2021

STATISTICS

[HONOURS]

Paper : II

Full Marks : 75   Time : 4 Hours

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

Answer all the questions.

1. Answer any **five** questions: 1×5=5

    a) What kind of diagram is usually used to represent time series data?

    b) Define skewness.

    c) Give a practical example of bivariate situation where any one variable can be taken as response and the other as explanatory variable.

    d) Point out one use of rank correlation coefficient.

    e) Information is obtained regarding the number, of complete years of study of parents of the students studying in a school. Specify the scale of measurement.

    f) Mention one situation where harmonic mean is preferred.

    g) In a ratio chart if you see collinear points, what can be said about the data set?

2. Answer any **six** questions: 2×6=12

    a) Write down the expression for the angle between two regression lines, explaining all the symbols used.

    b) Define correlation ratio.

    c) What is 'Kurtosis' of a distribution? Give a measure of it.

    d) Write down a situation where geometric mean is the appropriate measure of average.

    e) Define complete association in case of a 2×2 cross classification.

    f) Suggest a measure of dispersion in terms of quartiles.

    g) What is stem-and-leaf diagram?

    h) What are secondary data? Give an example.

3. Answer any **three** questions: 6×3=18

    a) Distinguish between the correlation approach and the regression approach to the analysis of bivariate data.

b) Deduce a result to show that higher the absolute value of correlation coefficient better is the linear regression for the given set of bivariate data.

c) In case of two attributes each of which is dichotomous, write down Yule's measure of colligation. Defining odds ratio, establish a relationship between Yule's measure and odds ratio.

d) How can you construct box plot? Point out its use in analysing statistical data.

e) let x be a variable assuming the values 1,2,...k and let F_1 = n, F_2,..., F_k be the corresponding cumulative frequencies of 'greater than type'. Show that:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{k} F_i$$

f) Suppose that the variable x takes positive values only and that the deviations $(x_i - m)$ of the individual observations are small as compared to m, where m is the mean of x. Then show that $X_h \approx m\left(1 - s^2/m^2\right)$.

4. Answer any **four** questions: 10×4=40

a) i) Define correlation index. Show that as the degree of the polynomial increases, the value of the correlation index also increases. 7

ii) In the context of case-control studies, define the term odds-ratio. What does a high value of odds-ratio signify? 3

b) Define intra-class correlation (r_I). Derive the form of r_I when a variables x is measured for p families each containing members. Obtain the limits of r_I with their interpretations. 10

c) i) Derive the least-square linear regression of y on x based on n observations $(x_i, y_i), i = 1, L......n$. How would you judge the performance of this linear regression for predicting y on the basis of x? 5

ii) Write a note on Lorenz curves. 5

d) i) Define multiple correlation coefficient of x_1 on x_2, x_3, ..., x_p. What does it measure? Show that it lies between 0 and 1. 6

ii) What is partial correlation coefficient? How does it differ form total correlation coefficient? Write down a formula for $r_{12.34...p}$ in terms of the cofactors of the correlation matrix. Indicate its range. 4

e) In case of no tie, write down the formula for Spearman's rank correlation coefficient. Show that this can be put in the form of a product moment correlation coefficient. If a tie of length k is found in one set of ranking, how would you modify the formula? 10

f) i) For k×l contingency table, discuss different measures of association. 6

ii) Show that, for a 2×2 cross classification,

$$Q = \frac{2Y}{1+Y^2},$$

and hence that Q is greater in absolute value than Y, except when both are zero or ± 1. 4

––––––––––––