University of Pune

STATISTICS

For First Year B. Sc. (Computer Science) Degree Course

(Formerly known as B. C. S. Course)

Syllabus

(To be implemented from Academic Year 2013-14)

Submitted by: Board of Studies, Statistics

1) **Title of the Course:** First Year B. Sc. (Computer Science)

2) **Preamble:** Statistics is a branch of science that can be applied practically in every walk of life. Statistics deals with any decision making activity in which there is certain degree of uncertainty and Statistics helps in taking decisions in an objective and rational way. The student of Statistics can study it purely theoretically which is usually done in research activity or it can be studied as a systematic collection of tools and techniques to be applied in solving a problem in real life.

   In last 5 to 7 years, computers are playing very crucial role in the society. The use of computers has horizontally spread and also penetrated vertically in the society. It has become a part and parcel of common man. Thus there is a huge demand for computer education.

   The University of Pune had done a pioneering work in this area and Three year degree course B. Sc. (Computer Science) of University of Pune (formerly known as B.C.S.) is very popular among the student community and I. T. Industry. This course covers various subjects which are required directly or indirectly for becoming computer professional. Statistics is one such important subject which is required and is extensively used in a vast spectrum of computer based applications. Data Mining and Warehousing, Theoretical Computer Science, Reliability of a computer Programme or Software, Machine Learning, Artificial Intelligence, Pattern Recognition, Digital Image Processing, Embedded Systems are just few applications to name where Statistics can be extensively used.

3) **Introduction:** The syllabus of Statistics for First Year of this course covers basic concepts and terminology in Statistics and covers basic tools and methods required for data analysis. The teachers teaching this syllabus and students should give emphasis on understanding the concepts and ability to apply statistical tools and techniques and not on the theoretical discussion. It is

expected that at the end of the course, a student should be well equipped to learn and apply acquired techniques in computer based applications.

4) Eligibility: 12th Science with Mathematics

Students admitted to F.Y.B.Sc.(C.S.) will be taking this as one of the compulsory course. Admissions to F.Y.B.Sc.(C.S.) will be given as per the selection procedure / policies adopted by the respective college keeping in accordance with conditions laid down by the University of Pune. Reservation and relaxation will be as per the Government rules.

5) Examination:

A) Pattern of examination and of question paper:

For Theory Papers (For Paper I and II):

Internal examination - 20 marks (10 marks for each semester)
Objective type/ short answer questions with maximum 2 marks for each question.

University Examination - 80 marks at the end of the year.
5 questions carrying 16 marks each.
Q1: Attempt all of the following: (2 marks each) (8 sub questions)
Q2, Q3, Q4, Q5: Attempt any four of the following (4 marks each) (any 4 out of 5 or out of 6)

For Practical paper in Statistics (Paper III):

Internal Evaluation of 20 marks -
(i) Statistics Journal &Attendance – 10 marks
(ii) Project Evaluation – 5 marks
(iii) Viva – 5 marks

External Examination of 80 marks – Total Duration 3 hours
(i) Questions based upon spreadsheet – 3 questions (1 question on diagrams) each of 10 marks should be asked. Total Duration – 1 hour, Total marks – 30.
(ii) Questions to be solved manually using scientific calculator – to solve any two questions out of 3 questions of 25 marks each. Total Duration – 2 hours, Total marks – 50.

B) Standard of Passing: In order to pass in the first year theory and practical examination, the candidate has to obtain 40 marks out of 100 in each course. (Minimum 32 marks must be obtained in the University Theory Examination.)

C) ATKT Rules: Not applicable, since Statistics is one of the compulsory courses taken at F.Y. level.

D) Award of Class: Not applicable, since Statistics is one of the compulsory courses taken at F.Y. level.

E) External Students: There shall be no external students.

F) Pattern of question paper: As specified in A)

G) Verification/Revaluation: As per the University rules

6) Structure of the Course:

F. Y. B. Sc.(C.S.) Statistics

Paper	Course Title	Marks	Lectures
Paper - I	Statistical Methods I	100	Three Hours/Week per Paper (Total 36/Paper per term)
Paper - II	Statistical Methods II	100	
Practical Course	Practical Course	100	Three Hours / Week

Medium of Instruction: The medium of instruction for the course shall be English

7) Equivalence of Previous Syllabus: No equivalence required at F. Y. B. Sc. level, the course titles are same as previous syllabus.

8) University Terms: Dates for commencement and conclusion for the first and second terms will be declared by the University authorities. Terms can be kept by only duly admitted students. The term shall be granted only on minimum 75 percent attendance at theory and practical course and satisfactory performance during the term.

9) Course wise Detail Syllabus

Detailed Syllabus for Statistics Paper I (Statistical Methods I)

1.	Data condensation and Graphical methods 1.1 Raw data, attributes and variables, discrete and continuous variables. 1.2 Presentation of data using frequency distribution and cumulative frequency distribution. (Construction of frequency is not expected) 1.3 Graphical Presentation of frequency distribution –histogram, stem and leaf chart, less than and more than type ogive curves. 1.4 Numerical problems related to real life situations.	5

2.	Review/Revision of Descriptive Statistics 2.1 Measures of Central tendency: Mean, Mode, Median. Examples where each one of these is most appropriate. 2.2 Partition values: Quartiles, Box-Plot. 2.3 Measures of Dispersion: Variance, Standard Deviation, Coefficient of Variation. (Section 2.1 to 2.3 should be covered for raw data, ungrouped frequency distribution and exclusive type grouped frequency distribution)	7
3.	Moments 3.1 Raw and Central moments: definition, computations for ungrouped and grouped data (only up to first four moments). 3.2 Relation between raw and central moments upto fourth order. 3.3 Numerical problems related to real life situations.	3
4.	Measures of Skewness and Kurtosis 4.1 Concept of symmetric frequency distribution, skewness, positive and negative skewness. 4.2 Measures of skewness-Pearson's measure, Bowley's measure, β_1, γ_1. 4.3 Kurtosis of a frequency distribution, measure of kurtosis(β_2,γ_2) based upon moments, type of kurtosis: leptokurtic, platykurtic and mesokurtic. 4.5 Numerical problems related to real life situations.	4
5.	Discrete Random variable 5.1 Definition of random variable and discrete random variable. 5.2 Definition of probability mass function, distribution function and its properties. 5.3 Definition of expectation and variance, theorem on expectation. 5.4 Determination of median and mode using p.m.f. 5.5 Numerical problems related to real life situations.	8
6.	Standard Discrete Distributions 6.1Discrete Uniform Distribution: definition, mean, variance. 6.2 Bernoulli Distribution: definition, mean, variance, additive property. 6.3 Binomial Distribution: definition, mean, variance, additive property. 6.4 Geometric Distribution (p.m.f $p(x) = pq^x$, x = 0,1,2.....): definition, mean, variance. 6.5 Poisson Distribution: definition, mean, variance, mode, additive property, limiting case of B(n, p) 6.6 Illustration of real life situations. 6.7 Numerical problems related to real life situations.	15
7.	Correlation (for bivariate raw data) 7.1 Bivariate data, Scatter diagram. 7.2 Correlation, Positive Correlation, Negative Correlation, Zero Correlation 7.3 Karl Pearson's coefficient of correlation (r), limits of r ($-1 \leq r \leq 1$), interpretation of r, Coefficient of determination (r^2), Auto-correlation upto lags 2. 7.4 Numerical Problems.	6

8	Regression (for ungrouped data) 8.1 Regression: illustrations, appropriate situations for regression and correlation. 8.2 Linear Regression. 8.3 Fitting of straight line using least square method. 8.4 Properties of regression coefficients: $b_{xy}.b_{yx} = r^2$, $b_{yx}.b_{xy} < 1$, $b_{yx} = r(\sigma_y/\sigma_x)$ and $b_{xy} = r(\sigma_x/\sigma_y)$ 8.5 Non Linear regression models: second degree curve, growth curve models. i) $Y = ae^{bX}$ ii) $Y = ab^X$ iii) $Y = aX^b$ iv) logistic model $Y = k / (1+e^{a+bx})$ 8.6 Residual plot, mean residual sum of squares (m. s. s) 8.7 Numerical problems related to real life situations.	9
9	Multiple and Partial Correlation and Regression (for trivariate data) 9.1 Yule's notation and concept of multiple regression. 9.2 Fitting of multiple regression plane. 9.3 Partial regression coefficient, interpretation. 9.4 Multiple correlation coefficient, concept, definition, computation and interpretation. 9.5 Partial correlation coefficient, concept, definition, computation and interpretation.	8
10	Time Series 10.1 Meaning and Utility. 10.2 Components of Time Series. 10.3 Additive and Multiplicative models. 10.4 Methods of estimating trend: moving average method, least squares method and exponential smoothing method. 10.5 Elimination of trend using additive and multiplicative models. 10.6 Simple time series models: AR (1), AR (2). 10.7 Numerical problems related to real life situations.	7
	Syllabus for 1st term is upto Binomial Distribution in Topic 6.	

Detailed Syllabus for Statistics Paper II (Statistical Methods II)

1	Detailed Review / Revision of Theory of Probability 1.1 Counting Principles, Permutation, and Combination. 1.2 Deterministic and non-determination models. 1.3 Random Experiment, Sample Spaces (finite and countably infinite) 1.4 Events: types of events, Operations on events. 1.5 Probability - classical definition, probability models, axioms of probability, probability of an event. 1.6 Theorems of probability (with proof) i) $0 \leq P(A) \leq 1$ ii) $P(A) + P(A') = 1$ iii) $P(A) \leq P(B)$ when $A \subset B$ iv) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ 1.7 Numerical problems related to real life situations.	5

2	Advanced Theory of Probability 2.1Concepts and definitions of conditional probability, multiplication theorem $P(A \cap B)=P(A).P(B\|A)$ 2.2 Bayes' theorem (without proof) 2.3 Concept of Posterior probability, problems on posterior probability. 2.4 Definition of sensitivity of a procedure, specificity of a procedure. Application of Bayes' theorem to design a procedure for false positive and false negative. 2.5 Concept and definition of independence of two events. 2.6 Numerical problems related to real life situations.	12
3	Continuous Random Variable 3.1 Definition of continuous random variable (r. v.), 3.2 Probability density function (p.d.f.), 3.3 Cumulative distribution function (c.d.f.), its properties. 3.4 Calculation of mean, mode, median, variance, standard deviation for continuous r. v. 3.5 Numerical problems related to real life situations.	6
4	Standard Continuous Probability Distributions 4.1 Uniform Distribution: statement of p.d.f., mean, variance, nature of probability curve. 4.2 Exponential Distribution: statement of p.d.f. of the form, $f(x) = (1/\theta)\, e^{(-x/\theta)}$, mean, variance, nature of probability curve, lack of memory property. 4.3 Normal Distribution: statement of p.d.f., identification of parameters, nature of probability density curve, standard normal distribution, symmetry, distribution of aX+b, aX+bY+c where X and Y are independent normal variables, computations of probabilities using normal probability table, normal approximation to binomial and Poisson distribution , central limit theorem (statement only), normal probability plot. 4.4 Pareto Distribution: p.d.f. of the form $f(x) = \frac{\propto}{x^{\propto+1}}$, $x \geq 1, \propto > 0$, mean, variance, applications. 4.5 Numerical problems related to real life situations.	13
	End of First term.	
5	Concepts and definitions related to testing of hypothesis 5.1Definitions: population, statistic, SRSWR, SRSWOR, random sample from a probability distribution, parameter, statistic, standard error of estimator. 5.2 Concept of null hypothesis and alternative hypothesis, critical region, level of significance, type I and type II error, one sided and two sided tests, p-value.	5

6	Large Sample Tests 6.1 H_o: $\mu = \mu_o$ Vs H_1: $\mu \neq \mu_o$, $\mu < \mu_o$, $\mu > \mu_o$ (One sided and two sided tests) 6.2 H_o: $\mu_1 = \mu_2$ Vs H_1: $\mu_1 \neq \mu_2$, $\mu_1 < \mu_2$, $\mu_1 > \mu_2$ (One sided and two sided tests) 6.3 H_o: $P = P_o$ Vs H_1: $P \neq P_o$, $P < P_o$, $P > P_o$ (One sided and two sided tests) 6.4 H_o: $P_1 = P_2$ Vs H_1: $P_1 \neq P_2$, $P_1 < P_2$, $P_1 > P_2$ (One sided and two sided tests) 6.5 Numerical problems related to real life situations.	7
7	Tests based on t-distribution 7.1 H_o: $\mu = \mu_o$ Vs H_1: $\mu \neq \mu_o$, $\mu < \mu_o$, $\mu > \mu_o$ (One sided and two sided tests) 7.2 H_o: $\mu_1 = \mu_2$ Vs H_1: $\mu_1 \neq \mu_2$, $\mu_1 < \mu_2$, $\mu_1 > \mu_2$ (One sided and two sided tests) 7.3 Paired t-test. 7.4 Test of significance of correlation coefficient for bivariate raw data. 7.5 Test of significance of regression coefficients for bivariate raw data. 7.6 Numerical problems related to real life situations.	8
8	Test based on Chi-square distribution 8.1 Chi square test for goodness of fit 8.2 Test for independence of attributes (m X n contingency table) 8.3 Test for significance of variation for a population. 8.4 Numerical problems related to real life situations.	3
9	Non parametric tests 9.1 Run test 9.2 Sign test. 9.3 Kolmogrov - Smirnov test 9.4 Mann – Whitney test 9.5 Numerical problems related to real life situations.	6
10	Simulation 10.1 Introduction to Simulation, merits and demerits and pitfall. 10.2 Pseudo-random number generator ,requisites of a good random number generator, Testing these requirements by using various test of hypothesis using Run test, goodness of fit test, Sign test etc. 10.3 Model Sampling from uniform and exponential distribution. 10.4 Model sampling from Normal distribution using Box-Muller transformation. 10.5 Numerical problems related to real life situations.	7

Detailed Syllabus for Statistics Paper III (Practical)

A) Practicals to be done manually using scientific calculator

1	Measures of Central Tendency and Dispersion.
2	Problems on simple probability, conditional probability, Baye's theorem and independence of events.
3	Measures of skewness and kurtosis

4	Correlation and Linear Regression Analysis. (for bivariate raw data)
5	Fitting of second degree and exponential type models. (for bivariate raw data)
6	Multiple and Partial Correlation and Regression Analysis. (for trivariate data) + Using spreadsheet with use of readymade function.
7	Time Series (Moving Average and Fitting of AR(1) and AR(2) models).
8	Fitting of Binomial and Poisson distributions.
9	Fitting of Normal Distribution.
10	Model Sampling from Simple Continuous Distributions
11	Large Sample Tests.
12	Tests based upon t distribution.
13	Tests based upon chi square distribution.
14	Non parametric tests.

B) Practicals to be done using any spreadsheet (like MS-Excel in MS-Windows or Open-Office in Linux etc.)

1	Diagrammatic Representation and Descriptive Statistics for raw data
2	For a bivariate raw data, fitting various models and finding the “best fit”. (3 problems to be solved in a slot)
3	Fitting of Geometric Distribution and Normal Distribution
4	Using random numbers, drawing of a sample form exponential distribution, normal distribution (Box Muller Transformation) etc.

C) Project –

Project is compulsory which is equivalent to 2 practicals.
Project will carry 5 marks as part of internal evaluation.
One project should be given to one practical batch of students.
The formal project report should be prepared by each student and it must be attached in Statistics journal.

10) Recommended books

Author Name	**Year of Publication**	**Title**	**Publisher**
Medhi J.	1992	Statistical Methods (An Introductory Text)	New Age International
Freund J.E.	2005	Modern Elementary Statistics	Pearson Publication
Trivedi K.S.	2001	Probability, Statistics, Design of Experiments and Queuing Theory with Applications of Computer Science	Prentice Hall of India, New Delhi

Gupta S. C.and Kapoor V. K.	1987	Fundamentals of Applied Statistics (3rd Edition)	S. Chand and Sons, New Delhi.
Ross S. M.	2006	A First Course In Probability 6th Edition	Pearson publication
Law A. M. and Kelton W. D.	2007	Simulation Modelling and Analysis	Tata McGraw Hill
Box G. E. P. and Jenkins G. M.	2008	Time Series Analysis, 4^{th} edition	Wiley
Brockwell P. J. and Davis R. A.	2006	Time Series Methods	Springer
Snedecor G. W. Cochran W. G.	1989	Statistical Methods	John Wiley & sons
Kulkarni M.B., Ghatpande S.B.,Gore S.D.	1999	Common Statistical Tests	Satyajeet Prakashan, Pune
Kulkarni M.B., Ghatpande S.B.	2007	Introduction to Discrete Probability and Probability Distributions	SIPF Academy
Sarma K.V.S.	2001	Statistics Made Simple. Do it Yourself on P.C.	Prentice Hall

11) Qualification of Teacher: As per the University rules