STATISTICS SYLLABUS

UNIT – I (Probability Theory)

Definition–Classical and axiomatic approaches.Laws of total and compound probability, conditional probability, Bayes Theorem. Random variable and its distribution function, mathematical expectation, generating functions (moment generating, characteristic and probability generating functions). Different modes of convergence laws of large numbers, central limit theorem. Joint distribution of two random variables, marginal and conditional distributions.

UNIT – II (Probability Distributions)

Discrete probability distributions–Geometric,Binomial, Poisson, Negative Binomial, Hypergeometric and Multinomial distributions.

Continuous probability distributions–Exponential, Normal, Uniform, Beta, Gamma, Cauchy and Bivariate Normal distributions.

UNIT – III (Statistical Methods)

Frequency distribution, graphical and diagrammatic representation of data.Measures of location and dispersion, moments, skewness and kurtosis. Curve fitting, association of attributes. Simple correlation and regression, partial and multiple correlations and regressions, correlation ratio. Distribution of sample mean and sample variance,t, F and Chi-square distributions.

UNIT – IV (Estimation and Testing of Hypothesis)

Characteristics of a good estimator.Estimation by the methods of maximum likelihood and least squares, properties of maximum likelihood estimator, Cramer-Rao inequality, sufficient statistic and Rao-Blackwell theorem.Interval estimation.

Testing of simple and composite hypotheses, types of errors, critical region.Neyman-Pearson fundamental lemma, power function, MP and UMP tests. Likelihood Ratio tests, Non-parametric tests-Sign, median, run and Mann-Whitney U- tests, large sample tests, tests based on t, F and Chi-square distributions.

UNIT – V (Sampling Techniques and Designs of Experiments)

Census versus sample surveys. Simple random sampling, stratified sampling, systematic sampling, sampling with probability proportional to size. Ratio and regression methods of estimation.

Principles of designs of experiment. Lay out and analysis of completely randomized, randomized block and Latin square designs. Factorial experiments ($2^2, 2^3$ and 3^2 experiments), Confounding (total and partial).

STATISTICS

(Sample Questions)

UNIT-I

1. If A, B and C are three events of the sample space S; pick up the incorrect statement from below:
   (a) $A^C B^C C^C$ means non-occurrence of all the three events A, B and C
   (b) $(A - B) \cup (B - A) = A \Delta B$
   (c) $(A \cup B)C = AC \cup BC$
   (d) $(A - B)$ is always a proper subset of A
2. The following statements relate to probability in sample space S
   (i) Probability is defined on some sets of S
   (ii) Probability is countably additive for disjoint subsets of S
   (iii) $P(A) = 0 \Longrightarrow A = \emptyset$

   Mark your answer with code:

   (a) If (i), (ii) only are correct
   (b) If (i), (iii) only are correct
   (c) If (ii), (iii) only are correct
   (d) If all the above statements are correct
3. The properties below relate to a distribution function $F(x)$ of a r.v. Pick up the incorrect statement from below:
   (a) $\lim_{x \to \infty} F(x) = 1$
   (b) $F(x)$ has at most a countable number of discontinuities
   (c) If $y > x \Longrightarrow F(y) > F(x)$
   (d) $F(x)$ is everywhere continuous to the right
4. The probability density function of a r.v. X is given by
   $$P(X = k) = e^{-t}(1 - e^{-t})^{k-1}, k = 1,2,3, \ldots\ldots$$
   $E(X)$ is given by
   (a) e^{-t} (b) e^{-2t} (c) e^{t} (d) e^{2t}
5. Which of the following statements is not a property of a characteristic function $\emptyset_X(t)$
   (a) $\emptyset_X(t)$ is a uniformly continuous function in t in the interval $(-\infty, +\infty)$
   (b) $\emptyset_X(t)$ and $\emptyset_X(-t)$ are (complex) conjugate functions
   (c) Every distribution function has a unique characteristic function
   (d) $|\emptyset_X(t)|^2 \leq 2^{-1}$
6. Which of the following statements is not true ?
   (a) Convergence in probability implies convergence in law
   (b) Convergence in law implies convergence in probability

(c) If $X_1, X_2, \ldots \ldots$ are i.i.d. random variables such that $E(X_i) = \mu$ and $V(X_i) < k$, then WLLN holds.

(d) If $X_1, X_2, \ldots \ldots$ are i.i.d. random variables such that $E(X_i) = m$, then WLLN holds.

ANSWERS : 1(d), 2(a), 3(c), 4(c), 5(d), 6(b)

UNIT-II

1. The statements below relate to binomial and Poisson distributions:

   (i) A binomial distribution is symmetrical implies $p = \frac{1}{2}$

   (ii) For a Poisson distribution, mean can be greater than variance

   (iii) If the mean and SD of a Poisson distribution are equal, then its variance is unity.

   Mark your answer with code:

   (a) If (i) only is incorrect
   (b) If (ii) only is incorrect
   (c) If (iii) only is incorrect
   (d) If all the statements are correct

2. If $B(n, p)$ stands for the binomial distribution with parameters n and p and $P(\lambda)$ for the Poisson distribution with parameter λ, which one of the following pairs is not correctly matched ?

	List – I	List – II
(a)	$B(10, 0.36)$	Mean ≈ 1.9
(b)	$B(15, 0.05)$	Mode at zero
(c)	$P(5.3)$	Mode at 5
(d)	$P(7)$	Mode at 7

3. For a negative binomial distribution with parameter (k, p), if variance is double the mean, then

   (a) $k = 4, p = \frac{1}{2}$ (b) $p = \frac{1}{2}$ (c) $k = 2$ (d) Such a distribution does not exist

4. The mean of a normal variable X is 30 and its variance is 4. If $\emptyset(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{-z^2/2}\, dz$, $t > 0$, express $P(28 \leq X \leq 32.4)$ in terms of $\emptyset(t)$ and pick up the correct answer from below:

   (a) $\emptyset(1.2) - \emptyset(1)$
   (b) $\emptyset(1.2) + \emptyset(1) - 1$
   (c) $\emptyset(0.6) + \emptyset(0.5) - 1$
   (d) This probability cannot be computed from the given data

5. The r.v. X has the following p.d.f.:

$$f(x,\lambda) = \frac{e^{-x}x^{\lambda-1}}{\Gamma(\lambda)}\ ;\ \lambda > 0, 0 < x < \infty$$

Pick up the incorrect statement from below:

(a) The mean and the variance of X are equal

(b) If X and Y are independent random variables having the same p.d.f. $f(x,\lambda)$, then $(X+Y)$ has the p.d.f. $f(x,2\lambda)$

(c) The distribution is bell-shaped for all $\lambda > 0$

(d) $Z = (x-\lambda)/\sqrt{\lambda}$ tends to the standard normal distribution as $\lambda \to \infty$

6. If $f(x,y;\ \mu_x,\mu_y,\sigma_x,\sigma_y,\rho)$ is the bivariate normal distribution in its most general form, which of the following statements is not a property of the distribution:

(a) The marginal distribution of X is normal $N(\mu_x,\sigma_x)$

(b) The conditional distribution of Y for given $X = x$ is normal $N(\mu_y + \frac{\rho\sigma_y}{\sigma_x}(x-\mu_x), \sigma_y\sqrt{1-\rho^2})$

(c) X and Y are independent if and only if $\rho = 0$

(d) $(X+Y)$ and $(X-Y)$ are independent random variables.

ANSWERS: 1(b), 2(d), 3(b), 4(b), 5(c), 6(d)

UNIT-III

1. The number of sides of the frequency polygon of a frequency distribution with n classes is-

(a) $(n-1)$ (b) n (c) $n+1$ (d) $n+2$

2. The mean and variance of 20 numbers are 6 and 3 respectively. The values of (i) the sum of the numbers, and (ii) the sum of squares of the numbers are respectively –

(a) (120, 780) (b) (120, 720) (c) (60, 720) (d) (60,780)

3. For a positively skewed distribution, the values of mean, median and mode, in some order, are 46, 48 and 42. Pick up the correct alternative from below:

(a) Mean = 48, median = 46, mode = 42

(b) Mean = 46, median = 48, mode = 42

(c) Mean = 42, median = 46, mode = 48

(d) Mean = 42, median = 48, mode = 46

4. If $U + 3X = 5, 2Y - V = 7$ and correlation coefficient between X and Y is $\frac{1}{2}$, the correlation coefficient between U and V is

(a) $\frac{1}{2}$ (b) -1 (c) $-\frac{1}{2}$ (d) 1

5. If the multiple correlation coefficient of X on Y and Z is zero, then the partial and simple correlation coefficients involving X are respectively

(a) 0 and 1 (b) 1 and 0 (c) 0 and 0 (d) $\frac{1}{2}$ and $\frac{1}{2}$

6. Which one of the following is not a true statement about the correlation ratio ?

(a) It measures curvilinear relationship between two variables

(b) It coincides with correlation coefficient when regression lines between the two variables are linear
(c) Its absolute value is less than that of the correlation coefficient
(d) It is independent of the change of origin and scale of the variables

ANSWERS: 1(d), 2(a), 3(a), 4(c), 5(c), 6(c)

UNIT-IV

1. For the sample mean to be an unbiased estimator of population mean, the condition of normality of population is:
   (a) Necessary but not sufficient
   (b) Necessary as well as sufficient
   (c) Sufficient but not necessary
   (d) None of these
2. Which of the following statements about a maximum likelihood estimator is not true ?
   (a) It is consistent
   (b) It is necessarily unbiased
   (c) It is sufficient if a sufficient statistic exits
   (d) It is asymptotically most efficient
3. Rao-Blackwell theorem enables us to obtain unique minimum variance unbiased estimator through
   (a) Unbiased estimators
   (b) Complete sufficient statistics
   (c) Efficient statistics
   (d) Sufficient statistics
4. A test which is at least as powerful as any other test of the same size is called
   (a) Best test (b) Optimum test (c) MP test (d) UMP test
5. If there are 10 symbols of two types, equal in number, the maximum possible number of runs in a sequence is:
   (a) 8 (b) 5 (c) 10 (d) 9
6. The hypothesis that the population variance has a specified value can be tested by
   (a) F −test (b) Chi-square test (c) Fisher's t −test (d) Z −test

ANSWERS: 1(d), 2(b), 3(b), 4(d), 5(c), 6(b)

UNIT-V

1. If a sample of 2 units is drawn from a population of 6 units by simple random sampling with replacement, then the probability of drawing a sample is

(a) $\frac{1}{15}$ (b) $\frac{1}{36}$ (c) $\frac{1}{12}$ (d) $\frac{1}{3}$

2. One of the drawbacks of the systematic sampling is that
   (a) It is operationally less convenient than simple random sampling
   (b) It is always less efficient than stratified random sampling
   (c) It fails to provide an unbiased estimate of its variance
   (d) None of these
3. Regression method of estimation is used when correlation between study variable and auxiliary variable is
   (a) Negative
   (b) Positive
   (c) Both positive and negative
   (d) None of these
4. In experimental designs, randomization is necessary to make the estimate
   (a) Valid
   (b) Accurate
   (c) Precise
   (d) Unbiased
5. In an RBD with 4 treatments and 6 replicates, error degrees of freedom is
   (a) 24 (b) 15 (c) 23 (d) 12
6. A Latin square design uses:
   (a) One-way classification
   (b) Two-way classification
   (c) Incomplete three-way classification
   (d) Does not classify the data

ANSWERS: 1(b), 2(c), 3(c), 4(a), 5(b), 6(c)