

University of Calcutta  
M. Sc Semester – III (CBCS), 2020  
CBCC STSG Gen 31 (Statistics)

**Time: 2.30 hours (including 30 minutes for scanning and sending the script)**

**Full Marks: 50**

**INSTRUCTIONS FOR STUDENTS**

1. Answer **any ten** questions, which are of equal values
2. Use your university roll number as the file name of the scanned pdf file. For example, if your roll number is **abc/def/181007**, your file name must be **abc\_def\_181007**.
3. The students are required to submit the scanned answer scripts by e-mail to the coordinator (**Email id: statmethodexam@gmail.com**) with a copy to the concerned authority (HOD or the Principal) within the stipulated time **1.30 pm**.
4. **Students who do not submit the paper within the stipulated time will be marked absent.**
5. **Email submission of the answer script will have the subject STATMETHODS.**
6. **All answers must be hand-written. No typed or copy-pasted answer will be accepted, and such answers, if submitted, will be marked zero.**

Candidates are required to give their answers in their own words as far as practicable

Notations and symbols are as usual

1. Distinguish between observational study and experimental study. Suppose one is interested to assess the health effect of exposure to lead. Describe the study design.
2. What do you mean by an ogive of a frequency distribution? Describe its utility in approximately finding various robust summarizing measures of a data, explaining the meaning of the word "robust" in this context.
3. "Box plot is a five number summary of the data" - Explain. Describe how box plot may be used to assess the skewness and kurtosis in the data.
4. Explain how a polynomial regression may be constructed for a bivariate data. How do you assess about the goodness of fit of such polynomial regression? What can be said about the performance of such assessment?
5. What do you mean by Simpson's paradox? Explain its concept with appropriate examples. Suggest your recommendation regarding the final findings when such paradox is present in the data.
6. Consider a set of three cards with (i) both sides of the first card are colored red (ii) both sides of the second card are colored black and (iii) one side of the third card is colored red and the other side is colored black. One of the cards is randomly selected and the color of the upper side happens to be red. What is the probability that the other side is colored black?
7. (a) Suppose that  $P(A) = 1/2$  and  $P(B) = 2/3$ . Give separate examples, where (i)  $P(A \text{ and } B) = 1/6$  and (ii)  $P(A \text{ or } B) = 5/6$ .  
(b) If  $P(A) = P(B) = P(C) = 1$ , determine  $P(A \text{ and } B \text{ and } C)$ .
8. Consider two tosses of a coin with probability of Head as  $p$  ( $0 < p < 1$ ). Define two events  $A$  and  $B$  such that  $P(A|B)$  is independent of  $p$  but  $0 < P(A|B) < 1$ . Calculation of  $P(A|B)$  needs to be shown.
9. Arrival of cars at a fuel refilling station is Poisson with average rate of 50 per hour. The station has only one refilling pump and all cars require one minute to get fuel. Find the probability that a car will have to wait for getting the fuel.
10. A box contains 5 tickets numbered from 1, 2, ..., 5. Two tickets are drawn at random and with replacement from the box. If  $X$  is the largest number of tickets drawn, find the probability mass function of  $X$ . Hence or otherwise compute  $E(X)$ .
11. What is meant by an unbiased estimator? For a  $N(50, \sigma^2)$  population, find an unbiased estimator of  $\sigma^2$ . Given that the mean-square-error of an estimator is 500 square-units and its standard deviation 20 units, what can you say about the bias of the estimator?
12. What is a confidence interval? The sample mean and the sample standard deviation of the lengths of 100 electrical cables came out as 80 meters and 4 meters respectively. Assuming that the lengths of the cables follow a  $N(\mu, \sigma^2)$  distribution, find the 95% confidence interval for the mean length of cables. ( $t_{0.05, 99} = 1.64$ ,  $t_{0.025, 99} = 1.96$ )
13. Mr. A and Mr. B are the only two candidates contesting an election in a constituency of 50000 voters. In an opinion poll with sample size 1500, Mr. B received 650 votes. Do you think Mr. B has less chance of winning the election than Mr. A?

14. What is the power of a test? Based on a sample of size 16 from a  $N(\mu, 25)$  population, the critical region of level 5% for testing  $H_0 : \mu = 50$  against  $H_0 : \mu < 50$ , was found to be  $W : \bar{x} < 47.95$ . Find the powers of the test at  $\mu = 48$  and  $\mu = 45$ .
15. To compare the summer day and night time pollution levels ( $PM_{2.5}$  measured in  $\mu\text{g}/\text{m}^3$ ) in a city, observations were taken on 30 days of April. The day and night time means came out as 210 and 150 respectively, while the variances were 170 and 80 respectively. The correlation coefficient was 0.8. Can we say that the pollution level during the nights is on the average 50 units less than that of the day?