

Seat No _____

Enrollment No _____

GUJARAT TECHNOLOGICAL UNIVERSITY
MCA - SEMESTER-III • EXAMINATION – WINTER 2016

Subject Code: 2630003

Subject Name: Statistical Methods (SM)

Time: 10.30 AM TO 01.00 PM

Date: 02/01/2017

Total Marks: 70

Instructions:

1. Attempt all questions.
2. Make suitable assumptions wherever necessary and specifically write your assumptions.
3. Figures to the right indicate full marks.
4. Statistical tables are permitted to students for reference.

Q. 1 (a) State whether the following statements are TRUE or FALSE. Give justification / reason. **07**

1. If correlation coefficient between two variables x and y is -0.8 , it implies that there is good (sufficiently strong) correlation between x and y .
2. For two events A and B , if $P(A) = 0.47$, it is possible that $P(A \cup B) = 0.46$.
3. Standard error will decrease when sample size increases.
4. The median of marks of 60 students is 56%. It will imply that Q1 will be half of 56%, i. e. 28%.
5. If Null Hypothesis is accepted when it is false, it will be Type-I error.
6. Distribution of sample means (also called Sampling Distribution) for a set of large samples is Normal Distribution, even though the Population is NOT Normally distributed.
7. (Algebraic) Sum of deviations from the mean can have a negative value.

Q. 1 (b)

1. State Central Limit Theorem. What is its importance? **02**
2. State Chebychev's theorem. Use it to compute the percentage of data lying between 34 and 62 for a sample with mean as 48 and standard deviation as 7. **02**
3. What is meant by (i) Mutually Exclusive Events, and (ii) Independent Events. For two events A and B , write down the conditions for these events being Mutually Exclusive, and these events being Independent. **03**

Q. 2 (a) An IT company wishes to visit only one MCA institute in North Gujarat for recruitment and the company asks all the interested institutes in that region to submit the result summary. Only two institutes submit the result summary. Inst-A reported the Average CPI of their students as 6.7 and Inst-B reported the Average CPI of their students as 6.1. Answer the following questions to help the IT company to decide on which institute to visit for placement:

1. Can it happen that Inst-B will have a higher percentage of students with CPI ≥ 6.5 than the Inst-A? Justify your answer by reasoning and by giving an example. Example may contain small data. **04**
2. Whether Percentile would be a better measure for reporting to the IT company by Inst-A and Inst-B? Justify your answer. **03**

- Q. 2 (b)** Thirty-six blood samples were selected by taking every 7th blood sample from racks holding 254 blood samples from the morning draw at a medical center. The White Blood Count (WBC) was measured and the mean WBC was 8.6 with a standard deviation of 1.8. Assume normality of WBC and answer the following questions based on the above inputs:
1. Within what range should 95% of WBC fall? **02**
 2. Within what range should 95% of sample means fall? Will this range be different from the one obtained in (1) above? Justify the reason for it being different. **03**
 3. What type of sampling method is used? Will the thirty-six samples so obtained be random samples? Justify. **02**

OR

- Q. 2 (b)** A test for ovarian cancer has a 5% rate of false positives and a 2% rate of false negatives. On average, 1 in every 2,500 American women over age 35 actually has ovarian cancer. **07**
1. If a woman over 35 years of age tests positive, what is the probability that she actually has cancer? Explain your reasoning.
 2. If a woman over age 35 years of age tests negative, what is the probability that she actually has ovarian cancer?

- Q. 3 (a)** A study by the Parents' television Council showed that 80% of movie commercials aired on network television between 8 and 9 pm (the prime family viewing hours) were for R-rated films. Answer the following questions:
1. Find the probability that in 10 commercials during this time slot, at least 9 will be for R-rated films. **05**
 2. Find the probability that in 10 commercials during this time slot, at most 8 will be for R-rated films. **02**

- (b)** A sample of 25 minivan electrical warranty repairs showed a mean repair cost of \$45 with a standard deviation of \$27
1. Construct a 95% confidence interval for the true mean repair cost. **04**
 2. How could the confidence interval be made narrower? Give reasons. **02**
 3. Whether 90% confidence interval will be a subset of or a superset of 95% confidence interval? **01**

OR

- Q. 3 (a)** The caffeine content of a cup of home-brewed coffee is a normally distributed random variable with a mean of 115 mg with a standard deviation of 20 mg. Answer the following questions:
1. What is the probability that a randomly chosen cup of home-brewed coffee will have more than 130 mg of caffeine? **04**
 2. A very strong cup of tea has a caffeine content of 91 mg. What is the probability that a cup of home-brewed coffee will have less caffeine than a very strong cup of tea? **03**
- (b)** A pediatrician's records showed the mean height of a random sample of 25 girls at age 12 months to be 29.530 inches with a standard deviation of 1.0953 inches.
1. Construct a 95% confidence interval for the true mean height. **03**
 2. What sample size would be needed for 95% confidence and an error of ± 0.20 **04**

inches?

- Q. 4 (a)** The average weight of a package of rolled oats is supposed to be at least 18 ounces. A sample of 18 packages shows a mean of 17.78 ounces with a standard deviation of 0.41 ounces. **05**
1. At the 5% level of significance (α), is the true mean smaller than the specification? Clearly state your hypotheses and decision rule. **02**
 2. Is this conclusion sensitive to (independent of) the choice of α ? **02**
- (b)** A 20-minute consumer survey mailed to 500 adults aged 25-34 included a \$5 Starbucks gift certificate. The same survey was mailed to 500 adults aged 25-34 without the gift certificates. There were 65 responses from the first group and 45 responses from the second group. **04**
1. Perform a two-tailed test comparing the response rates (proportions) at $\alpha = 0.05$. **04**
 2. Form a confidence interval for the difference of proportions at $\alpha = 0.05$, without pooling the samples. **03**

OR

- Q. 4 (a)** An automaker states that its cars equipped with electronic fuel injection and computerized engine controls will start on the first try (hot or cold) 99% of the time. A survey of 100 new car owners revealed that 3 had not started on the first try during a recent cold snap. **05**
1. At $\alpha = 0.025$ does this demonstrate that the auto-maker's claim is incorrect? **02**
 2. Calculate the p-value at $\alpha = 0.025$ and interpret it. **02**
- (b)** At $\alpha = 0.05$, does the following sample show that daughters are taller than their mothers? **07**

Family	Daughter's Height (cm)	Mother's Height (cm)
1	167	172
2	166	162
3	176	157
4	171	159
5	165	157
6	181	177
7	173	174

- Q. 5 (a)** Part-time weekly earnings by college students are given below:
- | | | | | | |
|-------------------|----|-----|-----|-----|-----|
| Hours Worked (x): | 10 | 15 | 20 | 20 | 35 |
| Weekly Pay (y): | 91 | 171 | 204 | 156 | 261 |
1. Develop the estimated regression equation for these data. **03**
 2. Compute SSE, SSR, and SST. **03**
 3. Compute Coefficient of Determination. **01**

- (b) For the following data on Number of Orders (x) and Shipping Cost (y):
x: 1068, 1026, 767, 885, 1156, 1146, 892, 938, 769, 677, 1174, 1009
y: 4489, 5611, 3290, 4113, 4883, 5425, 4414, 5506, 3346, 3673, 6542, 5088

1. Develop the estimated regression equation. **04**
2. Use the estimated regression equation to predict the value of y for x = 1200 **01**
3. Compute the residuals for individual values. **02**

OR

- Q. 5 (a)** Shown below are monthly food and beverage sales in the United States during a recent year over an 8-month period (\$ million).

Month	Sales (\$ Million)
January	32,569
February	32,274
March	32,583
April	32,304
May	32,149
June	32,077
July	31,989
August	31,977

1. Develop the equation of a trend line through these data and use the equation to forecast sales for October month. **05**
 2. Compute SSE. **02**
- Hint: Take numerical values for month

- (b) Shown below is the data on US exports of fertilizers to Indonesia over a 5-year period provided by the US Census Bureau.

Year	Fertilizers (\$ Millions)
2001	11.9
2002	17.9
2003	22.0
2004	21.8
2005	26.0

1. Compute Coefficient of Correlation. Can we conclude that fertilizers exports depend on year of export? **03**
2. Determine the equation of the trend line for the data given above. **02**
3. Compute MSE, & MSR **02**
