Course Title: Applied Statistics **Credit: 3**
Course No: CSIT.221 **Number of period per week: 3+3**
Nature of the Course: Theory+Lab **Total hours: 45+45**
Year: Second
Level: B.Sc.CSIT

1. Course Introduction

This course covers applied statistics for computer and information technology, which makes students able to understand the scope of applied statistics including non-parametric tests, correlation and regression models, sample survey, survey techniques, design of experiment and inferential statistics.

2. Objectives

At the end of this course the students should be able:

- To know the scope and concepts of applied statistics.
- to know basic concepts of non-parametric tests, correlation and regression models, sampling survey, survey techniques, design of experiment and inferential statistics, and their applications.
- to apply statistical tools and techniques in rational ways.
- to interpret statistical inferences meaningfully.

3. Specific Objectives and Contents

Specific Objectives	Contents
• Understand the meaning, scope, importance and limitations of applied statistics • Explain the relations of applied statistics with information technology and e-methods.	**Unit I: Concepts of Applied Statistics (01 hr)** 1.1. Scope, importance and limitations of applied statistics 1.2. Relations of applied statistics with information technology and e-methods.
• To know population, sample and need of sampling, distinguished between censuses and sample survey, and sampling and non-sampling errors, and to prepare questionnaires to select sample and determine sample size sampling and non-sampling errors. • To know random sample, and to apply simple random sampling with and without replacement, stratified random sampling and systematic sampling, ratio and regression method of estimation under simple and stratified sampling, cluster sampling multistage sampling, probability proportion to size (pps) sampling, and also to estimate mean and population total and variance.	**Unit II: Sample survey and Sampling Techniques (07 hrs)** 2.1 Concept of population and sample, need of sampling, censuses and sample survey, questionnaire design, sample selection and determination of sample size, sampling and non-sampling errors. 2.2 Definition of a random sample, types of sampling, uses and applications of simple random sampling with and without replacement, stratified random sampling and systematic sampling, ratio and regression method of estimation under simple and stratified sampling, cluster sampling, multistage sampling, probability proportion to size (pps) sampling, estimation mean and population total and variance (proof is not required)

• Compute Karl Pearson's correlation, Spearman rank correlation, Kendal Tau correlation, partial and multiple correlations for real data and interpret them. • To understand OLS, multiple linear regression and their assumptions, and compute coefficient estimation, fitting of first & second degree regression equations, exponential curves, residuals; and to calculate total sum of squares, coefficient of determination and interpret them, test of significance of regression coefficients, coefficient of determination, and analysis of variance (up to three variables). • To fit Cobb-Douglas production function in real data, and to understand Growth model, Logistic regression model, and Autoregressive model of order one, to understand the heteroscedasticity, multicolinearity and autocorrelation.	**Unit III: Correlation and Regression Models (08 hrs)** 3.1 Concept of simple correlation, Karl Pearson's correlation, Spearman rank correlation, Kendal Tau correlation, partial and multiple correlations. 3.2 Meaning of data modelling, Principles of Ordinary Least Squares (OLS), multiple linear regression, assumptions, coefficient estimation, methods of fitting of first and second degree equations, exponential curves, analysis of residuals, Fisher decomposition of total sum of squares, coefficient of determination and its interpretation. Test of significance of regression coefficients and analysis of variance (only application in real data up to three variables). 3.3 Concepts of Cobb-Douglas production function, growth model, logistic regression model, Autoregressive model of order one; fitting of Cobb-Douglas production function, and introduction of heteroscedasticity, multicolinearity and autocorrelation.
• To apply Run test, Sign test, Wilcoxon signed rank test and Kolmogorov-Smirnov test for real data sets. • Use Kolmogorov-Smirnov two sample test, Median test, Mann-Whitney U test, and to test Kruskal-Wallis one way ANOVA. • To measure the association such as Kendall's tau coefficient, Spearman's coefficient, contingency coefficient, coefficient of concordance, Friedman's two way analysis of variance by ranks. • Use chi-square test for independence of attributes and test for goodness of fit to numerical problems.	**Unit IV: Non-parametric test (07 hrs)** 4.1. Needs of applying non-parametric tests, Run test, Sign test, Wilcoxon signed rank test, Kolmogorov-Smirnov test. 4.2. Kolmogorov-Smirnov two sample test, Median test, Mann-Whitney U test, and Kruskal-Wallis one way ANOVA test. 4.3. Measures of association (Kendall's tau coefficient, Spearman's coefficient, contingency coefficient, coefficient of concordance, Friedman's two way analysis of variance by ranks 4.4. Chi-square test for independence of attributes and test for goodness of fit (only numerical problems)
• To know the design of experiments, Analysis of variance, F-statistics, linear model in ANOVA and their applications. • Carryout Analysis of one and two ANOVA with 1 and m observations per cell in fixed effect model. • Compute ANOVA of Completely Randomized Design (CRD), Randomized Block Design (RBD), Latin Square Design (LSD), and also to obtain the missing lot techniques for RBD and LSD for one observation. • Calculate main and interaction effects of 2^2, 2^3 and also analysis table.	**Unit V: Design of Experiments (07 hrs)** 5.1. Need and concepts of design of experiments, Analysis of variance, F-statistics and its applications, linear model in ANOVA. Analysis of one and two ANOVA with 1 and m observations per cell in fixed effect model. 5.2. Fundamental principles of design, Completely Randomized Design (CRD), Randomized Block Design (RBD), Latin Square Design (LSD), and their analysis, Missing lot techniques for RBD and LSD (one observation missing only). 5.3. Concepts of factorial design, 2^2 and 2^3, and compute main and interaction effects of factorial design.

• To know statistical quality control, and its use, importance and purposes. • Compute $\bar{x}$ and R charts and their constructions, to detect lack of control in $\bar{x}$ and R charts, their limits, uses and interpretations. • Compute p-chart and its limit, use and interpretations. • Compute c-chart and its limit, use and interpretation. • To know the acceptance quality level, consumers and producers risks.	**Unit VI: Statistical Quality Control (07 hrs)** 6.1 Concepts of statistical quality control, its use, importance and purposes, and control charts and control limits. 6.2 Control chart for variables, $\bar{x}$ and R charts, construction of $\bar{x}$ and R charts, detecting lack of control in $\bar{x}$ and R charts, limits, uses and interpretations. 6.3 Control chart for attributes, p-chart, its limit, use and interpretation. 6.4 Control chart for number of defects per unit (c-chart), its limit, use and interpretation. 6.5 Acceptance quality level, and consumers and producers risks.
• To understand point & interval estimation, confidence interval for mean and proportion, relationship of sample size with desired level of error. • To estimate parameters of binomial, Poisson and normal distribution using maximum likelihood estimation. Explain the properties of maximum likelihood estimate. Use method of moments and least squares techniques. • To estimate confidence interval and confidence coefficient, and confidence interval of mean, proportion, variance and difference between means. • To know null and alternative hypothesis, type I and type II errors, level of significance, critical value and critical region, p–value, one and two tail test, steps used in testing of hypothesis. • To test one sample case for mean of normal population, test for proportion, test for difference between two means and two proportions, paired sample t-test, two independent sample tests for variances of normal populations,	**Unit VII: Inferential Statistics (08 hrs)** 7.1. Concepts of Point & interval estimation, confidence interval for mean and proportion, relationship of sample size with desired level of error. 7.2. Estimation of parameters, likelihood function and its properties, maximum likelihood estimation of parameters of binomial, Poisson and normal distribution, properties of maximum likelihood estimate, method of moments and method of least squares techniques. 7.3. Interval estimation, confidence interval and confidence coefficient, method for obtaining confidence limits, confidence interval of mean, proportion, variance and difference between means. 7.4. Hypothesis Testing- Types of statistical hypotheses (null & alternative), type I & type II errors, level of significance, critical value and critical region, concept of p–value and use of p-value in hypothesis testing, one & two tail test, steps used in testing of hypothesis, one sample tests for mean of normal population (for known & unknown variance), test for proportion, test for difference between two means and two proportions, paired sample t-test, two independent sample tests for variances of normal populations.

Evaluation System:

Undergraduate Programs				
External Evaluation	**Marks**	**Internal Evaluation**	**Weightage**	**Marks**
End semester examination	60	Assignments	10%	20
(Details are given in the separate table at the end)		Quizzes	10%	
		Attendance	10%	
		Presentation	10%	
		Term papers	10%	
		Mid-Term exam	40%	
		Group work	10%	
Total External	60	Total Internal	100%	20
Full Marks 60+20+20 = 100				

External evaluation

End semester examination: It is a written examination at the end of the semester. The questions will be asked covering all the units of the course. The question model, full marks, time and others will be as per the following grid.

Full Marks: 100, Pass Marks: 50, Time: 3 Hrs

Nature of question	Total questions to be asked	Total questions to be answered	Total marks	Weightage	External exam marks
Group A: Multiple choice*	20	20	20×1 = 20	20%	12
Group B: Short answer type questions	7 questions	6	6×8 = 48	40%	24
Group C:Long answer type questions	3 questions	2	2×16 =32	40%	24
			100	100%	60

*Scoring scheme will not follow negative marking.

Each student must secure at least 50% marks in internal evaluation in order to appear in the end semester examination. Failing to get such score will be given NOT QUILIFIED (NQ) and the student will not be eligible to appear in the end semester examinations.

Practical examination: Practical examination will be taken at the end of the semester. Students must demonstrate the knowledge and skill of the subject matter.

Internal evaluation

Assignment: Each student must submit the assignment individually. The stipulated time for submission of the assignment will be seriously taken.

Quizzes: Unannounced and announced quizzes/tests will be taken by the respective subject teachers. Such quizzes/tests will be conducted twice per semester. The students will be evaluated accordingly.

Attendance in class: Students should regularly attend and participate in class discussion. Eighty percent class attendance is mandatory for the students to enable them to appear in the end semester examination. Below 80% attendance in the class will signify NOT QUALIFIED (NQ) to attend the end semester examination.

Presentation: Students will be divided into groups and each group will be provided with a topic for presentation. It will be evaluated individually as well as group-wise. Individual students have to make presentations on the given topics.

Term paper: Term paper must be prepared by using computer in a standard format of technical writing and must contain the required number of pages. It should be prepared and submitted individually. The

stipulated time for submission of the paper will be seriously taken as one of the major criteria of the evaluation.

Mid-term examination: It is a written examination and the questions will be asked covering all the topics in the session of the course.

Discussion and participation: Students will be evaluated on the basis of their active participation in the classroom discussions.

Instructional Techniques: All topics are discussed with emphasis on real-world application. List of instructional techniques is as follows:

- Lecture and discussion
- Group as well as individual work
- Self study and assignments
- Presentation by students
- Term paper writing
- Quizzes and guest lecture

Students are advised to attend all the classes and complete all the assignments within the specified time period. If a student does not attend the class(es), it is his/her sole responsibility to cover the topic(s) taught during that period. If a student fails to attend a formal exam/quiz/test, there won't be any provision for re-exam. Unless and until the student clears one semester he/she will not be allowed to study in the following semesters.

Laboratory Work

2. Specific Objectives	**Contents of the practical problems**
• To compute correlation coefficients in different situations and interpret the results. • To have a knowledge of fitting models of the given data also check the Heteroscedasticity, Multicolinearity and Autocorrelation, and interpret the results.	1. For given data, compute simple correlation, Karl Pearson's correlation, Spearman rank correlation, Kendal Tau correlation, partial and multiple correlations of the given data and interpret the results. 2. To fit multiple linear regression for first and second degree equations, exponential curves, and compute total sum of squares, coefficient of determination, test of significance of regression coefficients, analysis of variance up to three variables for given data and interpret the results. 3. To fit the Cobb-Dauglas production function, use the concept of Growth model, Logistic regression model, and Autoregressive model of order one to the given data and also to interpret the results, and check the Heteroscedasticity, Multicolinearity and Autocorrelation.
• To develop skills on preparing questionnaires and selecting appropriate sampling techniques and to compute mean and population total and variance.	4. To constructs the questionnaires for the survey, to determine the sample size in a sample survey, and to select appropriate sampling techniques for survey and estimate mean and population total and variance of a given data for different sampling techniques.
• To develop skills of applying non-parametric tests in different problems and interpret the results.	5. At least one numerical problem is carried out for each of the non-parametric tests (Run test, Sign test, Wilcoxon signed rank test, Kolmogorov-Smirnov test, Kolmogorov-Smirnov two sample test, Median test, Mann-Whitney U test, Kruskal-Wallis one way ANOVA test and Kendall's tau coefficient, Spearman's coefficient, contingency coefficient, coefficient of concordance, Friedman's two way analysis of variance by ranks, and also Chi-square test for independence of attributes and test for goodness of fit.
• Carry out one and two	6. Carry out one and two ANOVA and layout the Completely Randomized Design (CRD), Randomized Block Design (RBD) and Latin Square

ANOVA and layout the CRD, RBD and LSD for given data, and to calculate main and interaction effects of the design 2^2, and 2^3 designs for given data.	Design (LSD) for given data. 7. To estimate main and interaction effects of the design 2^2, 2^3 and to carry out problem related to factorial design for given data.
• To develop skills on statistical quality control related problems and use different charts.	8. To construct $\bar{x}$ and R charts, p-chart and c-chart for given data and interpret the results.
• To develop the skills on inferential statistics related problems and carry out the testing in different data in different situations and interpret them.	9. To calculate point & interval estimation, confidence interval and limits for mean, proportion and variance, and use maximum likelihood, moments and least square techniques to estimate the parameters of the distributions for given data. 10. Carryout one sample tests for mean of normal population (for known and unknown variance), test for proportion, test for difference between two means and two proportions, paired sample t-test, two independent sample tests for variances of normal populations.

Note: There will be practical examination after end-semester examination. An exteran examiner will be there for taking viva exam.

Prescribed Text

- Drpaer, N and H. Smith. *Applied Regression Analysis*, 2nd edution, New York, John Wiley & Sons, 1981.
- Hogg & Tanis, *Probability & Statistical Inference*, 6th edition, First Indian reprint, 2002
- Gujarati, D. *Basic Econometrics*, International Edition, 1995.
- Gibbons, J.D. *Nonparametric Statistical Inference*. International Student edition.
- Siegel, S. *Non-parametric Statistics for the Behavioural Sciences*. McGraw-Hill, New York.
- Hollander & Wolf. *Non-parametric Statistical Methods*. Johns Wiley & Sons, New York.
- Mukhopadhyay p. Theory and Methods of Survey Sampling, prentice Hall of India, New Delhi, 1998.
- Montgomery Douglas C. Design and Analysis of Experiments, 5th edition, John Wiley & Sons Inc., 2001.
- Cochran W.G. Sampling Techniques, 3rd edition, John Wiley and Sons Inc. New York, 1977.
- Kempthorane, O. Design and Analysis of Experiments, Wiley EASTERN, New York.
- Desraj, Pramod Chandhok, SamPle Survey Theory, Narosa Publishing House, 1998.

References

- Aryal, T.R. *Fundamental Statistics- Concepts and Practices*, Viddharthee Publication, Pvt. Ltd., 2010.
- Martin, A. *Research Methods, Statistics, IT and e-Methods*. Icon Publication Pvt. Ltd, 2004.
- Aryal, T.R. *Biostatistics-For Biology, Medical and Health Sciences*, Pinnacle Publication, Pvt. Ltd., 2011.
- Harry Frank & Steven C. Althoen. *Statistics Concepts and Applications*. Cambridge University Press (Low price edition), 1995.
- Murray R. Spiegel & Larry J. Stephens. *Statistics (Schaum's outlines),* Tata McGraw-Hill Publishing Company Ltd, New Delhi, India, 2000.
- Kapoor J. N. and H.C. Saxena. *Mathematical Statistics*, S. Chand & Company Ltd., New Delhi, India, 2001.

- Gupta S. C. and Kapoor V. K. *Fundamentals of Mathematical Statistics*, Sultan Chand and Sons, 2007.
- Rohatgi V. K. and Ehsanes Saleh, A. K. MD. *An Introduction to Probability and Statistics*, John Wiley & Sons, 2005.
- Miller and Fruend. *Modern Elementary Statistics*, Pearson Publication, 2007.
- Feller, W. *An Introduction to Probability Theory and its Applications*, Vol. 1, Third edition, John Wiley and Sons, Singapore, 2000.
- Mayer, P. L. *Introductory Probability and Statistical Applications*, second edition, Oxford and IBH Publishing Co. Pvt Ltd, New Delhi, 1970.
- Spiegel, M.R. *Theory and Problems of Statistics*, McGraw Hill Book Company, Singapore, 1992.

Note-

(i) Theory and practice should go side by side.

(ii) At least Excel and SPSS software should be used for data analysis.

(iii) It is recommended 45 hours for lectures and 15 additional hours for tutorial class for the completion of the course in the semester.

(iv) Home works and assignments covering the lecture materials will be given throughout the semester.