

University of Pune
Three Year B. A. Degree Program
Syllabus for F.Y.B.A. Statistics,
(With effect from Academic Year 2013-2014)

Submitted by

Prof. S. R. Deshmukh
Chairperson,
Board of Studies in Statistics

1) Title of the program: Three Year B. A. Degree

2) Preamble to the syllabus: The word **Statistics** is used in different ways in different contexts. To a cricket fan, Statistics is the information about runs scored or wickets taken by a player. To the manager of a manufacturing unit, Statistics may be the information about the process control. To a medical researcher investigating the effects of a new drug, Statistics are evidence of research efforts. To a college student, Statistics are the grades or marks scored in a course. Thus, in all these illustrations Statistics word 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.

Statistics provides tools for making decisions when conditions of uncertainty prevail. Hence these tools and techniques are used in almost all fields. Statistics is indispensable for people working in fields like agriculture, business, management, economics, finance, insurance, education, biotechnology and medical science etc. Since last two decade, with the help of computers large amount of data can be handled and more sophisticated statistical techniques can be used in an effective manner. Knowledge of different aspects of Statistics has become crucial. There is a continuous demand for statisticians in every field – education, industry, software and research. The syllabus of the three Year B. Sc. degree course in Statistics is framed in such a way that the students at the end of the course can apply judiciously the statistical tools to a variety of data sets to arrive at some conclusions.

Statistics can be divided into two broad categories, (1) exploratory statistics or descriptive statistics, which is concerned with summarizing data and describing these data, and (2) confirmatory statistics or inferential statistics, which is concerned with making decisions about the population based on the sample.

Up to higher secondary school, students are mostly exposed to descriptive statistics. At the first year a student can take any one of the four subjects related statistics, such as Statistics, Applied Statistics, Mathematical Statistics and Statistical Prerequisites. If the student continues with these subjects at the second year and third year, it is expected that at the end of the degree course a student is able to apply the statistical tools to real life data.

3) Introduction: Three Year B. A. degree program is of three years duration, with semester pattern for the second and third year and annual examination pattern for the first year.

The structure of **Bachelor of Arts (B.A.) is as follows.** The student joining the First Year B.A. Course has to take six subjects from 13 groups. The student cannot take more than one subject from one group. There are four subjects related to statistics. These are Statistics (Group L), Applied Statistics (Group L), Mathematical Statistics (Group J) and Statistical Prerequisites (Group K).

4) Eligibility

First Year B.A.

Higher secondary school certificate examination of the Maharashtra State Board of Higher Secondary Education or an equivalent examination of any other statutory Board or University with English as a passing subject.

Detailed Syllabus of F. Y. B. A. Statistics

Note: Mathematical derivations and proofs are not expected.

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 compute various measures of central tendency, dispersion, skewness and kurtosis.
- (ii) to analyze data pertaining to attributes and to interpret the results.
- (iii) to compute the correlation coefficient for bivariate data and interpret it.
- (iv) to fit linear, quadratic and exponential curves to the bivariate data to investigate relation between two variables.
- (v) to fit linear regression model to the bivariate data
- (vi) to compute and interpret various index numbers.

1. Introduction to Statistics: (4L)

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, 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 (6L)

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, directional data.

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

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

3. Summary Statistics. (16L)

Review / Revision of Presentation of Data

- 3.1 Classification: Raw data and its classification, ungrouped frequency distribution, Sturges' rule, grouped frequency distribution, cumulative frequency distribution, inclusive and exclusive methods of classification, Open end classes, and relative frequency distribution.
- 3.2 Measures of Central Tendency
 Review / Revision of following topics: Concept of central tendency of statistical data, Statistical averages, characteristics of a good statistical average.
 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.
 Mode and Median: Definition, formulae (for ungrouped and grouped data), merits and demerits. Empirical relation between mean, median and mode (without proof).
 Topics to be taught in detail
 Partition Values: Quartiles, Deciles and Percentiles (for ungrouped and grouped data), Box Plot.
 Geometric Mean (G.M.): Definition, formula, merits and demerits.
 Harmonic Mean (H.M.): Definition, formula, merits and demerits.
 Order relation between arithmetic mean, geometric mean, harmonic mean
 Weighted Mean: weighted A.M., G.M. and H.M.
 Situations where one kind of average is preferable to others.
- 3.3 Measures of Dispersion
 Review / Revision of following topics
 Concept of dispersion, characteristics of good measure of dispersion.
 Range, Semi-interquartile range (Quartile deviation): Definition, merits and demerits, Mean deviation: Definition, merits and demerits, minimality property (without proof), Variance and standard deviation: Definition, merits and demerits, effect of change of origin and scale, combined variance for n groups (derivation for two groups).
 Topics to be taught in detail
 Mean squared deviation: Definition, minimality property of mean squared deviation (with proof),
 Measures of dispersion for comparison: coefficient of range, coefficient of quartile deviation and coefficient of mean deviation, coefficient of variation (C.V.)

4. Moments, Skewness and Kurtosis (12L)

- 4.1 Raw moments (m'_r) for ungrouped and grouped data.
- 4.2 Central moments (m_r) for ungrouped and grouped data, Effect of change of origin and scale.
- 4.3 Relations between central moments and raw moments, upto 4-th order (without proof).
- 4.4 Concept of skewness of frequency distribution, positive skewness, negative skewness, symmetric frequency distribution.
- 4.5 Bowley's coefficient of skewness : Bowley's coefficient of skewness lies between -1 to 1 (with proof), interpretation using Box plot.
- 4.6 Karl Pearson's coefficient of skewness.
- 4.7 Measures of skewness based on moments (β_1, γ_1).

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

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

5 Theory of Attributes (10L)

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

5.2 Consistency of data upto 2 attributes.

5.3 Concepts of independence and association of two attributes.

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

End of First Term

6. Correlation (12L)

6.1 Bivariate data, Scatter diagram and interpretation.

6.2 Concept of correlation between two variables, positive correlation, negative correlation, no correlation.

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

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

6.5 Spearman's rank correlation coefficient: Definition, derivation of formula, computation and interpretation (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.)

7. Fitting of curves to the bivariate data (12L)

7.1 Fitting of line ($Y = a + b X$),

7.2 Fitting of second degree curve ($Y = a + b X + c X^2$),

7.3 Fitting of exponential curves of the type $Y = a b^X$ and $Y = aX^b$.

In all these curves parameters are estimated by the method of least squares.

8. Linear Regression Model (14L)

8.1 Meaning of regression, difference between correlation and regression,

8.2 Concept of error in regression, error modeled as a continuous random variable.

Simple linear regression model: $Y = a + b X + \epsilon$, where ϵ is a continuous random variable with $E(\epsilon) = 0$, $V(\epsilon) = \sigma^2$. Estimation of a, b by the method of least squares. Interpretation of parameters. Statement of the estimator of σ^2 .

8.3 Concept of residual, plot of residual against X , concept of coefficient of determination.

9 Index Numbers (10L)

- 9.1 Introduction.
- 9.2 Definition and Meaning.
- 9.3 Problems/considerations in the construction of index numbers.
- 9.4 Simple and weighted price index numbers based on price relatives.
- 9.5 Simple and weighted price index numbers based on aggregates.
- 9.6 Laspeyre's, Paasche's and Fisher's Index numbers.
- 9.7 Consumer price index number: Considerations in its construction. Methods of construction of consumer price index number-(i) family budget method (ii) aggregate expenditure method.
- 9.8 Shifting of base, splicing, deflating, purchasing power.
- 9.9 Description of the BSE sensitivity and similar index numbers.

Recommended Books

1. Agarwal, B. L. (2003). Programmed Statistics, Second Edition, New Age International Publishers, New Delhi.
2. Draper, N. R. and Smith, H. (1998). Applied Regression Analysis Third Edition, John Wiley and Sons
3. Goon, A.M., Gupta, M.K. and Dasgupta, B. (1983). Fundamentals of Statistics, Vol. 1, Sixth Revised Edition, The World Press Pvt. Ltd., Calcutta.
4. Gupta, S.C. and Kapoor, V. K. (1983). Fundamentals of Mathematical Statistics, Eighth Edition, Sultan Chand and Sons Publishers, New Delhi.
5. Gupta, S.C. and Kapoor, V. K. (1997). Fundamentals of Applied Statistics, Third Edition, Sultan Chand and Sons Publishers, New Delhi.
6. Freund, J.E. (1977). Modern Elementary Statistics. Fourth Edition, Prentice Hall of India Private Limited, New Delhi.
7. Montgomery, D. C; Peck, E. A.; Vining, G. G. (2006). Introduction to Linear Regression Analysis, John Wiley and Sons
8. Purohit, S. G., Gore S. D., Deshmukh S. R. (2008). Statistics Using R, Narosa Publishing House, New Delhi.
9. Sarma, K. V. S. (2001). Statistics Made it Simple: Do it yourself on PC. Prentce Hall of India, New Delhi.
10. Snedecor G.W. and Cochran W.G.(1989). Statistical Methods, Eighth Ed. East-West Press.

Reference Websites

1. www.stats.unipune.ac.in (100 Data sets for Statistics Education by Dr. Anil P. Gore, Dr. Mrs. S. A. Paranjpe and Madhav B. Kulkarni available in ISPS folder).
2. www.freestatistics.tk (National Statistical Agencies)
3. www.psychstat.smsu.edu/sbk00.htm (Online book)
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 (International journal on teaching and learning of statistics)
7. www.amstat.org/publications/chance (Chance magazine)
8. www.statsci.org/datasets.html (Data sets)
9. www.math.uah.edu/stat (Virtual laboratories in Statistics)
10. www.amstat.org/publications/stats (STATS : the magazine for students of Statistics)
11. www.stat.ucla.edu/cases (Case studies in Statistics).
12. www.statsoft.com
13. www.statistics.com
14. www.indiastat.com
15. www.unstat.un.org
16. www.stat.stanford.edu
17. www.statpages.net
18. www.wto.org
19. www.censusindia.gov.in
20. www.mospi.nic.in
21. www.statisticsofindia.in