Seat No.: _____ Enrolment No._____

GUJARAT TECHNOLOGICAL UNIVERSITY

MCA - SEMESTER–3 • EXAMINATION

Subject Code: 2630003

Subject Name: Statistical Methods (SM)

Sample Question Paper

Instructions:

1. Attempt all questions.
2. Make suitable assumptions wherever necessary.
3. Figures to the right indicate full marks.
4. Statistical tables for various distributions are permitted to be used.

Q. 1 (a) State whether the following statements are TRUE or FALSE. Justify your answer. [7]

1. Correlation between two variables will be high if the correlation coefficient (r) value is positive and it will be low if r is negative.
2. The variance of a set of random values is minus (–) 7.
3. For some events A and B, it is possible that $P(A) = 0.15$ and $P(A \cup B) = 0.12$.
4. The middle 50% data lie in the inter quartile range.
5. If $P(A \mid B) = P(A)$, the events A and B are independent.
6. Confidence interval will increase if the confidence level is reduced.
7. Degree of freedom in case of two sample of sizes 40 and 50 is 45.

Q. 1 (b) The probability that a fluorescent tube light has a useful life of at least 600 hours is 0.90. Find the probability that among 20 such lights,

1. Exactly 16 will have a useful life of at least 600 hours. [3]
2. None will have a useful life of at least 600 hours. [3]
3. At least one tube light will have a useful life of at least 600 hours.

[1]

Q. 2 (a) The average rejection rate of an Item-A is 1% and that of an Item-B is 0.8%. I purchased 1,000 pieces of Item-A and 1,000 pieces of Item-B. Answer the following:

1. Can it happen that I get higher rejections of Item-B as compared to that of Item-A? Justify your answer through example(s).
   [5]
2. What additional information should be provided to describe the quality of Item-A and Item-B more realistically? Justify by using the above example's data. [2]

Q. 2 (b) If a person finds that when he wears a particular yellow color shirt, he meets with a minor road accident most of the times. Answer the following:

1. Are the two events (“Wearing a particular yellow shirt” and “Meeting with a minor road accident”) correlated? Positively or negatively correlated? Give reasons.
   [2]
2. Are these events causally related? Justify your answer.
   [2]
3. “If two events are causally correlated then these will definitely be statistically correlated”. Is this statement always TRUE? If YES, justify and give an example. If NO, give a counter example.
   [3]

OR

Q. 2 (b) Prof. Hardtack gave three exams last semester in a large lecture class. The standard deviation $\sigma = 7$ was the same on both the exams, and scores were normally distributed. Below are scores for 10 randomly chosen students on each exam.

Exam-1: 71, 69, 78, 80, 72, 76, 70, 82, 76, 76

Exam-2: 77, 66, 71, 73, 94, 85, 83, 72, 89, 80

1. Find the 95% confidence interval for the mean score on each exam. [4]
2. Do the confidence intervals overlap? If so, what does this suggest? [3]

Q. 3 (a) Phone calls arrive at the rate of 48 per hour at the reservation desk for an organization. [7]

1. Compute the probability of receiving three calls in a five-minute interval of time.
2. Compute the probability of receiving exactly 10 calls in 15 minutes.

Q. 3 (b) Samples of two types of batteries were tested for length of life and the following data are obtained:

	Type-1	Type-2
Sample Size	n1=9	n2=16
Sample means	12,000 hrs.	11,200 hrs.
Standard Deviation	32 hrs.	40 hrs.

Test whether the difference in the means is significantly different regarding length of life of two types of batteries at 2% level of significance.

[7]

OR

Q. 3 (a) The population standard deviation is given as 15. If sample sizes is 100 and confidence level is 95%,

1. Compute the Margin of Error and Standard Error of the mean.
   [4]
2. What shall be the sample size if the standard error of the mean is to be reduced to half? [1]
3. How the Margin of Error will vary (i) if the sample size is increased, and (ii) if the confidence level is increased? [2]

Q. 3 (b) The sales data of an items in 8 shops before and after a special promotional campaign are as under:

Shop	1	2	3	4	5	6	7	8	
Sales Before campaign	51	29	30	49	52	41	37	44	
Sales After campaign		57	31	28	54	57	46	43	52

Use the '*paired t-test*' to check whether it can be concluded with 95% level of confidence that the sales increased after the promotional campaign?

[7]

Q. 4 (a) An airline serves bottles of Galena Spring Water that are supposed to contain an average of 10 ounces. The filling process follows a normal distribution with process

standard deviation 0.07 ounce. Twelve randomly chosen bottles had the weight shown below (in ounces). Set up a two-tailed decision rule to detect quality control violations using the 5% level of significance. State any assumptions you are making.

[7]

Q. 4 (b) Engineers are testing company fleet vehicles fuel economy (miles per gallon) performance by using different types of fuel. One vehicle of each size is tested. Does this sample prove that there is a significant difference in treatment means?

[7]

	87 Octane	89 Octane	91 Octane	Ethanol 5%	Ethanol 10%
Compact	27.2	30.0	30.3	26.8	25.8
Mid Size	23.0	25.6	28.6	26.6	23.3
Full Size	21.4	22.5	22.2	18.9	20.8
SUV	18.7	24.1	22.1	18.7	17.4

OR

Q. 4 (a) A study of post-coronary recovery showed that 11 of 39 patients without pets died within a year, compared with only 3 of 53 pet-owning patients. Calculate the survival rates and determine whether these are significantly different at 10% level of significance.

[7]

Q. 4 (b) Scrap rates per thousand parts (whose defects cannot be reworked) are compared for 5 randomly selected days at three plants. Does the data prove a significant difference in mean scrap rates?

[7]

Scrap Rate (Per 1000 Units)

Plant-A	Plant-B	Plant-C
11.4	11.1	10.2
12.5	14.1	9.5

10.1	16.8	9.0
13.8	13.2	13.3
13.7	14.6	5.9

Q. 5 (a) A high school guidance counselor collected the following data about the grade point averages (GPA) and the SAT mathematics test scores for the six seniors.

GPA	2.7	3.5	3.7	3.3	3.6	3.0
SAT	450	560	700	620	640	570

(i) Compute and interpret the sample covariance.

[4]

(ii) Compute the sample correlation coefficient and interpret the result.

[3]

Q. 5 (b) For the following data,

X	2	3	5	1	8
Y	25	25	20	30	16

Calculate

(1) Compute the estimated regression equation. [3]

(2) Compute SSE, SST AND SSR. [2]

(3) Compute the coefficient of determination r^2 and comment on the goodness of fit.

[2]

OR

Q. 5 (a) Following data shows prices for books and the number of pages that each book contains.

Book	A	B	C	D	E	F	G
Pages (x)	500	700	750	590	540	650	480
Price (y)	7	7.5	9	6.5	7.5	7	4.5

- Develop a least-squares estimated regression line.

  [4]

- Determine point estimate of mean value of price for 600 pages.

  [2]

- Compute the coefficient of determination and explain its meaning. [1]

Q. 5 (b) Given the data for two variables x and y.

xi	6	11	15	18	20
yi	6	8	12	20	30

a. Develop an estimated regression equation for these data.

[3]

b. Compute the residuals. [2]

c. Develop a plot of the residuals against the independent variable x.

[2]

High Bloom Questions are asked in Q. 2 only. These questions are taken from: (a) Descriptive Statistics including Correlation (Unit-1), (b) Probability & Probability Distribution, and (c) Sampling & Sampling Distribution

<u>Additional Questions under Q. 2</u>

From Descriptive statistics

Addl. Q. 1. MCA College-A reported average CPI of their students as 6.8 while College-B reported average CPI of their students as 5.9. If the eligibility criteria of an IT company to shortlist students is minimum 6.5 CPI, answer the following:

1. Can it happen that more students from College-B will be shortlisted than from College-A? If NO, give reasons. If YES, give an example to justify your statement.

   [3]

2. Whether any additional information (statistics) should be asked by the IT company from the colleges to ensure that the statistics submitted by the colleges describe their result more realistically? If yes, what should it be? Justify your statement by solving the above example.

   [4]

Addl. Q. 2. Hiren took a random sample of the number of steps per minutes from the

electronic readout of his aerobic climbing machine during a 1-hour workout.

(D)

90 110 97 144 54 60 156 86 82 64 100 47 80 164

93

(a) Calculate mean, median, and mode. [3]

(b) Which of these are the best and the worst measures of central tendency? Why? [2]

(c) What measure of Dispersion will be most appropriate for this data? Why? [2]

Addl. Q. 3. Analysis of annualized returns over the period 1991-2001 showed that prepaid tuition plans had a mean return of 6.3% with a standard deviation of 2.7%, while the Standard & Poor's 500 stock index had a mean return of 12.9% with a standard deviation of 15.8%. (D)

(a) Calculate and compare the coefficients of variation. [4]

(b) Why would we use a coefficient of variation? Why not just compare standard deviations? [1]

(c) What do the data say about risk and return of these investments at that time? [2]

From Probability & Probability Distribution

Addl. Q. 4. Suppose that a man is charged with a gambling offense, namely that he was using a ‘fixed’ die in which five of the six sides were 6’s. Let Hp be the (prosecution) hypothesis that the die was fixed, and let Hd be the alternative (defense) hypothesis that the die was not fixed (i. e. it was a ‘fair’ die). A key piece of evidence E presented by the prosecution is the observation that the outcome of two consecutive rolls of the die were two 6s. Calculate the prosecution likelihood, i. e. the probability P(E | Hp) and the defense likelihood, i. e. the probability P(E | Hd), stating any assumptions you are making. Hence, calculate the likelihood ratio of the evidence. What can you conclude about the (posterior) probability of guilt if (a) the prior probability P(Hp) = 0.5; (b) the prior probability P(Hp) =

0.001? [7]

From Sampling & Sampling Distribution

Addl. Q. 5. The diameter of busings turned out out of a manufacturing process is normally distributed random variable with a mean of 4.035 mm and a standard deviation of 0.005. The inspection procedure requires a sample of 25 bushings once an hour.

(a) Within what range should 95% of the bushing diameters fall?
[3]

(b) Within what range should 95% of the sample means fall?
[2]

(c) What conclusion would you reach if you saw a sample mean of 4.020?
[2]