INDIAN STATISTICAL INSTITUTE
Research and Training School
B.Stat. Part IV: 1970-71 [401]

PERIODICAL EXAMINATION

Probability

Date: 28.9.70 Maximum Marks: 100 Time: 3 hours

Note: All questions may be attempted. Marks allotted for each question are given in brackets [ ].

1. Q is a nonnegative countably additive interval function on $(-\infty, \infty)$ and $Q(-\infty, \infty) < \infty$. $f(x)$ is a bounded function defined on $(-\infty, \infty)$.

(a) Define $\overline{\int} f\,dQ$ and $\underline{\int} f\,dQ$. When do we say that f is integrable w.r.t. Q? [12]

(b) $f(x)$ is bounded and is continuous at all points except at the point $x = \alpha$. Also, $Q(\alpha) = 0$. Prove that $\int f dQ$ exists. [14]

Recall that since Q is countably additive, there is an interval J containing α in its interior and such that $Q(J)$ is arbitrarily small.

(c) Prove that if $f(x)$ is integrable (w.r.t. Q), then $|f(x)|$ is integrable. [12]

(d) Give an example to show that the converse of the above statement is false. [10]

(e) B is a bounded set of nonnegative real numbers. $\sup B = d$, $\inf B = c$. C is the set of squares of the numbers in B. Show that $(\sup C - \inf C) \leq 2d(d-c)$. [10]

Hint: First suppose that B is an interval.

(f) Show that if $f(x)$ is bounded, nonnegative and integrable w.r.t. Q, so is $\{f(x)\}^2 = g(x)$. [14]

2. a) Show that if γ and are any two real numbers, $||\gamma| - |\omega|| \leq |\gamma - \omega|$. [10]

b) Q is such that $Q(-\infty, -A] = Q[A, \infty) = 0$ for a suitable finite A. Also $Q[-A, A] < \infty$.
Put $\phi(t) = \int \frac{1}{1+|tx|}\,dQ$. Prove that $\phi(t)$ is uniformly continuous on $-\infty < t < \infty$. Express δ as a specific function of ϵ. [14]

$\phi(x)$ is complex-valued, bounded and continuous. Prove that

$$\left|\int \phi\, dQ\right| \leq \int |\phi|\, dQ.$$ Here Q is as in question 1.

You may assume the corresponding inequality for real-valued functions. [12]

INDIAN STATISTICAL INSTITUTE
Research and Training School
B.Stat. Part IV: 1970-71

[403]

PERIODICAL EXAMINATION

Statistics-7: Econometrics

Date: 5.10.70. Maximum Marks: 100 Time: 3 hours

Note: Answer all questions. Marks allotted for each question are given in brackets [ ].

1. A regression of q_t (quantity of a particular commodity sold during the period t) on p_t (price of the commodity during the same period) is obtained. Should the regression line in general be taken as the estimated demand curve? Give reasons for your answer. [20]

2. How should the general form of individual demand function be derived from the theory of consumer behaviour. How would you go over to the market demand curve from the individual? Discuss the practical difficulties in this connection. [20]

3. What is the nature of cross section data in relation to the theory of consumer behaviour? What form of relationship can be studied from these cross section dat? Discuss with examples. [20]

4. The table below gives the indices of price and quantity of consumption (production) of an agricultural commodity for six consecutive years. Assuming that a cobweb model of demand and supply holds good, plot the scatters for both the demand and supply curves and then fit straight lines to these scatters by inspection. Give equilibrium solutions for price and quantity.

Year	Indices (base 1949 = 100) of price	quantity
1951	150	100
1952	100	175
1953	135	125
1954	120	160
1955	140	130
1956	130	150

[40]

INDIAN STATISTICAL INSTITUTE
Research and Training School
B.Stat. Part IV: 1970-71

[404]

PERIODICAL EXAMINATION

Statistics-5: Statistical Methods Theory and Practical

Date: 19.10.70 Maximum Marks: 100 Time: 3 hours

Note: Answer as many questions as you can. Five questions carry full marks. Marks allotted for each question are given in brackets [ ].

1.a) Let X have a Poisson distribution

$$P(X = x) = e^{-\mu}\frac{\mu^x}{x!} \quad \text{for } x = 0,1,\ldots$$
$$= 0 \quad \text{otherwise.}$$

Obtain the probability density of $Y = X^2$. [8]

b) Let X_1 and X_2 have the joint probability density function

$$f(x_1,x_2) = (\tfrac{3}{4})^{x_1+x_2}\,(\tfrac{1}{4})^{2-x_1-x_2}$$

for $(x_1,x_2) = (0,0)$, $(0,1)$, $(1,0)$ and $(1,1)$ and zero elsewhere. Find the joint density function of $Y_1 = X_1+X_2$ and $Y_2 = X_1-X_2$. [12]

2.a) Let $(X_1,\ldots,X_k)$ be continuous random variables with joint probability density function $f(x_1,\ldots,x_k)$. Indicate a method of obtaining the joint probability density function of a function $Y = u(X_1,\ldots,X_k)$ of the random variables $X_1,\ldots,X_k$. [8]

b) Let X_1 and X_2 be two independent $N(0,1)$ random variables. Use the above method to show that the density function of $Y = X_1/X_2$ is the Cauchy probability density function given by

$$g(y) = \frac{1}{\pi(1+y^2)} \quad \text{for } -\infty < y < \infty. \qquad [12]$$

3. Let X and Y be two independent random variables following χ^2 distributions with k_1 and k_2 degrees of freedom respectively. Obtain the density function of the random variable

$$\frac{(X/k_1)}{(Y/k_2)}.$$ [20]

4. Let $X_1,\ldots,X_n$ be n independent $N(0,1)$ random variables. Let $\underset{\sim}{X}' = (X_1,\ldots,X_n)$. Let $Q_1,\ldots,Q_p$ be $p(\leq n)$ quadratic forms with ranks $n_1,\ldots,n_p$ respectively such that $\underset{\sim}{X}'\underset{\sim}{X} = Q_1+\ldots+Q_p$. Then show that a necessary and sufficient condition that Q_i's are distributed as χ^2 with n_i degrees of freedom for $i = 1,\ldots,p$ and are independent, is $n = \Sigma n_i$. [20]

5.a) Samples of different sizes were selected from lots supplied by two manufacturers and the number of defective articles detected are given below. Do you consider the lots supplied by both the manufacturers to be of the same quality?

	Sample size	Defectives
A	120	20
B	250	37

[10]

b) The following table gives the frequency distribution of the red blood corpuscles per cell of a haemocytometer.

Number of rbc	Number of cells
0	145
1	156
2	68
3	27
4	5
5	1
above 5	0

Assuming that a Poisson distribution fits the data examine whether the data is consistent with the hypothesis that the mean number of rbc per cell is 1. [10]

6. The following table shows the distribution of 6800 persons drawn from a population by colour of hair and colour of the eye. Test whether there is any association between the two attributes hair colour and eye colour.

Eye colour	Hair colour: Fair	Brown	Black	Red
Blue	1768	807	189	47
Grey	946	1387	746	53
Brown	115	438	288	16

[20]

INDIAN STATISTICAL INSTITUTE
Research and Training School
B.Stat. Part IV: 1970-71 [405]

PERIODICAL EXAMINATION

Statistics-7: Industrial Statistics

Date: 26.10.70 Maximum Marks: 100 Time: 3 hours

Note: Answer any five questions. Marks allotted for each question are given in brackets [ ].

1.a) Explain with illustrative example the importance of rational subgroup in quality control.

Following table provides the summarised data on a measurable characteristic with specification limits 0.7528 to 0.7571, the sample size being 4.

Sample	$\bar{X}$	R	Sample	$\bar{X}$	R	Sample	$\bar{X}$	R
1	0.7540	0.0011	6	0.7539	.0009	11	0.7541	0.0017
2	0.7542	0.0014	7	0.7541	.0012	12	0.7542	0.0010
3	0.7542	0.0009	8	0.7543	.0011	13	0.7545	0.0011
4	0.7546	0.0010	9	0.7547	.0007	14	0.7548	0.0009
5	0.7550	0.0008	10	0.7549	.0015	15	0.7551	0.0012

Examine the data for control, compare the process capability with the specification and hence give a control scheme which in your opinion, is most advantageous.

b) Suppose in (a) average level of the process is set at the mid point of the specification limits and it is desired to evolve an $\bar{X}$-chart for which the average run length is 1000 with respect to any one limit when the process is at the target level. Obtain the control limits for such a chart if the lack of control is indicated by two successive points falling beyond a control limit and process variability remains stable. [4+8+8]=[20]

2.a) The lower and upper specification limits for a measurable characteristic (x) are given as L and U respectively. The process s.d. is in a state of control. If c_1, c_2 and c_3 denote the net profit in Rs. per unit product respectively when $x > U$, $L < x < U$ and $x < L$, determine most profitable level for the process.

In the above case if the s.d. remains stable and the process average is subject to a linear upward trend and c_r be the cost of resetting the process, formulate and indicate how to obtain optimal resetting period 't'. Assume normal distribution for x.

b) The tolerance specified for the outer diameter of a shell is 14" ± 0.067". The oversized shells are reworked while the undersized ones are scrapped. The cost of reworking 5 shells is equal to the loss incurred by scrapping one shell. The process s.d. was found to be 0.028". Obtain the average level by minimizing the total cost of rework and scrap. [6+8+6]=[20]

3.a) Define the following terms:

(i) AQL (ii) AOI (iii) ASN

Explain briefly how the single sampling LTPD Plan of Dodge and Romig has been constructed.

b) Derive an expression for AOQL of single sampling plan by attributes. [15+5]=[20]

GO ON TO THE NEXT PAGE

4.a) Derive general expression for O.C. and ASN of a double sampling plan by attributes.

b) Construct a single sampling plan for AQL = 0.03, Producer's Risk = 0.05 LTPD = 0.08 and Consumer's Risk = 0.10. Plot the AOQ curve of this plan and read the approximate value of AOQL. [5+15]=[20]

5.a) The specification limits on the gross weight of an ink bottle are 110 ± 3 gms. It was found that the weight of an empty bottle has a mean of 53.9 gms. and a s.d. of 0.7 gms. Empty bottles are fed into the filling machine in a random order. Find out what is the maximum allowable s.d. of the weight of ink filled in a bottle so that the final product meets the specification limits on gross weight.

b) Briefly describe how to construct

i) control chart for number defectives

ii) control chart for defects per unit

when the sample size varies.

6. Write short notes on following :

a) Process capability and specification

b) Screening under acceptance/reel plan

c) Statistical Control.

d) O.C. of a control chart. [5 ×4]=[20]

INDIAN STATISTICAL INSTITUTE
Research and Training School
B.Stat. Part IV: 1970-71

406

PERIODICAL EXAMINATION

Statistics-8: Demography

Date: 2.11.70. Maximum Marks: 100 Time: 3 hours

Note: Answer all the questions. All questions carry equal marks.

1. What is the index of fertility you know that can be obtained from India population census data? Explain it indicating the possible drawbacks of such an index and suggesting some remedial measures. Why is it called an indirect index and not a measure of fertility? Define a cumulative fertility function and express it in mathematical symbols.

2. Why is mortality table called a life table? What special purpose does it serve in population studies? What is called the stationary population of a life table? Why is death rate calculated from the stationary population of a life table gives a higher value than that calculated from enumerated population? How do you calculate the mean after life in a life table? What special meaning does such an index convey?

3. Calculate the cumulative fertility rate of the following distribution and express it graphically.

Age in completed years	Total no. of women	Total no. of livebirths.
15 - 19	314056	5878
20 - 24	269340	13980
25 - 29	236187	12722
30 - 34	203477	8374
35 - 39	176534	5630
40 - 44	145037	1450
45 - 49	122946	250

4. Calculate the male and female population at ages 0,1,2,3 and 4 last birth day in the year 1940 from the following data.

Period	Number of reported births in (000's)	
	Male	Female
March 7, 1939 to March 6, 1940	446 (. 139)	421 (.959)
March 7, 1938 to March 6, 1939	430 (.8206)	405 (.9679)
March 7, 1937 to March 6, 1938	426 (.8436)	402 (.9054)
March 7, 1936 to March 6, 1937	408 (.8465)	384 (.8807)
March 7, 1935 to March 6, 1936	396 (.8233)	373 (.8271)

(The figures in brackets give the completeness of enumeration)

Age last birth day	Survival rate from age 0 to midpoint of specified age	
	Male	Female
0	.88815	.90163
1	.73521	.75595
2	.67303	.69344
3	.64049	.65867
4	.62194	.63885

Calculate the mean population for the period March 7, 1939 to March 6, 1940 if the crude birth rate is 40 per 1000 population. Find the crude death rate if the population has increased by 347625 and if net migration rate is - 0.5 per cent.

INDIAN STATISTICAL INSTITUTE
Research and Training School
B.Stat. Part IV: 1970-71
and
M.Stat. Part II: 1970-71 [407]/[601]

Statistics-6: Design of Experiments

Date: 9.11.70 Maximum Marks: 100 Time: 3 hours

Note: Answer Question 5 and as much as you can from the rest. Marks allotted for each question are given in brackets [ ].

1. Show that

a) For any positive integer n, the number of mutually orthogonal Latin Squares (m. o. l. s.) of order n is $\leq n-1$.

b) Any set of $n-2$ m.o. l. s. of order n can be extended to a complete set of $n-1$ m.o. l.s.

c) A complete set of m.o. l.s. of order n exists, if n is a prime power. [2[illegible]]

2.a) Define a Balanced Incomplete Block Design (BIBD). With usual meaning for v, b, r, k, λ show that

(i) $vr = bk$, (ii) $\lambda(v-1) = r(k-1)$, (iii) $b \geq v$.

b) When is a BIBD said to be resolvable? Prove that for a resolvable BIBD, $b \geq v + r - 1$.

c) Show that in a symmetrical BIBD, the number of treatments common between any pair of blocks is the same. [8+8+9]=[2[illegible]]

3.a) What are the situations where adjustment for concomitant variation is called for?

b) Sketch the statistical analysis of a Randomised Block Design when one observation on a concomitant variable is available on every plot, starting with the linear model and stating the underlying assumptions you have to make.

c) Comment on the efficiency of the experiment in (b). [4+14+7]=[2[illegible]]

4. Let v treatments be arranged in b blocks such that (i) i-th treatment is replicated r_i times, (ii) the size of j-th block is k_j and (iii) i-th treatment occurs in j-th block n_{ij} times, $i = 1,2,\ldots, v$; $j = 1,2,\ldots, b$.

a) Obtain C-matrix and indicate the analysis of variance of this design, assuming Rank $(C) = v-1$.

b) A Balanced Incomplete Block Design with parameters v, b, r, k, λ is modified to include a new treatment by increasing the block size to $k+1$ plots, so that the new treatment occurs once in each of the b blocks. Write the C-matrix and the subsequent method of analysis of this modified experiment. [14+11]=[25

5. A manurial experiment with sugarcane was conducted in a 5 X 5 Latin Square. Five treatments were as follows:

A : no manure
B : an organic manure
C, D, E : three levels of farm-yard manure.

The plan and yield figures (in suitable units) are shown below, there being one missing plot.

GO ON TO THE NEXT PAGE

Row	Column 1	Column 2	Column 3	Column 4	Column 5
1	A 52.5	E 46.3	D 44.1	C 48.1	B 40.9
2	D 44.2	B 42.9	A 51.3	E 49.3	C 32.6
3	B 49.1	A 47.3	C –	D 41.0	E 47.2
4	C 43.2	D 42.5	E 67.2	B 55.1	A 45.3
5	E 47.0	C 43.2	B 46.7	A 46.0	D 43.2

(a) Analyse the above data to find out if there are significant treatment effects.

(b) Compare the treatments C, D and E pairwise. [24+6]=[30]

INDIAN STATISTICAL INSTITUTE
Research and Training School
B.Stat. Part IV: 1970-71

MID-YEAR EXAMINATION

Statistics-4: Probability

Date: 21.12.70 Maximum Marks: 100 Time: 3 hours

Note: The paper carries 115 marks. You may attempt any part of any question.
Marks allotted for each question are given in brackets [ ].

1.a) We define the distribution function of a probability distribution P by $F(x) = P(-\infty, x]$

Prove that if two distributions have the same distribution function, the distributions are the same. [10]

b) Suppose the distribution function is defined (as is done by many authors) by $F(x) = P(-\infty, x)$.

Prove that this $F(x)$ is left-continuous at every point. [10]

2.a) Prove that if a and b are real numbers

$$||a| - |b|| \leq |a - b|.$$ [4]

b) Q is a probability distribution such that $Q(-\infty, -A) = Q(A, \infty) = 0$; here A is some positive number.

$$\phi(t) = \int_{-A}^{A} \frac{1}{\sqrt{1+|tx|}} dQ(x)$$

Prove that $\phi(t)$ is uniformly continuous on $-\infty < t < \infty$. [10]

3. You may use standard theorems on characteristic functions.

(a) a, b, c, d are real numbers. We know that $e^{iat} + 2e^{ibt} = e^{ict} + 2e^{idt}$ for all real t. Show that $a=c$ and $b=d$. [8]

(b) $a_1, a_2, a_3, \ldots$ are all positive real numbers such that $(a_1 + a_2 + \ldots)$ is convergent. Then prove that

$$g(t) = \sum_{n=1}^{\infty} a_n \cdot e^{int}$$

is absolutely convergent for every real number t and that $g(t)$ is uniformly continuous on $-\infty < t < \infty$. [10]

4.a) A is a bounded set of real numbers ≥ 1; its supremum is a. B is the set of square-roots of the numbers in A. Prove rigorously that $\sup B = \sqrt{a}$. [8]

b) Q is a probability distribution on $(-\infty, \infty)$. $f(x)$ is ≥ 1 for all x and is integrable w.r.t. Q. Prove that

$\sqrt{f(x)} = g(x)$ is integrable w.r.t. Q. [12]

c) A finite closed interval $[a, b]$ is divided into countably many disjoint subintervals. Show that the sum of the lengths of the subintervals is $= (b-a)$. [9]

5.a) State Levy's basic inversion formula: If $(a-h)$ and $(a+h)$ are continuity points of the distribution P, then $P(a-h, a+h) = \lim_{T \uparrow \infty} J(T)$ where $J(T)$ is an expression involving the characteristic function of P. What exactly is $J(T)$? [5]

b) Deduce that if two probability distributions have the same characteristic function, the distribution functions are the same. [9]

c) $\frac{1}{1+t^2}$ is the characteristic function of a continuous probability distribution P. Prove that for $h > 0$, $P(-h, h) \leq h$.

Deduce this directly from Levy's formula. Even if you happen to know the frequency function of P, do not use it. Use the inequality

$$\left|\frac{\sin \theta}{\theta}\right| \leq 1$$

which holds for all real θ. [10]

d) It is known that the frequency function of P above is

$$\frac{1}{2} e^{-|x|}.$$

Show by direct elementary computation that

$$\int_{-h}^{h} \frac{1}{2} e^{-|x|} dx \leq h.$$ [10]

INDIAN STATISTICAL INSTITUTE
Research and Training School
B.Stat. Part IV: 1970-71

MID-YEAR EXAMINATION

Statistics-5: Statistical Methods Theory and Practical

Date: 22.12.70 Maximum Marks: 100 Time: 3 hours

Note: You may attempt any part of any question. Not more than 60 marks will be awarded for either Group. Marks allotted for each question are given in brackets [ ].

GROUP A

1.a) Let X_1, X_2, X_3 denote a random sample from the distribution having the density function $f(x) = e^{-x}$ for $0 < x < \infty$, and zero else where. Show that

$$Y_1 = \frac{X_1}{X_1 + X_2} ; \quad Y_2 = \frac{X_1 + X_2}{X_1 + X_2 + X_3} ; \quad Y_3 = X_1 + X_2 + X_3$$

are mutually stochastically independent. [12]

b) If $f(x) = \frac{1}{4}$ for $-2 < x < 2$ and 0 elsewhere is the density function of a random variable X find the density function of $Y = X^2$. [8]

2. Let π_i (> 0), $i = 1, \ldots, k$ be the probability of the i^{th} class in a multinomial distribution with k classes. Suppose n independent events are observed resulting n_i in the i^{th} class for $i = 1, \ldots, k$. Let

$$\underset{\sim}{V}' = \left(\frac{n_1 - n\pi_1}{\sqrt{n\pi_1}}, \ldots, \frac{n_k - n\pi_k}{\sqrt{n\pi_k}}\right)$$

and $\underset{\sim}{\phi}' = (\sqrt{\pi_1}, \ldots, \sqrt{\pi_k})$.

Show that a sufficient condition for the asymptotic distribution of the quadratic form $\underset{\sim}{V}' \underset{\sim}{C} \underset{\sim}{V}$ to be χ^2 is that $\underset{\sim}{C}^2 = \underset{\sim}{C}$ and $\underset{\sim}{C} \cdot \underset{\sim}{\phi} = \alpha \cdot \underset{\sim}{\phi}$ where α is a constant in which case the degrees of freedom of χ^2 is $R(\underset{\sim}{C})$ if $\alpha = 0$ and $(R(\underset{\sim}{C}) - 1)$, if $\alpha \neq 0$.

(Hint: If $\underset{\sim}{A}$ is a $k \times (k-1)$ matrix such that the matrix $(\underset{\sim}{\phi} \mid \underset{\sim}{A})$ of order $(k \times k)$ is orthogonal, then the asymptotic distribution of the (k-1) linear functions $\underset{\sim}{G} = \underset{\sim}{A}' \underset{\sim}{V}$ is that of (k-1) independent normal variables each with mean zero and variance unity.) [20]

3. Consider the Gauss-Markov set up $\underset{\sim}{Y}_{(n \times 1)} \cap N(\underset{\sim}{X}_{(n \times m)} \underset{\sim}{\beta}_{(m \times 1)}, \sigma^2 I_{(n \times n)})$, where $\underset{\sim}{X}_{(n \times m)}$ is a known matrix. Obtain a necessary and sufficient condition for the estimability of a linear function $\underset{\sim}{P}' \underset{\sim}{\beta}$ of the parameters $\underset{\sim}{\beta}$. [10]

b) Show that if $\underset{\sim}{P}'\beta$ is estimable, $\underset{\sim}{P}' \hat{\underset{\sim}{\beta}}$ where $\hat{\underset{\sim}{\beta}}$ is a least squares estimate of β, has minimum variance in the class of all linear unbiased estimates of $\underset{\sim}{P}'\underset{\sim}{\beta}$. [10]

4. Consider the Gauss-Markov set up $\underset{\sim}{Y}_{(n \times 1)} \cap N(\underset{\sim}{X}_{(n \times m)} \underset{\sim}{\beta}_{(m \times 1)}, \sigma^2 \underset{\sim}{I})$ where $\underset{\sim}{X}$ is a known matrix of rank r. Show that $R_0^2 = \min_{\beta} (\underset{\sim}{Y} - \underset{\sim}{X} \underset{\sim}{\beta})'(\underset{\sim}{Y} - \underset{\sim}{X} \underset{\sim}{\beta})$ follows a χ^2 with (n-r) degrees of freedom. [illegible]

GO ON TO THE NEXT PAGE

5.a) Defining a generalized inverse of $\underset{\sim}{A}$, an $(n \times m)$ matrix, as that matrix $\underset{\sim}{A}^-$ such that for any vector $\underset{\sim}{Y}$ for which $\underset{\sim}{A}\,\underset{\sim}{X} = \underset{\sim}{Y}$ is a consistent equation, $\underset{\sim}{X} = \underset{\sim}{A}^- \underset{\sim}{Y}$ is a solution, show that $\underset{\sim}{A}^-$ exists $\Longleftrightarrow \underset{\sim}{A}\,\underset{\sim}{A}^-\underset{\sim}{A} = \underset{\sim}{A}$. [10]

b) Defining $\underset{\sim}{H} = \underset{\sim}{A}^- \underset{\sim}{A}$ where $\underset{\sim}{A}^-$ is any g-inverse of $\underset{\sim}{A}$ show that a general solution of the homogeneous equations $\underset{\sim}{A}\,\underset{\sim}{X} = 0$ is $(\underset{\sim}{I} - \underset{\sim}{H})\,\underset{\sim}{Z}$ where $\underset{\sim}{Z}$ is an arbitrary vector. [10]

GROUP B

6. The distribution in four blood group classes O, A, B, AB of 140 Christians who were army cadets and 295 other Christians are given below. Can the two samples be regarded as coming from the same population?

Samples	Blood group			
	O	A	B	AB
Army cadets	56	60	18	6
Others	120	122	42	11

[20]

7.a) The following table gives the result of an experiment to compare the effect of a newly discovered drug on a certain disease with that of the prevailing treatment. Do the data confirm superiority of the new medicine?

	Cured	Not cured
Control	4	6
New Medicine	4	2

[12]

b) Two series of measurements on the plankton organism in the water of a lake are made. The first series of measurements is made in succession at the same point in the lake. The second series is made at different points scattered over the lake. Do the observations suggest that there is more variability in the plankton content at different places in the lake?

Series I: 80, 96, 102, 77, 97, 110, 98, 88, 103, 108.

Series II: 74, 122, 92, 81, 104, 92, 119. [8]

8.a) The following data were collected in an experiment on jute in a village of West Bengal in 1953 in which the weights of green plants and the dry jute fibre were recorded. Do the data show any correlation between the two characteristics studied?

Plant No.	Wt. in Tolas of Green Plant	Wt. in Tolas of Dry fibre
1	37	2.8
2	36	2.3
3	47	3.5
4	33	2.5
5	105	7.0
6	17	1.3
7	59	4.5
8	87	6.0
9	62	4.2
10	54	3.9

[12]

b) The correlation coefficient between the scores in two halves of a psychological test applied on a sample of 30 students was 0.63. Can you conclude that the correlation coefficient in the population could be 0.6?

9. To determine the effect of dilution of the electrolyte and the strength of the current on the thickness of the coating of aluminium foil in an electroplating experiment, four different dilutions and three different current strengths were used. In each experimental set up the thickness of the coating on two aluminium foils were recorded.

Analyse the data and give your comments.

Dilution	Current strength		
	1.0	1.5	2.0
5.0	10.5, 11.3	12.9, 11.2	5.9, 7.6
6.0	10.6, 11.6	12.5, 13.4	10.3, 7.6
7.0	7.5, 9.4	10.6, 12.0	5.0, 5.6
8.0	6.8, 7.2	8.8, 8.0	3.2, 2.7

[20]

INDIAN STATISTICAL INSTITUTE
Research and Training School
416/ 863

B. Stat. Part IV & 1 M. Stat. Part II: 1970-71
MID-YEAR EXAMINATION
Statistics-6: Design of Experiments Theory

Date: 23.12.70 Maximum Marks: 100 Time: 3 hours

Note: You can answer any part of any question.
The paper carries 118 marks.
Marks allotted for each question are given in brackets [ ].

1. Define an m-ple Lattice Design.

Starting with the normal equations of general Incomplete Block Design, write down clearly the procedure for obtaining the best linear unbiased estimate of a treatment contrast from an m-ple Lattice Design. [18]

2.a) What is a Youden Square? [3]

b) Suppose from a Latin Square of order v one column is omitted. Show that the resulting arrangement gives a Youden Square. Obtain the parameters of the resulting design. [6]

c) Briefly indicate the method of analysis of a Youden Square Design. [9]

3.a) Given a symmetrical BIBD with parameters $v = b$, $r = k, \lambda$, describe the procedures for obtaining BIBD's with the following parameters:

i) $v^* = v-k$, $b^* = v-1$, $r^* = r$, $k^* = r-\lambda$, $\lambda^* = \lambda$,

ii) $v^* = k$, $b^* = v-1$, $r^* = r-1$, $k^* = \lambda$, $\lambda^* = \lambda-1$. [8]

b) Show that for a symmetrical BIBD, $(r-\lambda)$ must be a perfect square if v is even. [8]

4. Prove the following:

In a factorial experiment involving n factors, each at s levels, the totality of (s^n-1) degrees of freedom belonging to s^n observations can be split into

$$\frac{s^n-1}{s-1}$$

orthogonal components, each carrying $(s-1)$ degrees of freedom, provided s is a prime power.

Hence or otherwise, give a procedure for laying out an s^n Experiment in s blocks, confounding only a component of the highest order interaction. [20]

5.a) Describe any suitable method for laying out an s^n factorial experiment in s^2 blocks, accounting for all the confounded degrees of freedom. [Assume s to be a prime power.] [14]

b) A 3^3 Experiment is to be conducted in 9 blocks such that no main effect is confounded. Give the layout of the experimental plan. [8]

6.a) Give a balanced scheme of partial confounding, for studying five factors each at two levels, using five replications in blocks of eight plots each. It is desired that no main effect or two-factor interactions should be confounded in any replication.
[Give complete layout for one replication and principal blocks for the remaining ones.] [12]

6.b) Construct a half replicate of a 2^7 design, laid out in 4 blocks so that all main effects and first order interactions are estimable. Write down the aliases of main effects. Assuming interactions of 3 or more factors to be negligible, give the partitioning of the total degrees of freedom in the analysis of variance table. [illegible] -[12]

INDIAN STATISTICAL INSTITUTE
Research and Training School

[411/604]

B. Stat. Part IV and M. Stat. Part II: 1970-71

MID-YEAR EXAMINATION

Statistics-6: Design of Experiments Practical

Date: 24.12.70 Maximum Marks: 100 Time: 3 hours

Note: Answer all the questions. Marks allotted for each question are given in brackets [ ].

1. The following design was used to test 9 rations fed to rats. The gains in weight of the rats after the feeding experiment were as follows: (The ration numbers are in brackets.)

Replication 1							Replication 3						
Block 1:	(1)	20	(4)	15	(7)	11	Block 7:	(1)	13	(9)	19	(5)	14
" 2:	(3)	08	(6)	18	(9)	26	" 8:	(8)	14	(4)	34	(3)	02
" 3:	(2)	18	(5)	16	(8)	02	" 9:	(6)	14	(2)	20	(7)	14

Replication 2							Replication 4						
Block 4:	(7)	08	(8)	12	(9)	16	Block 10:	(5)	19	(7)	23	(3)	06
" 5:	(1)	20	(2)	02	(3)	02	" 11:	(1)	22	(6)	12	(8)	02
" 6:	(4)	20	(5)	06	(6)	02	" 12:	(9)	27	(2)	07	(4)	20

Analyse the data and compute the efficiency of this design relative to randomised block design. [30]

2. An experiment was conducted to estimate the optimum position of the primary fan in a furnace to ensure minimum irrecoverable loss, believed to be due to scaling. Four positions of the primary fan were tried and these were denoted as A, B, C and D respectively. The experiment was carried out in four days (columns) and at four intervals on each day (rows). It was felt necessary to consider one more factor, namely, chimney damper which might influence the scaling. Four levels of the chimney damper were chosen, denoted as α, β, γ and δ. A Graeco-Latin Square arrangement of the type given below was considered for the purpose. The percentage irrecoverable loss has been obtained for each of the sixteen positions. These are given in brackets.

Intervals in a day	Days 1	2	3	4
1	A α (3.0)	B β (7.2)	C γ (4.6)	D δ (3.0)
2	B γ (2.5)	A δ (3.0)	D α (3.7)	C β (4.6)
3	C δ (3.9)	D γ (3.3)	A β (10.4)	B α (5.0)
4	D β (3.6)	C α (3.6)	B δ (3.6)	A γ (5.7)

Analyse the data and interpret the results. [30]

GO ON TO THE NEXT PAGE

3. The table below gives the yields (in some unit) along with the plan of a factorial experiment involving 3 factors N, P and K each at two levels, conducted in 2 replications. Each replication consists of two blocks of 4 plots.

	Block 1 Treatment	Block 1 Yield	Block 2 Treatment	Block 2 Yield
Replication 1:	nk	159	p	153
	(1)	179	npk	202
	pk	139	n	153
	np	130	k	182

	Block 3 Treatment	Block 3 Yield	Block 4 Treatment	Block 4 Yield
Replication 2:	npk	155	n	191
	np	129	nk	138
	k	151	p	188
	(1)	159	pk	210

[Standard notation has been used for the treatment combinations.]

Test for the presence of different main effects and interactions. [30

4. Practical Note Book [10

INDIAN STATISTICAL INSTITUTE
Research and Training School
B.Stat. Part IV: 1970-71 |412|

MID-YEAR EXAMINATION

Statistics-7: Econometrics Theory and Practical

Date: 25.12.70 Maximum Marks: 100 Time: 3 hours

Note: Answer all the questions. Marks allotted for each question are given in brackets [].

1. What is an Engel curve? Give examples of alternative formulation of the Engel curve with possible economic interpretation. [20]

2.a) What is the homogeneity hypothesis in connection with the formulation of Engel curve? How does it affect the formulation of the Engel curve?

b) What is 'economies of scale' in household consumption? How is 'economies of scale' introduced in the formulation of the Engel curve? [20]

3. Assuming that individual income is distributed log-normally, derive the equation for the lorenz curve as also the expression for the associated concentration coefficient. [20]

4. The table below shows percentage distribution of persons by monthly percapita total expenditure classes along with the average monthly per capita total expenditure in each class for rural India for the periods 1961-62 and 1964-65.

Examine if the inequality of distribution of monthly per capita total expenditure has changed; if so in what direction?

Monthly per capita expenditure classes in Rs.	1961-62		1964-65	
	% distribution of persons	av. monthly per capita total expenditure Rs. (0.00)	% distribution of persons	av. monthly per capita total expenditure Rs. (0.00)
0 - 8	4.32	6.52	3.18	6.25
8 - 11	11.37	9.66	5.99	9.71
11 - 13	9.70	12.02	7.18	12.06
13 - 15	11.46	14.04	8.68	13.97
15 - 18	14.23	16.36	12.92	16.44
18 - 21	11.75	19.43	13.35	19.45
21 - 24	9.93	22.45	11.06	22.41
24 - 28	7.92	25.90	10.96	25.97
28 - 34	7.49	30.59	10.27	30.51
34 - 43	5.76	37.83	8.29	37-86
43 - 55	3.08	48.39	4.16	48.23
55 - 75	1.73	63.29	2.29	[illegible]
75 - above	1.26	119.85	1.67	130.01

[40]

INDIAN STATISTICAL INSTITUTE
Research and Training School
B.Stat. Part IV: 1970-71

413

MID-YEAR EXAMINATION

Statistics-8: Demography Theory and Practical

Date: 26.12.70 Maximum Marks: 100 Time: 3 hours

Note: Attempt Question 5 and any three of the rest. Marks allotted for each question are given in brackets [ ].

1.a) Name the important sources of demographic data. [5]

b) Define vital statistics. Discuss the uses and value of vital statistics in general and to a health administrator in particular. [20]

2.a) Distinguish between fertility rate and a reproduction rate. [5]

b) Starting with age-sp-fertility rate how do you calculate.

(i) T.F.R, (ii) G.R.R. and (iii) N.R.R. [12]

c) Explain the conditions under which a population may remain stationary may be increasing or may be decreasing. [5]

d) Indicate some relevent use of reproduction rate in population studies. [3]

3.a) What is the usefulness of standardisation for studying population events? [8]

b) What are the different methods of standardisation you know of and how are their uses related to actual data? [10]

c) What is called a standard million? [3]

d) Define a standardised mortality ratio (S.M.R.). [4]

4.a) Distinguish between m_x, q_x and μ_x, the three functions of mortality and deduce an explicit relation between m_x and μ_x. [18]

b) Show that a life table mortality rate is the weighted average of the specific rates of mortality. [7]

5. In distinguishing between the fertility differential in some of the East, West and Northern district of India, the data available from a survey are represented below.

Age. gr.	All India: a.S.M.F.R.	All India: Population	All India: Prop. currently married	Currently married population: Assam	Currently married population: U.P.	Currently married population: Bombay
15-19	323.0	166845	.6972	31389	30640	30974
20-24	310.2	145931	.9180	47275	39332	45494
25-29	280.3	121151	.9420	48965	33186	46171
30-34	222.6	100675	.9145	37192	27910	35174
35-39	158.7	83726	.8703	26066	20380	27822
40-44	69.0	68009	.7767	20991	18442	20605
45-49	27.3	53800	.6972	12865	12554	15498

The records available for these districts showing that the M.F.R. (crude) for these districts are in the order of 297.5, [illegible]05.0 and 212.1. Does the fertility differential computed from the two sources show the same order of differential? If not, draw your conclusion regarding the actual picture.

INDIAN STATISTICAL INSTITUTE
Research and Training School
B.Stat. Part IV: 1970-71 [414]

MID-YEAR EXAMINATION

Statistics-7: Industrial Statistics Theory and Practical

Date: 28.12.70. Maximum Marks: 100 Time: 3 hours

Note: Answer any five questions. Marks allotted for each question are given in brackets [ ].

1.a) Explain what is meant by a process being in a state of statistical control. [4]

b) The following are the $\bar{X}$ and R values, for 20 sub-groups of five observations. The specification for this product Characteristic is 0.4037 ± 0.0010. The values given are the last two figures of the observed readings (i.e. 31.6 means the number 0.40316)

Sub-group No.	$\bar{x}$	R	Sub-group No.	$\bar{x}$	R
1	34.0	4	11	35.8	4
2	31.6	4	12	35.8	4
3	30.8	2	13	34.0	14
4	33.0	3	14	35.0	4
5	35.0	6	15	33.8	7
6	32.2	2	16	31.6	6
7	33.0	5	17	33.0	5
8	32.6	13	18	33.2	3
9	33.8	12	19	31.8	9
10	35.8	6	20	35.6	6

i) Draw suitable control chart to find if the process is in control.

ii) Under suitable assumptions (to be stated) estimate the proportion of items that will fail to meet the specifications. [16]

2.a) Distinguish between natural tolerance limits, control limits and confidence limits.

In the following two cases, derive the necessary expressions to construct tolerance limits such that it can be asserted with 100β percent confidence that they will include at least 100α percent of the population.

i) a normal parent population with known mean

ii) a parent population with unknown but continuous p.d.f.

where α and β are preassigned numbers between 0 and 1. [14]

b) Find sample size n for which it can be stated with 95% confidence that at least 90% of the items in the population of unknown form but with continuous p.d.f. lie below the maximum in the sample. [6]

3.a) Derive an expression for AOQL of single sampling plan by attributes. [4]

b) A product in lots is submitted for inspection and the following double sampling plan is used.

$$n_1 = 50, \quad c_1 = 0, \quad n_2 = 60, \quad c_2 = 3$$

GO ON TO THE NEXT PAGE

Assuming that the lot is very large compared to the sample size compute $L(p)$ the probability of acceptance for 5 suitable values of p and draw the OC curve. [16]

4.a) Give some of the advantages and disadvantages of acceptance sampling plan by attributes and variables. [4]

b) Obtain an unknown sigma acceptance sampling plan by variables for the case-one-sided specification (upper) limit U, such that under the plan, lots with $100p_1$ and $100p_2$ per cent defectives $(p_2 > p_1)$ would be accepted with probabilities $(1-\alpha)$ and β respectively. Write out the expression for the O.C. function of the plan. [16]

5.a) Obtain graphical item by item sequential sampling plan using inspection by attributes.

Derive expressions for all the terms used in describing the sample plan from the following given parameters

AQL $= p_1$ Producer's risk $= \alpha$

LTPD $= p_2$ Consumer's risk $= \beta$ [12]

b) Derive also an expression for the O.C. function of the plan and hence obtain any five points on the O.C. curve. [8]

6. Write short notes on following:

a) Group control chart. [6]

b) Mil-Std 105D. [10]

c) Average Run Length of a control chart. [4]

INDIAN STATISTICAL INSTITUTE
Research and Training School
B.Stat. Part IV: 1970-71

415

PERIODICAL EXAMINATION

Statistics-7: Planning Techniques

Date: 29.3.71 Maximum Marks: 60 Time: 2 hours

Note: Answer any three questions. All questions carry equal marks.

1. Consider a firm trying to maximise income from given resources. It possesses a fixed supply, b, of resources and knows only two ways of producing goods, a_1 and a_2 where

$$b = \{2, 3\} \; ; \; a_1 = \{6, 2\} \; ; \; a_2 = \{1, 5\} .$$

The rates of income associated with a_1 and a_2 are respectively 8 and 4.

i) Write down the linear programming problem of the above firm.

ii) What should be the possible interpretation of the dual to the above problem? Solve the dual problem graphically.

2. State and prove the fundamental duality theorem for standard linear programs.

3.a) Show (graphically) that the following linear program is feasible but has no optimal solution: Find $\xi_1, \xi_2 \geq 0$ such that

$$\xi_1 + \xi_2 \text{ is a maximum}$$

subject to

$$-3\xi_1 + 2\xi_2 \leq -1$$
$$\xi_1 - \xi_2 \leq 2 .$$

b) Write the dual to the above problem. In view of the above result and the fundamental duality theorem what must be true of this dual problem? Verify this graphically.

4.a) State and prove the equilibrium theorem for standard linear programs.

b) Consider the following problem:

$$\text{maximise } \xi_1 + \xi_2 + \xi_3 + \xi_4$$

subject to

$$\begin{aligned} \xi_1 + \xi_2 \qquad\qquad &\leq 3 \\ \xi_3 + \xi_4 &\leq 1 \\ \xi_2 + \xi_3 \qquad &\leq 1 \\ \xi_1 \qquad + \xi_3 \qquad &\leq 1 \\ \xi_3 + \xi_4 &\leq 3 \\ \text{all } \xi_j &\geq 0 . \end{aligned}$$

GO ON TO THE NEXT PAGE

4. (contd.)

Show that this problem has the optimal solution

$$\xi_1 = 1, \quad \xi_2 = 1, \quad \xi_3 = 0, \quad \xi_4 = 1$$

by finding a solution of the dual problem making use of the equilibrium theorem.

5. Consider a system of m simultaneous linear equations in n unknowns $(n \geq m)$, $Ax = b$, with $r(A) = m$.

Prove that if there is a feasible solution, there is a basic feasible solution.

6. Solve the following simultaneous equations by 'replacement operation':

$$\begin{aligned} \xi_1 \quad\quad - \xi_3 &= 1 \\ \xi_1 + 2\xi_2 \quad\quad &= 8 \\ 2\xi_1 - \xi_2 + 3\xi_3 &= 4 . \end{aligned}$$

INDIAN STATISTICAL INSTITUTE
Research and Training School
B.Stat. Part IV: 1970-71

[416]

PERIODICAL EXAMINATIONS

Statistics-8: Educational Statistics
(Theory and Practical)

Date: 5.4.71 Maximum Marks: 100 Time: 3 hours

Note: Answer Q.No.1 and any five questions from the rest. Marks allotted for each question are given in brackets [ ].

1. Write short notes on any four of the following:

(a) Parallel tests
(b) Spearman's Two-Factor Theory
(c) Coefficient of Discrimination of a test
(d) Thurstone's method of attitude scale construction
(e) Rank-Order Correlation. [5+5+5+5]=[2

2.a) What is a normalised score? How would you derive normalised scores from raw scores? Compare normalised score with other standard scores and discuss its advantages and disadvantages.

b) Determine the reliability of the difference score obtained from the following pair of tests A and B, where the reliability of test A = .90, the reliability of test B = .80 the correlation between tests A and B = .30. Assume that the variances of test A and test B are both equal to one. [8+8]=[1

3.a) Illustrate geometrically the following:

i) the intercorrelations among the tests
ii) common factor loadings
iii) the concept of communality

(considering the case when there are three tests and two common factors).

b) What do you understand by Speed and Power tests? Explain why odd-even reliability should not be used in the case of speed tests. [8+8]=[1

4.a) What is 'correction for guessing'? How is item difficulty value corrected for guessing? Discuss the commonly used formulae giving the underlying assumptions.

b) Test X is reported as having a mean of 68.1, a standard deviation of 24.8, a reliability of 0.97 and a validity of 0.75 with the criterion. The mean value of the criterion score is 117.8 and the corresponding standard deviation is 20.1. If we screen a group using scores on test X and obtain a selected group with standard deviation 15.0 on test X, what will the validity be for this test on this selected group? [8+8]=[16

5.a) What do you understand by item discrimination? Describe the item discrimination indices which involve the slope of the regression of item on test.

GO ON TO THE NEXT PAGE

5.b) In a sample of 80 seventeen-year-old high school students of whom 35 were male and 45 female, the mean weights in kilograms were 68.8 and 56.6 respectively. The standard deviation of the weights for the combined group was 15.2. Find the point-biserial correlation between sex and body weight for seventeen-year-old school students. [8+8]=[16]

6.a) Describe different types of validity of a test. Derive the formula which shows the relation among the reliabilities of the test and the criterion, the original validity coefficient and the maximum validity that the test can attain when both the criterion and the test are perfectly reliable.

b) Describe how you would obtain the confidence interval (at the α % level) of the true score corresponding to a given observed score. [8+8]=[16

7.a) Write brief notes including equation on the estimation of factor loadings by the Centroid method.

b) Prove that if a test of n items is a sub-test of a test with m items $(n < m)$ the correlation r_{nm} is

$$r_{nm} = \sqrt{\frac{\frac{1-r}{m} + r}{\frac{1-r}{n} + r}}$$

where r is the reliability of a unit test. [8+8]=[16

INDIAN STATISTICAL INSTITUTE
Research and Training School
B.Stat. Part IV: 1970-71
PERIODICAL EXAMINATIONS

[417]

Statistics-7: Econometrics Theory and Practical

Date: 12.4.71 Maximum Marks: 100 Time: 3 hours

Note: Answer all the questions. Marks allotted for each question are given in brackets [ ].

1.a) Examine the methodological problems involved in fitting a production function to

i) time series data,

ii) cross-section data.

b) Show, stating necessary assumptions, that the exponents of the Cobb-Douglas production function equal the shares of the factors of production. [12+8]=[20]

2. Outline some of the major criticisms against the use of Cobb-Douglas production function. [20]

3. Production function for different units of a certain industry is given by $Q = AL^{\alpha} K^{\beta}$, where Q = output, L = labor and K = capital, A, α and β being constants. What will be the form of the aggregate production function for the whole industry, assuming that Q, L and K are lognormally distributed?

Describe how you would use such production functions for estimating the total capital employed in Indian manufacturing industries. [20]

4. The following gives the indices (base = 1899) of volume of production (Q), average number of wage earners (L), and of physical volume of capital (K), in manufacturing industries in U.S.A. for some selected years:

Year	Q	L	K
1900	102	105	107
1903	124	123	131
1906	152	133	163
1909	155	140	198
1912	177	152	226
1915	189	154	266
1918	223	200	366

Estimate α and β of the production function

$$Q = A L^{\alpha} K^{\beta} .$$ [40]

INDIAN STATISTICAL INSTITUTE
Research and Training School
B.Stat. Part IV: 1970-71

PERIODICAL EXAMINATION

Statistics-5: Statistical Methods (Theory and Practical)

Date: 19.4.71 Maximum Marks: 100 Time: 3 hours

Note: Answer as many questions as you can. Five questions carry full marks. All questions carry equal marks.

1.a) Let X, Y, Z be three random variables. Define clearly the partial correlation $r_{XY.Z}$ between X and Y with respect to Z. Obtain explicit expression for $r_{XY.Z}$ in terms of the coefficients of correlations $r_{..}$ between the variables X, Y and Z.

b) If $r_{XY.Z} = 0$ show that $r_{XZ.Y} = r_{XZ}\sqrt{\dfrac{1-r_{YZ}^2}{1-r_{XY}^2}}$.

2.a) Let $x_1,\ldots,x_n$ be a random sample of size n from a continuous distribution with d.f. F(x). Let Z_p denote the p^{th} quantile of the sample. Show that Z_p is asymptotically normal and obtain the parameters.

b) Obtain as a particular case, the asymptotic distribution of the median of a sample of size n from the normal distribution with mean μ and variance σ^2.

3.a) Define precisely when a random vector $\underset{\sim}{X}$ of p components is said to have a multivariate normal distribution.

b) Using the definition you give show that $\underset{\sim}{Y} = \underset{\sim}{D}\,\underset{\sim}{X}$ where $\underset{\sim}{D}$ is any $r \times p$ matrix, also has a multivariate normal distribution.

c) Show that any subset of the components of $\underset{\sim}{X}$ has a multivariate normal distribution and obtain its parameters.

4. Let $\underset{\sim}{X}$ follow a multivariate normal distribution. Show that every linear combination of the components of $\underset{\sim}{X}$ follows a univariate normal distribution. Conversely, show that if every linear combination of the components of a random vector $\underset{\sim}{X}$ follows a univariate normal distribution, $\underset{\sim}{X}$ has a multivariate normal distribution.

b) Let $\underset{\sim}{X}_1,\ldots,\underset{\sim}{X}_n$ be a random sample from a multivariate normal distribution. Let $\bar{\underset{\sim}{X}}$ denote the sample mean; deduce from above that $\bar{\underset{\sim}{X}}$ follows a multivariate normal distribution.

5. Each of three different samples of earthworm casting was divided into a number of portions. Each portion was given to one of two analysts for determination of percentage ash-content. The results are given below. Analyze the data to detect differences, if any, between samples and between analysts.

GO ON TO THE NEXT PAGE

Analysts	Samples		
	1	2	3
1	57.30 57.65 57.70	58.19 56.59 57.57 57.28	56.21 56.15 56.03
2	56.81 58.44 58.00 58.31	58.78 56.65 57.88	57.52 56.40 57.45 57.16 56.52

6. The components of X correspond to scores on tests in arithmetic speed X_1, arithmetic power X_2, memory for words X_3 and memory for meaningful symbols X_4. The observed correlations in a sample of 140 are

1.0000	0.4248	0.0420	0.0215
		0.1487	0.2489
			0.6693

a) Find the partial correlation between X_1 and X_2 holding X_3 and X_4 fixed.

b) Test whether this partial correlation coefficient is 0.

c) Test whether the correlation between X_1 and X_2 is 0.5.

INDIAN STATISTICAL INSTITUTE
Research and Training School
B.Stat. Part IV: 1970-71

[419]

PERIODICAL EXAMINATION

Statistics-4: Probability

Date: 26.4.71 Maximum Marks: 100 Time: 3 hours

Note: The question paper carries 110 marks. You may attempt any part of any question. The maximum you can score is 100. Marks allotted for each question are given in brackets [ ].

Proofs are to be given only when they are specifically asked for. If a question is followed by '(half a page)', for example, your answer should not take more than half a page.

Rough work must be clearly segregated from material meant for the examiner.

1.a) P is a probability distribution; a is a real number. Prove that we can construct a sequence $h_1 > h_2 > h_3 > \ldots$ of positive numbers such that

$$\text{i)} \quad \lim_n h_n = 0,$$

and ii) $(a + h_n)$ and $(a - h_n)$ are continuity points of P, for $n = 1,2,3,\ldots$

(half a page) [15]

b) If Q is any probability distribution symmetric w.r.t. zero, its characteristic function $\phi(t)$ is given by

$$\int_{-\infty}^{\infty} A(t, x)\, dQ$$

where $A(t, x)$ is a specific real-valued function. What is $A(t, x)$? [5]

(No proof needed. One line)

c) Write down Levy's fundamental inversion theorem ($\frac{1}{2}$ page). [5]

d) Prove that if the characteristic function $\phi(t)$ is such that

$$\int_{-\infty}^{\infty} |\phi(t)| dt < \infty,$$

then the distribution must be continuous.

Start as follows: Let P be the distribution. Let a be any point. We shall prove that $P(a) = 0$. In fact, take any $\epsilon > 0$, we shall prove that for a suitable $h > 0$, $P(a - h, a + h) \leq \epsilon$.

Hint: $\left|\frac{\sin \theta}{\theta}\right| \leq 1$ for all $\theta \neq 0$. Use the result in (a).

($\frac{2}{3}$ page altogether). [15]

e) Prove that if the characteristic function $\phi(t)$ is such that $\int_{-\infty}^{\infty} |\phi(t)| dt < \infty$, then $\phi(t) \to 0$ as $t \to \infty$.

Use uniform continuity. The inversion theorem is not needed.

($\frac{2}{3}$ page). [15

GO ON TO THE NEXT PAGE

2.a)

$$P = \begin{pmatrix} \theta & 0 & 1-\theta \\ \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{pmatrix}.$$

Here $\theta = \frac{1}{2}$ if your roll number is a multiple of 3; $= \frac{1}{3}$ if your roll number is of the form 3n+1 ; $= \frac{2}{3}$ if your roll number is of the form 3n+2.

Write down the matrix $\lim_n P^n$. Briefly explain the method. Calculations need not be shown. ($\frac{1}{3}$ page). [10

b) Find the greatest weighted average of 8, 9, 12, and 15 if the weights are required to satisfy the following conditions:

i) The weight of 9 is $\geq \frac{1}{10}$.

ii) (The greatest weight MINUS the smallest weight) $\leq \frac{1}{2}$.

Briefly explain how you get your answer. ($\frac{1}{4}$ page). [10

3.a) In a Markov chain with 8 states, the transition matrix P has positive entries in the following places and 0 elsewhere.

(1,4), (1,8), (2,1), (2,6), (2,7), (3,3), (3,6), (4,4), (4,8), (5,1), (5,2), (6,3), (7,2), (7,5), (8,1), (8,8).

For example, $p_{11} = 0$, $p_{14} > 0$.

Draw the corresponding network. What are the transient states? What are the ergodic classes? How many ergodic classes are there? ($\frac{1}{2}$ page). [10

b) In the above network, add as many new arrows as you please so as to maximize the number of transient states (the given arrows must all be retained). Answer as follows: The new arrows are. The transient states are.....(4 lines). [10

4. What is the least positive integer M such that every integer $\geq M$ can be expressed as $8\alpha + 10\beta + 15\gamma + 1000\delta$ where the Greek letters represent non-negative integers? Briefly explain your method.

($\frac{1}{2}$ page) [10

INDIAN STATISTICAL INSTITUTE
Research and Training School
B. Stat. Part IV and M. Stat. Part I [420/520]

PERIODICAL EXAMINATION

Inference

Date: 3.5.71 Maximum Marks: 100 Time: 3 hours

Note: Questions 1 and 5 are compulsory. Attempt any two of the remaining three questions. Marks allotted for each question are given in brackets [ ].

1.a) X is a Poisson variable with an unknown parameter $\lambda(>0)$ truncated at zero, i.e.,

$$P_\lambda (X = x) = \frac{e^{-\lambda} \lambda^x}{(1-e^{-\lambda})x!}, \quad x = 1,2,3, \ldots .$$

Show that on the basis of a single observation X_1 on X, the only unbiased estimator of $1-e^{-\lambda}$ is given by

$$\delta(X_1) = \begin{cases} 0, & \text{when } X_1 \text{ is odd;} \\ 2, & \text{when } X_1 \text{ is even.} \end{cases}$$

Comment on the above estimator. [6+2]=[8]

b) When is a sequence of estimators said to be 'consistent'? Does the sequence of maximum likelihood estimators always form a consistent sequence of estimators? Let $X_1, X_2, \ldots$ be a sequence of independent and identically distributed random variables with finite k^{th} moment $\mu'_{2k} = E(X_1^{2k})$. Show that there exists a sequence of consistent estimators for $\mu'_k = E(X_1^k)$. [2+1+4]=[7]

2.a) Prove the Cramér-Rao inequality stating carefully the regularity conditions under which it is valid. [12

b) $X_1, \ldots, X_n$ are independent and identically distributed random variables with a common probability density function

$$p_\theta(x) = \frac{1}{\theta} e^{-\frac{x}{\theta}}, \quad \theta > 0, \quad 0 < x < \infty .$$

Show that there exists an unbiased estimator of θ based on $X_1, \ldots, X_n$, the variance of which attains the Cramér-Rao lower bound. [7]

c) Give an example of a uniformly minimum variance unbiased estimator of a parameter θ ($\theta \in \Theta$), the variance of which is strictly greater than the Cramér-Rao lower bound. Justify your conclusions. (You may use here any results proved in the class.) [2+4]=[6]

3.a) Let (X_1, X_2) have a joint density $p_{\theta_1,\theta_2}(x_1, x_2)$, where θ_1 and θ_2 are unknown parameters. Define

$$S_i = \frac{\partial \log p_{\theta_1,\theta_2}(X_1, X_2)}{\partial \theta_i} \quad (i = 1,2).$$

Also, let $\underset{\sim}{\Lambda} = ((\lambda_{ij})) = ((\text{Cov}(S_i, S_j)))$ denote the variance-covariance matrix of (S_1, S_2). Assume $\underset{\sim}{\Lambda}$ to be positive definite. Let $T = T(X_1, X_2)$ be any unbiased estimator of θ_1. Show that

GO ON TO THE NEXT PAGE

$$\mathrm{Var}_{\theta_1,\theta_2}(T) \geq \frac{\lambda_{22}}{\lambda_{11}\lambda_{22} - \lambda_{12}^2}$$

(You may assume any regularity conditions needed.) [10]

3.b) Take the case when (X_1, X_2) has a trinomial distribution with unknown parameters (π_1, π_2), i.e.,

$$p_{\pi_1,\pi_2}(n_1,n_2) = P_{\pi_1,\pi_2}(X_1 = n_1, X_2 = n_2) =$$

$$= \frac{n!}{n_1!\, n_2!\,(n-n_1-n_2)!}\, \pi_1^{n_1}\, \pi_2^{n_2}(1-\pi_1-\pi_2)^{n-n_1-n_2},$$

$(0 < \pi_1;\ \pi_2;\ \pi_1 + \pi_2 < 1)$. Use (a) to obtain lower bounds for the variance of unbiased estimators of π_1 and π_2. Show that there exist unbiased estimators of π_1 and π_2, the variances of which attain these bounds. [12+3]=[15]

4.a) Define 'sufficient statistics'. State carefully (without proof) the Factorization Theorem. [2+3]=[5]

b) Show that if a complete sufficient statistic exists, there exists a unique (a.e.) minimum variance unbiased estimator for an estimable parametric function. [10]

c) Let $X_1, \ldots, X_n$ be $n(\geq 2)$ observations on a Poisson variable with parameter λ $(\lambda > 0)$. Show that there exists a unique (a.e.) uniformly minimum variance unbiased estimator for $e^{-\lambda}$. [10]

5. Let $X_1, \ldots, X_n$ be n independent observations on a random variable X with a density

$$p_\theta(x) = \begin{cases} \frac{1}{\theta}, & \theta \leq x \leq 2\theta \\ 0, & \text{otherwise.} \end{cases}$$

Define $T_n = \max(X_1, \ldots, X_n)$, $U_n = \min(X_1, \ldots, X_n)$.

a) Show that (T_n, U_n) is jointly sufficient for θ. [2]

b) Write down the joint density of (T_n, U_n), and also the marginal densities of each of T_n and U_n. [3]

c) Show that $E(T_n) = \frac{2n+1}{n+1}\theta$, $E(U_n) = \frac{n+2}{n+1}\theta$,

$$\mathrm{Var}(T_n) = \mathrm{Var}(U_n) = \frac{n\theta^2}{(n+2)(n+1)^2}, \quad \mathrm{Cov}(T_n, U_n) = \frac{\theta^2}{(n+2)(n+1)^2}.$$

[10

d) Show that for any constants a and b, $E(aT_n + bU_n) = \frac{\theta}{n+1}[(2n+1)a + (n+2)b]$. Hence, show that (T_n, U_n) is not complete and any unbiased estimator of the form $aT_n + bU_n$ for θ must satisfy $(2n+1)a + (n+2)b = n+1$. [1+2+1]=[4

e) Show that the maximum likelihood estimator for θ is $\frac{1}{2}T_n$. [5

GO ON TO THE NEXT PAGE

f) Show that the maximum likelihood estimator given in (c) is biased, and an unbiased estimator of θ based on

$$T_n \text{ only is } Y_n = \frac{n + 1}{2n + 1} T_n.$$ [2]

g) Show that $Z_n = \frac{n + 1}{5n + 4}(2T_n + U_n)$ is another unbiased estimator of θ and

$$\mathrm{Var}\,(Z_n) = \frac{\theta^2}{(n + 2)(5n + 4)}.$$ [1+3]=[4]

h) Show that $\frac{\mathrm{Var}\,(Z_n)}{\mathrm{Var}\,(Y_n)} \leq 1$ for $n \geq 2$, the inequality being strict for $n > 2$. [2]

i) Comment on the result (h). [3]

INDIAN STATISTICAL INSTITUTE
Research and Training School
B.Stat. Part IV: 1970-71 [421]

ANNUAL EXAMINATION

Statistics-4: Probability

Date: 7.6.71 Maximum Marks: 100 Time: 3 hours

Note: The paper carries 113 marks. Attempt as many questions as you like. The maximum you can score is 100. Marks allotted for each question are given in brackets [].

The maximum permitted length of the answer is indicated after each question. Rough work must be clearly segregated from material meant for the examiner.

1.a) State Levy's basic inversion formula for the characteristic function of a probability distribution on the real line.
($\frac{1}{3}$ page) [3]

b) Give a specific counter-example which shows that if at least one of the end-points of a finite open interval is a discontinuity point (of the probability distribution function), then Levy's formula may NOT give the probability situated on that open interval. [10]
Calculations must be shown. ($\frac{1}{2}$ page)

c) What is the spectrum of a probability distribution (on the real line)? State a basic property of the spectrum.
(5 lines) [5]

d) Complete the sentence: The spectrum of a probability distribution consists of exactly one point if and only if - -.
(3 lines) [3]

e) X and Y are independent, identically distributed random variables. (X- Y) takes the value 0 with probability one. Prove rigorously that X takes a single value with probability one.
(Do not use characteristic functions)($\frac{2}{3}$ page) [12]

f) The characteristic function $\phi(t)$ of a probability distribution P satisfies the condition: $|\phi(t)| = 1$ for every (real) t. Prove that the whole probability is concentrated at a single point.
You may use the result of the previous question.
(1/2 page) [12]

g) What is the characteristic function of the rectangular (that is, uniform) distribution on (-1, +1)? Show your calculations.
(1/3 page) [5]

h) Use the above result, together with Levy's formula, to calculate

$$\int_0^\infty \frac{\sin^2 t}{t^2}\, dt.$$

(1/3 page) [8]

GO ON TO THE NEXT PAGE

2. a) In a Markov chain with 9 states, the transition matrix has positive elements in the following cells and zero in the other cells. (1,7), (1,9), (2,4), (2,6), (3,1), (3,3), (3,4), (4,8), (5,4), (5,6), (6,8), (7,7), (7,9), (8,2), (8,5), (9,1).

For example, $p_{11} = 0$ and $p_{17} > 0$.

Find out the transient states. Divide the nontransient states into ergodic classes. Divide each ergodic class into cyclically moving sub-classes.

(1 page) [20]

b)

$$P = \begin{Bmatrix} 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{3}{4} & 0 & \frac{1}{4} & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{2}{3} & 0 & \frac{1}{3} & 0 \end{Bmatrix}$$

What is the least positive integer k such that

$\lim_{n \to \infty} P^{kn}$ exists? Give reasons.

(1/2 page) [10]

c) Calculate any one element in $\lim_{n \to \infty} P^{kn}$. (P is the P above)

Briefly explain how you got the answer.

(1/2 page) [15]

d) What is the greatest weighted average of 6, 8 and 12 subject to the condition:

Greatest weight MINUS least weight is $\leq \frac{1}{2}$.

(2/3 page) [10]

INDIAN STATISTICAL INSTITUTE
Research and Training School
B.Stat. Part IV: 1970-71 [400]

ANNUAL EXAMINATION

Statistics-4: Inference

M. Stat. Part I: 1970-71 [505]

Inference-1

Date: 8.6.71 Maximum Marks: 100 Time: 3 hours

Note: Answer Groups A and B in separate answerscripts. Marks allotted for each question are given in brackets [].

Group A

Answer as much as you can. You can score a maximum of 55 marks.

1. Let $X_1, X_2, \ldots, X_n$ be $n(\geq 2)$ independent and identically distributed random normal variables with mean μ and variance σ^2 (>0). Define

$$\bar{X}_n = n^{-1} \sum_{i=1}^{n} X_i,\quad s_n^2 = (n-1)^{-1} \sum_{i=1}^{n} (X_i - \bar{X}_n)^2,$$

$$G_n = \frac{\sqrt{\pi}}{n(n-1)} \sum_{1 \leq i <} \sum_{j \leq n} |X_i - X_j|.$$

a) Show that $E(s_n^2) = \sigma^2$. Do you require the normality assumption for proving the result? Justify your answer. [2+1+2]=[5]

b) Find an expression for $E(s_n)$. Show that there exists some constant a_n (depending on n) such that $a_n s_n$ is the minimum variance unbiased estimator of σ. [3+3]=[6]

c) Use (a) to prove that $a_n > 1$. [4]

d) Show that G_n is an unbiased estimator of σ. [4]

e) Take the particular case $n = 2$. Find $\mathrm{Var}\,(G_2)$ and $\mathrm{Var}\,(a_2 s_2)$. [2+2]=[4]

f) Compute $\dfrac{\mathrm{Var}(a_2 s_2)}{\mathrm{Var}(G_2)}$ and comment on the result. [1+2]=[3]

2. Let $\{X_1, X_2, \ldots\}$ be a sequence of independent and identically distributed random variables with a common density function

$$p_\theta(x) = \begin{cases} e^{-(x-\theta)} & , x \geq \theta \\ 0 & , \text{otherwise}. \end{cases}$$

Let T_n denote the maximum likelihood estimator of θ based on $X_1, X_2, \ldots, X_n$.

a) Find an expression for T_n. [3]

b) Show that the sequence $\{T_n, n \geq 1\}$ of estimators is biased but consistent for θ. [4+5]=[9]

GO ON TO THE NEXT PAGE

c) Write down the expression for the minimum variance unbiased estimator of θ based on $X_1, X_2, \ldots, X_n$. Justify your answer. [1+3]=[4]

Group B Max. Marks: 65

3.a) Show that a most powerful randomized level α test always exists for a simple hypothesis against a simple alternative. [10]

b) Construct the m.p. randomised level α test for $\theta = \theta_0$ against $\theta = \theta_1 > \theta_0$, where θ is the parameter of a Poisson distribution. Examine if this test is u.m.p. for testing $\theta = \theta_0$ against $\theta > \theta_0$. [8]

4.a) Describe the method of 'least favourable distribution' for testing a composite hypothesis against a simple alternative. Prove a theorem which states that, under stated conditions, a distribution on the set of parameter values specified by the hypothesis (i) is a 'least favourable' distribution, and (ii) provides a m.p. size α test of the hypothesis. [10]

b) From a random sample of n observations from $N(\theta, 1)$, obtain a test for

H: $\theta \leq -\theta_0$ or $\geq \theta_0$, $\theta_0 > 0$

Alt: $\theta = 0$.

[Try the distribution $P(\theta = -\theta_0) = P(\theta = \theta_0) = 1/2$.] [8]

5.a) Prove that the level α m.p. test of a simple hypothesis against a simple alternative is necessarily unbiased. [4]

b) Let Θ denote the parameter space, Θ_0 the hypothesis and Δ the common boundary of Θ_0 and $\Theta - \Theta_0$. Prove that if a level α m.p. similar test of Δ against $\Theta - \Theta_0$ is also unbiased for Θ_0 against $\Theta - \Theta_0$, then it is necessarily a m.p. unbiased test for Θ against $\Theta - \Theta_0$. [8]

c) Show that the right-tail t test of level α is a m.p. similar test for $\theta = 0$ against $\theta > 0$ for $N(\theta, \sigma^2)$, where σ^2 is unknown. [10]

d) Examine if the same test is also unbiased. [7]

INDIAN STATISTICAL INSTITUTE
Research and Training School
B.Stat. Part IV: 1970-71 [423]

ANNUAL EXAMINATION

Statistics-5: Statistical Methods Theory

Date: 10.6.71 Maximum Marks: 100 Time: 3 hours

Note: Five questions carry full marks. You may attempt any 6 questions, but the maximum you can score is 100. All questions carry equal marks.

1. Let $X_1, \ldots, X_n$ be a random sample of n observations from a continuous distribution with distribution function $F(x)$. Let $x_{(1)} \ldots x_{(n)}$ denote the ordered sample.

a) Obtain the sampling distributions of the ν^{th} value from the bottom and ν^{th} value from the top in the sample.

b) Let $\xi = n(1 - F(x_\nu))$ and $\eta = n.F(x_{(n-\nu+1)})$. Show that ξ and η are, in the limit, as $n \to \infty$, independent.

2.a) Let $X_1, \ldots, X_m$ be m random variables. Define precisely the multiple correlation coefficient of X_1 on the remaining variables. Show that this is the maximum of the correlation betwee X_1 on the one hand and a linear combination of $X_2, \ldots, X_m$, on the other.

b) Show that this multiple correlation coefficient is zero if and only if the correlation coefficients of X_1 with X_i, for $i = 2, \ldots, m$, are all zeroes.

3. Let $\underset{\sim}{X}' = (X_1, X_2, \ldots, X_p)$ have a p-variate normal distribution with mean vector 0 and correlation matrix $((\rho_{ij}))$. Let $\underset{\sim}{X}_1, \ldots, \underset{\sim}{X}_n$ be a random sample from the population. Show that if the distribution of the sample correlation coefficient r_{ij} between X_i and X_j based on a random sample of size n is denoted by $F(r|n, \rho_{ij})$, then the distribution of the sample partial correlation coefficient $r_{ij.\,q+1,\ldots,p}$ between the i^{th} component and the j^{th} component, $1 \leq i, j \leq q$ eliminating the the effects of the components $X_{q+1}, \ldots, X_p$, based on a sample of n is $F(r|n-(p-q), \rho_{ij.q+1,\ldots p})$ where $\rho_{ij.q+1,p}$ denotes the partial correlation coefficient between X_i and X_j eliminating the effect of $X_{q+1}, \ldots, X_p$ in the population.

4. Show that the mean of a sample $\underset{\sim}{X}_1, \ldots, \underset{\sim}{X}_N$ of size N from $N(\mu, \Sigma)$ is distributed according to $N(\mu, \frac{1}{N}\Sigma)$ and is independent of $A = \sum_{\alpha=1}^{N} (\underset{\sim}{X}_\alpha - \bar{\underset{\sim}{X}})(\underset{\sim}{X}_\alpha - \bar{\underset{\sim}{X}})'$

where $\bar{\underset{\sim}{X}} = \frac{1}{N} \sum_{\alpha=1}^{N} \underset{\sim}{X}_\alpha$. Also show that A is distributed as $\sum_{\alpha=1}^{N-1} \underset{\sim}{Z}_\alpha \underset{\sim}{Z}_\alpha'$ where $\underset{\sim}{Z}_\alpha'$s are independent and each is distributed according to $N(0, \Sigma)$.

GO ON TO THE NEXT PAGE

5.a) Define Hotelling's T^2 and show its relation with Mahalanobis' D^2.

b) Obtain T^2-statistic as a function of the likelihood ratio criterion of the test of the hypothesis $\mu = \mu_0$ on the basis of a random sample from $N(\mu, \Sigma)$.

c) Show that the T^2 statistic is invariant under a non-singular transformation of the random vector.

6.a) Let $x_{\alpha}^{(i)}$, for $\alpha = 1, \ldots, n_i$, $i = 1, \ldots, k$, be random samples from $N_p(\mu^{(i)}, \Sigma)$, $i = 1, \ldots, k$, respectively. Obtain a test statistic for the hypothesis

$$\sum_{i=1}^{k} \beta_i \mu^{(i)} = \mu, \quad \text{where } \beta_1, \ldots, \beta_k$$

are given scalars and μ is a given vector.

b) Let $X_1, \ldots, X_n$ be a random sample from $N_p(\mu, \Sigma)$, where $\mu' = (\mu_1, \ldots, \mu_p)$. Obtain a test statistic for the hypothesis $\mu_1 = \ldots = \mu_p$.

7. Suppose $Z_1, \ldots, Z_n$ are independent $N_p(0, I)$ and $A = \Sigma Z_\alpha Z'_\alpha$. Show that A can be expressed as the product of a lower triangular matrix B_p with its transpose where the diagonal elements of B_p are independent χ-variables and the remaining $p(p-1)/2$ non-diagonal elements of B_p are independent $N(0, 1)$ variables.

8. Define Fisher's linear discriminant function which provides maximum discrimination between two populations denoted I and II. Show that this maximizes

$$\frac{[E_I(X'd) - E_{II}(X'd)]^2}{V(X'd)}$$

for all vectors d.

INDIAN STATISTICAL INSTITUTE
Research and Training School
B.Stat. Part IV: 1970-71 [424]

ANNUAL EXAMINATION

Statistics-5: Statistical Methods Practical

Date: 11.6.71 Maximum Marks: 100 Time: 3 hours

Note: Four questions carry full marks. You may attempt any part of any question. The maximum you can score in Questions 1 to 5 is 80. Marks allotted for each question are given in brackets [ ].

1.a) The following results were obtained at Syracuse University in an investigation of the factors influencing academic success. The sample consisted of 450 students and the variates were X_0 (honor points), X_1 (general intelligence) and X_2 (hours of study per week).

$$r_{01} = 0.60 \quad r_{02} = 0.32 \quad r_{12} = -0.35$$

where r_{ij} denotes the correlation coefficient between X_i and X_j.

Find to what extent honor points were related to general intelligence, when the effect of varying study periods was eliminated. [10]

b) In the bivariate case, show that the distance function D^2 reduces to the square of the geometrical distance between the two sample means in the x-y plane, when the sample correlation coefficient is zero.

Hence test the hypothesis of equality of mean vectors of the two groups for the following data:

Matrix of pooled variances and covariances

	x	y
x	4.7350	0
y	0	0.1451

Sample	Size	Values of Means x	y
1	25	25.80	7.81
2	40	28.35	7.41

Also test whether inclusion of y in addition to x increases the distance between the groups. [10]

2. Two populations of students taking university physics are distinguished as (1) those taking the standard elementary course and (2) those taking a somewhat more advanced course intended for better-prepared students. Discrimination is made on the basis of three measurements, X_1 (mathematics test score), X_2 (the A.C.E test score) and X_3 (the students honor-point ratio). For 111 students in the first course and 257 students in the second, the following results were recorded:

GO ON TO THE NEXT PAGE

	Course (1)	Course (2)
N	111	257
$\bar{x}_1$	87.640	92.397
$\bar{x}_2$	31.081	56.074
$\bar{x}_3$	1.1586	1.2689
$S(x_1 - \bar{x}_1)^2$	53136	194356
$S(x_2 - \bar{x}_2)^2$	11616	15864
$S(x_3 - \bar{x}_3)^2$	51.85	120.39
$S(x_1-\bar{x}_1)(x_2-\bar{x}_2)$	4863	17878
$S(x_1-\bar{x}_1)(x_3-\bar{x}_3)$	485.5	1844.0
$S(x_2-\bar{x}_2)(x_3-\bar{x}_3)$	243.8	836.6

Calculate the best linear discriminant function for distinguishing between the courses. If a new student comes along with the scores $x_1 = 80$, $x_2 = 40$ and $x_3 = 1.5$, to which course should he be assigned? [20]

3. The following data give the observations from two 3-variate normal distributions with mean vectors μ_1 and μ_2 and covariance matrices Σ_1 and Σ_2 respectively. Use Scheffe's solution to test the hypothesis that $\mu_1 = \mu_2$.

[Hint: Use the transformation

$$Y_\alpha = X_\alpha^{(1)} - \sqrt{\frac{n_1}{n_2}}\, X_\alpha^{(2)} + \frac{1}{\sqrt{n_1 n_2}} \sum_{\beta=1}^{n_1} X_\beta^{(2)} - \frac{1}{n_2} \sum_{\gamma=1}^{n_2} X_\gamma^{(2)},$$

where n_1 and n_2 are the sample sizes and $n_1 < n_2$]

$X_1^{(1)}$	$X_2^{(1)}$	$X_3^{(1)}$	$X_1^{(2)}$	$X_2^{(2)}$	$X_3^{(2)}$
7.0	2.7	2.2	5.1	1.4	2.5
6.4	1.2	2.2	4.9	1.4	2.0
6.9	2.1	2.8	4.7	1.3	2.2
5.5	1.3	2.7	4.6	1.5	2.1
6.5	1.8	3.1	5.0	1.4	2.6
5.7	1.8	3.0	5.4	1.7	2.9
6.3	2.3	2.9	4.6	1.4	3.1
4.9	1.9	2.6	5.0	1.5	3.4
			4.4	1.3	1.9
			4.9	1.2	2.8

[20]

4. Let the components of X correspond to scores on tests in arithmetic speed (X_1), arithmetic power (X_2), memory for words (X_3) and memory for meaningful symbols (X_4). The observed correlations in a sample of size 140 are:

1.0000	0.4248	0.0420	0.0215
	1.0000	0.1487	0.2489
		1.0000	0.6915
			1.0000

Obtain the best set of regressor variables for regression of X_1 on two of X_2, X_3 and X_4. [20]

GO ON TO THE NEXT PAGE

5. The following give the mean values of measurements on 100 samples of four morphological characters of a species of plant, and also the matrix of covariances:

	X_1	X_2	X_3	X_4	Means
X_1	13.86	6.97	1.08	2.32	36.58
X_2		4.74	0.56	1.46	40.00
X_3			0.14	0.22	7.81
X_4				0.57	6.74

Test whether the sub-vectors $X^{(1)} = \binom{X_1}{X_3}$ and $X^{(2)} = \binom{X_2}{X_4}$ have equal means. [20]

Practical Record [10]

Viva Voce [10]

INDIAN STATISTICAL INSTITUTE [425]
Research and Training School
B.Stat. Part IV: 1970-71

ANNUAL EXAMINATION

Statistics-7: Econometrics Theory and Practical

Date: 14.6.71 Maximum Marks: 100 Time: 3 hours

Note: Answer Q.6 and any three of the rest.
Marks allotted for each question are given in brackets [].

1.a) 'The production function is a technical engineering relation between inputs and output'. Is the purity of this relation always retained in the empirical estimation of the production function? Give reasons for your answer.

b) How do you propose to solve the problem of aggregation of the Cobb-Douglas function in order to arrive at a relation between aggregate inputs and aggregate output? Can you use this relation to estimate the value of capital for all industries? What are the assumptions underlying your method? [15]

2. Discuss how you would find the best linear unbiased estimates of the parameters in a regression model in the presence of auto-correlated disturbances. Explain a practical procedure in which simple least squares method may be applied when the disturbances follow a first-order auto-regressive structure. [15]

3. Why does the log-normal distribution usually fit the distribution of personal income better than the normal distribution?

Indicate a method of testing whether a given grouped empirical income distribution obeys the log-normal law. How will you proceed with the business of fitting the log-normal distribution to the empirical distribution if your test is satisfied? [3+5+7]=[15]

4. Explain the problem of multi-collinearity as it appears in econometric analysis. What are the difficulties created by its presence and how do you propose to remove them? [15]

5. Write a note on the problem of heteroscedasticity, mentioning situations where the problem is met and the steps which may be taken to overcome this problem. [15]

6. Draw the concentration curves of per capita monthly total expenditure and per capita monthly expenditure on milk and milk products and also compute the related concentration ratios from the following data:

GO ON TO THE NEXT PAGE

Table: Average consumption expenditure in Rs. per person per 30 days on all items ($\bar{x}_j$) and on milk and milk products ($\bar{y}_j$) by persons in different classes of per person total expenditure per 30 days.

expenditure class (Rs.)	percent of population (p_j)	$\bar{x}_j$	$\bar{y}_j$
0 - 8	9.03	6.18	0.08
8 - 11	14.73	9.49	0.21
11 - 13	11.33	12.01	0.46
13 - 15	9.87	13.98	0.54
15 - 18	13.05	16.44	0.95
18 - 21	9.48	19.53	1.32
21 - 24	8.49	22.46	1.73
24 - 28	6.96	25.66	2.03
28 - 34	6.17	30.34	2.56
34 - 43	5.34	38.26	4.24
43 - 55	2.75	48.23	5.65
55 -	2.80	89.45	7.01
All classes	100.00	20.13	1.43

[35]

Practical Record [10]

Viva Voce [10]

INDIAN STATISTICAL INSTITUTE [426]
Research and Training School
B.Stat. Part IV: 1970-71
ANNUAL EXAMINATION
Statistics-7: Planning Techniques

Date: 15.6.71

Maximum Marks: 100 Time: 3 hours

Note: Answer Q.1 and any three of the rest.
Marks allotted for each question are given in brackets [ ].

1.a) Solve the following linear program by either Charnes' - M technique, or, two-phase method:

$$\begin{aligned} \text{maximise} \quad & 2x_1 + x_2 + 4x_3 \\ \text{subject to} \quad & x_1 + 3x_2 + 2x_3 \leqq 20 \\ & 2x_1 + 16x_2 + x_3 \geqq 4 \\ & 3x_1 - x_2 - 5x_3 \geqq -10 \\ & \text{all } x_j \geqq 0 \end{aligned}$$ [10

b) Is the optimal solution of (a) unique? [4

c) Find the optimal solution of the dual to (a) and explain the special feature of this dual solution. [

2. Consider a system of m simultaneous linear equations in n unknowns $(n \geqq m)$, $Ax = b$, with $r(A) = m$, $x \geqq 0$.

a) Define a basic feasible solution of the system. [

b) What is the connection between the basic feasible solutions of the system and the extreme points of the set of its feasible solutions? [2

3.a) Consider the following standard maximum problem:

$$\begin{aligned} \text{maximise} \quad & x_1 - x_2 \\ \text{subject to} \quad & -2x_1 + x_2 \leqq 2 \\ & x_1 - 2x_2 \leqq 2 \\ & x_1 + x_2 \leqq 5. \end{aligned}$$ [

Show that this problem has the optimal solution

$$x_1 = 4, \quad x_2 = 1$$

by finding a solution of the dual problem making use of the equilibrium theorem.

b) Let $x = \{x_j\}$ be a non-negative solution of $Ax = b$ and let $y = (y_i)$ be a solution of $yA \geqq c$.

Prove that x maximises cx and y minimises yb if and only if

$$x_j = 0 \text{ whenever } ya_j > c_j$$

where a_j is the j-th column of A and c_j is the j-th component of c. [

GO ON TO THE NEXT PAGE

3.c) Solve the following canonical maximum problem by computing all of its basic (not necessarily non-negative) solutions:

$$\begin{array}{ll} \text{maximise} & 2x_1 + 3x_2 \\ \text{subject to} & 4x_1 + 2x_2 + x_3 = 4 \\ & x_1 + 3x_2 \quad = 5 \end{array}$$

check by solving the dual. [9]

4. Consider a Leontief system: $(I - A)x = c$ where A is an $n \times n$ non-negative matrix, x and c are n-component vectors. Prove that for a given $c > 0$, the above system has a non-negative solution if and only if the Hawkins-Simon condition is satisfied, namely that the matrix $(I - A)$ has the n positive upper left-hand corner principal minors. [25]

5. Give a linear programming interpretation of the Leontief Static Open Model. [25]

6. 'It is a remarkable implication of the Leontief system that even if there were available several different processes for each industry, only one of them would ever be observed'. Elucidate. [25]

INDIAN STATISTICAL INSTITUTE
Research and Training School
: ANNUAL EXAMINATION. [427]

B.Stat. Part IV: 1970-71

Statistics-8: Genetics

M. Stat. Part I: 1970-71 [551]

Applied Statistics-2: Biostatistics
Theory and Practical

Date: 16.6.71 Maximum Marks: 100 Time: 3 hours

Note: The paper carries 114 marks. You may answer any part of any question. The maximum you can score is 100. Marks allotted for each question are given in brackets [ ].

1. Each of the following statements is either true or false. Indicate, without reasons, which is true and which is false.

   a) Mendel's laws of heredity are not at all applicable to higher animals like man.

   b) Lamark's theory of evolution was based upon inheritance of acquired characters.

   c) A population attains equilibrium with respect to any autosomal character in one generation of random mating.

   d) The frequency of a dominant gene gradually increases at the cost of its recessive allele just by virtue of its dominance.

   e) A defect which is controlled by a recessive gene on the X-chromosome occurs equally frequently in males and females.

   f) If both the parents have the blood group 'A' (referring to ABO blood group system), their children may have the blood groups 'A' or 'O'.

   g) All blood group systems such as ABO, MN etc. are controlled by the same locus.

   h) The relative frequencies of different alleles at a locus need not necessarily add to one.

   i) Haemophilia is an autosomal character.

   j) Random mating means no inbreeding. [10]

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

   a) Hardy-Weinberg law for autosomal characters.

   b) Bernstein's method of estimating ABO gene frequencies.

   c) Effect of a mixture of random mating communities on genotypic frequencies.

   d) One-way recurrent mutation.

   e) Mendel's laws of heredity. [8 X 3]=[24

3. Discuss the equilibrium for an X-linked character under Hardy-Weinberg law. [15

4. Describe, in detail, the maximum likelihood method of estimating ABO gene frequencies under random mating. [15

GO ON TO THE NEXT PAGE

5. Show that selfing eventually leads to complete homozygosity. [10

6. For an autosomal locus with two alleles, derive the genotypic frequencies under Wright's equilibrium law, starting from first principles. Hence write down the phenotypic frequencies for ABO blood group system. [12+3]=[15

7. Consider two gametes A and a with relative fitness coefficients (1 - S) and (1 - s) respectively, where s > S.

   a) Show that the gamete a eventually becomes extinct in the absence of mutation. [10]

   b) Introducing mutation from A ⟶ a at a constant rate of μ per generation, derive the stable equilibrium gametic frequencies. (You may assume that s is much larger than S.) [15

INDIAN STATISTICAL INSTITUTE [428]
Research and Training School
B.Stat. Part IV: 1970-71

ANNUAL EXAMINATION

Sample Surveys (Theory and Practical)

Date: 17.6.71 Maximum Marks: 100 Time: 3 hours

Note: Answer all the questions. Marks allotted for each question are given in brackets [].

1. From the following list of villages in a tehsil, draw 5 villages, one by one with replacement, the probability of selecting a particular village in any one draw being proportional to the number of households in the village as obtained from district census handbooks.

[You should use the table of random sampling numbers supplied to you and present the details of the procedure used by you in a neat tabular way.]

Village-wise data for a tehsil

Sl. No.	No. of households	Sl. No.	No. of households	Sl. No.	No. of households
1	757	8	984	16	718
2	278	9	523	17	1334
3	548	10	1197	18	812
4	1064	11	455	19	472
5	873	12	993	20	240
6	763	13	758		
7	541	14	198		
		15	445		

[30

2. The sample of five villages drawn above is to be used for estimating the following parameters:

1) the total population of the tehsil,
2) the total expenditure on food in a month in the tehsil and
3) the per capita monthly expenditure on food.

It is not possible to investigate all the households in each of the five selected villages and second stage sampling of households is to be undertaken, using the method of simple random sampling with replacement. For this purpose, a list is to be prepared of all the households in each of the five selected villages. Note that the actual number of households in these villages is likely to be different from the number obtained from district census handbooks which are generally out of date.

a) It has been decided to draw a total of 500 households from these five selected villages. How will you determine the number of households to be drawn from each of the selected villages?

Discuss in this connection the advantages and disadvantages of (i) selecting equal number of households from each village (ii) equal proportion of households from each village (iii) choosing the number to ensure a self-weighting design. [25]

b) How will you estimate the three parameters? Are these estimates unbiased? Work out the sampling variance of these estimators. [25]

c) How will you estimate the sampling variance of the estimated total population? [20]
