Reg. No. ☐☐☐☐☐☐☐☐☐

BCACAC 385

**Choice Based Credit System Sixth Semester B.C.A. Degree
Examination, September 2022
(2021-2022 Batch Onwards)
E2 : BUSINESS STATISTICS WITH R**

Time : 3 Hours   Max. Marks : 80

Note : 1) Answer **any ten** questions from Part – **A** and **one full** question
from **each** Unit in Part – **B**.
2) **Scientific** calculator is **allowed**.

PART – A   **(10×2=20)**

1. a) What are Vectors in R ? Give example.

b) Define population and sample mean. Write its formulae.

c) What are ordinal data ? Give an example.

d) What is an Experiment ? Give example.

e) Define Mutually Exclusive Events. Give an example.

f) What are t Distribution ? Write the formula for the t statistic.

g) What is Hypothesis ? Give an example.

h) What are One Sample t Test ? List its uses.

i) Write the formulae of SSC and SSE for ONE WAY ANOVA.

j) What is Linear Correlation ? Give an example.

k) What is a Correlation matrix ? Give an example.

l) What is sum of square error ? Write its formulae.

PART – B

Unit – I

2. a) With an example explain how to create a matrix using :
i) Vector
ii) List
iii) matrix() function

b) The number of U.S. cars in service by top car rental companies in a recent year according to Auto Rental News follows. Compute the mode, the median and the mean.

Company	Number of Cars in Service
National/Alamo	233,000
Avis	204,000
Dollar/Thrifty	167,000
Budget	144,000
Advantage	20,000
U-Save	12,000
Payless	10,000
ACE	9,000
Fox	9,000
Rent-A-Wreck	7,000
Triangle	6,000

c) Compute the 35th percentile, the 55th percentile for the following data.

16 28 29 13 17 20 11 34 32 27 25 30 19 18 33 **(6+5+4)**

3. a) Explain the following :

   i) Data frames   ii) Factors

b) Shown here is a list of the top five industrial and farm equipment companies in the United States, along with their annual sales ($ millions). Construct a pie chart and a bar graph to represent these data, and label the slices with the appropriate percentages.

Firm	Revenue ($ million)
Caterpillar	30,251
Deere	19,986
Illinois Tool Works	11,731
Eaton	9,817
American Standard	9,509

c) Construct a stem and leaf plot using two digits for the stem.

212	239	240	218	222	249	265	224
257	271	266	234	239	219	255	260
243	261	249	230	246	263	235	229
218	238	254	249	250	263	229	221
253	227	270	257	261	238	240	239
273	220	226	239	258	259	230	262
255	226						

(5+5+5)

Unit – II

4. a) Explain Marginal, Joint and Conditional Probabilities with example.

   b) Client company data from the Decision Dilemma reveal that 155 employees worked one of four types of positions. Shown here again is the raw values matrix (also called a contingency table) with the frequency counts for each category and for subtotals and totals containing a breakdown of these employees by type of position and by sex. If an employee of the company is selected randomly, what is the probability that the employee is female or a professional worker ?

COMPANY HUMAN RESOURCE DATA

		Sex		
		Male	Female	
	Managerial	8	3	11
Type of Position	Professional	31	13	44
	Technical	52	17	69
	Clerical	9	22	31
		100	**55**	**155**

   c) GMAT test that has a mean of 494 and a standard deviation of 100. Assume MAT scores are normally distributed. What is the probability of randomly obtaining a score between 300 and 600 on the GMAT exam ?

   P $(300 < x < 600 | \mu = 494$ and $\sigma = 100) = ?$ **(6+5+4)**

5. a) Determine the mean, the variance and the standard deviation of the following discrete distribution.

Discrete Distribution of Occurrences of Daily Crises	
Number of Crises	Probability
0	.37
1	.31
2	.18
3	.09
4	.04
5	.01

b) Assuming x is normally distributed; use the following information to compute a 99% confidence interval to estimate the population mean. And 99% confidence level. (Using the t statistic) the sample mean is 2.14 and the sample standard deviation is 1.29.

3   1   3   2   5   1   2   1   4   2   1   3   1   1

c) A bank has an average random arrival rate of 3.2 customers every 4 minutes. What is the probability of getting exactly 10 customers during an 8-minute interval ? (Using Poisson Formula)   **(6+5+4)**

Unit – III

6. a) Use chi-square test to determine whether the observed frequencies are distributed the same as the expected frequencies. ($\alpha = .05$)

Category	f_o	f_e
1	53	68
2	37	42
3	32	33
4	28	22
5	18	10
6	15	8

b) To test the difference in the two methods, the managers randomly select one group of 15 newly hired employees to take the three-day seminar (method A) and a second group of 12 new employees for the two-day DVD method (method B). Table shows required data using $\alpha = .05$, the managers want to determine whether there is a significant difference in the mean scores of the two groups.

Method A	Method B
$\overline{x_1} = 47.73$	$\overline{x_2} = 56.5$
$S_1^2 = 19.495$	$S_2^2 = 18.273$
$n_1 = 15$	$n_2 = 12$

c) A random of sample size 20 is taken resulting in sample mean of 25.51 and a sample standard deviation of 2.1933. Assume data is normally distributed use this information and $\alpha = 0.05$ to test the following hypothesis. 4 or 5.

$H_0 : \mu = 25$ pounds

$H_a : \mu \neq 25$ pounds   **(6+5+4)**

7. a) Construct one way ANOVA table for following data.

1	2	3	4	5
14	10	11	16	14
13	9	12	17	12
10	12	13	14	13
	9	12	16	13
	10		17	12
				14

b) Suppose a stock market investor is interested in determining whether there is a significant difference in the P/E (price to earnings) ratio for companies from one year to the next. Assume $\alpha = .01$. Assume that differences in P/E ratios are normally distributed in the population.

Company	Year 1 P/E Ratio	Year 2 P/E Ratio
1	8.9	12.7
2	38.1	45.4
3	43.0	10.0
4	34.0	27.2
5	34.5	22.8
6	15.2	24.1
7	20.3	32.3
8	19.9	40.1
9	61.9	106.5

(9+6)

Unit – IV

8. a) From following information find the correlation coefficient between advertisement expenses and sales volume using Karl Pearson's coefficient of correlation method (Direct Method).

Firm	1	2	3	4	5	6	7	8	9	10
Advertisement Exp. (Rs. In Lakhs)	11	13	14	16	16	15	15	14	13	13
Sales Volume (Rs. In Lakhs)	50	50	55	60	65	65	65	60	60	50

b) Determine the regression line for the following data.

Number of Beds	FTEs
23	69
29	95
29	102
35	118
42	126

c) Compute the coefficient of determination (r^2) using direct and computation formulae. Given that SSE = 2448.86, b1 = 2.232, SSxx = 3838.667 and Syy = 21,564. (n = 12). **(5+5+5)**

9. a) Compute the predicted values of y and residuals, sum of squares of error (SSE) and the standard error of the estimate (Se).

EQUATION OF REGRESSION LINE 2.9550 + 0.0202x

Number of Passengers x	Cost ($1000) y
61	4.280
63	4.080
67	4.420
69	4.170
70	4.480

b) Test the slope of the regression model, determine whether there is a significant positive slope (One Tailed). Use a α = .01. Given that $\hat{Y} = 30.888 + 2.232x$, b1 = 2.232, Se = 15.65, Σx = 592, Σx^2 = 33044, n = 12.

c) If r_{12} = 0.60, r_{13} = 0.50 and r_{23} = 0.45 then calculate coefficients of Partial Determination of $r_{13.2}$. **(5+5+5)**

_____