SLR-SW-1

Seat No.	

Set P

M.Sc. (Sem-I) (New) (CBCS) Examination: March/April-2023
BIOSTATISTICS
Probability Distributions (MSC22101)

Day & Date: Monday, 19-07-2023 Max. Marks: 80
Time: 03:00 PM To 06:00 PM

Instructions: 1) Question no. 1 and 2 are compulsory.
2) Attempt any three questions from Q. No. 3 to Q. No. 7.
3) Figure to right indicate full marks.

Q.1 A) Choose the correct alternative: **10**

1) In a hospital, sex of newly born baby is recorded as male or female is a real-life situation where _____ distribution is used.
   a) Binomial b) Discrete Uniform
   c) Bernoulli d) None of these

2) A distribution in which the probability at each successive draw varies is _____
   a) Hypergeometric b) Geometric
   c) Binomial d) Discrete uniform

3) If X and Y are independent random variables then _____
   a) $Cov(X,Y)=0$ *b)* $Corr(X,Y)=0$
   c) $E(XY)=E(X)E(Y)$ *d)* All the above

4) We get standard normal distribution from normal distribution if _____
   a) $\mu=1,\sigma=0$ b) $\mu=0,\sigma=1$
   c) $\mu=0,\sigma=0$ d) $\mu=1,\sigma=1$

5) Let $\underline{X}=(X_1,X_2\ldots,X_k)$ be a multinomial random variable with parameters $n,p_1,p_2\ldots,P_k,\ \sum_{i=1}^{k} pi=1$ Then marginal distribution of X_1 is _____
   a) $B(n,1-p_1)$ b) $B(n,P1)$
   c) $B(n-1,p_1)$ d) not binomial

6) Which one of the following distributions has memory less property?
   a) Normal b) Binomial
   c) Exponential d) Uniform

7) If random variable X has standard exponential distribution then _____
   a) $E(X)=2\,Var(X)$ b) $E(X)=Var(X)$
   c) $Var(X)=2E(X)$ d) none of these

8) If $X>0$ then _____
   a) $E[\log X]\le\log[E(X)]$ b) $E[\log X]\ge\log[E(X)]$
   c) $E[\log X]=\log[E(X)]$ d) None of these

9) If $\mu_1'=2,\mu_2'=8$ and $\mu_3=3$ then value of μ_3' _____
   a) 15 25
   c) 35 45

10) Let X has $B(n,p)$ distribution then PGF of X is _____
   a) $(p+q\,S)^n$ b) $(q+p\,S)^n$
   c) $(q-p\,S)^n$ d) $(p-q\,S)^n$

B) Fill in the blanks. **06**

1) If a random variable X has mean 3 and standard deviation 4 then the variance of variable $Y = 2X + 5$ is _____.

2) If Z is standard normal variate then variance of Z^2 is _____.

3) The pdf of random variable X is $f(x) = 2x, 0 \le x \le 1$ then $P(X = 0.5)$ is _____

4) Let X and Y be two iid random variables with $pdf\ f(x) = 2e^{-2x}, x \ge 0$. The distribution of $Z = X - Y$ is _____

5) Let X has a continuous distribution with $cdf\ F\ (X)$, then the distribution of $Y = F(x)$ is _____

6) If X has uniform $U(0,1)$ distribution, then the distribution of the r^{th} order statistic is _____.

Q.2 Answer the following **16**

a) What do you mean by (i) a Bernoulli trial (ii) binomial experiment? Give an illustration of each.

b) Define symmetric random variable. Give one example of symmetric random variable.

c) Define power series distribution. Show that Geometric distribution is power series distribution.

d) Define Poisson distribution giving an example and find its mean.

Q.3 Answer the following. **16**

a) Define cumulative distribution function (cdf) of a random variable and state and prove its important properties.

b) Define location family of distributions. Examine which of the following are in location family.
   1) $X \sim N(\theta, 1)$
   2) $X \sim Exp(\theta, 1)$

Q.4 Answer the following. **16**

a) Define moment generating function (MGF) of a random variable X. Explain how it is used to obtain moments of a random variable X.

b) Define multinomial distribution. Obtain its MGF. Hence or otherwise obtain its variance-covariance matrix.

Q.5 Answer the following. **16**

a) State and prove Markov's inequality.

b) Let X be $U(0,1)$ distribution. Find the distribution of
   1) $Y = 1 - X$
   2) $Y = -2\log X$

Q.6 Answer the following. **16**

Let X is a non-negative random variable with $pmf\ P(X = x) = P_x, x = 1,2, \ldots$

a) then show that $E(X) = \sum_{x=1}^{\infty} P[X \geq x]$

b) The joint probability distribution of (X, Y) is given by.

$$P(x, y) = \begin{cases} (x + y)/21, & x = 1,2,3, y = 1,2 \\ 0, & otherwise \end{cases}$$

1) Find marginal probability mass functions of X and Y.
2) Conditional distribution of X given $Y = 2$

Q.7 Answer the following. **16**

a) Define order statistics. Derive the distributions of smallest and largest order statistics based on random sample of size n from a continuous distribution.

b) Let (X, Y) has $BVN(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$. Obtain the conditional distribution of X given $Y = y$.

SLR-SW-7

Seat No.	

Set P

M.Sc. (Semester - II) (New) (CBCS) Examination: March/April-2023
BIOSTATISTICS
Statistical Inference - I (MSC22202)

Day & Date: Sunday, 23-07-2023 Max. Marks: 80
Time: 11:00 AM To 02:00 PM

Instructions: 1) Question no. 1 and 2 are compulsory.
2) Attempt any three questions from Q. No. 3 to Q. No. 7.
3) Figure to right indicate full marks.

Q.1 A) Choose the correct alternative: **10**

1) A sufficient statistic contains all the information which is contained in _____.
   a) population b) sample
   c) parameter d) none of the above

2) Cramer-Rao inequality with regards to the variance of an unbiased estimator provides _____.
   a) lower bound b) upper bound
   c) asymptotic variance d) Fisher information

3) A statistic $T(X)$ for θ is said to be ancillary if _____
   a) $T(X)$ is independent of θ
   b) $T(X)$ is dependent on θ
   c) The distribution of $T(X)$ is independent of θ
   d) The distribution of $T(X)$ is depends on θ

4) Which of the following is not a member of one-parameter exponential family of distributions?
   a) Bernoulli $(1, \theta)$ b) Cauchy $(1, \theta)$
   c) Normal $(\theta, 1)$ d) Poisson (θ)

5) Which of the following statements is correct _____.
   a) MLE always exists b) MLE is always unique
   c) MLE is always unbiased d) None of the above is true.

6) If T_n is sufficient statistic for θ based on random sample of size n, then $\frac{\partial \log L}{\partial \theta}$ is a function of _____.
   a) θ only b) T_n only
   c) both T_n and θ d) none of the above

7) Let T_n be an unbiased estimator of θ .Then _____.
   a) T_n^2 is unbiased estimator of θ^2
   b) $\sqrt{T_n}$ is unbiased estimator of $\sqrt{\theta}$
   c) e^{T_n} is unbiased estimator of e^{θ}
   d) $3T_n + 4$ is unbiased estimator of $3\theta + 4$

8) Let $X_1, X_2, \ldots, X_n$ is a random sample from $U(0, \theta), \theta > 0$ The MLE of θ is _____.
   a) $\bar{X}$ b) $X_{(1)}$
   c) $X_{(n)}$ d) sample median

9) If T_n is consistent estimator of θ then $\phi(T_n)$ is consistent estimator of $\phi(\theta)$ if _____.
   a) ϕ is linear function
   b) ϕ is continuous function
   c) ϕ is differentiable function
   d) none of these

10) Which of the following is true?
   a) Consistent estimator is always unbiased
   b) Every consistent estimator is CAN
   c) Consistent estimator need not be unique
   d) Consistent estimator is always MLE

B) Fill in the blanks. **06**

1) Based on random sample of size n from $N(\mu, 1), \mu \in R$ population. MLE of μ is _____.

2) For power series family of distribution _____ is sufficient statistic for θ

3) Let $X_1, X_2 \ldots X_n$ be iid Poisson (θ), UMVUE for θ is _____.

4) If an estimator T_n of population parameter θ converges in probability to θ as n tends to infinity is said to be _____.

5) For Cauchy distribution with location θ, the consistent estimator of θ is _____.

6) Minimal sufficient statistic is a function of _____.

Q.2 Answer the following

a) Define: **04**
   i) Sufficient statistic
   ii) Minimal sufficient statistic

b) Let random variable X has *Poisson* (θ) distribution. Show that distribution of X is complete. **04**

c) Define a maximum likelihood estimator for a parameter θ, and state the large sample properties of this type of estimator under regularity conditions, to be stated clearly. **04**

d) Let $X_1, X_2, \ldots, X_n$ be iid $U(0, \theta)$, computing the actual probability show that $X_{(n)}$ is consistent estimator of θ. **04**

Q.3 Answer the following.

a) Define Power series family of distributions. Show that $B(n, \theta)$ distribution belong to power series family. **08**

b) Let $X_1, X_2 \ldots, X_n$ be a random sample from $U(0, \theta), \theta > 0$ distribution. Show that $X_{(n)}$is sufficient statistic for θ, but $X_{(1)}$ is not sufficient statistic. **08**

Q.4 Answer the following.

a) Define consistent estimator. State and prove invariance property of consistent estimator of a real valued parameter θ. **08**

b) Let $X_1, X_2, \ldots, X_n$ be a random sample from $N(\theta, \sigma^2)$. Find two consistent estimators of σ^2. **08**

Q.5 Answer the following.

a) State and prove Rao-Blackwell theorem. **08**

b) Use Rao-Blackwell theorem to derive UMVUE of $P(X_1 = 0)$ based on sample $X_1, X_2, \ldots, X_n$ from *Poisson* $(\lambda), \lambda > 0$ distribution. **08**

Q.6 Answer the following.

a) Define Fisher information matrix. Obtain Fisher information matrix in case of normal distribution with parameters μ and σ^2 . **08**

b) State and prove Cramer-Rao inequality with necessary regularity conditions. **08**

Q.7 Answer the following.

a) Describe method of moments and method of minimum chi-square. **08**

b) Let $X_1, X_2, \ldots, X_n$ be iid $U(0, \theta), \theta > 0$ Find **08**

i) Moment estimator θ

ii) MLE of θ

SLR-SW-10

Seat No.	

Set P

M.Sc. (Semester - III) (New) (CBCS) Examination: March/April-2023
BIOSTATISTICS
Statistical Inference – II (MSC22301)

Day & Date: Monday, 10-07-2023 Max. Marks: 80
Time: 11:00 AM To 02:00 PM

Instructions: 1) Q. Nos 1 and 2 are compulsory.
2) Attempt any Three questions from Q. No.3 to Q. No.7
3) Figures to the right indicates full marks.

Q.1 A) Choose the correct alternatives. **10**

1) If we reject the null hypothesis when it is true, we might be making _____.
   a) type I error b) type II error
   c) a correct decision d) unpredictable

2) In a large sample test for the significance of population mean, the level of significance is increased. The new critical value will _____.
   a) decrease b) increase
   c) remain unchanged d) change but nothing can be said

3) If λ is the likelihood ratio test statistic, which one of the following has got its asymptotic distribution as χ^2 distribution?
   a) $\log_e(\lambda)$ b) $\log_e(1/\lambda)$
   c) $\log_e(\lambda^2)$ d) $\log_e(1/\lambda^2)$

4) To decide about hypothesis H_0 , SPRT involves _____.
   a) one region only b) two regions only
   c) three regions d) four regions

5) In Kruskal-Wallis test of k samples, the appropriate degrees of freedom are _____.
   a) k b) $k-1$
   c) $k+1$ d) $n-k$

6) The range of Kendall's rank correlation τ is _____.
   a) 0 to 1 b) 0 to ∞
   c) -1 to 1 d) $-\infty$ to ∞

7) In testing $H_0: \sigma = \sigma_0$ in $N(0, \sigma^2)$ the critical region based on n observations is $\sum_{i=1}^{n} X_i^2 < k$ For which alternative hypothesis does this provide UMP test?
   a) $\sigma \neq \sigma_0$ b) $\sigma = \sigma_0$
   c) $\sigma > \sigma_0$ d) $\sigma < \sigma_0$

8) Which of the following is a simple hypothesis for $N(\mu, \sigma^2)$?
   a) $H_0: \mu = 5, \sigma = 2$ b) $H_0: \mu = 10$
   c) $H_0: \mu = 0, \sigma > 1$ d) $H_0: \mu \neq 3, \sigma = 1$

9) Based on random sample of size *n* from $N(0, \sigma^2)$ distribution, the pivotal quantity for construction of confidence interval for σ^2 is _____.

a) $\left(\sum_{i=1}^{n} X_i^2\right)/\sigma^2$ b) $\left(\sum_{i=1}^{n} X_i\right)/\sigma^2$

c) $\sigma^2\left(\sum_{i=1}^{n} X_i^2\right)$ d) $\sigma^2\left(\sum_{i=1}^{n} X_i\right)$

10) A 95% confidence interval estimate for the difference between two population means, $\mu_1 - \mu_2$ is determined to be (62.75, 68.52). If confidence level is reduced to 90%, confidence interval _____.

a) becomes wider b) remains same
c) becomes narrower d) more information is needed

B) Fill in the blanks: **06**

1) The degrees of freedom for a χ^2 statistic in case of contingency table of order 2×2 are _____.
2) The approximate distribution of Kruskal-Wallis test statistic is _____.
3) A hypothesis is to be tested with possible rejection is known as _____ hypothesis.
4) If there are 10 symbols of two types, equal in number, the maximum possible number of runs is _____.
5) When testing for randomness, we can use _____ test.
6) If $k = 2$ then Kruskal-Wallis H test reduces to _____.

Q.2 Answer the following. **16**

a) What is goodness of fit test? Give its application.
b) Distinguish between sign test and signed-rank test.
c) Analyze 2×2 contingency table when the cell frequencies are sufficiently large.
d) Explain in brief the test of significance for testing $H_0 : \mu = \mu_0$ for $N(\mu, \sigma^2)$ distribution, where σ is known, on the basis of large sample.

Q.3 Answer the following. **16**

a) Describe the test of significance for testing equality of variances of two normal populations.
1) for large samples
2) for small samples.

b) A sample of size one is drawn from the exponential distribution having pdf $f(x, \theta) = \frac{1}{\theta} e^{-x/\theta}, x > 0$ to test the hypothesis $H_0 : \theta = 1$ against $H_1 : \theta = 10$. The hypothesis H_0 is accepted if the observed value is ≤ 2. Find the probabilities of committing type I and type II errors and also find power of test.

Q.4 Answer the following. **16**

a) Define two kinds of errors and power of a test. Which error is minimized in statistical test? Why not both errors?
b) Use Neyman-Pearson lemma to obtain most powerful test for testing $H_0 : \sigma = \sigma_0$ against $H_1 : \sigma = \sigma_1 (> \sigma_0)$ based on random sample of size n from $N(0, \sigma^2)$ distribution.

Q.5 Answer the following. **16**

a) What is a Run? Explain how the hypothesis of the randomness of a sample of numerical observations may be tested based on the number of runs.

b) Describe Mann-Whitney U test. How is it carried out for large samples?

Q.6 Answer the following. **16**

a) Define Wald's SPRT of strength (α, β) for testing simple null hypothesis against a simple alternative hypothesis. In what respect SPRT differs from the fixed sample test.

b) Describe sequential procedure for testing the hypothesis $H_0 : \theta = \theta_0$ against the alternative $H_1 : \theta = \theta_1 (\theta_1 > \theta_0)$, where θ is parameter of Poisson distribution.

Q.7 Answer the following. **16**

a) Explain the technique of likelihood ratio test. Use it to test $H_0 : \sigma^2 = \sigma_0^2$ against $H_1 : \sigma^2 \neq \sigma_0^2$, where σ^2 *is* the variance of $N(\mu, \sigma^2)$ distribution and μ is unknown.

b) Obtain $100(1-\alpha)\%$ confidence intervals for normal location parameter μ,

1) when σ^2 known and
2) when σ^2 unknown, a random sample of size n being given.

SLR-SW-12

Seat No.	

Set P

M.Sc. (Semester - III) (New) (CBCS) Examination: March/April-2023
BIOSTATISTICS
Multivariate Statistical Methods (MSC22306)

Day & Date: Wednesday, 12-07-2023 Max. Marks: 80
Time: 11:00 AM To 02:00 PM

Instructions: 1) Q. Nos. 1 and. 2 are compulsory.
2) Attempt any three questions from Q. No. 3 to Q. No. 7
3) Figure to right indicate full marks.

Q.1 A) Fill in the blanks by choosing correct alternatives given below. **10**

1) The _____ distribution is a multivariate generalization of chi-square distribution.
   a) Hotelling's T^2 b) Multivariate Normal
   c) Wishart distribution d) None of these

2) The range for Canonical correlation is _____.
   a) [-1,1] b) [0,1]
   c) [0, ∞] d) [0.5,1]

3) While applying _____ clustering algorithm, the distance between two clusters is taken to be the largest distance between observations from two clusters.
   a) average linkage b) complete linkage
   c) single linkage d) none of these

4) If $\underline{X}$ has $N_p\left(\underline{\mu}, \Sigma\right)$ distribution then moment generating function of vector X is _____.
   a) $Exp\left(\underline{t}'\,\underline{\mu} - \frac{1}{2}\underline{t}'\,\Sigma\underline{t}\right)$ b) $Exp\left(\underline{t}'\,\underline{\mu} + \frac{1}{2}\underline{t}'\,\Sigma\underline{t}\right)$
   c) $Exp\left(\underline{t}'\underline{\mu} + \frac{1}{2}\underline{t}'\,\Sigma^{-1}\underline{t}\right)$ d) $Exp\left(\underline{t}'\underline{\mu} - \frac{1}{2}\underline{t}'\,\Sigma^{-1}\underline{t}\right)$

5) To classify a given multivariate observation to either of two populations, we use _____.
   a) Principle components analysis
   b) Discriminant analysis
   c) Cluster analysis
   d) None of these

6) Let p dimensional vector $\underline{X}$ has $N_p\left(\underline{\mu}, \Sigma\right)$ distribution. Let us partition $\underline{X} = (\underline{X}_{(1)}, \underline{X}_{(2)}$ in q and p-q component sub vectors. Then conditional covariance of $\underline{X}_{(2)}$ given $\underline{X}_{(1)}$is _____.
   a) $\Sigma_{11} - \Sigma_{12}\,\Sigma_{22}^{-1}\Sigma_{21}$ b) $\Sigma_{22} - \Sigma_{21}\,\Sigma_{11}^{-1}\Sigma_{12}$
   c) $\Sigma_{12} - \Sigma_{12}\,\Sigma_{22}^{-1}\Sigma_{21}$ d) $\Sigma_{21} - \Sigma_{21}\,\Sigma_{11}^{-1}\Sigma_{12}$

7) A principal component analysis was run and the following eigen values were obtained: 3.8, 2.1, 0.4. How many components would you retain so that 50% of the variation present in the old variables will be explained?
   a) 1 b) 2
   c) 3 d) 0

8) Let A has $W_p(n,\Sigma)$ distribution then E(A) = _____.
   a) $\frac{\Sigma}{n}$ b) Σ
   c) $n\Sigma$ d) $\frac{\Sigma}{n-1}$

9) Total variation explained by all principal components is _____ that by the original variables.
   a) equal to b) greater than
   c) less than d) none of these

10) Marginal distribution of any single variable from multivariate normal vector follows _____.
   a) Univariate normal b) Beta
   c) Gamma d) None of these

B) Fill in the blanks, **06**

1) Wishart distribution is a multivariate generalization of _____.
2) The eigen values of the matrix $\begin{bmatrix}1 & 0\\ 3 & 4\end{bmatrix}$ are _____.
3) In case of single linkage, the _____ distance between various units of two clusters is taken to be the distance among these clusters.
4) Let vector $\underline{Y}$ has $N_p(\mu\Sigma)$ distribution. For a constant matrix $A_{q\times p}$ and vector $b_{q\times 1}$ the distribution of $\underline{X} = A\,\underline{Y} + b$ is _____.
5) The _____ principal component explains maximum variation of the data.
6) The trace of variance covariance matrix is always _____.

Q.2 Answer the following **16**

a) What do you mean by discriminant analysis?
b) Obtain characteristic function of multivariate normal distribution.
c) Write a note on canonical correlation.
d) Write a note on divisive clustering.

Q.3 Answer the following

a) Obtain the rule for discrimination for two multivariate populations with densities $f_1(\underline{X})$ and $f_2(\underline{X})$. **08**

b) Explain single linkage method in detail with the help of illustration. **08**

Q.4 Answer the following

a) With usual notations, obtain the maximum likelihood estimator of Σ for $N_P\left(\underline{\mu}, \Sigma\right)$ **08**

b) Obtain the distribution of sample mean vector of the sample of size n drawn from $N_P\left(\underline{\mu}, \Sigma\right)$ **08**

Q.5 Answer the following

a) Discuss the concept of clustering. Also explain, in detail, hierarchical and non-hierarchical clustering. **08**

b) Discuss the procedure of k-means clustering. **08**

Q.6 Answer the following

a) Obtain the distribution of linear combination of components of a multivariate normal vector. **08**

b) Discuss agglomerative clustering in detail. **08**

Q.7 Answer the following

a) Explain, in detail, complete linkage and average linkage methods of calculating distance. **08**

b) Describe the mechanism of k-means clustering in detail. **08**

SLR-SW-14

Seat No.	

Set P

M.Sc. (Semester - IV) (New) (CBCS) Examination March/April-2023
BIOSTATISTICS
Demography and Health Statistics (MSC22401)

Day & Date: Monday, 10-07-2023 Max. Marks: 80
Time: 03:00 PM To 06:00 PM

Instructions: 1) Question 1 and 2 are compulsory.
2) Attempt any Three from Q.3 to Q.7
3) All questions carry equal marks.

Q.1 A) Choose the correct alternatives from the given options. **10**

1) Demography is not related to _____.
   a) Economics b) Geography
   c) Social Biology d) None of these

2) The difference between population at two census is known as _____.
   a) Population constant b) Population change
   c) Population increase d) Population decrease

3) Death rate of infants having age 0 to 4 weeks is known as _____.
   a) Infant mortality rate b) Neo natal mortality rate
   c) Post natal mortality rate d) Specific death rate

4) Which of the following is not a component of population change?
   a) Fertility b) Mortality
   c) Migration d) None of these

5) In _____ method, the census was being conducted in one day.
   a) De Facto method of census
   b) De Jure method of census
   c) Regular method of census
   d) In Facto method of census

6) There are _____ and _____ standardized death rates.
   a) correct, incorrect b) direct, indirect
   c) complete, incomplete d) absolute, negative

7) The National Rural Health Mission was conducted during _____.
   a) 2007-12 b) 2002-09
   c) 2005-12 d) 2005-08

8) The NFHS -I was conducted during _____.
   a) 1992-93 b) 1993-94
   c) 1998-99 d) 1999-2000

9) The Malthusian theory have which of the following postulate?
   a) The law of diminishing returns applies to agricultural yield
   b) Food is essential for man's existence
   c) There is a natural instinct to increase the population
   d) All the above

10) Which of the following is not a Socio-Economic theories of population?
   a) Leibenstein's Motivational theory of population growth
   b) Karl Marx theory of surplus population
   c) Dumont's theory of social capillarity
   d) Pearl and Reeds Logistic Curve Theory of Population

B) Fill in the blanks: **06**

1) The base year is the year of _____.
2) Projection horizon means number of years between _____ and _____.
3) Geometric change is _____ extrapolation method of population projection.
4) The IMR of certain population of size 72,000 where there are 5,800 infant deaths during the year is _____.
5) Weight of infants at birth is as _____ cause of Mortality.
6) The people migrated from TN to Maharashtra are _____to the TN.

Q.2 Answer the following. **16**

a) Define direct and indirect standardized death rates.
b) State important factors affecting Mortality.
c) State the assumptions for Pearl and Reed's Logistic curve theory.
d) How Demography is related to Geography? Explain.

Q.3 Answer the following.

a) What are subject matter of Demography? Discuss in detail. **08**
b) Discuss in detail the Migration as a component of population change. **08**

Q.4 Answer the following.

a) Explain Malthusian Theory of Population. **08**
b) Discuss the Cohart-Component method of Population projection. **08**

Q.5 Answer the following.

a) Write in detail salient features and Merits of NFHS 1 and 2. **08**
b) Explain Demography as a scientific discipline. **08**

Q.6 Answer the following.

a) What is Dumont's theory of social capillarity? Explain criticism on it. **08**
b) Discuss the Government policy to control population growth. **08**

Q.7 Answer the following.

a) Explain Pearl and Reed's logistic curve theory. What is criticism on it? **08**
b) Discuss in detail Optimum theory of Population. **08**

SLR-SW-15

Seat No.	

Set P

M.Sc. (Semester - IV) (New) (CBCS) Examination: March/April-2023
BIOSTATISTICS
Clinical Trials (MSC22402)

Day & Date: Wednesday, 12-07-2023 Max. Marks: 80
Time: 03:00 PM To 06:00 PM

Instructions: 1) Q. Nos. 1 and. 2 are compulsory.
2) Attempt any three questions from Q. No. 3 to Q. No. 7
3) Figure to right indicate full marks.

Q.1 A) Fill in the blanks by choosing correct alternatives given below. **10**

1) What do you mean by a randomized design?
   a) Randomly assigning subjects either for placebo or active dose
   b) The subjects do not know which study treatment they receive
   c) Patients injected with placebo and active doses
   d) Signed document of the recruited patient for the clinical trial procedures

2) If the Relative Bioavailability is 1, it indicates: _____.
   a) Bioavailability of dosage form of one drug is same as that of the other dosage form
   b) Complete binding of the drugs to the proteins as compared to the standard drug
   c) Complete bioavailability of the drug
   d) Complete distribution of the drug

3) Three brands of coffee are rated for taste on a scale of 1 to 10. Six persons are asked to rate each brand so that there is a total of 18 observations. The appropriate test to determine if three brands taste equally good is _____.
   a) One way analysis of variance
   b) Wilcoxon rank-sum test
   c) Spearman rank difference
   d) Kruskal-Wallis test

4) When Phase IV clinical trials Conducted?
   a) After FDA Approval to marketing the drug
   b) Before FDA Approval to marketing the drug
   c) During Phase III trial Process
   d) None of these

5) Which of the following would occur in a longitudinal study?
   a) Measures are taken from different participants over an extended period
   b) Participation is expected to last for a minimum of 24 hours
   c) Measures are taken from the same participants on different occasions usually over extended periods of time
   d) Measures are taken from participants in at least six different countries

6) What is ANDA?
   a) Abbreviated New Drug Application
   b) Anything as New Drug Application
   c) All New Drug Application
   d) None of the above

7) Which of the following is NOT associated with Phase 4 clinical trials?
   a) Also known as post-marketing surveillance
   b) A mixture of populations involved
   c) small numbers of participants
   d) Monitoring of long-term safety

8) In clinical trials, 'Treatment' can be _____.
   a) Placebo
   b) any pharmaceutical identity
   c) Any medical device
   d) All of the above

9) Which type of trials are needed for the approval of generic drug?
   a) Superiority trials
   b) Combination trials
   c) Equivalence trials
   d) None of these

10) Carryover effect can be estimated in _____.
   a) Parallel design
   b) Crossover design
   c) Both a) and b)
   d) none of these

B) State whether the following statements are True or False: **06**

1) We can estimate carry over effect in 'Crossover Design'.
2) In clinical trials, 'Treatment' can be placebo.
3) Randomization is effective tool to prevent the selection bias.
4) The whole market will be under Phase III surveillance.
5) To avoid experimenter bias, when the experimenter nor the participant is aware of which group the participant is in, this is known as single blind.
6) The aim of post marketing studies is Safety and comparisons with other medicines.

Q.2 Answer the following **16**

a) Write notes on run in and washout period.
b) Write the note on: Investigation New Drug Application (INDA).
c) What are the major objectives behind conduction of the clinical trials (CTs)
d) Write a note on:
   i) Active control and equivalence trials
   ii) Combination trials

Q.3 Answer the following

a) Explain the concept of Bioavailability and Bioequivalence. **08**
b) What is the meaning of blinding? Way it is used in clinical trials? Explain the type of bindings. **08**

Q.4 Answer the following

a) What is randomization? Why randomization is needed? What are the types of randomizations involved in clinical trials? **08**
b) Explain concept of protocol and process of protocol developments in clinical trials. **08**

Q.5 Answer the following

a) What are crossover designs? In which situations crossover designs are useful? **08**

b) Explain the all phases involve in the development of clinical trials. **08**

Q.6 Answer the following

a) What is longitudinal design? Write advantages and disadvantages of longitudinal design. **08**

b) Explain the overall clinical drug development process. **08**

Q.7 Answer the following

a) Explain the difference between the Multicenter trails and Meta analysis. **08**

b) Explain the concept of sample size. Discuss are the factors necessary to calculate the appropriate sample size. **08**

SLR-SW-16

Seat No.	

Set P

M.Sc. (Semester - IV) (New) (CBCS) Examination: March/April-2023
BIOSTATISTICS
Survival Analysis (MSC22403)

Day & Date: Friday, 14-07-2023 Max. Marks: 80
Time: 03:00 PM To 06:00 PM

Instructions: 1) Question no. 1 and 2 are compulsory.
2) Attempt any three questions from Q. No. 3 to Q. No. 7.
3) Figure to right indicate full marks.

Q.1 A) Multiple choice questions. **10**

1) A series system is a special case of k-out-of-n system when _____.
   a) k = 1 b) k=2
   c) k = n-l d) k = n

2) Which of the following rate function corresponds to DFR distribution?
   a) h (t) = t b) h (t) = e^t
   c) h (t) = e^{-t} d) h (t) = t e^t

3) If $R_s(t)$ and $R_p(t)$ are the reliabilities of series and parallel systems respectively then for a system having identical components, which of the following is true?
   a) $\frac{R_s(t)}{R_p(t)} > 1$ b) $\frac{R_s(t)}{R_p(t)} < 1$
   c) $\frac{R_s(t)}{R_p(t)} = 1$ d) none of the above

4) Let p_i is the reliability of i^{th} component then reliability of parallel system of n independent components is _____.
   a) $1 - \coprod_{i=1}^{n} p_i$ b) $1 - \prod_{i=1}^{n} p_i$
   c) $1 - \prod_{i=1}^{n} (1 - p_i)$ d) $1 - \sum_{i=1}^{n} (1 - p_i)$

5) A life time distribution F having finite mean is said to be NWUE for $t \geq 0$, if _____.
   a) $\mu_t \leq \mu_0$ b) $\mu_t \geq \mu_0$
   c) $\mu_t = \mu_0$ d) none of the above

6) Which of the following is a particular case of random censoring?
   a) Type I b) Type II
   c) both type I and type II d) none of the above

7) Which of the following is not true?
   a) K-M estimator is parametric in nature
   b) K-M estimator is generalized maximum likelihood estimator
   c) K-M estimator is consistent
   d) K-M estimator is also known as product limit estimator

8) In survival analysis, the data set may contain_____.
   a) only left censored observations
   b) only right censored observations
   c) both left and right censored observations
   d) none of the above

9) Which of the following is an example of right censored observation?
   a) patient decided to move elsewhere
   b) patient become non-cooperative
   c) person may not experience the event before the study ends
   d) all the above

10) Which of the following is true?
   a) The hazard function has no upper bound
   b) The hazard function has an upper bound
   c) The hazard function is a probability
   d) $S(t) = P(T > t)$ is a hazard function

B) Fill in the blanks. **06**

1) If ∅(x) is a structure function then dual of ∅ (x) is _____.
2) DFRA property is preserved under_____.
3) For _____ lifetime distribution, each member has non-monotonic failure rate.
4) The scaled TTT transform for exponential distribution with mean λ. is _____.
5) In time censoring experiment with $n = 10, t_0 = 15$ two units failed at 9 and 13. The total time on TTT statistics has value equal to_____.
6) Product limit estimator of survival function is developed by____.

Q.2 Answer the following. **16**

a) Define k out of n system. Obtain the reliability function of this system.
b) Give two definitions of star shaped function and prove their equivalence.
c) Define TTT transform. Show that for an IFR distribution TTT transform is a convex function.
d) Obtain the nonparametric estimator of survival function based on complete data.

Q.3 Answer the following.

a) Define reliability of a component and reliability of a system. Obtain the reliability of series and parallel systems of an independent components. **08**
b) Obtain the structure function of a coherent system by using minimal path sets. Illustrate the same by an example. **08**

Q.4 Answer the following.

a) Define type-II censoring. Derive the likelihood function of observed data under type II censoring hence obtain MLE of mean of exponential distribution. **10**
b) Discuss maximum likelihood estimation of parameters of a Weibull distribution based on complete data. **06**

Q.5 Answer the following.

a) Define IFR and IFRA class of distributions. If F ϵ IFR then show that Fϵ IFRA. **08**
b) Define star shaped function. Prove that Fϵ IFRA if and only if – log R(t) is star shaped. **08**

Q.6 Answer the following.

a) Describe Kaplan-Meier estimator and derive an expression for the same. **08**

b) Describe in detail Mantel-Haenzel test. Indicate the null distribution of test statistic. **08**

Q.7 Answer the following.

a) If failure time of an item has gamma distribution obtain the failure rate function. **08**

b) Derive Greenwood's formula for an estimate of variance of actuarial estimator of survival function. **08**

SLR-SW-17

Seat No.	

Set P

M.Sc. (Semester - IV) (New) (CBCS) Examination: March/April-2023
BIOSTATISTICS
Time Series Analysis (MSC22407)

Day & Date: Sunday, 16-07-2023 Max. Marks: 80
Time: 03:00 PM To 06:00 PM

Instructions: 1) Q. Nos. 1 and 2 are compulsory.
2) Attempt any three questions from Q. No. 3 to Q. No. 7
3) Figure to right indicate full marks.

Q.1 A) Choose the correct alternative: **10**

1) Which of the following is not the component of time series?
   a) Level b) Secular Trend
   c) Seasonal d) None of these

2) The mean of a stationary process is _____
   a) always zero b) constant
   c) non-constant d) None of these

3) A strictly stationary process is _____ weakly stationary
   a) always
   b) only for multivariate Gaussian process
   c) Never
   d) Sometimes

4) Let $\{X_t\}$ be an autoregressive process of order two then the condition for its stationarity is _____.
   a) $\phi_2 - \phi_1 < 1, \phi_2 + \phi_1 < 1, |\phi_2| < 1$
   b) $\phi_2 - \phi_1 > 1, \phi_2 + \phi_1 < 1, |\phi_2| < 1$
   c) $\phi_2 - \phi_1 < 1, \phi_2 + \phi_1 > 1, |\phi_2| < 1$
   d) $\phi_2 - \phi_1 < 1, \phi_2 + \phi_1 < 1, |\phi_2| > 1$

5) The PACF of AR (1) process is zero after lag
   a) One b) two
   c) Three d) zero

6) The singe exponential smoothing equation is _____.
   a) $S_t = \alpha Y_{t-1} + (1-\alpha) S_{t-1} \quad t \geq 2$
   b) $S_t = \alpha^2 Y_{t-1} + (1-\alpha) S_{t-1} \quad t \geq 2$
   c) $S_t = \alpha Y_{t-1} + (1-\alpha)^2 S_{t-1} \quad t \geq 2$
   d) $S_t = \alpha^2 Y_{t-1} + 2(1-\alpha) S_{t-1} \quad t \geq 2$

7) In a given time series there is only trend component present, then which exponential smoothing best fit for the series?
   a) Single exponential smoothing
   b) Double exponential smoothing
   c) Triple exponential smoothing
   d) Quadratic smoothing

8) Holt - Winter smoothing method is used when there is _____.
   a) Trend component present only
   b) Seasonal component present only
   c) Trend and seasonal both present
   d) There is level component present only

9) Box - Jenkin's suggested to use _____ as an alternative method to find best model for given time series
   a) Interpolation b) Differencing
   c) Extrapolation d) Exponential smoothing

10) The process $X_t = \phi_1 X_{t-1} + Z_t$ where $\{Z_1\} \sim WN(0, \sigma^2)$ is process if _____.
   a) $|\phi_1| < 1$ b) $|\phi_1| > 1$
   c) $|\phi_1| = 1$ d) $|\phi_1| < 1.5$

B) Fill in the blanks. **06**

1) The autocovariance function of given time series at lag zero is nothing but _____ of the series
2) The ACF of time series with seasonal component shows _____ pattern
3) ACF of MA (q) process has insignificant autocorrelations after lag _____
4) A real-valued function defined on the integers is the autocovariance function of a stationary time series if and only if it is even and _____
5) Differencing method can be used to estimate _____ and _____ component in the given time series if present
6) The causal representation of ARMA (p, q) process $\{X_t\}$ is _____.

Q.2 Answer the following. **16**

a) Define the autocovariance function and state is characterizing properties.
b) Write a short note on preliminary transformations in time series analysis.
c) Write a short note on classical decomposition model of time series.
d) Define AR (1) process. Obtain its autocorrelation function.

Q.3 Answer the following.

a) Define MA (1) process and hence obtain its partial autocorrelation function. **08**
b) Define AR (2) process and hence obtain the condition of stationarity of the same process. **08**

Q.4 Answer the following.

a) Describe the diagnostic checking methods in time series analysis **08**
b) Describe any two tests for stationarity of a time series **08**

Q.5 Answer the following.

a) Define ARMA (1,1) process and hence obtain its autocorrelation function. **08**
b) Write in detail model identification methods. **08**

Q.6 Answer the following.

a) Determine which of the following processes are causal and/or invertible **08**
   i) $X_t + 0.6X_{t-1} = Z_t + 0.04Z_{t-1}$
   ii) $X_t + 1.6X_{t-1} = Z_t - 0.4Z_{t-1} + 0.04Z_{t-1}$
   in both process $\{Z_t\} \sim WN(o, \sigma^2)$
b) Describe analysis of Seasonal ARIMA $(p, d, q) \times (P, D, Q)$ process. **08**

Q.7 Answer the following.

a) Explain moving average as a method of estimation and elimination of trend. **08**

b) Explain double exponential smoothing method in detail. **08**

SLR-SW-18

Seat No.	

Set P

M.Sc. (Semester - IV) (New) (CBCS) Examination: March/April-2023
BIOSTATISTICS
Data Mining (MSC22408)

Day & Date: Sunday, 16-07-2023 Max. Marks: 80
Time: 03:00 PM To 06:00 PM

Instructions: 1) Question no. 1 and 2 are compulsory.
2) Attempt any three questions from Q. No. 3 to Q. No. 7.
3) Figure to right indicate full marks.

Q.1 A) Choose the correct alternative. **10**

1) In supervised learning, if 90% data is of positive class label value, and just 10% data is of negative class label value, such a data will be called _____.
   a) Imperfect data b) Imbalanced data
   c) Scattered data d) Bad data

2) The classification problem based on available data with known class labels is ____.
   a) Supervised learning b) Semi-supervised learning
   c) unsupervised learning d) All of these

3) The data used to check accuracy of the built model is called _____.
   a) Training data b) Testing data
   c) irrelevant data d) residual data

4) In ____ linkage method, the distance between two clusters is considered as average of all the possible pairwise distances between the observations of these clusters.
   a) Single linkage b) Average linkage
   c) Complete linkage d) Centroid linkage

5) Market-basket problem was formulated by ____.
   a) Agrawal et al. b) Toda et al.
   c) Steve et al. d) Simon et. Al

6) Which of the following can be considered as a generalization of regression technique?
   a) ANN b) SVM
   c) kNN d) Decision Tree

7) In k- nearest neighbor algorithm, k stands for _____.
   a) Number of neighbors that are investigated
   b) Number of Iterations
   c) Number of total records
   d) Random number

8) Which of the following can be considered to be generalization of decision tree?
   a) kNN b) Bayes' classifier
   c) Random forest d) None of these

9) Each neuron is made up of a number of nerve fibers called ____.
   a) Molecules b) Dendrites
   c) Atoms d) Sigmoid

10) In clustering, if the distance between two clusters is considered as distance between their centroids, then it is called as ____.
   a) Single linkage b) Average linkage
   c) Complete linkage d) Centroid linkage

B) State whether following statements are True or False. **06**

1) kNN is a supervised learning tool.
2) In clustering analysis, class labels are provided.
3) kNN classifier is also called as lazy classifier.
4) Most of the data should be used to check performance of the data.
5) Data used to verify performance of the built model is called training data.
6) Process of removing duplicate records is called data cleaning.

Q.2 Answer the following. **16**

a) Write a note on problem of classification.
b) Why kNN classifier is also called as lazy classifier?
c) Explain classification and regression in detail.
d) Explain the steps Involved in Supervised Learning.

Q.3 Answer the following.

a) Write a short note on clustering. **08**
b) Describe k-nearest neighbor classifier in detail. Also discuss the drawbacks of kNN classifier. **08**

Q.4 Answer the following.

a) Describe Artificial Neural Network (ANN) in detail. **08**
b) Write down the algorithm for Bayesian classifier. **08**

Q.5 Answer the following.

a) Explain logistic regression classifier in detail. **08**
b) What are the different metrics for Evaluating Classifier Performance? **08**

Q.6 Answer the following.

a) Explain the class imbalance problem in detail. **08**
b) Describe unsupervised learning. Also explain in detail, association rules and prediction. **08**

Q.7 Answer the following.

a) Explain decision tree in detail. Also explain the idea of random forest. **08**
b) Describe- **08**
   i) Accuracy of a model
   ii) Precision of a model