This question paper contains 8 printed pages]

HPAS (MAIN)—2017

STATISTICS

Paper I

Time : 3 *Hours* *Maximum Marks* : 100

Note :— Attempt Question No. 1 which is compulsory and any other *four* questions from the rest. *All* questions carry equal marks. Symbols used in questions have their usual meanings.

1. (*a*) A survey conducted over 100 persons living in a city centre and observed that among these 100 persons 65 persons read the newspaper A, 40 persons read the newspaper B, 70 persons read the paper C, 30 persons read both A and B, 40 persons read both A and C, 25 persons read both B and C and 20 persons read all the three papers. A person is selected at random. What is the probability that he reads :

   (*i*) only A

   (*ii*) only A or B

   (*iii*) none of the three papers

   (*iv*) at least one paper.

P.T.O.

(b) Give axiomatic definition of probability. Prove that for n events $A_1, A_2, \ldots\ldots, A_n$

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i) - \sum_{1 \le i} \sum_{j \le n} P(A_i \cap A_j) + \sum_{1 \le i <} \sum_{j} \sum_{k \le n} P(A_i \cap A_j \cap A_k) - \ldots\ldots + (-1)^{n-1} P(A_1 \cap A_2 \cap \ldots\ldots \cap A_n)$$

(c) X and Y are two random variables having the joint density function $f(x, y) = \frac{1}{27}(x + 2y)$ where x and y can assume only the integer values 0, 1 and 2. Find conditional distribution of Y for X = x.

2. (a) Let (X, Y) be a two-dimensional non-negative continuous random variable having the joint density :

$$f(x, y) = \begin{cases} 4xy\ e^{-(x^2+y^2)} & ; x \ge 0, y \ge 0 \\ 0; \text{elsewhere} \end{cases}$$

Prove that the density function of $U = \sqrt{X^2 + Y^2}$ is :

$$h(u) = \begin{cases} 2u^3 e^{-u^2}; & 0 \le u < \infty \\ 0 \quad ; & \text{elsewhere} \end{cases}$$

(b) Define Poisson distribution and find its mean and variance. Show that Poisson distribution is a limiting case of binomial distribution.

3. (a) Establish relationship between root mean square deviation and standard deviation. If n_1, n_2 are the sizes, $\bar{x}_1, \bar{x}_2$ the means, and σ_1, σ_2 the standard deviations of two series, then show that the standard deviation σ of the combined series is given by :

$$\sigma = \frac{1}{n_1 + n_2}\left[n_1(\sigma_1^2 + d_1^2) + n_2(\sigma_2^2 + d_2^2)\right]$$

where $d_1 = \bar{x}_1 - \bar{x}$, $d_2 = \bar{x}_2 - \bar{x}$ and

$$\bar{x} = \frac{n_1\bar{x}_1 + n_2\bar{x}_2}{n_1 + n_2}$$, is the mean of the combined series.

(b) Explain skewness and give its measures. In a frequency distribution, the coefficient of skewness based upon the quartiles is 0.6. If the sum of the upper and lower quartiles is 100 and median is 38, find the values of upper and lower quartiles.

4. (a) Define Karl Pearson's coefficient of correlation and find its limits. Let the variables X and Y are connected by the equation $aX + bY + c = 0$, then show that the correlation between them is -1 if the sign of a and b are alike and $+1$ if they are different.

(b) Define normal distribution and give its general properties. If X is a normal variate with mean 30 and S.D. 5, then find the probabilities that :

(i) $26 \leq X \leq 40$

(ii) $X \geq 45$

(iii) $|X - 30| > 5$

[Given $P(0 \leq z \leq 0.8) = 0.2881$, $P(|z| \leq 2) = 0.9544$, $P(|z| \leq 1) = 0.6826$ and $P(|z| \leq 3) = 0.9973$, where z is standard normal variate]

5. (a) Explain the important properties of regression coefficients. Given two lines of regression :

$$8x - 10y + 66 = 0 \quad \text{and} \quad 40x - 18y = 214.$$

Find :

(i) the mean values of x and y

(ii) correlation coefficient between x and y

(iii) standard deviation of y when $\sigma_x^2 = 9$.

(b) Explain the term 'correlation ratio'. How is it measured ? If η_{yx} is the correlation ratio between the variables X and Y and r is the correlation co-efficient, then prove that :

$$\eta_{yx}^2 \geq r^2 .$$

6. (a) Define chi-square (χ^2) distribution and obtain its moment generating function. If χ_1^2 and χ_2^2 are independent χ^2 variates with n_1 and n_2 d.f. respectively, then show that :

$$U = \frac{\chi_1^2}{\chi_1^2 + \chi_2^2} \quad \text{and} \quad V = \chi_1^2 + \chi_2^2$$

are independently distributed, U as a $\beta_1\left(\frac{n_1}{2}, \frac{n_2}{2}\right)$ variate and V as a χ^2 variate with $(n_1 + n_2)$ d.f.

(b) Define Snedecore's F-distribution and establish its relationship with t and χ^2 distributions.

7. (a) Explain consistency and unbiasedness as properties of good estimators. A random sample $(x_1, x_2, x_3, x_4, x_5)$ of size 5 is drawn from a

normal population with unknown mean μ. Consider the following estimators to estimate μ :

$$t_1 = \frac{x_1 + x_2 + x_3 + x_4 + x_5}{5},\ t_2 = \frac{x_1 + x_2}{2} + x_3,$$

$$t_3 = \frac{2x_1 + x_2 + \lambda x_3}{3}$$

where λ is such that t_3 is unbiased estimator. Find λ. Are t_1 and t_2 unbiased ? State giving reasons, the estimator which is best among t_1, t_2 and t_3.

(b) Explain Nayman's criterion for a distribution to admit sufficient statistic. Let $x_1, x_2, \ldots\ldots\ldots, x_n$ be a random sample of size n from $N(\mu, \sigma^2)$ population. Find sufficient estimators for μ and σ^2.

8. (a) Define maximum likelihood estimator and explain its properties. Find the MLE for the parameter λ of a Poisson distribution on the basis of sample of size n. Also find its variance.

(*b*) Explain the terms sampling distribution and standard error with suitable examples. A dice is thrown 9000 times and a throw of 3 or 4 is observed 3240 times. Show that the dice can not be regarded as an unbiased one and find the limits between which the probability of a throw of 3 or 4 lies.