Savitribai Phule Pune University

(Formerly University of Pune)

Faculty of Science & Technology

F.Y.B.Sc. (Computer Science) Statistics

Choice Based Credit System Syllabus

To be implemented from Academic Year 2019-2020

Title of the Course: B. Sc. (Computer Science) STATISTICS

Preamble of the Syllabus:

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 asystematic collection of tools and techniques to be applied in solving a problem inreal life.

In last 15 to 20 years, computers are playing very crucial role in the society. Theuse of computers has horizontally spread and also penetrated vertically in thesociety. It has become a part and parcel of common man. Thus there is a hugedemand 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 is very popular among the student community and I. T. Industry. This course covers various subjects which are required directly or indirectly forbecoming 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, Big Data Analytics, Theoretical Computer Science, Reliability of a computer Program or Software, Machine Learning, Artificial Intelligence, Pattern Recognition, Digital Image Processing, Embedded Systemsare just few applications to name where Statistics can be extensively used.

Introduction: The syllabus of Statistics for First Year of this course covers basic concepts and terminology in Statistics and covers basic tools and methodsrequired 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 isexpected that at the end of the course, a student should be well equipped tolearn and apply acquired techniques in computer based applications.

Structure of the Subject

Structure of the subject and the pattern of examination and question papers are as specified below.

Structure of F. Y. B. Sc. (Computer Science)Statistics

Semester	Paper code	Paper	Paper title	credits	Marks		
					CIA	ESE	Total
1	**CSST 111**	**I**	Descriptive Statistics I	**2**	**15**	**35**	**50**
	CSST 112	**II**	Mathematical Statistics	**2**	**15**	**35**	**50**
	CSST113	**III**	Statistics Practical Paper I	**1.5**	**15**	**35**	**50**
2	**CSST121**	I	Methods of Applied Statistics	**2**	**15**	**35**	**50**
	CSST122	**II**	Continuous Probability Distributions and Testing of Hypothesis	**2**	**15**	**35**	**50**
	CSST123	**III**	Statistics Practical Paper II	**1.5**	**15**	**35**	**50**

Semester I

Paper-I

CSST 111 :Descriptive Statistics

No. of Credits :2No. of lectures: 40

TOPICS/CONTENTS:

UNIT1: Data Condensation and Presentation of Data (9L)

1.1 Definition, importance, scope and limitations of statistics.

1.2 Data Condensation: Types of data (Primary and secondary), Attributes and variables, discrete and Continuous variables.

1.3 Graphical Representation: Histogram,Ogive Curves, Steam and leaf chart. [Note: Theory paper will contain only procedures. Problems to be included in practical]

1.4 Numerical problems related to real life situations.

UNIT2: Descriptive Statistics(14L)

2.1 Measures of central tendency:Concept of central tendency, requisites of good measures of central tendency.

2.2 Arithmetic mean: Definition, computation for ungrouped and grouped data, properties of arithmetic mean (without proof) combined mean, weighted mean, merits and demerits.

2.3 Median and Mode: Definition, formula for computation for ungrouped and grouped data, graphical method, merits and demerits. Empirical relation between mean, median and mode (without proof)

2.4 Partition Values: Quartiles, Box Plot.

2.5 Concept of dispersion, requisites of good measures of dispersion, absolute and relative measures of dispersion.

2.6 Measures of dispersion : Range and Quartile Deviation definition for ungrouped and grouped data and their coefficients, merits and demerits,

Variance and Standard deviation: definition for ungrouped and grouped data, coefficient of variation, combined variance & standard deviation, merits and demerits.

2.7 Numerical problems related to real life situations.

UNIT3: Moments, Skewness and Kurtosis (10L)

3.1 Concept of Raw and central moments: Formulae for ungrouped and grouped data (only first four moments), relation between central and raw moments upto fourth order. (without proof)

3.2 Measures of Skewness: Types of skewness, Pearson's and Bowley's coefficient of skewness, Measure of skewness based on moments.

3.3 Measure of Kurtosis: Types of kurtosis, Measure of kurtosis based on moments.

3.4 Numerical problems related to real life situations

UNIT4:Theory of Attributes (7L)

4.1 Attributes: Concept of a Likert scale, classification, notion of manifold classification, dichotomy, class- frequency, order of a class, positive classfrequency, negative class frequency, ultimate class frequency, relationship among different class frequencies (up to two attributes), 4.2 Consistency of data upto 2 attributes.

4.3 Concepts of independence and association of two attributes.

4.4 Yule's coefficient of association (Q), $-1 \leq Q \leq 1$, interpretation.

References:

1. Statistical Methods, George W. Snedecor, William G, Cochran, John Wiley &sons
2. Programmed Statistics, B.L. Agarwal, New Age International Publishers.
3. Modern Elementary Statistics,Freund J.E. 2005, PearsonPublication
4. Fundamentals of Applied Statistics(3rd Edition), Gupta and Kapoor, S.Chand and Sons, New Delhi, 1987.
5. An Introductory Statistics ,Kennedy and Gentle
6. Fundamentals of Statistics, Vol. 1,Sixth Revised Edition,Goon, A. M., Gupta, M. K. and Dasgupta, B. (1983). The World Press Pvt. Ltd., Calcutta

Semester I

Paper-II

CSST 112 :Mathematical Statistics

No. of Credits : 2 No. of lectures: 40

TOPICS/CONTENTS:

UNIT 1:Theory of Probability (10L)

1.1 Counting Principles, Permutation, and Combination.

1.2 Deterministic and non-determination models.

1.3 Random Experiment, Sample Spaces (Discrete and continuous)

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 (without proof)

i) $0 \leq P(A) \leq 1$ ii) $P(A) + P(A') = 1$ iii) $P(\Phi) = 0$ iv)$P(A) \leq P(B)$ 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.

UNIT 2: Conditional Probability and Independence (8L)

2.1Concepts and definitions of conditional probability, multiplication theorem $P(A \cap B)=P(A).P(B|A)$

2.2 Bayes' theorem (without proof). True positive , false positive and sensitivity of test as application of Bayes' theorem.

2.3 Concept of Posterior probability, problems on posterior probability.

2.4 Concept and definition of independence of two events.

2.5 Numerical problems related to real life situations.

UNIT 3: Random Variable (10L)

3.1 Definition of random variable (r.v.) , discrete and continuous random variable.

3.2 Definition of probability mass function (p.m.f.) of discrete r.v. and Probability density function of continuous r.v..

3.3 Cumulative distribution function (c.d.f.) of discrete and continuous r.v. and their properties. (Characteristic properties only)

3.4 Definition of expectation and variance of discrete and continuous r.v., theorem on expectation and variance (statement only).

3.4 Determination of median and mode using p.m.f. only.

3.5 Numerical problems related to real life situations.

UNIT 4 : Standard Discrete Distributions (12L)

4.1Discrete Uniform Distribution: definition, mean, variance.

4.2 Binomial Distribution: definition, mean, variance, additive property, Bernoulli distribution as a particular case with n = 1.

4.3 Geometric Distribution (p.m.f $p(x) = pq^x$, x = 0,1,2.....): definition,mean, variance.

4.4 Poisson Distribution: definition, mean, variance, mode, additive property, limiting case of B(n, p)

4.5 Illustration of real life situations.

4.6 Numerical problems related to real life situations.

* Only statement of mean and variance, derivation is not expected.

References:

1. A First course in Probability, Sheldon Ross.Pearson Education Inkc.
2. Statistical Methods (An IntroductoryText), Medhi J. 1992 , New Age International.
3. Modern Elementary Statistics ,Freund J.E. 2005, Pearson Publication.
4. Probability, Statistics, Design of Experiments and Queuing Theory with Applications of Computer Science Trivedi K.S. 2001, Prentice Hall of India, New Delhi.
5. Fundamentals of Mathematical Statistics(3rd Edition), Gupta S. C. and Kapoor V. K.1987 S. Chand and Sons, New Delhi.
6. Mathematical Statistics (3rd Edition), Mukhopadhyay P. 2015, Books And Allied (P), Ltd.
7. Introduction to Discrete Probability and Probability Distributions, Kulkarni M.B., Ghatpande S.B. 2007 , SIPF Academy
8. Programmed Statistics, B.L. Agarwal, New Age International Publishers.

Semester I

Paper-III

CSST113: Statistics Practical

No. of Credits : 1.5

TOPICS/CONTENTS

Pre-requisites: Knowledge of the topics in theory papers I and II

Objectives: At the end of the course students are expected to be able

i) To tabulate and make frequency distribution of the given data.
ii) To use various graphical and diagrammatic techniques and interpret.
iii) To compute various measures of central tendency, dispersion, Skewness and kurtosis.
iv) To fit the Binomial and Poisson distributions.
v) To compute the measures of attributes.
vi) The process of collection of data, its condensation and representation for real life data.
vii) To study free statistical softwares and use them for data analysis in project.

Sr. No.	Title of the practical
1	Tabulation and construction of frequency distribution. (Use of at least two data sets more than 50 observations- each for constructing frequency distribution)
2	Diagrammatic and graphical representation using EXCEL and data interpretation. (problems on the basis of SET and NET examination in Paper I to be taken)
3	Summary statistics for ungrouped data and comparison for consistency using EXCEL.
4	Summary statistics for grouped frequency distribution. (Problems based on central tendency, dispersion, measures of skewness: Karl Pearson's and Quartile measure to be covered)
5	Measure of Skewness and kurtosis based on moments.
6	Fitting of Binomial distribution and computation of expected frequencies. (Use the observed and expected frequencies for the next semester χ^2 test)
7	Fitting of Poisson distribution and computation of expected frequencies. (Use the observed and expected frequencies for the next semester χ^2 for test.) (Give one data set for fitting both Poisson and Binomial distributions.)
8	Measure of attributes. (Two attributes only)
9	Study of free statistical softwares and writing a report on it. (individual activity)
10	Project(Part-I) -Data collection, its condensation and representation.

Notes:

1) For project, a group of maximum 8 students be made.
2) All the students in a group are given equal marks for project.
3) Different data sets from primary or secondary sources may be collected.

Semester II

Paper-I

CSST 121 :Methods of Applied Statistics

No. of Credits: 2 No. of lectures: 40

TOPICS/CONTENTS:

UNIT 1:Correlation (For ungrouped data) (10L)

1.1Concept of bivariate data, scatter diagram, its interpretation, concept of correlation, Positivecorrelation, negative correlation, zero correlation.

1.2 Karl Pearson's coefficient of correlation, properties of correlation coefficient, Interpretation of correlation coefficient, coefficient of determination with interpretation.

1.3 Spearman's rank correlation coefficient (formula with and without ties).

1.4Numerical problems

UNIT 2: Regression (for ungrouped data) (12L)

2.1Concept of linear and nonlinear regression.

2.2 Illustrations, appropriate situations for regression and correlation

2.3 Linear regression :Fitting of both lines of regression using least square method.

2.4 Concept of regression coefficients.

2.5 Properties of regression coefficients : $b_{xy} \cdot b_{yx} = r^2$, $b_{xy} * b_{yx} \leq 1$, $b_{xy} = r\,(\sigma_x / \sigma_y)$ and$b_{yx} = r\,(\sigma_y / \sigma_x)$.

2.6 Nonlinear regression models: Second degree curve, exponential curves of the type $Y=ab^x$ and $Y=ax^b$.

2.7 Numerical problems related to real life situations

UNIT3: Multiple Regression and Multiple, partial Correlation (For Trivariate Data)(10L)

3.1 Concept of multiple regressions, Yule's Notations.

3.2 Fitting of multiple regression planes.[Derivation of equation to the plane of regression of X_1on X_2 and X_3 is expected. Remaining two equations to be written analogously.]

3.3 Concept of partial regression coefficients, interpretations.

3.4 Concept of multiple correlation: Definition of multiple correlation coefficientand its formula..

3.5 Concept of partial correlation. Definition of partial correlation coefficient and its formula.

UNIT4: Time series **(8L)**

4.1 Meaning and utility

4.2 Components of time series

4.3 Additive and multiplicative models

4.4 Methods of estimating trend : moving average method, least squares method and exponential smoothing method(with graph and interpretation).

4.5 Numerical problems related to real life situations

References:

1 Introduction to Linear Regression Analysis,Douglas C. Montgomery, Elizabeth A. Peck, G. Geoffrey Vining, Wiley

2 Time Series Methods, Brockwell and Davis, Springer, 2006.

3 Time Series Analysis,4th Edition, Box and Jenkin, Wiley, 2008.

4 Fundamentals of Applied Statistics(3rd Edition), Gupta and Kapoor, S. Chand and Sons, New Delhi, 1987.

5 Fundamentals of Statistics, Vol. 1, Sixth Revised Edition, Goon, A. M., Gupta, M. K. and Dasgupta, B. (1983). The World Press Pvt. Ltd., Calcutta

Semester II

Paper-II

CSST122: Continuous Probability Distributions and Testing of Hypotheses

No. of Credits : 2 No. of lectures: 40

TOPICS/CONTENTS:

UNIT 1:Standard Continuous Probability Distributions (10L)

1.1 Uniform Distribution: statement of p.d.f., mean, variance, nature ofprobability curve. Theorem (without proof): The distribution function of any continuous r.v. if it is invertible follows U(0, 1) distribution

1.2 Exponential Distribution: statement of p.d.f. of the form,$f(x) = (1/\theta)\, e^{(-x/\theta)}$, mean, variance, nature of probability curve, lack ofmemory property.(with proof)

1.3 Paratodistribution :Form of pdf $f(x):\alpha/ x^{(\alpha+1)}$; $x \geq 1$, $\alpha > 0$. Mean, variance, symmetry, applications

1.3 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 areindependent normal variables, computations ofprobabilities usingnormal probability table, normal approximation to binomial and Poissondistribution, central limit theorem (statement only), normal probabilityplot. Box Muller transformation

1.4 Numerical problems related to real life situations.

UNIT 2:Concepts and definitions related to testing of hypothesis (4L)

2.1 Concepts of population and sample.

2.2 Definitions: random samplefrom a probability distribution, parameter, statistic, standard error ofestimator.

2.3 Concept of null hypothesis and alternative hypothesis (Research hypothesis), critical region,level of significance, type I and type II error, one sided and two sided tests, Test of hypothesis,p-value.

UNIT 3:Parametric Tests (20L)

1.1 Large Sample Tests

3.1.1Ho: $\mu = \mu_o$Vs H1: $\mu \neq \mu_o$, $\mu < \mu_o$, $\mu > \mu_o$ (One sided and two sided tests)
3.1.2 Ho: $\mu_1 = \mu_2$ Vs H1: $\mu_1 \neq \mu_2$, $\mu_1 < \mu_2$, $\mu_1 > \mu_2$ (One sided and two sidedtests)
3.1.3 Ho: P = Po Vs H1: P ≠Po, P < Po, P > Po (One sided and two sided tests)
3.1.4 Ho: P1 = P2 Vs H1: P1≠P2, P1 < P2, P1 > P2 (One sided and two sidedtests)
3.1.5 Numerical problems related to real life situations.

3.2 Test based on F- distribution

3.2.1 F-test for testing significance of equality of two population variances.

3.3 Tests based on t – distribution

3.3.1 Ho: $\mu_1 = \mu_2$ Vs H1: $\mu_1 \neq \mu_2$, $\mu_1 < \mu_2$, $\mu_1 > \mu_2$ (One sided and two sided tests)
3.3.2 Paired t-test.

3.4 Tests based on Chi square distribution

3.4.1 Chi-square test for goodness of fit

3.4.2 Test for independence of attributes (mxn and 2x2)

3.5 Numerical problems related to real life situations.

UNIT 4 :Simulation (6L)

4.1 Introduction, concept of simulation , random numbers, pseudo random numbers , Advantages , Disadvantages of Simulation. Applications

4.2 Methods of simulation, Linear congruential generator and simulation from Uniform, Exponential and Normal Distribution.

References

1. A First course in Probability, Sheldon Ross.Pearson Education Inc.
2. Statistical Methods (An IntroductoryText),Medhi J. 1992 , New AgeInternational.
3. Modern Elementary Statistics ,Freund J.E. 2005, PearsonPublication.
4. Probability, Statistics, Design of Experiments and Queuing Theory with Applications of Computer Science,Trivedi K.S. 2001,Prentice Hallof India, NewDelhi.
5.Gupta S. C.andKapoor V. K.1987 Fundamentals of Mathematical Statistics(3rd Edition)S. Chandand Sons,New Delhi.
6. Mukhopadhyay P. 2015, Mathematical Statistics (3rd Edition), Books And Allied (P), Ltd.
7. Simulation Modelling and Analysis Law A. M. and Kelton W.D. 2007,Tata McGraw Hill.
8. Programmed Statistics, B.L. Agarwal, New Age International Publishers.
9. Common Statistical Tests Kulkarni M.B.,Ghatpande, S.B.,Gore S.D.1999 Satyajeet Prakashan,

Semester II

Paper-III

CSST 123: Statistics Practical

No. of Credits : 1.5

Pre-requisites: Knowledge of the topics in theory papers I and II

Objectives: At the end of the course students are expected to be able

i) To understand the relationship between two variables using scatter plot.
ii) To compute coefficient of correlation, coefficient of regression.
iii) To fit various regression models and to find best fit.
iv) To fit the Normal distribution.
v) To understand the trend in time series and how to remove it.
vi) To apply inferential methods for real data sets.
vii) To generate model sample from given distributions.
viii) To understandthe importance and functions of different statistical organizations in the development of nation.

Sr. No.	Title of the Practical
1	Linear correlation and regression (use of scatter plot for explaining the linear relationship between two variables)
2	Fitting of non-linear regression. (use of scatter plot for explaining the non-linear relationship between two variables)
3	Fitting of normal distribution and computation of expected frequencies.
4	Fitting of linear regression model (Simple and Multiple) and non-linear regression models and finding the best fit byusing EXCEL.
5	Modelsampling from continuous uniform, exponential and normal distributions using Excel.
6	Large sample tests.
7	F test, t test, χ^2 test using EXCEL (one problem each with equal and unequal variance)(χ^2 test – for goodness of fit-use fitted problems of Binomial, Poisson and Normal distribution in previous practical problems)
8	Time Series- Estimation of trend by using the method of moving averages
9	Write a report on application of some statistical technique in the field of computers.(individual activity)
10	Project (Part-II) - Analysis of data collected in semester-I

Notes:

i) For project, a group of maximum 8 students be made.
ii) All the students in a group are given equal marks for project.
iii) Students will be asked to use Statistical methods which they have learnt and use of free statistical software for data analysis.