

Register No:

Date:



ST. JOSEPH'S COLLEGE (AUTONOMOUS), BANGALORE-27
M.SC. BIG DATA ANALYTICS - II SEMESTER
SEMESTER EXAMINATION: APRIL 2018
BDA 2216: ADVANCED STATISTICAL METHODS

TIME: 2 ½ HRS

MAX MARKS 70

This Question Paper Contains Two Printed Pages

Statistical tables will be provided on request

Answer any seven of the following

7 X 10 = 70

1. A) Mention the criteria of good estimator and define estimator. (3)
B) Verify whether sample mean is unbiased estimator of population parameter θ in case of exponential distribution with mean $1/\theta$. (3)
C) Let $X \sim B(1,p)$ distribution, Let X_1, X_2 be a random sample of size 2 from above population. Choose the best estimator by comparing mean square errors of following estimators (Use Graph) (4)
a. $T_1 = X_1$ b. $T_2 = X_1 + X_2$ c. $T_3 = \frac{X_1 + X_2}{2}$ d. $T_4 = \frac{2X_1 - X_2}{3}$
2. A) The monthly family income in rupees for a random sample of 16 is drawn from a normal population.
594 689 585 1034 720 929 737 1026
791 836 952 898 1043 806 577 994
Obtain a MLE of average monthly income per family and variance of income (5)
B) Let $X_1, X_2, \dots, X_n \sim NB(r,p)$, where r is known (negative binomial distribution), then obtain moment estimator of p (5)
3. A) Explain types of errors involved in testing of hypotheses with an example (4)
B) Explain steps involved in testing null hypothesis against alternative hypothesis. (6)
4. A) Explain difference between level of significance and size of the test (4)
B) One important factor in selecting software for word processing and database management systems is the time required to learn how to use a particular system. In order to evaluate three database management systems, a firm devised a test to see how many training hours were needed for five of its word processing operators to become proficient in each of three systems. (6)
System A 16 19 14 13 18 hours
System B 16 17 13 12 17 hours
System C 24 22 19 18 22 hours
Using a 5% significance level, is there any difference between the training time needed for the three systems?
5. A) Define Multiple linear regression and Fit a regression line for it. (1+5)
B) Derive any two properties of Multiple linear regression. (4)
6. State and proof Gauss–Markov Model. (10)

7. A) Derive an expression for mean square error and explain how it is useful? (4)
 B) Explain the steps involved in testing equality of variance. (6)

8. A) A mouth wash distributors states that the average cost of processing the sales order is 15.25 \$ but the cost controller fears that it is more than 15.25 \$ and he would like to take some action. If it is so a random sample of 80 orders had a sample mean of 15.15 \$. Assuming $\sigma=0.05$ \$. Conduct a test at 0.01 level of significance to help the cost controller (5)

B) The nicotine content in mg is found in sample of 5 cigarette is 24, 27, 26, 21, 25 which were manufactured by company of A. In another sample of 6 cigarette of B the nicotine content found to be 27, 30, 28, 31, 22, 36. Can it be said that the variability in nicotine content of cigarette manufactured by A is more than company B (use $\alpha=0.01$) (5)

9. A) Explain backward and forward selection approach using p squared model (5)
 B) When Is Stepwise Regression Appropriate? Explain with an example (5)

10.A) Define Logistic Function. Give any two advantages of it. (4)

B) A group of 20 students spend between 0 and 6 hours studying for an exam.

The table shows the number of hours each student spent studying, and whether they passed (1) or failed (0).

Hours	0.5	0.75	1	1.25	1.5	1.75	1.75	2	2.25	2.5	2.75
Pass	0	0	0	0	0	0	1	0	1	0	1
Hours	3	3.25	3.5	4	4.25	4.5	4.75	5	5.5		
Pass	0	1	0	1	1	1	1	1	1		

- i) Which regression is appropriate to apply? Why?
 ii) Explain how does the number of hours spent studying affect the probability that the student will pass the exam? (6)