

2020

## STATISTICS — HONOURS

Paper : CC-1

Full Marks : 50

*The figures in the margin indicate full marks.**Candidates are required to give their answers in their own words as far as practicable.*

(Notations and symbols are of usual significance)

1. Answer *any ten* questions :

1×10

- (a) Give an example of a discrete variable assuming non-integral values.
- (b) What is relative frequency density?
- (c) “Cumulative frequency of less-than type of the class interval with class boundaries 40 and 50 is 173” – interpret.
- (d) Suppose a variable is of the form ‘ $x$  per unit  $y$ ’ and equal amounts of  $x$  are considered. What would be the appropriate average?
- (e) What do you mean by decile of order  $k$  for a set of observations?
- (f) Suppose the mutual absolute differences of observations in a set are 1, 1, 1, 1, 1, 1, 1, 1, 2, 2, 2, 2, 2, 2, 3, 3, 3, 3, 4, 4. What will be the standard deviation of the observations?
- (g) State, without calculation, whether the following statement is correct and why?

$$\sum_{i=1}^{10} |i - 5.2| = \sum_{i=1}^{10} |i - 5.5|$$

- (h) Suppose the mean of a set of 10 observations is 25, and the observations more than 25 are 31, 35, 43, 39 and 37. Find the mean deviation about mean.
- (i) Suppose  $x$ ,  $y$  and  $z$  be three variables such that  $y = 3x - 2$ ,  $z = 2 - 3x$ . If  $V(x) = 1$ , find a relation between the standard deviations of  $y$  and  $z$ .
- (j) Suppose two variables  $x$ ,  $y$  are related as  $y = 2 - 3x$ . If the lower and upper quartiles of  $x$  are 5 and 15, respectively, find the semi-interquartile range of  $y$ .
- (k) Suppose the correlation coefficient between two variables  $x$  and  $y$  is twice the ratio of their standard deviations. Find the regression coefficient of  $y$  on  $x$ .
- (l) State when or why the moment-based measures of skewness and kurtosis, viz.  $b_1$  and  $b_2$  can or cannot be equal.
- (m) What does  $(e_{yx}^2 - r^2)$  measure?

Please Turn Over

- (n) Suppose  $r_s$  be the Spearman's rank correlation coefficient based on the ranks  $(x_i, y_i), i = 1(1)n$ , of  $n$  paired observations with no tie. If  $x_i$ 's are replaced by  $u_i$ 's, where  $u_i = n - x_i + 1, i = 1(1)n$ , what would be the Spearman's rank correlation coefficient between  $u$  and  $y$  in terms of  $r_s$ ?
- (o) Define odds ratio.

2. Answer **any four** questions : 5×4

- (a) Suppose  $x_1, x_2, \dots, x_n$  are  $n$  positive observations of a variable  $x$ , and  $A$  is their arithmetic mean. Then show that the arithmetic mean of the logarithms of these observations cannot exceed  $\log_e A$ .
- (b) Suppose  $M_1$  and  $M_2$  be the medians of two groups having  $n_1$  and  $n_2$  observations, respectively. If  $M$  be the median of the combined group, then show that  $M$  lies between  $M_1$  and  $M_2$ .
- (c) In a bivariate data analysis with two variables  $x$  and  $y$ , suppose the two regression lines are  $y = a + bx$  and  $y = c + dx$ , where  $a, b, c, d$  are non-zero constants. Describe how one can identify the one of predicting  $x$ .
- (d) Define correlation index of order  $p$  ( $r_p$ ) and correlation ratio ( $e_{yx}$ ). Show that

$$r_p^2 \leq e_{yx}^2.$$

- (e) Suppose two qualitative characters are categorized in a  $2 \times 2$  contingency table. Suggest a suitable measure of association between them. Justify whether your suggested measure can be useful when one of the characters is quantitative.
- (f) Describe how one can use Box plots to compare two frequency distributions.

3. Answer **any two** questions :

- (a) (i) Suppose a variable  $x$  assumes only two distinct values with equal frequency. Find a relation between the standard deviation and range of  $x$ . Does this relation hold universally? Justify your answer with appropriate reasoning.
- (ii) Suppose a variable  $x$  assumes the values  $a, b$  and  $(n - 2)$  other values all equal to  $\frac{a+b}{2}$ . Find a relation between the standard deviation and range of  $x$ . Does this relation hold universally? Give appropriate explanation to justify your answer. 5+5
- (b) (i) Express  $r$ -th order moment about an arbitrary origin  $A$  in terms of  $r$ -th and lower order moments about another arbitrary origin  $B$ . Hence provide the expression for the  $r$ -th order central moment in terms of  $r$ -th and lower order raw moments.
- (ii) Define skewness and kurtosis of a frequency distribution. Suggest two measures of skewness, one ranging between  $-1$  and  $+1$ , and the other between  $-3$  and  $+3$ . Justify the range for any one of them. 5+5
- (c) Describe the motivation behind considering correlation coefficient ( $r_{xy}$ ) as a measure of association between two variables  $x$  and  $y$ . Interpret the cases when  $r_{xy} = +1, -1, 1 - \epsilon, \epsilon - 1, 0; 0 < \epsilon < 0.2$ . 5+5

---