Savitribai Phule Pune University

(Formerly University of Pune)

REVISED SYLLABUS

OF
S. Y. B. A. STATISTICS

(General and Special)

Choice Based Credit System Syllabus

With Effect from June 2020

Preamble:

Statistics is used in different ways in different contexts. For a cricket fan, Statistics is the information about runs scored or wickets taken by a player. For the manager of a manufacturing unit, Statistics may be the information about the production/manufacturing process. For a medical researcher, investigating the effects of a new drug, Statistics is the evidence of research efforts. For a college student, Statistics shows the grades or marks scored in a course. Thus, in all these examples, Statistics refers to quantitative data in the area under study. Statistics as a subject is an important branch of knowledge and is devoted to various techniques of collection, presentation, analysis and interpretation of data. It is a science of learning from data. The subject provides tools for making decisions when conditions of uncertainty prevail. Hence Statistical tools and techniques are used in almost all fields such as agriculture, business, management, economics, finance, insurance, education, sports, biotechnology, medical science, etc. For the last two decades, large amount of data has been collected with the help of computers and more sophisticated statistical techniques are needed for the effective analysis and meaningful conclusions from these data. Knowledge of different aspects of Statistics has become crucial in the present technologically advanced scenario. There is a continuous demand for statisticians in fields of education, industry, software and research. The syllabi of three-year B.A. (special) degree course in Statistics are framed in such a way that the students at the end of the course are thorough in basic statistical techniques and are ready to pursue a Master degree and/or simultaneously able to seek jobs involving statistical analysis related to a variety of data sets in order to arrive at some valid conclusions.

Note: (1) A student of the Three-Year B.A. Degree Course offering 'Statistics' at the special level must offer 'Mathematical Statistics' as a General level subject in all the three years of the course. Further, students of the three-year B.A. Degree Course are advised not to offer 'Statistics' as the General level unless they have offered 'Mathematical Statistics' as a General level subject in all the three years of the course.

(2) A student of three-year B.A. Degree Course offering 'Statistics' will not be allowed to offer 'Applied Statistics' and 'Statistical Pre-requisites' in any of the three years of the course.

(3) A student offering 'Statistics' at the Special level must complete all practicals in Practical Paper to the satisfaction of the teacher concerned.

(4) He / She must produce the laboratory journal along with the completion certificate signed by the Head of the Department at the time of Practical Examination.

(5) Preparation by Internal Examiner for Section I (Online examination/Computer based):

1. Keep at least 15 **computers** with latest configuration ready with battery backup and necessary software, printers, scientific calculators, necessary statistical tables, normal probability paper during the practical examination in the Statistics Laboratory.
2. Appropriate data sets for time series: linear, quadratic, exponential trend fitting, exponential smoothing be entered in spreadsheet.
3. Any other type of data required as per slip also be entered in computer spreadsheet.

(6) Structure of evaluation of practical paper at S.Y.B.A. Statistics:

A) Continuous Internal Assessment (CIA):

Section	Marks
i) Journal	20
ii) Viva-voce	05
iii) Project	05
Total of CIA	30

B) End of Semester Examination (ESE):

Section	Nature	Marks	Time
I	On line examination: Note: Question No.1 is compulsory. Q. 1: Execute the commands and write the same in answer book along with answers using: (A) $Ms - EXCEL$ (For Sem.-III) (B) $R - Software$ (For Sem.-IV)	 10 10	Maximum 30 minutes
II	Using Calculator / Computer Note: Attempt any two of the four questions (each of 25 marks): Q2, Q3, Q4 and Q5:	50	*3 hours And ^2hrs 30
III	Viva-voce	10	10 minutes
	Total of B	**70**	#

Foot note:

* For calculator user

^ For computer user

Total examination time 3 hours 40 minutes for calculator user and 3 hours 10 minutes for computer user.

Instructions to Examiners:

1. Students are not expected to fill data items at the time of examination. They are expected to use and commands (whichever is applicable) to operate on the data set which are already fed.
2. The questions on section I (On line examination Using Excel / R commands (whichever is applicable)) are compulsory and there is no internal option.
3. The slips made available for Section I shall be allotted to the candidates at random so that the total marks of all asked commands will be exactly 10.

(7) Structure of evaluation of theory paper at S.Y.B.A. Statistics (special and general), Applied statistics and Statistical pre-requisites:

A) Continuous Internal Assessment (CIA) for theory subjects:

Section	Marks
i) Theory examination	20
ii) Home assignment	05
iii) Seminar/class test etc.	05
Total of CIA	30

B) End of Semester Examination (ESE) for theory subjects:

Question	Nature	Marks
1	a) Choose correct alternative: i) ii) iii) iv) v) (each with four multiple choice answer A, B, C, D) b) True or false: i) ii) iii) iv) v)	5 5
2	Attempt any four of the following: a) b) c) d) e) f)	20
3	Attempt any four of the following: a) b) c) d) e) f)	20
4	Attempt any two of the following: a) b) c) d)	20
	Total ESE	**70***

Foot note:

* Numerical problem should not exceed 40% of total marks with option questions.

Instruction for Examination:

1. The theory question paper for each paper shall cover all the topics in the pertaining syllabus with proportional weightage to the number of hours of instruction prescribed.
2. The practicals are to be conducted in batches as per the University norms for the faculty of science.
3. Medium of Instruction: English
4. Examination:

A) Pattern of examination: Semester wise

B) Standard of passing: As per norms of University

Objectives:

1. To fit various discrete and continuous probability distributions and to study various real life situations.
2. To identify the appropriate probability model that can be used.
3. To use forecasting and data analysis techniques in case of univariate and multivariate data sets.
4. To use statistical software packages.
5. To test various hypotheses of significance like means, proportions, independence of attributes, variance etc. included in theory (using calculators, software).
6. To compute probabilities of discrete and continuous probability distributions using MS- Excel and/ or R software (whichever is applicable).
7. To study applications of statistics in the field of demography, time series, queueing theory etc.
8. To apply various sampling techniques for conducting real life surveys and draw inferences about parameters.
9. To use various control charts in quality control scheme and able to decide acceptance or rejection of a lot.

S.Y.B.A. STATISTICS Syllabus
For Choice Based Credit System-2019 pattern
To be implemented from the Academic year 2020-2021
Structure of the course:

Table with code and Title for CBCS 2019 pattern:

Subject	Semester III			Semester IV		
	Code	**Title**	**Credit**	**Code**	**Title**	**Credit**
Statistics General-I	ST-23873	Sampling Techniques **(CC 1C)**	3	ST-23874	Statistical Quality Control **(CC 1D)**	3
Statistics Special-I	ST-23883	Continuous Probability Distributions **(DSE 1A)**	3	ST-23884	Sampling Distributions and Inference **(DSE 1B)**	3
Statistics Special-II	ST-23893	Statistics Practical **(DSE 2A)**	3	ST-23894	Statistics Practical **(DSE 2B)**	3
Mathematical Statistics General-II	ST-23273	Discrete Probability Distributions and Time Series **(CC 2C)**	3	ST-23274	Tests of Significance and Statistical Methods **(CC 2D)**	3
Skill Enhancement Course**(SEC)**		**General subject other than Statistics (SEC 1A)**	3		**General subject other than Statistics (SEC 1B)**	3
Skill Enhancement Course**(SEC)**		Data Handling Through MS-Excel **(SEC 2A)**	2		Data Handling Through R-software **(SEC 2B)**	2
Statistical Prerequisites	ST-23573	Applications of Statistics and Theory of Prob. **(SEC 1A)**	3	ST-23574	Theory of Probability Distributions **(SEC 1B)**	3
Applied Statistics	ST-24173	Applications of Statistics and Theory of Probability **(SEC 1A)**	3	ST-24174	Discrete Probability Distributions and Demography **(SEC 1B)**	3

SYBA
SEMESTER -III

Subject: Statistics (General-I)
ST-23873: Sampling Techniques (CC-1C)

Unit 1. Sampling: (14)

1.1 Simple Random Sampling from Finite Population of size(N) with replacement(SRSWR) and without replacement(SRSWOR). Population mean and Population total as parameters, inclusion probabilities.

1.2 (a) Sample mean $\bar{y}$ as an estimator of population mean $\bar{Y}$, derivation of its expectation and standard error.

(b) $N\bar{y}$ as an estimator of population total, derivation of its expectation and standard error of $N\bar{y}$

(c) Estimator of above standard errors, in case of SRSWR and SRSWOR.

1.3 Sampling for proportion as an application of simple random sampling with Y_i taking value zero or one.

(a) sample proportion (p)as an estimator of population proportion(P) of units possessing a certain attribute, derivation of expectation and standard error of (p).

(b) Np as an estimator of total number of units in the population possessing a certain attribute, derivation of expectation and standard error of (Np).

(c) Estimator of above standard error both in case of SRSWR and SRSWOR.

Unit2. Determination of the sample size: (8)

2.1 Determination of the sample size for variable:

(i) Margin of error and confidence coefficient.

(ii) Coefficient of variation and confidence coefficient.

2.2 Determination of the sample size for attributes:

(i) Margin of error and confidence coefficient.

(ii) Coefficient of variation and confidence coefficient.

Unit 3. Stratified Random Sampling: (16)

3.1 Stratification, basis of stratification, real life situation where stratification can be used.

3.2 Stratified sampling as a sample drawn from individual strata by SRSWOR in each stratum

(a) $\bar{y}_{st} = \frac{N_i \bar{y}_i}{N}$ as an estimator of population mean $\bar{Y}$. Derivation of expectation and standard error of $\bar{y}_{st}$

(b) $N\bar{y}_{st}$ as an estimator of population total, derivation of expectation and standard error of $N\bar{y}_{st}$

(c) Estimator of standard errors of (a) and (b)

3.4 Derivation of allocation (n_i)under proportional allocation, optimum allocation with constant function $C = c_0 + \sum_i c_i n_i$ and Neyman's allocation with the expressions for the standard errors when these allocations are used.

3.5 Comparison between these three allocations (Statement Only).

Unit 4. Systematic Sampling (Population size divisible by sample size) (6)

4.1 Meaning of systematic sampling. Method of drawing a sample using systematic sampling, Real life situations where systematic sampling is appropriate

4.2 Estimators of the population mean and population total, standard error of these estimators **(without proof).**

Unit 5. Cluster Sampling: (4)

5.1Introduction to cluster sampling, Real life situations where the cluster sampling is appropriate

5.2 Comparison between systematic sampling and cluster sampling.

References:

1. Arnab, R. (2017). Survey Sampling: Theory & Applications, Academic Press
2. Chaudhuri, A. (2014). Modern Survey Sampling, CRC Press
3. Cochran, W.G. (1984). Sampling Techniques, Wiley.
4. Des Raj and Chandhok, P. (1998). Sample Survey Theory, Narosa
5. Singh, D. and Chaudhary F.S (1986). Theory and Analysis of Sample Survey Designs, Wiley Eastern Limited.
6. Singh, S. (2003). Advance Sampling Theory and Applications (Volume I and II), Kluwer Academic Publishers.
7. Sukhatme, P.V, Suktatme, B.V.,Sukhatme, S. and Asok, C. (1984). Sampling Theory of Surveys with Applications, Indian Society forAgricultural Statistics, New Delhi.
8. Thmpson, S. K. (2012). Samplig, 3rdEdn., Wiley
9. P. Mukhopadhyay: Sampling theory and methods of survey sampling.

SEMESTER -IV

Subject: Statistics (General –I)

ST-23874: Statistical Quality Control (CC-1D)

Unit 1. Statistical Process Control (On line methods): (10)

1.1 Introduction: Concepts: quality, dimensions of quality

1.2 Seven Tools of Statistical Process Control (SPC):

(i) Check Sheet, (ii) Cause and effect diagram (CED), (iii) Pareto Diagram,

(iv) Histogram, (v) Control chart, (vi) Scatter Diagram, (vii) Design of Experiments (DOE).

1.3 Chance causes and assignable causes of variation.

1.4 Statistical basis of control charts (Connection with tests of hypotheses is NOT expected).

1.5 Probability limits, 3σ limits, justification for the use of limits based on Chebychev's inequality and large sample theory.

1.6 Criteria for detecting lack of control:

i) a point outside the control limits with justification

ii) Non-random variation within the control limits of the following type:

(a) A run of seven or more points above or below the control lines.

(b) Presence of trend and cycles. (Mathematical justification is NOT expected for (ii) only).

Unit 2. Control charts for variables: **(10)**

2.1 R chart and $\overline{X}$ chart:

2.1.1: Purpose of R and $\overline{X}$ chart, normal probability plot for checking normality assumption. Construction of R chart when the process standard deviation is specified: control limits, drawing of control chart, plotting of sample ranges, drawing conclusion, determination of state of control process, corrective action if the process is out of control.

2.1.2: Construction of $\overline{X}$ chart when the process average is specified: control limits, drawing of control chart, plotting of sample means. Drawing conclusion, determination of state of control process, corrective action if the process is out of control.

2.1.3: Construction of R chart when the process standard deviation (σ) is not given: control limits, drawing of control chart, plotting sample range values, revision of control limits if necessary, estimate of σ (σ) for future use. Construction of $\overline{X}$ chart when the process average μ is not given: control limits based on σ, drawing of control chart, plotting sample means, revision of control limits of chart $\overline{X}$, if necessary.

***Note* : To find revised control limits of any control chart, delete the sample points above UCL and points below LCL (assuming a search for assignable causes at those points), in case of R and $\overline{X}$ charts, first of all, revisions of control limits of R is to be completed and then by using the observations for which R chart shows the process is under control, the**

control limits for $\overline{X}$ chart should be determined. Revision of control limits of $\overline{X}$ chart be continued without revising the value of $\overline{X}$ Estimate of μ and σ for further use. Determination of state of control of the process. Identification of real life situations where this technique can be used.

2.2 Limitations of $\overline{X}$, R chart:

Unit 3. Control charts for Attributes: (10)

3.1 Construction and working of p – chart:

3.1.1: when subgroup sizes are same and value of the process fraction defective *p* is specified: control limits, drawing of control chart, plotting of sample fraction defectives, revision of control limits if necessary, estimation of *p* for future use. Determination of state of control of the process. Interpretation of high and low spots. Probability of detecting the shift in process fraction defective (or signal) using normal approximation.

3.1.2: when subgroups sizes are different and value of the process fraction defective *p* is not specified (different cases of control limits):

(i) Separate control limits, (ii) control limits based on average sample size,

(iii) Stabilized (standardized P) control limits, drawing of control chart, plotting sample fraction defective, determination of state of control of the process. Identification of real life situations. Limitations of p- chart.

3.2 Construction and working of c – chart:

3.2.1: Construction of c-chart when standard is given:

control limits, justification of 3 sigma limits, drawing of control chart, plotting number of defects per unit.

3.2.2: Construction of c-chart when standard is not given:

control limits, justification for the use of 3 sigma limits, drawing of control chart. Plotting number of defects per unit, revision of control limits, if necessary, estimate of process parameter for future use. Determination of state of control, interpretation of high and low spots in above cases. Identification of real life situations. Probability of detecting shift (or signal) in parameter λ. Comparison between p and C charts. Limitations of c- chart.

Unit 4. Statistical Process Control (Off line methods): (12)

4.1 Concept, comparison between 100 percent inspection and sampling inspection. Procedure of acceptance sampling with rectification – single sampling plan, double sampling plan, Explanation of the terms – producer's risk, consumer's risk, AQL, LTPD, AOQ, AOQL, ASN, ATI, OC and AOQ curves.

4.2 Single sampling plan:

Expressions of Probability of acceptance using:

(i)Hypergeometric (ii) Binomial (iii) Poisson and (iv) Normal distributions. Expressions of the formula of AOQ and ATI, Graphical determination of AOQL, Simple problems to compute probability of acceptance by using single sampling plan

4.3 Double sampling plan:

Expressions of probability of acceptance using Poisson approximation. Statement of ASN and ATI (with complete inspection of second sample). Expressions of the approximate

formula of AOQ. Simple problems to compute probability of acceptance by using double sampling plan.

4.4 Comparison of single sampling plan and double sampling plan.

Unit 5. Capability Studies: (6)

5.1 Specification limits, natural tolerance limits and their comparisons, decisions based on these comparisons, estimate of percent defective.

5.2 Capability ratio and capability indices *(Cp*), capability performance indices *Cpk* with respect to machine and process, interpretation, relationship between (i) *Cp* and *Cpk* (ii) defective parts per million and *Cp*.

References:

1) A. J. Duncan: Quality Control and Industrial Statistics, Taraporewala Sons and Co. Pvt. Ltd., Mumbai.
2) D.C. Montgomery: Statistical Quality Control, John Wiley and Sons, Inc., NewYork.
3) D.H. Besterfield, C.B. Michna etc. (3rd edition 2009): Total Quality Management, Pearson Education, Delhi.
4) E. L. Grant and Leavenworth : Statistical Quality Control, Mc-Graw Hill Kogakusha Ltd., New Delhi.
5) Johnson, N. L. and Kotz : Capability Studies, Chapman and Hall Publishers

SYBA
SEMESTER -III

Subject: Statistics (Special-I)
ST-23883: CONTINUOUS PROBABILITY DISTRIBUTIONS (DSE-1A)

Unit. 1 Continuous Univariate Distributions: (10)

1.1 Continuous sample space: Definition, illustrations.

1.2 Continuous random variable: Definition, probability density function (p.d.f.), cumulative distribution function (c.d.f.), properties of c.d.f. (without proof), probabilities of events related to random variable.

1.3 Expectation of continuous r.v., expectation of function of r.v. $E[g(X)]$, mean, variance, geometric mean, harmonic mean, raw and central moments, skewness, kurtosis, mean deviation about mean.

1.4 Moment generating function (MGF): Definition, properties. Cumulant generating function (CGF): Definition, properties.

1.5 Mode, quartiles (Q_1, Q_2, Q_3)

1.6 Probability distribution of function of r. v. $Y = g(X)$ using i) Jacobian of transformation for $g(.)$ monotonic function and one-to-one, on to functions,

ii) Distribution function for $Y = X^2$, $Y = |X|$ etc., iii) M.G.F. of $g(X)$.

Unit. 2 Continuous Bivariate Distributions: (12)

2.1 Continuous bivariate random vector or variable (X, Y): Joint p. d. f., joint c. d. f, properties (without proof), probabilities of events related to r.v. (events in terms of regions bounded by regular curves, circles, straight lines). Marginal and conditional distributions. Independence of r.v.s X & Y and also its extension to k r.v.s.

2.2 Expectation of of function of r.v. $E[g(X,Y)]$, joint moments, $Cov(X,Y), Corr\ (X,Y)$, conditional mean, conditional variance, $E[E(X|Y=y)] = E(X)\ and\ E[E(Y/X=x) = E(Y)$, regression as a conditional expectation.

2.3 Theorems on expectation:

i) $E(X + Y) = E(X) + E(Y)$, $(ii)\, E(XY) = E(X)\, E(Y), if\ X\ and\ Y$ are independent, generalization to k variables. $E(aX + bY + c), Var\ (aX + bY + c)$ (statement only proof not expected).

2.4 M.G.F.:$M_{X,Y}\ (t_1, t_2)$, properties, M.G.F. of marginal distribution of r.v.s., properties

i) $M_{X,Y}\ (t_1, t_2) = M_X\ (t_1, 0)\ M_Y\ (0, t_2)$ if X and Y are independent r. v.s.,

ii) $M_{X+Y}\ (t) = M_{X,Y}\ (t, t)$

iii) $M_{X+Y} = M_X\ (t)\ M_Y\ (t)$ if X and Y are independent r.v.s.

2.5 Probability distribution of transformation of bivariate r. v. $U = \phi_1(X,Y)$, $V = \phi_2(X,Y)$.

Unit.3 Uniform or Rectangular Distribution: (06)

3.1 Probability density function (p.d.f.)

$$f(x) = \begin{cases} \frac{1}{b-a}, & a \le x \le b \\ 0 & , otherwise \end{cases}$$

Notation : $X \sim U[a,b]$.

3.2 Sketch of p. d. f. , c. d. f. and sketch of c.d.f., mean, variance, symmetry, MGF.

3.3 Distribution of i) $\frac{X-a}{b-a}$,ii) $\frac{b-X}{b-a}$,iii) $Y = F(X)$, where $F(X)$ is the c. d. f. of continuous r.v. X. Application of the result to model sampling. (Distributions of $X+Y, X-Y, XY$ and X/Y are not expected.)

Unit.4 Normal Distribution: (14)

4.1Probability density function (p.d.f.):

$$f(x) = \begin{cases} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} & ; -\infty<X<\infty, -\infty<\mu<\infty, \ \sigma>0 \\ 0 & ; \quad otherwise \end{cases}$$

Notation: $X \sim N(\mu\, \sigma^2)$.

4.2 p. d. f. curve, identification of scale and location parameters, nature of probability curve, mean, variance, MGF, CGF, central moments, cumulants, skewness, kurtosis, mode, quartiles(Q_1,Q_2,Q_3), points of inflexion of probability curve, mean deviation, additive property.

4.3 Probability distribution of : i) $\frac{X-\mu}{\sigma}$,standard normal variable (S.N.V.), ii) $aX+b$, iii) $aX+bY+c$, where X and Y are independent normal variates. Probability distribution of $\bar{X}$, the mean of n $i.i.d.$ $N(\mu,\sigma^2)$ r. v s.,

4.4 Computations of normal probabilities using normal probability integral tables. Central limit theorem (CLT) for $i.i.d.$ r.v.s. with finite positive variance (statement only), its illustration for Poisson and Binomial distributions.

Unit.5 Exponential Distribution: (06)

5.1 Probability density function (p. d. f.):

$$f(x) = \begin{cases} \alpha e^{-\alpha x} ; & x \ge 0, \alpha > 0 \\ 0 ; & otherwise \end{cases}$$

Notation : $X \sim Exp(\alpha)$.

5.2 Nature of density curve, interpretation of α as a scale and $\frac{1}{\alpha}$ as mean, mean, variance, MGF, CGF, skewness and kurtosis.

5.3 c.d.f., graph of c.d.f., lack of memory property, quartiles(Q_1, Q_2, Q_3), mean deviation about mean, additive property.

5.4 Distribution of $min(X,Y)$ and $max(X,Y)$ with X, Y $i.i.d.$ exponential r.v.s.

References:

1. Berger, R. and Casella G. (2002). Statistical Inference, Duxbury Resource Center, Second Edition.
2. Dasgupta, A. (2010) Fundamentals of Probability: A First Course, Springer, New York.
3. Goon, A. M., Gupta, M. K and Dasgupta, B.: Fundamentals of Statistics,Vol. 2,World Press, Calcutta.
4. Hogg, R. V. McKean, J. W. and Craig, T. T. (2005). Introduction to Mathematical Statistics, Sixth Edition, Pearson Prentice Hall, New Jersey.
5. Mood, A. M. Graybill, F. A. and.Boes, F. A Introduction to Theory of Statistics (Third Edition), McGraw – Hill Series G A 276.17
6. S. C. Gupta and V. K. Kapoor : Fundamentals of Mathematical Statistic, Sultan Chand and Sons, 23, Daryaganj, New Delhi 110002.
7. S. P. Gupta : Statistical Methods, Sultan Chand and Sons, 23, Daryaganj, New Delhi 110002.
8. S.C. Gupta and V. K. Kapoor : Fundamentals of Applied Statistics. S.Chand and Sons, New Delhi.

SYBA

SEMESTER -IV

Subject: Statistics (Special-I)

ST-23884: SAMPLING DISTRIBUTIONS AND INFERENCE (DSE-1B)

Unit.1 Gamma Distribution: (6)

1.11Probability density function (p. d. f.):

$$f(x) = \begin{cases} \frac{\alpha\lambda}{\Gamma\lambda} x^{\lambda-1} e^{-\alpha x} & ; x > 0, \ \alpha, \lambda > 0 \\ 0 & ; \ otherwise \end{cases}$$

= 0 , otherwise.

Notation: $X \sim G(\alpha, \lambda)$,

1.2 Nature of probability curve, special cases: i) $\alpha = 1$, ii) $\lambda = 1$,MGF, CGF, moments, cumulants, skewness, kurtosis, mode, additive property.

1.3 Distribution of sum of n i.i.d. exponential variables. Relation between distribution function of Poisson and Gamma variates.

Unit. 2 Chi-square Distribution: (10)

2.1 Definition as a sum of squares of i.i.d. standard normal variables. Derivation of the p.d.f. of Chi-square variable with n degrees of freedom (d.f.) using MGF technique.

2.2 Mean, variance, MGF, CGF, central moments skewness, kurtosis, mode, additive property. Use of chi-square tables for calculations of probabilities. Normal approximation (statement only)

2.3 Distribution of $\frac{nS^2}{\sigma^2} = \frac{1}{\sigma^2}\sum_{i=i}^{n}(X_i - \bar{X})^2$for a sample from a normal distribution using orthogonal transformation, independence of $\bar{X}$ and S^2.

Unit. 3 Student's t −distribution: (8)

3.1Definition of T r.v. with n d.f. in the form of $T = \frac{U}{\sqrt{\frac{V}{n}}}$, where $U \sim N(0,1)$ and V is chi-square with n d.f., where U & V are independent random variables.

Notation: $T \sim t_n$

3.2 Derivation of the p.d.f of t distribution, nature of probability curve, mean, variance, moments, mode.

3.3 Use of t-tables for calculations of probabilities, statement of normal approximation.

Unit.4 Snedecore's F −distribution: (08)

4.1 Definition of F r.v. with n_1 and n_2 d.f. as $F_{n_1,n_2} = \frac{\chi^2_{n_1}/n_1}{\chi^2_{n_2}/n_2}$ Independent chi-square variables with n_1 and n_2 d.f.

Notation: $X \sim F_{n_1,n_2}$

4.2 Derivation of the p.d.f, nature of probability curve, mean, variance, moments, mode.

4.3 Distribution of $\frac{1}{F_{n_1,n_2}}$,use of F −tables for calculation of probabilities.

4.4 Interrelationship between Chi-square, t and F distributions

Unit.5 Test of Hypothesis: (16)

5.1Tests based on chi-square distribution:

a) Test for independence of two attributes arranged in $r \times s$ contingency table, Mc Nemar's test (to be covered in practical only).

b) Test for goodness of fit. (to be covered in practical only)

c) Test for variance against one-sided and two-sided alternatives i) for known mean, ii) for unknown mean.

5.2 Tests based on t −distribution:

a) Tests for population means:

i) one sample with unknown variance and two sample for unknown equal variances tests for one-sided and two-sided alternatives.

ii)$100(1-\alpha)\%$ two sided confidence interval for population mean and difference of means of two independent normal populations.

b) Paired t-test for one-sided and two-sided alternatives.

5.3 Test based on F distribution:

Test for H_0: $\sigma_1^2 = \sigma_2^2$ against one-sided and two-sided alternatives when i) means are known and ii) means are unknown.

References:

1. Goon, A. M., Gupta, M. K. and Dasgupta, B. Fundamentals of Statistics, Vol. 2, World Press, Calcutta.
2. Gupta, S. C. and Kapoor, V. K. Fundamentals of Applied Statistics. S. Chand and Sons, New Delhi.
3. Gupta, S. C.and Kapoor, V. K. Fundamentals of Mathematical Statistic, Sultan Chand and Sons, 23, Daryaganj, New Delhi 110002.

4. Gupta, S. P Statistical Methods, Sultan Chand and Sons, 23, Daryaganj, New Delhi 110002.
5. Hogg, R. V. McKean, J. W. and Craig, T. T. (2005). Introduction to Mathematical Statistics, Sixth Edition, Pearson Prentice Hall, New Jersey.
6. Kulkarni, M. B., Ghatpande, S. B. and Gore, S. D. Common Statistical Tests, Satyajeet Prakashan, Pune 411029.
7. Mayer, P. L. Introductory Probability and Statistical Applications, Addision Weseley Pub. Comp. London.
8. Mood, A.M., Graybill, F. A. and Boes, F. A. Introduction to Theory of Statistics (Third Edition), McGraw – Hill Series G A 276.17
9. Mukhopadhyay, P. Applied Statistics, New Central Book Agency, Pvt. Ltd. Calcutta.
10. Sanjay A. and Bansilal, New Mathematical Statistics (First Edition), Satya Prakashan, 16/7698 New Market, New Delhi – 5.

SYBA
SEMESTER -III

Subject: Statistics Practical (Special-II)
ST-23893: PRACTICALS (DSE-2A)

Pre-requisites: Knowledge of the topics in theory.

Objectives:

1. To fit various discrete and continuous probability distributions and to study various real life situations (Using calculators and $Ms - EXCEL$).
2. To identify the appropriate probability model, that can be used.
3. To use forecasting and data analysis techniques in case of univariate and multivariate data sets.
4. To use statistical software packages.
5. To test the hypotheses particularly about mean, variance, correlation, proportions, goodness of fit.
6. To study applications of statistics in the field of economics, demography etc.

Notes:

1. Students must complete all the practicals to the satisfaction of the teacher concerned.
2. Students must produce the laboratory journal along with the completion certificate signed by the Head of the Department at the time of practical examination
3. Use of computer software whenever possible to be encouraged.

Preparation by Internal Examiner for Section I (Online examination):

1) Keep at least 15 computers with latest configuration ready with battery backup and necessary software at the examination laboratory.
2) Trivariate and bivariate data set of 10 to 20 items be fed in computer MSEXCEL spreadsheet (Trivariate data set for multiple regression plane) before the commencement of examination. Appropriate data set for time series: linear, quadratic, exponential trend fitting, exponential smoothing be entered in spreadsheet.
3) Any other type of data required for time to time also be entered in computer spreadsheet.

Instructions to Examiners:

1. Students are not expected to fill data items at the time of examination. They are expected to use and commands (whichever is applicable) to operate on the data set which are already fed.
2. The questions on section I (On line examination Using Excel based on computer) are compulsory and there is no internal option.
3. The slips made available for Section I shall be allotted to the candidates at random so that the total marks of all asked commands will be exactly 10.

Sr. No.	Title of the Practical	No. of Practicals
1	Fitting of negative binomial distribution and computation of expected frequencies.	1
2	Fitting of normal distribution and computation of expected frequencies.	1
3	Applications of negative binomial and multinomial distributions.	1
4	Applications of exponential and normal distributions.	1
5	Model sampling from normal distribution using i) Distribution function, ii) Box-Muller transformation.	1
6	Estimation of trend by moving averages.	1
7	Estimation and forecasting of trend by exponential smoothing, curve fitting (Linear and second degree), plotting of residuals.	1
8	Fitting of $AR\,(1)$ model	1
9	Estimation of seasonal indices by simple averages and ratio to trend using moving averages	1
10	Estimation of seasonal indices by ratio to linear trend by least square method	1
11	Stratified sampling I and II	2
12	Finding probabilities , Fitting of negative binomial distribution & probability Bar Diagram using $Ms - EXCEL.$	1
13	Finding probabilities, Fitting of normal distribution & normal probability curve tracing using $Ms - EXCEL.$	1
14	Fitting of linear, quadratic, exponential trends to time series data. Finding the best fit using R^2., Moving averages and Exponential smoothing using $Ms - EXCEL.$	2
15	**Project**: Project based on analysis of data collected by students in groups of maximum **6** students. (Project is equivalent to five practical's)	5

SYBA

SEMESTER -IV

Subject: Statistics Practical (Special-II)

ST-23894: PRACTICALS (DSE-2B)

Pre-requisites: Knowledge of the topics in theory.

Objectives:

1. To test various hypotheses of significance like means, population proportions, independence of attributes, variance etc. included in theory (using calculators, R software).
2. To study statistical tools for quality control management through control charts.
3. To compute probabilities of discrete and continuous probability distributions using R software.
4. To use R software for finding basic summary statistics.

Notes:

1. Students must complete all the practicals to the satisfaction of the teacher concerned.
2. Students must produce the laboratory journal along with the completion certificate signed by the Head of the Department at the time of practical examination
3. Use of computer software whenever possible to be encouraged.

Preparation by Internal Examiner for Section I (Online examination):

1) Keep at least 15 computers with latest configuration ready with battery backup and necessary software at the examination laboratory.
2) Any other type of data required for time to time also be entered in computer spreadsheet.

Instructions to Examiners:

1) Students are not expected to fill data items at the time of examination. They are expected to use R commands to operate on data set which are already fed.
2) The question from Section I are compulsory and there is no internal option.
3) The slips made available for Section I shall be allotted to the candidates at random so that the total marks of all asked commands will be exactly 10.

Sr. No.	Title of the Practical	No. of Practicals
1	Computations of GRR and NRR	1
2	Test for proportions and construction of confidence interval	1
3	Test for means and construction of confidence interval i) $H_0: \mu = \mu_0$, σ^2 known and σ^2 unknown ii) $H_0: \mu_1 = \mu_2$, σ_1^2 and σ_2^2 both known iii) $H_0: \mu_1 = \mu_2$, $\sigma_1^2 = \sigma_2^2 = \sigma^2$ unknown iv) $H_0: \mu_1 = \mu_2$ (paired t-test)	2
4	Tests based on χ^2 distribution i) Goodness of fit ii) Independence of attributes (2 × 2, r ×s contingency table) iii) Mc Nemar's test$H_0: \sigma^2 = \sigma_0^2$, μ unknown, confidence interval for σ^2	2
5	Tests based on F distribution $H_0: \sigma_1^2 = \sigma_2^2$for i) means known ii) means unknown	1
6	Control charts for variables ($\overline{X}$ and R chart)	2
7	Control charts for attributes (p-chart) with sample size fixed and sample size variable.	2
8	Control charts for attributes (c-chart)	1
9	Use of basic R software commands c( ), scan( ), rep( ), seq( ), min, max, sort, extract, data. frame, matrix, accessing resident data sets etc.	1
10	Finding summary statistics using summary ( ) and fivenum( ). Calculate arithmetic mean (AM), geometric mean (GM), harmonic mean (HM), median, mode, quantiles, range, quartile deviation (QD), variance, coefficient of variation (CV) using R software.	1
11	Computation of probabilities of negative binomial, multinomial, normal, exponential, gamma, t, χ^2, F using R software	1
12	Tests based on proportions, means, χ^2 distribution, F distribution using R software.Fitting of trivariate regression plane using R software	1
13	**Project**: Project based on analysis of data collected by students (primary data) in groups of maximum **6** students. (Project is equivalent to five practical's)	5

Savitribai Phule Pune University
STATISTICS Syllabus
for S. Y. B. A. (Credit System)
form the Academic Year 2020-2021
Skill Enhancement Courses (SEC) – (2 Credits)

Notes:

1.The University Grants Commission has made it compulsory for students to earn two credits from a Skill Enhancement Course (SEC) in each semester second year onwards.

2.It is mandatory for the student to complete one Skill Enhancement Course (SEC) in each semester from Semester III to Semester VI.

3.It must be noted that student has to choose any one of the four Skill Enhancement Courses (SEC) for each Semester.

4.Each Skill Enhancement Course (SEC) will have two (2) credits only.

Semester III

Data Handling Through MS-Excel (SEC-2A) (2 Credits)

Course Objectives:

1. This course is designed to introduce MS-Excel to the students.
2. It will enable students to understand basic concept of MS-Excel.
3. It will help students to represent the data in pictorial forms.

Course Outcomes:

1. Students will have learned to open the Excel spread sheet and are able to enter the data in worksheet.
2. They will be able to represent data into charts, diagrams, graphs, etc.
3. They can perform various mathematical calculations and can learn the use of excel as calculator.
4. They can perform various statistical calculations.

Course Content:

Unit 1: Introduction to MS-Excel (2)

Ribbon tabs, Ribbon bar, Understanding the worksheet (Rows and Columns, Sheets, Work-books), Active Cell, Columns, Rows, Fill Handle, Address Bar, Formula Bar, Title Bar, File Menu, Quick Access Toolbar, Ribbon Tab, Worksheet Tab, Status Bar.

Unit 2: Microsoft Excel Basic Functions (2)

SUM, COUNT, AVERAGE, TIME, DATE, LEFT, RIGHT, VOOKUP, IF, NOW, etc.

Unit 3: Visualization of Data (4)

Diagrammatic representation of statistical data: simple and subdivided bar diagrams, multiple bar diagram, percentage bar diagram, pie diagram.

Unit 4: Presentation of Data (4)

Graphical representation of statistical data: Histogram, frequency curve and ogive curves. Determination of mode and median graphically.

Unit 5: Statistical Computations (4)

Computation of summary statistics, mean , mode, median , partition values, variance, standard deviation, absolute deviation, maximum, minimum, etc.

Unit 6: Curve Fitting (4)

Scatter diagram, correlation coefficient, fitting of a line of regression, fitting of second degree curve.

Unit 7: Problem solving with MS-Excel (16)

a) Computations using basic mathematical and Statistical functions
b) Diagrammatic representation of data
c) Graphical representation of frequency data
d) Computations of correlation coefficients and curve fitting.

References:

1.Michael Alexander and John Walkenbach (2013), Microsoft Excel Dashboards and Reports, 2nd Edition, Wiley.

2. Greg Harvey (2019). Microsoft Excel 2019 All-in-one for Dummies, Wiley

3. John Walkenbach (2018), Excel 2016 Bible ,Wiley

4. Schmuller, Joseph (2020), Statistical Analysis with Excel, 4th Edition, Wiley

Savitribai Phule Pune University
STATISTICS Syllabus
for S. Y. B. A.(Credit System)
form the Academic Year 2020-2021
Skill Enhancement Courses (SEC) – (2 Credits)

Notes:

1.The University Grants Commission has made it compulsory for students to earn two credits from a Skill Enhancement Course (SEC) in each semester second year onwards.

2.It is mandatory for the student to complete one Skill Enhancement Course (SEC) in each semester from Semester III to Semester VI.

3.It must be noted that student has to choose any one of the four Skill Enhancement Courses (SEC) for each Semester.

4.Each Skill Enhancement Course (SEC) will have two (2) credits only.

Semester IV

Data Handling Through R-software (SEC-2B) (2 Credits)

Course Objectives:

1. This course is designed to introduce R-software to the students, which freeware software.
2. It will enable students to understand basic concept and uses of R-software.
3. It will help students to represent the data in pictorial and graphical forms.
4. It will able the students to perform various statistical analysis tools through R- software.

Course Outcomes:

1. Students have learned the fundamental of R and are able to enter the data in R-worksheet.
2. They will be able to draw charts, diagrams, graphs, etc. using R-software.
3. They can perform various mathematical calculations and can learn the use R as calculator.
4. They can perform various Statistical analysis using R-software.

Course Content:

Unit 1. Fundamentals of R-Software: (6L)

Introduction to R, features of R, starting and ending R session, getting help in R, R commands and case sensitivity. Vectors and vector arithmetic:

a) creation of vectors using functions c, seq, rep

b) Arithmetic operations on vectors using operators +, - , * , / , ^ .

c) Numerical functions: log10, log, sort, max, min, unique, range, length, var, prod, sum, summary, fivenum etc. d) accessing vectors

Data frames: creation using data.frame , subset and transform commands. Resident data sets: Accession and summary

Unit 2. Diagrams: (4 L)

Simple bar diagram, Subdivided bar diagram, multiple bardiagram, Pie diagram, Stem and leaf chart

Unit 3. Graphs: (8 L)

Box plot for one and more than one variables, rod or spike plot, histogram for raw data with prob=T option and for both equal and unequal class intervals, frequency polygon, ogive curves, empirical distribution function Saving the diagram and graph in MS-Word file.

Unit 4. Measures of central tendency, dispersion, skewness and kurtosis: (12 L)

Computations of following measures for all types of data

(a) central tendency mean, mode, median, quartiles, deciles, percentiles, geometric mean and harmonic mean.

(b) Dispersion: variance, standard deviation, coefficient of variation, mean deviation

(c)Skewness: Bowley's coefficient and Karl Pearson's coefficient of skewness

(d) Moments: Computations of raw and central moments, measure of skewness and kurtosis based on moments

Unit 5. Probability distributions: (6 L)

Simulation from distributions, computations of probabilities, cumulative probabilities, quantiles and drawing random sample using d,p,q,r functions for following distributions.

Binomial, Poisson, Hypergeometric, normal, exponential, gamma, uniform.

Fitting of Poisson and normal distribution, testing normality of data by Shapiro-Wilks test.

References:

1. Crawley, M. J. (2006): Statistics - An introduction using R. John Wiley,London
2. Purohit, S.G.; Gore, S.D. and Deshmukh, S.R. (2015): Statistics using R,2nd Edition, Narosa Publishing House, New Delhi.
3. Randall, P. (2016), Foundations and Applications of Statistics: An Introduction using R, American Mathematical Society
4. Shahababa , B. (2011). Biostatistics with R, Springer, New York
5. Verzani, J. (2005). Using R for Introductory Statistics, Chapman and Hall CRC Press, New York