



Register Number:

DATE:

ST. JOSEPH'S COLLEGE (AUTONOMOUS), BANGALORE-27

M.Sc. BIG DATA ANALYTICS – II SEMESTER

SEMESTER EXAMINATION – APRIL 2019

BDA2218: ADVANCE STATISTICAL METHODS

Time: 2 ½ Hours

Max: 70 marks

This question paper has **ONE** part and **TWO** printed pages

Note: Statistical table and graph sheet will be provided on request

I Answer any SEVEN of the following:

7x 10 = 70

1. A) Differentiate between estimate and estimator with an example (3)
B) Verify whether sample mean is unbiased estimator of population parameter p in case of Binomial distribution (3)
C) Consider a random sample of size 15 from Poisson (λ). 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) A Suppose the weights of randomly selected American female college students are normally distributed with unknown mean μ and standard deviation σ . A random sample of 10 American female college students yielded the following weights (in pounds): 115 122 130 127 149 160 152 138 149 180 identify the likelihood function and the maximum likelihood estimator of μ , the mean weight of all American female college students. (5)
B) Define moments and estimate the moments for the following example
Suppose we have a random sample X_1, X_2, \dots, X_n where:
 $X_i = 0$ if a randomly selected student does not own a sports car, and
 $X_i = 1$ if a randomly selected student does own a sports car.
Assuming that the X_i are independent Bernoulli random variables with unknown parameter p , find the moment estimator of p , the proportion of students who own a sports car. (5)
3. A) Define size and power of the test with an example. (4)
B) An engineer measured the Brinell hardness of 25 pieces of ductile iron that were subcritically annealed. The resulting data were:

170	167	174	179	179	187	179	183	179
156	163	156	187	156	167	156	174	170
183	179	174	179	170	159	187		

The engineer claims that the mean Brinell hardness of *all* such ductile iron pieces is greater than 170. Test whether the statement is true or not. (6)

4. A) Define: i) Parameter ii) Statistic iii) Sample (3)

B) A study is designed to test whether there is a difference in mean daily calcium intake in adults with normal bone density, adults with osteopenia (a low bone density which may lead to osteoporosis) and adults with osteoporosis. Adults 60 years of age with normal bone density, osteopenia and osteoporosis are selected at random from hospital records and invited to participate in the study. Each participant's daily calcium intake is measured based on reported food intake and supplements. The data are shown below. Is there a statistically significant difference in mean calcium intake in patients with normal bone density as compared to patients with osteopenia and osteoporosis? (7)

Normal Bone Density	Osteopenia	Osteoporosis
1200	1000	890
1000	1100	650
980	700	1100
900	800	900
750	500	400
800	700	350

5. State and proof Gauss–Markov Model. (10)

6. A) Derive the OLS estimator of Regression Coefficient. Show that it is an Unbiased estimator of Regression Coefficient (8)

B) What is the importance of ϵ in regression model (2)

7.A) Derive an expression for mean square error and explain how it is useful? (5)

B) Explain the test procedure for testing equality of mean when variance is unknown (5)

8. A) Mention four assumptions of one-sample t-test. (4)

B) Of all of the individuals who develop a certain rash, suppose the mean recovery time for individuals who do not use any form of treatment is 30 days with standard deviation equal to 8. A pharmaceutical company manufacturing a certain cream wishes to determine whether the cream shortens, extends, or has no effect on the recovery time. The company chooses a random sample of 100 individuals who have used the cream, and determines that the mean recovery time for these individuals was 28.5 days. Does the cream have any effect? (6)

9. A) Write a note on backward and forward selection approach (5)

B) When Is Stepwise Regression Appropriate? Explain with an example (5)

10.A) Write a note on logistic regression.

(4)

B) Suzuki et al. (2006) measured sand grain size on 10 beaches in Japan and observed the presence or absence of the burrowing wolf spider *Lycosa ishikariana* on each beach.

Grain size (mm)	0.245	0.247	0.285	0.299	0.327	0.347	0.356	0.36	0.363	0.364
Spiders	absent	absent	present	present	present	present	absent	present	absent	present

i) Which regression is appropriate to apply? Why?

ii) Determine whether there was a relationship between sand grain size and the presence or absence of the species