

M. Sc. EPIDEMIOLOGY
SEMESTER - II EXAMINATION, JUNE - 2026
(CMEP507) ADVANCED BIOSTATISTICS AND DATA HANDLING

[Time Allotted: 3:00 Hours]**[Max. Marks: 100]****Note:** Read all the instructions carefully

Answer all questions in serial order, Draw suitable diagrams wherever necessary

- Q. 1. Objective type questions: (Attempt All)** **(10 x 02 = 20)**
 (Very short/ Short answer/ MCQ/ One liners/ Fill in the blanks/ True/False)
- Wilcoxon signed test (CO1 PO1)
 - Level of significance (CO1 PO1)
 - Scatter plot of correlation (CO 1 PO2)
 - Journal club (CO 2 PO5)
 - Imputation is just making up data to artificially inflate results. It is better to just drop cases with missing data than to impute. **(True/False)** (CO 2 PO 1)
 - Excel does not format the entry in the formatting bar. **(True/False)** (CO 3 PO2)
 - The appropriate test for comparing the means of two populations using ordinal-level data from two independent samples is the Mann-Whitney U test. **(True/False)** (CO 1 PO1)
 - Write down Spearman's with tied rank correlation formula (CO 1 PO 5)
 - What are the continuous random variables? (CO 1 PO1)
 - Addition theorem of probability for three events, say A, B, and C. (CO 2 PO1)
- Q. 2. Short notes / Short Answer Type (Attempt any five out of seven)** **(5 × 4 = 20)**
- What are the key features of primary data to that of secondary data? (CO-3, PO-1)
 - Procedure to calculate Mann Whitney U test statistic. (CO-1, PO-2)
 - What are the key points a researcher should keep in mind while doing epidemiological data analysis. (CO-1, PO-2)
 - If A and B are two independent events such that $P(A \cup B) = 0.3$ and $P(A) = 0.6$. Find $P(B)$. (CO-2, PO-2)
 - Open Epi (CO-3, PO-5)
 - Discuss distribution function of two-dimensional random variables. (CO-2, PO-2)
 - Explain the term "research" and define characteristics of good research. (CO-1, PO-2)
- Q. 3. Structured Question (Attempt any three out of four)** **(3 × 10 = 30)**
- Explain data cleaning steps in MS Excel. (CO-3, PO-5)
 - What are different ways to deal with missing data? (CO-3, PO-2)
 - Define correlation. Find Karl Pearson's correlation coefficient for the given data. (CO-1, PO-2)
 Height of father (in inches): 69, 58, 65, 71, 69, 64, 69, 68
 Height of son (in inches): 72, 68, 57, 66, 71, 67, 70, 72
 - There are two identical bottles (A and B) containing respectively 8 red eyed and 4 white eyed Drosophila and 5 red eyed and 3 white eyed Drosophila. A bottle is chosen randomly, and a Drosophila fly is taken from it. If the fly is white, what is the probability that the red fly is taken from the second box (B)? (CO-2, PO-5)
- Q. 4. Long Answer/Essay type (Attempt any two out of three)** **(2 × 15 = 30)**
- Data Management of Hospital-based Registry System (CO-4, PO-5)
 - Explain Binomial distribution. What are the probabilities that a family (parents are heterozygous) of five will consist of blue eyed and 2 brown eyed brown eyed children (in any order) (Blue is dominant over brown). (CO-2, PO-5)
 - An IQ test was administered to 5 persons before and after they are trained. The results are given below. Test whether there is any change in IQ after training. It is given that the tabulated value at 1% level of significance for 4 df is 4.6. (CO-1, PO-1)

IQ before training	123	120	113	115	135
IQ after training	119	123	128	133	134