INDIAN STATISTICAL INSTITUTE

STATISTICAL METHODS I

B-Stat I, First Semester 2025-2026

Mid Semester Examination

Date: 09.09.2025 Maximum marks: 60 Duration: 2 hours.

1. The mean and standard deviation of 100 items are 50 and 5, respectively; that of another distinct set of 150 items are 40 and 6, respectively Find the combined standard deviation of all 250 items taken together.

   [Note: Here the standard deviation is based on a divisor of n, the sample size, not $(n-1)$.] [10]

2. Let $X_1, X_2, \ldots, X_n$ be n independent Uniform$(0, \theta)$ random variables, where θ is a positive real constant. Let $M = X_{(n)} = \text{Maximum}(X_1, X_2, \ldots, X_n)$ be the largest of these n random variables. Show that the probability density function of M is given by

$$f_M(x) = \frac{nx^{n-1}}{\theta^n}, \quad 0 < x < \theta.$$

[10]

3. Suppose that X and Y have a continuous joint distribution with the following joint pdf:

$$f(x, y) = c(x^2 + y) \text{ for } 0 \leq y \leq 1 - x^2.$$

   Determine

   (a) The value of the constant c.

   (b) $P(Y \leq X + 1)$.

[5+5=10]

4. A shipment of $N = 240$ tubes (used in the manufacture of televisions), contains, unknown to us, 15 defective items.

   Seven tubes are selected at random (without replacement) from this lot.

   (a) What is the probability that two out of these seven tubes are defective?

   (b) What is the expected number of defective items among these seven tubes?

[5+5=10]

5. Suppose X and Y are independent continuous random variables, and let $F_X(\cdot)$, $f_X(\cdot)$ be the cumulative distribution function and the probability density function of X, respectively. Similarly, let $F_Y(\cdot)$ and $f_Y(\cdot)$ be the same measures for Y.

   Let $F_{X+Y}(\cdot)$ represent the cumulative distribution function of $X + Y$. Show that

   $$F_{X+Y}(a) = \int_{-\infty}^{\infty} F_X(a - y) f_Y(y) dy.$$

   (The cumulative distribution function F_{X+Y} is called the convolution of the distributions F_X and F_Y.) [10]

6. Consider the linear regression problem. Normally we consider Y to be the response variable, and X to be the explanatory variable, so statistically, the linear regression model has the form, conditional on $X = x$,

   $$Y = \alpha + \beta x + \epsilon,$$

   where we need to estimate the parameters α and β to fit the regression model.

   Given a set of n random pairs of values of X and Y, consider the two possible regression lines: the first is our usual regression line where Y is the response and X is explanatory; the second is where we treat X as the response variable, and consider Y as the explanatory variable. We fit both models by the method of least squares.

   Will these produce identical or (possibly) distinct regression lines? If they are not necessarily identical, explain why they are not so. Can these two regression lines be parallel? Can you think of some restriction under which the regression lines must be identical?

[5+3+2=10]