(b) Show that the angle θ, between the two lines of regression is

$$\tan\theta = \frac{1-r^2}{|r|}\left(\frac{\sigma_X\sigma_Y}{\sigma_X^2+\sigma_Y^2}\right)$$

where σ_1 and σ_2 are the standard deviations of X and Y respectively, and r is the correlation coefficient between X and Y.

Also, interpret the cases when $r = 0$ and $r = \pm 1$.

(8,7)

[This question paper contains 8 printed pages.]

Your Roll No.....

Sr. No. of Question Paper : 1031 **D**

Unique Paper Code : 2372011101

Name of the Paper : Descriptive Statistics

Name of the Course : **B.Sc. (Hons.) Statistics, (NEP-UGCF)**

Semester : I

Duration : 3 Hours Maximum Marks : 90

Instructions for Candidates

1. Write your Roll No. on the top immediately on receipt of this question paper.
2. Attempt **six** questions in all.
3. Question No. **1** is compulsory.
4. Attempt **five** more questions selecting at least **two** questions from each section.
5. Use of non-programmable calculator is allowed.

1. (a) Fill in the blanks :

(i) A histogram helps in determining the value of _____ .

(ii) The sum of squares of deviations of a set of values is minimum when taken about _____ .

(iii) The G.M. and A.M. of a distribution are 27 and 30 respectively, then H.M. is _____ .

(iv) The mean of 6 observations is 8. A new observation 8 is added, then the mean of 7 observation is _____ .

(v) Median can be determined graphically from _____ .

(vi) The standard deviation of 15 items is 6. If each item is increased by 2, then the new standard deviation will be _____ .

(vii) If both the regression lines are perpendicular to each other, then we say that there is _____ correlation between X and Y.

(viii) For a platykurtic distribution, γ_2 is _____.

(ix) For n attributes, the total class frequencies are _____, and the total ultimate class frequencies are _____ .

7. (a) Define Spearman's rank correlation coefficient. If d_i be the difference in the ranks of the i^{th} individual in two different characteristics, then show that the maximum value of $\sum_{i=1}^{n} d_i^2$ is $\frac{1}{3}(n^3 - n)$. Hence or otherwise, show that the rank correlation coefficient lies between -1 and $+1$.

(b) Let X and Y be two random variables with variances σ_X^2 and σ_Y^2, respectively and 'r' be the coefficient of correlation between them. If $U = X + kY$ and $V = X + (\sigma_X/\sigma_y)Y$ then find the value of 'k', so that U and V are uncorrelated.

(8,7)

8. (a) Obtain the equation of the line of regression of Y on X and hence write the equation of the line of regression of X on Y. Also, find the two regression coefficients. Show that the coefficient of correlation is the geometric mean of coefficients of regression.

(i) In terms of positives frequencies

(ii) In terms of ultimate class frequencies. (7,8)

6. (a) Define Yule's coefficient of association (Q) and coefficient of colligation (Y). Show that : $Q = \frac{2Y}{1+Y^2}$. Also, find the limits of Q and Y.

(b) In an investigation relating to health and nutrition of children between the age of one and five years, two group of children were compared, one belonging to the well-to-do class with 125 children, and the other belonging to the poor class with 124 children. The following results were obtained:

	Poor children	Well-to-do children
Below normal weight	75	23
Above normal weight	5	42

Find the coefficient of association between the weight of children and their parents' financial condition. (8,7)

(b) If $r(X,Y) = 0.8$, $\sigma_X = 2.5$ and $\sigma_Y = 3.5$, then find $Var(3X - 2Y)$.

(c) The Karl Pearson's coefficient of skewness based on measures of central tendency of a distribution is 0.32. The standard deviation and mean are 6.5 and 29.6 respectively. Find the mode of the distribution.

(d) Given $(A) = (\alpha) = (B) = (\beta) = \frac{N}{2}$, show that $(AB) = (\alpha\beta)$. (9×1,3×2)

SECTION – A

2. (a) Explain the term dispersion. Write its different measures. Show that for any discrete distribution, the standard deviation is never less than mean deviation about mean.

(b) In a frequency table, the upper boundary of each class interval has a constant ratio to the, lower boundary. Show that the geometric mean G may be expressed by the formula

$$\log G = x_0 + \frac{c}{N}\sum_i f_i(i-1)$$

where x_0 is the logarithm of the mid-value of the first interval and 'c' is the logarithm of the ratio between upper and lower boundaries. (8,7)

3. (a) Define Pearson's coefficient β_1, and β_2. Discuss their utility in statistics. Also, show that for any discrete distribution $\beta_2 > 1$.

(b) Show that in a discrete series if deviation are small compared with mean M so that $(x/M)^3$ and higher power of (x/M) are neglected, then

$$H = M\left(1 - \frac{\sigma^2}{M^2}\right)$$

where, M is the arithmetic mean, H is the harmonic mean, and σ is the standard deviation of the distribution. (7,8)

4. (a) Find the mean deviation from the mean and standard deviation of the following series

a, a + d, a + 2d,....., a + 2nd.

Also, verify that the latter is greater than the former.

(b) The first four moments of a distribution about the origin of variable X are −1.5, 17, −30, and 108 respectively. Find the first three moments about mean. Hence find mean, variance, coefficient of variation, and coefficient of skewness. Interpret the result. (8,7)

SECTION – B

5. (a) Use the principle of least squares to fit the curve $y = ab^x$ to a given set of n points,

$\{(x_1, y_1); i = 1,2,.....n\}$.

(b) What is Meant by independence of attributes? Give a criterion of independence for attributes of A and B :