UNIVERSITY OF PUNE

Syllabus for F.Y.B.Sc.

(With effect from Academic Year 2010-2011)

Statistics/Statistical Techniques

Note :

1. A student of three-year B.Sc. degree course will not be allowed to offer **Statistics** and **Statistical Techniques** simultaneously in any of the three years of the course.

2. Students offering Statistics at the First year of the three-year B.Sc. course may be allowed to offer **Statistical Techniques** as one of their subjects in the second year of the three-year B.Sc. in place of **Statistics**.

3. Students offering **Statistical Techniques** at the first year of the three-year B.Sc. course may be allowed to offer **Statistics** as one of their subjects in the second year of the three-year B.Sc. course in place of Statistical Techniques.

4. Students must complete all the practicals to the satisfaction of the teacher concerned.

5. Students must produce at the time of practical examination, the laboratory journal along with the completion certificate signed by the Head of the Department.

6. **Structure of the evaluation of Practical Paper**

   (A) **Continuous internal evaluation** :

   (i) Journal 10 marks

   (ii) viva-voce 10 marks

(B) **Annual practical examination** :

Duration : 3 hours + additional 10 minutes for viva during practical examination.

(i) Questions on MSEXCEL to be performed on computer during examination	10 marks
(ii) Questions based on other practicals to be performed using calculators	60 marks
(iii) Viva-voce	10 marks
Total :	100 marks

F.Y.B.Sc. Statistics/Statistical Techniques

Paper - I : Descriptive Statistics

Objectives :

The main objective of this course is to acquaint students with some basic concepts in Statistics. They will be introduced to some elementary statistical methods of analysis of data.

At the end of this course students are expected to be able,

(i) to tabulate statistical information given in descriptive form.

(ii) to use graphical techniques and interpret.

(iii) to compute various measures of central tendency, dispersion, skewness and kurtosis.

(iv) to compute the correlation coefficient for bivariate data and interpret it.

(v) to analyze data pertaining to attributes and to interpret the results.

(vi) to summarize and analyze the data using computer.

(vii) to apply statistics in the various fields.

1. Introduction to Statistics : (3)

1.1 Meaning of Statistics as a Science.

1.2 Importance of Statistics.

1.3 Scope of Statistics : In the field of Industry, Biological Sciences, Medical Sciences, Economics Sciences, Social, Sciences, Management Sciences, Agriculture, Insurance, Information Technology, Education and Psychology.

1.4 Statistical organizations in India and their functions : CSO, ISI, NSS, IIPS (Devnar, Mumbai), Bureau of Economics and statistics.

2. Population and Sample (4)

2.1 Types of characteristics :

Attributes : Nominal scale, ordinal scale, Variables : Interval scale, ratio scale, discrete and continuous variables, difference between linear scale and circular scale.

2.2 Types of data : (a) Primary data, Secondary data.

(b) Cross-sectional data, time series data, failure data, industrial data, directional data.

2.3 Notion of a statistical population : Finite population, infinite population, homogeneous population and heterogeneous population. Notion of sample, random sample and non-random sample.

2.4 Methods of sampling (Description only) : Simple random sampling with and without replacement (SRSWR and SRWOR) stratified random sampling, systematic sampling, cluster sampling and two-stage sampling.

3. Presentation of Data (4)

3.1 Classification : Raw data and its classification, Discrete frequency distribution, Sturge's rule, continuous frequency distribution, inclusive and exclusive methods of classification, Open end classes, cumulative frequency distribution and relative frequency distribution.

3.2 Graphical Presentation of Data : Histogram, frequency curve, frequency polygon, Ogive curves, stem and leaf chart.

3.3 Check sheet, Parato diagram.

3.4 Examples and Problems.

4. Measures of Central Tendency (10)

4.1 Concept of central tendency of statistical data : Statistical average, characteristics of a good statistical average.

4.2 Arithmetic Mean (A.M.) Definition, effect of change of origin and scale, combined mean of a number of groups, merits and demerits, trimmed arithmetic mean,

4.3 Mode : Definition, formula for computation (with derivation) graphical method of determination of mode, merits and demerits.

4.4 Median : Definition, formula for computation (with derivation) graphical method of determination of median, merits and demerits.

4.5 Empirical relation between mean, median and mode.

4.6 Partition Values : Quartiles, Deciles and Percentiles, Box Plot, Percentile ranks.

Means of transformed data :

4.7 Geometric Mean (G.M.) Definition, merits and demerits.

4.8 Harmonic Mean (H.M.) Definition, merits and demerits.

Order relation between arithmetic mean, geometric mean, harmonic mean (proof for $n = 2$).

4.9 Weighted Mean : Weighted A.M., G.M. and H.M.

4.10 Situations where one kind of average is preferable to others.

4.11 Examples and Problems.

5. Measures of Dispersion (10)

5.1 Concept of dispersion, characteristics of good measure of dispersion.

5.2 Range : Definition, merits and demerits.

5.3 Semi-interquartile range (Quartile deviation).

5.4 Mean deviation : Definition, merits and demerits, minimality property (without proof).

5.5 Mean square deviation : Definition, minimality property of mean square deviation (with proof), Variance and standard deviation : Definition, merits and demerits, effect of change of origin and scale, Combined variance (derivation for 2 groups), Combined standard deviation, generalization for n groups.

5.6 Measures of dispersion for comparison : coefficient of range, coefficient of quartile deviation and coefficient of mean deviation, coefficient of variation (C.V.)

5.7 Examples and Problems.

6. Moments (5)

6.1 Raw moments (m'_r) for grouped and ungrouped data.

6.2 Moments about an arbitrary constant for grouped and ungrouped data $m_r(a)$.

6.3 Central moments (m_r) for grouped and ungrouped data, Effect of change of origin and scale, Sheppard's correction.

6.4 Relations between central moments and raw moments (upto 4-th order).

7. Skewness and Kurtosis (6)

7.1 Concept of skewness of frequency distribution, positive skewness, negative skewness, symmetric frequency distribution.

7.2 Bowley's coefficient of skewness : Proof of Bowley's coefficient of skewness lies between -1 to 1, interpretation using Box plot.

7.3 Karl Pearson's coefficient of skewness.

7.4 Measures of skewness based on moments (β_1, γ_1).

7.5 Concepts of kurtosis, leptokurtic, mesokurtic and platykurtic frequency distributions.

7.6 Measures of kurtosis based on moments, (β_2, γ_2).

7.7 Examples and Problems.

End of First Term

8. Correlation (7)

8.1 Bivariate data, bivariate frequency distribution.

8.2 Concept of correlation between two variables, positive correlation, negative correlation, zero correlation.

8.3 Scatter diagram, conclusion about the type of correlation from scatter diagram.

8.4 Covariance between two variables (m_{11}) : Definition, computation, effect of change of origin and scale.

8.5 Karl Pearson's coefficient of correlation (r) : Definition, computation for grouped and ungrouped data and interpretation.
Properties : (i) $-1 \leq r \leq 1$ (with proof), (ii) Effect of change of origin and scale (with proof).

8.6 Spearman's rank correlation coefficient : Definition, computation and interpretation (without ties), Spearman's rank correlation coefficient (derivation of formula in case of without ties). In case of ties, compute Karl Pearson's correlation coefficient between ranks. (Spearman's rank correlation coefficient formula with correction for ties not expected.)

8.7 Examples and Problems.

9. Regression (12)

9.1 Concept of regression, lines of regression, fitting of lines of regression by the least squares method, interpretation of slope and intercept.

9.2 Regression coefficient (b_{yx}, b_{xy}) : Definition, computation, properties (with proof).

(1) $b_{yx}b_{xy} = r^2$, (2) $b_{yx}b_{xy} \leq 1$, (3) $b_{yx} = r\frac{\sigma_y}{\sigma_x}, b_{xy} = r\frac{\sigma_x}{\sigma_y}$,, (4) Effect of change of origin and scale, (5) Angle between the two lines of regression.

9.3 Mean residual sum of squares (s.s.) $\frac{1}{n-2}\Sigma(y_i - \hat{y}_i)^2$, Residual plot and its interpretation.

9.4 Explained and unexplained variation, coefficient of determination.

9.5 Non-linear regression : (1) Second degree curve, (2) Exponential curve of type $y = ab^x$, fitting of such curves by the least square method after logarithmic transformation, (3) Logistic curve $y = \frac{k}{1+e^{a+bx}}$, Interpretation of $b < 0, b > 0$. Illustrations of logistic curve. (Fitting of logistic curve is not expected).

Mean residual s.s. as a criteria to decide the best fit of the curve.

9.6 Examples and Problems.

10. Theory of Attributes (10)

10.1 Attributes : classification, notion of manifold classification, dichotomy, class-frequency, order of class, positive class-frequency, negative class frequency, quanta class frequencies, ultimate class frequency, relationship among different class frequencies (up to three attributes), dot operator to find the relation between frequencies, fundamental set of class frequencies.

10.2 Consistency of data upto 3 attributes.

10.3 Concepts of independence and association of two attributes.

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

10.5 Examples and Problems.

11. Applications of Statistics in Economics (13)

11. Index Numbers

11.1 Introduction

11.2 Definition and Meaning

11.3 Points to be considered in construction of Index numbers.

11.4 Simple and weighted price index numbers.

11.5 Laspeyre's, Passche's and Fisher's Index numbers.

11.6 Description of following index numbers; CPI, BSE, SENSEX.

11.7 Examples and Problems.

Reference Books

1. Goon Gupta and Das Gupta : Fundamentals of Statistics, Vol. 1, The World Press Pvt. Ltd., Kolkata.
2. Miller and Fruend : Modern Elementary Statistics.
3. Snedecor and Cochran : Statistical Methods, Oxford and IBH Publishers.
4. Mukhopadhyay, P. : Mathematical Statistics (1996), New Central Book Agency, Calcutta, Introduction to Mathematical Statistics, Ed. 4 (1989), MacMillan Publishing Co. New York.
5. Gupta and Kapoor : Fundamentals of Mathematical Statistics, Sultan Chand and Sons, New Delhi.
6. Neil Weiss : Introductory Statistics : Pearson Publishers
7. Gupta and Kapoor : Fundamentals of Applied Statistics, Sultan Chand and Sons, New Delhi.
8. Amir D. Aczel and Jayael Soundarpandiyan, Complete Business Statistic : McGraw Hill Education (6th Edition).
9. B. L. Agarwal : Programmed Statistics, New Age International Publishers, New Delhi.
10. D. C. Montgomory : Introduction to Statistical Quality Control, Wiley Eastern Publishers.
11. K. V. S. Sarma : Statistics Made Simple : Do it yourself on PC. Prentice Hall of India Pvt. Ltd., New Delhi.

F.Y.B.Sc. Statistics / Statistical Techniques

Paper - II : Discrete Probability and Probability Distributions

Objectives :

The main objective of this course is to introduce to the students the basic concepts of probability, axiomatic theory of probability, concept of random variable, probability distribution (univariate and bivariate) discrete random variables, expectation and moments of probability distribution.

By the end of the course students are expected to be able :

1. to distinguish between random and non-random experiments.
2. to find the probabilities of events.
3. to obtain a probability distribution of random variable (one or two dimensional) in the given situation, and
4. to apply standard discrete probability distribution to different situations.

F.Y.B.Sc. Statistics/Statistical Techniques

Paper - II : Discrete Probability and Probability Distributions

Pre requisite : Permutation and Combination theory, Binomial theorem, Algebra of sets.

1. Sample Space and Events (5)

1.1 Experiments/Models, Ideas of deterministic and non-deterministic models.

1.2 Definitions of - (i) Sample space, (ii) Discrete sample space : finite and countably infinite, (iii) Event, (iv) Elementary event, (v) Complement of an event.

1.3 Concepts of occurrence of an event.

1.4 Algebra of events and its representation in set theory notations.

Occurrence of :

(i) at least one of the given events,

(ii) none of the given events,

(iii) all of the given events,

(iv) mutually exclusive events,

(v) mutually exhaustive events,

(vi) exactly one event out of the given events.

1.5 Examples and Problems.

2. Probability (for finite sample space only) (9)

2.1 Equiprobable sample space, probability of an event, certain event, impossible event, classical definition of probability and its limitations, relative frequency approach.

2.2 Non-equiprobable sample space, probability with reference to a finite sample space : probability assignment approach, probability of an event.

2.3 Axioms of probability.

2.4 Probability of union of two events. Theorem of total probability

$P(A \cup B) = P(A) + P(B) - P(A \cap B)$ and its generalization to three events (with proof).

2.5 To prove

(i) $P(A') = 1 - P(A)$, (ii) If $A \subset B, P(A) \leq P(B)$, (iii) $P(\bigcup_{i=1}^{k} A_i) \leq \sum_{i=1}^{k} P(A_i)$. (Boole's inequality).

2.6 Examples and Problems.

3. Conditional Probability and Independence (9)

3.1 Definition of independence of two events

$$P(A \cap B) = P(A) \cdot P(B)$$

3.2 Pairwise independence and mutual independence for three events.

3.3 Definition of conditional probability of an event.

3.4 Multiplication theorem $P(A \cap B) = P(A) \cdot P(B/A)$.

Generalisation to $P(A \cap B \cap C)$.

3.5 Bayes' Theorem (with proof).

3.6 Examples and Problems.

4. Univariate Probability Distributions (defined on Discrete Sample Space) (10)

4.1 Concept and definition of a discrete random variable.

4.2 Probability mass function (p.m.f.) and cumulative distribution function (c.d.f.), $F(\cdot)$ of discrete random variable, properties of (c.d.f.).

4.3 Mode and median of a univariate discrete probability distribution.

4.4 Examples and Problems.

5. Mathematical Expectation (Univariate Random Variable) (9)

5.1 Definition of expectation of a random variable, expectation of a function of a random variable.

5.2 Definitions of mean, variance of univariate probability distribution, effect of change of origin and scale on mean and variance.

5.3 Probability generating function (PGF), Simple properties, mean and variance using PGF.

5.4 Definition of raw, central and factorial moments of univariate probability distributions and their interrelations.

5.5 Examples and Problems.

End of the First Term

6. Bivariate Probability Distribution (defined on Finite Sample Space) (12)

6.1 Definition of two-dimensional discrete random variable, its joint p.m.f. and its distribution function and their properties.

6.2 Computation of probabilities of events in bivariate probability distribution.

6.3 Concepts of marginal and conditional probability distributions.

6.4 Independence of two discrete random variables.

6.5 Examples and Problems.

7. Mathematical Expectation (Bivariate Random Variable) (6)

7.1 Definition.

7.2 Theorems on expectations of sum and product of two jointly distributed random variables.

7.3 Conditional expectation.

7.4 Definitions of conditional mean and conditional variance.

7.5 Definition of raw and central moments.

7.6 Definition of covariance, correlation of coefficient (ρ), independence and uncorrelatedness of two variables.

7.7 Variance of linear combination of variables.

7.8 Examples and Problems.

8. Some Standard Discrete Probability Distribution (24)

8.1 Uniform discrete distribution on integers 1 to n: - p.m.f., c.d.f., mean, variance, real life situations, comment of mode and median.

8.2 Bernoulli Distribution : p.m.f., mean variance, moments distribution of sum of independent identically distributed Bernoulli variables.

8.3 Binomial Distribution : p.m.f.

$$P(x) = \begin{cases} \binom{n}{x} p^x q^{n-x} & x = 0, 1, \cdots, n, \quad 0 < p < 1, \quad q = 1 - p \\ 0 & \text{otherwise} \end{cases}$$

Notation : $X \sim B(n, p)$.

Recurrence relation for successive probabilities, computation of probabilities of different events, computation of median for given parameters, mode of the distribution,

mean, variance, moments, skewness (comments when $p = 0.5, p > 0.5, p < 0.5$), P.G.F. additive property of binomial variables, conditional distribution of X given $X + Y$, where X and Y are independent, $B(n_1, p)$ and $B(n_2, p)$ variables.

8.4 Hypergeometric Distribution : p.m.f.,

$$p(x) = \begin{cases} \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}}, & x = a, a+1`, \cdots, b \\ 0, & \text{otherwise} \end{cases}$$

$$\text{where } a = \max(0, n - N + N), b = \min(n, M)$$

Notation : $X \sim H(N, M, n)$.

Computation of probability, situations where this distribution is applicable, binomial approximation to hypergeometric probabilities, mean and variance of the distribution.

8.5 Poisson Distribution : p.m.f. $p(x) = e^{-m}\frac{m^x}{x!}, x = 0, 1, 2, \cdots;\ \ m > 0$. State the mean, variance, additive property (no derivation). Derivation of Poisson distribution as a limiting case of binomial distribution.

8.6 Example and Problems.

Reference Books

1. Hogg, R. V. and Craig R. G. : Introduction to Mathematical Statistics, Ed. 4. (1989), MacMillan Publishing Co., New York.
2. Hoel, P. G. : Introduction to Mathematical Statistics (1962), John Wiley and Sons, New York.
3. Feller, W. : Introduction to Probability Theory and Its Applications, Vol. I (1963), Asian Publishing House, Bombay.
4. Mood, A. M. and Graybill, F. A. and Boes D.C. E. : Introduction to Theory of Statistics, Ed. 3 (1974), McGraw Hill and Kagakusha Ltd. London.

5. Mayer, P. N. : Introduction to Probability and Statistical Applications, Addison Wesley Publishing Co., Massachusetts).

6. Gupta and Kapoor : Fundamentals of Mathematical Statistics, Sultan Chand and Sons, New Delhi.

7. Ross : Probability theory, Pearson Publishers.

8. M. B. Kulkarni and S. B. Ghatpande : Discrete Probability and Probability Distributions, SIPF Academy, Nashik.

9. B. L. Agarwal : Programmed Statistics, New Age International Publishers, New Delhi.

10. K. V. S. Serma : Statistics Made Simple : Do it yourself on PC.

Reference Websites :

1. www.stats.unipune.ernet.in (100 Data sets for Statistics Educagtion by Dr Anil P. Gore, Dr. Mrs. S. A. Paranjpe and Madhav B. Kulkarni available in ISPS folder).
2. www.freestatistics.tk
3. www.psychstat.smsu.edu/sbk00.htm
4. www.bmj.bmjournals.com/collections/statsbk/index.shtml
5. www.statweb.calpoly.edu/bchance/stat-stuff.html
6. www.amstat.org/publications/jse/jse-data-archive.html
7. www.statpages.org (Webpages that perform statistical calculations)
8. www.amstat.org/publications/chance (Chance magazine)
9. www.statsci.org/datasets.html (Data sets)
10. www.math.uah.edu/stat (Virtual laboratories in Statistics)
11. www.amstat.org/publications/stats (STATS : the magazine for students of Statistics)
12. www.stat.ucla.edu/cases (Case studies in Statistics)

Paper III - Practical

Pre-requisites. Knowledge of the topics in theory and papers.

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

1. to compute various measures of central tendency, dispersion, skewness and kurtosis,
2. to compute correlation coefficient, regression coefficients,
3. to fit binomial distribution,
4. to analyse data pertaining to discrete and continuous variables and to interpret the results,
5. to compute probabilities of bivariate distributions,
6. to interpret summary statistics of computer output.

Sr. No.	Title of the Experiment	No. of Experiments
1.	Use of random number tables to draw SRSWOR, SRSWR stratified, systematic sampling. (Also using MSEXCEL/Spread sheet)	1
2.	Graphical and diagrametic representation of statistical data (also using MS-EXCEL/spread sheet) problems based on simple and subdivided bar diagrams, pie diagram	2
3.	Tabulation, Parato Diagram	1
4.	Computation of measures of central tendency (ungrouped and grouped data). Use of an appropriate measure and interpretation of results and computation of partition values. (Also using MS-EXCEL/spread sheet).	2
5.	Computation measures of dispersion (ungrouped and grouped data) (Also using MS-EXCEL/spread sheet). Box plot	2
6.	Raw and central moments (with Sheppard's connection)	1
7.	Computation of measuring skewness and Kurtosis, Box plot	1
8.	Scatter diagram, correlation coefficient (ungrouped data) Fitting of lines of regression. (Results to be verified by using computer), Rank correlation	2

Sr. No.	Title of the Experiment	No. of Experiments
9.	Fitting of lines of regression and computation of correlation coefficient (grouped data), Mean residual s.s., residual plot. Also using MSEXCEL/Spread sheet	1
10.	Fitting of second degree curve, exponential curve of type $y = ab^X$. Comparison finding the best fit using mean residual s.s., also using MSEXCEL/Spread sheet	2
11.	Fitting of binomial distribution and computation of probabilities.	1
12.	Applications of (a) binomial distribution, (b) hypergeometric distribution, (c) Poisson distribution Also using MSEXCEL/Spread sheet	1
13.	Index numbers	1
14.	Practical based on analysis of data collected by students in a batch of size not exceeding 15 students	2