: Non proportional Hazards Models, Models with time dependent covariates, Stratified Proportional Hazard Models. Competing Risk models, Recurrent event models, models for related observations. Sample size and power in survival analysis.

(10 hours)

Unit-V: Fully parametric analysis of dependency – accelerated life model – simple form, log logistic accelerated life model, proportional hazards model – relation with accelerated life model. Restricted Mean Survival Models.

(08 hours)

References:

1. Cox D.R. and Oakes D, Analysis of Survival Data. New York :Chapman and Hall;1984
2. Elisa T.Lee and John Wenyu Wang, Statistical Methods for Survival Data Analysis. John Wiley & Sons INC; 2003
3. Hosmer D.W, Lemeshow S. and May S. Applied Survival Analysis: Regression Modelling of Time-to- Event Data. 2nd Edition. John Wiley& Sons Inc;2008
4. Kalbfeisch J.D. and Prentice R.L, The Statistical Analysis ofFailure Time Data, John Wiley& Sons. Inc. 2nd Edition;2002
5. Lawless J.F, Statistical Models and Methods for Lifetime Data. John Wiley & Sons Inc; 2002
6. Miller R.G, Survival Analysis. John Wiley& Sons Inc; 1981
7. Klienbalm & Klienbalm, Survival Analysis: A Self Learning Text. 3rdedition .New York: Springer Science+Business Media, LLC, Springer; 2012

Hard Core	FS02BT-3C2**:** **Methods in Clinical Trials**	**No. of credits: 4** **(56 L)**

Learning Objectives: To acquaint the students with

- Basic knowledge on various concepts in Clinical trials
- Statistical principles, concepts and methods for the analysis ofdata in a clinical trial.

Learning outcome: At the end of the Course the student will be able to

- Explain the general principles of clinical trials and its implications for valid inference.
- To design and carry out analysis of measurements at Phase 1, 2,3, and 4 clinical trials.

UNIT I: Introduction to clinical trials: Historical background- The need and Ethics of clinical trials. Types of bias and random error in clinical studies. Types of clinical trials, bioequivalence trials. Concept of blinding, randomization technique. Organisation and planning: Main features of the study protocol- selection of patients-Treatment schedule- Evaluation of patient response-follow-up study- GCP/ICH guidelines. Intervention Study population- Definitions, baseline assessment and recruitment.

(12 hours)

Unit II : Phases of Clinical trials: Phase I, II, III, and IV trials: Randomised control study, Non randomised concurrent study, Historical controls. Cross over design, withdrawal studies, group allocation designs, hybrid designs and studies of equivalence (description only). Methods of randomisation: Fixed allocation randomisation, stratified randomisation, Adaptive randomisation- Unequal randomisation,

(12 hours)

Unit III: Blinding and Placebos : Unblinded(open label), single blind trials, double blind trials and triple blind trials. Conduct of double blind trials, assessment and reporting of blindness, advantages and disadvantages.

(08 hours)

Unit IV: Single end point/ outcomes, multiple endpoints/outcomes, Surrogate response variables and their desirable properties and usefulness in clinical trials. Meta- analysis of clinical trials, paired and unpaired tests, significance of clinical tests. Introduction to superiority and non-inferiority, equivalence trials.

Sample size determination for a clinical trial, data collection and quality control, assessing and reporting adverse events.

(12 hours)

UNIT V: Issues in data analysis. Non adherence, poor quality or missing data, competing events. Covariate adjustments- Surrogate as covariates, baseline characteristics as covariates.

Analysis of categorical outcome from Phase1 to Phase 3 trials. Multiple test and Subgroup analysis including adjustments of significance level and p values. Intention to treat analysis, per protocol analysis and as treated analysis. Comparison of multiple variables. Termination procedures, early stopping due to intervention activity/ toxicity, post study follow up, data cleanup and verification, storage of study material. Reporting of clinical trials CONSORT guidelines.

(12 hours)

References:

1. Chow S.C. and Liu J.P, Design and Analysis of Bioavailability and bioequivalence. 3rd Ed. CRC Press;2009
2. Chow S.C. and Liu J.P ,Design and Analysis of Clinical Trials. 2nd Ed. Marcel Dekkar;2004
3. Fleiss J. L , The Design and Analysis of Clinical Experiments. Wiley; 1989
4. Friedman L.M, Furberg C. D, and Demets DL, Fundamentals of clinical trials. 4th edition. Springer;2010
5. Jennison .C. and Turnbull B. W , Group Sequential Methods with Applications to Clinical Trails. CRC Press;1999
6. Marubeni .E. and Valsecchi M. G, Analyzing Survival Data from Clinical Trials and Observational Studies. Wiley; 1994
7. Piantadosi S, Clinical Trials a Methodological Perspective. 2nd edition, John Wiley&Sons;2005
8. Pocock SJ, Clinical trials: a practical approach. John Wiley & Sons;2013

Hard Core	FS02BT-3C3**:** **Hierarchical linear and Hierarchical Generalized linear models.**	**No. of credits: 4** **(56L +28P)**

Learning Objectives: To acquaint the students with

- Hierarchical linear and Hierarchical generalised linear models.
- Understand the concepts, requirement of mixed models and its difference from usual linear and generalised linear models.

Learning outcome: At the end of the Course the student will be able to

- Identify appropriate design as Hierarchical linear and Hierarchical generalised linear models for the biomedical research and other health related study.
- Apply the methods discussed in the analysis of data from longitudinal studies, complex surveys, cluster randomised trials and other sources of hierarchical data.

Unit I: Introduction to Hierarchical data

Introduction to Hierarchical data. Designs which lead to hierarchical data. Multistage surveys, longitudinal studies, cluster randomised trials. Concept of design effect and intra cluster correlation coefficient.

(4 hours)

Unit II: Univariate analysis of clustered randomised trials and longitudinal data

Univariate analysis of continuous outcome in a cluster randomised trials. Individual level analysis, weighted average analysis, cluster level analysis, analysis corrected for design effect. Sample size determination for cluster randomised trials. Practical choice of intra cluster correlation and design effect. Summary measure approach for the analysis of longitudinal data.

(8 hours)

Unit III: Hierarchical linear models: Linear mixed models

Introduction to fixed and random effect models. Random intercept and random coefficients model. Estimation of regression parameter in linear mixed model using restricted maximum likelihood estimation procedure. Examples of application of linear mixed models in cluster randomised trials and longitudinal data. Variance component approach for the analysis of longitudinal data. regression model.

(12 hours)

Unit IV: Hierarchical linear models: Covariance pattern model

Repeated measures ANOVA-Design, Analysis: assumption of sphericity, test for sphericity, alternative test when sphericity condition is violated. Variance ratio for accounting correlation arising from each level in random effects

Introduction to covariance pattern model. Different covariance structure used in covariance pattern model. Selection of different covariance structures from nested covariance structure using likelihood ratio

(12hours)

Unit V: Hierarchical generalised linear models Random effect models, marginal models and transition models. Introduction to generalised linear mixed models (GLMMIX) and generalised estimating equations method (GEE)-Methodology, hypothesis testing and inference.

Population averaged logistic regression and Poisson regression (GEE for binary and count outcome variable) and subject specific Logistic regression (GLMMIX for binary and count outcome variable). Estimation and inference with example. Extension to nominal and ordinal variables. Diagnostics used in GLMMIX and GEE.

(20 hours)

References:

1. Brown H and Prescott R, Applied mixed models in medicine. John Wiley & Sons ;2015

2. Berry,.D.A, Statistical Methodology in the pharmaceutical sciences. CRC Press;1989

3. Diggle, P, Heagerty, P., Liang K.Y., & Zeger S, Analysis of longitudinal data. Oxford University Press; 2002

4. Davis,C.S, Statistical Methods for the analysis ofrepeated measurements. Springer. Science and Business Media; 2002

5. Everitt B., & Pickles A , Statistical aspects of the design and analysis of clinical trials. London Imperial College Press; 2004

6. Rencher, A.. and Schaalje, G.B, Linear mixed models, Linear models in Statistics. 2nd Edition.John Wiley & Sons Inc. Publication; 2007

7. Song,X.K, Correlated data analysis. modelling, analysis and applications. Springer Science& Business Media; 2007

Soft Core	FS02BT-3C4.1a: Genetic Epidemiology & Bioinformatics	No. of credits: 3 (42L+14P)

Learning Objective: To acquaint students with

- Basic concepts of genetics epidemiology and Bioinformatics
- Applications of statistical methods in Genetic Epidemiology and Bioinformatics
- Basics of bioinformatics and difference between bioinformatics and computational biology
- Knowledge on common bioinformatics (nucleotide, protein, and structure) databases and file formats (plain text files)

Learning Outcome: At the end of the Course the student will be able to

- Interpret the basic concepts in Genetic Epidemiology and Bio informatics.
- Carryout Statistical analysis of data from studies related to genetic epidemiology and Bioinformatics.

UNIT I: **Introduction to Genetics**

Basic concepts of genetics - gene, allele, genotype & phenotype, Mendel's laws - pea plant experiment, law of dominance, law of segregation and law of independent assortment, monohybrid cross, dihybrid cross, test cross and back cross, methods to solve genetic crosses - punnet square and probability method, goodness-of-fit.

Extensions and modifications of basic principles: Co-dominance and incomplete dominance, lethal alleles, multiple alleles; gene interactions –with Epistasis, interaction; interaction between sex and heredity, sex influenced, sex limited, sex linked characteristics; interaction between gene and environment.

(12 Hours)

UNIT II: **Population Genetics**

Definition and description, allelic frequency and genotypic frequency, calculating allelic frequency, the Hardy-Weinberg law, forces affecting gene frequency; mutation – forward and backward mutation, net change in the gene frequency due to mutation, migration – immigration & emigration, effect of migration on gene frequency, natural selection – fitness, selection coefficient, general selection model, selection under different situations, selection against a recessive trait, selection against a dominant trait, selection in the absence of dominance, selection against both homozygotes. Fisher's fundamental theorem of natural selection, genetic drift – magnitude of genetic drift, effective population size, causes and consequences of genetic drift.

Quantitative genetics: quantitative characteristics, heritability, heritability types, methods for calculation of heritability phenotypic variance, limitations of heritability, repeatability and genetic correlation.

(11 Hours)

UNIT III: Genetic epidemiology:

Introduction to genetic epidemiology, Familial aggregation, Segregation analysis, Linkage analysis, Association studies, Applications of genomics.

(07 Hours)

Unit IV: Bioinformatics:

Introduction and History of Bioinformatics, Bioinformatics concepts- Functional Genomics, Comparative genomics, Structural biology, Medical information. Sequence alignment, Markov Model, Next Generation sequencing, Mapping genetic variants. Careers in Bioinformatics, Major databases & tools.

(12 hours)

References:

1. Pierce BA. Genetics, A Conceptual Approach. New York:WF Freeman and Company; 2012
2. Liu BH , Statistical genomics: linkage. Mapping. and QTL analysis. CRC press; 2017
3. Thomas DC, Statistical methods in genetic epidemiology. Oxford University Press; 2004
4. Guttman B, Griffiths A, Suzuki D, Cullis T., Genetics : A beginner's guide. Oneworld Publications; 2002
5. Winchester AM, Genetics: a survey of the principles of heredity. Houghton Mifflin companyBoston; 1977
6. Laird NM, Lange C, The fundamentals of modern statistical genetics. Springer Science &Business Media; 2010
7. Lin S, Zhao H, Handbook on Analyzing Human Genetic Data Computational Approaches and Software. Springer-Verlag, Berlin Heidelberg; 2010
8. Bassi S , Python for Bioinformatics. CRC Press; 2017
9. Lesk AM, Introduction to Bioinformatics. Oxford University Press Inc; 2002
10. Mount DW , Bioinformatics - Sequence and Genome Analysis. CSHL Press; 2001

Soft Core	FS02BT-3C4.2 b: Psychometrics	No. of credits : 3 (42L+14P)

Learning Objective: To acquaint students with

- Basic Concepts in Psychometrics
- Multivariate methods and structural equation modeling.
- Applications of multivariate methods, structural equation modeling and other statistical methods in the analysis of problems related Psychometric study.

Learning outcome: At the end of the Course the student will be able to

- Apply multivariate methods for the analysis of data related to Psychometrics and interpret the outcomes.
- Apply the structural equation methods to identify the hidden factors in the analysis of Psychometric problems.

Unit I:

Theory of data, Metric properties and the problems of scale, Principal component analysis: Extraction of Components-properties and characteristics of components-total variation, relative importance, standardization of variables and components, special covariance structures- interpretation of principal components.

(12 hours)

Unit II:

Factor analysis: orthogonal factor model- parameter estimation by principal component and maximum likelihood method factor rotation- factor scores.

(11 Hours)

Unit III:

Simplex structures and item difficulty, Types and measures of Reliability, item response theory, types and measures of validity, scale construction(types and methods).

(11 Hours)

Unit IV:

Structural equation modeling, goodness of fit and problems with SEM.

(08 Hours)

References:

1. Ben Hui Liu, Statistical Genomics: Linkage. mapping and QTL analysis. CRS press LLC;1998
2. Lesk AM, Introduction to Bioinformatics. Oxford University Press Inc; 2002
3. Raiph Singleton W , Elementary Genetics. D. Van Nostrand Company;1962
4. Alvin C Rencher, Methods of multivariate analysis. 2nd ed. Wiley interscience ;2002
5. Tenko Raykov & George A Marcoulides, An introduction to applied multivariate analysis. Taylor& Francis group; 2008
6. Jurg Olt, Analysis of human genetic Linkage, John Hopkin's University Press. Baltimore
7. DS Falconer, Introduction to Quantitative Genetics. ELBS ; 2004
8. Winchester AM, Genetics. Boson: Houghton MIFFLIN company; 1977

Soft Core	FS02BT-3C5.1**a: TIME SERIES ANALYSIS**	**No. of credits : 3 (42L+28P)**

Learning Objectives: To Acquaint the student with:

- The concept of time dependent phenomenon and use of time series models in biomedical and health science research.
- Various components of time series, their estimation and elimination.
- Identification and analysis of appropriate time series model for the biological data.
- Diagnostic checking of time series models and forecasting.

Learning Outcome: At the end of the Course the student will be able to

- Identify appropriate time series model for the health science data and carry out diagnostic check.
- Develop improved model depending upon the biomedical problem for better forecasted values.
- Take up case studies and choose an appropriate time series model for the analysis.

Unit-I

Introduction to time series. Examples from biomedical studies. Simple Descriptive Techniques: time series plots, trend, seasonal effect. Auto covariance and autocorrelation function and their properties. Partial autocorrelation function. Exploratory time series

analysis. Tests for trend and seasonality.
Elimination of trend and seasonality from the observed time series. Smoothing the time series

(10 Hours)

Unit-II

Stationary processes: White noise model, Moving Average (MA), Autoregressive (AR), ARMA and l ARIMA models. Invertibility. ACF and PACF of these models. Sample ACF and PACF with examples from health science. Distribution of sample ACFand PACF

(14 Hours)

Unit-III

Model Building: Estimation of mean, autocovariance function and autocorrelation function. Model identification. Yule-Walker equations, estimation of AR and MA model. Seasonal ARIMA models.

(10 Hours)

Unit-IV

Forecasting: Forecast mean square error (FMSE), Least squares prediction. Box-Jenkins forecasting. Forecasting through exponential smoothing and Holt-Winters smoothing. Residual analysis and diagnostic checking. Nonstationary time series models and their identification. Unit root test.

(08 Hours)

References:

1. Box GEP and Jenkins G.M, Time Series Analysis: Forecasting and Control. Holden-day. San Francisco; 1976
2. Brockwell P.J and Davis R.S, Introduction to Time Series and Forecasting. 2nd Ed. Springer; 2002
3. Chatfield C, The Analysis of Time Series An Introduction. Chapman & Hall; 1996
4. Dilip M. Nachane Econometrics- Theoretical Foundations and Empirical Perspectives. OUP India; 2006
5. Janacek G, Practical Time Series Arnolds Texts in Statistics; 2001
6. Kendall M.G. and Ord J.K, Time Series. 3rd Ed. Edward Arnold; 1990
7. Montgomory D.C. and Johnson L.A, Forecasting and Time Series Analysis. McGraw Hill. New York; 1977
8. K. Tanaka, Time Series Analysis. Wiley Series; 1996

Soft Core	FS02BT-3C5.2 b : NON-PARAMETRIC REGRESSION	No. of credits: 3 42L +14P

Learning Objectives: Acquaint the student with

- Basic idea of smoothing and different nonparametric regression methods.
- Fitting nonparametric regression models for the biological data.

Learning Outcome: At the end of the Course the student will be able to

- Work on the application of nonparametric regression methods for the biostatistical analysis of health science problems.
- Choose an appropriate nonparametric method for the analysis of the data on hand.
- Apply lack of fit tests for the suitability of the non parametric method.

Unit-I

Nature and scope of nonparametric regression: Basic idea of smoothing, Smoothing histograms and nonparametric probability density function. Random design and fixed design model. Bin smoothers, running mean and running line smoothers. Univariate Kernel density estimation. Local regression estimate.

(10 Hours)

Unit-II

Kernel Regression: Kernel smoothing, Local linear kernel estimators, Local polynomial Kernel estimators, Properties of kernel estimators. Kernel estimators of derivatives, computational aspects of Kernel smoothing, K-nearest neighbor (KNN) estimates. Computational aspects of K-NN estimators.

(12 Hours)

Unit-III

Spline Smoothing: Roughness penalties, quantifying roughness of a curve, cubic splines, computational aspects of natural cubic splines, Orthogonal series estimators. Wavelet method.

Introduction to nonlinear regression models, fitting quadratic, cubic and polynomial regression models.. Introduction to growth curves, logistic and Gomperz growth curve. Introduction to hurdles, zero inflated regression models, Zero inflated Poisson and Hurdle Poisson Mixed models.

(10 Hours)

Unit-IV

Lack of Fit Tests: Testing the fit of a linear model. Lack of fit tests based on linear smoothers: Smoothing residuals, comparing parametric and nonparametric models, Introduction to additive models. Semi parametric regression models.

(10 Hours)

References:

1. Clive Loader, Local Regression and Likelihood. Springer; 1999
2. Hardle, Applied Non-parametric Regression. Cambridge University Press; 1990
3. Hart J.D, Non-parametric Smoothing and Lack of Fit Tests. Springer Verlag; 1997
4. Hastie T.J. and Tibshirani R.J, Generalised Additive Models. Chapman & Hall; 1990
5. John Fox, Nonparametric Sample Regression. Sage Publications; 2000
6. Takezawa K, Introduction to Non-parametric Regression - Wiley Series in Probability and Statistics. John Wiley and Sons; 2005
7. Wand and Jone, Kernel Smoothing. Chapman & Hall; 1995

Hard Core	**FS02BT-3P1: Practical IV:** **Data management and Statistical Computing III** **(Based on all the core theory papers)**	**No. of credits: 3** **(84P)**

Learning Objective: To provide the students hands on working knowledge on

- Use of statistical software like SPSS, STATA and Python.
- Selection of appropriate design for Phase 1, 2,3, and 4 clinical trials and its analysis.
- Epidemiological studies and Hierarchical linear and hierarchical generalized models.
- Genetic epidemiology, Bioinformatics and Psychometrics.

Learning outcome: At the end of the course the students will be able to

- Carry out statistical analysis of biomedical data on clinical trials and Statistical genetics.
- Analyse the data related to biomedical study with the help of statistical software such as SPSS, STATA and Python.

Examples from Genetic epidemiology, Bioinformatics, Psychometrics and health related problems are considered chosen for the practical exercise

Nonparametric Regression

1. Density Estimation through histogram functions for univariate data.
2. Kernel density estimation for univariate data
3. Nadaraya–Watson Kernel Regression estimator
4. Local linear Kernel Regression.
5. Spline smoothing
6. Fitting polynomial regression model
7. Zero inflated models.

Time Series Analysis

1. Estimation and elimination of trend component and Seasonal Component
2. Examining Stationarity of the time series. Sample ACF and PACF.
3. Model Identification and estimation of moving average (MA), Autoregressive(AR) and ARMA process.
4. Residual Analysis and Diagnostic checking.
5. Identification of ARIMA (p d q) process and order selection.
6. Goodness of fit of the model based on AIC and Ljung-Box criteria.
7. Seasonal ARIMA models
8. MMSE Forecasting.

Survival Analysis:

1. Estimation of Survival function
2. Nonparametric methods used in survival analysis
3. Parametric regression models in survival analysis
4. Construction of life table
5. Cox proportion Hazard model.

Fourth Semester M.Sc., Biostatistics

Subject	Details	Credit
FS02BT-4C1	Project	24
Total		24

The aim of this project is to make the students understand issues involved in designing a study-data collection, data management, analysis, interpretation and reporting. This posting is expected to make students confident to independently tackle any real life problems.

At the end of the fourth semester the students need to submit two copies of their project report in the format prescribed by the department. Also, the students have to submit a copy of their internship report along with the project report to the department.

****