



Register Number:

DATE: 22-10-18

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

M.Sc. BIG DATA ANALYTICS- I SEMESTER
SEMESTER EXAMINATION - OCTOBER 2018

BDA:1118 – BASIC STATISTICAL METHODS

Time: 2½Hours

Max: 70Marks

This question paper has **TWO** printed pages and **TWO** parts

Note: Students are allowed to use scientific calculators (non-programmable only)

Statistical tables will be provided only on request

Students are expected to provide justification for selection of methods, along with meaningful interpretations of final results wherever applicable

PART – A

I Answer any FIVE of the following: 10 x 5 = 50

1. A) Design a data collection format for collecting feedback information about services from customers visiting **ABCD** bank (5)
B) Briefly explain the data collection system that you are planning to use to collect feedback information in above situation (1.A) (5)
2. A) Write a note on different scales of measurements with examples (5)
B) Briefly explain various data imputations methods (5)
3. A) Briefly explain various steps involved in data cleaning (5)
B) Write a note on principles of data visualization. (5)
4. A) Define Statistics and mention any three functions of statistics. (3)
B) Define Dispersion and mention any three measures of dispersion. (3)
C) Show that standard deviation is independent of change in scale. (4)
5. A) What do you mean by bivariate data? Give an example (2)
B) Differentiate between correlation and regression (3)
C) What is scatter diagram? Interpret various types of correlation coefficient using scatter diagram. (5)
6. A) Define central tendency and explain any two measures of it. (3)
B) What are partition values? How are they determined graphically? (4)
C) List out any two merits and demerits of geometric mean (3)
7. A) Derive least squares estimates of parameters of simple linear regression model. (7)
B) Write a note on coefficient of determination (3)
8. A) Write down mathematical model for simple linear regression model along with three assumptions (5)
B) Explain the procedure of testing independence of attributes (5)

CASE STUDY

I Answer anyone among the following:

20 x 1 = 20

9. A) A small study is conducted involving 8 infants to investigate the association between gestational age at birth, measured in weeks, and birth weight, measured in grams.

Infant ID	1	2	3	4	5	6	7	8
Gestational Age(wks)	34.7	36	29.3	40.1	35.7	42.4	40.3	37.3
Birth weight (gm)	1895	2030	1440	2835	3090	3827	3260	2690

- Find the correlation between Gestational Age and Birth Weight (2)
 - Fit a regression model for birth weight depending on gestational age (4)
 - Estimate weight of new baby when Gestational age is 34.3 weeks (1)
 - Comment on validity of estimates regression coefficients(r^2) (2)
 - Comment on assumption about normality (2)
- B) Suicide is the third-leading cause of death for adolescents between 15 and 22 years of age in the United States and the following data related to number of suicides in the year 1962, is obtained from an research article

Type of suicide	Age group					
	14-16		17-19		20-22	
	Boys	Girls	Boys	Girls	Boys	Girls
Firearms	42	37	32	24	14	25
Poison	123	160	178	125	20	24
Hanging	10	16	28	41	10	15
Others	25	36	38	27	16	15

- What is the most common method of suicide among adolescents? (1)
 - What percent of girls chose poison? (2)
 - Test the relationship between age group and type of suicide at 1 % l.os. (6)
10. A) Following data is collected from forty two students about number of *WhatsApp* groupsthat they are member of (active)

14	13	12	14	13	15	13	14	12	13	12	12	14	15	12	13	14	15	12	14	13
13	15	12	13	12	12	12	16	14	12	14	12	12	13	17	13	12	16	15	12	16

- Fit a suitable distribution to above data (4)
 - Test it for the goodness of fit at 5% level of significance. (3)
 - Estimate the number of students who are members of more than 15 groups in a college of total strength 5000 (2)
- B) The following data provides measurements of the girth, height and volume of timber in 10 felled black cherry trees. Note that girth is the diameter of the tree (in inches) measured at 4 ft 6 in above the ground.

Trees	1	2	3	4	5	6	7	8	9	10
Girth (diameter in inches)	8.3	8.6	8.8	10.5	10.7	10.8	11	11	11.1	11.2
Height (Height in ft)	70	65	63	72	81	83	66	75	80	75
Volume (in cubic ft)	10.3	10.3	10.2	16.4	18.8	19.7	15.6	18.2	22.6	19.9

Study above data using appropriate statistical methods and comment on the interrelationship between Girth and Height (7)

- C) A political scientist suspects that there is a relationship between the number of promises a political candidate makes and the number of promises that are fulfilled once the candidate

is elected. Following data is the "track record" of 10 politicians, compute rank correlation and comment on the result (3)

Politician	1	2	3	4	5	6	7	8	9	10
Promises Made	21	40	31	62	28	50	55	43	61	30
Promises Kept	7	5	6	1	5	3	2	6	3	5

BDA-1118-A-18

OCT 2018