STATISTICS — PAPER - I

Time – Three hours *Full Marks – 100*

Answer to ***Question No. 1*** *i.e.* ***MCQ*** *type questions under Section-A must be written in English only. Answers to other questions must be written either in English or in Bengali. It must not be answered partly in English and partly in Bengali.*

This instruction should be followed scrupulously.

The figures in the margin indicate full marks for the questions.

Candidates are required to give their answers in their own words as far as practicable.

Scientific calculator is allowed in this paper.

SECTION–A

Answer *all* questions.

1. Select the best correct answer and write it in the answer-script putting option (i) or (ii) or (iii) or (iv) as applicable. 20×1=20

Example :

Question : For two random variables X and Y the relation E(XY) = E(X) E(Y) holds good

(i) if X and Y are independent

(ii) for all X and Y

(iii) if X and Y are identical

(iv) none of these

Answer : (i) if X and Y are independent.

(a) Which of the following is not a characteristic of the mean ?

(i) It is affected by extreme values.

(ii) It minimizes the sum of squared deviations.

(iii) The sum of the deviations about the mean is 0.

(iv) It is best used with ordinal data.

(b) Ogive curve occur for :

(i) more than type distribution (ii) less than type distribution

(iii) both (i) and (ii) (iv) none of (i) and (ii)

(c) The point of intersection of two cumulative frequency curves provides :

(i) mean (ii) mode

(iii) median (iv) Q_1

(d) If the two lines of regression are perpendicular to each other, the relation between the two regression coefficients is :

(i) $b_{yx} = b_{xy}$ (ii) $b_{yx}\, b_{xy} = 1$

(iii) $b_{yx} \leq b_{xy}$ (iv) $b_{yx} = -\, b_{xy}$

(e) If the correlation coefficient between the variables X and Y is 0·5, then the correlation coefficient between the variables $2X - 4$ and $3 - 2Y$ is

(i) 1 (ii) 0·5

(iii) – 0·5 (iv) 0

(f) If $A \subset B$, the probability, $P(A|B)$ is equal to

(i) zero (ii) one

(iii) $P(A)/P(B)$ (iv) $P(B)/P(A)$

(g) The probability of throwing an odd sum with two fair dice is :

(i) 1/4 (ii) 1/16

(iii) 1 (iv) 1/2

(h) If X is a random variable such that $X \geq 0$, then

(i) $E(X^2) = [E(X)]^2$ (ii) $E(X^2) \geq [E(X)]^2$

(iii) $E(X^2) \leq [E(X)]^2$ (iv) none of these

(i) If X is a random variable, E (e^{itx}) is known as :

(i) characteristic function (ii) moment generating function

(iii) probability generating function (iv) all of these

(j) The daily rainfall of a region is a random variable of the type :

(i) continuous

(ii) discrete

(iii) neither discrete nor continuous

(iv) continuous as well as discrete

(k) If the m.g.f of a distribution is $(q + pe^t)^n$, the variance of the distribution is :

(i) 2n (ii) pq

(iii) npq (iv) pq/n

(l) Normal distribution was invented by :

(i) Laplace (ii) De-moiver

(iii) Gauss (iv) none of them

(m) If a variable $X \sim \gamma(\alpha, l)$, the p.d.f of X is same as that of

(i) chi-square distribution (ii) exponential distribution

(iii) normal distribution (iv) Beta distribution

(n) The skewness of binomial distribution will be zero if

(i) $p < \frac{1}{2}$ (ii) $p > \frac{1}{2}$

(iii) $p = \frac{1}{2}$ (iv) $p < q$

(o) A normal random variable has mean = 2 and variance = 4. Its fourth central moment μ_4 will be :

(i) 16 (ii) 64

(iii) 80 (iv) 48

(p) If T is consistent estimator of θ, then e^T is a

(i) unbiased estimator of e^θ (ii) consistent estimator of e^θ

(iii) MVUE of e^θ (iv) none of these

(q) Least square estimators of the parameters of a linear model are :

(i) unbiased (ii) BLUE

(iii) UMVU (iv) all of these

(r) If the sample mean $\bar{x}$ is an estimate of population mean μ, then $\bar{x}$ is

(i) unbiased and efficient (ii) unbiased and inefficient

(iii) biased and efficient (iv) biased and inefficient

(s) Size of critical region is known as :

(i) power of the test

(ii) size of type II error

(iii) critical value of the test statistics

(iv) size of the test

(t) The hypothesis under test is :

(i) simple hypothesis (ii) alternative hypothesis

(iii) null hypothesis (iv) none of these.

SECTION – B

Attempt any *six* questions. 6×5=30

2. What do you understand by median ? What are the points in favour of median ? What are the demerits of median ?

3. If X and Y are two variables such that $Y=7^{\frac{1}{x}}$ and HM of X is 7, find GM of Y.

4. For any distribution, prove that $\beta_2 \geq \beta_1 + 1$.

5. State and prove the Cauchy-Schwartz inequality.

6. A rod of length 'a' is broken into three parts at random. What is the probability that a triangle can be formed from these parts ?

7. A continuous random variable X has the c.d.f

$$F(x) = \begin{cases} 0 & ; \ x \leq 1 \\ k(x-1)^4 & ; \ 1 < x \leq 3 \\ 1 & ; \ x > 3 \end{cases}$$

Find (i) K, (ii) mean and median of X.

8. If X is a Poisson variate such that $P(X=2)=9\,P(X=4) + 90\,P(X = 6)$. Find (i) λ, (ii) the coefficient of skewness.

9. If $X_1, X_2, \ldots\ldots, X_n$ is a random sample of size 'n' from $N(\mu, \sigma^2)$, where μ is known and if $T = \frac{1}{n}\sum |X_i - \mu|$, examine if T is unbiased for σ. If not, obtain an unbiased estimator of σ.

SECTION – C

Attempt any *five* questions. $5 \times 10 = 50$

10. Show that in a discrete series if deviations are small compared with mean 'M' so that $(x/M)^2$ and higher powers of (x/M) are neglected, prove that $MH = G^2$, where 'G' is geometric mean and 'H' is harmonic mean.

11. Show that mean deviation about median is least.

12. If X is a standard normal variate and $Y = a + bX + cX^2$ where a, b, c are constants, find the correlation coefficient between X and Y. Hence or otherwise obtain the conditions where (i) X and Y are uncorrelated and (ii) X and Y are perfectly correlated.

13. If $X_1, X_2, \ldots\ldots, X_n$ are random variables each with the same expected value 'μ' a SD 'σ'. The correlation coefficient between any two X's is 'ρ'. Show that

(i) $\text{Var}(\bar{X}) = \frac{\sigma^2}{n} + \left(1 - \frac{1}{n}\right)\rho\sigma^2$

(ii) $E\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2 = (n-1)(1-\rho)\sigma^2$

14. If $X \sim N(0, 1)$ then obtain the distributions of (i) X^2 and (ii) $|X|$.

15. State and prove Cramer-Rao inequality.

16. Explain the following terms :

(i) Simple and composite statistical hypothesis.

(ii) Errors of first and second kinds.

(iii) Power of a test.

(iv) Null and alternative hypothesis.

STATISTICS — PAPER - II

Time – Three hours *Full Marks – 100*

Answer to ***Question No. 1*** *i.e.* ***MCQ*** *type questions under Section-A must be written in English only. Answers to other questions must be written either in English or in Bengali. It must not be answered partly in English and partly in Bengali.*

This instruction should be followed scrupulously.

The figures in the margin indicate full marks for the questions.

Candidates are required to give their answers in their own words as far as practicable.

Scientific calculator is allowed in this paper.

SECTION–A

Answer *all* questions.

1. Select the best correct answer and write it in the answer-script putting option (i) or (ii) or (iii) or (iv) as applicable. 20×1=20

Example :

Question : Chance or random variation in the manufactured product is

(i) controllable

(ii) not controllable

(iii) partially controllable

(iv) none of these

Answer : (ii) not controllable.

(a) What sampling design is most appropriate for cluster sampling ?

(i) SRSWOR (ii) SRSWR

(iii) Stratified random sampling (iv) Quota sampling

(b) Double sampling is known as

(i) two stage sampling (ii) two phase sampling

(iii) two directional sampling (iv) all of these

(c) Under proportional allocation the size of the sample from each stratum depends on

(i) total sample size (ii) size of the stratum

(iii) population size (iv) all of these

(d) The number of possible samples of size 'n' from a population of size 'N' units without replacement is

(i) $\binom{N}{n}$ (ii) ${}^{N}P_{n}$

(iii) n^2 (iv) n!

(e) Sampling frame is a term used for

(i) a list of random numbers (ii) a list of voters

(iii) a list of sampling units (iv) none of these

(f) Replication in an experiment means :

(i) the number of blocks

(ii) total number of treatments

(iii) the number of times a treatment occurs in an experiment

(iv) none of these

(g) Local control is a device to maintain

(i) homogeneity among blocks (ii) homogeneity within blocks

(iii) both (i) and (ii) (iv) neither (i) and (ii)

(h) Local control in the field is maintained through

(i) uniformity trials (ii) randomization

(iii) natural factors (iv) all of these

(i) Which of the following is not a contrast ?

(i) $T_1 + 2T_2 - T_3$ (ii) $T_1 - T_3$

(iii) $T_1 - 2T_2 + T_3$ (iv) $-T_1 + 2T_2 - T_3$

(j) A randomize block design has

(i) two way classification (ii) one way classification

(iii) three way classification (iv) no classification

(k) The component of a time series which is attached to short-term fluctuations is

(i) seasonal variation (ii) cyclic variation

(iii) irregular variation (iv) all of these

(l) A group for moving average consists of

(i) 5-years period

(ii) 3-years period

(iii) a period which forms a cycle

(iv) none of these

(m) The weights used in Paasche's formula belong to

(i) the base period

(ii) the given period

(iii) to any arbitrary chosen period

(iv) none of these

(n) If the index number is independent of the units of measurements, then it satisfies

(i) time reversal test (ii) factor reversal test

(iii) unit test (iv) all of these

(o) Fisher's ideal formula does not satisfy

(i) time reversal test (ii) circular test

(iii) factor reversal test (iv) unit test

(p) The value of N.R.R > 1 will result into

(i) increase in population

(ii) negative increase in population

(iii) zero increase in population

(iv) none of these

(q) A population have constant size and composition is called a

(i) stable population (ii) stationary population

(iii) continuous population (iv) discrete population

(r) The control limits delimited by the consumer are called

(i) modified control limits (ii) natural control limits

(iii) specified control limits (iv) none of these

(s) Main tools of statistical quality control are

(i) Shewhart charts (ii) acceptance sampling plans

(iii) both (i) and (ii) (iv) none of these

(t) The relation between expected value of 'R' and SD 'σ' with usual constant factors is

(i) $E(R) = d_1 \sigma$ (ii) $E(R) = d_2 \sigma$

(iii) $E(R) = D_1 \sigma$ (iv) $E(R) = D_2 \sigma$

SECTION–B

Attempt any *six* questions. 6×5=30

2. What is sample ? What is meant by sampling method ? Define sampling unit and give its two examples.

3. Prove that in simple random sampling without replacement, sample mean square is an unbiased estimate of population mean square.

4. What do you understand by local control and in what way does it increase the efficiency of an experimental designs ?

5. What is meant by a randomised block design ? Obtain the efficiency of this design compared to completely randomised design.

6. What events are covered under vital statistics ? What are various uses of vital statistics for a country ?

7. What is time series ? Discuss in brief the moving average method for ascertaining the trend.

8. What is meant by quality of a product ? What do you understand by control charts in statistical quality control ?

9. Write short note on T-scores.

SECTION – C

Attempt any *five* questions. 5×10=50

10. What is stratified random sampling ? Describe the advantages of stratified random sampling. Obtain the efficiency of the optimum allocation compared to proportional allocation.

11. What is meant by randomised block design? Give the analysis of variance for the design.

12. Explain the statistical basis and construction of 'p' and 'np' charts. How is the choice between 'p' and 'np' charts made ?

13. Distinguish between defect and defective. Give some examples of defects for which the C-chart is applicable. How do you calculate control limits for a C-chart ?

14. Explain how the 'principle of least squares' used to estimate trend equation $y = ab^x$ in a time series.

15. Define index numbers. What are the uses of index numbers ? Describe the limitations of index numbers.

16. Write short notes on :
    (i) Age specific fertility rate
    (ii) Net reproduction rate.