INDIAN STATISTICAL INSTITUTE
Research and Training School
B.Stat.(Hons.) Part IV: 1974-75

FINAL EXAMINATION

.OFFICIAL STATISTICS (Conducted by CSO)

[40/533]

Date: 2.8.1974 Maximum Marks: 100 Time: 3 hours.

[Note: Attempt any five questions. All questions carry equal marks].

1. "The statistical system in each country represents that country's solution to meet the general functions of a statistical system in terms of the national needs, national interest and resources, and the conditions - economic, technical and political - in the country". Discuss with particular reference to the Indian Statistical System.

OR

Describe the statistical system in any foreign country.

2. Describe the uses of population data to an administrator, planner, sociologist and the like. How sampling was made use of in 1961 and 1971 population censuses ?

3. Write an explanatory note on the present availability of agricultural statistics in India mentioning, inter-alia, major gaps and the steps being undertaken in filling up these.

4. Describe the different sources of information on employment/unemployment statistics in India bringing out the limitations of data along with suggestions for putting them on a firm footing.

5. Briefly describe the scope, content and methods adopted in the collection of statistics of manufacturing industries. What attempts are being made to fill in the gaps in data in the unorganised industrial sector ?

6. Describe the nature of official data collected, compiled and the agencies responsible in three of the following fields :

(i) Statistics of Joint Stock Companies, (ii) Financial Statistics,
(iii) Foreign Trade Statistics, and (iv) Income Tax Statistics.

7. Describe the organisational set up, methods of data collection, classification as also the coverage of any three of the following :

(i) Family Planning Statistics, (ii) Housing and Construction Statistics,
(iii) Health Statistics and (iv) Occupational Classification.

8. Give an account of any one of the official price index numbers mentioning the agencies responsible, the base period, the item coverage and the mo. of determination of weighting diagram. How is the index used in the official system ?

9. Describe the methods of national income estimation with particular reference to India.

INDIAN STATISTICAL INSTITUTE
Research and Training School
B. Stat.(Hons.) Part IV: 1974-75

[401]

PERIODICAL EXAMINATION

Statistical Methods

Date: 25.11.74 | Maximum Marks: 100 | Time: $2\frac{1}{2}$ hours

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

1.a) Let T_n be a sequence of statistics such that

$$\sqrt{n}\,(T_n - \theta) \xrightarrow{w} N(0, \sigma^2(\theta))$$

Show that, if $g(x)$ is differentiable at θ and $g'(\theta) \neq 0$, then

$$\sqrt{n}\,\left(g(T_n) - g(\theta)\right) \xrightarrow{w} N(0, [g'(\theta)\sigma(\theta)]^2).$$

b) State the multivariate version of the result in (a).

c) Using the result in (a) show that the asymptotic variance of

$$\sin^{-1}\sqrt{r/n}\,,$$

where r/n is the binomial proportion, is independent of π.

2.a) Prove that if $f_n(x) \to f(x)$ a.e. and $f_n(x)$ and $f(x)$ are probability densities, then $P_n(B) \to P(B)$ uniformly over all Borel sets.

b) State the discrete version of the above theorem.

3.a) Derive the likelihood ratio test to test the hypothesis $\mu_1 = \mu_2 = \ldots = \mu_k$ on the basis of samples of size n_i from $N(\mu_i, \sigma^2)$, $i = 1,2,\ldots,k$ when the common σ^2 is unknown. Show that the test reduces to the analysis of variance test for one-way classification.

b) Derive the likelihood ratio test for testing the homogeneity of variances on the basis of samples of size n_i from $N(\mu_i, \sigma_i^2)$ $i = 1,2,\ldots,k$.

4. Consider the trinomial with cell probabilities

$$\pi_1 = \pi_2 = \theta, \quad \pi_3 = 1 - 2\theta\,, \quad 0 < \theta < \tfrac{1}{2}\,.$$

(a) Derive the maximum likelihood and minimum chi-square estimators of θ.

(b) Show that in both cases,

$$Q^2 = \frac{\Sigma\,(n_i - n\pi_i(\hat{\theta}))^2}{n\,\pi_i(\hat{\theta})}$$

is asymptotically distributed as χ^2 with 1. d.f.

Consider the multinomial with cell probabilities $\pi_1, \pi_2, \ldots, \pi_k$. In the standard notation, let

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

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

p.t.o.

5. i) Show that $V_n \xrightarrow{w} N(0, I - \phi\phi')$.

ii) Obtain necessary and sufficient conditions for $V_n' C V_n$ to be asymptotically distributed as a χ^2.

iii) For the case $k = 2$, $\pi_1 = \pi_2 = 1/2$, $C = \begin{pmatrix} a & b \\ b & c \end{pmatrix}$ Show that the necessary and sufficient condition for $V_n' C V_n \overset{a.d.}{\sim} \chi_1^2$ is that $b = \frac{a+c}{2}$ or $\frac{a+c}{2} - 1$.

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

i) Kolmogorov-Smirnov Test

ii) Cramer-Wold Theorem and its applications.

iii) Large sample test for testing the homogeneity of correlations.

iv) Weak convergence of probability measures - various equivalent definitions.

INDIAN STATISTICAL INSTITUTE [402]
Research and Training School
B. Stat.(Hons.) Part IV: 1974-75

PERIODICAL EXAMINATION

Statistical Quality Control

Date: 2.12.74 Maximum Marks: 50 Time: 2 hours

Note: The paper carries 56 marks. Answer any part of any question. The maximum you can score is 50. Marks allotted for each question are given in brackets [ ].

1.a) When do you say that a process is under 'Statistical Control' ?

b) What is a control chart ? What is the purpose of setting up a control chart ?

c) Write down clearly, the steps involved in setting up a c-chart.

OR

Describe various out of control situations in a control chart. [3+5+7]=[15]

2.a) It is known that a certain measurable dimension of a product is normally distributed with parameters μ and σ (known). U and L are upper and lower specifications for the product dimension. If dimension of any item is above U it is to be scrapped at a cost of Rs.c_1 per item. If the dimension is below L rework is possible at a cost of Rs.c_2 per item. Assume that $c_1 > c_2$. Derive a suitable expression to determine optimum setting level. [7]

b) The resistance in ohms of a certain electrical device is specified as 200 $\pm$ 15. A control chart is run on the manufacturing process with samples of 4 taken from the production line every hour for 20 hours.

$$\Sigma \bar{X} = 4,140 \quad \text{and} \quad \Sigma R = 288$$

i) What should be the control limits on $\bar{X}$ and R charts?

ii) Assume that (a) all the points on both the charts fall within control limits (b) resistances follow normal distribution. What percentage of defective articles will be produced ?

iii) Determine the process capability. [4+4+3]=[11]

3. A single sampling attributes inspection plan is n = 200 and c = 4.

a) Compute the probabilities of acceptance of lots 0.5%, 1%, 1.5%, 2%, 2.5%, 3%, 4% and 5% defective.

b) Plot the O.C. curve for the plan and determine IQL.

c) Determine the lot qualities corresponding to C.R.=10% and P.R. = 5% .

d) Draw the AOQ curve and determine AOQL.

[4+5+4+10]=[23]

INDIAN STATISTICAL INSTITUTE [403]
Research and Training School
B.Stat.(Hons.) Part IV: 1974-75

PERIODICAL EXAMINSTION

Inference

Date: 6.1.75 Maximum Marks: 100 Time: 3 hours

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

1.a) Let $F(x, \theta)$ $\theta \in \underline{\text{(H)}}$ be a family of distributions. Let $x_1, \ldots, x_n$ be random sample of size n and $T(x_1, \ldots, x_n)$ is a mapping from $(x_1 \ldots x_n)$ to $\underline{\text{(H)}}$. Explain when T is said to be sufficient for the above family of distributions.

b) If $F(x, \theta)$, $\theta \in \underline{\text{(H)}}$ is a family of discrete distributions and $\underline{\text{(H)}}$ is a subset of the real line obtain the Neyman factorization of the joint distribution of $X_1, \ldots, X_n$ corresponding to the sufficient statistic $T(X_1, \ldots, X_n)$.

c) For the family of Normal distributions $N(\mu, \sigma^2)$ $-\infty < \mu < \infty$ $0 < \sigma^2 < \infty$ show that

$$\bar{X} = \frac{1}{n} \Sigma X_i \quad \text{and} \quad s^2 = \frac{1}{n-1} \Sigma (X_i - \bar{X})^2$$

are jointly sufficient.

2. i) When is a family of distributions $F(x, \theta)$, $\theta \in \underline{\text{(H)}}$ said to be (a) complete, (b) boundedly complete ?

ii) Show that the family of binomial distributions $\underline{B}(n, p)$ $0 \leq p \leq 1$ is complete.

iii) Let a random variable X take values 1 and 0 with $P(X = 1) = p$ and $P(X = 0) = 1 - p$ where $0 < p \leq 1$ and p is unknown. Based on a sample of n observations obtain a UMVU estimate of $p(1-p)$ and show that it is unique.

3.a) Let $F(x, \theta)$, $\theta \in \underline{\text{(H)}}$ be a family of distributions. Let $g(\theta)$ be a parametric function to be estimated on the basis of a sample of size n. Under suitable regularity conditions (to be clearly stated by you) obtain the Cramér-Rao lower bound for the variance of any unbiased estimate with finite variance of $g(\theta)$.

b) Under the regularity conditions you assumed in (a) is it always possible to obtain an unbiased estimate with variance equal to the Cramér-Rao lower bound for any parametric function $\lambda(\theta)$? If your answer is 'Yes' give a proof and if it is 'No' give an example to support your answer.

4.a) In the usual general set up, Let $T_1(X_1, \ldots, X_n)$ and $T_2(X_1, \ldots, X_n)$ be two unbiased estimates of a parameteric function $g(\theta)$ with variances σ_1^2 and σ_2^2 respectively and correlation coefficient ρ. Obtain that linear function of T_1 and T_2 which is minimum variance unbiased estimate of $g(\theta)$. Obtain the minimum variance. ...

p.t.o.

4.b) If in the above case T_1 is an unbiased minimum variance estimate and T_2 is any other unbiased estimate with finite variance show that the correlation between T_1 and T_2 is given by

$$\sqrt{\frac{\text{Var } T_1}{\text{Var } T_2}} .$$

5.a) Let $F(x, \theta)$, $\theta \in \textcircled{H}$ be a family of distributions. Let $X_1 \ldots X_n$ be a sample of size n. Under suitable regularity conditions to be stated by you, show that the likelihood equation has a root with probability going to 1 as n tends to ∞ which is a consistent estimate of θ.

b) If the family in (a) admits a sufficient statistic $T(X_1, \ldots, X_n)$ for all n show that the maximum-likelihood estimate will be an explicit function of T.

c) Consider the usual linear model set up

$$\underset{\sim}{Y}_{(n \times n)} = X_{(n \times m)} \underset{\sim}{\beta}_{(m \times 1)} + \underset{\sim}{\epsilon}_{(n \times 1)}$$

where $E(\underset{\sim}{\epsilon}) = \underset{\sim}{0}$

and $D(\underset{\sim}{\epsilon}) = \sigma^2 I_{(n \times n)}$,

D standing for the dispersion and I for the unit matrix.

$X_{(n \times m)}$ is a matrix of known coefficients and is of rank $m (\leq n)$.

(a) Obtain the least squares estimate $\hat{\underset{\sim}{\beta}}$ of $\underset{\sim}{\beta}$.

(b) If $\underset{\sim}{P}' \underset{\sim}{\beta}$ is estimable, show that $\underset{\sim}{P}' \hat{\underset{\sim}{\beta}}$ has minimum variance in the class of linear unbiased estimates of $\underset{\sim}{P}' \underset{\sim}{\beta}$.

INDIAN STATISTICAL INSTITUTE
Research and Training School
B.Stat.(Hons.) Part IV: 1974-75

[403]

PERIODICAL EXAMINATION

Measures and Probability

Date: 13.1.75 | Maximum Marks 100 | Time: 3 hours

Note: The whole paper carries 110 marks. Answer as much as you can. The maximum you can score is 100. Marks allotted for each question are given in brackets [ ].

1.a) Define a σ-ring. Show that given any class $\underline{C}$ of sets there is a σ-ring $\underline{S}$ containing $\underline{C}$ such that any σ-ring containing $\underline{C}$ is contained in $\underline{S}$. Show that $\underline{S}$ is unique. [8]

b) Let $\underline{S}$ be a σ-ring of subsets of X and μ be a measure on $\underline{S}$. Show that the class of sets $E \in \underline{S}$ with σ-finite measure (i.e., there exists a sequence of sets with finite measure whose union is E) is a σ-ring [7]

c) If $\underline{C}$ is any class of subsets of X and E belongs to the σ-ring generated by $\underline{C}$, then show that E belongs to the σ-ring generated by a countable collection of sets from $\underline{C}$. [8]

2.a) Define a measurable function on a measurable space. [3]

b) If f and g are measurable functions, show that $f+g$ is measurable. Is the converse true ? Why ? [6]

c) If f is measurable, show that $|f|$ is measurable. Is the converse true ? Why ? [4]

d) If the σ-field of a measurable space has only finitely many sets, show that a function is measurable iff it is simple. [5]

3. What is the biggest class of functions on a measure space for which the integral is defined? Give all the important steps (without proofs) in defining the integral for functions in this class. [10]

4. Z is the set of positive integers, $\underline{F}$ = the class of all subsets of Z and $\mu(E)$ = number of elements of E

a) When is a function $f : Z \to R$ measurable ? [4]

b) Show that a function f is integrable iff $\sum_{n=1}^{\infty} |f(n)|$ is finite. Also then show that $\int f d\mu = \sum_{n=1}^{\infty} f(n)$ [12]

c) Deduce from the result in (b) that the sum of an absolutely convergent series is unaffected by a rearrangement of the terms. [6]

d) If $\sum_{n=1}^{\infty} f(n)$ converges then show by examples that $\int f d\mu$ may or may not exist. [6]

p.t.o.

5. State and prove the monotone convergence theorem. [10]

6. a) Show that if f_n is a sequence of measurable functions such that $f_n \geq g$ where g is integrable then

$$\int \liminf f_n \, d\mu \leq \liminf \int f_n \, d\mu$$ [8]

b) Show by an example that the existence of integrable g can not be dropped in (a). [4]

c) Deduce from the result in (a) that if A_n is any sequence of measurable sets then

$$\mu(\liminf A_n) \leq \liminf \mu(A_n).$$ [5]

State the analogous result for lim sup of a sequence of measurable sets. [3]

INDIAN STATISTICAL INSTITUTE
Research and Training School
B.Stat.(Hons.) Part IV: 1974-1975

[405]

PERIODICAL EXAMINATION
Multivariate Analysis

Date: 20.1.75 Maximum Marks: 100 Time: 3 hours

Note: (1) This is an OPEN-BOOK examination.
(2) Correct answers to about 5/6 of the paper will carry full marks. You may attempt any part of any question.
(3) Numbers in [ ] indicate marks allotted to the questions.

1.a) Show that the dispersion matrix of a p-dimensional random variable is non-negative definite. [6]

b) Show that in a p-variate distribution, all the correlations are equal, say ρ, only if

$$\rho \geq -\frac{1}{p-1}.$$ [11]

c) Show that the dispersion matrix of a p-dimensional random variable is singular if and only if there exists a constant linear relationship among the p components. [10]

2. Let random variables X_i, $i = 1,2,\ldots, k$ be independent, each X_i having a univariate normal distribution with mean $\beta + rz_i$, and variance σ^2, where the z_i's are given numbers such that

$$\sum_{i=1}^{k} z_i = 0.$$

Find the joint distribution of $\bar{X}$ and Y, where

$$\frac{\sum_{i=1}^{k} z_i X_i}{\sum_{i=1}^{k} z_i^2}, \quad \text{for} \quad \sum_{i=1}^{k} z_i^2 > 0$$

and $\bar{X}$ is the mean of the X_i's. [12]

3. Let a p-dimensional random variable $(X_1, X_2, \ldots, X_p)$ be distributed as a multivariate normal with mean vector μ and a nonsingular dispersion matrix $((\sigma_{ij}))$, $i,j = 1,2,\ldots, p$.

a) Find the distribution of the (p+1)-dimensional random variable $(X_1, X_2, \ldots, X_p, \lambda_1 X_1 + \lambda_2 X_2 + \ldots + \lambda_p X_p)$ where $\lambda_1, \lambda_2, \ldots, \lambda_p$ are given constants. [6]

b) Find the conditional joint distribution of $X_1, X_2, \ldots, X_p$ given $\lambda_1 X_1 + \lambda_2 X_2 + \ldots + \lambda_p X_p = c$, c being a constant. [8]

c) Show that the covariance between X_i, X_j conditional upon $\lambda_1 X_1 + \lambda_2 X_2 + \ldots + \lambda_p X_p = c$, is

$$\sigma_{ij} - \frac{\left(\sum_{k=1}^{p} \lambda_k \sigma_{ik}\right)\left(\sum_{k=1}^{p} \lambda_k \sigma_{jk}\right)}{\sum_{m=1}^{p} \sum_{k=1}^{p} \lambda_k \lambda_m \sigma_{km}}$$

for $i, j = 1,2,\ldots, p$. [12]

p.t.o.

4. If a symmetric matrix $S = ((s_{ij}))$, $i, j = 1, 2, \ldots, p$ has a Wishart distribution with m degrees of freedom based on a matrix Σ, find the distribution of the matrix $((s_{ij}))$, $i, j = 1, 2, \ldots, r < p$, with its parameters. [9]

5. Show that based on a random sample from a p-dimensional normal distribution with mean vector μ and non-singular dispersion matrix Σ, the sample mean vector and the sample dispersion matrix are jointly sufficient for μ, Σ. [13]

6. Let X, Y, Z be each a $1 \times p$ random variable. Let (X, Y, Z) have a 3p-dimensional normal distribution with mean vector (λ, μ, ν) and non-singular dispersion matrix Σ, where $E(X) = \lambda$, $E(Y) = \mu$, $E(Z) = \nu$. We have a random sample of size n from (X, Y, Z). Derive a suitable test of the hypothesis

$$H_1 : 3\lambda = \mu + 2\nu$$

$$\text{against } H_2 : 3\lambda \neq \mu + 2\nu .$$

[13]

7. In an anthropological study of Egyptian skulls, the researcher has four series of skulls, 91 predynastic, 162 from the sixth to the twelfth dynasties, 70 from the twelfth and the thirteenth dynasties and 75 from Ptolemaic dynasties, 398 in all. On each skull, four measurements (in mm.)

x_1 = maximum breadth

x_2 = basi-alveolar height

x_3 = nasal height

x_4 = basi-bregmatic height

are made.

Making suitable assumptions, formulate the observational set-up as a multivariate linear model. Discuss with the help of clearly stated formulae, etc., how you would analyse data from this study to find out if there are differences in the mean vectors of the skull measurements in the four series. [20]

INDIAN STATISTICAL INSTITUTE
Research and Training School
B.Stat. (Hons.) Part IV: 1974-75

[406]

PERIODICAL EXAMINATION

Economics

Date: 27.1.75 Maximum Marks: 50 Time: 2 hours

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

1. In what sense, and to what extent, did the process of modern economic development in the developed countries of the world to-day involve a 'movement away from agriculture'? Suggest some explanations of this movement with a clear statement of the logical connections presumed in each.

2. Clearly bring out the logical inter-relationship between economic growth in the modern era and 'industrialisation'. Would you assign the same relative importance to the different factors in this inter-relationship for viewing the historical association between the two in today's developed countries on the one hand, and for viewing the future prospects of the underdeveloped countries on the other? Defend your stand.

3. How would you define the so-called 'tertiary' sector of an economy? What are its principal components and their relation to different economic activities? Briefly describe the behaviour, over the period of modern economic growth in the present-day developed countries, of the share of this sector in labour force and national product, and comment on the significance of the composition within the sector for an interpretation of the behaviour.

INDIAN STATISTICAL INSTITUTE [408]
Research and Training School
B.Stat.(Hons.) Part IV: 1974-75

MID-YEAR EXAMINATION

Inference

Date: 13.2.75 Maximum Marks: 100 Time: 3 hours

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

1.a) Obtain a set of sufficient conditions for a sequence of estimates $\{T_n(X_1, \ldots, X_n) \; n \geq 1\}$ to be consistent for a parametric function $\vartheta(\theta)$ when the variables are i.i.d. with frequency function $\{f(X, \theta), \theta \in (\underline{H})\}$ and show the sufficiency of the conditions.

b) Show by means of examples that a consistent sequence of estimates $\{T_n\}$ need not be unbiased for any n and conversely a sequence of estimates which is unbiased for all n need not be consistent. [10+10]=[20]

2.a) Suppose the density function of a random variable is $\{f(X, \theta), \theta \in (\underline{H})\}$. It is required to estimate the parametric function $\vartheta(\theta)$. The loss function $L(\partial(X) \; \vartheta(\theta))$ is assumed to be a continuous convex function in ∂ for each θ. Suppose the family of distributions admits a sufficient statistic. T. Suppose $g(X)$ is an estimate of $\vartheta(\theta)$. Show that there exists an estimate which is a function of T alone whose risk is no larger than that of $g(X)$.

b) Use the above result to obtain a uniformly minimum variance unbiased estimate of $e^{-\theta}$ from a sample of size n from a poisson distribution with parameter θ. [12+8]=[20]

3. Under the regularity conditions of Cramer-Rao inequality obtain the form of the density function admitting unbiased estimates with variance equal to the Cramer-Rao lower bound. What is the form of the parametric function for which such an estimate exists? Illustrate your answer by an example. [10+4+6]=[20]

4.a) Under the usual regularity conditions show that any consistent solution of the likelihood equation provides a maximum of the likelihood function with probability tending to unity as the sample size tends to infinity and it is unique.

b) Indicate briefly the method of scoring to obtain a solution of the likelihood equation. [14+6]=[20]

5. Suppose $\{f(X,\theta), \theta \in (\underline{H})\}$ be a class of frequency functions Let

$$\pi(x, \theta) = \frac{f(x, \theta)}{f(x, \theta_0)}$$

where $\theta_0 \in (\underline{H})$ is a fixed point. Let $h(\theta)$ be a parametric function to be estimated and let the loss function be

$$L(T, h(\theta)) = | T(X) - h(\theta)|^s$$

where $s > 1$. Let $\underline{m}_s$ denote the class of unbiased estimates of $h(\theta)$. Show that

p.t.o.

5.(contd.)

i) a necessary condition that $\underline{\underline{m}}_s$ be non-empty is that there exists a constant c such that for every set of n functions $\pi(x,\theta_1),\ldots,\pi(x,\theta_n)$ and every set of n real numbers $a_1,\ldots,a_n$, one has

$$\left|\sum_{i=1}^{n} a_i h(\theta_i)\right| \leq c \left\|\sum_{i=1}^{n} a_i \pi(x, \theta_i)\right\|_r$$

where $\| \pi(x, \theta)\|_r = \left[\int \left|\frac{f(x, \theta)}{f(x,\theta_0)}\right|^r f(x, \theta_0)dx \right]^{1/r}$ is assumed to be finite for all θ ;

ii) for every $\varphi \in \underline{\underline{m}}_s$ one has $\| \varphi \|_s \geq c_0$

where $\| \varphi \|_s = [\int|\varphi|^s f(x, \theta_0)dx]^{1/s}$ and c_0 is the infimum of the set of admissible constants c in (i);

iii) if $\underline{\underline{m}}_s$ is non-empty there is a unique

$\varphi_0 \in \underline{\underline{m}}_s$ with $\| \varphi_0 \|_s = c_0$. [10+5+5]=[20]

6.a) Let $f(x, \theta)$, $\theta \in (\underline{H})$ be a family of density functions. Let $\lambda(\theta)$ be the density function of a prior distribution of θ on $(\underline{H})$. Let $g(\theta)$ be a parametric function to be estimated with squared error loss function. Describe how a Bayes estimate is obtained for $g(\theta)$.

b) Let the family of distributions in (a) be Binomial with parameters n, and θ, $0 \leq \theta \leq 1$.

Let $\lambda(\theta) = \frac{1}{\beta(\lambda, m)} \theta^{\lambda-1}(1 - \theta)^{m-1}$ for some fixed λ and m both > 0. Obtain the Bayes estimate of θ.

[12+8]=[20]

INDIAN STATISTICAL INSTITUTE
Research and Training School
B.Stat. (Hons.) Part IV: 1974-75

[409]

MID-YEAR EXAMINATION

ECONOMICS

Date: 14.2.1975. Maximum marks: 100 Time: 3 hours 10 minutes.

[Note: Choose between the two alternatives given below]

Alternative I: Answer any three questions from Group A and one question from Group B. In this case the maximum mark for any question in Group A is 20 (twenty) and for that in Group B is 40 (forty).

Alternative II: Answer any five questions from Group A and one question from Group B. In this case the maximum mark for any question in Group A is 15 (fifteen) and for that in Group B is 25 (twenty-five).

Please indicate on the top of the answerscript the alternative chosen. 10 (ten) minutes extra time is allowed for making the choice.

GROUP A

1. Do you think that effective demand and productive capacity can remain in balance in an economy where (a) consumption depends upon income, and (b) investment is maintained at a steady level over time ? Defend your answer very briefly.

2. Clearly being out the two-way significance of profits for economic growth under capitalism, given that the essential factor behind growth is technical progress.

3. Pinpoint the essential role of marketable agricultural surplus in the process of industrialisation, and briefly indicate the significance of different means of raising it.

4. What is "Engel's Law" ? For what particular feature of the economic development of to-day's advanced industrial countries is it cited as an explanatory factor ? Indicate the logic of this explanation very briefly.

5. What is "demonstration effect" ? In what way does it help one understand the problem of capital formation in underdeveloped countries in a historical perspective ?

6. Summarize very briefly the behaviour of relative sectoral labour productivities over the course of economic development of to-day's advanced industrial countries, and contrast this with the pattern of relative sectoral labour productivities that emerges from a cross-section study of countries at different levels of development in the world to-day.

7. In what way do you think is the phenomenon of population growth in the present underdeveloped countries different from that in the early phase of development following the industrial revolution in many advanced countries of the world to-day ?

8. Examine the logical validity of any one of the following arguments:

   (a) The simple reason why an underdeveloped country must industrialise in order to raise the level of its percapita income is that the average productivity of labour is much higher in industry than in agriculture in these countries.

   (b) Since any technical progress requires resources for its concrete application and the rate of technical progress is, on the whole, faster in industry then in agriculture, a country always benefits more from technical progress by allocating greater amounts of resouces to industry.

GROUP B

1. Clearly bring out the relation of industrialisation to economic development (a) from the standpoint of the history of economic development in to-day's economically advanced countries, and (b) from the standpoint of planning for economic development in to-day's under-developed countries.

2. What, in your understanding, is the principal purpose behind growth in a socialist economy ? Given this purpose, do you think that the mechanism of growth undersocialism is different in any essential respect from that under capitalism ? What role do the terms of trade between agriculture and industry play in this mechanism in the early phase of socialist development in a backward agricultural country ?

INDIAN STATISTICAL INSTITUTE [411]
Research and Training School
B.Stat. (Hons.) Part IV: 1974-75

MID-YEAR EXAMINATION

ECONOMETRICS (THEORY & PRACTICAL)

Date: 17.2.1975. Maximum Marks: 100 Time: 3 hours.

1. Discuss the problems of multicollinearity and effects of income distribution in the context of estimation of a demand from time series data. To what extent these problems can be solved by pooling cross-section and time series data ? [20]

2. What is a Loreng Curve ? Derive the expressions for the Loreng Curve and for the Loreng ratio for the lognormal distribution. [20]

3. The following table shows the per capita consumption of butter and margarine in sweden during 1921-1939 along with figures of per capita national income, consumer price index and average retail price of butter and margarine.

Year	per capita annual consumption of butter and margarine (Kg.)	average retail price of butter and margarine (per Kg.)	national income per capita	consumer price index
(1)	(2)	(3)	(4)	(5)
1921	12.16	4.62	909	241
1924	14.12	3.03	743	174
1927	15.65	2.52	786	171
1930	18.04	2.13	860	164
1933	18.77	1.94	726	153
1936	20.38	2.06	907	158
1939	20.44	2.53	1114	171

Assuming the constant elasticity form of the demand function, estimate the price elasticity of demand given that income elasticity is 0.4. Calculate also the staward error of the estimated price elasticity. [35]

4. The following table shows the distribution of persons by monthly per capita consumer expenditure based on a survey of rural India.

monthly per capita expenditure class per 30 days in Rs.	estimated % of population	average per capita consumer expenditure per 30 days (in Rs.)
(1)	(2)	(3)
0 - 11	14.12	7.30
11 - 15	15.26	12.40
15 - 21	23.07	17.70
21 - 34	19.55	23.50
34 - 55	18.19	35.80
55 -	9.81	62.30

p.t.o.

Calculate the Lorenz ratio and draw the Lorenz Curve for the above data. Assuming the distribution to be Lognormal, estimate the variance from the Lorenz ratio. Also estimate the mean of the distribution by method of quantiles.

[25]

INDIAN STATISTICAL INSTITUTE [412]
Research and Training School
B.Stat.(Hons.) Part IV : 1974-75

MID-YEAR EXAMINATION

Multivariate Analysis

Date: 19.2.75 Maximum Marks: 100 Time: 3 hours

Note: a) This is an OPEN-BOOK examination. You are allowed to use notes (any hand-written material) and the book C.R.Rao: Linear Statistical Inference and its Applications.

b) Correct answers to about 2/3 of the paper will carry full marks. You may attempt any part of any question. However, more credit will be given to complete answers than to fragmentary ones.

c) Numbers in [ ] indicate marks allotted to the questions.

1.a) When do you say that a p-dimensional random variable has a multivariate normal distribution? Under what conditions does a random variable so distributed has a probability density function? [6]

b) Let (X_i, Y_i), $i = 1,2,3$ be distributed independently and identically as

$$N_2\left[\binom{\mu}{\nu}, \begin{pmatrix}\sigma_{xx} & \sigma_{xy}\\ \sigma_{xy} & \sigma_{yy}\end{pmatrix}\right].$$

i) Find the joint distribution of $(X_1, Y_1, X_2, Y_2, X_3, Y_3)$. [6]

ii) Find the joint distribution of $(\bar{X}, \bar{Y})$ where

$$\bar{X} = \frac{X_1 + X_2 + X_3}{3}, \qquad \bar{Y} = \frac{Y_1 + Y_2 + Y_3}{3}$$ [6]

2. Let

$$\underset{1\times p}{Y'} = (\underset{1\times p_1}{Y^{(1)'}}, \underset{1\times p_2}{Y^{(2)'}}) \sim N_p(\mu, \Sigma)$$

where $p = p_1 + p_2$ and Σ is <u>nonsingular</u>.

a) <u>Derive</u> the conditional distribution of $Y^{(1)}$ given $Y^{(2)} = y^{(2)}$. [9]

b) Show that $E(Y^{(1)} | Y^{(2)} = y^{(2)})$ is the solution of the simultaneous equations

$$\frac{\partial Q}{\partial Y_i^{(1)}} = 0, \qquad i = 1, 2, \ldots, p,$$

where $Y^{(1)'} = (Y_1^{(1)}, Y_2^{(1)}, \ldots, Y_{p_1}^{(1)})$ and

$$Q = (Y - \mu)' \Sigma^{-1} (Y - \mu).$$ [9]

3. Let $\underset{p\times 1}{U} \sim N_p(\mu, \Sigma)$. Let A be a symmetric $p \times p$ <u>non-singular</u> matrix.

p.t.o.

a) State and prove a necessary and sufficient condition for $(U-\mu)' A (U-\mu)$ to be distributed as chi-square. When the condition is satisfied, how is the degree of freedom of the chi-square related to A? [7]

b) Let $\bar{U}$ be the sample mean vector based on a random sample of size n from $N_p(\mu, \Sigma)$ where Σ is non-singular. Find the distribution of

$$n(\bar{U}-\mu)' \Sigma^{-1}(\bar{U}-\mu).$$ [6]

c) Derive a $100\alpha\%$ confidence region for μ when Σ is known, $(0 < \alpha < 1)$. [7]

d) Discuss a method of deriving a $100\alpha\%$ confidence region for μ when Σ is not known for $n > p$, $(0 < \alpha < 1)$. [11]

4.a) Define Hotelling's T^2 statistic and show how it is a multivariate analogue of the Student's t statistic. Find the distribution of the T^2 statistic. [8]

b) Let $\underset{p\times1}{X} \sim N_p(\mu, \Sigma)$ where $\mu = (\mu_1, \mu_2, \ldots, \mu_p)$. Explain a suitable method of testing the hypothesis

$$H_o : \mu_1 = \mu_2 = \ldots = \mu_p$$

against an alternative in which at least one of the equalities is violated, when Σ is not known. [8]

c) Let $\underset{2p\times1}{X} \sim N_{2p}(\mu, \Sigma)$ where $\mu = (\mu_1, \mu_2, \ldots, \mu_p, \mu_{p+1}, \ldots, \mu_{2p})$
Explain a suitable method of testing the hypothesis

$$H_o: \mu_1 = \mu_{p+1}, \mu_2 = \mu_{p+2}, \ldots, \mu_p = \mu_{2p} .$$

against an alternative in which at least one of the equalities is violated, when Σ is not known. [10]

5.a) Explain what is meant by a multivariate linear model. [7]

b) Consider a linear model

$$E(\underset{n\times1}{Y}) = \underset{n\times m}{X}\ \underset{m\times1}{\beta}$$

where the elements of Y are correlated and the dispersion matrix of Y is Σ, a positive definite matrix.
Show that a value of β that minimizes

$$(Y-X\beta)' \Sigma^{-1} (Y-X\beta)$$

is

$$\beta = (X' \Sigma^{-1} X)^{-} (X' \Sigma^{-1} Y).$$

[Hint: Use the positive definiteness of Σ (and of Σ^{-1}) to reformulate the problem in the usual frame-work (where the dispersion matrix is of the form $\sigma^2 I$) and use the results thereof.] [12]

6. Consider the problem of discrimination between two groups using p variables $X = (X_1, X_2, \ldots, X_p)$ distributed as $N_p(\mu^{(i)}, \Sigma)$, i = 1,2 in the two groups.

p.t.o.

INDIAN STATISTICAL INSTITUTE [413]
Research and Training School
B.Stat. (Hons.) Part IV: 1974-75

MID-YEAR EXAMINATION

Statistical Methods

Date: 20.2.75 Maximum Marks: 100 Time: 3 hours

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

1. Let $F(x)$ be a distribution function admitting the p.d.f. $f(x)$ and let p, $0 < p < 1$, be a number. Define the population pth-fractile ξ_p and the sample fractile $\hat{\xi}_p$. Show that under suitable conditions (to be stated by you),

   i) $\hat{\xi}_p \to \xi_p$ wp 1 .

   ii) $\sqrt{n}\,(\hat{\xi}_p - \xi_p) \xrightarrow{W} N(0, \dfrac{p(1-p)}{[f(\xi_p)]^2})$. [20]

2. Let T_n be a k-dimensional statistic $(T_{1n}, T_{2n}, \ldots, T_{kn})$ such that

$$\sqrt{n}\,(T_{1n} - \theta_1), \sqrt{n}(T_{2n} - \theta_2), \ldots, \sqrt{n}\,(T_{kn} - \theta_k) \xrightarrow{W} N(0, \Sigma).$$

Further let g be a function of k variables which is totally differentiable at $\underset{\sim}{\theta}$. Prove that

$$\sqrt{n}[g(T_{1n}, T_{2n}, \ldots, T_{kn}) - g(\theta_1, \theta_2, \ldots; \theta_k)] \xrightarrow{W} N(0, v(\theta))$$

where $v(\theta) = \sum_i \sum_j \sigma_{ij} \frac{\partial g}{\partial \theta_i} \frac{\partial g}{\partial \theta_j}$ provided $v(\theta) \neq 0$. [20]

3.a) Let $x_1, \ldots, x_n$ be independent observations on a random variable X and let

$$\bar{x} = \frac{1}{n} \Sigma x_i , \qquad m_r = \frac{1}{n} \Sigma (x_i - \bar{x})^r .$$

Derive the joint asymptotic distribution of (m_r, m_s), $r \neq s$. State the assumptions you make about the distribution of X.

b) Let r be the sample correlation coefficient based on n independent observations from a bivariate normal population with correlation coefficient ρ.

Show that $\sqrt{n}\,(r - \rho) \xrightarrow{W} N(0, \frac{(1-\rho^2)^2}{n})$. [20]

4. The correlations obtained from 6 samples of sizes 10, 14, 16, 25, 28, 20 were found to be 0.318, 0.106, 0.253, 0.116, 0.112 and 0.340 respectively. Can these be considered homogeneous? If so, obtain a combined estimate of ρ. [20]

p. t. o.

5. The following table gives Bateson's distribution of sweet pea plants obtained from an intercross. The marginal frequencies are expected to be in the ratio 3:1, and if the two characters, flower colour and pollen shape, are independently inherited, then the cell frequencies are expected to be in the ratio 9:3:3:1 .

Pollen Shape	Flower Colour Purple	Red	Total
Long	296	27	323
Round	19	85	104
Total	315	112	427

Carry out a suitable statistical analysis of the data and write a short note on your findings. [20]

INDIAN STATISTICAL INSTITUTE
Research and Training School
B.Stat. (Hons.) Part IV: 1974-75

[414]

MID-YEAR EXAMINATION

Measures and Probability

Date: 22.2.75 Maximum Marks: 100 Time: 3 hours

Note: The whole paper carries 105 marks. Answer as many questions as you can. The Maximum you can score is 100. Marks allotted for each question are given in brackets [].

1. Let μ be a non-negative finitely additive set function defined on a ring $\underline{R}$ of sets and write $E \sim F$ if $\mu(E \triangle F) = 0$. Prove the following three statements and show that the three statements are false if non-negativity of μ is not assumed.

   a) $\sim$ is an equivalence relation. [6]

   b) The class $\underline{C}$ of sets $C \sim \emptyset$ is a ring. [8]

   c) $\mu(E \triangle F) \leq \mu(E \triangle G) + \mu(F \triangle G)$. [6]

2.a) If $F : R \to R$ is an increasing right continuous function, show that there exists a unique measure μ on the Borel σ-field such that $\mu(a, b] = F(b) - F(a)$. [12]

b) Is the correspondence $F \to \mu$ given in (a) 1 - 1? Is it onto all measures on the Borel σ-field ? [5]

3. State and prove Lebesgue's dominated convergence theorem. (You may assume Fatou's lemma). [10]

4. Show that a bounded function on a bounded interval is Riemann integrable iff the set of its discontinuity points is a Lebesgue measurable set of Lebesgue measure 0. Also then show that the Riemann and Lebesgue integrals coincide. [10]

5.a) Define the product of two measurable spaces $(X, \underline{F})$ and $(Y, \underline{G})$. [3]

b) Show that if $A \subseteq X$ and $B \subseteq Y$ then $A \times B$ belongs to the product σ-field iff $A \in \underline{F}$ and $B \in \underline{G}$. [5]

c) If μ and ν are finite measures on $\underline{F}$ and $\underline{G}$ respectively, prove the existence of the 'product measure'. [10]

d) State and prove Fubini's theorem for non-negative measurable functions on the product space. [1..

6. If $(X, \underline{F}, \mu)$ is a measure space, $(Y, \underline{G})$ is a measurable space and $f : X \to Y$ is measurable function, define ν on $\underline{G}$ by $\nu(G) = \mu(f^{-1}(G))$.

   a) Show that ν is a measure. [5]

   b) If $g : Y \to R$ is quasi-integrable (w.r.t. ν) then show that

$$\int g \, d\nu = \int g \cdot \, d\mu$$

where $(g \circ f)(x) = g(f(x))$. [10]

INDIAN STATISTICAL INSTITUTE
Research and Training School
B.Stat.(Hons.) Part IV: 1974-75
M. Stat. Part I: 1974-75

[4326/17]
[6426/15]

PERIODICAL EXAMINATION

Advanced Design and Analysis of Experiments

Date: 7.4.75 Maximum Marks: 100 Time: 2 hours

Note: a) This is an OPEN-BOOK EXAMINATION.
b) Answer all the questions. All questions carry equal marks.

1. An experiment was conducted to study the effects of five anaesthetic drugs A, B, C, D, E, the first three being sedatives and the last two being barbiturates. One hundred subjects were divided at random into five equal groups and each group was administered one drug.

Write down linear contrasts that would be suitable to estimate the following comparisons: [Make your notations clear.]

Sedatives vs. Barbiturates
Drug A vs. Drug C
Drug D vs. Drug E
Drug B vs. other sedatives.

Do the contrasts you suggest form an orthogonal set? [Justify your answer.]

Is the linear contrast 'Mean of Drug A - Mean of Drug B' orthogonal to these five?

With suitable notations write down the sum of squares in an analysis of variance, in which the treatment sum-of-squares is subdivided into suitable one-degree-of-freedom components.

2. Given the following ten units of varying 'size' (symbolising their innate error response to the treatments) to test five treatments, each to be replicated twice, how would you divide the units into two groups (blocks) of five units each (so that the sum of squares between the groups is as large as possible)?

Unit	1	2	3	4	5	6	7	8	9	10
Size index	2.3	3.2	4.2	2.6	3.6	4.5	2.8	3.9	4.8	2.1

With suitable assumptions, obtain some tentative conclusions regarding blocking as a means of control of experimental error.

Also suggest suitable models for analysis of data obtained from an experiment with these units

i) without blocking
ii) with blocks as suggested by you.

Explain the concepts of main effects and interactions in an experiment with two factors.
[Try and limit to your answer to two pages of relevant material presented in your own words.]
In an experiment with two factorseach at two levels with the condition of 'proportional frequencies', derive the analysis (estimators and expressions for sums of squares for error and for main effects and interactions) with the interaction terms present.

p.t.o.

4. It is desired to compare two rolls of cloth by means of their performance when made up into men's trousers and used in a 'weaver trial'. The two rolls are identical in appearance but differ in the proportion of synthetic fibres incorporated; each is sufficient to make 12 pairs of trousers (i.e. 12 pants!) Devise various plans for carrying out this experiment. [Two or three plans of really different types will be adequate.]

Discuss their relative merits and state which one you consider to be preperable and why.

INDIAN STATISTICAL INSTITUTE
Research and Training School
B.Stat.(Hons.) Part IV: 1974-75
PERIODICAL EXAMINATION

[447/14]

Inference

Date: 14.4.75 Maximum Marks: 50 Time: 2 hours

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

1.a) Let $\{f(X, \theta), \theta \in \textcircled{H}\}$ be a family of density functions of a random variable X. Show that (i) there exists a most powerful test of size $\alpha (< 1)$ for testing the simple hypothesis $H_o: \theta = \theta_o \in (\underline{H})$ against the simple alternative H_1: $\theta = \theta_b \in (\underline{H})$ and obtain a necessary and sufficient condition for a test to be most power for testing H_o against H_1.

b) Use the above result to obtain the most powerful test for testing $H_o: \mu = 0$ against $H_1: \mu = 1$ in a family of Normal distributions $\{N(\mu, 1) - \infty < \mu < \infty\}$ based on a sample of size 9.

2.a) Define a Monotone likelihood ratio family of distributions.

b) Show that the family of distributions $\{N(0, \sigma^2), 0 < \sigma^2 < \infty\}$ is an MLR family.

c) Obtain the most powerful test for $H_o: \sigma^2 \leq 4$ against $H_1: \sigma^2 > 4$ based on a sample of size 5.

d) Obtain the power of the test for $\sigma^2 = 2$ and $\sigma^2 = 8$.

3.a) Let X follow a poisson distribution with unknown mean λ. Obtain a UMP test of size α for $H_o: \lambda \leq 1$ or $\lambda \geq 2$ against the alternative $H_1: 1 < \lambda < 2$ based on a sample of size 3.

b) State clearly the principal result you used in obtaining the UMPT in (a) above.

4. Let X have a Koopman's exponential distribution with density function belonging

$$\{c(\theta)h(x)\, e^{\theta T(x)}, \theta \in \mathbb{R}\}$$

with unknown θ. Show that there exists a least favourable distribution λ on w for $H_o: \{\theta \in w : \theta \leq \theta_1$ or $\theta \geq \theta_2\}$ against $H_1: \theta = \theta_3$ $(\theta_1 < \theta_3 < \theta_2)$. Also show that the MPT level α test for testing H_λ: the density function of x is $g_\lambda(x) = \int_w h(x)c(\theta) \cdot e^{\theta T(x)}\, d\lambda(\theta)$, against H_1: the density of x is $c(\theta_3)h(x)\, e^{\theta_3 T(x)}$ is also most powerful level α test for H_o against H_1.

INDIAN STATISTICAL INSTITUTE
Research and Training School
B.Stat.(Hons.) Part IV: 1974-75 [412/17]

PERIODICAL EXAMINATION

Application of Probability

Date: 21.4.75 Maximum Marks: 100 Time: 3 hours

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

1. Give two examples each, from different areas of application, of the following types of stochastic processes: (a) discrete state space and discrete parameter; (b) discrete state space and continuous parameter; (c) continuous state space and discrete parameter; and (d) continuous state space and continuous parameter. [2x3x4]=[24]

2.a) Derive the Chapman-Kolmogorov equations for a countable state space and continuous parameter Markov process. [6]

b) Stating the requisite assumptions regarding 'instantaneous' probability functions, derive the set of Kolmogorov 'forward' differential equations for a Markov process of the above type.

Write these equations in matrix form, after defining the requisite matrices. [15]

3.a) If $A = [a_{ij}]$ and $B = [b_{ij}]$ are $n \times n$ matrices and the 'norm' of a matrix A is defined as

$$\|A\| = \sum_{i,j} |a_{ij}|,$$

show that

i) $\|A+B\| \leq \|A\| + \|B\|$

ii) $\|AB\| \leq \|A\| \cdot \|B\|$

iii) $\|Ax\| \leq \|A\| \cdot \|x\|$

where the norm $\|x\|$ of an n-vector x is similarly defined. [9]

b) Let $X(t) = [x_{ij}(t)]$ be an $n \times n$ matrix of differentiable functions $x_{ij}(t)$, and denote by $\frac{dX(t)}{dt}$ the matrix [$[\frac{dx_{ij}(t)}{dt}]$. Show that

i) $\frac{d}{dt} X(t)Y(t) = \frac{dX(t)}{dt} Y(t) + X(t) \frac{dY(t)}{dt}$.

ii) $\frac{d}{dt} [X(t)]^{-1} = - [X(t)]^{-1} \frac{dX(t)}{dt} [X(t)]^{-1}$. [8+8]=[16]

4.a) If the functions $x_{ij}(t)$ are differentiable in the interval $0 \leq t \leq T$, and $A = [a_{ij}]$ and $C = [c_{ij}]$ are $n \times n$ matrics of constant elements, show that there exists a solution of the differential equations

$$\frac{dX(t)}{dt} = AX(t), \quad X(o) = C \quad \text{in this interval.}$$ [15]

b) Assuming the requisite properties of the matrix

$$e^{At} = \sum_{j=o}^{\infty} \frac{A^j t^j}{j!},$$

show that e^{At} is a solution of the set of differential equa equations $\frac{dY(t)}{dt} = AY(t)$, $Y(o) = I$. [8]

c) Using the above result, show that Y(t) has the property

$$Y(t+s) = Y(t)Y(s).$$ [7]

INDIAN STATISTICAL INSTITUTE
Research and Training School
B.Stat.(Hons.) Part IV: 1974-75

[422/18]

PERIODICAL EXAMINATION

Non-Parametric Methods and Sequencial Analysis

Date: 28.4.75 Maximum Marks: 100 Time: $2\frac{1}{2}$ hours

Note: Any <u>three</u> questions carry full marks.

1.a) Describe the SPRT (b,a) of a simple hypothesis versus a simple hypothesis, based on i.i.d. observations.

b) Given α and β, obtain Wald's approximations for the stopping bounds. Show that the resulting error probabilities α^* and β^* satisfy

i) $\alpha^* + \beta^* \leq \alpha + \beta$ (ii) $\alpha^* \leq \frac{\alpha}{1-\beta}$, $\beta^* \leq \frac{\beta}{1-\alpha}$.

2.a) Show that SPRT (b, a), $-\infty < b < 0 < a < \infty$, based on i.i.d. observations terminates wp 1 if

$$z = \log\left(\frac{f_1(x)}{f_0(x)}\right)$$

is not degenerate at the origin.

b) Show that, for all $k \geq 0$, $E(\pi^k) < \infty$, where π is the stopping variable of the SPRT.

3. X is a discrete random variable such that, under

$$H_0: P(X = 0) = \frac{4}{5} \text{ and } P(X=1) = \frac{1}{5}$$

and under H_1: $P(X=0) = \frac{2}{5}$, $P(X=1) = \frac{2}{5}$ and $P(X=2) = \frac{1}{5}$.
Consider the SPRT (-J log 2, K log 2), J and K being positive integers, of H_0 vs. H_1.

i) Find the exact values of α and β.

ii) Find the exact values of $E(\pi/H_0)$ and $E(\pi/H_1)$.

4. Consider testing $H_0: N(0, 1)$ vs. $H_1 = N(1,1)$.

a) Find the stopping bounds b and a of the SPRT such that $\alpha = \beta = .01$. (Use Wald's approximations).

b) Compute $E(\pi/H_0)$ and $E(\pi/H_1)$. (Use Wald's approximations).

c) Find the smallest n for which the fixed sample size MPT of H_0 vs. H_1 has $\alpha = \beta = .01$.

d) Compute the efficiency of the SPRT w.r.t. the fixed sample size MPT

(i.e. $\frac{n}{E(\pi)}$) under H_0 and H_1.

INDIAN STATISTICAL INSTITUTE
Research and Training School [462/6]
M. Stat. Part I and B. Stat. Part IV: 1974-75

PERIODICAL EXAMINATION

Operations Research

Date: 5.5.75 Maximum Marks: 100 Time: 3 hours

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

1. Define the following terminologies.

a) Slack and surplus variables

b) Basic feasible solution

c) Degenerate basic feasible solution [9]

2. Explain the use of artificial variables in solving linear programming problems. [6]

3. Consider the linear programming problem: maximize $z = CX$ subject to $AX = b$ and $X \geq 0$, where c, b and X are respectively $1 \times n$, $m \times 1$ and $n \times 1$ vectors and A is $m \times n$ matrix with rank of A equal to m. Let B be a basis of matrix A such that $X_B = B^{-1}b \geq 0$. Suppose there exists a non-basic vector a_j such that $y_j - c_j < 0$ and $y_{ij} \leq 0 \;\forall\; i = 1, 2, \ldots, m$. Prove that this linear programming problem is unbounded (notations are as used in the class) [10]

4. Consider the linear program given in question 3, and when it is solved by Charne's M-method, prove the following. In the final simplex tableau an artificial variable appears in the basis,

a) at positive level => the original ℓ.p. problem is inconsistent

b) at zero level in the ith row such that the entries in the ith row are all zero => the ith constraint of the original ℓ.p. problem is redundent. [15]

5. By Charne's M. method solve the following ℓ. p.

$$\text{maximize} \quad y = 8x_1 + x_2 + 3x_3 + 5x_4$$

$$\text{subject to} \quad 2x_1 + 3x_2 - 5x_3 + 4x_4 = 7$$

$$-3x_1 + 2x_2 + 3x_3 - 9x_4 = 8$$

$$1.75x_1 - x_2 + 3.25x_3 - 4.25x_4 = -4.75$$

$$x_i \geq 0, \quad i = 1, 2, 3, 4.$$

(After writing the initial simplex tableau make only one iteration i.e. compute the next simplex tableau). [20]

6. Solve the following ℓ.p. by graphical method

$$\text{maximize} \quad z = x_1 + 5x_2$$

$$\text{subject to} \quad 3x_1 + 4x_2 \leq 6$$

$$x_1 + 3x_2 \geq 2$$

$$x_1, \; x_2 \geq 0$$ [10]

7. Home assignment on formulation of linear programming problems [to be submitted before 9th May 1975]. [30]

INDIAN STATISTICAL INSTITUTE
Research and Training School

M.Stat. I Year /B.Stat.(Hons.) IV Year

Periodical Examination : Sample Surveys

Answer as many questions as you can. The paper carries 125 marks. Maximum you can score is 100

12.5.75 Time: 3½ hrs.

1. Describe, in non-mathematical terms, the purpose, necessity, basic framework and methodology of inference and problems arising therefrom in sample surveys. [10]

2. From a finite population a simple random sample with replacement is drawn until one obtains just three distinct units. The repetitions among the units selected are then knocked off and the truncated sample is treated as the final sample.

   (a) Prove that this procedure is equivallent to drawing a simple random sample, without replacement, of size three. [6]

   (b) Prove that the mean over distinct units is an unbiassed estimator of the population mean. [8]

3. (a) Define a sampling design and a sampling proceedure and state (without proof) the main result connecting the two. Discuss the implications and use of this result through an illustrative example. [8]

   (b) Characterise the sampling designs generated by

   (i) SRS without replacement of size n ,

   (ii) PPS with replacement of size n_1 in the first stratum and SRS with replacement, of size n_2 in the second stratum, for stratified sampling with two strata. [3+5+8]

4. (a) In a general sampling design. define a general homogeneous linear estimator of the population total, and derive the conditions for unbiassedness of this estimator. [8]

   (b) Write down the 'conventional' unbiassed estimators of Y in the case of PPS sampling with replacement and identify it as a special case of general homogeneous linear unbiassed estimator. [6]

5. What are the situations and considerations that motivate one to adopt (a) PPS sampling and (b) stratified sampling. How do the conventional estimates of population total meet requirements in these cases? [8]

6. Describe the basic theoretical principles in stratified sampling strategies. [6]

7. For the sampling design obtained by SRS <u>with</u> replacement of size n.

   (a) evaluate the inclusion probabilities π_i and π_{ij} [8]

   (b) write down $\hat{Y}_{HT}$, the Horvitz-Thompson estimator of Y, and prove its unbiassedness. [8]

   (c) Write down (no derivation needed) $V(\hat{Y}_{HT})$ and $\hat{V}_{HT}(Y_{HT})$. [6]

8. For a general sampling design setup, write down (no proofs are needed) clearly and rigorously the results concerning

   (a) existence of linear unbiassed estimator $\hat{Y}$ of Y.

   (b) existence of best linear unbiassed estimator of Y.

   (c) existence of unbiassed estimator of $V(\hat{Y})$, the variance of Y.

   What are the consequences of (a), (b) and (c) of above for the special case of linear systematic sampling? [3+3+4+5=15]

9. From a finite population consisting of two strata, a PPS with replacement sample of size n_1 is taken from stratum I and an SRS with replacement of size n_2 is taken from stratum II. Conventional unbiassed estimator of population total Y is used.

   (a) For a fixed value of $n = n_1+n_2$, what is the optimum allocation of n to the two strata? [8]

   (b) Estimate the gain due to the strategy adopted over the hypothetical strategy of drawing an unstratified SRS <u>without</u> replacement of size n from the population and adopting the conventional estimator of Y. [12]

INDIAN STATISTICAL INSTITUTE
Research and Training School
B.Stat.(Hons.) Part IV: 1974-75 [482b]

PERIODICAL EXAMINATION

Demography

Date: 19.5.75 Maximum Marks: 100 Time: 3 hours

Note: Answer Q.No.5 and three of the rest. Marks allotted for each question are given in brackets [ ].

1.a) Briefly enumerate the uses of vital statistics.

b) Define de facto and de jure population enumeration. [20+5]=[25]

2. Indicate the nature of deficiencies in the Indian vital registration data. What are the factors underlying them? [25]

3.a) Why is Life Table death rate is assumed to describe a standardised measure of current mortality? Enumerate some of the practical applications of a life table in population reproduction and estimates.

b) Considering a life table, calculate the following probabilities:

i) a life aged 30 years will survive 20 years more,

ii) on the average, how many years of life will a group of children who has just attained age 5 years live next 15 years,

iii) average age at death of females dying between 30 and 50. [10+15]=[25]

4. Write short notes on:

a) Fertility measures from census data.

b) Total Fertility rate as a hypothetical index of fertility.

c) Adjusted I.M.R.

d) Measure of mortality to delineate the influence of biological factors from environmental factors.

e) $M_x \simeq q_{x+\frac{1}{2}}$. [5x5=[25]

5. Summary of data collected from a sample survey of Jute Mill Workers of Howrah conducted in 1961 and those collected from All India registration and Census statistics are shown below:

Age group	Survey of Jute Mill Workers (1961): No. of Male workers /1000 in Pay Rolls	Survey of Jute Mill Workers (1961): D.R./1000 in the age group	All India (1961): No.of Male/ 1000 persons	All India (1961): D.R./1000 in the age group
20 - 29	240	8	93	8
30 - 39	330	14	72	12
40 - 49	300	22	49	16
50 - 59	130	40	30	32

Crude D.R. (A.I) 1961 - 22/1000 Population

From the above data compute a single index for the general mortality condition expressed by Jute Mill Workers during their employment. Compare it with the A.I. average. [25]

INDIAN STATISTICAL INSTITUTE
Research and Training School
M.Stat. Part I: 1974-75
and
B.Stat.(Hons) Part IV: 1974-75
ANNUAL EXAMINATION
O R (Theory and Practical)

Date: 11 7 75 Maximum Marks: 100 Time: 3 hours

Note: The paper carries 110 marks. Answer as many questions as you like. Maximum you can score is 100. Marks allotted for each question are given in brackets [ ]

1 Put the following linear program into the standard form

Maximize $x_1 - 2x_2 + 3x_3$

Subject to $x_1 + x_2 + x_3 \leq 7$

$x_1 - x_2 + x_3 \leq -2$

$3x_1 + 2x_3 = 5$

$x_2 - x_3 \geq 1$

$x_1 \geq 0$, $x_2 \leq 0$, x_3 is unrestricted in sign. [7]

2 Let $X_B = B^{-1}b$ be a basic feasible solution to $AX = b$ and a_j a vector from A not in B Then if $Y_j = B^{-1}a_j$ show that $B(X_B - \theta Y_j) + \theta a_j = b$. is a feasible solution to $AX = b$ for any θ, $0 \leq \theta \leq \theta_{max}$ where

$$\theta_{max} = \min_i \left\{ \frac{x_{B_i}}{y_{ij}} / y_{ij} > 0 \right\}$$

For which values of θ, is the new solution basic ? [10]

3 Given a basic feasible solution $X_B = B^{-1}b$ with $Z_o = C_B X_B$ to the linear program $AX = b$, $X \geq 0$, max $Z = CX$. such that $z_j - c_j \geq 0$ for every column vector a_j in A Then Z_o is the maximum value of Z subject to the constraints and the basic feasible solution is an optimal basic feasible solution. Prove the above statement [10]

4 Apply charnes Big M-method to the following linear program

Maximize $x_1 + x_2$

Subject to $-x_1 + x_2 \leq -1$

$x_1 - x_2 \leq -1$

$x_1 \geq 0$, $x_2 \geq 0$ [13]

5 Describe the method of getting alternate optimum Basic feasible solutions of a linear program from the final simplex tableau of the program. [10]

p t o

6 Describe the two basic systems of Inventory Control and compare them [7]

7 A contractor has to supply 10,000 bearings per day to an automobile manufacturer He finds that, when he starts a production run, he can produce 25,000 bearings per day The cost of holding a bearing in stock for one year is 2 cents, and the setup cost of a production run is $ 18.00 How frequently should production runs be made? What is the economic batch size production ? Before solving, derive the relevant mathematical model [20]

8 Derive the relevant mathematical relationship to test the sensitivity of the Wilson's square root formula [10]

9 Describe the method of developing an ordering table for a store [13]

10 Machine C costs $ 10,000 Operating costs are $ 500 per year for the first five years In the sixth and succeeding years operating costs increase $ 100 each year Assuming a 10 % cost of money per year find the optimum length of time to hold the machine before replacing it. [20]

INDIAN STATISTICAL INSTITUTE
Research and Training School
B.Stat.(Hons.) Part IV: 1974-75
and
M.Stat. Part I: 1974-75

ANNUAL EXAMINATION

Biostatistics

Date: 22.7.75 Maximum Marks: 100 Time: 3 hours

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

1.a) State clearly Mendel's laws of segregation and independent assortment, and describe how these laws follow from the mechanisms of meiosis and gametic union.

b) The locus of a gene pair (A, a) is on an autosomal pair of chromosomes and that of (B, b) on the sex-chromosomes, A and B being dominant over a and b respectively. Enumerate the possible genotypes and phenotypes of males and females.

c) (A, a) is a gene pair located on the sex-chromosomes, a being recessive and the genotype aa being absent in the population. Draw up a table showing possible mating types of parents and the corresponding genotypes of male and female offspring.

d) Establish the 9:3:3:1 law for the cross between two double heterozygotes Aa Bb, the gene pairs being located on different autosomal pairs of chromosomes. [8+6+6+5]=[25]

2.a) If random mating is practised in a population in which the male and female frequencies in the parental generation for AA, Aa and aa are u, v and w $(u+v+w = 1)$, show that genotypic equilibrium is reached from the first filial generation onwards.

b) Describe the nature of co-dominance of the ABO blood-group system, derive the probabilities of the phenotypes A, B, AB and O for the gene-frequencies p, q, r, and describe two methods of estimation of p, q, r for a random sample of individuals whose phenotypes are observed.

c) In a gene pair (A, a) the gene A is dominant and the gene-frequency of a is q. Prove that, under random mating, in the offspring population

i) the frequency of the gene a among offspring of Dominant x Dominant matings is $[q/(1+q)]^2$, and

ii) the frequency of the gene a, among offsprings of Dominant x Recessive matings is $q/(1+q)$. [5+8+7+7]=[27]

3.a) A phenotype carrying a mutant gene has probability $1-s$ to survive till reproduction. Find the probability that such a phenotype survives for exactly n generations and show that $E(n) = 1/s$.

p.t.o

3·b) For a gene pair (A, a), where a is the mutant gene, assume that selection takes place at the zygotic stage and mutation takes place subsequently till maturity, and derive a general expression for the frequency of the mutant gene in the offspring generation as a function of the gene frequency in the parental generation, the fitness coefficients and the mutation rate.

c) In (b), if selection operates on the genotype aa only, show that the equilibrium frequency of the mutant gene is $\sqrt{\mu / s}$, where μ is the mutation rate and s the selection coefficient.

d). The age-distribution(at the time of birth of a child) of fathers in the fertile ages is given by the p.d.f. p(x), with mean value $\bar{x}$ and variance σ^2. The probability that a father of age x produces a mutant child is given by

$$\mu_x = \alpha x + \mu_0 .$$

The age-distribution of the fathers of mutant children therefore has a p.d.f.

$$q(x) = \mu_x p(x) / \int_x \mu_x p(x) dx .$$

If $\bar{x}_m$ is the mean value of q(x), show that

$$\bar{x}_m = [\alpha(\sigma^2 + \bar{x}^2) + \mu_0 \bar{x}] / [\alpha \bar{x} + \mu_0] .$$

[6+9+5+8]=[28]

4·a) Describe (i) selfing and (ii) assortative phenotypic mating and show that in each case the proportion of heterozygotes decreases progressively with generations.

b) Two unequal but large populations in Hary-Weinberg equilibrium have different gene frequencies of the gene pair (A, a). Show that if the two populations are mixed and continue random mating, the proportion of heterozygotes in the offspring generation will be smaller than that in the mixed parental generation. [7+8+5]=[20]

INDIAN STATISTICAL INSTITUTE
Research and Training School
B. Stat. (Hons.) Part IV: 1974-75

ANNUAL EXAMINATION

Statistical Inference

Date: 25.7.75 | Maximum Marks: 100 | Time: 3 hours

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

1. a) Let X be a random variable with probability density $f(x, \theta)$ θ being a real unknown parameter, with monotone likelihood ratio in $T(x)$. For testing the null hypothesis $H_0 : \theta \leq \theta_0$ against the alternative $H_1 : \theta \geq \theta_1$ where $\theta_0 < \theta_1$, at level of significance α, show that there exists a uniformly most powerful test and obtain its form. Also show that the power function of this test is an increasing function of θ.

b) Let X be a random variable taking values 1 and 0 with probabilities θ and $1 - \theta$ respectively where $0 < \theta < 1$ is unknown. Based on a sample of size 10 obtain a UMP test of size 0.05 for the null hypothesis $\theta \leq 1/3$ atainst the alternative $\theta > 1/3$.

2. a) Let $f_0(x), f_1(x), \cdots, f_m(x)$ be real valued integrable functions defined on an Euclidean space $\mathcal{X}$. Suppose $\mathcal{C}$ is the class of all functions $\varphi(x)$ such that $0 \leq \varphi(x) \leq 1$ for all x and $\int \varphi(x) f_i(x) dx = c_i$ for $i = 1, \cdots, m$ where $c_1, \cdots, c_m$ are given constants. Show that a necessary and sufficient condition for a member of $\mathcal{C}$ to maximize $\int \varphi(x) f_0(x) dx$ is the existence constants $k_1, \cdots, k_m$ such that

$$\varphi(x) = 1 \quad \text{if} \quad f_0(x) > \sum_1^m k_i f_i(x)$$

and

$$\varphi(x) = 0 \quad \text{if} \quad f_0(x) < \sum_1^m k_i f_i(x)$$

b) Let X be a random variable whose density function is of the form $c(\theta)\, h(x) \exp \theta T(x)$ where θ is real. By using the result in (a) obtain the form of an uniformly most powerful test of size α for the hypothesis $H_0 : \theta \leq \theta_1$ or $\theta \geq \theta_2$ $(\theta_1 < \theta_2)$ against the alternative $H_1 : \theta_1 < \theta < \theta_2$.

3. a) Explain clearly when a test of a hypothesis is said to be (i) similar, (ii) of Neyman structure. Obtain a necessary and sufficient condition for all similar tests to have Neyman structure.

b) Let X be a random variable with density function $f(x, \theta, \underset{\sim}{\eta})$ of the form

$$c(\theta, \underset{\sim}{\eta})\, h(x) \cdot \exp \quad \theta U(x) + \sum_{i=1}^{k} \eta_i T_i(x)$$

where (θ, η) belongs to the $(k+1)$ dimensional Euclidean space. Show that there exists a uniformly most powerful test for the hypothesis $H_0 : \theta \leq \theta_0$ against the alternative $H_1 : \theta \geq \theta_1$ $(\theta_0 < \theta_1)$ of level α given the values $T_i(x) = t_i$ for $i = 1, \cdots, k$.

4.a) Let the family of possible distributions of a random variable X be $\mathcal{P} = \{P(X, \theta), \theta \in (\underline{H})\}$ and let $T(x)$ be a sufficient statistic for $\mathcal{P}$. Obtain a necessary and sufficient condition that a statistic $V(x)$ is statistically independent of $T(x)$ for all θ.

b) Use the result in (a) to show that, if a random variable X has density function of the form $c(\theta, \underset{\sim}{\eta}') h(x)$

$\exp\left\{\theta U(x) + \sum_{1}^{k} \eta_i T_i(x)\right\}$ where $(\theta, \underset{\sim}{\eta}')$ belongs to the (k+1)

dimensional Euclidean space, there exists a uniformly most powerful unbiased test of size α for the null hypothesis $H_0: \theta_1 \leq \theta \leq \theta_2$ against $H_1: \theta < \theta_1$ or $\theta > \theta_2$ provided there exists a function $h(U, T_1 \cdots T_k)$ increasing in U for fixed $T_1, \cdots, T_k$.

5.a) Let a random variable X have a Normal distribution $N(\mu, \sigma^2)$ $-\infty < \mu < \infty$ and $0 < \sigma^2 < \infty$. Based on a sample of size 10 derive a uniformly most powerful unbiased test of size 0.05 for the hypothesis $\sigma = 2$ against the alternative $\sigma \neq 2$. (State the relevant result which shows that the test you propose is uniformly most powerful unbiased).

b) Explain how you would compute the power function of the above test.

c) Obtain a 95% confidence interval for σ^2 in the above example.

6.a) Let $P(x, \theta, \underset{\sim}{\eta})$ be a family of distributions of a random variable X, where (θ, η') belong to (k+1) dimensional Euclidean space. Define the concept of an unbiased confidence set for the parameter θ.

b) Show that uniformly most accurate unbiased confidence sets of a given confidence coefficient $1-\alpha$ exist for θ if there exist UMP unbiased tests of size α for testing suitable hypothesis about θ.

c) Let a random variable X have Normal distribution $N(\mu_1, \sigma^2)$ and another random variable Y be independent of X and have Normal distribution $N(\mu_2, \sigma^2)$. Based on a sample of size 9 from the X population and a sample of size 16 from the Y population, obtain a 95% confidence interval for $\mu_1 - \mu_2$.

7.a) Let a random variable X have distribution $P(x, \underset{\sim}{\theta})$ where $\underset{\sim}{\theta} = (\theta_1, \cdots, \theta_k)$ belongs to the k dimensional Euclidean space. Under suitable regularity conditions (to be clearly stated by you) derive Rao's test statistic for testing the hypothesis $\theta_1 = \theta_1^0, \cdots, \theta_k = \theta_k^0$ based on a sufficiently large sample.

b) What is the likelihood ratio test for the problem in (a) and what is its asymptotic distribution?

--- -------

INDIAN STATISTICAL INSTITUTE
Research and Training School
B Stat.(Hons.) Part IV: 1974-75

ANNUAL EXAMINATION

Statistical Inference

Date: 25.7.75 | Maximum Marks: 100 | Time: 3 hours

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

1 a) Let X be a random variable with probability density $f(x, \theta)$ θ being a real unknown parameter, with monotone likelihood ratio in $T(x)$. For testing the null hypothesis $H_0 : \theta \leq \theta_0$ against the alternative $H_1 : \theta \geq \theta_1$ where $\theta_0 < \theta_1$, at level of significance α, show that there exists a uniformly most powerful test and obtain its form. Also show that the power function of this test is an increasing function of θ.

b) Let X be a random variable taking values 1 and 0 with probabilities θ and $1 - \theta$ respectively where $0 < \theta < 1$ is unknown. Based on a sample of size 10 obtain a UMP test of size 0.05 for the null hypothesis $\theta \leq 1/3$ atainst the alternative $\theta > 1/3$.

2.a) Let $f_0(x), f_1(x), \cdots, f_m(x)$ be real valued integrable functions defined on an Euclidean space $\mathcal{X}$. Suppose $\mathcal{C}$ is the class of all functions $\varphi(x)$ such that $0 \leq \varphi(x) \leq 1$ for all x and $\int \varphi(x) f_i(x) dx = c_i$ for $i = 1, \cdots, m$ where $c_1, \cdots, c_m$ are given constants. Show that a necessary and sufficient condition for a member of $\mathcal{C}$ to maximize $\int \varphi(x) f_0(x) dx$ is the existence constants $k_1, \cdots, k_m$ such that

$$\varphi(x) = 1 \quad \text{if} \quad f_0(x) > \sum_1^m k_i f_i(x)$$

and

$$\varphi(x) = 0 \quad \text{if} \quad f_0(x) < \sum_1^m k_i f_i(x)$$

b) Let X be a random variable whose density function is of the form $c(\theta)\, h(x) \exp \theta T(x)$ where θ is real. By using the result in (a) obtain the form of an uniformly most powerful test of size α for the hypothesis $H_0 : \theta \leq \theta_1$ or $\theta \geq \theta_2$ $(\theta_1 < \theta_2)$ against the alternative $H_1 : \theta_1 < \theta < \theta_2$.

3.a) Explain clearly when a test of a hypothesis is said to be (i) similar, (ii) of Neyman structure. Obtain a necessary and sufficient condition for all similar tests to have Neyman structure.

b) Let X be a random variable with density function $f(x, \theta, \underset{\sim}{\eta})$ of the form

$$c(\theta, \underset{\sim}{\eta})\, h(x) \cdot \exp \quad \theta U(x) + \sum_{i=1}^{k} \eta_i T_i(x)$$

where (θ, η) belongs to the $(k+1)$ dimensional Euclidean space. Show that there exists a uniformly most powerful test for the hypothesis $H_0 : \theta \leq \theta_0$ against the alternative $H_1 : \theta \geq \theta_1$ $(\theta_0 < \theta_1)$ of level α given the values $T_i(x) = t_i$ for $i = 1, \cdots, k$.

p.t.o

4.a) Let the family of possible distributions of a random variable X be $\mathcal{P} = \{P(X, \underset{\sim}{\theta}),\ \theta \in (\underline{H})\}$ and let $T(x)$ be a sufficient statistic for $\mathcal{P}$. Obtain a necessary and sufficient condition that a statistic $V(x)$ is statistically independent of $T(x)$ for all θ.

b) Use the result in (a) to show that, if a random variable X has density function of the form $c(\theta, \underset{\sim}{\eta}')\ h(x)$

$\exp\left\{\theta U(x) + \sum_{1}^{k} \eta_i T_i(x)\right\}$ where $(\theta, \underset{\sim}{\eta}')$ belongs to the $(k+1)$ dimensional Euclidean space, there exists a uniformly most powerful unbiased test of size α for the null hypothesis H_0: $\theta_1 \leq \theta \leq \theta_2$ against H_1: $\theta < \theta_1$ or $\theta > \theta_2$ provided there exists a function $h(U, T_1 \cdots T_k)$ increasing in U for fixed $T_1, \cdots, T_k$.

5.a) Let a random variable X have a Normal distribution $N(\mu, \sigma^2)$ $-\infty < \mu < \infty$ and $0 < \sigma^2 < \infty$. Based on a sample of size 10 derive a uniformly most powerful unbiased test of size 0.05 for the hypothesis $\sigma = 2$ against the alternative $\sigma \neq 2$. (State the relevant result which shows that the test you propose is uniformly most powerful unbiased).

b) Explain how you would compute the power function of the above test.

c) Obtain a 95 % confidence interval for σ^2 in the above example.

6.a) Let $P(x, \theta, \underset{\sim}{\eta})$ be a family of distributions of a random variable X, where $(\theta, \underset{\sim}{\eta}')$ belong to $(k+1)$ dimensional Euclidean space. Define the concept of an unbiased confidence set for the parameter θ.

b) Show that uniformly most accurate unbiased confidence sets of a given confidence coefficient $1-\alpha$ exist for θ if there exist UMP unbiased tests of size α for testing suitable hypothesis about θ.

c) Let a random variable X have Normal distribution $N(\mu_1, \sigma^2)$ and another random variable Y be independent of X and have Normal distribution $N(\mu_2, \sigma^2)$. Based on a sample of size 9 from the X population and a sample of size 16 from the Y population, obtain a 95 % confidence interval for $\mu_1 - \mu_2$.

7.a) Let a random variable X have distribution $P(x, \underset{\sim}{\theta})$ where $\underset{\sim}{\theta} = (\theta_1, \cdots, \theta_k)$ belongs to the k dimensional Euclidean space. Under suitable regularity conditions (to be clearly stated by you) derive Rao's test statistic for testing the hypothesis $\theta_1 = \theta_1^0, \cdots, \theta_k = \theta_k^0$ based on a sufficiently large sample.

b) What is the likelihood ratio test for the problem in (a) and what is its asymptotic distribution?

--- -------

INDIAN STATISTICAL INSTITUTE
Research and Training School
B.Stat. (Hons.) Part IV: 1974-75

ANNUAL EXAMINATION

Nonparametric Methods and Sequential Analysis

Date: 28.7.75 Maximum Marks: 100 Time: 3 hours

Note: Answer question 6 and any three of the remaining. The notation throughout is same as used in the class.

1.a) State and prove Wald's Fundamental Identity in sequential analysis.

b) Assuming that differentiation under the expectation sign is valid in the above identity, show that

$$E(Z_n, \theta) = E(n, \theta)\, E(z, \theta).$$
[15+10]=[25]

2.a) Show that the OC function of $S(b,a)$, $-\infty < b < o < a < \infty$ has the bounds

$$\frac{e^{ah(\theta)}-1}{e^{ah(\theta)}-\eta(\theta)e^{bh(\theta)}} \leq Q(\theta) \leq \frac{\delta(\theta)e^{ah(\theta)}-1}{\delta(\theta)e^{ah(\theta)}-e^{bh(\theta)}} \quad \text{if } h(\theta)>0.$$

In the above, define $h(\theta)$, $\delta(\theta)$ and $\eta(\theta)$ and state clearly the assumptions you make.

b) Let $x_1, x_2, \ldots,$ be i.i.d. random variables with common probability mass function

$$f(x,\theta) = \begin{cases} \theta^x & x = 1,2,\ldots \\ \dfrac{1-2\theta}{1-\theta} & x = 0 \end{cases}$$

where $0 < \theta < \frac{1}{2}$, and consider the SPRT $(b\Lambda, a\Lambda)$ of the hypothesis H_0: $\theta = \frac{1}{5}$ against the alternative H_1: $\theta = \frac{3}{8}$. The constants b and a, $b < 0 < a$, are integers and $\Lambda = \log(15/8)$. Compute the exact values of $Q(\theta)$ and $E(n, \theta)$ under an arbitrary value of θ. [10+15]=[25]

3. Let $x_1, x_2, \ldots$ be a sequence of random variables whose specified joint probability distribution is given by either P or P'. Let g_n and g_n' respectively be the probability density functions of $(x_1, \ldots, x_n)$ taken w.r.t. some σ-finite measure μ_n and let $z_n = g_n'/g_n$.

a) Prove that, for any $\epsilon > 1$,
$$P\left\{ z_n \geq \epsilon \text{ for some } n \geq 1\right\} \leq \frac{1}{\epsilon}$$

b) Let P_θ, $-\infty < \theta < \infty$, denote the probability distribution under which $x_1, x_2, \ldots,$ are i.i.d. $N(\theta, 1)$. Taking $P = P_0$ and $P'(\cdot) = \int_{-\infty}^{\infty} P_\theta(\cdot)\, d\Phi(\theta)$ where Φ is the distribution function of $N(0,1)$, show that, for $a > 0$,

p t o

3. (b) (contd.)

$$P\left\{|S_n| \geq \sqrt{(n+1)(a^2 + \log(1+n))} \text{ for some } n \geq 1\right\} \leq e^{-\frac{1}{2}a^2}$$

where $S_n = x_1 + x_2 + \dots + x_n$.

c) Use the result in (b) to construct a confidence sequence for θ based on a sample sequence $\{x_n\}$ from $N(\theta, 1)$. [5+10+10]-[25]

4. Let $X_1, X_2 \dots X_N$ be i.i.d. having common distribution function F which is assumed to be continuous, but otherwise arbitrary. Let R_i be the rank of X_i, $i = 1, 2, \dots N$.

a) Show that $R = (R_1, R_2, \dots, R_N)$ has a uniform distribution over $\mathcal{R}$, the class of all permutations of $(1, 2, \dots, N)$.

b) Define $S = \sum_1^N c_i\, a(R_i)$, where $c = (c_1, \dots, c_N)$ and $a = (a_1, \dots, a_N)$ are vectors of constants. Show that

$$E(S) = \frac{1}{N}\left(\sum_1^N c_i\right)\left(\sum_1^N a_i\right) \text{ and } Var(S) = \frac{1}{N-1}\left(\sum_1^N (c_i - \bar{c})^2\right)\left(\sum_1^N (a_i - \bar{a})^2\right)$$

[10+15]-[25]

5. Let $\varphi(u)$, $0 < u < 1$, be a square-integrable function. Put $\bar{\varphi} = \int_0^1 \varphi(u)du$ and assume $\int (\varphi - \bar{\varphi})^2 du > 0$. Define $a_N^{\varphi}(i) = E[\varphi(U^{(i)})]$, where $U^{(i)}$ is the i-th order statistic in a random sample of size N from Uniform (0, 1). Let $X_1, X_2, \dots, X_N$ be i.i.d. having common distribution function F which is assumed to be continuous but otherwise arbitrary. Let $R_{N,i}$ be the rank of X_i, $i = 1, 2, \dots, N$.

Define $S_c = \sum_1^N c_i\, a_N^{\varphi}(R_{N,i})$ where $c = (c_1 \dots c_N)$ is a vector of constants such that $\sum_1^N (c_i - \bar{c})^2 > 0$.

Prove that, as $\dfrac{\sum (c_i - \bar{c})^2}{\max\limits_{1 \leq i \leq N} (c_i - \bar{c})^2} \to \infty$ the statistics S_c are asymptotically normal (μ_c, σ_c^2) with $\mu_c = \bar{c} \sum_1^N a_N^{\varphi}(i)$, and $\sigma_c^2 = \sum_{i=1}^N (c_i - \bar{c})^2 \int_0^1 (\varphi(u) - \bar{\varphi})^2 du$.

(State, but do not prove, all the supplementary results required to prove the above.) [illegible]

[illegible]

6. The following are two random samples of sizes $m = n = 12$.

Sample I :	799	709	442	1109	418	749
	1181	1045	1154	1122	537	370
Sample II :	251	59	253	818	981	810
	710	491	308	166	620	599

Apply the two-sample Wilcoxon test to the above data to test H_0: The two samples come from the same population, having distribution function F which is continuous, but otherwise arbitrary.

versus H_1: the d.f. of the first population is $F(x-\Delta)$ and that of the second is $F(x)$, where $\Delta > 0$ and F is an arbitrary, but continuous, d.f. [25]

--- --------

INDIAN STATISTICAL INSTITUTE
Research and Training School
B. Stat.(Hons.) Part IV: 1974-75

ANNUAL EXAMINATION

Applications of Probability

Date: 30.7.75 | Maximum Marks: 100 | Time 3 hours

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

1. $A = [a_{ij}]$ is an infinite matrix of constants, with $a_{ii} \le 0$, $a_{ij} \ge 0$, $i \ne j$, and $\Sigma a_{ij} = 0$ for each i. $A^{(n)}$ is an $n \times n$ matrix $[a_{ij}^{(n)}]$ defined by

$$a_{ij}^{(n)} = a_{ij} \; ; \; i, j = 1, \ldots, n.$$

If $P^{(n)}(t) = [P_{ij}^{(n)}(t)]$ is a matrix of functions of t satisfying the differential equations

$$\frac{dP^{(n)}(t)}{dt} = P^{(n)}(t).A, \quad P^{(n)}(o) = I^{(n)},$$

show that

i) $P^{(n)}(s+t) = P^n(s)P^n(t)$,

ii) $P_{ij}^{(n)}(t) \ge 0$ for all $t \ge 0$,

iii) $\sum_{j=1}^{n} P_{ij}^{(n)}(t) \le 1$ for all i and all $t \ge 0$. [4+6+6]=[16]

2.a) For a linear birth and death process with birth and death coefficients λ and μ and transition probabilities $P_{ij}(t)$, the probability generating function $\Pi_i(s,t) = \sum_j s^j P_{ij}(t)$ is given by

$$\Pi_1(s,t) = [\alpha(t) + \{1 - \alpha(t) - \beta(t)\} s]^1 [1 - \beta(t)s]^{-1},$$

where $\alpha(t) = \mu(1 - e^{(\lambda-\mu)t})/[\mu - \lambda e^{(\lambda-\mu)t}]$

and $\beta(t) = \lambda(1 - e^{(\lambda-\mu)t})/[\mu - \lambda e^{(\lambda-\mu)t}]$.

i) Derive an expression for $P_{ij}(t)$.

ii) Show that the extinction probabilities are $(\mu/\lambda)^1$ or 1 according as $\mu < \lambda$ or $\mu \ge \lambda$.

b) Prove Whittle's probabilistic threshold theorem that the size of an epidemic (with relative removal rate ρ) does not exceed a proportion i of the initial number of susceptibles. [4+6+14]=[24]

3.a) If $\{X_j\}$ is a sequence of positive-valued independent random variables with $E(\log X_j) = \mu_j$ and $V(\log X_j) = \sigma_j^2$,

i) State a set of sufficient conditions that $\prod_{j=1}^{n} X_j$ will have a log normal distribution in the limit when $n \to \infty$.

ii) If, in the above case, each X_j has a lognormal distribution, derive the distribution of $X_1 \ldots X_k / X_{k+1} \ldots X_n$.

p t o

3.b) State the Gibrat-Kapetyn 'law of proportionate' effect for a sequence of random variables $\{X_j\}$ and show that, under specified conditions, X_n tends to follow a lognormal distribution as $n \to \infty$.

c) In a process of successive breakages, assume that the probability $G_j(x|u)$ for an element of size u at the $(j-1)$st stage of breakage to give rise to an element of size $\leq x$ at the jth stage is a function of $\frac{x}{u}$ i.e., $G_j(x|u) = \pi_j(\frac{x}{u})$.

Prove that in the long run the distribution of sizes will be lognormal. [5+5+5+5]=[20]

4.a) For any finite probability system E with states $E_1, \ldots, E_n$ and $P(E_i) = p_i$, $p_i \geq 0$, $\sum_{i=1}^{n} p_i = 1$, it is desired to define 'entropy' as a function $H(p_1, \ldots, p_n)$ [which may also be denoted as $H(E)$] which would be continuous in each variable and have the properties

i) For given n, $H(p_1, \ldots, p_n)$ is maximum when

$$p_1 = \cdots = p_n = \frac{1}{n};$$

ii) $H(EF) = H(E) + H(F|E)$; and

iii) $H(p_1, \ldots, p_n, 0) = H(p_1, \ldots, p_n)$.

Prove that $H(p_1, \ldots, p_n)$ must be of the form

$$H(p_1, \ldots, p_n) = -\lambda \sum_{i=1}^{n} p_i \log p_i .$$

b) Show that, for the function H so defined,

$$H(F|E) \leq H(F).$$

[13+5]=[18]

5.a) Define an 'instantaneous uniquely decodable' code and give an example of such a code for binary coding.

b) Given a source alphabet $S : \{s_1, \ldots, s_t\}$ and a code alphabet $X : \{x_1, \ldots, x_r\}$, prove that a necessary and sufficient condition for the existence of an instantaneous code with word lengths $\ell_1, \ldots, \ell_t$ is

$$\sum_{i=1}^{t} r^{-\ell_i} \leq 1.$$

c) The following table shows the word lengths in three codes with code alphabet $X : \{0, 1, 2\}$.

word length l_i	no. of words with word length l_i		
	code A	code B	code C
1	2	1	3
2	2	4	2
3	2	6	3
4	3	0	2
5	2	0	1

i) Examine in each case if the lengths are acceptable for an instantaneous uniquely decodable code.

ii) Construct an instantaneous code in each acceptable case. [3+7+2+2+2+6]=[22]
