UNIVERSITY OF PUNE

Syllabi for First Year Bachelor of Science (Computer Science)

With Effect From Academic Year 2008-2009

Subject : Statistics

(1) Statistics Paper I - Statistical Methods - I (Total Marks : 100)
(2) Statistics Paper II - Statistical Methods - II (Total Marks : 100)
(3) Statistics Paper III - Practical Course in Statistics (Total Marks : 100)

Notes :

1. A student of First Year B.Sc. (Computer Science) course must complete all practicals in Statistics to the satisfaction of teacher concerned.

2. A student must produce at the time of practical examination the laboratory journal along with completion certificate signed by head of the department.

3. Duration of practical examination will be extended by 10 minutes to compensate for the loss of time for viva-voce of the candidate.

4. In all papers, emphasis shall be on studying statistical concepts and techniques in the field of Computer Science. Mathematical derivations and proofs are not expected. It is expected to use MS-EXCEL/spreadsheet commands (s) wherever possible.

5. Students are allowed to use non-programmable scientific calculators at the time of theory and practical examination and also for regular practicals.

6. Theory question paper should not contain questions on software packages, e.g., MS-EXCEL/ Spreadsheet.

Revised Syllabus for F. Y. B.Sc. (Computer Science)

PAPER - I : Statistical Methods - I

1. Data condensation and Graphical methods **(6)**

1.1 Raw data, attributes and variables, discrete and continuous variables.

1.2 Presentation of data using exclusive frequency distribution and cumulative frequency distribution.

(Construction of frequency distribution is not expected).

1.3 Graphical Presentation of frequency distribution - histogram, stem and leaf chart, less than and more than type ogive curves.

1.4 Numerical Problems.

2. Measures of Central Tendency **(10)**

2.1 Central Tendency : concept, illustrations, scope and limitations.

2.2 Measures of central tendency -

(a) Arithmetic Mean (A. M.) : definition, formula for computations of A. M. for ungrouped and grouped, data combined mean, weighted mean, merits and demerits of A. M., Trimmed mean.

(b) Median ; definition, computation of median for ungrouped and grouped data, graphical methods, merits and demerits.

(c) Mode : definition, computation of mode for ungrouped and grouped data, graphical representation, merits and demerits.

2.3 Partition Values : Quartiles, Deciles and Percentiles by formula and by graph, percentile rank, Box Plot.

2.4 Numerical Problems.

3. Measures of Dispersion **(10)**

3.1 Dispersion : concept and utility

3.2 Measures of dispersion

(a) Range : definition, computations for ungrouped and grouped data, merits and demerits

(b) Quartile deviation

(c) Variance and Standard Deviation : definition, computations for ungrouped and grouped data, combined variance for two groups, merits and demerits.

3.3 Measures of dispersion for comparison - Coefficient of range, coefficient of quartile deviation, coefficient of variation (C.V.)

3.4 Numerical Problems.

4. Moments **(4)**

4.1 Raw and Central Moments : definition, computations for ungrouped and grouped data (only up to first four moments).

4.2 Relation between raw and central moments up to fourth order.

4.3 Numerical Problems.

5. Measures of Skeweness and Kurtosis **(4)**

5.1 Concept of symmetric frequency distribution, skewness, positive and negative skewness

5.2 Measures of skewness - Pearson's measure, Bowley's measure, β_1, γ_1

5.3 Kurtosis of a frequency distribution, measure of kurtosis (β_2, γ_2) based upon moments, type of kurtosis : leptourtic, platykurtic and mesokurtic

5.4 Numerical Problems

6. Correlation (for ungrouped data) (8)

6.1 Bivariate data ; scatter diagram

6.2 Correlation, positive correlation, negative correlation, zero correlation

6.3 Karl Pearson's coefficient of correlation (r), limits of $r(-1 \leq r \leq 1)$, interpretation of r, coefficient of determination (r^2) and interpretation as strength of relation

6.4 Karl Pearson's coefficient of correlation between ranks

6.5 Numerical problems.

END OF THE FIRST TERM

7. Regression (for ungrouped data) (12)

7.1 Regression, illustrations, appropriate situations for regression and correlation

7.2 Linear regression

7.3 Fitting of straight line using least squares method

7.4 Properties of regression coefficients : $b_{xy} \cdot b_{yx} = r^2, b_{xy} * b_{yx} \leq 1, b_{yx} = r\frac{\sigma_x}{\sigma_y}$ and $b_{xy} = r\frac{\sigma_y}{\sigma_x}$

7.5 Non-linear regressions models : second degree curve, growth curve models.

(i) $Y = ae^{bX}$, (ii) $Y = ab^X$, (iii) $Y = aX^b$ and

(iv) logistic model $Y = \frac{K}{1+e^{a+b}}$.

7.6 Residual plot, mean residual sum of squares (s.s.)

7.7 Numerical Problems

8. Multiple and Partial Regression and Correlation (For trivariate data) (10)

8.1 Yule's notation and concept of multiple regression

8.2 Fitting of multiple regression plane

8.3 Partial regression coefficient, interpretation

8.4 Multiple correlation coefficient, concept, definition, computation and interpretation

8.5 Partial correlation coefficient, concept, definition, computation and interpretation

8.6 Numerical Problems

9. Time Series (12)

9.1 Meaning and utility

9.2 Components of time series

9.3 Additive and multiplicative models

9.4 Methods of estimating trend : moving average method, least squares method and exponential smoothing method

9.5 Elimination of trend using additive and multiplicative models

9.6 Measurement and estimation of seasonal variations using Link Relative method and Ratio to trend method

9.7 Simple time series models : AR(1), AR(2)

9.8 Numerical problems

10. Statistical Quality Control (SQC) (8)

10.1 Introduction to seven process control (PC) tools

10.2 Quality, causes of variation, lot and process control, control charts $(\overline{X}, R)$, control limits, specification limits, capability index (C_p), capability performance index (C_{pk}), interpretation.

PAPER - II : - Statistical Methods - II

Note : Pre-requisites of permutations, combinations, binomial theorem and algebra of sets.

1. Theory of Probability **(14)**

1.1 Deterministic and non-deterministic models

1.2 Random Experiment, Sample Spaces (finite and countably infinite)

1.3 Events : types of events, operations on events

1.4 Probability - classical definition, relative frequency approach, probability models, axioms of probability, probability of an event

1.5 Theorems of probability (with proof)

(i) $0 \leq P(A) \leq 1$ (ii) $P(A) + P(A') = 1$ (iii) $P(A) \leq P(B)$ when $A \subset B$

(iv) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

1.6 Concept and definitions of conditional probability, multiplication theorem $P(A \cap B) = P(A)P(B/A)$.

1.7 Bayes' theorem (without proof)

1.8 Concept and definition of independence of two events

1.9 Numerical Problems

2. Discrete Random variables **(6)**

2.1 Definition of random variable and discrete random variable

2.2 Definition of probability mass function, distribution function and its properties

2.3 Definition of expectation and variance, theorems on expectation

2.4 Determination of median and mode using p.m.f.

2.5 Numerical Problems

3. Standard Discrete Distributions **(16)**

3.1 Uniform Distribution : definition, mean, variance

3.2 Bernoulli Distribution : definition, mean, variance, additive property

3.3 Binomial Distribution : definition, mean, variance, additive property

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

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

3.6 Illustrations of real life situations

3.7 Numerical Problems

4. Continuous Random Variables **(6)**

4.1 Definition of continuous random variable (r.v.), probability density function (p.d.f.)

4.2 Distribution function and its properties (statements only)

4.3 Numerical Problems

END OF THE FIRST TERM

5. Standard Continuous Probability Distributions **(14)**

5.1 Uniform Distribution : statement of p.d.f., mean, variance, nature of probability curve.

5.2 Exponential Distribution : Statement of p.d.f. of the form, $f(x) = (1/\theta) \cdot e^{-x/\theta}, x \geq 0, \theta > 0$, nature of probability curve, mean, variance, lack of memory property.

5.3 Normal Distribution : Statement of p.d.f., identification of parameters, nature of probability density curve, standard normal distribution, symmetry, distribution of $aX+b, aX+bY+c$ where X and Y are independent random variables, computations of probabilities using normal probability table, normal approximation to binomial and Poisson distribution, central limit theorem (statement only), normal probability plot.

5.4 Pareto Distribution : p.d.f. of the form $f(x) = \frac{\alpha}{x^{\alpha+1}}$ for $x \geq 1$ and $\alpha > 0$, mean, variance, applications.

5.5 Numerical Problems

6. Statistical Inference : The idea of estimation and testing of hypothesis (6) Tests of Hypothesis

6.1 Definitions : population, sample, SRSWR, SRSWOR, random sample from a probability distribution, parameter, statistic, standard error of estimator

6.2 Concept of null and alternative hypothesis, critical region, level of significance, type I and type II error, one sided and two sided tests, p-value.

7. Large Sample Tests (8)

7.1 $H_0 : \mu = \mu_0$ Vs $H_1 : \mu \neq \mu_0, \mu, \mu < \mu_0, \mu > \mu_0$;
$H_0 : \mu_1 = \mu_2$ Vs $H_1 : \mu_1 \neq \mu_2, \mu_1 < \mu_2, \mu_1 > \mu_2$

7.2 $H_0 : P = P_0$ Vs $H_1 : P \neq P_0, P < P_0, P > P_0$;
$H_0 : P_1 = P_2$ Vs $H_1 : P_1 \neq P_2, P_1 < P_2, P_1 > P_2$

7.3 Numerical Problems

8. Tests based on t, χ^2 **(6)**

8.1 Chi square test for goodness of fit, test for independence of attributes ($m \times n$ contingency table)

8.2 t-test for testing $H_0 : \mu = \mu_0$ Vs $H_1 : \mu \neq \mu_0, \mu < \mu_0, \mu > \mu_0$

$H_0 : \mu_1 = \mu_2$ Vs $H_1 : \mu_1 \neq \mu_2, \mu_1 < \mu_2, \mu_1 > \mu_2$, paired test.

8.3 Numerical Problems

9. Simulation **(8)**

9.1 Introduction to simulation, merits and demerits

9.2 Pseudo-random number generator (Linear congruential generator), model sampling from uniform and exponential distributions as simulation technique

9.3 Model sampling from Normal distribution using Box-Muller transformation

9.4 Run test for testing randomness of the sample and sign test for testing symmetry of the sample

9.5 Numerical problems.

STATISTICS PAPER III - Practical Course in Statistics

Note : The practicals on all topics should be done manually even if they are performed using MS-Excel.

(A) List of Practicals to be performed (Using both calculators and computer software).

1. Measure of central tendency
2. Measures of dispersion
3. Computation of raw and central moments. Measures of skewness and kurtosis, box plot.
4. Fitting of binomial and Poisson distribution
5. Fitting of normal distribution
6. Computation of correlation coefficient, coefficient of determination and fitting of lines of regression, residual plot
7. Fitting of second degree and exponential curve and determination of mean residual s.s.
8. Fitting of multiple regression plane and computation of multiple and partial correlation coefficients.
9. Estimation of trend in a time series data using moving average and straight line fitting.
10. Large sample tests
11. Small sample tests
12. Model sampling from uniform and exponential distributions

(B) List of Practicals to be performed by using exclusively MS-EXCEL/Spreadsheet or any other Statistical software package as SYSTAT, R, SPSS, MINITAB

13. Pie chart, Histogram, bar diagram, ogives, box-plot, stem and leaf plot and descriptive statistics

14. Computation of summary statistics and further analysis of data using statistical tools. Practical based on analysis of data collected by students in a batch of size not exceeding 15 students using statistical tools such as (i) diagrams and graphs, (ii) descriptive statistics, correlation, regression (iii) multiple regression.

15. Fitting of regression lines, multiple regression planes, scatter diagram and residual plots for bivariate data.

16. Fitting of curves (i polynomial, (ii) logarithmic (iii) exponential.

17. Computations of binomial Poisson, normal etc. probailities. Simulation of M/M/1 and M/G/1 queues where G : as uniform over $[a, b]$.

(C) Developing C Programs and getting successful out of the program. (Program Code and successful outputs should be included in the journal.)

18. Computation of mean and standard deviations for ungrouped data using programs in C.

19. Computation of correlation coefficient and fitting of regression lines for ungrouped data using program in C.

20. Computation of moving average using program C.

Books Recommended

1. **Fundamentals of Statistics**, (5-th Edition, 1986), Vol. 1 and 2 : Goon Gupta Dasgupta, World Press, Calcutta.

2. **Statistical Methods (An Introductory Text)** : J. Medhi, New Age International

3. **Modern Elementary Statistics** : J. E. Freund

4. **Probability, Statistics, Design of Experiments and Queuing Theory with Applications of Computer Science**, (2001) : K. S. Trivedi, Prentice Hall of India, New Delhi.

5. **Programmed Statistics** ; B. L. Agarwal, New Age International Publishers, New Delhi

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

7. **A First Course In Probability** 6th Edition : Ross, Pearson publication.

8. **An Introductory Statistics** : Kennedy and Gentle

10. **Simulation and Modelling and its Applications** : Keltan and Law, Tata McGraw Hill.

11. **Time Series Analysis** : Box and Jenkin

12. **Time Series Methods** : Brockwell and Devis

13. **System Simulation with Digital Computer** : Narsingh Dev, Prentice Hall

14. **Statistical Methods** : G. W. Snedecor, W. G. Cochran

15. **Common Statistical Tests** : M. B. Kulkarni, S. B. Ghatpande, S. D. Gore, Satyajeet Prakashan, Pune (1999).

16. **Introduction to Discrete Probability and Probability Distributions** : M. B. Kulkarni, S. B. Ghatpande, SIPF Academy (2007).

17. **Introduction to Statistical Quality Control** : D. C. Montgomery, Wiley Eastern.

18. **Statistics Made Simple. Do it Yourself on P.C.** : K. V. S. Sarma; Prentice Hall.

Websites Recommended

1. **www.stats.unipune.ernet.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**
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** (Web pages that perform statistical calculations)
8. **www.amstat.org/publications/chance** (Chance magazine)
9. **www.statsci.org/datasets.html** (Data sets)
11. **www.amstat.org/publications/stats** (STATS : the magazine for students of statistics)
12. **www.stat.ucla.edu/cases** (Case studies in statistics)

Structure of Evaluation of Practical Paper

	Marks
(A) Continuous Internal Evaluation	
Viva-Voce	10
Journal	10
(B) Annual Practical Examination (duration 3 hours)	
Questions based on practicals to be solved using MS-EXCEL/ Spreadsheet and Program in C	20
Questions based on other practicals using calculators	50
Viva-Voce	10
Total	100