****

Register Number:

DATE:

**ST. JOSEPH’S COLLEGE (AUTONOMOUS), BENGALURU-27**

**M.Sc. BIG DATA ANALYTICS – II SEMESTER**

**SEMESTER EXAMINATION: APRIL 2017**

**BDA 2216 Advance Statistics**

**Time 2.5 Hours Maximum Marks 70**

**This Question Paper Contains TWO Printed Paper And ONE Part**

**Answer Any Seven questions 7 x10 = 70**

1. a) What is mean squared error? Show that mean squared error is equal to the sum of variance and square of bias. (2)

b) State the difference between an estimator and an estimate. (2)

c) A population is known to be normally distributed with mean μ and variance σ2. We draw a random sample X1, X2, …, Xn such that these variables are identically and independently distributed. We have three candidate estimators

$T1=\frac{1}{n}\sum\_{i=0}^{n}Xi ,$ T2 = (X1 + Xn)/2 , T3 = $\overbar{X }$+ 3

Which estimator is the best? (6)

1. a) Differentiate between type I error and type II error? (2)
2. Define level of significance and power of a test? (2)
3. A sample of 900 members is found to have a mean of 3.47 cm. Can it be regarded as a simple sample from a large population with mean 3.23 cm and standard deviation2.31 cm. (6)
4. Derive the maximum likelihood estimate of mean and variance of a random variable that follows a normal distribution (10)
5. Derive least square estimators for the simple linear regression model with only one explanatory variable. (10)
6. Using the information given below, fit a linear regression line by considering Y as the dependent variable and X as the independent variable. Compute R square and adjusted R square. (10)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Y | 4 | 7 | 3 | 9 | 17 |
| X | 2 | 3 | 1 | 5 | 9 |

1. State and prove the Gauss Markov theorem in case of a multiple variable regression analysis (10)
2. What are the assumptions of classical linear regression model. Explain each one of these assumptions focusing on the problem that may arise if these assumptions are relaxed. (10)
3. To assess the teaching quality of class teachers, a random sample of examination marks was obtained from three different class. Use ANOVA to see if there is any significant difference in marks among the three class. (10)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Class I | 85 | 75 | 82 | 76 | 71 | 85 |
| Class II | 71 | 75 | 73 | 74 | 69 | 82 |
| Class III | 59 | 64 | 62 | 69 | 75 | 67 |

1. What are the problems of using a linear probability model in case of a dichotomous dependent variable? Explain how the logistic regression address these problems. (10)