INDIAN STATISTICAL INSTITUTE
Research and Training School

Periodical Examination
B.Stat. IV year
PROBABILITY

Duration: 3 hours | Date: 3 October 1963

1. Let ζ be the number of successes in n independent Bernoulli trials with probability of success p. Prove that for any α and β with $\alpha < \beta$

$$P\left\{np + \alpha (npq)^{\frac{1}{2}} \leq \zeta \leq np + \beta (npq)^{\frac{1}{2}}\right\} \longrightarrow F(\beta) - F(\alpha)$$

as $n \longrightarrow \infty$, where

$$F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2}\, dt.$$

2. Let ζ be the number of successes in n Bernoulli trials with probability of success p. If $r \geq np$, show that

$$P\left\{\zeta \geq r\right\} \leq \left\{\frac{(r+1)q}{r+1-(n+1)p}\right\} \cdot P\left\{\zeta = r\right\}$$

3. Discuss briefly the construction of the sample space corresponding to <u>infinite</u> sequences of Bernoulli trials (with probability of success in one trial equal to p).

Prove that if ζ is the number of successes in n trials,

$$P\left\{\frac{1}{n}\zeta \longrightarrow p\right\} = 1 \qquad (*)$$

Explain briefly the differences between the above statement (*) and the statement that for each $\epsilon > 0$,

$$P\left\{\left|\frac{1}{n}\zeta - p\right| > \epsilon\right\} \rightarrow 0$$

as $n \rightarrow \infty$.

INDIAN STATISTICAL INSTITUTE
Research and Training School

Periodical Examination
INFERENCE.
B. Stat. IV Year

Duration : 2 hrs. 30 mints. Maximum Marks:100 Date: 10. 10. 63

1. Answer the following questions in the context of a statistical decision problem.

(a) What are the domain and range of the loss function? (5)

(b) What are the domain and range of a decision function? (5)

(c) What is the risk of a given decision function for a particular value of the parameter? (5)

(d) What is a Bayes solution with respect to a given prior distribution of the parameter? (5)

(e) What is an admissible decision function? (5)

(f) What is a minimax decision function? (5)

2(a) X is a real-valued random variable which has a density function $f(x,\theta)$ where the parameter θ is unknown. ϕ is a real-valued function of θ. What is the Cramer-Rao lower bound for the variance of an arbitrary unbiased estimator of $\phi(\theta)$. State a set of regularity conditions which is sufficient for the validity of this bound. Under what conditions does there exist an unbiased estimator for which the above minimum variance is attained? How will the Cramer-Rao lower bound be modified for the mean - squared error of a biased estimator? (No proof is needed.) (25)

(b) Suppose $X_1, \ldots, X_n$ are independent normal variates with mean θ and variance 1. Find the Cramér-Rao lower bound for the variance of an unbiased estimator for θ^2 and examine whether it is attained. (10)

3. X is a real-valued random variable with distribution function F for which there exists a density function f. f_0 and f_1 are two completely specified probability density functions. Consider the problem of deciding whether $f = f_0$ or $f = f_1$ where there is no loss in making a correct decision but the loss in deciding $f = f_1$ when $f = f_0$ and the loss in deciding $f = f_0$ when $f = f_1$ are 1 and $w > 0$ units respectively.

(a) Find a Bayes solution for this problem in the class of all randomized decision functions with respect to a prior distribution under which $f = f_0$ with probability p and $f = f_1$ with probability $1 - p$, $0 < p < 1$. (10)

(b) Let ϕ_p be the Bayes solution obtained in (a) which leads to the decision $f = f_1$ with probability $\phi_p(x)$ when x is observed. Show that if a randomized decision function ϕ satisfies

$$\int \phi(x)\, f_0(x)\, dx \leq \int \phi_p(x)\, f_0(x)\, dx,$$

then $\int \phi(x)\, f_1(x)\, dx \leq \int \phi_p(x)\, f_1(x)\, dx$ holds. (5)

(PLEASE TURN OVER)

(c) State Neyman - Pearson lemma about the necessary and sufficient conditions for the most powerful test of level α for $H_0 : f = f_0$ against $H_1 : f = f_1$.

(d) $H_0 : f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$

against $H_1 : f(x) = \frac{1}{2\sqrt{2\pi}} \left(e^{-(x-1)^2/2} + e^{-(x+1)^2/2}\right)$

INDIAN STATISTICAL INSTITUTE
Research and Training School

Periodical Examination
B. Stat. IV year

STATISTICAL METHODS

Duration : $2\frac{1}{2}$ hours   Maximum Marks 100   Date: 7 Nov. 1963

Answer as many questions as you can. Each question carries 20 marks.

1. Explain how statistical methods can be applied to estimate the cranial capacity of a skull from a set of external measurements on the skull. Explain what data you need and derive the formula to be used.

2(a). Let $y_1 \ldots y_n$ be independent normal with $E(y_i) = \alpha + \beta x_i$ and $V(y_i) = \sigma^2$ where $x_1 \ldots x_n$ are given. Show that under the hypothesis $\beta = 0$, the statistic $\frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$ follows a t-distribution.

(b). If in addition to (a), $x_1 \ldots x_n$ are themselves independent and identical normal variates, then write down the joint distribution of x and y. What is the relation between β and the population coefficient ρ. Show that if $\rho = 0$, the over-all distribution of $\frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$ is still a t-distribution.

3. Given the heights of son (y_i) and of further (x_i) for n pairs of fathers and sons, explain how you will set up confidence interval for (a) the expected value of the height of a son of a father whose height is given (b) the actual value of the height of the son whose father's height is given.

4. Explain briefly how you will set up the analysis of variance table in the case of a two way classification with unequal and multiple observations in a cell? What tests will you perform and how?

5. Given measurements on p characteristics $x_1 \ldots x_p$ for n individuals, explain how you will fit the multiple regression

$$x_1 = \alpha + \beta_2 x_2 + \ldots + \beta_p x_p$$ and test for the

hypothesis $\beta_2 = \ldots = \beta_p = 0$. What is the physical meaning of this hypothesis?

Show that the statistic you use for this test is also equal to

$$\frac{R^2_{1.23\ldots p}/(p-1)}{(p - R^2_{1.23\ldots p})/(n-p)}$$

where $R_{1.23\ldots p}$ is the sample multiple correlation coefficient of x_1 on $x_2 \ldots x_p$.

6. Describe Bartlett's test for homogenity of variances of k normal populations, giving all the details regarding the procedure.

INDIAN STATISTICAL INSTITUTE
Research and Training School

Periodical Examination

B. Stat. IV Year

DESIGN OF EXPERIMENTS

Duration : $2\frac{1}{2}$ hours   Date: 31.10.63

Each question carries 20 marks

1. Write a letter to the editor of a newspaper making a case for appointment of a statistician in a newly organised government institute of experimental biology.

2. Explain the role of the techniques of randomisation, replication and error control in planning an experiment.

3. Describe a randomised blocks layout, explaining how the blocks are formed. Give the structure of the analysis of variance table for an experiment in randomised blocks.

4. Select a random 7 x 7 latin square. Explain how in a Latin square design, heterogeneity is eliminated in two orthogonal directions.

5. Explain the advantages of a factorial type of experiment over the classical one-factor-at-a-time type of experiment.

-.........-

INDIAN STATISTICAL INSTITUTE
Research and Training School

Periodical Examination

B.Stat. IV Year

Statistical Quality Control

Duration: 3 hours | Maximum Marks: 100 | Date: 14.11.65

1. An industrialist, who has not got much knowledge of mathematics, wants to know what problems can be solved in industry by S.Q.C techniques. Prepare a note to be sent to him. (15)

2. Explain, with examples, how quality characteristics can be classified, mentioning also the types of inspection from which each class results. (6)

3. What are the different ways in which item quality can be specified? (5)

4. What are the relative merits of sampling by attributes and sampling by variables? (5)

5. What are group control charts? When can you draw them? What are the advantages of such charts? (8)

6. Mention a few circumstances in which you will draw a control chart for individuals. (3)

7. Explain briefly what is meant by lot acceptance sampling. (3)

8. What are the different criteria that can be used to classify acceptance sampling plans? (5)

9. (a) Get a suitable single sampling plan from Mil-std-105A, to meet the following requirements.

   Lot size 2500, AQL = 4 percent defective. (4)

   (b) What are the plans to be used, if you decide to

      (i) tighten the inspection (2)
      (ii) reduce the inspection? (2)

   (c) The plan chosen in (a) is in operation and the first 10 lots are sampled under normal inspection. The number of defectives found are 0, 15, 9, 8, 3, 10, 11, 7, 6 and 11 respectively. Should you like to change over to tightened inspection? (5)

   (d) Suppose the plan you have chosen in 9(b) - (i) is in operation. The number of defectives found in the last 10 samples are 12, 7, 5, 0, 3, 11, 8, 2, 6 and 7 respectively. Can normal inspection be reinstated? (5)

10. Obtain 95% confidence interval for

   (i) Binomial proportion, given that in a sample of 200 items 9 defectives were observed. (3)

   (ii) ~~Arithmetic~~ Poisson average given that on 15 items 23 defectives were found. (3)

   (iii) Arithmetic mean of a normally distributed characteristic, given that in a sample of 15 items the average $\bar{x}$ was 13.75 and s.d.s (divisor (n-1)) was 1.23. (5)

11. 25 samples of five items each are taken from a process at regular intervals. $\bar{x}$ and R are calculated for each sample for a certain characteristic, x. We get $\sum \bar{x} = 358.50$ and $\sum R = 9.80$.

(a) Compute the control limits for the $\bar{x}$ and R charts. (6)

(b) Assuming that the process is in control at the level found in (a), what are the natural tolerance limits of the process? (6)

(c) If the specification limits are 14.40 ± 0.45 what conclusions can you draw concerning the ability of the process to produce items within these specifications? (3)

(d) What percentage of the lot fall outside the specification limits if the process is in control as stated in (b). (5)

12. The specification limits on a measurable characteristic are 100 ± 30 units. The process is such that in the course of production the process mean is likely to shift, but the process variability remains almost stable 10 samples of size 5 each taken at random from the process have given the following values of the sample range 9.4, 4.5, 6.7, 11.4, 12.3, 3.6, 8.4, 10.4, 7.6, and 12.9.

(a) Estimate the process capability (5)

(b) Find out the interval within which the process mean can wander while still giving a proportion of outside specification items ≤ 0.0027. Technical action for resetting the process level is quite expensive. Hence the management would like to leave the process untouched as long as the process mean is such that the proportion of out of specification items does not exceed 0.0027. (10)

(c) Explain how to set up a control chart for $\bar{x}$ $(n = 5)$ in such a case. (5)

....

INDIAN STATISTICAL INSTITUTE
Research and Training School

Periodical Examination

B. Stat. IV year

ECONOMICS

Duration : 3 hours | Date : 17 October 1963

Note: Separate booklet should be used for each Group

GROUP A

Attempt any two questions.

1. Comment on the following :

   "Unlike minimum reserve and open market policy, discount policy by itself does not guarantee the Central bank's control over the commercial banks' lending potential." (20)

2. Classify inflation with respect to its causal factors and explain the characteristics of different kinds of inflation. (20)

3. Explain the structure of the London discount market. How does the Bank of England play its part in this market? (20)

GROUP B

1. Define the concentration curve for an 'income' distribution, and discuss its properties. What are its uses? How is it related to the Gini Mean Difference? How is concentration different from variability? (24)

   OR, Indicate briefly the different types of uses of family budget data. Describe in detail the estimation of Engel curves for different items.

2. Suppose you have carried out a family budget enquiry covering all middle class households living in a certain region. Describe in detail how you do the following (attempt any two) :

   (a) Examine, in a preliminary way, whether the distribution of persons by per capita household expenditure on all items, follows the Pareto or the lognormal distribution.

   (b) Construct the specific concentration curve and estimate the specific concentration coefficient for an item like sugar.

   (c) Examine whether the Engel curve for sugar (say) has the constant elasticity curve to the observed data.

   You may assume that the sample of households is simple random. But the data are available only in a grouped form by intervals of per capita household expenditure. (36)

ooooooooooo

INDIAN STATISTICAL INSTITUTE
Research and Training School

Periodical Examination

B.Stat. IV Year

DEMOGRAPHY

Duration: 2 hours Maximum Marks: 100 Date: 26. 9. 63

1. a) Enumerate the items of information recorded in the Indian birth and death registers.

b) State some of the important recommendations of the expert bodies set up in recent years to examine the question of improving the collection, registration and compilation of vital statistics in this country.

2. a) Enumerate clearly the essential features of an Abridged Life Table.

b) Define "force of mortality". The usual relationship between $\overset{o}{e}_x$ and e_x is given by $\overset{o}{e}_x = e_x + \frac{1}{2}$. For which value of x this relationship is absolutely untenable? Give reasons.

INDIAN STATISTICAL INSTITUTE
Research and Training School

Mid-term examination, 1963
B.Stat. IV Year

PROBABILITY

Duration: $2\frac{1}{2}$ hours   Maximum Marks: 100   Date: 5 December 1963

1(a). If $A_1, A_2, \ldots$ is a sequence of events and if $\sum_n P(A_n) < \infty$, prove that the probability is one that only finitely many of the events occur.

(b) Let Ω be the sample space of sequences $(a_1, a_2, \ldots)$ (where each $a_j = 0$ or 1) associated with an unlimited sequence of Bernoulli trials, the probability of success being p $(0 < p < 1)$. Write, for a sequence $\omega = (a_1, a_2, \ldots)$

$$s_n(\omega) = a_1 + \ldots + a_n$$

$$s_n^*(\omega) = \frac{s_n(\omega) - np}{\sqrt{npq}}$$

Prove that

$$\limsup_{n \to \infty} \frac{s_n^*(\omega)}{\sqrt{2 \log \log n}} = +1$$

for almost all sequences ω.

(c) Give an example of a sequence $\omega = (a_1, a_2, \ldots)$ in Ω such that

$$\liminf_{n \to \infty} \frac{s_n(\omega)}{n} = \frac{1}{4}, \quad \limsup_{n \to \infty} \frac{s_n(\omega)}{n} = \frac{3}{4}.$$

2(a). Let X be a finite set, having elements $x_1, x_2, \ldots, x_k$ $(k \geq 2)$. Let X_∞ be the space of all sequences $(y_1, y_2, \ldots)$ where each y_j is in X. Define what is meant by primitive cylinder subsets of X_∞.

(b). Give an example of a primitive cylinder set of rank 3. How many primitive cylinder sets (excluding the empty set) are there, for a given rank m?

(c). Prove that the class of all cylinder sets is closed under all finite set operations.

(P.T.O.)

(d). Let $X = \{0,1\}$ so that $k = 2$, $x_1 = 0$, $x_2 = 1$ and X_∞ is the space of all sequences which are composed entirely of zeros and ones. Find out which of the following sets are cylinder sets.

(1) $A = \left\{(a_1, a_2, \ldots\ldots) : \Sigma \frac{a_n}{n} < \infty\right\}$

(2) $A = \left\{(a_1, a_2, \ldots\ldots) : a_1 + a_2 + 3a_3 = 4\right\}$

(3) $A = \left\{(a_1, a_2, \ldots\ldots) : a_{2k} = a_{2k+1} \text{ for all } k\right\}$

3(a). Let X be the number of successes in n Bernoulli trials. Find $E(X - np)^4$.

(b). Prove, using Chebyshev's inequality that

$$P\left\{\left|\frac{X - np}{\sqrt{npq}}\right| \geq \delta\right\} \leq \frac{M}{\delta^4}$$

where M is a constant independent of n and δ.

Hence show that

$$P\left\{\left|\frac{X}{n} - p\right| \geq \frac{1}{n^{1/8}}\right\} \leq \frac{M}{n^{3/2}}$$

(Hint: Choose $\delta\sqrt{pq} = n^{3/8}$)

(c). Deduce from (b) that

$$P\left(\frac{X}{n} \longrightarrow p\right) = 1.$$

oooooOooooo

INDIAN STATISTICAL INSTITUTE
Research and Training School
Mid-term examination, 1963

B. Stat. IV Year

STATISTICAL INFERENCE

Duration: $2\frac{1}{2}$ hours | Maximum Marks: 100 | Date: 5 December 1963

Note: Separate Answer Book should be used for each group. Answer two questions from each group.

GROUP A

1(a). $f(x;\theta)$ is the density function of a population with unknown parameter θ. $T(x_1, \ldots, x_n)$ is an unbiased estimator of θ computed from a random sample $x_1, \ldots, x_n$. Explain how you would get an unbiased estimator with less variance than T. (10)

(b). Let $f(x;\theta) = \frac{1}{\theta}$ $\quad 0 \leq x \leq \theta$

$= 0$ $\quad$ Otherwise.

$(x_1, \ldots, x_n)$ is a random sample drawn from the above population. $2x_1$ is an unbiased estimator for θ. Obtain an estimator which has smaller variance than this estimator. (15)

2(a). The frequency function of a distribution is given as $f(x;\theta)$ where θ ranges over a non-degenerate interval A and the domain of x is independent of θ. It is also given that a real valued function $T(x_1, \ldots, x_n)$ calculated from a random sample of size n drawn from the population is sufficient for θ. Obtain the form of the likelihood function of the sample $x_1 \ldots x_n$. (8)

(b). Let $P(X = 0) = p$

$P(X = 1) = 1-p$

A random sample $(x_1, \ldots, x_n)$ has been drawn from the above population. Show that $T(x_1 \ldots x_n) = \sum_1^n x_i$ is a sufficient statistic for p. (9)

(c). Examine whether the statistic in (b) is complete. (9)

3(a). $f(x;\theta)$ is the density function of a one parameter family of distributions, θ ranging over a non-degenerate interval A. $x_1, \ldots, x_n$ are independent observations from the population, $T(x_1, \ldots, x_n)$ is an unbiased estimator of θ whose variance attains the Cramér - Rao lower bound. Prove that the joint density function of $x_1, \ldots, x_n$ is of the form

$$g(T(x_1, \ldots, x_n); \theta). h(x_1, \ldots, x_n)$$

(P.T.O.)

where $g(T;\theta)$ satisfies

$$\frac{d \log g(T;\theta)}{d\theta} = k(\theta)\ (T-\theta),$$

with $k(\theta)$ not depending on T.

(b). Give an example of an unbiased estimator of θ, which is sufficient for θ but does not attain the Cramer – Rao lower bound.

GROUP B

4(a). Explain each of the following concepts in a single sentence :

(i) Statistical model.

(ii) Null hypothesis and alternative hypothesis.

(iii) Critical region of a test.

(iv) Level of significance of a test.

(v) Power of a test.

(b). A manufacturer of electric bulbs claims that the length of life (in hours) of a randomly selected bulb produced in his factory will have the probability density function

$$\frac{1}{\theta}\, e^{-x/\theta}\ , \quad x > 0$$

with $\theta = \theta_0 > 0$. A purchaser accepts the form of the density function but wants to test whether $\theta = \theta_0$ or $\theta < \theta_0$ on the basis of a sample. He selects N bulbs at random (independent selections) and lights them simultaneously. At the end of time t(hours) he finds that n out of N bulbs are still burning. Find the uniformly most powerful test of level α for testing $H_0 : \theta = \theta_0$ against $H_1 : \theta > \theta_0$.

5. $X_1, X_2, \ldots$ is an infinite sequence of independent random variables with common unknown probability density function f.

(a) State the necessary and sufficient conditions for the most powerful test ϕ_n of level α for $H_0 : f = f_0$ against $H_1 : f = f_1$ based on $(X_1, \ldots, X_n)$ where f_0 and f_1 are given distinct probability density functions.

(b). Let ϕ_0 be the test which rejects H_0 with probability α irrespective of the observations. Restricting to the sample space of $(X_1, \ldots, X_n)$ use the necessary condition stated in a) to compare the power γ_{n-1} of ϕ_{n-1} with the power γ_n of ϕ_n and prove that

$$\alpha < \gamma_1 < \gamma_2 < \cdots\cdots < \gamma_n$$

(c) Show that for any given $0 < \alpha < 1$, $0 < \beta < 1$, there exists an integer N and a test for H_0 against H_1 based on $(X_1, \ldots, X_n)$ for which the probability of first kind of error is less than α and the probability of second kind of error is less than β. (10)

6. $X_1, \ldots, X_m$, $Y_1, \ldots, Y_n$ are independent normal random variables, X_i's having mean θ_1 (unknown), variance 1 and Y_i's having mean θ_2 (unknown), variance 2. The null hypothesis $H_0 : \theta_1 = \theta_2$ is to be tested against the alternative $H_1 : \theta_1 = \theta_2 + \delta$, $\delta > 0$.

a) Find $C_{m,n}$ such that the test with critical region

$$R_{m,n} = \left\{ (x_1, \ldots, x_m, y_1, \ldots, y_n) : \frac{1}{m} \sum_1^m x_i - \frac{1}{n} \sum_1^n y_i > C_{m,n} \right\}$$

is of level α for testing H_0. (5)

b) For this problem,

$$P(x_1, \ldots, x_m, y_1, \ldots, y_n) = \Pr\left[\frac{1}{m} \sum_1^m X_i - \frac{1}{n} \sum_1^n Y_i > \frac{1}{m} \sum_1^m x_i - \frac{1}{n} \sum_1^n y_i / H_0\right]$$

is called the significance probability for the sample $(x_1, \ldots, x_m, y_1, \ldots, y_n)$ where $\Pr[E/H_0]$ denotes the probability of the event E when H_0 holds. Write down the expression for the significance probability for a given sample and show that for a proper choice of $d_{m,n}$, the critical region

$$S_{m,n} = \left\{ (x_1, \ldots, x_m, y_1, \ldots, y_n) : P(x_1, \ldots, x_m, y_1, \ldots, y_n) < d_{m,n} \right\}$$

is the same as $R_{m,n}$. (5)

c) Show that when $\theta_1 = \theta_2 + \delta$, the test with critical region $R_{m,n}$ accepts H_0 with probability

$$\beta_{m,n} = \Phi\left(\Phi^{-1}(1-\alpha) - \delta\sqrt{\frac{mn}{2m+n}} \right),$$

where $\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-t^2/2}\, dt$ and $\Phi^{-1}(u)$ is the solution of the equation $\Phi(x) = u$. (8)

d) If you are allowed a given total sample size N (i.e. $m+n=N$), how would you choose m and n so as to maximise $\beta_{m,n}$? (7)

INDIAN STATISTICAL INSTITUTE
Research and Training School

Mid-term examination, 1963

B. Stat. IV Year

STATISTICAL METHODS I (Theory)

Duration: $2\frac{1}{2}$ hours | Maximum Marks: 100 | Date: 3 December 1963

1(a). State and prove Cochran's theorem regarding distribution of several quadratic forms in standard normal variates and explain its usefulness in analysis of variance.

(b). Show that the quadratic form $\Sigma(x_i-\bar{x})^2$ can be expressed as the sum of squares of (n-1) linear functions of the x's. What do you conclude from this regarding the rank of the quadratic form?.

(c). What can you conclude from the identity $\Sigma x_i^2 = \Sigma (x_i-\bar{x})^2 + n\bar{x}^2$ regarding the rank of $\Sigma(x_i - \bar{x})^2$.

(Hint:- Both (b) and (c) are required to establish the rank of $\Sigma(x_i-\bar{x})^2$)

(d). Use Cochran's theorem to establish the independence of $\bar{x}$ and s in sampling from normal population. Also show that Cochran's theorem enables us to obtain their actual distributions also. (30)

2(a). Define the multiple correlation coefficient $R_{1.23}$ of x_1 on x_2 and x_3 and derive an expression for $R_{1.23}$ in terms of ordinary correlation coefficients r_{ij}.

(b). Define the partial correlation coefficient $r_{12.3}$ between x_1 and x_2 eleminating x_3 and obtain an expression for $r_{12.3}$ in terms of ordinary correlation coefficient r_{ij}.

(c). From (a) and (b) deduce that

$$(1-R^2_{1.23}) = (1-r^2_{1.23})\ (1-r^2_{13})$$

What is the significance of this relation? (20)

3. Given n observations $(y_i\ x_i\ t_i)$ on 3 associated variables y, x, and t, let a+bx+ct be the best linear regression of y on x and t. Let $\alpha + \beta t$ be the best linear regression y on t. Let $Z = y-\alpha-\beta t$. Show that if the best linear regression of Z on x is A+Bx, then B = b. What is the statistical significence of this result? (18)

4(a). Define the bivariate normal distribution giving their joint probability density function. Show that

i) the marginal distribution of x and y are normal

ii) the conditional distribution of y given x (as well as x given y) is also normal

iii) the regression of y and x (as well as of x and y) is linear

(P.T.O.)

iv) the conditional distributions are homoscedatic.

v) If $\rho = 0$, then x and y are independent.

(b). Define the multivariate normal distribution. Find the constant of integration. Show that the variance, covariance matrix is the inverse of the matrix occuring in the definition.

5(a). If $x_1\ x_2 \dots x_m$ be independent $N(\mu_1, \sigma_1)$ and $y_1 \dots y_n$ independent $N(\mu_2, \sigma_2)$. If $(m-1)\ s_1^2 = \Sigma (x_i - \bar{x})^2$ and $(n-1)\ s_2^2 = \Sigma (y_i - \bar{y})^2$, then what is the distribution of

$$\frac{s_1^2 / \sigma_1^2}{s_2^2 / \sigma_2^2}$$

? From this obtain 95% confidence limits for the ratio of the variance $\frac{\sigma_1^2}{\sigma_2^2}$?

(b). Give an example of a pair of random variables x and y having $\rho = 0$ but which are not independent.

-.....-

INDIAN STATISTICAL INSTITUTE
Research and Training School
Mid-term examination, 1963

B. Stat. IV Year

STATISTICAL METHODS II (Practical)

Duration: 3 hours   Maximum Marks: 100   Date: 3 December 1963

Answer as many as you can.

1. A steel bar 18 inches long is subjected to a carefully regulated hardening process. The hardness is determined at the extremities of the bar, and at nine positions in between. The following results are obtained :

Distance from end of bar (d) inches	0	1.8	3.6	5.4	7.2	9.0	10.8	12.6	14.4	16.2	18.0
Vickers hardness number (h)	250	276	298	335	374	414	454	503	558	604	671

It is required to determine a mathematical function to graduate the change in hardness along the bar. Two forms of function are suggested

i) $h = A + Bd + Cd^2$

ii) $h = \alpha e^{\beta d}$

Which of these formulas appear to give the better representation of the changes in hardness along the bar?

(Hint:- Fit both these curves to the data. Compare the S.S. of residuals. Or graphical compare the goodness of fit. Work with $x = \frac{d}{1.8}$ or $\frac{d-9}{1.8}$ to simplify computations). (40)

2(a). Two experiments A and B take repeated measurements (in mm) on the length of a copper wire and the data are given below. Test whether B's measurements are more accurate than A's (It may be supposed that the readings taken by both are unbiased).

A's readings: 12.47, 12.44, 11.90, 12.13, 12.77, 11.86, 12.78, 11.96, 12.25, 12.27

B's readings: 12.06, 12.34, 12.39, 12.16, 12.23, 11.98, 12.22, 12.46

(b). The coefficient between nasal length and stature for a group of 20 Indian adult males was found to be·203. Test whether there is any correlation between the characters in the population. (20)

3. From measurements for each of 18 cinchona plants on y (the yield of dry park in oz.), x_1 (the height in inches), and x_2 (the girth in inches), at a level 6" above ground, the following quantities were computed.

	y	x_1	x_2
Sums	581	179	66
S.S.	22,293	2,133	278
	yx_1	x_1x_2	x_2y
S.P.	6,636	715	2387

(P.T.O.)

a) Obtain the multiple regression equation of y on x_1 and x_2.

b) Examine whether x_1 and x_2 are jointly useful for predictery y.

c) Examine whether x_2 is useful when x_1 is already there.

d) For $x_1 = 8$ and $y_2 = 4$ obtain 95% confidence interval for the corresponding mean value of y. (25)

5. The following table gives the moisture content (in gms.) of a certain food product. The levels of factor A are 3 kinds of salts those of B are amounts of acid and those of C are two different additives. Perform analysis of variance test for main effects and interactions. Describe your findings in simple language.

	level of B			
	1		2	
level of A	level of C 1	level of C 2	level of C 1	level of C 2
1	8	5	8	4
2	17	11	13	10
3	22	16	20	15

(10)

00000O00000

INDIAN STATISTICAL INSTITUTE
Research and Training School
Mid-term examination, 1963

B.Stat. IV Year

DESIGN OF EXPERIMENTS

Duration: 3½ hours. Date: 14 December 1963

Attempt Two questions from Group A
and Two questions from Group B.

GROUP A

1. Describe the Latin Square lay-out and discuss the advantages and disadvantages of this design. Show the structure (that is, the components with their degrees of freedom) of the analysis of variance appropriate for this design.

2. Define a balanced incomplete block design with parameters (b, v, r, k, λ). What is the need for such designs? Show that for such designs $b \geq v$.

3. Explain the terms: main effects and interaction with reference to a factorial experiment involving 2 factors, each at 2 levels.

   What are the advantages of a factorial type of experiment over the classical one-factor-at-a-time type of experiment?

GROUP B

4. What is meant by "confounding" in a factorial experiment? Why is it taken recourse to?

   Draw up a scheme of balanced partial confounding in a factorial experiment involving 5 factors each at 2 levels, in blocks of 8 plots.

5. The yields in lbs. and the lay-out of a varietal trial involving 5 varieties of wheat in 4 randomised blocks are given below.

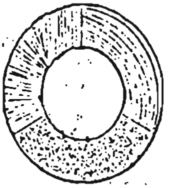
   Analyse the data and arrange the varieties in decreasing order of yield.

Block I :	(1) 28	(3) 27	(2) 31	(4) 16	(5) 22
Block II :	(3) 23	(2) 29	(1) 27	(5) 15	(4) 20
Block III :	(5) 15	(4) 20	(1) 39	(3) 37	(2) 43
Block IV :	(2) 31	(5) 18	(1) 29	(4) 14	(3) 35

(P.T.O.)

6. To compare the suitability of four different types (A, B, C, D) of rubber in the production of motor car tyres each of four tyres was built in three sections, using one type of rubber for each section. The tyres were fitted on a car and run for a fixed length of time. The loss in weight due to road friction for each of the twelve sections is shown below (in unspecified units). Analyse the data and give your comments.

	sections		
tyres	1	2	3
1	A 24	B 24	C 28
2	D 31	A 20	B 21
3	C 33	D 37	A 25
4	B 31	D 41	C 42



the three sections of a ty

INDIAN STATISTICAL INSTITUTE
Research and Training School

Mid-term examination, 1963

B.Stat. IV Year

SQC

Duration: 3 hours | Maximum Marks: 100 | Date: 7 December 1963

1. Write a note on Military Standard 105A bringing out all its salient features. (18)

2. Describe briefly Dodge's continuous sampling plan (CSP-1). In what ways CSP-2 and CSP-3 are different from it? Can you suggest any other variations that can be made in CSP-1? (18)

Or,

Derive the expression for AOQ in CSP-1 in terms of the elements of the plan, i and f. (18)

3. Lot quality is specified by proportion-defective. An individual product is defective if its quality $x > U$; otherwise it is non-defective. A sampling scheme is specified as follows :

Sample size. n

if $\bar{x} \leq U - k\sigma$, accept the lot.

$\bar{x} > U - k\sigma$, reject the lot.

$\bar{x}$ is the mean of the sample and σ is the known standard deviation of the population.

(a) Stating clearly the assumptions, if any, derive a formula to obtain points on the OC-curve of the plan. (10)

(b) Using the above formula, obtain the elements of the sampling plan given that P_1, P_2, α and β are AQL, LTPD and the two corresponding risks. (8)

Control charts for $\bar{X}$ and s are maintained on the weight in ounces of the contents of a certain container. The subgroup size is 10. After 18 subgroups, $\Sigma \bar{X} = 595.8$ and $\Sigma s = 8.24$.

(a) Compute the values of the 3σ limits for the $\bar{X}$, s charts. Estimate the value of σ, the process standard deviation. What will be the standard error of the average of 250 measurements? (12)

(b) The minimum specification for the weight is 32 ozs. It is desired to hold the overfill to as low a value as possible consistent with meeting this specification. What should be the aimed at value of the process average? (6)

(c) What was the average over-fill during the period covered by the control chart in (a)? If it were permissible for 5% to be below 32 ozs., how much could the average over-fill be reduced below its present value? (9)

(P.T.O.)

5. Derive an item-by-item sequential sampling plan and set up a graphical procedure given that $P_1 = 0.01$, $P_2 = 0.04$ $\alpha = 0.05$ and $\beta = 0.10$. (20)

6. A single sampling acceptance-rectification plan for attributes having elements $n = 10$ and $c = 1$ is under operation, the lot size being 500.

(a) Calculate the probability of accepting a lot with incoming lot quality $p = 0.2$ (0.2) 0.8. (8)

(b) Find the values of AOQ in the above cases. (3)

(c) How many items on the average will be inspected per lot in each of the above cases? (5)

INDIAN STATISTICAL INSTITUTE
Research and Training School

Mid-term examination 1963

B. Stat. IV Year

ECONOMICS -I

Duration: 2 hours | Date: 2 December 1963

Note: Answer any four

1. Do logarithmic demand functions of the form

$$\log E_{ij} = \alpha + \beta \log y_j + u_{ij}$$

provide good graduations to all components of family expenditures and savings in the estimation of Engel curves from family-budget data?

E_{ij} = expenditure on the i-th commodity by the j-th family unit

y_j = income of the j-th family unit

u_{ij} = random disturbance

α and β are constants

2. Carefully distinguish between the Lorenz curve and the specific concentration curve. Indicate how these may be used in the estimation of Engel elasticities.

3. Why are single cross-section samples not suitable for the estimation of price elasticities of demand? Would a time series of cross-section samples be more suitable for this purpose?

4. Outline some of the major criticisms against the use of the Cobb-Douglas type of production function for Indian manufacturing industries.

5. How would you devise a statistical test for the existence of returns to scale in production (increasing, constant, decreasing)?

6. Write short notes on the following :

i). Pareto $-\alpha$

ii). Durbin-Watson test

iii). Multicollinearity.

-.....-

INDIAN STATISTICAL INSTITUTE
Research and Training School

Mid-term examination, 1963

B.Stat. IV Year

ECONOMICS II

Duration: 2 hours | Date: 2 December 1963

Attempt Q.1 and any two of the rest. All questions carry equal marks.

1. Comment on the statement that the Quantity Theory of money is highly over-simplified but in times of inflation it comes into its own.

2(a). Represent graphically the different sacrifice principles stating clearly the assumptions you will make for such representation.

(b). If $\log_e y$ represents the utility of an income y and if t be the rate of taxation on income, discuss the nature of the tax system when

(i) there is equal absolute sacrifice of utility on all incomes

(ii) there is equal proportional sacrifice of utility on all incomes.

3. "The supply and demand for money-holding determine the rate of interest."

"Business demand and household supply of earning assets are the determinats of the rate of interest."

Develop the ideas contained in the two theories and show whether they are equivalent. Would you consider the rate of interest as a solely monetary phenomenon?

4. Explain the role of the Central bank as the lender of the last resort.

-.......-

INDIAN STATISTICAL INSTITUTE
Research and Training School

Mid-term examination 1963

B. Stat. IV Year

ECONOMICS - III

Duration: 3 hours | Maximum Marks: 100 | Date: 6 December 1963

Note: 20 marks are reserved for good performance in class work and take-home examinations.

1. With respect to the data supplied to you in Table 1 the following graphical analysis is expected.

   i) What distribution fits best to the data? You may experiment with the lognormal and Pareto.

   ii) Is there any indication that income distribution varies substantially from time to time? To what extent and in what way?

   iii) Is there any indication that the variation if any can be explained by the average level of income?

   iv) What additional data would you need in order to make a real comparison between the two distributions.

Table 1. Distribution of households in India by size classes of personal income : at curent prices.

annual household income (Rs.)	percentage of			
	1952-53		1956-57	
	households	income	households	income
(0)	(1)	(2)	(3)	(4)
less than 500	8.18	2.6	9.33	3.8
500 - 1000	31.37	16.9	38.03	23.6
1000 - 1500	27.61	25.2	29.94	28.9
1500 - 2000	13.94	17.8	13.62	19.7
2000 - 2500	12.20	18.5	4.05	6.9
2500 and over	6.70	19.0	5.03	17.1

(40)

(P.T.O.)

2(i). From the time-series data relating to Indian irons and steel industry during 1951-57, construct a Cobb-Douglas production function

$$x = A\, n^{\alpha} k^{\beta} u$$

where

x = net output (index)
n = labour input (")
k = capital input (")
u = random disturbance.

Table 2. Some data relating to Indian iron and steel industry (1951 - 57); base 1951 = 100

year	net output (x)	labour (n)	capital (k)
1951	100.00	100.00	100.00
1952	103.02	97.51	97.21
1953	97.05	95.14	110.92
1954	94.23	107.37	124.02
1955	94.01	109.76	137.44
1956	98.04	111.13	165.05
1957	95.06	111.31	209.72

(ii) Also, test the hypothesis of constant returns to scale against that of decreasing returns in the industry. (40)

INDIAN STATISTICAL INSTITUTE
Research and Training School

Mid-term examination, 1963

B.Stat. IV Year

DEMOGRAPHY

Duration: 3 hours | Date: 6 December 1963

Answer any THREE

1. Either,

i) Define 'underlying cause of death' (WHO) and discuss its importance in the field of public health.

ii) What is the necessity of standardizing death rates? Indicate the methods.

Or,

What are the items of information usually collected in a population census? Briefly indicate their significance in demographic studies.

2. Explain what is meant by the following terms :

(a) Specific fertility rate;
(b) Complete expectation of life at age x;
(c) Central rate of mortality;
(d) Net reproduction rate;
(e) General fertility rate.

Give your opinion about the validity of the following statement with reasons.

"In a certain country the net reproduction rate (N.R.R.) has been less than unity for some years past, but the population has been increasing. The N.R.R. is not therefore a suitable index of population growth".

3. The following table shows the growth of a bacterial colony as observed in a certain locality in square centimeters. Fit a logistic curve to the data. Estimate the area after a week.

Age of colony (in days)	Area in square centimeters
0	0.24
1	2.78
2	13.53
3	36.30
4	47.50
5	49.50

....

-.......-

INDIAN STATISTICAL INSTITUTE
Research and Training School
Periodical Examination
B. Stat. IV year
PROBABILITY

Duration : 2½ hour test — Date : 17 Feb., 1964.

Maximum marks : 100

1. Let X be a finite set having k elements $\{x_1, \ldots, x_k\}$. Let X_∞ be the space of all sequences of elements of X. Let C_1, C_2, ... be cylinder sets in X_∞ such that $C_1 \supseteq C_2 \supseteq \ldots$. If $\bigcap_n C_n = \emptyset$, prove that $C_n = \emptyset$ after a certain stage.

2. Let X be the set $\{0, 1\}$ and X_∞ mean as usual. For $\xi = (y_1, y_2, \ldots) \in X_\infty$ define

$$f(\xi) = \sum_1^\infty \frac{y_n}{2^n}.$$

Consider the following sets in X_∞.

(a) $\{\xi : 0 \leq f(\xi) \leq \frac{1}{2}\}$ (b) $\{\xi : f(\xi) \geq \frac{1}{2}\}$.

(c) $\{\xi : \frac{1}{4} \leq f(\xi) \leq \frac{3}{4}\}$ (d) $\{\xi : f(\xi) = 1\}$

Find out these sets and discuss which of them are cylinder sets.

3. A number ξ is selected at random (i.e. having uniform distribution) between 0 and 1 and expanded as the infinite decimal

$$\xi = 0 \cdot y_1 \; y_2 \; \ldots \; y_n \ldots$$

Prove that $y_1, y_2, \ldots$ are independent and identically distributed random variables with

$$\Pr\{y_k = r\} = \frac{1}{10} \qquad r = 0, 1, \ldots 9$$

for all k = 1, 2, ...

--- --- ---

INDIAN STATISTICAL INSTITUTE
Research and Training School

Periodical Examination

B.Stat. IV Year

STATISTICAL INFERENCE

Duration $2\frac{1}{2}$ hours | Maximum Marks: 100 | Date: 23 March 1964

Separate answer-book should be used for each group.

GROUP A

1(a). Let $f(x, \theta_1 \ldots \theta_k)$ be the frequency function of random variable x, where $\theta_1, \ldots, \theta_k$ are unknown parameters. Let $x_1, \ldots x_n$ be a random sample drawn from the population.

(i) Define the information matrix of the sample $x_1, \ldots, x_n$ [5]

(ii) Let $t_1(x_1, \ldots, x_n), \ldots, t_k(x_1, \ldots x_n)$ be the estimators of $\theta_1, \ldots, \theta_k$ respectively. When do you say that these estimators are (a) unbiased, (b) jointly sufficient, (c) jointly efficient. [15]

(iii) Explain what you mean by minimal set of sufficient statistics. [5]

2(a). Examine whether the following statements are true.

(i) If $t(x_1, \ldots x_n), \ldots t_k(x_1, \ldots, x_n)$ are jointly sufficient for $\theta_1, \ldots, \theta_k$ in Question then $t_1(x_1, \ldots x_n)$ is sufficient for θ_1. [7]

(ii) The method of maximum likelihood estimation always provides minimum variance unbiased estimates. [7]

(b). Write down the asymptotic distribution of the maximum likelihood for a single parameter, stating the underlying regularity assumptions. [11]

(Please go on to the next page)

GROUP B

3. $X_1, \ldots, X_n$ are independent random variables with common density function

$$e^{-(x-\theta)}, \quad x \geqslant \theta .$$

(a) Show that $Y_i = e^{-X_i}$ has uniform distribution on $(0, e^{-\theta})$.

(b) Find a uniformly most powerful test of level α for $H_0 : \theta = \theta_0$ against alternatives $\theta < \theta_0$.

4. State the generalized Neyman - Pearson lemma.

5. $X_1, \ldots, X_n$ are independent normal variates with mean 0 and variance σ^2. Show that the uniformly most powerful unbiased test of level α for $H_0 : \sigma^2 = 1$ against alternatives $\sigma^2 \neq 1$ accepts H_0 if and only if

$$C_1 < \sum_1^n X_i^2 < C_2$$

where C_1 and C_2 satisfy

$$\int_{C_1/2}^{C_2/2} \frac{1}{\Gamma(\frac{n}{2})} e^{-u} u^{\frac{n}{2}-1} du = 1 - \alpha = \int_{C_1/2}^{C_2/2} \frac{1}{\Gamma(\frac{n}{2}+1)} e^{-u} u^{\frac{n}{2}} du$$

ooooooooooo

INDIAN STATISTICAL INSTITUTE
Research and Training School
Periodical Examination
B.Stat. IV Year

STATISTICAL METHODS AND SAMPLE SURVEYS

10.2.64. Time: $2\frac{1}{2}$ hrs. Max. marks. 100

(Answer as many as you can.)

1. For estimating the total number of fruits in all the 1000 trees in a garden, the gardener goes to each tree and notes down his eye estimate (x) of the number of fruits. The owner of the garden wants to get a more accurate estimate. He choses a simple random sample (without replacement) of 100 trees and gets the exact number of fruits (y) in each of these trees by plucking out the fruits. Suggest a method of estimation of the total number of fruits, which makes use of the available informations. Justify your method. How can the owner get an idea of the accuracy of the estimate from the sample itself. [20]

2. A population is divided into two strata of sizes N_1 and N_2 respectively. A simple random sample of size n_i is taken from the ith stratum and observations are made on a characteristic Y.

   a) Suggest an unbiased estimate of the population total of Y and find out the variance of the estimate.

   b) What will be the size of the sample taken from the i-th stratum when the total sample size is $n_1 + n_2$, under optimal allocation.

   c) If R is the ratio of n_1 / n_2 to n_1' / n_2', where n_i' is the optimum i-th stratum sample size as obtained in (b), then show that the relative precision of the above allocation to the optimum allocation is never less than $4R/(1+R)^2$.

   [Hint. Express the relative precision as a function of R and $\dfrac{N_1^2 S_1^2}{N_2^2 S_2^2}$ and try to find the minimum value of the same w.r.t. $\dfrac{N_1^2 S_1^2}{N_2^2 S_2^2}$] [8+8+1[illegible]

3.a) What type of stratification will your suggest in designing a sample survey to study incomes of house hold in Calcutta City.

   b) In stratified sampling, what are the considerations to be taken into account in forming the strata.

   c) 'Stratification usually yields greater accuracy per unit and often reduces cost per unit as well'. Explain.

   d) 'Proportional sampling is usually more convenient than Neyman's optimum sampling, but sometimes Neyman's allocation is more advantageous'. Explain. [8+8+8+8]

4. To study the earnings of daily wage workers in all factories in Calcutta which of the two following procedures do your recommend? Why?

   i) Stratify the factories by size of workers and value of total output and use a stratified random sample.

   ii) Choose a pps sample of factories with probability proportional to the number of workers.

   We are given a list of the factories and the total number of daily wage workers in the factory and the value of the total daily output. [10]

P.T.O.

5.a) 'Sample should be used only when it is impossible to get a complete count of the population. Give your comments.

b) 'A sample is never so satisfactory as complete enumeration'. Comment.

c) What are the advantages of sample survey over complete enumeration. [ ]

6.a) Show that the variance of the r-th raw moment is given by

$$V(m_r') = \frac{1}{n}\left[\mu_{2r}' - \mu_r'^2\right]$$

Assuming $\mu_1' = 0$, hence deduce that the variance of the r-th central moment is given approximately by

$$V(m_r) = \frac{1}{n}\left[\mu_{2r} - \mu_r^2 + r^2\mu_2\mu_{r-1}^2 - 2r\mu_{r-1}\mu_{r+1}\right]$$ [ ]

b) On the basis of a sample of size n drawn from any given population with mean μ and s.d. σ, indicate the procedure you will adopt to test the following:

i) $\mu = 3\mu_0$ when σ is unknown.

ii) $\sigma = \sigma_0$.

What will happen in the case when the population is normal? [ ]

INDIAN STATISTICAL INSTITUTE
Research and Training School

Periodical Examination

B. Stat. IV Year
DESIGN OF EXPERIMENTS

Duration: 3 hours | Full Marks: 100 | Date: 2 March 1964

Attempt ALL the questions. Figures in the margin indicate full marks.

1. Either, Write a note to explain with examples the roles of the techniques of randomisation, replication and error control in planning a scientific experiment.

   Or, Explain the concepts of main effects and interaction with reference to an experiment involving 2 factors each at 2 levels. Why is a factorial type of experiment preferred to a one-factor-at-a-time type of experiment? (20)

2. Explain the use of the technique of analysis of covariance in increasing the precision of an experiment. Show the structure of the table and the computational details for an experiment in r randomised blocks involving t treatments and one auxiliary variable. (20)

3. What is a balanced incomplete block design? In what situations are they useful? How will you analyse the results of such an experiment? Give the layout of the (blank) analysis of variance table and the computational details. (20)

4. The following table gives the results of an experiment to compare the yields of 4 varieties of wheat M, S, V, T carried out in 6 randomised blocks with one plot missing.

   Analyse the data and give your comments.

	Varieties			
Blocks	M	S	V	T
1	81.0	105.4	119.7	*.
2	80.7	82.3	80.4	87.2
3	140.6	142.0	150.7	[illegible]
4	100.4	115.5	[illegible]	147.7
5	82.3	77.3	78.4	131.3
6	103.1	105.1	118.5	109.9

(40)

* missing.

oooooooo

INDIAN STATISTICAL INSTITUTE
Research and Training School

Periodical Examination

B. Stat. IV Year

ECONOMICS AND ECONOMETRICS

Duration : 2½ hours | Maximum Marks : 100 | Date: 13 April 1984

Separate answer-book should be used for each Group.

GROUP A

Answer any two.

1(a). Critically examine the effect of repayment of loans by government on national income. (15)

(b). State and prove the Haavelmo theorem. (10)

2(a). Show that an increase in government expenditure on goods and services, ceteris paribus, has a stronger expansionary effect on national income than an increase in transfer payments of the same amount. (10)

(b) Summarise the principal arguments for and against protection. (15)

3. Carefully examine the theory of comparative costs as applied to explain international trade. (25)

GROUP B

Answer any two

1. Give a brief account of the nature and uses of Leontief input-output models. Mention some of the major limitations of those models, and suggest improvements, if any. (25)

2. Outline the basic assumptions made by Milton Friedman in his permanent income hypothesis. If you agree with all his assumptions, indicate how you would proceed to estimate the overall 'true' marginal propensity to consume. (25)

3. In the linear, single-equation, errors-in-equation models show how you would estimate the individual influences of exogenous variables when the successive disturbances are unautocorrelated. How would you modify your method if the disturbances are autocorrelated and follow a simple linear autoregressive scheme? (25)

••••••••

INDIAN STATISTICAL INSTITUTE
Research and Training School

Periodical Examination

B. Stat. IV Year

EDUCATIONAL STATISTICS

Duration : 2½ hours | Maximum Marks: 100 | Date: 6 April 1964

Attempt as many questions as you can

1. What are the basic assumptions of Mental Test Theory? (10)

2. Define the following terms :

(a) Error of Measurement.
(b) Reliability.
(c) Validity.
(d) Parallel Tests.
(e) Speed Versus Power Tests. (20)

3. Examine whether the following statements are true or not.

(a) The sequare root of the difference between unity and reliability coefficient is equal to the correlation between observed scores and error scores.

(b) The true variance is equal to the reliability coefficient multiplied by the error variance. (16)

4. Find out the effect of increasing the test length k times on

i) Mean.
ii) Error variance.
iii) Reliability. (20)

5(a). Find out a function of test validity that is invariant with respect to changes in test length. (7)

(b). What is "correction for attenuation"? What does it signify? (7)

6(a). If the reliability of a test is raised from .60 to .90 by lengthening the test, a validity coefficient of .60 for this test would be expected to be increased to what value? (10)

(b)

Test	Mean	Standard deviation	Reliability
A	100	20	.81
B	200	40	.95

Estimate i) error of measurement of B.
ii) the correlation between true scores and observed scores of A. (10)

7. Describe two methods for eliminating the effects of guessing in a power test. (20)

INDIAN STATISTICAL INSTITUTE
Research and Training School

Annual Examination, 1964

B. STAT. IV Year

PROBABILITY

Duration : 3 hours   Maximum Marks : 100   Date: 25 May 1964

1(a). s_n is the number of successes in n Bernoulli trials with probability of success p $(0 < p < 1)$. Let $s_n^* = (s_n - np)/\sqrt{npq}$. Prove that there exists a constant C, not depending on n, such that

$$E\, s_n^{*4} < C$$

for all n. (5)

(b). Hence or otherwise prove that

$$\frac{s_n}{n} \longrightarrow p \qquad (15)$$

with probability one.

2(a). Let X be a finite set and let X_∞ be the space of all infinite sequences of elements from X. Define carefully the notion of a cylinder subset of X_∞. (5)

(b). Let $X = \{0, 1\}$ and for any $\omega = (a_1, a_2, \ldots\ldots)$ in X_∞ (i.e. each $a_j = 0$ or 1) write

$$f_n(\omega) = (a_1 + \ldots + a_n)/n$$

Prove that the set

$$\left\{\omega : f_n(\omega) \longrightarrow \frac{1}{2}\right\}$$

is a Borel set in X_∞ which is not a cylinder set. (10)

(c). Give an example of an ω in X_∞ for which $\lim\limits_{n \to \infty} f_n(\omega)$ does not exist. (5)

(Please turn over)

3. $\{a_n\}$ is a sequence of positive members such that $a_n \to +\infty$ but $a_n^3/\sqrt{n} \to 0$ as $n \to +\infty$. Let s_n be the number of heads in n independent tossings of an unbiassed coin and let $s_n^* = (s_n - \frac{1}{2}n) / \frac{1}{2}\sqrt{n}$. Assuming the normal approximation to the binomial, prove that

$$\Pr(s_n^* > a_n) \sim \frac{1}{\sqrt{2\pi}\, a_n} e^{-\frac{1}{2}a_n^2}$$

in the sense that the ratio of the two sides tends to 1 as $n \to \infty$. (25)

4. Suppose that in a ballot there are two candidates P and Q with P obtaining p and Q obtaining q of the votes. Assuming $p > q$, prove that the probability that P led throughout the counting is $(p-q)/(p+q)$. (15)

5(a). $A_1, A_2, \ldots$ are mutually independent events. Let $p_j = P_r(A_j)$. If $\sum_j p_j = \infty$, prove that, with probability one, infinitely many of the A_j's occur. Prove also that, if $\sum_j p_j < \infty$, with probability one, only finitely many of the A's occur. (8+8=16

(b). Is the independence assumption necessary in order to prove the second assertion in (a)? Give an example to illustrate your answer. (4)

....

INDIAN STATISTICAL INSTITUTE
Research and Training School
Annual Examination; 1964

B. Stat. IV Year

STATISTICAL INFERENCE

Duration : 3 hours | Maximum Marks: 100 | Date: 23 May 1964

Separate answer-book should be used for each Group.

GROUP A

Answer any three questions

1(a). State the asymptotic properties of the estimators obtained by the method of maximum likelihood mentioning the underlying assumptions. (10)

(b). $X_1, \ldots, X_n$ are independent Poisson variates with mean θ and $Y_1 \ldots Y_n$ are independent Poisson variates with mean 2θ. Find the maximum likelihood estimator of θ based on $X_1, \ldots, X_n$, $Y_1, \ldots, Y_n$. (6)

2(a). Define a Bayes solution for a general decision problem with respect to a given a priori distribution on the parameter space. (4)

(b). In the above context define the a posteriori risk of a decision procedure for a given sample. (4)

(c). Show that given an apriori distribution any decision procedure which minimises the a posteriori risk for each point in the sample space is a Bayes solution. (4)

(d). Show that if a decision procedure with constant risk is a Bayes solution with respect to some a priori distribution on the parameter space, then it is minimax. (4)

3(a). What are consistent estimators? (2)

(b). The probability distribution on a sample space involves an unknown parameter θ. $T_1, T_2, \ldots$ are statistics defined on this sample space.

If $\lim_{n \to \infty} E_\theta T_n = \theta$ and $\lim_{n \to \infty} Var_\theta T_n = 0$ for each θ show that T_n is a consistent estimator of θ. (5)

(c). Define the efficiency of an unbiased estimator. (3)

(d). If T_1 is an unbiased efficient estimator and T_2 is another unbiased estimator of efficiency e, show that the coefficient of correlation between T_1 and T_2 is $\sqrt{e}$. (6)

(Please turn over)

4(a). Define a sufficient statistic. (3)

(b). State and prove the factorisation theorem for sufficient statistics on a discrete sample space. (5)

(c). $X_1, \ldots, X_n$ are independent random variables with

$$P[X_i = 1] = \theta, \quad P[X_i = 0] = 1 - \theta$$

Show that $\sum_{i=1}^{n} X_i$ is sufficient for θ. (2)

(d). Show that $\sum_{i=1}^{n} X_i/n$ is the minimum variance unbiased estimator of θ. (6)

NEATNESS (2)

GROUP--B

Answer any THREE questions

5. $X_1, \ldots, X_n$ are independent random variables with common probability density function

$$f(x, \theta) = \frac{1}{\theta} e^{-x/\theta}, \quad x \geq 0$$
$$= 0, \quad x < 0,$$

$\theta > 0$. $0 < \theta_0 < \theta_1$ are given.

(a) Find the uniformly most powerful level α test for the null hypothesis $\theta = \theta_0$ against the alternative $\theta \geq \theta_1$. (8)

(b) Show that the power of this test is an increasing function of θ. (8)

6. $(X_1, Y_1), \ldots, (X_n, Y_n)$ are independent observations on a bivariate normal distribution with $EX = \xi$, $EY = \eta$, Var $X = \tau^2$, Var $Y = \gamma^2$ and Cov $(X, Y) = \rho$. Derive the uniformly most powerful unbiased level α test for the null hypothesis $\rho = 0$ against the alternative $\rho \neq 0$. (16).

(Please turn over)

7. 1, 0, 1, 0, 0 are independent observation on a Poisson population with mean μ and 0, 2, 1, 1, 2 are independent observations on another Poisson population with mean λ. Use the uniformly most powerful similar region test at level .10 to test the null hypothesis $\lambda = \mu$ against the alternative $\lambda > \mu$. (In case you cannot take a decision outright, give your rule for randomisation). (16)

8(a). $X_1, \ldots, X_m$, $Y_1, \ldots, Y_n$, $Z_1, \ldots, Z_r$ are independent normal variates with $EX_i = \xi$, $EY_i = \eta$, $EZ_i = \zeta$ and common variance σ^2. ξ, η, ζ and σ^2 are all unknown. Derive the likelihood ratio test for the null hypothesis $\xi = \eta = 2\zeta$. (10)

(b). Explain how tests of level α for simple hypotheses involving an unknown parameter can be inverted to obtain confidence sets of confidence coefficient $1 - \gamma$ for the same parameter. (6)

NEATNESS (2)

....

INDIAN STATISTICAL INSTITUTE
Research and Training School

Annual Examination, 1964

B. Stat. IV Year

STATISTICAL METHODS (Theory)

Duration : 3 hours | Maximum Marks: 100 | Date: 21 May 1964

(Answer as many questions as you can)

1(a). Explain the role of stand error in large sample tests of significance?

(b). What is the difference between large sample tests and small sample tests?

(c). What is the standard error of sample coefficient of variation? Sketch the derivation. (5+5+8)

2(a). Show that the sample quantile z_p is asymptotically normal with mean h_p (the corresponding population quantile) and S.D. $\frac{1}{f(z_p)}\sqrt{\frac{pq}{n}}$, clearly stating the assumptions which are made.

(b). Hence or otherwise obtain the standard error of the sample median for a normal population? What is its asymptotic effiency compared to the mean? (10+8)

3(a). Derive the distribution of the sample range, given n independent observations with p.d.f. $f(x)\,dx$.

(b). Deduce the explicit formula for the distribution of the range when $f(x) = \frac{1}{\sqrt{2}} e^{-\frac{1}{2}x^2}$ and n=2. (8+7)

4. Write short notes on each of the following :

(a) Bartlett's test for equality of variances.

(b) Tolerance limits

(c) Run test

(d) Uses of orthogonal polynomials in regression analysis. (6+6+6+6)

(Please turn over)

5(a). Define multiple and partial correlation coefficients. Explain their physical significance.

(b). How will you test on the basis of a sample of size n on p associated variates that the p^{th} variate is useless in predicting x_1 when $x_2, \ldots, x_{p-1}$ are already available? (9+6)

6. Explain the concept of interaction in analysis of variance. (Consider a two-way classification with multiple and equal no. of observations each cell). (9)

7. Give two examples of use of statistical methods in each of the following branches of science :

a) Biology ; b) Physics and c) Chemistry. (9)

****

INDIAN STATISTICAL INSTITUTE
Research and Training School
Annual Examination, 1964

B. STAT. IV Year

STATISTICAL METHODS (Practical)

Duration : 3 hours | Maximum Marks: 100 | Date: 22 May 1964

Answer as many questions as you can

1. The correlation coefficients between the scores in two halves of a psychological test applied on seven groups of 30, 20, 25, 40, 45, 35, 50 students were .63, .48, .71, .65, .57, .39, .51. Examine whether the groups are different in respect of the correlation coefficient. (30)

2. The following data relate to the heart weight in grams of 12 female and 15 male cats.

Heart weight in Grams

Males : 12.7, 15.6, 9.1, 7.6, 12.8, 8.3, 11.2, 9.4, 8.0, 14.9, 10.7, 13.6, 9.6, 11.7, 9.3

Females: 7.4, 7.3, 7.1, 9.0, 7.6, 9.5, 10.1, 10.2, 10.1, 9.5, 8.7, 7.2

Examine whether the heart weight of males and females follow the same distribution. [Use large sample median test]. (20)

3. The following table gives the test scores made by 10 salesmen on Intelligence test and their weekly sales.

Salesmen	Test scores	Sales (in hundreds of Rs.)
1	40	25
2	70	60
3	50	45
4	60	50
5	80	45
6	50	20
7	90	55
8	40	30
9	60	45
10	60	30

(a) Examine whether there is any correlation between the two series. (You need not calculate the correlation coefficient).

(b) Obtain the regression equation of weekly sales on the test scores.

(c) A new salesman makes a score of 70. Estimate his weekly sale.

(d) What is the standard error of the estimate in (c)? (35)

(Please turn over)

4. A study of the market for various commodities among 8000 readers of Collier's magazine reveal the following distribution of sample households by size of household as compared with the corresponding census estimates for all U.S. households.

Table showing the distributions of 8000 Collier families and all U.S. families by size of household.

Persons in household	Collier sample	U. S. Familes (percent)
1	576	10.0
2	2368	29.8
3	1968	24.2
4	1592	18.0
5	800	10.0
6	376	4.5
7	152	1.7
8 or more	168	1.8
Total	8000	100.00

Examine whether the Collier sample provides a representative picture of the distribution of the size of all U.S. households. (25)

•••••••

INDIAN STATISTICAL INSTITUTE
Research and Training School

Annual Examination, 1964

B. Stat. IV Year

SAMPLE SURVEYS (Theory & Practical)

Duration: 3 hours | Maximum Marks: 100 | Date: 20 May 1964

Answer as many questions as you can.
Answers must be brief and to the point.

1(a) What are the principal steps involved in planning and execution of sample surveys?

(b) What are the various fields of applications of sample surveys? Give examples. (8+7)

2. Write down the Horvitz and Thomson estimator for estimating the population total from a pps sample without replacement of size 2. Show that the estimator is unbiased. Obtain the variance of the estimate and an unbiased estimate of the variance. (4+6+9+

3. A finite population consists of N first stage units and the ith first stage unit consists of M_i second stage units. A simple random sample (without replacement) of n first stage units are taken. Simple random samples (with replacement) of second stage units are taken from the selected first stage units. (If ith first stage unit was one among the n selected, then m_i second stage units are included in the second stage sample). Suggest an unbiased estimate of the population total (You have to show that the estimate is unbiased). Derive the variance of the estimate. Also obtain a suitable of the variance. (8+10+7)

4. With the help of necessary mathematical derivations explain the use of double sampling (i) for stratification and (ii) for regression method of estimation (only one concomitant variable). [In both cases only the estimate of the population total and variance of the estimate are to be worked out]. (12+12)

5. A population of 112 villages was divided into 3 strata. From the first stratum which consists of 51 villages a simple random sample (without replacement) of 8 villages were taken. From the second stratum, which consists of 38 villages, 5 villages are chosen with probability proportional to the total cultivated area X (with replacement). From the third stratum which consists of 23 villages, two independent circular systematic samples of 4 villages each are taken. For each selected village the total area under wheat (in acres), y, was observed. The observed values are given below. Estimate the total area under wheat in each stratum separately and also in all the 3 strata together. Obtain suitable estimates of the variances of the estimates.

(Please turn over)

Stratum 1

y (in acres)	: 75, 101, 5, 78, 78, 45

Stratum 2

y (in acres)	: 247, 238, 359, 120, 223
x (in acres)	: 729, 617, 870, 305, 569

Total cultivated area for all the villages in the stratum = 29612 acres

Stratum 3

Subsample 1, y : 427, 326, 481, 445
(in acres)

Subsample 2, y : 335, 412, 503, 348
(in acres) (12+24)

ooooo

INDIAN STATISTICAL INSTITUTE
Research and Training School
Annual Examination, 1964

B. Stat. IV Year

ECONOMICS -I (Economic Theory)

Duration : 3 hours | Maximum Marks : 100 | Date: 18 May 1964

Attempt any FIVE questions. All questions are of equal value.

1. "We have to consider whether a change in the quantity of money leads to a change in effective demand. The second question is whether a change in effective demand brings about a change in prices." - Examine the Quantity Theory in the light of this statement. (20)

2. Classify the different types of inflation and explain their characteristics. (20)

3. "The supply of and demand for money-holding determine the rate of interest."

"Business demand and household supply of earning assets are the determinants of the rate of interest."

Develop the ideas contained in the two statements and discuss whether they are essentially the same. (20)

4. Discuss whether discount policy by itself can guarantee the Central Bank's control over the lending potential of commercial banks. (20)

5. Analyse the effects of a devaluation of a country's currency on its balance of payments. (20)

6. Discuss the statement that fiscal policy cannot exist independently of monetary policy but must be integrated with it. (20)

7(a). Show that, in a closed economy, when net taxes (taxes less transfers) are a rising function of national income at market prices, there is a tendency towards automatic equalisation of aggregate income and aggregate expenditure.

(b). In an economy the planned consumption expenditure of the community is the fraction b of its private disposable income. Its net taxes always bear the same ratio to its national income at market prices.

Assuming private investment expenditure to remain constant, show that the amount of budget deficit to be incurred by government in order to double national income at market prices is $Y_o(1-b)$ where Y_o is the initial equilibrium level of national income. (10+10)

8(a). Explain the interest-effect of open market operations.

(b). Discuss whether the rate of interest is solely a monetary phenomenon. (10+10)

••••••••

ad.

INDIAN STATISTICAL INSTITUTE
Research and Training School

Annual Examination, 1964

B. Stat. IV Year

ECONOMICS II :

Duration: 3 hours | Maximum Marks: 100 | Date: 18 May 1964

Separate Answer-book should be used for each Group.

GROUP – A (Economic Statistics)

Answer any THREE questions

1. Define the concentration curve for a size distribution, and state its properties. Why is the 'area of concentration'(x2) used as a measure of inequality? Prove the relation connecting this measure with the mean difference due to Gini.

2. Prove the moment distribution property of the lognormal distribution. Derive the expressions for the concentration curve and the concentration coefficient for the lognormal distribution.

3. Suppose you want to estimate the demand function for a commodity like coffee using time-series data. Explain clearly the nature of the different series you will collect for the purpose. Indicate the fitting of the constant elasticity demand function.

4(a). Define the specific concentration curve for consumption of a particular commodity and explain its different uses.

(b). How is the Engel curve used for forecasting future demand of a consumption item? What assumptions underlie such forecasts?

5. State briefly the properties of the Cobb-Doughlas production function. How do you measure labour, capital and output when fitting this function to cross-section data? How would you test whether the returns to scale are constant?

GROUP – B (Econometrics Theory)

Answer any THREE questions

6. Briefly describe an input-output table farxanxopen-outputxtable for an open economy. State the assumptions of Leontief's static open model and deduce the price conditions of production for such a system.

7. State clearly the assumption underlying the use of least squares regression methods. Illustrate the limitations of such assumptions in econometric analysis.

(Please turn over)

8. Consider the following model where X is exogenous :

$$Y_t = \alpha + \beta X_t + u_t$$

$$\left.\begin{array}{l} E(u_t) = 0 \\ E(u_t . u_{t'}) = \sigma_{tt'} \end{array}\right\} \begin{array}{l} \text{for all} \\ t, t' = 1, 2, \ldots, n \end{array}$$

Obtain generalised least squares estimates of α and β and show that these estimates are best linear unbiased.

9. Discuss fully the concept of multicollicollinearity in the estimation of economic relationships. For what purposes does multicollinearity not pose a serious problem?

10. Explain identification problem in simultaneous - equation models. Examine the identifiability of the parameters in the following model of a competitive market for an agricultural product.

$$q^d + \alpha p = u \text{ (demand)}$$

$$q^s + \gamma p + \delta r = v \text{ (supply)}$$

$$q^s - q^d = 0.$$

u and v follow bivariate normal distribution with zero means,

where p = price

q^d = quantity of demand

q^s = quantity of supply

r = rainfall during a critical period

Obtain the formulae for estimating the identifiable parameter(s)

oooooo

INDIAN STATISTICAL INSTITUTE
Research and Training School

Annual Examination, 1964

B. Stat. IV Year

ECONOMICS III

Duration : 3 hours | Maximum Marks: 100 | Date: 10 May 1964

Separate Answer-book should be used for each Group

GROUP A - Economic Statistics Practical

1) Answer any TWO.

2) Use of tables and computing machines permitted.

The following table is based on a family budget enquiry covering rural India.

per capita monthly exp. on all items, (Rs.)	percentage of persons	per capita monthly exp. (Rs.) on all items	on clothing
(1)	(2)	(3)	(4)
0 - 8	15.5	6.2	0.2
8 - 11	17.8	9.6	0.5
11 - 13	12.0	11.9	0.8
13 - 15	10.3	14.0	1.1
15 - 18	10.8	16.2	1.3
18 - 24	15.7	20.6	2.3
24 - 34	10.6	28.3	3.6
34 -	7.3	50.3	6.1

1. Plotting the ogive on log-probit scale examine whether the lognormal distribution should be fitted to the observed distribution of persons by per capita monthly expenditure on all consumer items. [Use the information in cols.(1) and (2) only].

2. Using cols.(3) and (4) of the table above, estimate the parameters of the constant elasticity Engel curve for clothing. (Unweighted least squares may be used).

3. Using cols. (2) - (4) of the same table, obtain the concentration coefficient for total expenditure as well as the specific concentration coefficient for expenditure on clothing.

(Please turn over)

GROUP B - Econometrics Practical

4. Consider the model

$$c_t = \alpha y_t + \beta + u_t$$

$$y_t = c_t + z_t$$

where c = consumption expenditures
y = disposable income
z = investment expenditures
u = random variable

From the following table estimate the parameters (α and β) of the consumption function by the method of (i) ordinary least squares, and (ii) by some suitable simultaneous - equation method.

TABLE 1 *

year	y	c	z
1927	498	447	51
1928	511	466	45
1929	534	474	60
1930	478	439	39
1931	440	399	41
1932	372	350	22
1933	381	364	17
1934	419	392	27
1935	449	416	33
1936	511	463	48
1937	520	469	51
1938	477	444	33
1939	517	471	46
1940	548	494	54
1941	629	529	100
total	7284	6617	667

* Figures are in dollars per capita, deflated. (30)

5. The following table shows the personal disposable income and personal consumption in the U.S. from 1948 to 1957 in constant 1954 dollars. On the assumption of the following model work out a prediction formula for consumption and predict the consumption level for 1958 on the assumption of an income figure of \$300 billion for the year.

Model :

$$c_t = \alpha + \beta y_t + u_t$$

$$u_t = \rho u_{t-1} + \epsilon_t$$

$$\left.\begin{aligned} E(\epsilon_t) &= 0 \\ E(\epsilon_t \cdot \epsilon_{t+s}) &= \sigma_t^2 \quad \text{if } s = 0 \\ &= 0 \quad \text{if } s \neq 0 \end{aligned}\right\} \text{ for all } t.$$

The value of ρ is assumed to be 0.457.

(Please turn over)

TABLE 2

Personal disposable income and personal consumption in the United States, 1948-1957)

(Billions , constant 1954 dollars)

year	consumption c	income y
1948	199	212
1949	204	214
1950	216	231
1951	218	237
1952	224	244
1953	235	255
1954	238	257
1955	256	273
1956	264	284
1957	270	290
total	2324	2497

(20)

....

INDIAN STATISTICAL INSTITUTE
Research and Training School

Annual Examination, 1964

B. Stat. IV Year

EDUCATIONAL STATISTICS (Theory and Practical)

Duration: 1½ hours | Maximum Marks: 100 | Date: 19 May 1964

Answer all questions

1.

Test	Mean	Standard deviation	Number of items	Reliability	Validity
A	16.5	4.4	30	.72	.68
B	12.6	3.5	20	.77	.60
C	53.2	10.7	100	.88	.68

Criterion reliability = .70

(a) If the test A is lengthened to a 100-item test what would you expect the new mean, standard deviation, reliability and validity to be? (Assume that the criterion test has not been altered).

(b) If test B is lengthened to increase its reliability to .90, how many new items will be needed? What will be the new validity be, assuming that the criterion test remains unchanged?

(c) Give the true and error variances of test C. Estimate them when C is increased to 300 items. (50)

2. Define four different types of error in test theory and find out the error of measurement in each case. (15)

3(a) Describe two methods of Standardisation of test scores, one linear and the other non-linear. (15)

(b) Prove that the correlation between true and observed scores for a test of double length is $\frac{2r}{(1+r)}$ where r is the reliability of the original test. List the assumptions used in deriving the formula. (10)

(c) Prove that if a test of n items is a subset of a test with m items $(n<m)$ the correlation $r_{n,m}$ is

$$r_{n,m} = \sqrt{\frac{\frac{1-r}{m}+r}{\frac{1-r}{n}+r}} \qquad (10)$$

where r is the reliability of a unit test.

******

INDIAN STATISTICAL INSTITUTE
Research and Training School

Annual Examination, 1964

B. Stat. IV Year

GENETICS (Theory and Practical)

Duration : 1½ hours | Maximum Marks: 100 | Date 19 May 1964

Answer all questions

1. Prove that a population which is under random mating with respect to an autosomal locus, attains equilibrium with respect to that locus, in one generation. Assume that the locus can contain one of two allelomorphs A and 'α' of a gene, and that 'A' is dominant over 'α'. (15)

2. Colour-blindedness is a sex linked character in human beings, controlled by two alleles and α of a gene, of which A is dominant over α, and recessives being colour-blind. In a sample of 2000 unrelated women, 72 are found to be colour-blind. If a sample of 300 unrelated men is taken, what is the expected number of colour blinded males in this sample? What is the variance of the number of colour blinded males in the sample? (20)

3. In three crosses of AA plants with $\alpha\alpha$ plants, the following results are obtained in the F_2-generation.

cross no.	number of observed/recessives i.e. $\alpha\alpha$'s.	Total number of progeny
1	62	271
2	77	319
3	133	512

Do the data support Mendel's first law? Are the data on three crosses homogeneous among themselves? (25)

4. In an experiment conducted to detect linkage between two factors the progeny of the cross AABB X $\alpha\alpha\beta\beta$ are backcrossed with the double recessive parent of the 417 of this progeny, the frequencies of the four phenotypes observed are as follows :

AB	152
Aβ	79
αB	44
and $\alpha\beta$	142

Test for linkage between A and B. If it is found that there is linkage, estimate the recombination fraction 'p' between the loci A and B, and test whether the data are in agreement with the genetic theory. (40)

oooooo

CENTRAL STATISTICAL ORGANISATION

Course for M.Stat. and B.Stat. Students
of the Indian Statistical Institute, 1963

Official Statistics
and Related Methodology

Time : 3 hours

Answer any five, not less than 2 from each part

A

1. Describe any one:

   (1) The U.N.Statistical System
   (2) The Statistical System in any one foreign country
   (3) The Indian Statistical System

2. In any scheme of data collection, explain the items that will engage your attention when you develop the plan of the survey,

OR

Write a note on the International Standard Industrial Classification of all Economic Activities.

3. Describe the salient feature of the 1961 Population Census of India,

OR

Write a note on the use of sampling methods in connection with the Population Census.

4. What are the recent improvements brought about in the sphere of-

   (1) Land-utilisation Statistics
   (2) Production Statistics

OR

What are the crop surveys in vogue in India at the present time? Describe any one of these to show that the sample survey method is being successfully employed in this sphere.

5. Give an account of the Annual Survey of Industries currently in vogue in India,

OR

Describe the present Indian Index of Industrial Production.

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

   1) Labour Statistics in India,
   2) Transport Statistics in India,
   3) Monetary and Banking Statistics in India,
   4) Main features of the quatum index for imports or exports,
   5) Main features of the index of whole-sale prices in India,
   6) The All-India consumer price Index,
   7) Mixed cropping is no impediment in the estimation of production of any of the components,
   8) The scheme for improvement of market intelligence,
   9) Technical aspects of the yield surveys in India (without proofs).

B

1. Distinguish between :

(1) The comparative Mortality Figure and the Standardised Mortality Ratio.

(2) The Total Fertility Rate, the gross reproduction rate and the net reproduction rate.

2. Explain how L_x obtains two interpretations, the one leading to the concept of Expectation of life at age x and the other, to the method of construction of the Life Table.

3(a) Explain by means of an example from the Indian Index Numbers, how the Chain-Base method can arise as part of the Fixed Base method.

(b) What is the circular test? What are the two conditions which are sufficient for the test being satisfied? Show that one aggregative index satisfies these two conditions and the circular test.

4. How would you use an aggregate econometric model, an inter-industry model and linear programming model for national economic planning? Indicate in broad terms the nature of these models, the data requirements thereof and the important computational steps involved in the use of these models.

5. Discuss the advantages and disadvantages of using various methods for the estimation of parameters in a simultaneous equations system where some of the equations contain a variable denoting random disturbance.

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

(a) Comparative Ratios

(b) Explosive, regular and damped oscillations in a cobweb model

(c) Identification

(d) Mixed strategy

(e) Traffic intensity

(f) Minimax

(g) Laspeyre and Paasche formulae for an index

(h) Errors in Index Number

(i) Stationary Population.

SPECIAL ADMISSION TEST

FOR M. STAT.

SUBJECT : STATISTICS

Duration: 3 hours.   Maximum marks: 100   Date: 15 July, 1964

Name ______________________________

Directions

1. Answers must be brief and to the point.
2. All answers must be written in the spaces provided for the purpose.
3. All scratchwork must be done on the test booklet.

1. The following computations have been made from a set of 10 observations on height (in cm.).

Mean = 160 cm.

Standard deviation = 2 cm.

Show that the data cannot contain a height observation of 170 cm. [4]

2. Consider two random variables X and Y.

i) Show that if $Y - 5 + 10X$ has mean value zero and is uncorrelated with X then the best fitting linear regression of Y on X is given by $Y = 5 - 10X$ [4]

ii) If further, the regression of X on Y is known to be $X = -15 - .009Y$, compute the product-moment correlation coefficient between X and Y [4]

3. The correlation coefficient between heights of brothers is r and the correlation coefficient between heights of first cousins is r'. There are two families, the offspring of two brothers, each containing two sons whose heights are X_1 and X_2 for the first family and X_3 and X_4 for the second. The height distribution for the generation has mean $\bar{X}$ and variance s^2. (These provide the mean and variance for each of X_1, X_2, X_3 and X_4).

a) Write down the normal equations for the multiple regression of X_1 on X_2, X_3 and X_4. [8]

3. Continued

b) If $b_{13.24} = b_{14.23} = \frac{r'(1 - r)}{1 + r - 2r'^2}$, find $b_{12.34}$ [5]

4. 2 successes were observed in 10 independent Bernoulli trials with probability π of success in each trial.

a) Estimate π. [2]

b) Obtain an unbiased estimate of the variance of the estimator used in (a). [5]

5. A population containing 1000 individuals was divided into two strata, A (size 250) and B (size 750). The following figures were computed from a stratified random sample (with replacement) from this population.

	stratum	
	A	B
sample size	5	15
mean	10	20
variance	4	5

a) Estimate the population mean. [4]

b) Estimate the variance of the estimator used in (a). [5]

c) Estimate the variance of the sample mean in a random sample of size 20 drawn from the same population with replacement but without using the stratification. [8]

6. Suppose (X, Y) follow a bivariate normal distribution:

a) If λ is a constant, express the mean and variance of $Y - \lambda X$ in terms of $E(X)$, $E(Y)$, Var (X), Var (Y) and Cov (X, Y). [5]

b) What is the distribution of $Y - \lambda X$? [3]

c) If $\lambda = E(Y) \div E(X)$, test if $\lambda = 2$ using a sample of n pairs of observations (X_i, Y_i), $i = 1, \ldots, n$ on (X, Y). [5]

d) Use the data to obtain a 95 percent confidence interval on the ratio of the mean values, $E(Y) \div E(X)$. [8]

7. X is a random variable that takes values in the range $[-1, +2]$. Its probability distribution has density function (fr. f) $k.x^2$ on the interval $[-1, +2]$; of course, outside this interval, the density function is zero.

a) Determine the value of k. [3]

/ continued

b) What is the range of values of the random variable X^2 ? [3]

c) If α is any real number, determine

$$\Pr\left\{X^2 \leq \alpha\right\} .$$ [6]

d) Hence obtain the density function of X^2. [3]

8. X and Y are two random variables X takes the values 1, 2, 3, 4 with probability 1/4 each. Y takes the values 1, 2, 3 with probability 1/3 each. The correlation coefficient between X and Y is $+\frac{1}{10}$. Examine whether this information is sufficient to determine.

a) the joint distribution of X and Y. [5]

b) E (XY) [5]

c) $E(X^2 Y)$? [5]

P.T.O.

Give reasons for your answer. Actually determine them (if the given information is sufficient).