INDIAN STATISTICAL INSTITUTE
Research and Training School
B. Stat. Part IV: 1966-67

Question Papers - Contents

Sl. No.	Exam. No.	Date	Subject
Periodical Examinations			
1	6	26. 9.66	Statistics-4: Probability
2	11	3.10.66	Statistics-8: Demography (Theory and Practical)
3	18	10.10.66	Statistics-6: Sample Surveys (Theory and Practical)
4	25	17.10.66	Statistics-7: Econometrics (Theory and Practical)
5	33	7.11.66	Statistics-5: Statistical Methods (Theory and Practical)
6	41	21.11.66	Statistics-4: Statistical Inference
Mid-Year Examinations			
7	62	21.12.66	Statistics-4: Probability
8	63	22.12.66	Statistics-4: Statistical Inference
9	64	23.12.66	Statistics-5: Statistical Methods (Theory)
10	65	23.12.66	Statistics-5: Statistical Methods(Practical
11	66	26.12.66	Statistics-6: Sample Surveys (Theory)
12	67	26.12.66	Statistics-6: Sample Surveys (Practical)
13	68	27.12.66	Statistics-7: Econometrics and Planning Techniques (Theory)
14	69	27.12.66	Statistics-7: Econometrics (Practical)
15	70	24.12.66	Statistics-8: Demography (Theory and Practical)
16	71	28.12.66	Statistics-8: Educational Statistics (Theory and Practical)
Periodical Examinations			
17	139	27. 2.67	Statistics-5: Statistical Methods (Theory)
18	140	6. 3.67	Statistics-4: Statistical Inference
19	141	13. 3.67	Statistics-7: Econometrics (Theory and Practical)
20	142	20. 3.67	Statistics-6: Design of Experiments (Theory and Practical)
21	143	27. 3.67	Statistics-7: Industrial Statistics (Theory and Practical)
22.	144	10. 4.67	Statistics-4: Probability
Annual Examination			
23	205	23. 5.67	Statistics-4: Probability
24	206	24. 5.67	Statistics-4:Inference
25	207	25. 5.67	Statistics-5:Statistical Methods (Theory and Practical)
26	208	26. 5.67	Statistics-6: Design of Experiments (Theory and Practical)
27	209	27. 5.67	Statistics-7: Industrial Statistics (Theory and Practical)
28	210	29. 5.67	Statistics-7: Econometrics (Theory and Practical)
29	211	30. 5.67	Statistics-8: Genetics (Theory and Practical
30.	119	12. 8.66	Statistics-6: Official Statistics

INDIAN STATISTICAL INSTITUTE
Research and Training School
B.Stat. Part IV: 1966-67

PERIODICAL EXAMINATION

Statistics-4: Probability

Date: 26.9.66 Maximum marks: 100 Time: 3 hours.

Note:- The whole paper carries 130 marks. You may attempt any part of any question.

1. In a sequence of independent tosses of an unbiased coin, let S_n be the total number of heads in n tosses minus the total number of tails in n tosses for $n \geq 1$ and let $S_0 = 0$.

(a) If $u_n = P(S_n = 0)$, $n \geq 1$ and $u_0 = 1$, show that the generating function of u_n is given by

$$U(s) = (1 - s^2)^{-1/2} \qquad (10)$$

(b) If $f_n = P(S_1 \neq 0, S_2 \neq 0, \ldots, S_{n-1} \neq 0, S_n = 0)$, $n \geq 1$ and $f_0 = 0$, show that the generating function of f_n is given by

$$F(s) = \frac{U(s) - 1}{U(s)} . \qquad (15)$$

(c) Let $v_n = P(S_1 \neq 0, S_2 \neq 0, \ldots, S_n \neq 0)$, $n \geq 1$, and $v_0 = 1$, show that the generating function of $\{v_n\}$ is $(1+s)\ U(s)$. (10)

(d) Let $k \geq 1$ and put $f_n^{(k)}$ = the probability that the k-th equalisation (of heads and tails) occurs at trial n and put $f_0^{(k)} = 0$. Prove that the generating function of $\{f_n^{(k)},\ n = 0, 1, \ldots\}$ is $F^k(s)$. (10)

(e) For $n \geq 1$, prove that the conditional probability

$$P(S_1 \geq 0, S_2 \geq 0, \ldots, S_{2n-1} \geq 0 \mid S_{2n} = 0) = \frac{1}{n+1} . \qquad (10)$$

(f) For $k \geq 0$, let $v_n^{(k)}$ be the probability that exactly k among the random variables $S_1, S_2, \ldots, S_n$ are zero (here $n \geq 1$) and let $v_0^{(k)} = 1$. Show that the generating function of $\{v_n^{(k)},\ n = 0,1,2,\ldots\}$ is $F^k(s)U(s)(1+s)$ [Hint: Use (c) and (d)]. (15)

(g) We shall say that Peter leads at trial i $(i \geq 1)$ if $S_i > 0$ or $S_i = 0$ and $S_{i-1} > 0$. Prove that the probability that Peter leads in exactly $2k$ out of $2n$ trials is $u_{2k} \cdot u_{2n-2k}$, for $0 \leq k \leq n$ and $n = 1,2,\ldots$ (10)

(h) For $n \geq 1$, prove that

$$P(S_{2n} > S_0, S_{2n} > S_1, \ldots, S_{2n} > S_{2n-1}) = \frac{1}{2} u_{2n} .$$

[Hint:- $P(S_1 > 0, S_2 > 0, \ldots, S_{2n} > 0) = \frac{1}{2} u_{2n}$]. (10)

P.T.O.

2. State clearly and prove the Ballot Theorem. [You may assume the Reflection Principle]. (15)

3. (a) In a sequence of independent tosses of a coin (probability of heads = p, probability of tails = $q = 1 - p$), let u_n be the probability that the combination HT (H = heads, T = tails) occurs for the first time at trials number $n-1$ and n. Prove that the generating function of

$$\{u_n\} \text{ is } pqs^2 / (1 - ps)(1 - qs). \qquad (15)$$

(b) Let a_n be the probability that in n tosses an even number of heads occurred, $n \geq 1$ and let $a_0 = 1$. Prove that $2a_n = 1 + (q-p)^n$, $n = 0, 1, \ldots$ (10)

INDIAN STATISTICAL INSTITUTE
Research and Training School
B.Stat. Part IV:1966-67

PERIODICAL EXAMINATION.

Statistics 8: Demography (Theory and Practical)

Date:3.10.66. Maximum marks: 100 Time: 3 hours

1. Discuss briefly the method adopted by the census actuary in calculating the survival probabilities ${}_{10}p_x$ and p_x in different age groups for the Census of India 1941-50 Life Tables. [30]

2. Write short notes on:

   i) Force of mortality
   ii) Complete expectation of life at age x. [10]

3.(a) Starting with $l_{15} = 10{,}000$ compute abridged life table values of $l_{20}, l_{25}, l_{30}, \ldots, l_{50}$ for a population of females in the child bearing age range from the table given below.

[${}_nP_x$ = population enumerated in the age group x to x + n.

${}_nD_x$ = number of deaths in the age group x to x + n during a year]

Age group x to x+5	Population in the age group x to x+5 ${}_nP_x$	Deaths ${}_5D_x$
15 - 20	295,516	3390
20 - 25	278,236	3342
25 - 30	251,724	3768
30 - 35	230,319	4216
35 - 40	191,236	4314
40 - 45	152,612	3677
45 - 50	127,422	4518

[20]

(b) Given that the complete expectation of life at ages 30 and 31 for a particular group are 21.39 and 20.91 years respectively and that the number living at age 30 is 41,176 find the number that attains the age 31. [10]

4.(a) Briefly examine the chief defects in the Indian vital registration data and the factors responsible for them.

(b) Discuss briefly the causes and trends in mortality during the neo-natal and post neo-natal periods. [30]

INDIAN STATISTICAL INSTITUTE
Research and Training School

B. Stat. Part IV: 1966-67

PERIODICAL EXAMINATION

Statistics-6: Sample Surveys Theory and Practical

Date: 10.10.66. Maximum marks: 100 Time: 3 hours

Answer Questions 1 and 5, and any <u>two</u> of the <u>remaining</u>.
<u>All questions carry equal marks</u>.

1.(a) State the situations in which complete enumeration is unavoidable, giving examples. Describe briefly the conditions under which sample survey could be prefered to complete enumeration, indicating the advantages of the former.

(b) Write short notes on any <u>three</u> of the following :

i) Relative standard error of the estimate.
ii) Centrally located systematic sample.
iii) Lahiri's method of probability proportionate to size (p.p.s) selection.
iv) Use of sampling in census.

2.(a) For estimation of production of wheat in a region, a sample of n fields is drawn with replacement following pps sampling procedure, size being the area under wheat in them and yield per acre is determined for each of the sample fields. Suggest an unbiased estimator of total wheat production in the region. Derive its sampling variance and also obtain an unbiased estimator of this variance.

(b) Explain briefly the concept of pps sampling. Taking each unit as made of sub-units equal to its size, show that the usual estimator of population total based on a simple random sample of n sub-units selected with replacement is the same as estimator in pps sampling with replacement and find variance of this estimator using sub-unit concept.

3.(a) Find an unbiased estimator of the population variance σ^2 both in case of simple random sampling with and without replacement.

(b) In a finite bivariate population of N units, the means and standard deviations of the variables x and y are $\bar{X}$, $\bar{Y}$ and σ_x, σ_y respectively and correlation coefficient between x and y is ρ. Derive the correlation coefficient $\rho(\bar{x},\bar{y})$ between the sample means $\bar{x}$ and $\bar{y}$ based on the same sample of n units selected with simple random sampling without replacement.

4.(a) In case of linear systematic sampling, suggest an unbiased estimator $\hat{Y}$ for the population $\bar{Y}$ when the sample size is not a sub-multiple of the number of units in the population and hence prove that $E(\hat{Y}) = \bar{Y}$.

(b) Suggest two methods of systematic sampling which make the sample mean unbiased for Y when the total number of population units is not a multiple of the sample size.

GO ON TO THE NEXT PAGE

4.(c) Derive the sampling variance of sample mean in case of circular systematic sampling and compare it with that of simple random sampling without replacement.

5. For estimating the total absentees in 325 factories situated in a district, a pps sample of 20 factories was drawn with replacement. Utilizing the data given below estimate the total absentees and its relative standard error. Total number of workers in the factories is 25,600.

Number of workers (x) and number of absentees (y) in 20 sample factories.

Sr. No.	x	y	Sr. No.	x	y
1	95	9	11	148	16
2	79	7	12	89	4
3	30	3	13	57	5
4	45	2	14	132	13
5	28	3	15	47	4
6	142	8	16	43	9
7	125	9	17	116	12
8	81	10	18	65	8
9	43	6	19	103	9
10	53	2	20	52	8

INDIAN STATISTICAL INSTITUTE
Research and Training School
B. Stat. Part IV: 1966-67

PERIODICAL EXAMINATION

Statistics-7: Econometrics Theory and Practical

Date: 17.10.66. Maximum marks: 100 Time: 3 hours.

1. EITHER
Mention the statistical measures commonly employed for studies on income distributions. What is the most important mathematical property of all these measures?

OR
Derive the equation for the Lorenz curve of the lognormal distribution. Discuss the properties of the Lorenz curves for the family of lognormal distributions. [15]

2. Give an outline of Champernowne's model leading to the Pareto distribution of income. Bring out the significance of the main assumptions as clearly as possible. [25]

3. Write short notes on any three:
   a) Engel curve,
   b) two definitions of inferior goods,
   c) Slutsky's relation,
   d) substitutes and complements,
   e) uses of family budget data. [3 × 5=15]

4. Below we give the estimated distribution of all persons in rural India by per capita monthly expenditure on all items.
   a) Find the Lorenz ratio and the share of the top 10% in the aggregate domestic consumer expenditure of the commodity. [15]
   b) Test graphically whether the Pareto curve can fit the distribution. [15]
   c) Assuming that the log-logistic distribution is adequate, find the estimates of the parameters and calculate the 'expected' frequency for the interval Rs.11-13.

per capita monthly expenditure (Rs.)	percentage distribution of persons	average per capita monthly expenditure on all items (Rs.)
0 - 8	15.47	6.20
8 - 11	17.80	9.65
11 - 13	12.04	11.92
13 - 15	10.31	13.96
15 - 18	10.83	16.21
18 - 21	8.75	19.06
21 - 24	6.94	22.26
24 - 28	5.77	26.06
28 - 34	4.82	30.53
34 - 43	3.73	36.89
43 - 55	1.97	48.98
55 -	1.57	83.05
all	100.00	17.24

[15]

INDIAN STATISTICAL INSTITUTE
Research and Training School
B. Stat. IV year:1966-67

PERIODICAL EXAMINATION

Statistics-5: Statistical Methods: Theory and Practical

Date:7.11.66. Maximum marks:100 Time: 3 hours.

1.(a) Define a generalised inverse of a matrix. [5]

(b) Show that if A^- be a generalised inverse of matrix A and $A^-A = H$, a general solution of the consistent equations $A\underset{\sim}{x} = y$ is given by $\underset{\sim}{x} = A^-y + (H - I)\underset{\sim}{z}$ where I is the identity matrix and $\underset{\sim}{z}$ an arbitrary vector. [15]

2\. Observations on 12 independent normal variables with expectations given as linear functions of $\theta_1, \theta_2, \theta_3, \theta_4$ and a common variance σ^2 led to the following normal equations

$$
\begin{aligned}
2\theta_1 + \theta_2 + \theta_3 + \theta_4 &= 6.9 \\
\theta_1 + 5\theta_2 + 2\theta_3 + 2\theta_4 &= 12.4 \\
\theta_1 + 2\theta_2 + 10\theta_3 + 7\theta_4 &= 17.2 \\
\theta_1 + 2\theta_2 + 7\theta_3 + 5\theta_4 &= 13.6
\end{aligned}
$$

Sweepout operations on these equations gave the following computed figures.

A particular solution: $\hat{\theta}_1 = 2.0,\ \hat{\theta}_2 = 1.6,\ \hat{\theta}_3 = 1.2,\ \hat{\theta}_4 = 0$

A generalised inverse of the matrix C of normal equation

$$C^- = \frac{1}{81}\begin{pmatrix} 46 & -8 & -3 & 0 \\ -8 & 19 & -3 & 0 \\ -3 & -3 & 9 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

Basis for the vector space generated by rows of matrix C:

$$(1, 0, 0, \tfrac{1}{9}), \quad (0, 1, 0, \tfrac{1}{9}), \quad (0, 0, 1, \tfrac{2}{3}).$$

The sum of squares of the 12 observations was computed as 54.44.

(a) Examine if the following linear functions of parameters are estimable

(i) $\theta_1 + 2\theta_2 + \theta_3 + \theta_4$; (ii) $\theta_2 - \theta_3$. [6]

(b) Test the hypothesis $\theta_1 = \theta_2$. [14]

(c) An independent observation on yet another normal variable $N(3\theta_1 + 3\theta_2 - \theta_3, \sigma^2)$ is now obtained as 9.3. Compute revised estimates for $\theta_1 - \theta_2$ and its estimated standard error, showing each step of computation. [20]

PLEASE TURN OVER

3. To determine the effect of the dilution of the electrolyte on the thickness of the coating on aluminuim foil an experiment was conducted using eight different dilutions at a constant current strength. In each experimental set up the thickness of coating on two aluminium foils were recorded.

dilution x	thickness of coating y
4.0	9.7 9.4
5.0	10.5 11.3
6.0	10.6 11.7
7.0	7.5 9.4
8.0	6.8 7.2

dilution x	thickness of coating y
9.0	5.7 6.3
10.0	5.1 4.8
11.0	4.6 3.8

Obtain the polynomial regression of y on x using a table of orthogonal polynomials and determine the value of x at which y is maximum. [34 + 6]= 40

Research and Training School
B. Stat. Part IV: 1966-67

PERIODICAL EXAMINATION

Statistics-4 : Statistical Inference

Date: 21.11.66. Maximum marks: 100 Time: 3 hours.

Answer Questions 1, 2, 4 and *any two* from the rest.

1. Explain the following:

   i) Simple hypothesis, (ii) composite hypothesis
   iii) randomized test, (iv) non-randomized test,
   v) level of a test, (vi) size of a test,
   vii) most powerful test, (viii) uniformly most powerful test, (ix) unbiased test. [13]

2. Let X be a random variable having the probability density functions f_o and f_1 under the hypothesis H and the alternative hypothesis K, respectively.

   i) Show that for any given α $(0 \leq \alpha \leq 1)$ there exists a test φ^* such that

$$\varphi^*(x) = \begin{cases} 1, & \text{if } f_1(x) > c\, f_o(x) \\ 0, & \text{if } f_1(x) < c\, f_o(x) \end{cases}$$

where $c \geq 0$, and

$$E_0\, \varphi^*\ (x) = \alpha$$ [12]

   ii) Prove that the above test φ^* is a most powerful level α test for testing H against K. [8]

   iii) If ψ is a most powerful level α test for testing H against K, then show that for some $c \geq 0$,

$$\psi(x) = \begin{cases} 1, & \text{if } f_1(x) > c\, f_o(x) \\ 0, & \text{if } f_1(x) < c\, f_o(x) \end{cases}$$

for (almost) all x, If

$$\alpha' = \int_{x:\ f_1(x) > 0} f_o(x)\, d\, x \geq \alpha,$$

then show that

$$E_o\ \psi(x) = \alpha\ .$$ [10]

   iv) Let $\beta(\alpha)$ denote the power of a MP level α test for the above testing problem. Show that

$$\alpha_1 < \alpha_2 \Rightarrow \beta(\alpha_1) \leq \beta(\alpha_2),$$

and the equality holds only when $\beta(\alpha_1) = 1$. [8]

3. With reference to the testing problem stated in Question 2, a test φ_1 is said to be better than a test φ_2, if

$$E_o\varphi_1(x) \leq E_o\varphi_2(x), \quad E_1\varphi_1(x) \geq E_1\varphi_2(x),$$

with at least one strict inequality.

[Please Turn Over]

Let Φ^* be the class of all tests φ given by

$$\varphi(x) = \begin{cases} 1, & \text{if } f_1(x) > c\, f_o(x) \\ 0, & \text{if } f_1(x) < c\, f_o(x), \end{cases}$$

where $c \geq 0$.

Deduce the following from the results stated in Question 2.

i) For any test φ not in Φ^* there exists a test φ^* in Φ^* which is better than φ.

ii) If φ_1 and φ_2 are in Φ^*, then neither of these is better than the other when [12]

$$f_o(x) > 0 \Rightarrow f_1(x) > 0.$$

4.(a) Define a 'Monotone likelihood ratio' family of distributions. Give examples, with special reference to one-parameter exponential family. [5]

(b) Let X be a random variable having density $f(x;\theta)$ which has a monotone likelihood ratio in $T(x)$ (θ is real). Obtain a uniformly most powerful level α $(0 < \alpha < 1)$ test of the hypothesis $H: \theta \leq \theta_o$ against $K: \theta > \theta_o$. [10]

5. Let $X_1, X_2, \ldots, X_n$ be mutually independent and identically distributed random variables with the common density function $f(.\,;\theta)$. In each of the following cases state whether a UMP test, based on these n observations, for testing H_o against H_1 at any level $\alpha (0 \leq \alpha \leq 1)$ exists. If it exists, then state its form ; otherwise, prove that it does not exist.

i) $f(x;\theta) = \frac{1}{\sqrt{2\pi}} \exp[-(x-\theta)^2/2]$ [6]

$$H_o: \theta = 0; \quad H_1: \theta \neq 0.$$

ii) $f(x;\theta) = e^{-(x-\theta)}, \; x \geq \theta$

$= 0,$ otherwise

$$H_o: \theta = 0; \quad H_1: \theta \neq 0.$$ [6]

6. Define: Sufficient statistic. [1]

State the 'factorization criterion' for sufficiency. [2]

Show that the 'factorization criterion' is equivalent to the definition of a sufficient statistic (prove only in the discrete case). [3]

Consider the three-parameter family of distributions having densities

$$f_n(\underset{\sim}{x};\, a, b, c) = \prod_{i=1}^{n} f(x_i, a, b, c)$$

where

$$f(x; a, b, c) = \begin{cases} \frac{(x-a)^{c-1}}{\Gamma(c)\, b^c}\, e^{-\frac{x-a}{b}}, & \text{if } x \geq a \\ 0, & \text{if } x < a. \end{cases}$$

$$-\infty < a < \infty, \quad 0 < b < \infty, \quad 0 < c < \infty, \quad n \geq 3.$$

For each one of the following sub-families find sufficient statistic from the list given below:

(A) $\min (x_1, x_2, \dots, x_n)$

(B) $\prod_{i=1}^{n} x_i$

(C) $\sum_{i=1}^{n} x_i$

(D) $\max (x_1, \dots, x_n)$.

The list of families is specified by the following:

i) $b = 1$, $c = 1$ [2]

ii) $a = 0$, $b = 1$ [2]

iii) $c = 1$, $a = 0$. [2]

7. Class assignments.
(. to be submitted not later than 28 November 1966.) [10]

INDIAN STATISTICAL INSTITUTE
Research and Training School
B. Stat. Part IV: 1966-67

MID-YEAR EXAMINATION

Statistics-4: Probability

Date: 21.12.66 Maximum marks: 100 Time: 3 hours

The whole paper carries 115 marks.
Attempt any questions, or parts thereof,
carrying a maximum of 100 marks.

1. Let $\mathcal{E}$ be a persistent repetitive pattern. Suppose N_k is the number of occurrences of $\mathcal{E}$ in k trials. Put $q_{k,n} = P\{N_k = n\}$. Prove that

(a) $E(N_r) = u_1 + u_2 + \dots + u_r$. [5]

(b) $q_{k,n}$ is the coefficient of s^k in

$$F^n(s)\,\frac{1 - F(s)}{1 - s}.$$ [10]

(c) $E(N_r^2)$ is the coefficient of s^r in

$$\frac{F^2(s) + F(s)}{(1-s)\{1-F(s)\}^2}.$$ [10]

2. In the Branching Process, prove that the probability of extinction is one if and only if

$$\mu = \sum_{k=1}^{\infty} kp_k,$$

the expected number of direct descendants of a single individual, is less than or equal to one. [20]

3. In N tosses of an unbiased coin, where N is a random variable with a Poisson distribution (with parameter $\lambda > 0$), prove that the numbers of heads and tails are stochastically independent random variables. [15]

4. In a sequence of independent tosses of a coin with probability of heads $= p$, prove that the repititive pattern 'equalisation of heads and tails' is persistent or transient according as $p = \frac{1}{2}$ or $p \neq \frac{1}{2}$. [15]

5. In the Ballot Problem, suppose candidate A gets p votes and candidate B gets q votes, where $p \geq q > 0$. Prove that in a random counting of the votes, the probability that at each stage A has at least as many votes as B is $(p + 1 - q)/(p + 1)$. [20]

In $2n$ tosses of an unbiased coin, prove that the probability that there are exactly r equalisations of heads and tails (where $r \leq n$) is

$$\frac{1}{2^{2n-r}}\binom{2n-r}{n}$$ [20]

INDIAN STATISTICAL INSTITUTE
Research and Training School
B. Stat. Part IV: 1966-67
MID-YEAR EXAMINATION

Statistics-4: Statistical Inference

Date: 22.12.66 Maximum marks: 100 Time: 3 hours

Answer Questions 1 and 2 and any two from the rest.

1. Let $X_1, \ldots, X_m$ and $Y_1, \ldots, Y_n$ be independent samples from $N(\xi, 1)$ and $N(\eta, 1)$, respectively.

(a) Show that the most powerful level α test ϕ^* for testing the simple hypothesis $H: \xi = \eta = (m\xi_1 + n\eta_1)/(m+n)$ against the simple alternative $K: \xi = \xi_1$, $\eta = \eta_1 . (\xi_1 < \eta_1)$ rejects H iff $\bar{Y} - \bar{X} > C$, where C is so chosen that ϕ^* is of size α. [12]

(b) Show that the power of ϕ^* is an increasing function of $(\eta - \xi)$. [3]

(c) Deduce from the above results that ϕ^* is the UMP level α test for testing $H^*: \eta \leq \xi$ against $K^*: \eta > \xi$. [7]

[Hint: Show that any level α test for H^* is also of level α for H and use the fact that ϕ^* is independent of ξ_1, η_1.]

2.(a) Explain the following concepts with illustrations:

i) A complete family of distributions. [3]
ii) A similar test. [3]
iii) A test having Neyman structure. [3]

(b) In each of the four following cases, a statistic T is defined. Consider the families of distributions of T.

i) Let $X_1, \ldots, X_n$ be a random sample from $N(a\sigma, \sigma^2)$, with 'a' fixed, and $0 < \sigma < \infty$.

$$T = \left(\sum_1^n X_1, \ \sum_1^n X_1^2 \right).$$

ii) Let T be a random variable taking on the values -1, 0, 1, 2, ... with probabilities given by

$$P_\theta[T = -1] = \theta \ ; \quad P_\theta[T = t] = (1-\theta)^2 \theta^t;$$

$$t = 0, 1, 2, \ldots$$

0 < [illegible] < 1.

iii) Let $X_1, \ldots, X_n$ be a random sample from $N(0, \sigma^2)$, $0 < \sigma < \infty$. $T = \sum_1^n X_1$.

iv) Let T be distributed according to binomial distributions $B(n, \theta)$, $0 \leq \theta \leq 1$.

Classify each of the above families of distributions of T into one of the following and justify your answer:

(I) Complete (II) Boundedly complete but not complete

2.(c) Let X be a random variable with distribution $F \in \mathcal{P}$, and let T be a sufficient statistic for $\mathcal{P}$. Prove that a necessary and sufficient condition for all similar tests (similar on $\mathcal{P}$) to have Neyman structure with respect to T is that the family of distributions of T (induced by $\mathcal{P}$) be boundedly complete. [13]

3.(a) Let X and Y be the number of successes in two independent sets of n Bernoulli trials with probabilities p_1 and p_2 of success. Consider the problem of testing the hypothesis H: max $(p_1, p_2) \leq C$ against the alternative K: max $(p_1, p_2) > C$, where C is a specified constant, $0 < C < 1$. Show that for every unbiased size α test φ

$$E_{p_1, p_2} [\varphi (X, Y)] = \alpha, \quad \text{for} \quad p_1 = C \quad \text{or} \quad p_2 = C. \qquad [8]$$

(b) On the basis of a random sample of size n from $N(\xi, \sigma^2)$ obtain (explicitly) the UMP unbiased level α test for testing the hypothesis H: $\sigma^2 = \sigma_0^2$ against $\sigma^2 \neq \sigma_0^2$; ξ is unknown. [illegible] [14]

4.(a) Let X be a random variable having density $f(x; \theta)$, where θ is a real parameter. It is known that the power function of every test is differentiable with respect to θ. How do you obtain a locally most powerful level α test for testing $\theta = \theta_0$ against $\theta > \theta_0$? [7]

(b) Let $X_1, \ldots, X_n$ be independent and identically distributed random variables with the common density

$$f(x; \theta) = \frac{1}{\pi} \frac{1}{1 + (x - \theta)^2}, \quad -\infty < x < \infty.$$

Find the locally MP level α test for testing $\theta = 0$ against $\theta > 0$. [7]

Show that the power of this test tends to 0 as θ tends to ∞ ($\alpha < \frac{1}{2}$). [8]

5.(a) Define a likelihood-ratio test. [2]

(b) Give an example where the likelihood-ratio test is useless in the sense that one can do better without making any observation. [10]

(c) Find the likelihood-ratio test for the problem stated in Question 3(b). [10]

INDIAN STATISTICAL INSTITUTE
Research and Training School
B. Stat. Part IV: 1966-67

MID-YEAR EXAMINATION

Statistics-5: Statistical Methods Theory

Date: 23.12.66 Maximum marks: 100 Time: 3 hours

Attempt any four questions. All questions carry equal marks.

1.(a) Define a p-variate normal distribution $N_p(\underset{\sim}{\mu}, \Sigma)$ and derive its characteristic function. [7]

(b) Show that if $\underset{\sim}{X}$ has a $N_p(\underset{\sim}{\mu}, \Sigma)$ distribution and $\underset{\sim}{Y} = A\underset{\sim}{X}$ defines a linear transformation from $\underset{\sim}{X}' = (X_1, X_2, \ldots, X_p)$ to new variables $\underset{\sim}{Y}' = (Y_1, Y_2, \ldots, Y_q)$ then $\underset{\sim}{Y}$ is $N_q(A\underset{\sim}{\mu}, A \Sigma A')$. [8]

(c) Show that, if in addition, the vector X' be partitioned as

$$\underset{\sim}{X}' = (\overset{p_1}{\underset{\sim}{X}'_{(1)}}, \overset{p_2}{\underset{\sim}{X}'_{(2)}}), \quad p_1 + p_2 = p$$

and the mean vector $\underset{\sim}{\mu}'$ and dispersion matrix Σ be correspondingly partitioned as

$$\underset{\sim}{\mu}' = (\overset{p_1}{\mu'_{(1)}}, \overset{p_2}{\mu'_{(2)}})$$

$$\Sigma = \begin{array}{ccc} p_1 & p_2 & \\ \Sigma_{11} & \Sigma_{12} & p_1 \\ \Sigma_{21} & \Sigma_{22} & p_2 \end{array}$$

then given $\underset{\sim}{X}_{(1)} = \underset{\sim}{x}_{(1)}$, the conditional distribution of $\underset{\sim}{X}_{(2)}$ is $N_{p_2}(\underset{\sim}{\mu}_{(2)} + \Sigma_{21}\Sigma_{11}^{-1}(\underset{\sim}{x}_{(1)} - \underset{\sim}{\mu}_{(1)}), \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12})$. [10]

2. Let $(x_{1\lambda}, x_{2\lambda}, \ldots, x_{p\lambda})$, $\lambda = 1, 2, \ldots, n$ represent a sample of size n from $N_p(\underset{\sim}{\mu}, \Sigma)$. Write

$$S_{ij} = \sum_{\lambda=1}^{n} (x_{i\lambda} - \bar{x}_i)(x_{j\lambda} - \bar{x}_j)$$

where $\bar{x}_i = \frac{1}{n}\sum_{\lambda=1}^{n} x_{i\lambda}$ and $\bar{x}_j = \frac{1}{n}\sum_{\lambda=1}^{n} x_{j\lambda}$.

Show that if $\{S^{ij}\} = \{S_{ij}\}^{-1}$ and $\{\sigma^{ij}\} = \Sigma^{-1}$, then σ^{pp}/S^{pp} has a chisquare distribution with (n-p) degrees of freedom.

3. Assuming the result stated in Question 2, or otherwise, derive the null distribution of the Hotellings T^2 statistic and discuss some of its applications to problems in multivariate analysis.

Please Turn Over

4. Let $Y_1, Y_2, \ldots Y_n$ represent independent normally distributed random variables with a common unknown variance σ^2 and mean values given by $E(\underset{\sim}{Y}) = A\underset{\sim}{\theta}$, where $\underset{\sim}{Y}' = (Y_1, Y_2, \ldots, Y_n)$, $A(n \times m)$ a known matrix of coefficients and $\underset{\sim}{\theta}' = (\theta_1, \theta_2, \ldots, \theta_m)$ a vector of unknown parameters. Show that

   (a) if $R_0^2 = \min_{\theta} (\underset{\sim}{Y} - A\underset{\sim}{\theta})(\underset{\sim}{Y} - A\underset{\sim}{\theta})'$, R_0^2/σ^2 has a chisquare distribution with $(n-r)$ degrees of freedom, where $r = \text{Rank}(A)$. [10]

   (b) If $B = HA$, rank $B = s$ and the equations $B\underset{\sim}{\theta} = \underset{\sim}{b}$ are consistent and if $R_H^2 = \min_{\theta} (\underset{\sim}{Y} - A\underset{\sim}{\theta})'(\underset{\sim}{Y} - A\underset{\sim}{\theta})$, subject to $B\underset{\sim}{\theta} = \underset{\sim}{b}$ then $(R_H^2 - R_0^2)/\sigma^2$ is independently distributed of R_0^2 as a chisquare with s degrees of freedom. [15]

5. Consider a two-way classification in m classes of A and n classes of B with the same number r of observations in each cell. It is known that while the m classes of A are fixed the n classes of B are determined by random sampling from a larger collection. Show that under certain conditions (which you are required to state in full) to test the hypothesis of no main effects of A, one can apply the variance ratio test to the ratio of mean square due to main effects of A and the mean square due to interactions $A \times B$. [25]

6. With the same set-up as in Question 5, compute the expected values of the various mean squares occurring in the usual analysis of variance table, namely those attributable to the main effects of A and B, interactions $A \times B$ and Error. [25]

|. 651

INDIAN STATISTICAL INSTITUTE
Research and Training School
B.Stat. Part IV: 1966-67

MID-YEAR EXAMINATION

Statistics-5: Statistical Methods Practical

Date: 23.12.66 Maximum marks: 100 Time: 3 hours

Answer any two questions. All questions carry equal marks

1. The following table gives the yield (y) of ten plants as observed in a randomised block experiment carried in four blocks of four plots each. Shown in parenthesis are the preliminary yields (x) recorded on the same plants.

Block	Treatments			
	A	B	C	D
1	91 (85)	88 (81)	83 (90)	102 (93)
2	118 (121)	94 (93)	110 (106)	109 (114)
3	109 (114)	105 (106)	115 (111)	94 (93)
4	102 (107)	91 (92)	96 (102)	88 (92)

Examine if the treatment effects are significantly different, correcting for differences in preliminary yields.

[Total (corrected) sum of squares and products

$$S_{yy} = 1526.0 \qquad S_{xx} = 2040.0$$

$$S_{xy} = 1612.00 \text{]}.$$

2. Measurements were taken on 86 individuals on each of four different characteristics (X_0, X_1, X_2 and X_3). The corrected sum of squares and products matrix for variables X_1, X_2 and X_3 was computed as

$$(s_{ij}) = \begin{pmatrix} 0.01875 & 0.00848 & 0.00684 \\ & 0.02904 & 0.00878 \\ & & 0.02886 \end{pmatrix}$$

and its inverse as

$$(s^{ij}) = \begin{pmatrix} 64.21 & -15.57 & -10.49 \\ & 41.71 & -9.00 \\ & & 39.88 \end{pmatrix}$$

Other figures available are

$$s_{00} = 0.02781, \quad s_{01} = 0.03030, \quad s_{02} = 0.4410$$

$$s_{03} = 0.03629$$

mean values:

$\bar{x}_0 = 3.1685$, $\bar{x}_1 = 2.2752$, $\bar{x}_2 = 2.1523$, $\bar{x}_3 = 2.1128$.

Please Turn Over

using these figures

(a) Compute the multiple (linear) regression equation of X_0 on X_1, X_2 and X_3.

(b) Test if the three regression coefficients occurring in this expression

  i) are significantly different from 0.

  ii) are significantly different from one another.

  iii) are significantly different from 1.

3. Yield of paddy (in tolas) were recorded for three concentric circular cuts of radii 2 ft., 4 ft., and 5 ft. 8 in., in each of 15 different fields. The mean values and corrected sums of squares and products computed from these data were as follows:

2 ft.	4 ft.	5 ft. 8 in.
	Mean Yield	
29.0	102.9	202.7
	Corrected SP matrix	
430	325	340
	1561	2251
		3707

Examine if the mean yield rates (that is, yield per unit area) obtained from circles of different sizes are significantly different from one another.

INDIAN STATISTICAL INSTITUTE
Research and Training School
B. Stat. Part IV: 1966-67

MID-YEAR EXAMINATION

Statistics-6: Sample Surveys Theory

Date: 26.12.66 Maximum marks: 100 Time: 3 hours

Answer all questions

1.(a) Describe briefly the steps involved in planning a large scale sample survey with special reference to measurement and control of non-sampling errors. [15]

(b) Write short notes on (i) self-weighting design and (ii) cluster sampling. [5 + 5]= 10

2.(a) Discuss briefly the principle of stratification and mention also the advantages of stratified sampling. [5]

(b) In case of a stratified uni-stage sampling design, where units in each stratum are selected with probability proportional to a given measure of size, and with replacement, derive the optimum allocation for a given total sample size. [10]

(c) Let there be two strata with N_1 and N_2 units in them. Suppose one unit, say, the jth unit in the ith stratum (U_{ij}) is selected with probability proportional to its size X_{ij} from the total population of N_1+N_2 units and then one unit from the remaining (N_1-1) units in the ith stratum and two units from the other stratum are selected with equal probability without replacement. Show that the estimator

$$\hat{Y} = \frac{N_1 \bar{y}_1 + N_2 \bar{y}_2}{N_1 \bar{x}_1 + N_2 \bar{x}_2} X, \quad (X = \sum_{i=1}^{2} \sum_{j=1}^{N_i} X_{ij}),$$

is unbiased for the population total Y, $\bar{y}_i$ and $\bar{x}_i$ being the sample means for the characteristics x and y $(i = 1,2,)$. [10]

3.(a) Under what circumstances would you recommend the use of a multistage sampling design? [5]

(b) In a two-stage sampling design, n first stage units (fsu's) and from each sample fsu (possibly of varying sizes) m second stage units are selected using simple random sampling without replacement at both the stages. Obtain an unbiased estimator of the population total of a given characteristic and derive its sampling variance. Assuming the cost function to be of the form

$$C = C_0 + C_1 n + C_2 nm,$$

where C_0 is the over-head cost and C_1 and C_2 are the costs per fsu and ssu respectively, determine the optimum values of n and m, when the total cost is fixed at C'. [20]

Please Turn Over

4.(a) Stating clearly the assumptions involved, derive the sampling bias and variance for the ratio estimator of the population total based on a sample of n units selected from a population of N units with equal probability and without replacement. [10]

(b) Derive the condition for the ratio estimator to be more efficient than the usual unbiased estimator. [5]

(c) What do you understand by 'product method of estimation'? Under what condition, if any, would you prefer it to the usual unbiased estimator and the ratio estimator? [10]

INDIAN STATISTICAL INSTITUTE
Research and Training School
B. Stat. Part IV: 1966-67

MID-YEAR EXAMINATION

Statistics-6: Sample Surveys Practical

Date: 26.12.66 Maximum marks: 100 Time:3 hours

Answer any three of the four questions, Questions 1-4. All questions carry equal marks.

1. The farms of a small country were divided into 7 strata based on their areas reported in the last census. A sample of 3,000 farms was selected from the 100,000 farms in that country using simple random sampling without replacement in each stratum and adopting proportional allocation. Using the data given in Table 1,

(a) obtain an unbiased estimate of the overall mean yield per farm and estimate its relative standard error. [15]

(b) estimate the gain due to stratification as compared to unstratified simple random sampling of 3000 farms without replacement.

Table 1

stratum number	proportion of farms	estimated	
		mean	standard deviation
1	0.50	0.13	0.5
2	0.23	0.72	1.7
3	0.20	3.34	8.5
4	0.053	18.03	35.0
5	0.015	68.85	95.0
6	0.0012	786	200.0
7	0.0008	434	170.0

[10]

2. Raw wool contains varying amount of grease, dirt and other impurities and its quality is measured by the percentage of the weight of clean wool to that of raw wool termed clean content. To estimate the clean content, an electrical core boring machine is used, which takes the cores of about 1/4 lb. from a bale, which is then subjected to laboratory analysis. In an experiment 5 bales were selected from a large lot with equal probability and from each bale 4 cores were taken at random and clean content was determined. The results of this experiment are given in Table 2.

Table 2: The clean content of wool for 20 cores.

core	bales				
	1	2	3	4	5
1	54.3	57.0	54.2	56.2	59.9
2	56.2	57.4	55.5	54.4	57.8
3	58.3	58.5	56.4	60.1	60.3
4	53.2	57.6	57.2	58.7	57.3

(a) Estimate the average clean content of wool for the lot. Also obtain an estimate of its relative standard error. [10]

(b) Obtain the efficiency of sampling 10 bales and 2 cores from each bale as compared to that of the above scheme. [15]

3. From an urban area consisting of 1840 households with a total population of 8846 persons, a sample of 10 per cent of households is drawn with equal probability without replacement for estimating the total income in that area. Using the data given in Table 3,

(a) Obtain a ratio estimate for the total income taking number of persons as the supplementary variable and estimate its relative standard error. [15]

(b) Also calculate the efficiency of this ratio estimator compared to that of the usual unbiased estimator. [10]

Table 3: Distribution of sample households by income groups and household size.

house-hold size	income groups (in rupees)						all groups
	1-50	51-100	101-150	151-200	201-300	301-500	
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
1	2	6	1	1	-	-	10
2	6	9	12	3	2	-	32
3	5	6	10	13	2	-	36
4	2	5	8	13	12	3	43
5	-	3	6	18	2	2	31
6	-	1	2	5	7	3	18
7	-	-	1	2	4	2	9
8	-	-	-	1	3	1	5
total	15	30	40	56	32	11	184

4. To estimate the total number of persons (P) and the average household size (P/H) in an urban area, 24 blocks are selected with probability proportional to size with replacement, the size being previous census population, and from each selected block a sample of households is selected linear systematically with a random start. The sampling interval to be used in each sample block for selecting households is so specified that the sampling design becomes self-weighting with a constant inflation factor 480. Using the data given in Table 4.

(a) estimate P unbiasedly and obtain its relative standard error by estimating its sampling variance unbiasedly. [10]

(b) Also estimate the ratio P/H and its relative standard error. [15]

Table 4: Number of sample households and the number of persons in them for 24 sample blocks.

sample villages	no.of sample		sample villages	no.of sample		sample villages	no.of sample	
	households	persons		households	persons		households	persons
1	8	35.	9	5	19	17	1	6
2	7	40.	10	9	35	18	13	54
3	5	22.	11	7	36	19	0	0
4	6	32	12	6	32	20	6	18
5	5	16	13	5	26	21	5	27
6	6	28	14	10	33	22	4	20
7	2	8	15	7	28	23	5	21
8	9	32	16	8	29	24	11	47

Practical Record [25]

(to be submitted to the Dean's Office by 26.12.1966)

INDIAN STATISTICAL INSTITUTE
Research and Training School
B. Stat. Part IV: 1966-67

MID-YEAR EXAMINATION

Statistics-7: Econometrics and Planning Techniques: Theory

Date:27.12.66 Maximum marks: 100 Time: 3 hours

Answer Groups A and B in separate answerscripts.

Group A: Econometrics Theory Maximum marks:50

Answer any one question from Part I and any two questions from Part II

Part I

1.(a) Define the Lorenz Ratio, and show how it is related to the Gini Mean Difference. [9]

(b) Show how the theory of proportionate effect leads to the lognormal distribution [9]

2. Write short notes on any two: The universality of the Pareto law, the graphical test of log-normality, the log-logistic distribution, Lydall's model leading to the Pareto distribution. [18]

Part II

3. Write short notes on any two:

(a) Choice of algebraic form of the Engel curve,

(b) uses of the specific concentration curve,

(c) household size and composition in Engel curve analysis,

(d) demand projections based on the Engel curve. [8+8]=[16]

4. Discuss fully the various difficulties in the estimation of demand functions from time-series of market statistics. [16]

5. Give an outline of the Cobweb model of demand and supply. What would be the method of estimating the demand and the supply functions of such a model? [16]

Please Turn Over

Please Turn Over

Group B: Planning Techniques Theory

Max. marks: 50

All questions carry equal marks.
Answer any two questions.

1.(a) State the basic assumptions of the static input-output system.

(b) Explain the economic considerations underlying the set of equations in the static Leontief system.

(c) Show how the net-output-possibility schedule can be obtained, given labour supply and capacity restrictions in different sectors of the economy.

2.(a) If there are alternative processes for the production of a given commodity, how would you define the concept of a technologically efficient combination of inputs required to produce a given volume of output? Illustrate graphically by a simple example.

(b) Demonstrate the proposition that if there is one primary factor, then the existence of technological choice will not affect the fixity of input-output coefficients, irrespective of changes in the final bill of goods.

3.(a) State the basic assumptions of the two-sector Mahalanobis model and deduce its implications regarding the allocation of investment for long-run economic growth.

(b) Show that in the Mahalanobis model, the rate of savings is determined by the allocation parameters and productivity coefficients.

4.(a) Explain clearly the problem which Leontief dynamic system seeks to answer. Set up the basic relations of this system and derive the efficiency locus.

(b) Show that there may not be a production programme exhausting all stocks, if they are given arbitrarily.

INDIAN STATISTICAL INSTITUTE
Research and Training School
B. Stat. Part IV:1966-67

MID-YEAR EXAMINATION

Statistics-7: Econometrics Practical

Date:27.12.66 Maximum marks: 100 Time: 3 hours

1. The following gives the size distribution of incomes liable to surtax in UK during 1953-54:

Incomes (£)	2000-2499	2500-2999	3000-3999	4000-5999	6000-9999	10000-
percentage of returns	28.6	19.9	21.7	16.8	8.8	4.2

Estimate the relative frequency of returns in the range £ 2500-2999 on the basis of a fitted Pareto distribution. What is the Lorenz ratio corresponding to the fitted distribution? [15]

2. The following table relates to all individual incomes in India assessed for income tax.

income (Rs.)	Cumulative percentage of					
	1950-51			1960-61		
	assessees	income before tax	income after tax	assessees	income before tax	income after tax
3500	18.26	6.16	6.80	-	-	-
5000	46.24	18.09	20.14	37.49	16.16	18.35
7500	67.42	31.32	34.61	63.59	32.99	37.21
10000	77.96	40.67	44.63	75.75	43.99	49.35
15000	87.81	53.04	57.49	87.27	58.75	65.26
20000	91.98	60.39	64.80	92.20	67.68	74.58
30000	95.90	70.25	74.16	96.34	78.18	84.91
50000	98.25	79.43	82.43	98.75	87.81	93.14
100000	99.47	88.00	89.85	99.98	98.43	99.59
200000	99.86	93.53	94.53	99.99	99.11	99.82
∞	100.00	100.00	100.00	100.00	100.00	100.00

By using suitable statistical methods examine how the concentration of incomes changed during 1950-51 to 1960-61, and bring out the role of taxation in this process. [25]

3. Estimate the demand function for butter and margarine in Sweden on the basis of the data presented below. You may assume the constant elasticity form of the demand function.

year	consumption in kg.per person	retail price	overall consumer price index	income per person (current prices)
1930	18.04	2.13	164	860
1931	18.44	1.99	159	816
1932	18.85	1.88	156	753
1933	18.77	1.94	153	726
1934	19.11	2.16	154	782
1935	19.91	2.09	156	858
1936	20.38	2.06	158	907
1937	20.44	2.30	162	994
1938	20.20	2.42	166	1062
1939	20.44	2.53	171	1114

Please Turn Over

Estimate the standard errors of the elasticities you got. Given that the income elasticity of demand is 0.4, re-estimate the price elasticity and its standard error. Compare the two residual sums of squares. [35]

4. The following data are based on a family budget enquiry in rural India:

total household expenditure per person (Rs.)	5-11	11-15	15-21	21-28	28-34	34-43	43-55	55-
average of total household expenditure per person (Rs.)	7.9	12.9	17.6	24.2	30.5	36.9	49.0	89.1
value of cereals consumed per person (Rs.)	4.5	6.8	7.9	8.8	9.0	9.4	12.7	11.2

Estimate the parameters of semilogarithmic Engel Curve by passing a straight line by judgment through the graph on arith-log scale. Obtain the elasticity at the average of total expenditure, viz., Rs 17.20. Use any method to estimate the aggregate demand for cereals at a future date when population would have risen by 10 percent, and total domestic expenditure by 20 percent. (State the assumptions you make). [25]

70

INDIAN STATISTICAL INSTITUTE
Research and Training School
B. Stat. Part IV: 1966-67

MID-YEAR EXAMINATION

Statistics-8: Demography Theory and Practical

Date: 24.12.66 Maximum marks: 100 Time: 3 hours

1. Explain the terms (a) Total Fertility rate (b) General Fertility rate (c) Net Reproduction rate. [3× 5] =[15]

2.(a) Explain clearly the different stages which led to the formulation of the law of population growth:

$$P = \frac{L}{1 + e^{(\beta - t)/\alpha}}$$

where L, β, α, t have their usual significance. [8].

(b) Describe one method of fitting the above growth curve to the population data. [7]

(c) Discuss the law of growth of population with reference to India, given the following table.

Year	1891	1901	1911	1921	1931	1941	1951	1961
Population of Indian Union (in millions)	236	235	249	248	276	313	356	436

[5]

3.(a) Given

Age x	40	45	50	55	60
Survival probability at age x (p_x)	.98206	.97677	.97039	.96164	.95036

using Gompertz Law, estimate the survival probabilities at ages 70, 80, 90. [15]

(b) Calculate the force of mortality at age 42 (μ_{42}) from the table given below.

Age x	Life table survivors at age x (l_x)
40	59868
41	58794
42	57679
43	56525
44	54110

[10]

(c) The following table gives the female population and the total children born to them for different age groups in a certain year and also the number of females in the stationary population. It is also given that male births: female births = 51.3 : 48.7. Calculate the net reproduction rate from the data.

Please Turn Over

Age group	No. of women	No. of births	Females in the stationary population
15 - 19	225,927	29,074	28,949
20 - 24	254,851	63,850	27,763
25 - 29	232,731	54,766	26,242
30 - 34	217,017	42,644	24,205
35 - 39	172,273	25,858	21,818
40 - 44	170,932	12,004	19,331
45 - 49	127,761	2,727	16,855

$\ell_0 = 10,000$ [15]

4.(a) Define 'underlying cause of death' (WHO) and discuss its importance in the field of public health- [8]

(b) Briefly discuss the uses of vital statistics. [7]

(c) It was stated in the Annual Report (1899) of the American Secretary of War that the death rate from disease among the American soldiers stationed in Phillipines was 17.2 per 1000. As this death rate was more or less similar to the one prevailing among the populations of the cities of Washington and Boston, it was argued that mortality was not excessive among the soldiers.

Comment on the above argument. [10]

INDIAN STATISTICAL INSTITUTE
Research and Training School
B. Stat. Part IV:1966-67

MID-YEAR EXAMINATION

Statistics-B: Educational Statistics Theory and Practical

Date: 28.12.66 Maximum marks: 100 Time: 3 hours

Answer Question 1 and any five of the rest.

1.(a) What do you mean by standard score? For what purpose are raw scores converted to standard ~~errors~~ scores? How would you calculate the normalised score and stanine grades? [10]

(b) The scores of two students on five tests A, B, C, D and E are presented below, along with the means and standard deviations of the scores on these tests of the group to which they belong.

Test / Student	A	B	C	D	E
a	28	26	30	17	35
b	15	32	15	32	41
Mean	22	15	28	33	26
s.d.	4	6	8	5	7

Give the rank order to each student in the five tests first in terms of raw score, then in terms of z-score. Explain the discrepancies in rank order. Derive the conversion equations for transforming scores in test A and test D into a scale that would give a mean of 100 and a standard deviation of 15. [10]

2.(a) What are the different methods of estimating reliability of a test? Discuss each method briefly. [8]

(b) Let x and y be the scores on two parallel tests and

$$V(x-y) = 73.28$$
$$V(x+y) = 841.56.$$

What is the reliability of $(x+y)$ score?
What is the reliability of x or y score? [8]

3.(a) What are parallel tests? What are the hypotheses H_{mvc}, H_{vc} and H_m? [6]

(b) A test consisting of 5 items was administered to a large number of subjects. The following table gives the value of

p_i = proportion of subjects answering the ith item correctly.

p_{ij} = proportion of subjects answering the ith and the jth items correctly.

Let X denote the total number of items correctly answered by a subject. Compute the mean and variance of X.

Item	Value of p_{ij}					p_i
	1	2	3	4	5	
1	-	-	-	-	-	0.80
2	0.68	-	-	-	-	0.73
3	0.58	0.48	-	-	-	0.65
4	0.40	0.38	0.25	-	-	0.44
5	0.42	0.35	0.32	0.40	-	0.53

[10]

4. Write short notes on <u>any four</u> of the following:
   a) Intelligence quotient.
   b) Correction for attenuation.
   c) Factor loadings and factor scores.
   d) Standard error of measurement .
   e) Speed and power test.
   f) Coefficient of discrimination of a test. [4 X 4]=[16]

5.(a) What do you mean by item difficulty and item validity? Briefly describe the biserial and point biserial correlation coefficients. [8]

(b) Show how you can estimate the reliability of a test from the test variance, the variances of the items and the number of items in the test. Clearly state the assumptions involved. [8]

6.(a) Determine the reliability of the difference score obtained from the following pair of tests A and B where

reliability of test A = .90
reliability of test B = .80
correlation between tests A and B = .70.

Assume that the variances of test A and B are both equal to 1. [8]

(b) Let S_Y^2 and s_x^2 be the variances of explicit variable X in the extended and the curtailed group respectively.
Let S_Y^2 and s_y^2 be the variances of incidental variable Y. in the extended and the curtailed group respectively.
r_{xy} = the correlation between X and Y calculated on the basis of the curtailed group.

Show that $\left(\frac{S_Y^2}{s_y^2} - 1 \right) = r_{xy}^2 \left(\frac{S_X^2}{s_x^2} - 1 \right)$

after clearly stating the assumptions involved. [8]

7.(a) Suppose a test contains p items and a sub-test is formed with q (q < p) items out of this test. If r be the reliability of each item show that

$$r_{pq} = \sqrt{\frac{q}{p} \; \frac{1 + (n-1)r}{1 + (q-1)r}}$$

where r_{pq} is the correlation between the test and the sub-test. Assume items to be parallel to one another. [8]

(b) A test has a validity coefficient of .50 and reliability coefficient of .90. The criterion has a reliability coefficient of .70. Estimate the validity when the test length is doubled and the criterion length is increased three times. What is the maximum amount of validity that could exist in this situation? [8]

INDIAN STATISTICAL INSTITUTE
Research and Training School
B. Stat. Part IV:1966-67

PERIODICAL EXAMINATIONS

Statistical Methods (Theory)

Date: 27.2.67 Maximum marks: 100 Time: 3 hours

Answer any three questions.

1.(a) Define Wilks Λ criterion.

(b) Show that the null distribution of Wilk's Λ is the same as that of the product of several independent Beta variables.

(c) Hence derive the t-th raw moment of Λ and the moment generating function of $\log_e \Lambda$.

2. Derive the density function of the Wishart distribution for the case where the population dispersion matrix is nonsingular.

3. Show that

(a) if the elements of a square matrix S of order p have a Wishart distribution $W_p(K, \Sigma, .)$, B be a fixed matrix of order p then the elements of BSB' also have a Wishart distribution;

(b) if the elements of square matrices S_1 and S_2 have independent Wishart distributions $W_p(K_1, \Sigma, .)$ and $W_p(K_2, \Sigma, .)$ respectively then the distribution of $S_1 + S_2$ is also Wishart ;

(c) if the elements of S have a Wishart distribution $W_p(K, \Sigma, .)$ and L be a fixed vector then $L'SL/L'\Sigma L$ is distributed as χ^2 with K d.f. If W_p is central then so is χ^2 ;

(d) if U' be a matrix whose columns (K in number) are distributed as independent p-variate normal $N_p(\mu_i, \Sigma)$ i=1,2,..., k, then a necessary and sufficient condition for T = U'AU to have a Wishart distribution $W_p(K, \Sigma, .)$ is that for some fixed vector L, $L'TL/L'\Sigma L$ is χ^2 with K d.f. and that T is $W_p(K, \Sigma)$, a central Wishart if and only if for each L (with $L'\Sigma L > 0$), $L'TL/L'\Sigma L$ is a central χ^2.

(e) If $\begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix}$ has a Wishart distribution $W_p(K, \Sigma)$ and $\Sigma = \begin{array}{c} p_1 \\ p_2 \end{array} \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$ with $|\Sigma_{11}| \neq 0$ then $S_{22} - S_{21}S_{11}^{-1}S_{12}$ has the Wishart distribution

$W_{p_2}(K - p_1, \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12})$.

Please Turn Over

4\. Suppose a given set of measurements (represented by the vector U) be distributed as $N_p(\mu_1, \Sigma)$ and $N_p(\mu_2, \Sigma)$ in two alternative populations.

(a) Derive the expression for the best linear discriminant function (BLDF) for discriminating between these two populations.

(b) Show that if L be the BLDF and T be a linear function of the measurements which is uncorrlated with L then the distribution of T is identical in both the populations.

(c) Hence suggest a test procedure for testing if a given linear function provides the BLDF, on the basis of samples drawn from these two population. and derive the null distribution for the suggested test criterion.

INDIAN STATISTICAL INSTITUTE
Research and Training School
B. Stat. Part IV: 1966-67

PERIODICAL EXAMINATIONS

Statistical Inference

Date: 6.3.67 Maximum marks: 100 Time: 3 hours

1.(a) Let the family of densities $f(x;\theta)$, $\theta \subset \Omega$, have monotone likelihood ratio in $T(x)$, and suppose that the c.d.f. of $T(x)$ is a continuous function for each fixed θ. Obtain the UMP level- α test for testing the hypothesis $H : \theta = \theta_0$ against the alternative $K : \theta > \theta_0$. [5]

(b) Show that there exists a uniformly most accurate lower confidence bound $\underline{\theta}$ for θ at each confidence level $1-\alpha$ for the situation described in Question 1(a). How do you obtain $\underline{\theta}$ given the value of α ? [9 + 3]= [12]

(c) State the UMP unbiased level α test for testing $\theta = \theta_0$ against $\theta \neq \theta_0$ on the basis of a random sample of size n from $N(\theta, 1)$. Use this to obtain the shortest (in Neyman's sense) unbiased confidence bound for θ. [3 + 9]=[12]

2. EITHER

(a) State and prove the Cramer-Rao inequality mentioning the regularity conditions involved. Obtain the Chapman-Robbins lower bound for the variance of an unbiased estimate. [12 + 8]=[20]

OR

(b) State and prove the Rao-Blackwell theorem along with its extension to multiparameter case. [10 + 10]= [20]

3. EITHER

(a) How do you define a consistent sequence of estimators ? [4]

(b) Let $\{T_n\}$ be a sequence of estimators with

$$E(T_n) = \theta + b_n, \quad V(T_n) = \sigma_n^2 .$$

Prove that T_n tends to θ in probability as $n \to \infty$, given that $b_n \to 0$ and $\sigma_n^2 \to 0$ as $n \to \infty$. [12]

(c) Let $\{X_n\}$ be a sequence of i.i.d. random variables the common density being

$$g(x ; \theta) = \frac{1}{\theta} \exp(-x/\theta), \text{ if } x \geq 0$$
$$= 0, \text{ otherwise.}$$

Obtain a consistent sequence of estimators for estimating θ. [5]

OR

(d) Show that a necessary and sufficient condition for an estimator T to be a best unbiased estimator of a parametric function is that the covariance between T and every unbiased estimator of zero having finite variance is 0. [11]

Please Turn Over

OR (contd.)

(e) consider a random variable X with the following probability distribution: For $0 < \theta < 1$,

$$P_\theta(X = -1) = \theta; \; P_\theta(X = k) = (1 - \theta)^2 \theta^k,$$

$$k = 0, 1, 2, \ldots$$

Find $E_\theta(X)$, and show that all unbiased estimators of 0 are of the form cX. Using the result in 3(d), prove that the following is the best unbiased estimate of $(1 - \theta)^2$:

$$T(X) = 1, \quad \text{if } X = 0$$
$$= 0, \quad \text{if } X \neq 0.$$

[3 + 4 + 3]=[10]

4. EITHER

(a) Give an example to show that the variance of the best unbiased estimator can be strictly greater than the corresponding Cramer-Rao lower bound. [10]

(b) Let $X_1, X_2, \ldots, X_n$ be i.i.d. random variables, with the common density

$$g(x; \theta) = \exp(\theta - x), \quad \text{if } x \geq \theta$$
$$= 0, \quad \text{otherwise.}$$

Let $T(x_1, \ldots, x_n) = \min(x_1, \ldots, x_n)$. Show that the density of T is

$$h(t; \theta) = ne^{n(\theta - t)} \quad \text{if } t \geq \theta$$
$$= 0, \quad \text{otherwise.}$$

Show that $(nT-1)/n$ is an unbiased estimator of θ but its variance is less than the corresponding Cramer-Rao lower bound. [5+4 + 7]=[16]

OR

(c) On the basis of random samples of sizes n_1 and n_2 from $N(\theta, \sigma_1^2)$ and $N(\theta, \sigma_2^2)$, respectively, obtain the least-square estimate of θ assuming that σ_1 and σ_2 are known. Is this the best unbiased estimate of θ? [7 + 7] = [14]

(d) On the basis of a random sample $X_1, \ldots, X_n$ from the Poisson distribution with mean θ, it is desired to estimate $\exp(-\theta)$. Show that the following estimator is unbiased for $\exp(-\theta)$

$$h(x_1, \ldots, x_n) = 1, \quad \text{if } x_1 = 0$$
$$= 0, \quad \text{if } x_1 \neq 0.$$

Prove that the conditional probability distribution of X_1, given $X_1 + \ldots + X_n = t$, is the binomial distribution $B(t; 1/n)$. Use the above results to show that the best unbiased estimate of $\exp(-\theta)$ is

$$\left(\frac{n-1}{n}\right)^{X_1 + \ldots + X_n}$$

(2 + 6+4]= [12]

For neatness and clarity : [4].

INDIAN STATISTICAL INSTITUTE
Research and Training School
B. Stat. Part IV: 1966-67

PERIODICAL EXAMINATION

Statistics-7: Econometrics Theory and Practical

Date: 13.3.67 Maximum marks:100 Time: 3 hours

Group A

Answer any two questions from this group.

1.(a) Discuss the different properties of the Cobb-Douglas form of the production function. [13]

(b) How are the different variables in this production function measured in different types of applications? [12]

2.(a) What are the main criticisms of the common applications of the Cobb-Douglas production function? [15]

(b) What could be done to meet these criticisms? [10]

3.(a) Discuss the concept of the elasticity of substitution between factors used in production. [15]

(b) Explain how the CES production function embraces, among other things, the input-output model and the Cobb-Douglas form of the production function. [10]

4. Write short notes on any two:

(a) The cost function
(b) The Supply Function
(c) Multicollinearity. [25]

Group B

5. The following shows the indices of (i) volume of industrial production (ii) labour input in man-years and (iii) machine capacity used in industry in Finland during 1943-1952:

year	indices (base: 1925)		
	output	labour input	machine capacity
1943	2.030	1.383	2.996
1944	1.909	1.366	2.830
1945	1.987	1.624	2.871
1946	2.409	1.765	2.915
1947	2.686	1.875	2.953
1948	3.049	1.948	3.095
1949	3.202	1.956	3.522
1950	3.401	1.970	3.877
1951	3.947	2.136	4.274
1952	3.788	2.053	4.402

Fit a Cobb-Douglas production function and test whether returns to scale are constant. [35 + 15] = [50]

INDIAN STATISTICAL INSTITUTE
Research and Training School
B. Stat. Part IV: 1966-67
PERIODICAL EXAMINATION

Statistics-6: Design of Experiments Theory and Practical

Date: 20.3.67. Maximum marks: 100 Time: 3 hours

1.(a) State three targets of comparative experiments. [3]

(b) Explain 'Randomisation' stating two purposes served by it, with an example. [4]

(c) What is sensitivity of an experiment ? Illustrate with examples how it can be increased by

i) qualitative methods and
ii) quantitative methods. [5]

2. In a garden, r of the N plants are chosen at random for which a treatment T_1 is given and for the remaining a different treatment T_2 is given. The yields are observed after a period of two months. Obtain an unbiased estimate of variance for estimating the difference of treatment effects by their observed mean yields. [20]

3.(a) Discuss relative merits of factorial experiments and one factor at a time experiments

i) when some effects are confounded and
ii) when none is confounded. [8]

(b) Obtain a plan for conducting a 2^5 experiment in blocks of 2^3 units each so that no main effect is confounded. [20]

(c) Show that allowing for a block size of 16 units, 15 factors each at two levels can be tested without confounding any interaction of less than three factors. [10]

4. What design do you suggest and why, given the following information? Eight roasts can be cut from each of 4 animals. The experimenter wishes to study the effect of freezing, length of freezing, storage temperatures and length of storage upon tenderness of roasts and to use the following in all combinations:

storage temperatures:	10°c	and	15°c
lengths of storage:	20	and	40 days
Freezing temperatures:	0°c	and	-10°c
length of freezing:	5	and	10 days.

Give the lay-out of the design and break down the total degrees of freedom. [1 + 1 + 10 + 8]=[20]

5. Practical Records. [10]

INDIAN STATISTICAL INSTITUTE
Research and Training School
B.Stat. Part IV: 1966-67

143

PERIODICAL EXAMINATION

Statistics-7: Industrial Statistics Theory and Practical

Date: 27.3.67. Maximum marks: 100 Time: 3 hours

Attempt any five questions.
All questions carry equal marks.

1.(a) Explain what is meant by a production process being in a state of statistical control? [3]

(b) The following table provides the results on some measurable characteristic for 15 samples of size 4, drawn in order of production.

Sample	$\bar{x}$	R
1	0.7540	0.0011
2	0.7542	0.0014
3	0.7542	0.0009
4	0.7546	0.0010
5	0.7550	0.0008
6	0.7539	0.0009
7	0.7541	0.0012
8	0.7543	0.0011
9	0.7547	0.0007
10	0.7549	0.0015
11	0.7541	0.0017
12	0.7542	0.0010
13	0.7545	0.0011
14	0.7548	0.0009
15	0.7551	0.0012

The specification was as 0.7524-0.7565.

(i) Setup modified $\bar{x}$ chart and an R-chart. [12]

(ii) Explain the advantage of a modified $\bar{x}$ chart and indicate alternative decisions which might be taken to take advantage of the above situation. [5]

2.(a) An item is inspected for visual defects. The defects are classified as minor and major. The major and minor defects are distributed with mean 0.5 and 1.5 respectively. The item will be accepted if the number of major defects in it does not exceed one or if the total number of defects does not exceed 5. Calculate the probability of acceptance for an item selected at random. [12]

(b) Suggest a suitable method to combine minor and major defects and therefrom give a control procedure for the total number of defects observed per item. [8]

3.(a) i) State the conditions for the applicability of group control chart. [2]

ii) How would you proceed to construct such a chart? [3]

iii) What are the advantages of a group control chart? [2]

(b) A control chart analysis indicates that the standard deviations of the distributions of dimensions of two mating parts, A and B, are 0.0008 inch and 0.0020 inch respectively. It is desired that the probability of a smaller clearance than 0.002 inch should be 0.005.

Please Turn Over.

3.(b) contd.

i) What distance between the average dimensions of A and B should be specified by the designer? [8]

ii) With this distance specified, what is the probability that two parts assembled at random will have a greater clearance than 0.012 inch? [5]

(Assume normal distribution and random assembly).

4.(a) - Explain in brief the terms

i) OC,
ii) AQL,
iii) Producer's risk,
iv) LTPD,
v) consumer's risk and
vi) AOQL for a single sampling acceptance rectification plan. [6]

(b) Write the general mathematical expression for O.C. of a double sampling plan. [4]

(c) Using Biometrika tables construct a single sampling plan having operating characteristic curve passing through the points (0.020, 0.95) and (0.061, 0.05). [10]

5.(a) Write a brief note on published Standard for acceptance sampling for attributes. [8]

(b) Construct a suitable double sampling plan with AQL of 1.5 percent with producer's risk 5 percent, LTPD of 5 percent with consumer's risk of 10 percent under

i) Tightened inspection. [6]
ii) Reduced inspection. [6]

6.(a) Discuss briefly the advantages and disadvantages of acceptance sampling by variables. [2]

(b) For an item, having upper specification U for a certain characteristic, a one sided variable sampling plan is to be used. The standard deviation is given to be σ. The plan requirements are such that a lot containing 100 p_1 percent defective will be rejected with a small probability α and a lot containing 100 p_2 percent defective will be accepted with a small probability β.

i) Derive the general expressions required to construct the plan. [15]

ii) Evaluate the suitable plan for

p_1 = .03 $\alpha = 0.05$
p_2 = .08 $\beta = 0.10$. [3]

INDIAN STATISTICAL INSTITUTE
Research and Training School
. B. Stat. Part IV:1966-67

144

PERIODICAL EXAMINATION

Statistics-4: Probability

Date:10.4.67. Maximum marks: 100 Time: 3 hours

Attempt all questions.

1. Let (p_{ij}), $i, j = 1,2,\ldots$ be the transition matrix of a Markov chain with stationary transition probabilities. Prove that for all i, j

$$\lim_{N \to \infty} \frac{1}{N} \sum_{n=0}^{N-1} p_{ij}^{(n)}$$

exists. If the limit is denoted by π_{ij}, show that

$$\sum_{k=0}^{\infty} p_{ik}\, \pi_{kj} = \sum_{k=0}^{\infty} \pi_{ik}\, p_{kj} = \sum_{k=0}^{\infty} \pi_{ik}\, \pi_{kj} = \pi_{ij}$$

for all i, j and that

$$\sum_{j=0}^{\infty} \pi_{ij} \leq 1$$

for all i. [30]

2. Let C be an essential class of period d. Show that the only solutions of the system of equations

$$u_i = \sum_{j \in C} p_{ji}\, u_j \quad , \qquad i \in C$$

such that $\sum_{i \in C} |u_i| < \infty$ are of the form $u_i = e\pi_i$, $i \in C$ for some constant e. [25]

3. Consider a Markov chain whose stationary transition probabilities are given by

$$p_{00} = r_0, \qquad p_{01} = p_0$$

$$p_{i,i-1} = q_i,\ p_{i,i} = r_i,\ p_{i,i+1} = p_i \quad \text{for } i=1,2,\ldots$$

where $p_0 + r_0 = 1$, $p_i + r_i + q_i = 1$ for $i = 1,2,\ldots$ and $0 < p_i, q_i, r_i < 1$ for all i.

Prove that the chain is positive recurrent if and only if

$$\sum_{k=1}^{\infty} \frac{p_0 \cdots p_{k-1}}{q_1 \cdots q_k} < \infty. \qquad [20]$$

4. Prove that, for any countable state Markov chain with stationary transition probabilities, if i is recurrent and $i \rightarrow j$, then j is also recurrent. [15]

5. Prove that a finite irreducible Markov chain with stationary transition probabilities is aperiodic if and only if there exists a $n > 0$ such that

$$p_{ij}^{(n)} > 0$$

for all i, j. [10]

INDIAN STATISTICAL INSTITUTE
Research and Training School
B.Stat. Part IV: 1966-67

205

ANNUAL EXAMINATION

Statistics-4: Probability

Date: 23.5.67. Maximum marks: 100 Time: 3 hours

The number of marks allotted to each question is given in brackets [ ].

Answer Groups A and B in separate answerscripts.

Group A

Answer any two questions.

1.(a) Define a two-dimensional symmetric random walk. Describe clearly the state space and the 1-step transition probabilities. [3]

(b) Prove that every state of a two-dimensional symmetric random walk is null recurrent. [10]

(c) In a two-dimensional symmetric random walk starting at the origin, let

$$D_n^2 = x^2 + y^2$$

be the square of the distance of the particle from the origin at time n. Prove that $E(D_n^2) = n$.
[Hint: Compute $E(D_{n+1}^2 - D_n^2)$.] [7]

2.(a) Consider a Markov chain with state space $I = \{0,1,2,\dots\}$ and stationary transition probabilities given by the following 1-step transition matrix:

$$\begin{pmatrix} p_0 & p_1 & p_2 & p_3 & \cdots \\ p_0 & p_1 & p_2 & p_3 & \cdots \\ 0 & p_0 & p_1 & p_2 & \cdots \\ 0 & 0 & p_0 & p_1 & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \end{pmatrix}$$

where $\{p_k\}$ is a probability distribution with probability generating function P(s). Prove that the chain is positive recurrent if and only if $P'(1) < 1$. When the chain is positive recurrent, find the probability generating function of its stationary distribution. [14]

(b) Consider a 3-state Markov chain with stationary transition probabilities given by the 1-step transition matrix

$$\begin{pmatrix} 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & 0 \end{pmatrix}$$

Discuss the nature of the states. Compute the stationary distribution, if any. Find the quantities ~~π_{ij}~~

$$\pi_{ij} = \lim_{n \to \infty} \frac{1}{n} \sum_{k=0}^{n-1} p_{ij}^{(k)}$$

for each i, j. [6]

Go on to the next page

3.(a) Consider an irreducible Markov chain with state space $I = \{0, 1, \ldots\}$ and stationary transition probabilities (p_{ij}). Prove that the chain is non-recurrent if and only if the system of equations

$$y_i = \sum_{j=1}^{\infty} p_{ij} \, y_j, \quad i = 1, 2, \ldots$$

admits of a non-zero bounded solution. [15]

(b) In a finite state Markov chain with stationary transition probabilities, prove that a state is positive recurrent if and only if it is essential. [5]

Group B

Answer any three questions

Ω stands for the whole space, $\emptyset$ for the empty set.

4.(a) Define a field of subsets of Ω. [2]

(b) Define a σ-field of subsets of Ω. [3]

(c) Give an example of a field which is not a σ-field. [5]

(d) Give an example of a finitely additive set function on a field which is not countably additive. [10]

5.(a) Define an outer measure. [2]

(b) Given an outer measure μ^*, define a μ^*-measurable set. [2]

(c) For an outer measure μ^*, prove that $\mathcal{A}^*$, the class of μ^*-measurable sets, is a σ-field and that μ^* restricted to $\mathcal{A}^*$ is a measure. [10]

(d) Give an example of an outer measure. [6]

6.(a) Let φ be a σ-additive set function on a σ-field $\mathcal{A}$ with $\varphi(\emptyset) = 0$. Prove that φ admits a decomposition of the form: $\varphi = \varphi^+ - \varphi^-$, where φ^+ and φ^- are measures. [You may assume the existence of sets $C, D \in \mathcal{A}$ such that $\varphi(C) = \sup_{A \in \mathcal{A}} \varphi(A)$ and $\varphi(D) = \inf_{A \in \mathcal{A}} \varphi(A)$.] [12]

(b) Prove that a finitely additive set function φ on a σ-field $\mathcal{A}$ with $\varphi(\Omega) < \infty$ is countably additive if and only if φ is continuous from above at $\emptyset$. [8]

7.(a) Define the Borel σ-field of subsets of the real line. [3]

(b) Let $\mathcal{J}_0$ denote the class of sets of the form $[a, b)$ with $-\infty < a \leq b < +\infty$. Define a set function μ on $\mathcal{J}_0$ by : $\mu([a, b)) = b-a$. Show that μ is a measure on $\mathcal{J}_0$. [11]

(c) Let $Tx = x+c$ for all real x, where c is a fixed real number. Prove that if

i) E is a Borel set, so is TE,

ii) E is Borel, then $\mu(E) = \mu(TE)$, where μ is Lebesgue measure on the Borel σ-field. [6]

206

INDIAN STATISTICAL INSTITUTE
Research and Training School
B.Stat. Part IV: 1966-67

ANNUAL EXAMINATION

Statistics-4: Inference

Date: 24.5.67. Maximum marks: 100 Time: 3 hours

The number of marks allotted to each question is given in brackets [ ].

Answer Groups A and B in separate answerscripts.

Group A

Answer any two questions.

1.(a) Define the following terms with illustrations:
  i) Sufficient statistic.
  ii) Complete family of distributions.
  iii) Test of Neyman structure. [3 X2]=[6]

(b) Obtain a necessary and sufficient condition for all similar tests to be of Neyman structure. [9]

(c) On the basis of a random sample of size n from $N(\theta, \sigma^2)$ obtain the UMP unbiased level α test for testing $\theta = 0$ against $\theta > 0$, σ^2 being assumed to be unknown. [10]

2.(a) Explain how the principle of confidence region is connected with testing of hypothesis. Define uniformly most accurate confidence bounds and shortest unbiased confidence region (in Neyman's sense). [12]

(b) Let x_1 and x_2 be two independent observations from the rectangular distribution:

$$f(x;\ \theta) = \frac{1}{\theta}, \text{ if } 0 \leq x \leq \theta$$
$$= 0, \text{ otherwise.}$$

Find out the confidence interval for θ of confidence coefficient $1-\alpha$ from the following procedures of choosing the critical region $w(\theta_0)$ of size $1-\alpha$, for testing $H_0 : \theta = \theta_0$.

Procedure I . $w(\theta_0) : |x_1 + x_2 - \theta_0| > \Delta$

where Δ is to be chosen so that $w(\theta_0)$ is of size α for H_0.

Procedure II: $w(\theta_0) : q\,\theta_0 > L$ or $L > \theta_0$

where q is to be chosen suitably so that $w(\theta_0)$ is of size α for H_0; $L = \max(x_1, x_2)$.

By computing $P_\theta\,[w(\theta_0)]$ in both the cases, or otherwise, show that Procedure II provides a shorter (in Neyman's sense) confidence region for θ than Procedure I. [13]

3.(a) Explain the 'maximum likelihood' method of point estimation with illustrations. [3]

(b) Under suitable regularity conditions (to be stated), state and prove the well-known asymptotic properties of a maximum likelihood estimate based on sequence of i.i.d. random variables. [16]

Go on to the next page

3.(c) Give an example where $\sqrt{n}\,(\hat{\theta}_{(n)} - \theta)$ is not asymptotically normally distributed, when $\hat{\theta}_{(n)}$ is the maximum likelihood estimate of θ based on a random sample of size n. [6]

4.(a) State and prove Cramér-Rao inequality mentioning the regularity conditions involved. [8]

(b) Let $X_1, \ldots, X_m$ be a random sample from $N(\xi, \sigma^2)$ and $Y_1, \ldots, Y_n$ be a random sample from $N(\xi, \tau^2)$.

i) Show that the best unbiased estimate of ξ, when σ^2 and τ^2 are known, is

$$t(x_1, \ldots, x_m, y_1, \ldots, y_n;\ \sigma^2, \tau^2)$$

$$= [\frac{1}{\sigma^2} \sum_{i=1}^{m} x_i + \frac{1}{\tau^2} \sum_{i=1}^{n} y_i] / (\frac{m}{\sigma^2} + \frac{n}{\tau^2}).$$ [8]

ii) Show that there does not exist a uniformly minimum variance unbiased estimate of ξ when σ^2 and τ^2 are not known but such an estimate exists if only the ratio σ^2/τ^2 is known. [3+6]=[9]

Group B

Answer any two questions.

5.(a) i) Describe Wald's sequential probability ratio test for testing a simple hypothesis against a simple alternative. [4]

ii) Illustrate the method for testing $\theta = \theta_0$ against $\theta = \theta_1$ based on random samples from $N(\theta, 1)$. [5]

(b) How are the constants involved in SPRT related with the probabilities of the two kinds of error? [3]

(c) Show that Wald's SPRT terminates with probability 1 under both the hypothesis and the alternative. [13]

6. Let m(F) be a median of a distribution whose c.d.f. is F.

(a) Obtain the UMP level α test for testing $m(F) = 0$ against $m(F) > 0$ based on a random sample of size n from the distribution F. [10]

(b) Is the above test consistent? [5]

(c) Obtain an unbiased confidence interval for m(F) with confidence level $1-\alpha$. [10]

7. Write short notes on any three of the following:

(a) Tolerance region.
(b) U statistic.
(c) Rank-sign test.
(d) Stein's two-sample test.
(e) Kolmogorov-distance test. [25]

INDIAN STATISTICAL INSTITUTE
Research and Training School
B.Stat. Part IV: 1966-67

[207]

ANNUAL EXAMINATION
Statistics-5: Statistical Methods Theory and Practical

Date: 25.5.67. Maximum marks: 100 Time: 4 hours

The number of marks allotted to each question is given in brackets [ ].

Answer Groups A and B in separate answerscripts.

Group A

Answer any three questions.

1.(a) Let the n observations in a sample be classified in k classes. Denote the observed frequency in class i by O_i, and the corresponding expected frequency by E_i $(i=1, 2, \ldots, k)$. Show that under certain conditions which you are required to state in full, the statistic $T = \sum_{i=1}^{k} (O_i - E_i)^2/E_i$ is, in the limit as $n \to \infty$, distributed as chisquare with k-1 d.f. [10]

(b) Suppose some of the classes in (a) above are merged together to form a new system involving fewer (k') classes. If the observed and expected frequencies in the ith new class be denoted by O_i' and E_i' respectively and if

$$T' = \sum_{i=1}^{k'} (O_i' - E_i')^2/E_i'$$

show that as $n \to \infty$, T-T' is distributed in the limit as chisquare with k-k' d.f. Suggest some practical application of this result. [6]

2.(a) Explain the term 'standard error of a statistic'. [3]

(b) Obtain the expression for the standard error of the square root of sample β_1. [8]

(c) The sample β_1 was computed as 0.45 in a sample of size 150. Can this be considered as sufficient evidence of non-normality of the population sampled? [5]

3.(a) Show that under certain conditions the sample quartile Q_1 has asymptotically a univariate normal distribution in a sense which you are required to explain in full. [10]

(b) How does the average of first and third quartiles compare with the sample mean as an estimator of the population mean, in samples from a normal population? [6]

4. In many text books on Sample Surveys theory and methods the authors on large sample considerations make use of the normal approximation to the sampling distribution of the mean even though it is computed from a sample drawn without replacement from a finite population. What should be the nature of the population so that the approximation could be applied with some degree of confidence. State and prove any theorem that you may need to defend your statement. [16]

Go on to the next page

Group B

Answer any two questions from Questions 5 to 7

5. Given below are the mean values of 3 characters for 3 castes of Uttar Pradesh.

Caste	Sample size	mean values head length x_1	head breadth x_2	bizygomatic breadth x_3
Brahmin	86	191.92	139.88	133.36
Chhattri	139	192.58	131.72	131.70
Bhil	187	181.87	137.62	131.18

The within dispersion matrix computed from these 412 observations are

$$\begin{array}{cccc} x_1 & 43.65 & 5.89 & 8.44 \\ x_2 & & 20.25 & 11.14 \\ x_3 & & & 20.98 \end{array}$$

Examine if these three castes differ significantly among themselves in respect of measurements x_1, x_2 and x_3. [16]

6. The following table gives the observed frequency of different combinations of colour and pollen shape in sweat pea in a certain genetical experiment. Given in brackets are the relative frequencies, expected on the basis of a certain genetical theory, expressed as functions of an unknown parameter θ, $0 \leq \theta \leq 1$.

Pollen shape	Colour Purple	Red
Long	$296\ (\frac{1}{2} + \frac{\theta}{4})$	$27\ (\frac{1}{4} - \frac{\theta}{4})$
Round	$19\ (\frac{1}{4} - \frac{\theta}{4})$	$85\ \frac{\theta}{4}$

Test if:

(a) the data are in agreement with the theory. [8]
(b) $\theta = 0.80$. [8]

7. For a certain normal population with unknown mean μ and unknown variance σ^2 the proportion of observations greater than U is denoted by P. Note that if the upper 100P ·/· point of the standard normal distribution be λ_p, we have $U = \mu + \lambda_p\sigma$.

(a) If $\bar{x}$ and s be the mean and the standard deviation computed from a sample of size n drawn from the above population, using large sample approximations derive an expression for the probability π_p that $\bar{x} + ks$ would exceed the value U. [6]

(b) Hence determine numerically the values of n and k that would satisfy the equations

$$\pi_{.01} = .05$$
$$\pi_{.05} = .95.$$ [4]

(c) Obtain an exact expression for π_p in terms of the concentral t probability integral. [6]

8. Viva Voce [10]

9. Practical Record [10]

INDIAN STATISTICAL INSTITUTE
Research and Training School
B.Stat. Part IV: 1966-67

[208]

ANNUAL EXAMINATION

Statistics-6: Design of Experiments Theory and Practical

Date: 26.5.67 Maximum marks: 100 Time: 4 hours

The number of marks allotted to each question is given in brackets [ ].

Answer Groups A and B in separate answerscripts.

Group A

1.(a) 'Randomisation is a method by which every experimental unit has an equal chance of receiving a treatment'. Discuss this statement in the context of different designs. [6]

(b) What is sensitivity of an experiment and how can it be increased without increasing the size of the experiment? [1+2]=[3]

2.(a) Explain 'total confounding' and 'partial confounding'. [2]

(b) Four surface treatments each at 2 different intensities of application are to be studied in all combinations with respect to the durability of surface of motor car tyres. It is decided to consider each tyre to be of homogeneous material and different tyres to be heterogeneous. On four consecutive parts on the surface of a tyre any four of the 2^4 treatment combinations can be applied. You are allowed to use as many automobiles as you need. Interest lies in information on all the treatment combinations. Write down a plan for conducting the experiment with the smallest number of cars which consists of driving the car for 100 miles on a standard road and noting down the difference in resistences of the surface before and after the experiment. [10]

(c) Present a mathematical model for analysing the above experiment stating the assumptions. [2]

3.(a) Discuss the technique of analysis of covariance. [2]

(b) 'Missing plot technique is a scientific way of getting back the lost data'. Criticize this statement. [2]

(c) Discuss long term experiments. [2]

(d) Explain the role of transformations on experimental data. [2]

4.(a) Define a balanced incomplete block design. [1]

(b) State and prove Fisher's inequality for balanced incomplete block designs. [8]

(c) Define an association scheme with m associates. [2]

(d) Obtain an estimate of treatment effect in a balanced incomplete block design stating your assumptions on the mathematical model for the yields. [8]

Go on to the next page

Group B

5. Seeds of a variety are stored at two different temperatures namely 20°c and 30°c for a month before sowing. Potash is applied at four different levels in the field:

0.25 lb./plot : p_1 ; 0.50 lb./plot : p_2
1.25 lbs./plot: p_3 ; 2.00 lbs./plot: p_4

The plan and yields are given below:

Seeds treated at 20°c

Block 1		Block 2		Block 3		Block 4	
315	p_1	265	p_2	195.	p_2	85	p_4
265	p_2	273	p_3	219	p_1	180	p_3
252	p_4	150	p_1	205	p_4	190	p_2
165	p_3	215	p_4	305	p_3	165	p_1

Seeds treated at 30°c

Block 5		Block 6		Block 7		Block 8	
195	p_4	180	p_2	175	p_1	305	p_4
265	p_3	175	p_1	180	p_3	192	p_2
150	p_1	165	p_3	185	p_2	250	p_1
305	p_2	260	p_4	195	p_4	295	p_3

(a) Write down the confounded effects. [1]

(b) Show the decomposition of total degrees of freedom. [2]

(c) What storage temperature do you recommend to obtain more yield and give the precision of the estimate of higher yield. [8]

(d) Is there interaction between storage temperatures and levels of potash significant? [6]

(e) Test if the highest and lowest levels of potash have different effects on the yield. [6]

6. The following is a field plan for a 2^3 factorial experiment.

I		II		III	
ac	c	ab	c	(1)	bc
(1)	abc	a	(1)	ab	a
bc	b	abc	bc	c	b
ab	a	ac	b	abc	ac

(a) What effects are confounded? [3]

(b) Write down the analysis of variance table. [4]

7. Viva Voce [10]

8. Practical Record [10]

INDIAN STATISTICAL INSTITUTE
Research and Training School
B.Stat. Part IV: 1966-67

209

ANNUAL EXAMINATION

Statistics-7: Industrial Statistics Theory and Practical

Date: 27.5.67. Maximum marks: 100 Time: $3\frac{1}{2}$ hours

The number of marks allotted to each question is given in brackets [ ].

Answer Groups A and B in separate answerscripts.
Npte that you may attempt questions carrying a total of 110 marks.

Group A

Answer any three questions.

1.(a) In a simplex tableau let a_{ij} be the matrix element in the i^{th} row and j^{th} column. Suppose the variable in the r^{th} column enters the basis and that in the k^{th} row departs. Denote by a'_{ij} an arbitrary element in the new matrix. Express a'_{kj} and a'_{ij} $(i \neq k)$ in terms of the old elements a_{ij}. [6]

(b) Maximise $5x - 3y + 2z$

subject to

$$\begin{aligned} 2x + 2y - z &\geq 2 \\ 3x - 4y &\leq 3 \\ y + 3z &\leq 5 \\ \text{and } x, y, z &\geq 0 \end{aligned}$$ [9]

2.(a) State and prove the duality theorem in Linear Programming. Explain clearly the relationships between the final simplex tableaux of the primal and dual problems. [8]

(b) Find the minimum of

$$\begin{aligned} &4x_1 + 3x_2 + 8x_3 \text{ subject to} \\ &x_1 + x_3 \geq 2 \\ &x_2 + 2x_3 \geq 5 \\ &x_1, x_2, x_3 \geq 0 \end{aligned}$$

by solving its dual problem. [7]

3. A furniture company manufactures four models of desks. Each desk is first constructed in the carpentry shop and is next sent to finishing shop, where it is varnished waxed, and polished. The number of man-hours of labour required in each shop is as follows.

	Desk 1	Desk 2	Desk 3	Desk 4
Carpentry shop	4	9	7	10
Finishing shop	1	1	3	40

Because of limitations in capacity of the plant, no more than 6,000 man-hours can be expected in the carpentry shop and 4,000 in the finishing shop in the next 6 months.

The profit (in Rupees) from the sale of each item is as follows:

Desk :	1	2	3	4
Profit (Rs.):	12	20	18	40

Assuming that raw materials and supplies are available in adequate supply and all desks produced can be sold,

Go on to the next page

determine the optimal product mix i.e., the quantities of each type of product to be made, which will maximise profit. [15]

4. Determine $Y' = [x_1, x_2, \ldots, x_{12}]$

to maximise $Z = \sum_{j=1}^{12} c_j x_j$ subject to

$$\begin{array}{llll} x_1+x_2+x_3+x_4+\cdots & & x_{10}+x_{11}+x_{12} & \leq A_1 \\ x_1+x_2+x_3 & +x_5+x_6+x_7+ & +x_9+x_{10}+x_{11} & \leq A_2 \\ x_1+x_2 & +x_5+x_6 & +x_9+x_{10} & \leq A_3 \\ x_1 & +x_5 & +x_9 & \leq A_4 \end{array}$$

$$x_j \geq 0 \quad j = 1, 2 \ldots 12$$

and $c_1 < c_6 < c_{11} < c_4 < c_5 < c_{10} < c_3 < c_8 < c_9 < c_2 < c_{12}$

$$A_1 > A_2 > A_3 > A_4$$ [15]

Group B

Answer any three questions from Questions 5-8

5.(a) Show that in a transportation problem vector p_{ij} corresponding to any non-basic variable x_{ij} can be expressed as $\sum_{\alpha\beta} \pm p^B_{\alpha\beta}$ where $p^B_{\alpha\beta}$ is a vector corresponding to a basis variable $x_{\alpha\beta}$ in the basis and the summation is extended over all $\alpha\beta$ in the basis. [5]

(b) A manufacturer has distribution centres located at Atlanta, Chicago, and New York city. These centres have available 40, 20 and 40 units of his product respectively. His retail outlets require the following number of units; Cleveland 25; Louisville 10; Memphis 20; Pittsburgh 30 and Richmond 15. The shipping costs per unit in dollars between each centre and outlet is given in the following table:

	Cleaveland	Louisville	Memphis	Pittsburgh	Richmond
Atlanta	55	30	40	50	40
Chicago	35	30	100	45	60
New York	40	60	95	35	30

Find an allocation of products from each distribution centre to each retail outlet which minimises shipping costs. Are there other minimum feasible solutions? [10]

6. A law firm has 3 law suits to handle in the near future and 3 lawyers are available to take charge of the suits (one to handle each suit). For each lawyer-suit combination, the firm's president has estimated the probability of winning the suit in favour of their client as follows:

	Suit	I	II	III
Lawyer	1	.3	.2	.1
	2	.7	.5	.4
	3	.8	.7	.4

(a) What assignment maximises the expected number of suits won? [8]

Go on to the next page

(b) If the fees collected for each suit are given by the following table, how should the men be assigned to maximise the expected fees collected.

Suit	Fee (in thousands of Rs.) If won	 If suit lost
I	2	1
II	4	2
III	8	4

[7]

7.(a) Customers arrive at a public telephone booth in a Post office in a Poisson fashion with mean arrival rate λ per minute. The call time of each customer is exponentially distributed with mean rate μ per minute. Show that the density function $\theta(u)$ for the total time an arrival spends in the system (waiting time plus service time) is given by

$$\theta(u) = (\mu-\lambda)\, e^{-(\mu-\lambda)u}.$$

[9]

(b) If $\lambda = 1/12$ and $\mu = 1/4$ in the above problem, then

i) what is the probability that a fresh arrival will not have to wait for the phone?

ii) for what fraction of the time is the phone busy?

iii) what is the probability that an arrival who goes to the post office to make a phone call will take less than 15 minutes to complete his job? [6]

8. A factory has got 4 machines of a special type which are working all the time except when they are under breakdown and repair. There are two mechanics to look after the maintenance of these machines. As soon as a machine breaks down a mechanic attends to its repair if he is free. If several machines breakdown simultaneously they are taken up on the basis of first-in-first-out for repair by mechanics.

The time interval between successive breakdowns of each machine is distributed negative exponentially with an average of 100 hours. Thus if r machines are working at any instant of time, the breakdown rate at that instant is r(1/100) per hour. The time taken for repair of a machine by a mechanic follows the negative exponential distribution with average of 10 hours.

(a) Find out the steady state probabilities P_r, that there are r machines under repair at any instant of time, for r=0, 1,... 4.

(b) What is the percentage of machine time spent on repair on an average? [15]

9. Viva Voce [10]

10. Practical Record [10]

INDIAN STATISTICAL INSTITUTE
Research and Training School
B.Stat. Part IV: 1966-67

210

ANNUAL EXAMINATION

Statistics-7: Econometrics Theory and Practical

Date: 29.5.67.　　Maximum marks: 100　　Time: 4 hours

The number of marks allotted to each question is given in brackets [ ].

Answer Groups A and B in separate answerscripts.

Group A

Answer four questions choosing one from Questions 1 and 2 and any three of the rest.

1.(a) Suppose you have fitted the Cobb-Douglas production function to data from a sample of firms in an industry. Assuming that size distributions of labour, capital and output are lognormal, how would you use the fitted function to estimate total capital in the industry from the corresponding totals of labour and output? [7]

(b) Why and how is time introduced as an additional explanatory variable in the formulation of a production function? What alternative approaches are available for such purposes? [7]

2. Explain fully the concept of elasticity of substitution between two factors used in production and discuss the extreme cases where this elasticity is zero and infinity. [14]

3. Obtain the main results on the generalized least squares method and show that this method is equivalent to the classical least squares method applied to a transformed problem. [12]

4. Explain clearly the problem of identification in a simultaneous structural linear equation model and obtain the order condition of identifiability. [12]

5. Give an account of the two-stage least squares method of estimation of any identified relation in a simultaneous structural linear equation model. [12]

6. Write short notes on any two of the following:

(a) Heteroscedasticity and weighted least squares.

(b) Wald's and Bartlett's methods of grouping when both variables are subject to error.

(c) least squares bias

(d) k-class estimators. [2×6]=[12]

Group B

Answer either Question 7 or Question 8.

7. Utilising the data given below,

(a) fit the regression $Y = \alpha + \beta X$ by ordinary least squares and examine the autocorrelation of residuals. [10]

(b) estimate the same regression working with appropriate autoregressive transformations of both variables. [10]

Go on to the next page

(c) Compare the standard errors of the two estimates of β. [10]

Year	1947	1948	1949	1950	1951	1952	1953	1954	1955
Y	156	163	177	201	216	218	227	238	248
X	352	373	411	441	462	490	529	577	641

8.(a) Consider the following data on personal disposable income and personal consumption in the US in billion dollars at 1954 prices.

Year	1948	1949	1950	1951	1952	1953	1954	1955
Y = consumption	199	204	216	218	224	235	238	256
X = income	212	214	231	237	244	255	257	273

Assume the standard errors-in-variables model where the true values of X and Y are connected by an exact linear relation. Estimate this underlying linear relation assuming that the ratio of the two error variances, i.e., $\sigma^2_{e_X} / \sigma^2_{e_Y}$ is $\frac{1}{2}$. [15]

(b) Examine the identifiability of the relations in the following simultaneous structural equation model:

$$y_{1t} + \beta_{12} y_{2t} = u_{1t}$$

$$\beta_{21} y_{1t} + y_{2t} + \gamma_{21} x_{1t} + \gamma_{22} x_{2t} = u_{2t}$$

with the assumption that $\operatorname{cov}(u_{1t}, u_{2t}) = 0$. [15]

9. Viva Voce [10]

10. Practical Record [10]

INDIAN STATISTICAL INSTITUTE
Research and Training School
B.Stat. Part IV: 1966-67

211

ANNUAL EXAMINATION

Statistics-8: Genetics Theory and Practical

Date: 30.5.67. Maximum marks: 100 Time: 3 hours

The number of marks allotted to each question is given in brackets [ ].

Answer any three questions from Questions 1 to 4.

1. What is meant by Linkage? [30]

   Compare the informations on linkage contained in intercross data in coupling and repulsion phases.

2. Show that under random mating the frequencies of O, A, B, AB blood group types attain stable equilibrium. How do you estimate O, A, B gene frequencies from observed data? [30]

3. What is the purpose of inbreeding? Examine the consequences of selfing and sibmating inbreeding programs. [30]

4. Examine the following intercross data regarding the segregation of two factors and draw inferences.

		Flower colour	
		Purple	red
Pollen shape	Long	296	27
	Round	19	85

   (Hint: Test forconsistency of segregation with respect to individual factors and independence of factors etc. Estimate linkage if tests show that the factors are linked.) [30]

5. Viva Voce [10]

CENTRAL STATISTICAL ORGANISATION
(TRAINING DIVISION)

Training Course in 'official statistics'
for B.Stat. and M.Stat. students of the
Indian Statistical Institute, Calcutta (1966).

FINAL EXAMINATION

August 12, 1966. Maximum marks: 100 Time: 3 hours

Attempt any five questions

1. Describe the Organisational set up and the activities of the Central Statistical Organisation. Name a few important publications of the Central Statistical Organisation with short notes thereon.

2. Discuss the salient features of the 1961 Population Census of India.

3. Describe the agencies concerned and the scope, content and coverage of Labour Statistics in India.

4. Discuss the recent improvements in Agricultural Statistics particularly the Production Statistics or Land Utilisation Statistics.

5. Explain the organisational set up for collection of statistics of Manufacturing Industries (organised sector) in India and also describe the coverage and content of statistics in this field. Name the statistical publication.

6. Discuss the role of statistics in Planning with particular reference to India.

7. Describe the nature of data collected and published in the Transport Sector by Official agencies in India.

8. Write notes on any two of the following mentioning the scope, content and coverage of statistics available:-

   i) Monetary and Banking statistics
   ii) Unorganised manufacturing sector
   iii) Educational statistics
   iv) External Trade Statistics.

9. Describe the various methods of National Income Estimation. Explain the method followed in India for estimating income from the Agriculture or Industry Sector.
