BSc (Honours) Semester -I Examination 2020

Subject: Statistics

Paper-CC1 (Descriptive Statistics-Theory)

Full Marks: 40 Time: 3 hours

Answer ***any four*** questions:

(Notations carry usual meanings)

1. a) Explain, with suitable examples, the distinction i) between an attribute and a variable, and ii) between discrete variable and a continuous variable. 6

   b) Suppose each value of a variable x lies between p and q, both values inclusive. Show that $p \leq \bar{x} \leq q$, where $\bar{x}$ is the mean of x. 4

2. a) Define different types of average. Cite examples where which one should be used. 6

   b) Let x be a variable assuming the values 1, 2, 3, …, k and let $F_1 = n, F_2, F_3, \ldots, F_k$ be the corresponding cumulative frequencies of the 'greater than type'. Prove that 4

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{k} F_i$$

3. a) The variance of $1, 2, 3, \ldots, n$ is 24. Find n. 4

   b) What is skewness and kurtosis of a frequency distribution? Give at least two suitable measures for skewness and kurtosis. 6

4. a) When do we say two variables are correlated? How to measure its extent? Derive the limits of the product moment correlation coefficient. Hence interpret the cases $r = \pm 1$. 6

   b) Prove that regression coefficients do not depend on change of origin but depend on change of scale. 4

5. a) Derive the formula of Spearman’s rank correlation coefficient. 5

b) Given that $X = 4Y + 5$ and $Y = kK + 4$, are the two regression lines of X on Y and Y on X respectively, show that $0 < k < \frac{1}{4}$. If $k = \frac{1}{16}$, find the means of the two variables and coefficient of correlation between them. 5

6. a) Explain ‘price index number’ with an example and write its different uses. 5

b) For a distribution, the mean is 10, variance is 16, $\gamma_1 = +1$ and, $\beta_2 = 4$.

Obtain the first four moments and comment upon the nature of the distribution. 5

BSc (Honours) Semester -I Examination 2020

Subject- Statistics

Paper- CC1B (Descriptive Statistics-Practical)

Full Marks: 20 Time: 2 hours

Answer all questions:

(Notations have usual meanings)

1. For the frequency distributions given in the adjoining table, the mean calculated from the first was 25.4 and that from the second was 32.5. Find the values of x and y. 4

Class	Distribution-I Frequency	Distribution-II Frequency
10-20	20	4
20-30	15	8
30-40	10	4
40-50	x	2x
50-60	y	y

2. Obtain the correlation coefficient between the heights of fathers (x) and of the sons (y) from the following data.

x:	65	66	67	68	69	70	71	67
y:	67	68	64	72	70	67	70	68

4

3. Out of the two lines of regression given by: X+2Y-5=0 and 2X+3Y-8=0 which one is the regression line of X on Y?
   Use the equations to find the mean of X and the mean of Y. If the variance of X is 12 calculate the variance of Y.
   When X=15 what would be the value of Y? 4

4. Compute Fisher's Ideal Index Number for the following data and interpret suitably. 4

Commodity	Base year Price	Base year quantity	Current year price	Current year quantity
A	6	30	15	40
B	5	40	10	55
C	10	25	12	20
D	4	15	3	30
E	2	50	5	28

5. The mean of 5 observations is 4.4 and variance is 8.24. If three of the five observations are 1, 2 and 6, find the other two. 4

B.Sc. (Honours) Examination, 2020
Semester-I
Statistics
Course: CC-2
(Calculus [Theory and Tutorial])
Time: 3 Hours Full Marks: 60

Questions are of value as indicated in the margin
Notations have their usual meanings

$\underline{Group - A}$ $10 \times 1 = 10$

1. Answer any ten of the following questions with proper justification.

(a) Find $\lim\limits_{(x,y)\to(0,0)} \dfrac{xy^3}{x^2+y^6}$.

(b) If $y = \frac{1}{x^2-3x+2}$, then find $\frac{d^n y}{dx^n}$.

(c) Give an example of a function which has no local maxima or, minima.

(d) If $f(x,y) = \frac{x^2+y^2}{x^2-y^2}$ then find $x\frac{\partial f}{\partial x} + y\frac{\partial f}{\partial y}$.

(e) Give an example of a curve which has triple point.

(f) Give example of two curves which has arcnode.

(g) Plot the curve $y = |[x]|$.

(h) Find the length of the curve $y = \sqrt{4-x^2}$ from $x = -\sqrt{2}$ to $x = \sqrt{2}$.

(i) Find $\int_0^2 [x]^2 \, dx$.

(j) Find $\int_{-\infty}^{\infty} e^{-x^2} \, dx$.

(k) What do you mean by the explicit solution of an ordinary differential equation?

(l) Find the singular solution of the differential equation: $3+\left(\frac{dy}{dx}\right)^2 = x^2$.

(m) Find the general solution of $\frac{d^2y}{dx^2} + 2y = 0$.

(n) Find the type, order and the degree of the differential equation: $\dfrac{\partial^2 z}{\partial x^2} + \dfrac{\partial^2 z}{\partial y^2} + \dfrac{\partial^2 z}{\partial w^2} = 0$.

$\underline{Group - B \ (Answer\ any\ five\ questions)}$ $5 \times 6 = 30$

2. (a) Let $f(x,y) = \begin{cases} \sqrt{x^2+y^2}\sin\left(\dfrac{1}{\sqrt{x^2+y^2}}\right), & if\ (x,y) \neq (0,0) \\ 0, & if\ (x,y) = (0,0) \end{cases}$

Then check for the existence of f_x, f_y at $(0,0)$.

(b) Find $\lim\limits_{x\to 0}\left[\dfrac{1}{x^2} - \dfrac{1}{x\tan x}\right]$. 3+3

3. (a) If $u = \cos^{-1} \frac{x+y}{\sqrt{x}+\sqrt{y}}$, then show that $xu_x + yu_y + \frac{1}{2}\cot u = 0$.
(b) Show that points of inflexion of the curve $y = x \sin x$ lie on $y^2(4+x^2) = 4x^2$.

3+3

4. (a) Find the area of the smaller of the two regions enclosed between $\frac{x^2}{9} + \frac{y^2}{2} = 1$ and $y^2 = x$.
(b) Show that the length of the loop of $9y^2 = (x-2)(x-5)^2$ is $4\sqrt{3}$. 3+3

5. (a) Find the value of $\lim_{n\to\infty}\left[\left(1+\frac{1}{n^2}\right)^{\frac{2}{n^2}}\left(1+\frac{2^2}{n^2}\right)^{\frac{4}{n^2}}\cdots\left(1+\frac{n^2}{n^2}\right)^{\frac{2n}{n^2}}\right]$.
(b) Show that $\int\int_R e^{y/x}\,dxdy = \frac{e-1}{2}$, where R is the triangle bounded by $y = x,\ y = 0,\ x = 1$. 3+3

6. (a) Show that $\int\int_E (x^2+y^2)dx\ dy = 6/35$, where E is the region bounded by $y = x^2,\ y^2 = x$.
(b) Show that $\int_0^\infty \frac{x^{m-1}}{(a+bx)^{m+n}}dx = \frac{\beta(m,n)}{a^n b^m},\ a,b,m,n > 0$. 3+3

7. (a) Check whether the initial value problem: $\frac{dy}{dx} = y^{1/3},\ y(0) = 0$ has unique solution or not.
(b) Show that infinity is not a regular singular point for the Bessel equation: $x^2y'' + xy' + (x^2-n^2)y = 0$. 3+3

8. (a) Find the general solution of the ordinary differential equation $y^{(iv)} + 2y = e^x$.
(b) Solve the ordinary differential equation $x^2\frac{d^2y}{dx^2} - 4x\frac{dy}{dx} + 6y = 0$. 4+2

9. (a) Find PDE by eliminating the arbitrary constants: $2z = \frac{x^2}{a^2} + \frac{y^2}{b^2}$.
(b) Find the complete integral for the PDE: $z^2 = pqxy$. 3+3

Group – C (Answer any two questions) $2 \times 10 = 20$

10. (a) Find the general solution of : $y'' + y = \tan^3 x$.
(b) Use Lagrange's Multiplier method to find the shortest distance between $(-1,3)$ and the straight line $12x - 5y + 71 = 0$. 5+5

11. (a) Find the volume common to the cylinder $4(x^2+y^2) = 1$ and $4(x^2+z^2) = 1$.
(b) Show that the surface area of the solid obtained by rotating the cycloid $x = a(t - \sin t),\ y = a(1-\cos t)$ about $x-axis$ is $\frac{64\pi a^2}{3}$. 5+5

12. (a) Find a power series solution of the initial value problem: $(1-x)y'' + xy' - y = 0,\ y(0) = 1,\ y'(0) = 1$.
(b) Find a complete, singular and general integrals of the partial differential equation $2xz - px^2 - 2qxy + pq = 0$. 5+5

B.Sc. (Honours) Examination, 2020
Semester-III
Statistics
Course: CC 5
(Sampling Distribution)
Time: 3 Hours   Full Marks: 40

Questions are of value as indicated in the margin
Notations have their usual meanings

Answer **any four** questions

1. 1. a. Show that $X_{(k)}$ in a random sample of size n from a $R(0,1)$distribution has a beta distribution with parameters $(k, n-k+1)$.
   b. Show that the pdf of the sample range from an $R(0,1)$ distribution is given by $n(n-1)r^{n-2}(1-r), 0 \le r \le 1, r(range) = x_{(n)} - x_{(1)}$. 5+5
2. a. Write down the test procedure to perform a large sample test for comparing two independent binomial proportions.
   b. Hence or otherwise find a $100(1-\alpha)\%$ confidence interval for the difference of proportions. Find the expected length of the interval. 6+4
3. a. Let X_1 and X_2 be independently binomially distributed random variables, with parameters $(n_1, \frac{1}{2})$ and $\left(n_2, \frac{1}{2}\right)$, respectively. Show that $X_1 - X_2 + n_2$ has the binomial distribution with parameters $(n_1 + n_2, \frac{1}{2})$.
   b. Let X and Y be independently distributed, each in the form $N(0,1)$. Show that $Z = X/Y$ has the Cauchy distribution with pdf
   $$f(z) = \frac{1}{\pi[1+z^2]}$$
   What would be the distributions of $W_1 = X/|Y|$ and $W_2 = X/|X|$? 4+6
4. (a) Let X_1 and X_2 be two independently distributed random variables following $R(0,1)$ distributions. Then find the distributions of the following two random variables:
   $$U_1 = \sqrt{-2lnX_1}\cos 2\pi X_2\, , U_2 = \sqrt{-2lnX_1}\sin 2\pi X_2$$
   (b) State and prove WLLN for Bernoulli variable.
5. (a) Define χ^2 distribution. Find its mean and variance. Prove the additive property of this distribution.
   (b) State and prove DeMoivre-Laplace Central Limit theorem (CLT). 5+5

6. (a) Derive the pdf of an F-distribution.
   (b) If X and Y are independent random variables each distributed uniformly over $(0,1)$, find the distributions of
   (i) $\frac{X}{Y}$ (ii) XY and (iii) $\sqrt{X^2+Y^2}$ 3+7

B. Sc. Examination 2020-2021

Semester: III

Paper: CC5B (Practical)

Subject: Sampling Distribution

Time: 2 Hours **Full Marks: 20**

Questions are of value as indicated in the margin

1. A researcher believes that the proportion of boys at birth should be 50%. To test this belief randomly selected birth records of 10,000 babies born during a period were examined. It was found in the sample that 52.55% of the newborns were boys. Determine whether there is sufficient evidence, at the 5% level of significance, to support the researcher's belief. 4

2. To determine whether there was a difference between the ranges of deer located in two different geographical areas, some deer were caught, tagged, and fitted with small radio transmitters. Several months later, the deer were tracked and identified, and the distance X from the release point was recorded. The mean and standard deviation of the distances from the release point were:

Location	Sample Size	Sample Mean	Sample SD
1	40	2985 ft	1130 ft
2	50	3200 ft	960 ft

Do the data provide sufficient evidence to indicate that the mean distances differ for the two geographical regions? 5

3. In a study it is reported that of 569 study participants who regularly used aspirin after being diagnosed with colorectal cancer, there were 81 colorectal cancer-specific deaths, whereas among 720 similarly diagnosed individuals who did not subsequently use aspirin, there were 131 colorectal cancer-specific deaths. Does this data suggest that the regular use of aspirin after diagnosis will decrease the incidence rate of colorectal cancer-specific deaths? Also find 95% confidence interval for the difference of proportions.

5

4. For developing countries in Asia and Africa, let p_1 and p_2 be the respective proportions of preschool children with chronic malnutrition. If respective random samples of $n_1 = 1300$ and $n_2 = 1100$ yielded $y_1 = 520$ and $y_2 = 385$ children with chronic malnutrition.
   a. Give a point estimate of p_1
   b. Find 95% confidence interval for p_1
   c. Find point estimate of $p_1 - p_2$
   d. Find 95% CI for $p_1 - p_2$ 6

BSc (Honours) Semester -III Examination 2020

Subject- Statistics

Paper- CC6A (Statistical Inference-Theory)

Full Marks: 40 Time: 3 hours

Answer any four questions: (Notations have usual meanings)

1. What do you understand by Point Estimation? Define the following terms and give one example for each: Sufficient Statistic, Unbiased Estimator, Consistent Estimator, Efficient Estimator. 10

2. a) Show that if T is an unbiased estimator of a parameter θ, then $\lambda_1 T + \lambda_2$ is an unbiased estimator of $\lambda_1\theta + \lambda_2$, where λ_1 and λ_2 are known constants, but T^2 is a biased estimator of θ^2. 6

b) Let T_n be an estimator of θ with variance σ_n^2 and $E(T_n) = \theta_n$. Prove that if $\theta_n \to \theta$ and $\sigma_n^2 \to 0$, as $n \to \infty$ then T_n is a consistent estimator of θ. 4

3. a) Let $x_1, x_2, \ldots, x_n$ be a random sample from a population with pdf

$$f(x,\theta)=\theta\, e^{-\theta x}\ ;\ x > 0,\ \theta > 0$$

Find Cramer-Rao lower bound for the variance of the unbiased estimator of θ. 4

b) State Neyman-Pearson Lemma for testing simple versus simple hypothesis. If $x \geq 1$ is the critical region for testing H_0: $\theta=2$ against the alternative H_1: $\theta=1$, on the basis of a single observation from the population

$$f(x,\theta) = \begin{cases} \theta\, x^{\theta-1}, & \text{if } 0 < x < 1 \\ 0, & \text{otherwise} \end{cases}$$

where $0 < \theta < \infty$.

Obtain the values of type-I and type-II errors and power function of the test. 6

4. a) Let $x_1, x_2, \ldots, x_n$ be a random sample from the Bernoulli population with parameter θ, $0 < \theta < 1$. Obtain a sufficient statistic for θ and show that it is complete. Hence find minimum variance unbiased estimator (MVUE) of θ. 7

b) $x_1, x_2, \ldots, x_{10}$ is a random sample of size 10 from a Poisson distribution with mean λ. Show that the critical region W defined by $\sum_{i=1}^{10} x_i \geq 3$, is the best critical region for testing H_0: $\lambda=0.1$ against the alternative H_1: $\lambda=0.5$. 3

5. What are simple and composite statistical hypotheses? Give examples. Explain the following terms in the context of testing of statistical hypothesis:

Most Powerful Test, Uniformly Most Powerful Test, Power function of a test, Level of significance. 10

6. a) An urn contains 6 marbles of which θ are white and others are black. In order to test the null hypothesis H_0: $\theta=3$ against the alternative H_1: $\theta=4$, two marbles are drawn at random (without replacement) and H_0 is rejected if both the marbles are white; otherwise H_0 is accepted. Find the probabilities of committing type-I and type-II errors. 5

b) Given a random sample $x_1, x_2, \ldots, x_n$ of size n from the distribution with pdf

$$f(x,\theta)=\theta\, e^{-\theta x}\ ;\ x > 0,\ \theta > 0$$

show that UMP test for testing H_0: $\theta = \theta_0$ against H_1: $\theta < \theta_0$ is given by $W=\{\mathbf{x}: \sum x_i \geq (1/2\theta_0)\chi^2_{\alpha,\, 2n}\}$. 5

BSc (Honours) Semester -III Examination 2020

Subject- Statistics

Paper- CC6B (Statistical Inference-Practical)

Full Marks: 20 Time: 2 hours

Answer all questions:

(Notations have usual meanings)

1. Let $0.3, 0.8, 0.2, 0.9, 0.2, 0.4, 0.8$ are random sample from $U(0, \theta)$. Compute an unbiased estimate of θ. 5

2. Let $-2, 5, -6, 9, -5, -9$ be the observed values of a random sample of size 6 from a discrete distribution having probability density function

$$f(x, \theta) = \begin{cases} e^{-(x-\theta)} & , if\ x > \theta \\ 0, & otherwise \end{cases}$$

Then find the maximum likelihood estimate of θ. 5

3. Let X be a random variable with probability density function $f \in \{f_0,\ f_1\}$, where

$$f_0(x) = \begin{cases} 2x, & if\ \ 0 < x < 1 \\ 0, & otherwise \end{cases}$$

and

$$f_1(x) = \begin{cases} 3x^2, & if\ \ 0 < x < 1 \\ 0, & otherwise \end{cases}$$

Let we wish to test the null hypothesis H_0: $f = f_0$ against the alternative hypothesis H_1: $f = f_1$. Find the most powerful test at level of significance $\alpha = 0.19$. 5

4. Let X be a binomial distribution with parameter n and $p, n = 3$. For testing the hypothesis $H_0: p = 2/3$ against $H_1: p = 1/3$, let a test be: "Reject H_0 if $X \geq 2$ and accept H_0 if $X \leq 1$". Then find the probabilities of Type-I and Type-II errors. 5

B.Sc. (Honours) Examination, 2020
Semester-IV
Statistics
Course: CC-7
(Mathematical Analysis [Theory and Tutorial])
Time: 3 Hours Full Marks: 60

Questions are of value as indicated in the margin
Notations have their usual meanings

Group – A (Answer any ten questions) $10 \times 1 = 10$

1. Answer the following questions with proper justification.

   (a) Give an example of a set which is not ordered field.

   (b) How do you represent $\sqrt{8}$ on directed line?

   (c) Write down the definition of equivalent sets and give an example.

   (d) Find $Sup(S)$ and $Inf(S)$ of $S = \{|x| : x^2 < 1, x \in R\}$.

   (e) Discuss if $\{\tan \frac{10\pi}{n}\}$ is a sequence or not.

   (f) Justify whether every convergent sequence is bounded or not.

   (g) Suppose $\sum u_n$ is an infinite series with $\lim\limits_{n\to\infty} u_n = 0$. What can you conclude about the series convergence?

   (h) If $\sum u_n$ be a convergent series of positive real numbers then what can you say about the convergence of series $\sum u_n^4$.

   (i) How to represent a step function by using indicator function?

   (j) Is $f(x) = [x]$, $x \in \boldsymbol{R}$ a bijective function?

   (k) Find $\lim\limits_{x\to 0} Sgn(x)$.

   (l) What do you mean by uniform continuity?

   (m) How you can derive Lagrange's mean value theorem from Cauchy's mean value theorem?

   (n) What is the relation between ∇ and Δ operator?

Group – B (Answer any five questions) $5 \times 6 = 30$

2. (a) Show that $2.7^n + 3.5^n - 5$ is divisible by 12.
   (b) Show that the set of rational numbers $\boldsymbol{Q}$ is not order complete. 3+3

3. (a) Show that $\{n^{\frac{1}{n+1}}\}$ converges to 1.
   (b) Find $\lim\limits_{n\to\infty} \left[\frac{1^2}{n^3} + \frac{2^2}{n^3} + \cdots + \frac{n^2}{n^3}\right]$. 3+3

4. (a) State and prove Cauchy's first limit theorem on sequence.
(b) Show that the sequence $\{x_n\}$ where $x_n = 1 + \frac{1}{2!} + \frac{1}{3!} + \cdots + \frac{1}{n!}$ is convergent. 4+2

5. (a) Discuss the statement: Every bounded sequence is a Cauchy sequence.
(b) Suppose $x_n = (1 - \frac{1}{n^2})\sin(\frac{n\pi}{4})$, then find two subsequences of $\{x_n\}$ one of which converges to the $\overline{\lim}\, x_n$ and other converges to the $\underline{\lim}\, x_n$. 3+3

6. (a) Comments on the series $\sum(\sqrt[3]{n^4+1} - \sqrt[3]{n^4-1})$.
(b) Discuss the convergence of the series $\frac{2+k}{3+k} + \frac{2^2+k}{3^2+k} + \frac{2^3+k}{3^3+k} + \cdots$. 3+3

7. (a) If $\sum u_n$ be a convergent series of positive real numbers, then show that $\sum \frac{u_n}{1+u_n}$ is also convergent.
(b) State the alternative form of Gauss test. Discuss the convergence of the series $1 + \frac{\alpha}{1!} + \frac{\alpha(\alpha+1)}{2!} + \frac{\alpha(\alpha+1)(\alpha+2)}{3!} + \cdots$. 2+4

8. (a) State and prove Leibnitz's test.
(b) Discuss the conditional convergence of the series $\sum \frac{(-1)^{n-1}}{(n+1)^2 \log(n+1)}$. 3+3

9. (a) Show that $\lim_{x\to 0} \sin^3\left(\frac{1}{x}\right)$ does not exists.
(b) Let $f : (0, \infty) \to \boldsymbol{R}$ be defined by $f(x) = x(e^{\frac{1}{x^3}} - 1 + \frac{1}{x^3})$. Then find out correct statement(s). (i) $\lim_{x\to\infty} f(x)$ exists. (ii) $\lim_{x\to\infty} xf(x)$ exists. (iii) $\lim_{x\to\infty} x^2 f(x)$ exists. (iv) There exists $m > 0$ such that $\lim_{x\to\infty} x^m f(x)$ does not exists. 3+3

Group – C (Answer any two questions) $2 \times 10 = 20$

10. (a) If f and g be two functions which are continuous at c, then show that $f(x)g(x)$ is continuous at c.
(b) Let $f(x) = x|x| + |x-1|,\ x \in \boldsymbol{R}$. Then which of the following statements is true? (i) f is not differentiable at $x = 0, 1$. (ii) f is differentiable at $x = 0$ but not differentiable at $x = 1$. (iii) f is not differentiable at $x = 0$ but differentiable at $x = 1$. (iv) f is differentiable at $x = 0, 1$.
(c) If $a_0 + \frac{a_1}{2} + \frac{a_2}{3} + ... + \frac{a_n}{n+1} = 0$, where $a_i \in R$, show that the equation $a_0 + a_1x + a_2x^2 + ... + a_nx^n = 0$ has at least one real root between $(0, 1)$. 4+3+3

11. (a) Let a function $f : [0, 1] \to \boldsymbol{R}$ be continuous on $[0, 1]$ and differentiable on $(0, 1)$. If $f(0) = 1$ and $[f(1)]^3 + 2f(1) = 5$, then prove that $\exists$ a $c \in (0, 1)$ such that $f'(c) = \frac{2}{2+3[f(c)]^2}$.
(b) Using Taylor's theorem or otherwise show that $x - \frac{x^3}{3!} < \sin(x) < x - \frac{x^3}{3!} + \frac{x^5}{5!}$, for $x > 0$.
(c) Obtain the power series expansion of $\cos(x)$. 3+3+4

12. (a) Let a function $y = f(x)$ have the values $y_0, y_1, ...y_n$ corresponding to the values of the argument $x_0, x_0 + h, ..., x_0 + nh$, then for any non-negative integer m, show that $\Delta^m y_r = \sum_{i=0}^{m} (-1)^{m-i} \binom{m}{i} y_{r+i}$.
(b) State and prove Gauss's Forward Interpolation Formula.
(c) Find the solution of the difference equation $x(n+1) - 3^n x(n) = 0$. 3+4+3

B.Sc. (Honours) Examination, 2021 Semester-V
Statistics
Course: CC-11
(Stochastic Process and Queuing Theory)

Time: 3 Hours **Full Marks: 40**

Questions are of value as indicated in the margin

Notations have their usual meanings

Tick the appropriate option. $10 \times 1 = 10$

1. Consider a stochastic process(SP)$\{X_t : t \in IN\}$ where $X - t = t.A$ with A∼Uniform(3,7). The process is
   - (i) Discrete state discrete index stochastic process.
   - (ii) Continuous state space discrete index stochastic process.
   - (iii) Continuous state space continuous index stochastic process.
   - (iv) Discrete state space continuous index stochastic process.
2. Consider a stochastic process $\{X_t : t \in \mathbb{N}\}$ where $X - t = t.A$ with A∼Uniform(3,7).
   - (i) process is strong stationary
   - (ii) process is weak stationary
   - (iii) process is neither strong nor weak stationary
   - (iv) process is both strong nor weak stationary
3. State which one is correct
   - (i) A closed set is always irreducible.
   - (ii) An irreducible set is always closed.
   - (iii) A closed set is always absorbing.
   - (iv) In a closed set all the states must have some period.
4. For a simple random walk model
   - (i) all states are persistent for any choice of p.
   - (ii) all states are transient for $p = \frac{1}{2}$.
   - (iii) all states are persistent for $p = \frac{1}{2}$.
   - (iv) all states are transient for any choice of p.
5. Let $X(t)$ be a Poisson process with $\lambda = 2/hr.$. The distribution of $p[X(3) = 4/X(1) = 1]$ is
   - (i) Exponential(2)
   - (ii) Poisson(2)
   - (iii) Binomial(4, $\frac{1}{3}$)
   - (iv) Poisson(3)

6. In a shopping mall customers arrive in a Poisson flow with rate 2/min. Among them 20% come for actual shopping and the rest 80% are just window shopper. What is the expected time for having 10 serious shoppers ?

   (i) 5 minute
   (ii) 25 minute
   (iii) 10 minute
   (iv) 30 minute

7. In an interview, candidates are asked to enter the waiting room according to their roll numbers. But during interview they are called haphazardly. The queue discipline for this example is

   (i) first come first served.
   (ii) last come first served.
   (iii) served in random order.
   (iv) first come last served.

8. For the finite irreducible chain S where $i \in S$ and $j \in S$

   (i) i and j are transient.
   (ii) For state i and state j stationary distribution is same.
   (iii) i and j are non-null recurrent.
   (iv) $f_{ii} = 1$ but $f_{jj} < 1$.

9. Suppose John is in gambling zone. On each successive gambling either he wins Rs.1 or he loss Rs.1 with probability of winning 0.6. John starts playing with Rs.2. What is the probability that John obtains a fortune of Rs.4 without going broke?

   (i) 0.91
   (ii) 0.5
   (iii) 0.36
   (iv) 1

10. A queuing process is derived from

   (i) general birth and death process under steady state.
   (ii) pure birth process under steady state.
   (iii) pure birth and death process under steady state.
   (iv) generalized death process.

Answer in one sentence. $5 \times 2 = 10$

(1) The probability of weather condition (rainy=0, sunny=1) given the weather on preceding day can be represented by a transition matrix

$$\mathbf{P} = \begin{array}{c c} & \begin{array}{cc} 0 & 1 \end{array} \\ \begin{array}{c} 0 \\ 1 \end{array} & \begin{pmatrix} 0.9 & 0.1 \\ 0.5 & 0.5 \end{pmatrix} \end{array}$$

What is the probability that the weather on the day after tomorrow will be rainy?

(2) For the following transition probability matrix with state space $S = \{0, 1\}$ where $p(X = 0) = p(X = 1) = \frac{1}{2}$. Answer whether the following statements are **TRUE/FALSE**.

$$\mathbf{P} = \begin{array}{cc} & \begin{array}{cc} 0 & 1 \end{array} \\ \begin{array}{c} 0 \\ 1 \end{array} & \begin{pmatrix} 0.8 & 0.2 \\ 0.6 & 0.4 \end{pmatrix} \end{array}$$

(a) The transformation matrix is regular.

(b) Steady state distribution does not exist.

(3) Let $N(t)$ be the number of arrivals in a park with rate $\lambda = 30/hr.$. What is the probability that by $8.15 a.m.$ the park has 10 visitors given that in first half an hour only 25 visitors were there.

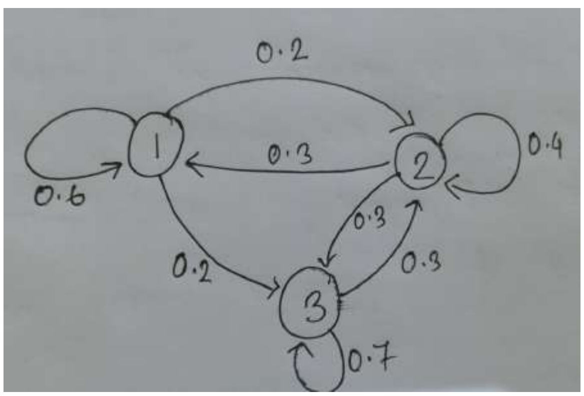
(4) Let $N(t)$ be the Poisson process with intensity $\lambda = 2$. Let T_i be the i^{th} interval time. You start watching the process at time $t = 10$minute. T_1 is the first arrival after $t = 10$. Fine $E(T_1)$.

(5) Let the variable be the number of cars waiting in a queue in front of petrol pump while the queue can hold at max n cars. Write the state space and index set.

Answer briefly. $4 \times 5 = 20$

(1) Define strict stationarity. Give an example of strict stationarity. Justify why it is strict. $2 + 1 + 2$

(2) The following transition graph shows the mood of a person-{cheerful (1), so-so (2), sad (3)}. $1 + 2 + 2$



Answer the following.

(a) How many classes are in the chain?

(b) Is it irreducible chain?

(c) What is the steady state probability for this process?

(3) State the necessary assumptions of Yule-Furry birth process. Next, deduce the model equation of Yule-Furry process, starting from the differential equation of generalized birth process. $2 + 3$

(4) Deduce the expression of expected number of customers in a system and expected number of customers in a queue for a single server queuing process where service rate is thrice of arrival rate.

B.Sc. (Honours) Examination, 2020
Semester-V
Statistics
Course: CC-12A (Statistical Computing using C/C++ Programming (Theory))
Full Marks: 40 Time: 3 Hours

(Question 1 is compulsory. Answer any five from the rest.)

1. State whether the following statements are TRUE or FALSE. 10

(a) = is used for comparison, whereas, == is used for assignment of two quantities.
(b) Blank spaces may be inserted between two words to improve the readability of the statement.
(c) ferror( ) reports any error that might have occurred during a read/write operation on a file.
(d) A file opened for writing, if already exists, its contents would be overwritten.
(e) Suppose that x is a one dimensional array, then *(x + n) is same as x[n] +1
(f) In c programming a function can return multiple values.
(g) The given expression P– = Q+1 is same as P=P – Q + 1
(h) The default parameter passing mechanism is called as call by value.
(i) The C language is a context free language.
(j) In C, keywords have predefined meanings and user can change its value at compile time.

2. Write down the outputs of the following codes, with proper explanations. 3+3

(a)

```
int main()
{
int a=5;
printf("%d %d %d", a ,a++, ++a);
return 0;
}
```

(b)

```
int main()
{
int x=-5,y,num=5;
y=x%-4;
y=(y?0:num*num);
```

```
printf("%d",y);
}
```

3. Write a program in C to print the first 25 prime numbers. 6

4. Write a program in C to print the following triangle of numbers. 6

```
5
4   5
3   4   5
2   3   4   5
1   2   3   4   5
```

5. Define pointer. What are the benefits of using pointers in programming? Can two pointer variables be added? 1+4+1

6. Give examples of functions with 3+3
(a) no argument but one return value.
(b) arguments but no return value.

7. Write a short note on the use of the following library functions. 6
(a) `rand()` (b) `strrev()` (c) `atoi()`

8. Find the errors in the following C codes.Assume that all variables have been declared. 6
(a)

```
while(count!=10);
{
count=1;
sum=sum+x;
count=count+1;
}
```

(b)

```
for(p=10;p>10;)
p=p-1;
printf(“%f”,p);
```

(c)

```
p1=&m;
p2=n;
*p1=&n;
p2=&*&m;
```

```
m=p2-p1;
p1=&p2;
m=*p1+*p2++;
```
(Here `m` and `n` are integers and `p1` and `p2` are pointers to integers.)

B.Sc. (Honours) Examination, 2020
Semester-V
Subject: Statistics (Practical)
Paper: CC-12B (Statistical Computing using C/C++ Programming (Practical))
Full Marks: 20 Time: 2 Hours

(Answer all questions. Save your program files and output files in the format Roll No._Prob no.cpp/dat e.g. BSC-Sem-5-Stat-01_Prob 1.cpp)

1. Write a program in C to draw 25 samples from a Binomial distribution with parameters n=20 and p=0.72. Save your output in a file. 10

2. Write a program in C to find the inverse of the following matrix. 10

$$\begin{pmatrix} 5 & 1 & 2 \\ 3 & 0 & 1 \\ -1 & 2 & 1 \end{pmatrix}$$

B.Sc. (Honours) Examination, 2020
Semester-V
Statistics
Course: DSE-2A
(Demography and Vital Statistics)
Time: Three Hours   Full Marks: 40

Questions are of value as indicated in the margin
Notations have their usual meanings

Answer **any five** questions

1. Explain the need of standardization in comparing death rates. Describe the methods of direct and indirect standardization of death rates. 2+6

2. Discuss Gross Reproduction Rate (GRR) and Net Reproduction Rate (NRR) and interpret the situations when for a society
   (i) GRR=NRR and (ii) NRR=1 3+3+2

3. With appropriate assumption and usual notation, show that

$$ {}_n q_x = \frac{2n._n m_x}{2 + n._n m_x} . $$ 8

4. Write short notes on (i) Whipple's index, and (ii) Age dependency ratio. 5+3

5. Define CBR, GFR and ASFR, and indicate why each is considered an improvement on the preceding measure of fertility. 2+3+3

6. Describe the structure of a complete life table. Explain how different columns of a complete life table may be computed based on the observed age specific mortality rates. 4+4

7. Describe few sources of raw data in Demography. Distinguish between errors of coverage and errors of response in connection with errors in census and registration data. 2+6

B.Sc. (Honours) Examination, 2020
Semester-V
Statistics
Course: DSE-2B (Practical)
(Demography and Vital Statistics)
Time: Two Hours Full Marks: 20

Questions are of value as indicated in the margin
Notations have their usual meanings

1. Complete the following life table:

Age(x)	l_x	d_x	$1000\,q_x$	L_x	T_x	e_x^0
25	78046					
26	77614	400				
27						
28	76723		6.06			
29						
30	75523				2750493	

Hence determine the probability that a person of age 25 lbd will die before reaching age 30 lbd. 4+2

2. With the help of the following data, determine the crude death rate and the age-specific death rates, separately for males and females. 2+6

Age	Population(000)		Number of deaths	
	Male	Female	Male	Female
0	29.8	28.5	807	609
1-4	109.3	104.9	192	138
5-9	126.1	120.7	88	65
10-19	198.2	189.7	182	82
20-29	150.8	142.7	247	117
30-39	156.9	151.0	284	203
40-49	139.5	138.3	565	425
50-59	110.0	106.7	1230	746
60-69	70.1	80.9	2083	1464
70-79	45.4	54.5	3308	2650
80-	13.7	18.1	2195	2621
Total	1149.8	1136.8	11181	9120

3. Compute a summary index of age preference of the following table in ending by “0” using Whipple index in the range 15-66. Also calculate the age dependency ratio. 4+2

Age(in years)	Number	Age(in years)	Number	Age(in years)	Number
10	841356	37	242462	64	34381
11	581400	38	316210	65	102440
12	796786	39	225207	66	26445
13	619293	40	434156	67	35311
14	596592	41	128632	68	40711
15	565714	42	217881	69	20921
16	566942	43	169167	70	136771
17	538891	44	151142	71	13000
18	651318	45	319118	72	28017
19	491441	46	160329	73	16662
20	565801	47	160855	74	14490
21	494895	48	237287	75	50558
22	515823	49	155094	76	15010
23	456892	50	313636	77	11878
24	425212	51	78534	78	23353
25	522203	52	128935	79	9212
26	358549	53	93279	80	73791
27	376221	54	95715	81	5532
28	395766	55	163093	82	9331
29	300610	56	87754	83	5653
30	535924	57	71828	84	5089
31	333086	58	93049	85	18604
32	318481	59	72206	86	4803
33	246260	60	275436	87	5617
34	233700	61	31299	88	4388
35	401936	62	49634	89	4000
36	242659	63	40154		

M.Sc. Examination, 2020
Semester-I
Statistics
Course: MSC-11
(Linear Models and Distribution Theory)
Time: 3 Hours Full Marks: 40

Questions are of value as indicated in the margin
Notations have their usual meanings

Answer **any four** questions

1. (a). State and proof Gauss-Markov theorem for following linear model
$$\boldsymbol{y} = \boldsymbol{x\beta} + \boldsymbol{\epsilon}$$
(symbols having their usual meanings)
Also show that for any linear parametric function $\boldsymbol{\lambda}'\boldsymbol{\beta}$ condition of uniqueness and estimability is the same.

6+4

2. (a). Define estimation space and error space. Show that the covariance between any linear function belonging to the error space and any BLUE is zero.
(b). Consider the following linear model:
$$y_1 = \beta_1 + \beta_2 + \epsilon_1$$
$$y_1 = \beta_1 + \beta_3 + \epsilon_2$$
$$y_3 = \beta_1 + \beta_2 + \epsilon_3$$
Is $5\beta_1 + 2\beta_2 + 3\beta_3$ estimable?

7+3

3. Show that for the linear model $\boldsymbol{y} = \boldsymbol{x\beta} + \boldsymbol{\epsilon}, \boldsymbol{\epsilon} \sim \boldsymbol{N}(\boldsymbol{0}, \boldsymbol{\sigma^2 I})$
$$\frac{\frac{(\boldsymbol{\Lambda}\widehat{\boldsymbol{\beta}} - \boldsymbol{\Lambda\beta})'(\Lambda S^{-}\Lambda')^{-1}(\boldsymbol{\Lambda}\widehat{\boldsymbol{\beta}} - \boldsymbol{\Lambda\beta})}{m}}{\left(\frac{SSE}{n-r}\right)} \sim F_{m,n-r}$$
where$\boldsymbol{\Lambda}^{m\times p}$ is of rank m, $\boldsymbol{S}^{-}$ is a generalized inverse of$(\boldsymbol{x}'\boldsymbol{x})$. Discuss the case when $m = 1$. (You need to prove all the results to be used) 8+2

4. (a) Let $X_i \sim N(\mu_i, 1), i = 1..n$ and they are independently distributed. Find the distribution of $\sum_{i=1}^{n} X_i^2$. Hence or otherwise find the mgf of the distribution.
(b) If $\boldsymbol{M} \sim \boldsymbol{W}_p(\boldsymbol{\Sigma}, m)$ and $\boldsymbol{B}$ is a $p \times q$ matrix, then show that
$$\boldsymbol{B}'\boldsymbol{MB} \sim W_q(\boldsymbol{B}'\boldsymbol{\Sigma B}, m)$$ 8+2

5. (a) If $\boldsymbol{M} \sim W_p(\boldsymbol{\Sigma}, m), m > p$ then show that the ratio

$\frac{\boldsymbol{a}'\boldsymbol{\Sigma}\boldsymbol{a}}{\boldsymbol{a}'\boldsymbol{M}^{-1}\boldsymbol{a}}$ has the χ^2_{m-p+1} distribution for any fixed p-vector $\boldsymbol{a}$

(b).Prove that the BLUE of any linear combination of estimable parametric function is the linear combination of their BLUEs. 7+3

6. (a) Show that error sum of squares is distributed independently of the BLUE of any estimable function.

(b). Show that for the liner model $\boldsymbol{y} = \boldsymbol{x\beta} + \boldsymbol{\epsilon}$ a necessary and sufficient condition for a linear parametric function $\boldsymbol{\lambda}'\boldsymbol{\beta}$ to be estimable is that $\boldsymbol{\lambda}'$ is a linear combination of rows of $\boldsymbol{x}'\boldsymbol{x}$. 5+5

B.Sc. (Honours) Examination, 2020
Semester-I
Statistics
Course: MSC-12
(Real Analysis and Measure Theory)
Time: 3 Hours Full Marks: 40

Questions are of value as indicated in the margin
Notations have their usual meanings

Answer any four questions $4\times 10 = 40$

1. Let $\{A_n\}$ be a sequence of subsets of Ω. Find the $\limsup\{A_n\}$ and $\liminf\{A_n\}$ for the following
i) $A_n = [0, \frac{n}{n+1})$, ii) $A_n = [1, 2-\frac{1}{n}]$, iii) $A_n = [1, 3+\frac{1}{n})$, iv) $A_n = C$ if n is odd, $A_n = D$ if n is even. 10

2. State and prove the continuity theorems for additive set function. 10

3. State and prove dominated convergence theorem. 10

4. Let ϕ, F be the characteristic function, distribution function of the random variable X. If $\int |\phi(t)|\, dt < \infty$ and F is absolutely continuous. Then show that the pdf of X can be written as: $f(x) = \frac{1}{2\pi}\int e^{-itx}\phi(t)\, dt$ 10

5. (a) Let S be a subset of $\boldsymbol{R}$ such that every infinite open cover of S has a finite subcover. Prove that S is closed and bounded.
(b) Exhibit an open cover of the set $\{\frac{1}{2n} : n \in \mathbf{N}\}$ that has no finite subcover. Is the set compact? 5+5

6. (a) Prove that the union of a finite number of open sets is an open set. Does this holds for an arbitrary number of open sets.
(b)Find the radius of convergence of the power series $\sum_{n=0}^{\infty} \frac{5^n}{n^2+1} x^n$ 5+5

M.Sc. Examination, 2021
Semester-I
Statistics
Course: MSC-13
(Statistical Inference I)

Time: 3 Hours **Full Marks: 40**

Questions are of value as indicated in the margin
Notations have their usual meanings

Answer **all** questions

1. State whether the following statements are **True/False** by writing T/F by the questions. $5 \times 1 = 5$
   (a) As n gets larger, the lower bound for $V(T(X))$ gets smaller where $T(X)$ being the unbiased estimator of θ.
   (b) Maximum likelihood estimate is always a function of minimal sufficient statistic.
   (c) A constant statistic $T = 12$ can be looked as an ancillary statistic for parametric function θ.
   (d) A set of order statistics $X_{(1)}, X_{(2)}, ..., X_{(n)}$ for any distribution $f(x, \theta)$ is always minimal sufficient.
   (e) A minimum variance unbiased estimator might not be an U statistic.

2. Choose out the correct alternative. $5 \times 1 = 5$
   (a) Let $Y_1, Y_2, ..., Y_n$ be a random sample from Geometric(p) with parameter p. The MLE for p is
      (i) $\dfrac{1}{n + \bar{Y}}$.
      (ii) $\dfrac{Y_i}{1 + Y_i}$.
      (iii) $\dfrac{\bar{Y}}{n + \bar{Y}}$.
      (iv) $\dfrac{\sum_{i=1}^{n} Y_i}{n + \sum_{i=1}^{n} Y_i}$.
   (b) Which of the following <u>is</u> correct?
      (i) Completeness implies that two different functions of a statistic T cannot have the same expectation.
      (ii) Sufficiency of a statistic ensures completeness of that statistic.
      (iii) Independence of ancillary statistic and sufficient statistic holds even when the family is <u>not</u> complete.
      (iv) Non-completeness of a statistic might ensure the presence of minimum variance unbiased estimator.
   (c) If T is sufficient for θ, $\frac{\delta \log f(x_1, x_2, .., \theta)}{\delta \theta}$ will be always
      (i) function of t and θ only.
      (ii) function of X only.
      (iii) can't be said.
      (iv) sometimes a function of t and θ only.

(d) Let X be a random variable with p.d.f.

$$f(x, \theta) = \frac{1}{2}, \qquad \theta \le x \le \theta + 2.$$

Sufficient statistic for θ is

(i) $X_{(1)}$
(ii) $X_{(n)}$
(iii) $\frac{X_{(1)}+X_{(n)}}{2}$
(iv) No sufficient exists.

(e) Let $X_{(1)}, X_{(2)}, ..., X_{(n)}$ be the order statistics of a distribution having p.d.f. $f(X-\theta), \theta \in \mathbb{R}$, then

(i) $X_{(n)} - 2$ is ancillary for θ.
(ii) $X_{(n)} - X_{(1)}$ is ancillary for θ.
(iii) $\frac{X_{(n)}}{X_{(1)}}$ is ancillary.
(iv) No ancillary statistic exists.

3. Answer any <u>five</u> questions. $5 \times 2 = 10$

(a) Define minimal sufficient statistic for a parameter θ.
(b) For the distribution $f(x) = e^{-(x-\lambda)}; x > \lambda$, find the Fisher's information for λ based on a single observation.
(c) Suggest a symmetric kernel for $\theta^2 + 2\theta$ based on two variables.
(d) For the distribution of Poisson(θ), suggest an MLE of $e^{-\theta}$.
(e) Define U statistic.
(f) Let X_1, X_2 be two random variables from Poisson(λ). Find a UMVUE of $e^{-\lambda}(1+\lambda)$.

4. Answer any <u>four</u> of the following. $4 \times 5 = 20$

(a) Show that for a U statistic of θ, the asymptotic variance is $m^2\xi_1 + o(\frac{1}{n})$ where $\xi_1 = E_{X_2,...X_n/X_1}[f(X_1, X_2, .., X_n)]$ and $f(X_1, X_2, .., X_n)$ being the kernel of θ.
(b) Let $X_1, X_2, ..., X_n$be i.i.d. $N(\mu, 1)$. Suppose that $\Theta = \{\mu \ge 0\}$. Find the MLE of μ.
(c) Propose a U statistic for $P(X_i, X_j > 0)$, where X_i, X_j are two random observations coming from a continuous distribution $f_\theta(x)$. Find the variance of your proposed U-statistic.
(d) Show that Fisher's information for θ is

$$I(\theta) = -\int [\frac{\delta^2}{\delta\theta^2} \log L(\theta, y)].L(\theta, y)dy.$$

Where $L(\theta, y)$ being the likelihood function.

(e) State and prove Rao-Blackwell theorem.
(f) Suppose $X_1, X_2 \sim f(x;\theta) = \frac{1}{\theta}e^{-\frac{x}{\theta}}, 0 < X < \infty$. Find a complete sufficient statistic for θ. Also show that $\frac{X_1}{X_1+X_2}$ is independent of that complete sufficient statistic.

M.Sc. Examination, 2020
Semester-I
Statistics
Course: MSC-14 (Sample Survey)
Full Marks: 40 Time: 3 Hours

(Answer any four questions.)

1. (a) Explain the problem you will encounter while asking people directly about their belongings to a dichotomous sensitive characteristic.

   (b) Discuss how will you estimate the proportion of individuals belonging to different political parties of West Bengal using unrelated questionnaire method, where the unrelated questions have 'Yes/No' type response.

   4+6=10

2. (a) Explain the role of auxiliary information in sample survey.

   (b) Describe how the double sampling technique can be applied for the regression method of estimation of the population total.

   (c) Find the approximate expressions for the bias and MSE of the estimator of the population total under this sampling scheme. Discuss its efficiency over uni-phase sampling.

   2+3+5=10

3. Find the expressions of the expectation and variance of the effective size of the sample obtained using (i) a SRSWR(n,N) design and (ii) a PPSWR(n,N) design. You need to prove all necessary results.

   10

4. (a) State and prove a necessary and sufficient condition for the existence of an unbiased estimator of the population total.

   (b) Show that among the class of HLUEs, no one exists with uniformly minimum variance.

   5+5=10

5. (a) Define Desh Raj's estimator . Show that it is unbiased for the population total.

   (b) Find the unbiased estimator of its variance.

   (c) Mention a drawback of the estimator and the way of improvement.

   3+4+3=10

6. 'Instead of applying PPSWR directly to draw samples from a population, if we classify the observations through SRSWOR and then select one unit from each group by PPSWR; there will be a gain in the efficiency.'– justify.

   10

M.Sc. Examination, 2020
Semester-I
Statistics
Course: MSC-15 (Practical)
(Linear Models and Statistical Distribution Theory)
Time: 3 Hours Full Marks: 40

Questions are of value as indicated in the margin
Notations have their usual meanings

1. Let $X_1, X_2, \ldots, X_n$ be a set of iid random sample observations from $N(\mu, \sigma^2)$ where $\sigma = 2$. What would be the value of n if it allows the maximum likelihood (ML) estimation process to generate estimate of μ within 0.01 distance of the true parameter value? Will the sample size increase if the ML estimate lies within 0.05 distance of the true parameter value? 3+2

2. There are four objects w_1, w_2, w_3, w_4 whose weights are to be determined.

Left pan	Right pan	Weight needed for equilibrium
w_1, w_2, w_3, w_4	-------------	20
w_1, w_2	w_3, w_4	10
w_1, w_3	w_2, w_4	5
w_1, w_4	w_2, w_3	1

(a). Obtain best estimate of all weights

(b). Find their dispersion matrix of the estimators and estimate of the variance of each measurement. 6+4

3. (a) Consider the following regression model

$$y_1 = \beta_1 - 3\beta_2 + 4\beta_3 + \epsilon_1$$
$$y_2 = \beta_1 - 2\beta_2 + \beta_3 + 3\beta_4 + \epsilon_2$$
$$y_3 = \beta_1 - \beta_2 + 2\beta_4 + \epsilon_3$$
$$y_4 = \beta_1 + \beta_3 + \beta_4 + \epsilon_4$$
$$y_5 = \beta_1 + \beta_2 + \epsilon\_5$$

Estimate the following linear parametric functions (if estimable)

$2\beta_1 + \beta_2 + \beta_4$, $2\beta_1 - \beta_2 - \beta_3$, $\beta_1 + \beta_2$, $\beta_1 + \beta_4$ and β_4 where $y_1 = 11, y_2 = 21, y_3 = 13, y_4 = 45, y_5 = 50$.

Also find their standard errors. 10

4. Suppose that the two observations on each of three treatments are as follows

Treatments

t_1	t_2	t_3
8	5	12
6	3	14

$y_{11} = \mu + t_1 + e_{11}$, $y_{12} = \mu + t_1 + e_{12}$, $y_{21} = \mu + t_2 + e_{21}$, $y_{22} = \mu + t_2 + e_{22}$, $y_{31} = \mu + t_3 + e_{31}$, $y_{32} = \mu + t_3 + e_{32}$

Find the best linear estimator for the following (if estimable):
$\frac{1}{2}(t_1 + t_2) - t_3$, $\mu + \frac{t_1+t_2+t_3}{3}$, $\mu + t_1$, $t_2 - t_3$, μ, $t_1 + t_2$ 10

5. For a linear model, the normal equations are:

$$10\,\beta_1 - 2\beta_2 - 8\beta_3 = 12$$
$$-2\beta_1 + 5\beta_2 - 3\beta_3 = 16$$
$$-8\beta_1 - 3\beta_2 + 11\beta_3 = -28$$

(a). Obtain any solution of the normal equations.
(b). Find the maximum number of linearly independent estimable parametric functions.

3+2

M.Sc. Examination, 2020
Semester-I
Statistics (Practical)
Course: MSC-16 (Practical on Sample Survey)
Full Marks: 40 Time: 4 Hours
(Answer all questions.)

1. Following is a sampling frame regarding study of household size:

Village	Area (km^2)	Previous census population	Number of households	Size of households
1	8.7	69	17	7,5,5,4,6,2,3,5,5,6,5,4,4,4,5,3,3
2	10.6	82	18	6,5,4,5,4,5,6,5,3,5,4,4,5,3,3,5,6,4
3	15.0	110	26	6,6,3,5,3,4,5,5,4,4,4,3,7,5,4,6,2,5,5,6,1,5,5,4,6,3
4	6.2	80	18	6,3,6,3,6,3,4,5,4,4,4,5,6,3,5,1,3,5
5	9.6	92	24	5,4,6,5,4,5,6,5,4,4,7,6,6,5,4,4,5,6,3,4,3,3,5,3
6	7.3	65	17	3,4,4,6,5,7,3,5,4,6,4,5,4,5,3,3,6
7	4.5	72	20	6,4,4,5,4,5,6,4,3,5,4,6,5,5,2,2,4,5,4,3
8	10.6	108	24	5,3,3,7,4,4,6,6,4,5,3,7,6,4,5,6,3,5,1,3,5,4,4,6
9	5.4	106	24	5,3,3,7,4,4,6,6,4,5,3,7,5,6,6,3,5,2,7,5,4,3,1,6
10	3.5	80	22	4,5,6,5,4,4,7,6,6,5,4,4,5,6,3,4,3,3,5,3,5,4
11	5.8	72	15	5,4,5,6,5,4,4,7,6,6,5,4,4,5,6

(i) Take a sample of 3 villages using SRSWOR and select 5 households from each villages using SRSWOR.
(ii) Estimate the estimate of average size per household and the estimate of the variance of the estimator.
(iii) How can an estimate obtained from a simple random sample of 15 plots be compared with the estimate obtained in (ii)?
(iv) Select 3 distinct villages by Lahiri's method and find an estimate of the average household size under this sampling scheme.

3+(3+4)+3+(3+4)=20

2. The following is the population of M.Sc. (Statistics) students of a certain univer- sity, together with the data on two variables, viz. Marks in Sample survey and Marks in Estimation (both out of 50).

Serial Number	Student Name	Marks in Sample Survey	Marks in Estimation
1	Ajoy	30	45
2	Bikram	24	32
3	Buddhadev	40	47
4	Anushree	36	40
5	Payel	31	25
6	Ramen	40	30
7	Ritika	34	29
8	Rima	28	43
9	Somnath	25	27
10	Riddhi	31	44
11	Sulagna	43	31
12	Surya	48	46
13	Dipanjan	32	34
14	Bishal	43	49
15	Sudip	27	39
16	Molay	40	22
17	Arnoneel	34	28
18	Arindam	36	33

(a) Select 3 students (distinct) from this list using simple random sampling. Estimate the mean marks of all students in Sample Survey, using their marks in Estimation as auxiliary information. Provide an estimate of the variance of your estimator.

(b) Next select 3 students (distinct) from this list using PPS sampling and estimate the mean marks of all students in Sample Survey. Provide an estimate of variance of your estimator.

(c) Now form 3 groups, each containing 6 students using SRSWOR. From each group, select one student by PPS sampling. Under this scheme, provide an estimate of the mean marks of all students in Sample Survey.

(d) Compare the estimate obtained in these three cases with the true mean.

6 + 6 +5 + 3 = 20

M.Sc. Examination, 2020
Semester-I
Statistics
Course: MSC-31
(Measure Theory and Real Analysis)
Time: 3 Hours Full Marks: 40

Questions are of value as indicated in the margin
Notations have their usual meanings

Answer **any four** questions

1. (a). Let $\{A_n\}$ be a sequence of subsets of Ω. Find the lim sup $\{A_n\}$ and lim inf $\{A_n\}$ for the following
   $(i) A_n = \left[0, \frac{n}{n+1}\right)$ (ii). $A_n = \begin{cases} C \text{ if } n \text{ is odd} \\ D \text{ if } n \text{ is even} \end{cases}$ $(iii). [0, 1 - \frac{1}{n}]$ (iv). $\left[0, 1 + \frac{1}{n}\right)$

   (b). Define σ −field. Give an example to show that all fields are not σ −field.

Show that a σ −field is a monotone field. Is the converse true? 4+6

2. (a). Show that any non-negative random variable can be expressed as a limit of a sequence of non-negative, non-decreasing simple random variables.

   (b). Are all additive set functions σ −additive? 8+2

3. (a) Using monotone convergence theorem, state and prove Fatou's lemma. Mention the strict inequality case in this regard.

   (b). If f and g are two integrable functions, show that $f \pm g$ is also integrable and
$$\int (f \pm g) d\mu = \int f\, d\mu \pm \int g\, d\mu$$

4. Define $f_n : [0, \infty) \to R$ by
$$f_n(x) = \frac{sin(nx)}{1 + nx}$$

(a) Show that f_n converges pointwise on $[0, \infty)$ and find the pointwise $limit\ f$.

(b) Show that $f_n \to f$ uniformly on $[a, \infty)$ for every $a > 0$.

(c) Show that fn does not converge uniformly to f on $[0, \infty)$.

5. (a) Show that a sequence of functions is convergent if and only if it satisfies Cauchy criterion.

(b) Prove that if a sequence $\{u_n\}$ converges to l the every subsequence of $\{u_n\}$ also converges to l.

(c) Prove that $\lim \left(1 + \frac{1}{2n}\right)^n = \sqrt{e}$ 5+3+2

6. (a) State and Prove Borel 0-1 law.

(b). Show that a necessary and sufficient condition for the convergence of a sequence $\{u_n\}$is that for a pre-assigned positive ϵ, there exists a natural number k such that

$$\left|u_{n+p} - u_n\right| < \epsilon \; \forall \; n \geq k \text{ and for } p = 1,2,3, \dots \dots$$ 5+5

M.Sc. Examination, 2020
Semester-III
Statistics
Course: MSC-32 (Categorical Data Analysis and Advanced Data Analysis Techniques)
Full Marks: 40 Time: 3 Hours

(Answer any four questions.)

Answer any four questions.

1. (a) Define Yule's Q-statistic for a 2×2 contingency table. Prove that it equals Goodman and Kruskal's Gamma for 2×2 tables. Find the limits of the measure and interpret the marginal cases.

   (b) Show that for multinomial sampling, the asymptotic variance of $\sqrt{n}\left(\widehat{Q} - Q\right)$ is $\sum_i \sum_j \pi_{ij}{}^{-1}\left(1 - Q^2\right)^2/4$, where $\widehat{Q}$ is the sample analogue of Q-statistic. (You need to prove the necessary results.)

   1+2+2+5=10

2. (a) Explain with suitable examples what do you mean by prospective and retrospective studies. Compare their relative advantages and disadvantages.

   (b) Explain why Odd's ratio is an appropriate measure in each of these two types of study.

   (c) For counts $\{n_i\}$, define the power divergence statistic for testing goodness of fit and find its limits as the power parameter λ tends to 0 and -1 respectively.

   (2+3)+2+3=10

3. (a) For a Hierarchical log-linear model with 3 variables X, Y and Z, define joint,mutual and conditional independence. Prove that joint independence implies conditional independence. Define an appropriate test statistic for testing conditional independence and state its distribution.

   (b) Interpret the parameters of the logistic regression model.

   (3+2+2)+3=10

4. (a) Explain how EM algorithm can be used to cluster data which can be modeled as a mixture of univariate normal populations.

   10

5. (a) Geometrically interpret bootstrap and jackknife resampling technique. What is the relationship between them?

   (b) Find the bootstrap and jackknife estimate of bias and standard error of sample variance.

   4+6=10

6. (a) Compare importance sampling with accept-reject method.

   (b) Prove the ascent property of EM algorithm.

   (c) Illustrate Gibbs sampling technique with an example.

   3+4+3=10

M.Sc. Examination, 2020
Semester-III
Statistics
Course: MSC-33(MSE-1)
(Operations Research and Optimization Techniques)
Time: Three Hours Full Marks: 40

Questions are of value as indicated in the margin
Notations have their usual meanings

Answer **any five** questions

1. What is an assignment problem? How it is different from traveling salesman problem? Show that if a constant be added to any row and/or any column of the cost matrix of an assignment problem, then the resulting assignment problem has the same optimal solution as the original problem. 2.5+2.5+3

2. For the M/M/c queuing system find the expected number of customers in the system in the steady state and also the expected queue length. Find the cumulative distribution function for the waiting time of a customer who has to wait in a M/M/c queuing system. 4+4

3. Briefly describe the phases in Operations Research study. What do you mean by models in OR? Discuss the different types of models that are usually encountered in OR. 2+2+4

4. What is two-person zero-sum game? Transform this game to a Linear Programming Problem. Prove that the value of a two-person zero-sum game is unique. 2+2+4

5. (a) Define an inventory. What are the advantages and disadvantages of having inventories?
(b) In a manufacturing situation the production is instantaneous and demand is D units per year. If no shortages are allowed, show that the optimal manufacturing quantity per run is $q=\sqrt{2C_p D/C_h(1-D/K)}$, where C_p =set-up cost per run, C_h =holding cost per unit per year, K =manufacturing rate per unit of time ($K>R$). 3+5

6. What is replacement? Describe some important replacement situations. Discuss replacement policy of equipment that deteriorates gradually with (i) no change in time value of money and (ii) change in time value of money. 1+2+(2+3)

7. Write short notes on any two of the following: 4+4

(a)Duality problem in LPP
(b)Graphical solution of game problem
(c)Queue discipline
(d)(s, S) inventory policy

MSc Semester -III Examination 2020

Subject- Statistics

Paper- MSC34 (Time Series Analysis)

Full Marks: 40 Time: 3 hours

Answer any four questions: (Notations have usual meanings)

1. a) When will you say a time series is stationary? 4

b) Let X_t=YCosθt + Z Sinθt, where Y, Z are two uncorrelated random variables each with 0 mean and variance unity and θ ε [-π, π]. Comment on the stationarity of the series. 6

2. a) Define auto-covariance and auto-correlation functions. 3

b) State and prove the properties of auto-corrleation function. 7

3. a) Examine the stationarity of the series y_t-1.3y_{t-1}+0.42y_{t-2}=e_t, where et is white noise with mean 0 and variance σ^2. Also find the value of ρ_1. 7

b) Find the AR representation of the MA (1) process y_t=(1-0.4L)e_t 3

4. a) Explain the exponential smoothing method of forecasting a time series. 4

b) When do we use single, double, and triple exponential smoothing techniques? Derive the exponential smoothing formula when the time series has a trend with constant increatsing growth rate. 6

5. Describe how will you estimate the hidden periodicity of a time series. 10

6. a) Describe the necessity of frequency domain analysis of time series. Find the power spectrum of the AR (1) process. 6

b) Justify the name White Noise with respect to frequency domain analysis. 4

M.Sc. Examination 2020
Semester III
Statistics (Practical)
Paper: MSC-35

Full Marks: 40 Time: 4 Hrs.

1. Draw a random sample of suitable size from the Beta (5, 7) distribution using acceptance-rejection sampling (Take your proposal density to be $U(0, 1)$). Construct the histogram of the sample and fit a Beta distribution using the same sample. Draw the actual density and the fitted density over the histogram (on the same graph). 6

2. The following are the marks secured by two batches of salesmen in the final test after completion of training.

   Batch A: 26, 27, 31, 26, 19, 21, 20, 25, 30
   Batch B: 23, 28, 26, 24, 22, 19.

   Use permutation test to decide whether there is any significant difference between the mean marks of these two batches. 4

3. Consider the following sample from a Poisson distribution.

   $7, 5, 6, 7, 3, 5, 7, 0, 9, 4$

   Based on the sample, find the bootstrap and jackknife estimate of the standard error of (i) sample mode, (ii) sample median, (iii) sample quartile deviation, (iv) sample range , (v) sample Gini's mean difference and (vi) sample standard deviation. Draw histograms of the replications in each case.

   6+2=8

4. **Instructions:**

   - Please find an MS-Excel file named "$msc - 35 - 2020 - problem4.xlsx$" in the attachment.
   - It contains the natural tip length of 12 different fishes taken from a pond, over 12 time points, together with the amount of dissolved oxygen present in the pond over the time points.
   - Denote the length of the ith fish at time point t by $Y_{it}, i = 1, 2, \dots 12, t = 1, 2, \dots 12$.
   - Denote the amount of dissolved oxygen present at time t by Z_t.
   - Define the relative growth rate of fish i at time point t as $R_{it} = \ln\left(\frac{Y_{i,t+1}}{Y_{it}}\right), i = 1, 2, \dots 12, t = 1, 2, \dots 11$.

   (a) Plot the values of R_t against Z_t values. Fit an appropriate curve $R_t = \phi(Z_t) + \epsilon_t$, based on the data points. Draw the fitted curve over the observed data points.

   (b) Find the bootstrap estimate of the standard error of the model parameters. Take the number of bootstrap replications (B) to be 50.

   (c) Draw the histograms of the bootstrap replications of each of the parameters. Also check graphically whether they can be regraded to come from a Normal distribution.

   5+6+4=15

5. There are five regions in a state and a cellular phone company offers 4 plans in some of the regions and 5 plans (The previous 4 plus one additional plan) in some regions. The customers have to select

the best plan out of 4 in some regions and out of 5 in some regions. The following data represents the number of customers opting various plans. Based on the data, find the best plan for the customers.

Plan / Area	1	2	3	4	5
1	26	63	20	0	-
2	31	14	16	51	-
3	41	28	34	10	-
4	24	25	5	34	13
5	23	71	10	0	18

7

x———— Best of Luck————x

M.Sc Semester III Examination, 2020

Statistics

MSC-36(Practical)

Time: Four Hours Full Marks: 40

Write answer of each group in separate answer sheet

Group-A (Full Marks: 20, Send Answer Script to sudhansu.maiti@gmail.com)

1. Use the concept of dominance to convert the following game into 2x2 game and hence solve the game. 7

	Player B			
	I	II	III	IV
I	2	1	4	0
II	3	4	2	4
III	4	2	4	0
IV	0	4	0	8

2. Find out the optimal assignment and minimum cost for the assignment with the following cost matrix.

	I	II	III	IV	V
A	160	130	175	190	200
B	135	120	130	160	175
C	140	110	155	170	185
D	50	50	80	80	110
E	55	35	70	80	105

7

3. A supper market has two girls ringing up sales at the counters. If the service time for each counter is exponential with mean 4 minutes, and if people arrive in a Poisson fashion at the rate of 10 an hour, (a) what is the probability of having to wait for service? (b) What is the expected percentage of idle time for each girl? (c) Find the average queue length and the average number of units in the system. 3

4. A contractor has to supply 10,000 bearings per day to an automobile manufacturer. He finds that, when he starts a production run, he can produce 25,000 bearings per day. The cost of holding a bearing in stock for one year is 20 paise and the set up cost of a production run is Rs. 180. How frequently should production run be made? 3

[Turn over]

Group-B (Full Marks: 20)

Statistical Package R may be used. Candidates are not allowed to use the auto-arima function for fitting and forecasting Time Series models. Explicit codes along with outputs should be provided in the answer script. Submit answer-script at onlinebidyacharcha.net

5. The following table gives the data on certain process; observations are being taken at every two hours interval. Plot the series and comment on its stationarity. Calculate and plot first 7 autocorrelations. 5

t	y_t	t	y_t	t	y_t	t	y_t	t	y_t
1	17.0	11	16.7	21	17.1	31	17.5	41	17.6
2	16.6	12	17.4	22	17.4	32	18.1	42	17.5
3	16.3	13	17.2	23	17.4	33	17.5	43	16.5
4	16.1	14	17.4	24	17.5	34	17.4	44	17.8
5	17.1	15	17.4	25	17.4	35	17.4	45	17.3
6	16.9	16	17.0	26	17.6	36	17.1	46	17.1
7	16.8	17	17.3	27	17.4	37	17.6	47	17.4
8	17.4	18	17.2	28	17.3	38	17.7	48	16.9
9	17.1	19	17.4	29	17.0	39	17.4	49	17.3
10	17.0	20	16.8	30	17.8	40	17.8	50	17.6

6. Comment on the stationarity of the following processes with reasons. 4

i) $y_t - 1.1y_{t-1} + 3y_{t-2} = 0$

ii) $y_t - y_{t-1} + y_{t-2} – y_{t-3} = 0$

7. Plot 250 simulated observations from the process $(1-0.5L)y_t = (1-0.7L)e_t$, e_t are iid $N(0,0.1)$. Also plot its autocorrelations for first 14 lags. 4

[Turn over]

8. Sales of papers (in thousands) from Jan 1963 on Dec 1972 are shown below. Fit an appropriate ARIMA model to the data and forecast for next two years. Plot the original and the forecasted values on the same graph with error band. Give all necessary codes and outputs in the answer script. 7

SN	y_t	SN	y_t	SN	y_t	SN	y_t	SN	y_t	SN	y_t
1	562.67	21	560.72	41	701.1	61	795.33	81	742	101	835.08
2	599	22	602.53	42	790.07	62	788.42	82	847.15	102	934.59
3	668.51	23	626.38	43	594.62	63	889.96	83	731.67	103	832.5
4	597.79	24	605.51	44	230.71	64	797.39	84	898.52	104	300
5	579.88	25	646.78	45	617.12	65	751	85	778.14	105	791.44
6	668.23	26	658.44	46	691.39	66	821.25	86	856.07	106	900
7	499.23	27	712.90	47	701.07	67	691.60	87	938.83	107	781.72
8	215.18	28	687.71	48	705.77	68	290.65	88	813.02	108	880
9	555.81	29	723.91	49	747.63	69	727.14	89	783.41	109	875.02
10	586.93	30	707.18	50	773.39	70	868.36	90	828.11	110	992.96
11	546.14	31	629	51	813.78	71	812.39	91	657.31	111	976.80
12	571.11	32	237.53	52	766.71	72	799.56	92	310.03	112	968.69
13	634.71	33	613.29	53	728.87	73	843.03	93	780.	113	871.67
14	639.28	34	730.44	54	749.19	74	847	94	860	114	1006.85
15	712.18	35	734.92	55	680.95	75	941.95	95	780	115	832.03
16	621.56	36	651.81	56	241.42	76	804.30	96	807.99	116	345.58
17	621	37	676.16	57	680.23	77	840.30	97	895.21	117	849.52
18	675.98	38	748.18	58	708.32	78	871.52	98	856.07	118	913.87
19	501.32	39	810.68	59	694.23	79	656.33	99	893.26	119	868.75
20	220.28	40	729.36	60	772.07	80	370.51	100	875	120	993.73

****