390

INDIAN STATISTICAL INSTITUTE

QUESTION PAPERS

for

Statistician's Diploma Examinations

April & September 1960

Price Rupee one on

INDIAN STATISTICAL INSTITUTE

STATISTICIAN'S DIPLOMA EXAMINATION, APRIL 1960

PAPER I : THEORETICAL STATISTICS, GENERAL

Time : 4 Hours   Full marks : 100

(a) Attempt *any five* questions.
(b) All questions carry equal marks.
(c) Use of calculating machines is not permitted.

(a) Discuss briefly the usefulness of the following descriptive measures in statistics:-
(i) Mean, Median and Mode
(ii) Mean deviation, Standard deviation and range.

(b) Show that for any distribution the standard deviation cannot be less than mean deviation about the median.

2. (a) A random variable x has the density function given below :

$$dF = K\, e^{-2x^2+10x}\, dx$$

(i) Find the constant K and the mean and the variance of the random variable x.

(ii) Find the upper 5 percent point of the distribution of the mean of a random sample size 25 from the above population.

(b) Let x_1, x_2 and x_3 be three independent observations from the above population. nd (i) the correlation coefficient between x_1+x_2 and x_1+x_3, (ii) $E(e^{x_1})$, (iii) E(minimum x_1, x_2)).

3. (a) Comment on the utility of Multiple and Partial correlation coefficients to scribe the relationship of x_1 to x_2 and x_3 where x_1, x_2 and x_3 are three variables supposed be related statistically to one another.

(b) Let the regressions of x_1 on x_2 and x_3, and of x_2 on x_1 and x_3 be given by

$$x_1 = b_{12\cdot3}\, x_2 + b_{13\cdot2}\, x_3$$

$$x_2 = b_{21\cdot3}\, x_1 + b_{23\cdot1}\, x_3$$

ow that the partial correlation coefficient between x_1 and x_2 eliminating x_3 is given by

$$r_{12\cdot3} = \sqrt{b_{12\cdot3}\, b_{21\cdot3}}$$

(c) Suppose the total correlation coefficients between three variables x_1, x_2 and are given by the following:

$r_{13} = \rho$, $r_{12} = 0.8$ and $r_{23} = 0.5$ show that
$-0.12 < \rho < 0.92$.

4. Let $(x_1 y_1) \ldots (x_n, y_n)$ be a sample of values of x and y which are distributed in the Bivariate normal form with equal variances σ^2 and the correlation ρ. Show that $u =$

$x+y$, and $v = x-y$ are independent Normal variates. Hence derive the distributi
of the ratio

$$w = \frac{1-\rho}{1+\rho} \cdot \frac{s_u^2}{s_v^2} \text{ where } s_u^2 = \sum_{j=1}^{n} (u_j - \bar{u})^2$$

$$s_v^2 = \sum_{j=1}^{n} (v_j - \bar{v})^2$$

5. Derive expressions for the probability of occurrence of (i) K out of m eve
(ii) at least K out of m events.

A lift containing m passengers leaves the ground floor of a building and serv
floors. If each passenger is equally likely to alight at any floor, what is the probab
that there will be just 't' floors at which none of the passengers alight ?

6. Expalin the technique of 'Analysis of Variance' and illustrate your answer
analysing the total variation of K classes of data into variance 'between' and 'within' cla

What are the principal assumptions underlying Analysis of Variance ? Expla
the consequences when one or more of these assumptions are not satisfied in the actu
data and how an improved analysis can be obtained under these circumstances.

7. *Either*,

(a) Derive any interpolation formula using divided differences and discuss the err
involved.

(b) Establish the following formulae for approximating $\int_0^4 u(x)dx$:

(i) $4(2u_0-4u_1+5u_2)/3$

(ii) $(u_0+u_4+4u_1+4u_3+2u_2)/3$

and in each case obtain the error term.

Or,

(a) Derive Simpson's one-third rule for numerical quadrature complete with impro
error term.

(b) Below are given the values of $y = \frac{1}{1+x}$ at intervals of 0.5 in x

x	0	0.5	1.0	1.5	2.0	2.5	3.0
y	1.0000	0.6667	0.5000	0.4000	0.3333	0.2857	0.2500

Obtain a value for $\int_0^3 \frac{dx}{1+x}$ by Simpson's rule and compare this value with the exact value of the integral.

PAPER II : APPLIED STATISTICS, GENERAL

Time: 4 Hours | Full marks : 100

(a) Attempt *any five* questions.
(b) All questions carry equal marks.
(c) Use of calculating machines is not permitted.

1. *Either,*

Give some illustrations of uses of sample surveys for collection of data in various fields. Suppose a survey is to be conducted for estimating the volume of rural saving in India, with a view to ascertain the resources available in the rural area for the Third Five Year Plan. What type of sampling design will you suggest ? Give reasons for the design suggested. Indicate the type of frame to be used and describe the method of selection.

Or,

You are required to investigate how indebtedness amongst Schedule caste compare with that amongst the other economically similar sections of society. Describe how you will design and carry out the survey and what information you will collect.

2. Give a critical review of statistics of land utilisation and agricultural production in India, indicating recent improvements, if any.

3. Why do we compute consumer price index numbers ? If you are asked to calculate the consumer price index number of working class people in an industrial area, what type of data will you need ? How will you collect these data ?

4. Expalin the role of statistical method in business forecasting, pointing out its limitations.

5. What is meant by a factorial experiment ? Why are such experiments peferred to those that study one factor at a time ? Suppose you are required to prepare a lay out for an experiment to study the effects of compost manure, chemical fertilizer and spacing, each at two levels, on rice yield. Give an outline of the design you will adopt and of the method of analysis.

6. What are the salient features of the 1951 population census of India ? What improvements would you like to suggest in the scope of the census proposed to be taken in 1961 ?

7. What do you mean by the 'reliability' and 'validity' of a psychological test ? How will you measure the reliability of a test and estimate the 'true score' ?

8. What methods will you adopt for estimating the intensity of linkage between two genes using (a) backcross data, (b) F_2 data ? Derive the formula for the variance of the estimate in each case.

9. Write short notes on *any three* of the following :-

(a) Structure and compilation in India of balance of payment statistics.
(b) Sources of error in primary data collected in a sample survey.
(c) Usefulness of family budget data for forecasting of demand.
(d) Fractional replication.
(e) Life table.
(f) Mendel's laws of inheritance.

PAPER III : STATISTICAL INFERENCE

Time : 4 Hours | Full marks : 100

(a) Attempt *any four* questions.
(b) All questions carry equal marks.
(c) Use of calculating machines is not permitted.

1. Let L be the likelihood function involving a single parameter θ ;

let
$$\psi(\theta) = \frac{\partial \log L}{\partial \theta} \quad \text{and}$$

$$I(\theta) = -E\left(\frac{\partial^2 \log L}{\partial \theta^2}\right).$$

Consider the random variable

$$S(\theta) = \theta + \frac{\psi(\theta)}{I(\theta)}.$$

Under usual regularity conditions :

(a) Show that $E[S(\theta)] = \theta$, $V[S(\theta)] = \dfrac{1}{I(\theta)}$

(b) Let T be any unbiassed estimate of θ; show that $\text{Cov}[T, S(\theta)] = \dfrac{1}{I(\theta)}$. Hence, or otherwise, prove that

$$V(T) \geqslant \frac{1}{I(\theta)}$$

(c) Hence show that if $S(\theta)$ does not involve θ, it is an essentially unique minimum variance unbiassed estimate of θ.

(d) Let $x_1, x_2, \ldots, x_n$ be n independent observations from a population with probability density function

$$\frac{1}{\theta}\, e^{-x/\theta} \; (0 < x < \infty \;;\; \theta > 0)$$

Making use of the results above, obtain the minimum variance unbiassed estimate of θ.

2. (a) Define a sufficient statistic in terms of conditional probability distributions on the sample space.

(b) X_1 and X_2 are independent random variables and $X_i = 1, 0$ with probability $p, 1-p$, $(i = 1, 2)$. By considering probability distributions in the sample space of (X_1, X_2), show that the statistic X_1+X_2 satisfies the above definition of sufficiency.

(c) How will you verify if the proportion of successes in n Bernoulli trials is a sufficient statistic for p, the probability of success ?

(d) State without derivation the general form of distributions which admit of a sufficient statistic. Verify if the normal distribution $N(\mu, 1)$ and the binomial distribution (n, p) can be expressed in this form.

3. (a) For the problem of estimating a single parameter θ by the method of maximum likelihood, in case the likelihood equation turns out to be too complicated to admit an algebraic solution, show how starting from an initial approximation θ_0, a solution can be obtained from the iterative scheme

$$\theta_{i+1} = S(\theta_i) \quad i = 0, 1, 2, \ldots$$

where $S(\theta)$ is the function defined in question 1.

(b) Let $x_1, x_2, \ldots, x_n$ be n independent random variables, each with the probability law

$$\text{Prob. } (X = x) = \frac{\theta^x}{(e^\theta - 1)x!} \quad (x = 1, 2, \ldots)$$

Give a computational scheme for obtaining the maximum likelihood estimate of θ What is the asymptotic variance of this estimate?

(c) Show that in this case an unbiassed estimate of θ is given by $t = \bar{x} - f_1/n$ where $\bar{x}$ is the sample mean and f_1 the frequency of '1'. What is the variance of t ?

4. (a) What is meant by a confidence-interval for a parameter θ with confidence-coefficient $(1-\alpha)$?

(b) State (without proof) how you will set up confidence intervals with confidence-coefficient $(1-\alpha)$ for each of the following parameters, on the basis of random samples drawn from the population concerned :-

(i) The variance of a Normal population

(ii) The proportion of defective articles in a lot

(iii) The coefficient of correlation in a bivariate Normal population.

(c) With a random sample of size $N = nm$ from a Normal population with unknown parameters, write down, without proof, the expression for the usual confidence-interval for the mean with confidence-coefficient $(1-\alpha)$. Compute λ_1 the expected length of this confidence interval.

Since this requires heavy computation (squaring of N items) an alternative procedure was suggested. This calls for drawing m independent random samples, each of size n. Let $\bar{x}_i$ be the mean of the i-th sample and let

$$\bar{\bar{x}} = \frac{1}{m}\sum_{i=1}^{m} \bar{x}_i, \quad e^2 = \frac{1}{m(m-1)}\sum_{i=1}^{m}(\bar{x}_i - \bar{\bar{x}})^2$$

Let $t(\alpha, m-1)$ denote the both-sided 100α percent point of the t-distribution with $(m-1)$ degrees of freedom. Show that

$[\bar{\bar{x}} \pm e\,t(\alpha, m-1)]$ provides another confidence-interval

for the mean with confidence-coefficient $(1-\alpha)$ and find λ_2 the expected length of this confidence-interval.

Give your comments on the magnitude of the ratio λ_2/λ_1 for moderately large values of n and m, say $m = 10$ and $n = 100$.

5. (a) What are locally most powerful unbiased tests ?

(b) How are such tests constructed for a hypothesis concerning a single parameter ?

(c) Obtain such a test for the hypothesis $\sigma = \sigma_0$ in a normal distribution $N(0, \sigma)$.

6. (a) What is meant by a composite hypothesis ?

(b) Describe the likelihood-ratio method for constructing tests of composite hypothesis.

(c) Consider k Normal populations with possibly different means $\mu_1, \mu_2, \ldots \mu_k$ but the same variance σ^2. Suppose that a sample of size n is available from each of these k populations. Obtain the likelihood-ratio test for the hypothesis that $\mu_1 = \mu_1 = \ldots = \mu_k$.

7. Suppose that n observations $(x_{1\lambda}, x_{2\lambda}, \ldots, x_{p\lambda})$ $\lambda = 1, 2, \ldots, n$ are available on p-variate $X_1, X_2, \ldots, X_p$ having a joint Normal distribution with $E(X_i) = \mu_i$ and $V(X_i) = \sigma_{ii}$, Cov $(X_i, X_j) = \sigma_{ij}$, $i, j = 1, 2, \ldots, p$.

(a) Show that the hypothesis $H : \sigma_{ij} = 0$ for $j = 2, 3, \ldots, p$ is equivalent to the hypothesis that X_1 is statistically independent of any linear function of the form $Y = l_2X_2 + \ldots + l_pX_p$. Hence derive a criterion for testing H by maximising the sample correlation between X_1 and Y with respect to $l_2, \ldots, l_p$. Show that this yields the sample multiple correlation coefficient of X_1 on $X_2, \ldots, X_p$. What is its null distribution ?

(b) Give the expression for the statistic you would use to test the hypothesis that

$$\mu_1 = \mu_2 = \ldots = \mu_p = 0$$

separately for the cases when the variance-covariance matrix $((\sigma_{ij}))$ is (i) known and (ii) unknown and has to be estimated from the sample. In each case state without proof the null and the non-null distribution of the statistic.

PAPER VI : (Practical)

Time : 6 Hours | Full marks : 100

(a) Attempt *any four* questions.
(b) All questions carry equal marks.
(c) Use of calculating machines is permitted.

1. (a) Find the present value of an annuity of one per annum for 27 years at 5 percent compound interest, given

number of years of annuity	present value at 5 percent
15	10.3797
20	12.4622
25	14.0939
30	15.3725
35	16.3742
40	17.1591

Also find the number of years of annuity for which the present value is 14.5.

(b) From the table below sub-tabulate to six places of decimal at intervals of 0.001, within the range $x = 0.02$ to $x = 0.03$.

x	y
0.00	1.5707963
0.01	1.5747456
0.02	1.5787399
0.03	1.5827803
0.04	1.5868678
0.05	1.5910035

2. The following table shows the production index of a factory in the U.S.A. during the period 1943-1954 with 1942 as base.

year	production index	year	production index
1943	101.1	1949	125.1
1944	107.1	1950	123.9
1945	111.0	1951	123.1
1946	112.6	1952	123.6
1947	114.7	1953	126.0
1948	119.9	1954	124.6

(a) Fit a polynomial trend of suitable degree using orthogonal polynomials.
(b) Represent the data and the trend on a graph paper.
(c) Predict the production index for 1960.

3. The table below shows the number of wheat seeds out of 50 which failed to germinate in 3 repetitions of an experiment under four different treatments.

number of experiments	treatments			
	1	2	3	4
1	10	11	8	9
2	8	10	3	7
3	5	11	2	8

Assuming that in any given class the frequency obeys a binomial distribution, test for significance of treatment differences using a suitable transformation of these observations, if necessary.

4. (a) A correlation $r = 0.562$ is reported to be significant at 1 percent level. Show that it must have been computed from at least 20 observed pairs.

(b) The correlation coefficient between fibre weight and staple length in six cotton crosses was estimated as

$$-0.1289, \quad 0.1158, \quad -0.2780, \quad 0.0033, \quad 0.2331, \quad 0.0550$$

based on sample sizes 73, 81, 67, 83, 71, 57 respectively. Can the samples be reasonably assumed to come from equally correlated populations ? If so, obtain a better estimate of the common correlation coefficient by combining the above estimates. Give the 95 percent confidence limits for the population correlation coefficient using the combined estimate you have obtained.

5. A number of recruits are given a preliminary test to ascertain their suitability for a certain course of training. At the end of the training course, they undergo a proficiency test. The marks for three groups of recruits from three different towns are as follows.

Group 1	Preliminary :	45, 50, 56, 58, 59, 60, 62, 64
	Proficiency :	46, 60, 52, 46, 48, 50, 55, 63
	Preliminary :	65, 75
	Proficiency :	58, 64
Group 2	Preliminary :	44, 49, 52, 58, 59, 60, 62
	Proficiency :	48, 55, 45, 60, 64, 69, 71
	Preliminary :	63, 63, 66, 69, 70, 72, 73
	Proficiency :	77, 70, 75, 80, 72, 75, 81
Group 3	Preliminary :	47, 52, 59, 60, 63, 66, 68
	Proficiency :	43, 56, 51, 72, 60, 61, 55
	Preliminary :	69, 74, 76
	Proficiency :	74, 72, 80

(i) Test separately in each group whether the relationship between preliminary (X), and proficiency (Y) can be represented by a linear relation of the form

$$Y-\mu_y=\beta(X-\mu_x)$$

(ii) Do the regressions differ in the three groups ?

(iii) Is the relation between group means linear ?

6. The following data were presented in an article published in a medical journal :—

(a)

treatment	total number of cases	number of cases with complications	percentage of complications
Drug A	189	13	6.88
Drug B	104	10	9.62
Totals	293	23	7.85

followed by a summary statement which read as

'It was found that there were 40 percent more complications when Drug B was used than when Drug A was used.' Is this conclusion valid ?

(b) In a political campaign a candidate claimed that at least 60 percent of the electorate would vote for him. A sampler of public opinion asked 1000 registered voters if they expected to vote for this candidate and 55 percent of them said 'yes'. Test whether this candidate is justified in his claim.

What statement about the electorate can the sampler make with 99 percent confidence?

(c) The data in the table given below are from an experiment on the use of drugs (sulfones and streptomycin) in the treatment of leprosy. The rows denote the 'change' in the overall clinical condition of the patient during 48 weeks of treatment; the columns indicate the degree of infiltration (a measure of certain type of skin disease) present at the beginning of the experiment. Test whether patients with much initial infiltration progressed differently from those with little infiltration.

TABLE : 196 PATIENTS CLASSIFIED ACCORDING TO CHANGE IN CONDITION AND DEGREE OF INFILTRATION

clinical change	infiltration		total
	0–7	8–15	
Improvement :			
Marked	11	7	18
Moderate	27	15	42
Slight	42	16	58
Stationery	53	13	66
Worse	11	1	12
Total	144	52	196

PAPER VII : (Practical)

Time : 6 Hours Full marks : 100

(a) Attempt *any four* questions.

(b) All questions carry equal marks.

(c) Use of calculating machines is permitted.

1. The following figures are available for the electricity generated for public supply, but the spaces marked with an asterisk are not available. Make appropriate estimates of the missing figures.

Year	Month			
	January	April	July	October
1952	6319	4775	4108	5526
1953	6609	5269	4489	*
1954	*	5634	4972	6241
1955	7988	6102	5115	7022
1956	8588	6952		

(Units are million killowatt hours)

[Hint :You are required to isolate the trend and the seasonal fluctuations, if any, and make proper use of these components for building up your estimates. The trend may be assumed to be linear].

2. (a) It is well-known that the consumption of goods and services varies with the person's age. In a certain study it was estimated that if the total consumption of a person of age 30 be unity the equivalent total consumption for other age-groups would be as follows:

Age	0–14	15–29	30–44	45–59	60–74	75 and above
Equivalent consumption	0.19	0.81	0.95	0.68	0.32	0.06

The following table gives the population of Great Britain, by age, in 1931 and al in 1951.

Age	Population (thousands)	
	1931	1951
0–14	11,174	11,388
15–29	11,680	10,395
30–44	9,891	11,305
45–59	7,979	9,576
60–74	4,357	6,126
75–above	957	1,766

Compute the rate of increase in the total population and also in total consumptior between the two dates. Account for the difference in these rates.

(b) In a certain community, a total of 5 deaths due to diptheria was reported in 1956 among children who have never been inoculated and among children who have bee inoculated 8 died of diptheria during the same year. In 1956 the number of inoculated an uninoculated children in the community was 24000 and 5000 respectively.

Compute the fatality rate per thousand children in each group and comment o the effectiveness of diptheria inoculation.

Are you prepared to revise your opinion if you are told that all the 24000 childre were inoculated in 1956 during an intensive inoculation program which commence on April 1, 1956 and was uniformly spread over the rest of the year.

[Hint : Compute the total number of child years of inoculated and uninoculatec exposure and recalculate the fatality rates].

3. *Either*,

A series of experiments was carried out on the keeping quality of milk stored by three different methods while in transit. Owing to the initial variations in the quality of the milk, samples of milk from the same farm were stored by all the three methods and their keeping qualities were separately assessed.

The keeping qualities (in hours) in samples taken from different farms are give below:-

Sample	Keeping qauality by method		
	A	B	C
1	18.5	17.0	18.0
2	15.0	15.5	16.0
3	31.5	32.0	33.0
4	29.5	28.5	29.5
5	17.0	16.0	16.5
6	29.0	29.5	29.5
7	28.0	26.5	27.5
8	20.5	20.0	21.0
9	16.0	15.5	17.0
10	33.0	33.5	32.5

Analyse this data and test for differences in storage methods. Can you recommend a method which would give the best keeping quality ?

Or,

The following table gives the yield (in lbs. per tree) of organges in three consecutive asons, for each plot of a randomised block experiment laid in three blocks of four plots ch:

season	block	variety			
		A	B	C	D
1924	1	130	152	118	134
	2	171	163	164	165
	3	159	173	170	164
1925	1	132	125	119	119
	2	146	138	149	146
	3	150	143	140	142
1926	1	93	109	85	95
	2	122	108	122	123
	3	112	120	124	104

nalyse the yields and test for varietal differences

(a) separately, in each season and

(b) in the total yield over all the three seasons.

4. The following are the means and ranges of 20 samples of 5 each. The data rtain to the overall lengths of a fragmentation bomb base manufactured by an ordnance tory. The measurement are in inches.

group number	X	R	group number	X	R
1	0.8372	0.010	11	0.8380	0.006
2	0.8324	0.009	12	0.8322	0.002
3	0.8318	0.008	13	0.8356	0.013
4	0.8344	0.004	14	0.8322	0.005
5	0.8346	0.005	15	0.8304	0.008
6	0.8332	0.011	16	0.8372	0.011
7	0.8340	0.009	17	0.8282	0.006
8	0.8344	0.003	18	0.8346	0.005
9	0.8308	0.002	19	0.8360	0.004
10	0.8350	0.006	20	0.8374	0.006

(a) From these data set up an $\bar{X}$ chart and an R chart to control the lengths of mb base produced in the future.

(b) The above samples were taken every 15 minutes in order of production after anging fixtures. The production rate was 350-400 per hour and the tolerances (speci-ation limits) were 0.820 and 0.840 inches.

(i) On the assumption that lengths of bomb bases are normally distributed, hat is the percent defective of the above process operating at the levels indicated by the bove data?

(ii) What would happen to the percent defective if the process average should ift to 0.8370?

(iii) What is the probability that you would catch such a shift on your control chart on the first sample following the shift ?

(iv) How many samples (of size 5 each) would you have to take to have a chance of approximately 0.95 of catching the shift on at least one of these samples?

(v) If this shift is not caught before the eighth sample, approximately how many defective bomb bases will have been produced in the interval ? (You may assume that the shift takes place immediately after a sample has been drawn.)

(vi) From the time the shift takes place and until it is detected on the control chart what is the expected number of defective articles that will be produced (Assumption: as in (v)).

(vii) From these considerations, would you prefer a control program where a sample of size ten are taken every half an hour and the sample mean and range are plotted on appropriate charts.

5. A consumer preference study involving three varieties of snap beans was conducted as follows :-

One lot of each varieties (V_1, V_2 and V_3) was displayed in retail stores and each of n consumers was asked to rank the beans according to first, second and third choices.

Let f_{ijk} be the observed frequency of consumers whose preferences in diminishing order are V_i, V_j and V_k [where (i, j, k) is a permutation of integers (1, 2, 3)] and let 0_{im} be observed number of consumers assigning rank m to V_i ($i = 1, 2, 3$, $m = 1, 2, 3$).

(a) Express 0_{im} in terms of the f_{ijk}'s

(b) Formulate clearly the null hypothesis H_0 that all the varieties are equally preferred in the population.

(c) What is the distribution of the f_{ijk}'s when H_0 is true ?

(d) Assuming H_0 to be true, obtain (i) Variance (0_{11}), (ii) Cov (0_{11}, 0_{12}), (iii) Cov (0_{11}, 0_{22}).

(e) The following data were obtained in one store on one day :-

TABLE : CONSUMER RANKING OF THREE VARIETIES OF SNAP BEANS

variety	rank 1	rank 2	rank 3	total
V_1	42	64	17	123
V_2	31	16	76	123
V_3	50	43	30	123
total:	123	123	123	

Why is the usual contingency chi-square not applicable in this case for testing H_0?

(f) Is it possible to test H_0, by the frequency chi-square, using the data for one rank only or for a single variety, which corresponds to one column or one row respectively of two way table. If so, why ? Apply such tests, considering separately the second row (Variety 2) and first column (Rank 1) of the given table.

(g) An overall test using the entire data can be performed as follows. Let Q^2 be the usual contingency chi-sauare computed from the two-way table of the 0_{im}'s. It can be shown that if the table be of r rows and r columns, that is, if altogether r varieties be under comparison, then under H_0

$$\frac{r}{r-1} Q^2$$

has asymptotically a chi-square distribution with $(r-1)^2$ degrees of freedom.

Apply this test to the given data.

6. *Either*,

In a stratified sample to estimate the crop in a region, it is given that the variance of the estimate, and the total cost involved W, are of the following forms.

$$V = \sum_i \frac{A_i^2 a_i}{n_i \; x_i^g}$$

$$W = b + \sum_i n_i (c_i + d_i x_i)$$

where A_i is the geographical area, n_i = the number of sample unit's and x_i = size of the sample unit in the i-th stratum, and the other letters denote some constants. The number of strata is only three and the numerical values of the constants are given below :-

b = Rs. 5000/-

A_i (sq. mile)	a_i	g_i	c_i (Rs.)	d_i (Rs.)
1. 7000	0.09	0.4	30	3.50
2. 9000	0.12	0.5	36	4.00
3. 8000	0.11	0.6	32	3.00

Find out the optimum values of n_i and x_i for W = Rs. 50,000/- and find out the corresponding value of V.

Or

The sugar beets on a 1-acre field were sampled for sugar percentage. This was done by dividing the field into fifteen strata, each of four rows. The sugar percentage was then estimated for two samples of 10 beets from each stratum with the following results:-

stratum	sample 1	sample 2
1	14.58	13.81
2	13.35	13.87
3	13.90	14.31
4	13.49	14.78
5	14.92	14.14
6	14.71	13.44
7	14.48	14.85
8	15.01	14.58
9	14.28	14.24
10	14.38	14.46
11	14.14	13.73
12	14.38	14.27
13	15.04	15.19
14	14.58	14.87
15	16.20	14.50

This constituted a 1 in 20 sample of the whole field. Assuming that beets are of uniform weight, estimate the mean sugar percentage of the whole crop and find the standard error of this estimate.

What would be the estimate and its standard error if the data are regarded as relating to a random sample with no stratification of the field ?

PAPER IV AND V : STATISTICAL QUALITY CONTROL (Theoretical)

Time : 4 Hours Full marks : 100

(a) Attempt *any four* questions.
(b) All questions carry equal marks.
(c) Use of calculating machines is not permitted.

1. Mention at least four types of situations on the control chart, on the basis of which action may be recommended. Briefly indicate in each case the theoretical justification for the recommendation.

2. Account for the following procedures in control chart analysis, stating clearly any assumption involved.

a) a sample size of 5 is ordinarily used in measurement inspection.

b) the Range is used in place of standard deviation.

c) the moving range is calculated when one measurement only per rational sub-group is made.

d) the mean-range is preferred to range in many cases.

e) Three-sigma limits are used for control limits instead of exact probability limits, whatever be the type of quality characteristic.

3. Discuss briefly the problems ordinarily faced in starting and maintaining control charts in a manufacturing plant, with special reference to the statistical principles useful in approaching such problems.

4. Reconcile the following statements with respect to acceptance sampling plans for attributes :-

'In constructing the plans, the approximation used for the probability of obtaining '*d*' defectives in a sample size '*n*' is not a function of '*N*' the lot size. But the Dodge-Roming plans are given for specific lot sizes.'
and examine whether the reasoning you put forward is applicable to the AQL attributes plans which are also given against specific lot sizes.

5. (a) Under what circumstances are AOQL plans applicable ?

(b) How far is the statement the 'AOQL value will never be exceeded in practice' correct ?

(c) Indicate the derivation of the AOQL function for a double sampling attributes plan.

6. Write notes on *any four* on the following :-

i) acceptance plans for variables,
ii) continuous sampling plans,
iii) the MIL-STD 414 and MIL-STD 105A,
iv) modified control charts,
v) acceptance procedure for two-sided specification limits,
vi) sequential sampling,
vii) interval estimates of lot quality.

PAPER IV AND V : SAMPLE SURVEYS, THEORY (Theoretical)

Time : 4 Hours Full marks : 100

a) Attempt *any four* questions.

(*b*) All questions carry equal marks.

(c) Use of calculating machines is not permitted.

1. (a) Discuss the advantages of stratified sampling in comparison with unstratified sampling.

(b) A population is divided into K strata. If m_i units are selected at random from the M_i units in the i-th stratum ($i = 1, 2, \ldots K$) find the unbiassed estimate of the population total and obtain its standard error.

(c) Discuss the relative efficiency of stratified proportional sampling with that of random sampling without stratification.

2. Describe the estimation procedures you will recommdnd in each of the following cases of two-stage sampling for estimating the population total of a character :-

(i) Select n primary units (p.u) with equal probability and from each selected p.u. select m elements also with equal probability;

(ii) Select n primary units (p.u) with probability proportional to their size (M_i) with replacement, and select with equal probability m elements from each selected p.u. (M_i is the number of elements in the i-th p.u.).

(iii) Select n primary units (p.u.) with equal probability and constant proportion of elements from each selected p.u. such that the overall sampling fraction is the same as in the two procedures given above.

Derive an expression for the variance of the estimate under (ii). Comment on the relative usefulness of the different procedures suggested.

3. (a) Define a ratio estimate for estimating the population total of a character y, and derive an expression for the standard error of the estimate.

(b) If the coefficient of variation of the auxiliary variate 'x' is more than twice the coefficient of variation of the character 'y', then show that in large samples, with simple random sampling, the ratio estimate is less precise than the 'mean per sampling unit' estimate. Is the converse true ?

(c) State the conditions under which the ratio estimate is 'best linear unbiassed estimate'.

4. (a) what are the main considerations in the choice of the size of a sampling unit ?

(b) In a survey on farm conditions, it was desired to group M neighbouring farms into clusters to form larger sampling units. Two components of field costs are distinguished; a component $c_1 M_n$ comprises costs that vary directly with the total number of farms, M_n, in the sample and the second component $c_2\sqrt{n}$ measures the cost of travel between the n sample clusters. The total field cost, c, is thus given by

$$C = c_1 M_n + c_2\sqrt{n}$$

If the variance between farms within clusters (s_w^2) follows the empirical law $s_w^2 = AM^g$, where A and g are positive constants, obtain the optimum size of the cluster.

5. (a) What is meant by multistage sampling ? Obtain an expression for th variance of the estimate of a 'total' for a suitable three-stage sampling when the units ar of unequal size at each stage of sampling.

(b) Give the structure of the analysis of variance in three stage sampling and ex plain how the analysis can be used in planning similar surveys subsequently.

6. Write critical notes on :-

(i) Partial replacement in sampling on successive occasions.

(ii) Double sampling

(iii) Systematic sampling.

PAPER IV AND V : SAMPLE SURVEYS, APPLIED (Theoretical)

Time : 4 Hours Full marks : 10

(a) Attempt *any five* questions.

(b) All questions carry equal marks.

(c) Use of calculating machines is not permitted.

1. (a) What are the various uses of pilot surveys ?

(b) What are meant by cost and variance functions with reference to large sca sample surveys and in what ways are they useful for such surveys ? What are the importar components of the cost function ?

2. Describe the various stages of a large scale sampling enquiry in the sequence i which you would proceed.

3. Explain clearly the concepts of sampling and non-sampling errors in surveys. Ho do you propose to control the latter ?

4. What is the difference between a questionnaire and a schedule ? In what situa tions are they used ? What points should you consider while writing instructions for in (i) schedule, (ii) questionnaire?

5. What are the different methods of data collection ? Explain their advantag and disadvantages and mention the situations in which they are useful.

6. Describe the different checking systems that you may adopt for ensuring accurac of work at the processing stages of a large scale survey and explain their relative advantag and disadvanages.

7. Write short notes on the following explaining their physical meaning, advantag and disadvantages. Also write down the expressions for the appropriate estimates (population total and its variance in each case :-

(i) Two-stage sampling with equal probability and with replacement at both stages.

(ii) Cluster sampling with equal probability and without replacement.

(iii) Sampling with probability proportional to a given measure of size.

(iv) Interpenetrating sampling.

PAPER IV AND V : DESIGN OF EXPERIMENTS, APPLIED (Theoretical)

Time : 4 Hours Full marks : 100

(a) Attempt *any four* questions.

(b) All question carry equal marks.

1. Discuss the role of randomisation, replication and local control in planning experimental investigations with special reference to experiments in randomised block design.

'The variance ratio test in a randomised block experiment has only a small chance of detecting overall differences in treatments equal to or smaller in magnitude than the block × treatment interactions.' Discuss.

2. (a) Describe a Graeco Latin square design.

(b) An experimenter while planning for a latin square experiment on mango trees foresees the possibility of losing interest in the original set of treatments and changing over to a new set for next year's experiment on the same collection of trees. Discuss how a Graeco latin square could be effectively used for the second year experiment to eliminate residual effects which the original treatments might have left on the trees. What restrictions should be imposed in such a case on the randomisation in the initial year ?

(c) After the completion of a 4×4 latin square experiment on wheat, it is discovered that two out of the four treatments used are actually identical. Suggest modifications that are needed in the analysis, indicating the appropriate degrees of freedom and the algebraic expressions for the 'treatment' and the 'error' sum of squares.

3. In a sugar beet experiment, dressings of 0,1, and 3 tons of chalk per acre were compared in 1932 in 5 randomised blocks of 4 plots each. The same crop was also grown in 1933 with no further treatments added to the soil. The weights of clean roots in lbs. per plot are available for each one of 20 plots separately for each year. Suggest a suitable analysis for this data for providing answers to the following questions :-

(a) Are there any differences among treatments in respect of their average yield in two years ? (The linear and the quadratic responses are required to be studied separately).

(b) Are there any treatment × year interactions?

Your analysis should take account of the fact that the yields of the same plot in consecutive years are likely to be correlated.

4. You are required to plan an experiment to determine whether there is an inequality in the arterial blood pressure of the two arms of man. Forty individuals are available for measurement. It is however desirable that gross differences in blood pressure that may exist between individuals should not enter into the comparisons and if multiple measurements are taken on the same individual the experiment should also take proper account of the 'order' of measurement since it is known that when repeated blood pressure readings are taken on the same individual at short intervals, there is usually a definite decrease in the second and subsequent readings.

Prepare a suitable design for obtaining measurements and suggest appropriate methods of analysis.

5. Describe the steps involved in the recovery of interblock information in incomplete block experiments discussing fully the methods of obtaining combined inter-and intra-block estimates and their standard errors and also suggesting appropriate methods for testing the the equality of treatment effects.

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

(a) Criss cross designs
(b) Change-over trials
(c) Youden squares
(d) Uniformity trials
(e) Interaction as valid error
(f) Fractional replication.

7. What is meant by partial confounding in a factorial experiment ?

Draw up a balanced scheme of partial confounding in a 2^5 factorial experiment in blocks of 8 plots each ? (Only the interactions confounded in each replication and the detailed treatment allocation in blocks of one replication need be shown).

Briefly explain how you will compute the sum of squares due to a partially confounded interaction.

PAPER IV AND V : MATHEMATICAL THEORY OF SAMPLING DISTRIBUTION (Theoretical)

Time : 4 hours | Full marks : 100

(a) Attempt *any four* questions.
(b) All questions carry equal marks.

1. Let $x_1, \ldots, x_p$ are p random variables having the joint probability density

$$\text{Const. } e^{-\Sigma\Sigma \lambda^{ij}(x_i-\mu_i)(x_j-\mu_j)/2}$$

Answer the following questions :—

(a) What is the value of the constant ?
(b) How is the matrix (λ^{ij}) related to the variance-covariance matrix of $x_1 \ldots, x_p$?
(c) What is the distribution of

$$\Sigma\Sigma \lambda^{ij}(x_i-\mu_i)(x_j-\mu_j) ?$$

(d) How is a subset, $x_1, \ldots, x_r$ of the variables distributed ?

2. (a) If $x_1, \ldots, x_p$ is a sample from a rectangular distribution in the range $(\theta, 2\theta)$ find the joint distribution of the minimum, η, and maximum, ξ, in the sample.

(b) Compute the variance-covariance matrix of ξ and η and find the linear function of ξ and η which is unbiased for θ and has minimum variance.

3. (a) Explain the importance of Hotelling's generalized T^2 test.

(b) Find the non-null distribution of the T^2 statistic when the number of variables is two.

4. (a) Define intra-class correlation. If from each family the heights of a brother and a sister (both adults) are obtained, then, to estimate the correlation between the heights of brothers and sisters, would you compute the intra or inter-class correlation. Give reason for your answer.

(b) Find the distribution of the intra-class correlation computed from n pairs of observations assuming bivariate normal distribution.

5. (a) What is a characteristic function ? How is it useful in the derivation of sampling distributions ?

(b) Give a proof of the law of large numbers using charactoristic functions.

6. Find the distributions of the following :-

(a) The distance between any two points chosen at random on an interval from to 1.

b) The distance from an arbitrary origin of the point at which an α-particle emanating in a random direction from an outside source strikes a straight line boundary.

c) The sum of the two distances corresponding to two α-particles in (b).

PAPER IV AND V : PROBIT ANALYSIS (Theoretical)

Time : 4 Hours | Full marks : 100

(a) Attempt *any three* questions.
(b) All questions carry equal marks. Four marks are reserved for neatness.
(c) Avoid verbiage; credit will be given for brief and pointed answers.

1. It is known that the percentage free-alkali content of cakes of soap manufactured by a certain company is distributed Normally with mean μ and standard deviation σ, but the numerical values of μ and σ are only approximately known. The problem is to estimate their values more precisely and the following experiment is suggested for this purpose.

It is a tedious job to determine the precise alkali content of a cake of soap by individual titration. It is comparatively easy to prepare an indicator solution set up at a fixed level, say x, such that on immersion of a cake in the solution, the indicator would change colour if and only if the free alkalinity of the cake is greater than x. Two such indicators would be arranged, one set up at $x_1 = \mu + k\sigma$ and the other at $x_2 = \mu - k\sigma$ for a suitably chosen value of k. A different sample of n cakes will be tried on each of these two indicators separately, so that the experiment will give the number r_i of cakes in the i-th sample of n cakes with free alkalinity greater than x_i $(i = 1, 2)$.

Obtain estimates of μ and σ and their standard errors. How would you choose k?

2. An insecticide is to be prepared by mixing two poisons A and B in the ratio $1 : r$. A is a stomach poison and B is a contact poison. What are the considerations in choosing a suitable value of r? How will you plan to collect information that you may require to solve this problem

3. In a biological assay of the quantal type in addition to k dosages tested with n_i subjects at the i-th dosage x_i $(i = 1, 2, \ldots, k)$ there is also a control group of n_0 subjects on which no dosage is given. Assuming that the probability of response in the control group is ρ and in a group where a dosage x is given, is

$$\rho + (1-\rho)\, P(\alpha + \beta x)$$

where P is a function of known form, obtain computational formulae for estimating the parameters α, β, ρ by the method of maximum likelihood.

4. Describe, compare, and give your critical comments on *any four* of the following procedures for estimating the median effective dose from a quantal type bio-assay :-

(a) Probit : maximum likelihood method
(b) Logit : minimum chi-square method
(c) Robbins-Monro sequential procedure
(d) Behren's method
(e) Karber's method

5. (a) Enumerate carefully the assumptions behind the probit analysis technique for estimation of the median effective dose and describe how experiments can be conducted to examine the validity of these assumptions.

(b) Explain how the fundamental principles of randomisation, replication and error control are made use of in planning bioassays of the quantal type.

PAPER IV AND V : PSYCHOLOGY AND EDUCATION (Theoretical)

Time : 4 Hours Full marks : 100

(a) Attempt *any four* questions.

(b) All questions carry equal marks.

1. Differential aptitude tests and general classification ('intelligence') tests are two types of tests used in educational and vocational selection and placement. Describe situations for which differential aptitude tests should be used and explain why. Also describe situations for which general classification tests should be used and explain why.

2. Compare projective and questionnaire tests of personality, in terms of method of response, objectivity of scoring, and reliability. Illustrate the comparison by reference to well-known or typical personality tests.

3. Describe the centroid method and the principal components methods of factor analysis. In what ways are the two methods different?

4. Plan a research project for the purpose of predicting B.Sc. performance from data to be collected before entering the university. The design of the project, types of data to be collected, sampling procedures, and statistical methods to be employed should be indicated.

5. Discuss the influence of factor analysis on educational and psychological tests, with reference to test construction, reliability, item analysis, and validity.

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

a) Effect of speed on a split-half reliability estimate.

b) Selection of items so that differences in difficulty are approximately equal.

c) Correlation of test scores for chance success.

d) 'I.Q.' and its measurement.

PAPER VIII AND IX : ECONOMIC STATISTICS (Practical)

Time : 4 Hours Full marks : 100

a) Figures in the margin indicate full marks.

b) Use of calculating machines is permitted.

1. *Either,*

The following is a time-series of prices p_t. Fit a linear trend and find for residual time-series the first-four serial correlations. State whether any conclusions can be drawn regarding the nature of the residual time-series, namely, whether it is a moving average type or an auto-regressive type or a series of independent random variables.

t	p_t
1	28.50
2	29.00
3	25.50
4	33.00
5	34.50
6	40.00
9	40.50
10	45.00
11	43.50
12	44.00
13	43.50
14	45.00
15	44.50
16	43.00

(60)

Or,

The following table gives the acreage a_t under jute and the parity price p_t of jute to rice for the years 1939-57 for West Bengal.

year	acreage under jute in West Bengal ('000 acres) a_t	parity price of jute to rice p_t
(1)	(2)	(3)
1939	174	2.39
1940	577	2.20
1941	181	1.65
1942	282	1.28
1943	241	0.87
1944	194	1.10
1945	203	1.07
1946	152	1.51
1947	229	2.29
1948	315	2.43
1949	498	2.43
1950	651	2.44
1951	876	4.14
1952	820	2.19
1953	535	1.88
1954	551	2.19
1955	854	2.31
1956	720	2.10
1957	759	2.18

Estimate simultaneously the parameters of the model :

$$a_t = \alpha p_{t-1} + \beta + \epsilon_t$$

$$p_t = \gamma a_{t-1} + \delta + \eta_t$$

assuming that ϵ_t and η_t are mutually independent random residuals. Interpret the results. (60)

2. With population in 1951 as 356.88 and the estimate for 1956 as 391.40 million respectively, calculate the index of calories available per head in 1956-57 with 1951-52 as base using the data given below. Assume, where necessary, that 7/8 of the production is available for human consumption. (40)

foodgrain	production in '000 tons		calories per ounce
	1951–52	1956–57	
rice	20,964	28,282	97
jowar	5,981	7,249	101
bajra	2,309	2,885	102
maize	2,043	3,009	97
ragi	1,291	1,715	98
small millets	1,885	1,964	87
Wheat	6,085	9,314	98
barley	2,330	2,827	95
gram	3,334	6,264	103
tur	1,801	1,954	95
other pulses	3,152	3,285	102

PAPER VIII AND IX : STATISTICAL QUALITY CONTROL (Practical)

Time : 4 Hours Full marks : 100

a) Answer *all* questions.
b) Figures in the margin indicate full marks.
c) Use of calculating machines is permitted.

1. A process is in control at $\mu = 18$ and $\sigma = 2$. A control chart for $\bar{x}$ based on a sub-group size of $n = 4$ is set up (using 3-sigma limits). If a shift in the mean of 2.5 units occurs, what is the probability of the next $\bar{x}$ falling outside the control limits ? (15)

2. Samples of 5 items each are taken from a process at regular intervals. The sums of the $\bar{x}$ and R values (calculated for each sample) for the first 20 samples are

$$\Sigma\bar{x} = 363.52 \quad \text{and} \quad \Sigma R = 10.70$$

a) compute the control lines for the $\bar{x}$ and R charts

b) assuming the process is in control at the level found in (a), estimate the 3-sigma tolerance limits of the process.

c) if the specification limits were 18.0 ± 0.5, what conclusions would you draw about the ability of the process to produce within specification limits?

d) What percentage of the products will fall outside the specification limits if the process were in control as in (b) ? (30)

3. (a) Find the boundaries for the graphical procedure of the item by item sequential plan for attributes which will satisfy the stipulations.

$$p_1 = 0.03, \qquad p_2 = 0.10$$
$$\alpha = 5 \text{ percent} \qquad \beta = 10 \text{ percent}$$

where the symbols $p_1, p_2, \alpha\ \beta$ have their usual significance.

(b) For the same stipulations, find out the sample size (n) and acceptance number (n) corresponding to the single sampling plan that may alternately be adopted.

(c) Sketch 5-point OC curves for the two plans and compare them. (40)

4. A single sampling variables plan is stated as follows :-

Take a sample of 20 items from the lot, measure characteristic x on each, and if $\bar{x}-K\bar{R} > L$, accept the lot, otherwise reject the lot, where $\bar{x}$ is the arithmetic mean of the 20-measurements, $\bar{R}$ the mean range based on 4 sub-groups each of size 5 and L the lower specification limit for x.

Determine the value of k, to satisfy an AQL value of 2.5 percent associated with a producer risk of 5 percent (x is Normally distributed and you may use the Normal approximation for the sampling distribution of $\bar{x}-k\bar{R}$). (15)

PAPER VIII AND IX : SAMPLE SURVEYS, THEORY (Practical)

Time : 4 Hours Full marks : 100

(a) Attempt *all* questions.
(b) All questions carry equal marks.
(c) Use of calculating machines is permitted.

1. In a sample survey for estimating the number of standards and yield of pepper in a taluka a random sample of 5 clusters, each of 20 'survey numbers' (of fields) was selected from each of the 12 randomly selected villages from the total of 70 villages in a taluka. The total number of clusters in each village along with the number of standards as enumerated in each of the selected clusters are also given in the table. Obtain the unbiased estimate of the total number of standards in the taluka along with its sampling error.

serial number of sample village	total number of clusters in the village	number of standards in each of the five selected clusters				
		I	II	III	IV	V
(1)	(2)	(3)	(4)	(5)	(6)	(7)
1	27	430	402	363	975	389
2	24	586	1234	100	368	344
3	14	1164	546	3060	17244	1274
4	116	693	218	836	1218	575
5	25	191	270	4502	4184	243.
6	118	1036	1333	1179	728	1957
7	147	1555	254	950	382	355
8	26	910	452	129	122	243
9	91	340	0	92	28	340
10	171	57	59	0	0	21
11	86	159	45	242	1075	539
12	88	84	462	147	16	10

Check the mean square between villages and mean square between clusters within villages in the following analysis of variance for the number of standards in the above taluka.

source of variation	D.F.	mean square
between villages	11	407700
between clusters (within villages)	48	235100
between survey number (within clusters)	1140	42400
Total	1199	

Utilising the above table (corrected, if necessary) estimate the relative efficiency of the cluster as a unit of sampling compared with that of the individual survey numbers.

2. With a view to estimating the total cattle population of a given area, twentyfive villages were slected at random out of 12381. Under (X) are given the number of cattle heads in the selected villages in 1945 Under (Y) are given the figures obtained for 1946 The total number of cattle heads in the area according to 1945 census were 13159660.

(X)	(Y)	(X)	(Y)
623	654	1795	1890
690	696	1406	1123
534	530	330	375
293	315	218	212
69	78	160	147
842	640	210	297
475	692	262	401
371	292	204	252
161	210	185	199
298	555	574	564
2045	2110	118	115
1069	592	320	316
706	707		

Estimate the total number of cattle heads in the year 1946 for the given area with and withou the use of previous year's figures. Compare the efficiency of ratio and regression method: of estimation for the above data.

PAPER VIII AND IX : SAMPLE SURVEYS,APPLIED (Practical)

Time : 4 Hours Full marks : 10(

(a) Attempt *all* questions.

(b) Figures in the margin indicate full marks for each question.

(c) Use of calculating machines is permitted.

1. A complete list of households in a village along with some auxiliary information is supplied to you (please see enclosure). It is proposed to draw a systematic sample of 8 households after rearranging the frame of households according to size of household 1, 2, 3-4 and above 4.

(i) Rewrite the column heading of the listing schedule providing suitable space r necessary computational work and for drawing the sample according to the above plan.

(ii) Write out the necessary instructions to the investigator as clearly and precisely possible, for drawing the sample as per plan above. (35)

2. From the data supplied in question 1, draw two systematic samples as per plan in question 1 and build up unbiased estimates of

(i) number of persons reading hindi newspapers in the villages.

(ii) number of persons reading English newspapers in the villages.

Also to estimate the sampling error of these estimates. All computational steps should be shown clearly. (35)

3. Draw a suitable questionnaire for mail enquiry to ascertain the preferences of radio listeners for the timings and items of programme such as news, music, etc. to be relayed by the All India Radio.

[*Note :* Necessary instructions for filling up the questionnaire should also be appropriately given]. (30)

List of Households

1. State : J and K; 2. District : D; 3. Tehsil : X; 4. Village : Y.

house number	household serial number	name of head of household	household size	newspaper readers: Hindi	newspaper readers: English
(1)	(2)	(3)	(4)	(5)	(6)
1	1	a	5	2	1
2	2	b	3	1	1
3	3	c	5	-	-
	4	d	5	1	1
4	5	e	5	2	1
5	6	*f*	4	1	-
6	7	g	9	1	-
7	8	h	6	-	2
8	9	i	4	4	2
9	10	j	1	-	1
	11	k	4	1	1
	12	l	7	3	2
10	13	m	3	1	1
	14	n	4	2	1
11	15	o	5	3	-
12	16	*p*	9	4	3
13	17	q	3	1	-
14	18	r	5	2	1
15	19	s	7	4	1
	20	t	6	2	2
16	21	u	4	1	-
	22	v	4	1	-
	23	w	5	2	2
17	24	x	4	2	2
18	25	y	11	1	4
19	26	z	5	1	-
20	27	aa	3	1	-
	28	bb	5	3	1
21	29	cc	21	16	2
22	30	dd	5	1	1
	31	ee	3	1	-
23	32	ff	7	2	-
24	33	gg	2	1	1
25	34	hh	10	4	4
26	35	ii	6	2	2
27	36	jj	4	1	1
	37	kk	4	1	-

PAPER VIII AND IX : DESIGN OF EXPERIMENTS,APPLIED (Practical)

Time : 4 Hours Full marks :

(a) Attempt *any two* questions.
(b) All questions carry equal marks.
(c) Use of calculating machines is permitted.

1. Milk filter disks—as the name suggests—are used to strain milk shortly after m ing. They are designed to remove dirt and debris, but not butter fat.

The following experiment was designed to compare the speed of flow of mil different makes of this disk. The test procedure was to make 'Farm tests', wherein 3 d were tested on a farm at a single milking. Experience had shown that not more tha disks could be satisfactorily tested at one time and that there were large differences not c between farms but also between different times on the same farm. Accordingly, it thought advisable to regard the 3 tests made at one time as a block and provide in design for the elimination of shifts between blocks. Attention was also given to the orde which the disks would be tested at one time since the order of pouring could also pos affect the speed of flow.

The layout corresponds to a two-associate partially balanced incomplete bl alys design with the following parameters :

$b = v = 16,\ r = k = 3,\ n_1 = 9,\ n_2 = 6,$

$\lambda_1 = 0,\ \lambda_2 = 1.$

$p'_{11} = 4,\ \ p'_{12} = p'_{21} = 4,\ \ p'_{22} = 2$

$p^2_{11} = 6,\ \ p^2_{12} = p^2_{21} = 3,\ \ p^2_{22} = 2.$

The association scheme may be specified by the following Latin square :-

A (1)	B (2)	C (3)	D (4)
D (5)	A (6)	B (7)	C (8)
C (9)	D (10)	A (11)	B (12)
B (13)	C (14)	D (15)	A (16)

Two treatments are first associates if they occur in the same row or column or corresp to the latin letter. They are second associates otherwise. Thus for treatment 10 the fi associates are 2, 4, 5, 6, 9, 11, 12, 14 and 15. The rest are second associates.

While randomising care was taken to ensure that every treatment occurred once in e ialy order. This is essentially an extention of the Youden square principle to a PBIBD. ect.

The experimental plan and observations are recorded below. The observati ns are the number of seconds required for a standard quantity of milk to pass through the d Treatments are indicated in parenthesis.

PLAN AND YIELDS IN THE MILK FILTER DISK EXPERIMENT

Farm test number	order of pouring		
	first	second	third
(1)	(2)	(3)	(4)
1	451 (10)	457 (7)	343 (16)
2	260 (11)	418 (8)	320 (13)
3	464 (12)	317 (5)	315 (14)
4	306 (9)	462 (6)	291 (15)
5	381 (13)	597 (4)	491 (6)
6	362 (14)	325 (1)	449 (7)
7	292 (15)	402 (2)	576 (8)
8	431 (16)	477 (3)	394 (5)
9	329 (7)	261 (9)	430 (4)
10	389 (8)	313 (10)	272 (1)
11	368 (5)	244 (11)	447 (2)
12	398 (6)	517 (12)	354 (3)
13	490 (2)	311 (16)	278 (9)
14	467 (3)	429 (13)	486 (10)
15	735 (4)	642 (14)	474 (11)
16	402 (1)	380 (15)	589 (12)

..alyse the data and give your comments. [You may use the following computations :-

Total sum of squares (corrected) = 521,099

sum of squares due to order (columns) = 767

sum of squares due to blocks (ignoring treatments) = 241,609].

.. A certain rocket program called for firing at three nominal slant ranges (R_1, R_2, R_3) with three levels of propellant temperature (T_1, T_2, T_3). The dependent variable the azimuth co-ordinate of 'miss distance'. Three groups of three rounds each were for each set of conditions spread over three days (one group for each day). The coded of miss distance are given below.

Slant Range

	(R_1) Days			(R_2) Days			(R_3) Days		
	1	2	3	1	2	3	1	2	3
(T_1)	−10	−22	−9	−5	−17	−4	11	−10	1
	−13	0	7	−9	6	13	−5	10	20
	14	−5	12	21	0	20	22	6	24
(T_2)	−15	−25	−15	−14	−3	14	−9	8	14
	−17	−5	2	15	−1	5	−3	−2	18
	7	−11	5	−11	−20	−10	20	−15	−2
(T_3)	−21	−26	−15	−18	−8	0	13	−5	−8
	−23	−8	−5	5	5	−13	−9	−18	3
	0	−10	0	−10	−10	3	−13	−3	12

..alyse the data to determine whether slant range and propellant temperature have any ..ct on miss distance and whether there is any interaction between them. State the assump..ns under which your analysis is valid. Note that

$$R_1 < R_2 < R_3 \quad \text{and}$$
$$T_1 < T_2 < T_3.$$

3. An experimental station, as a routine practice was conducting a varietal trial a certain crop involving 3 different varieties V_1, V_2 and V_3 arranged in 4 randomised blo of 3 plots each. After the seedlings had appeared, it was decided to combine one m objective in the same investigation and to study simultaneously the effect of a cert fertiliser (usually applied late) on the crop yield. Accordingly the experimental c plots where the seedlings had already appeared, were divided into 3 parts each and fertiliser was applied at 3 levels (0, 2 and 4 cwt per acre). The assignment of the leve the part was done at random independently for each crop plot.

The plan and the yields are given below. The levels are indicated by (0), and (2) respectively.

Block 1	(2) 118		(0) 111		(0) 117	
	(0) 140	V_3	(1) 130	V_1	(1) 114	V_2
	(1) 105		(2) 157		(2) 161	
Block 2	(2) 104		(0) 74		(1) 103	
	(0) 70	V_3	(1) 89	V_1	(0) 64	V_2
	(1) 89		(2) 81		(2) 132	
Block 3	(1) 108		(1) 124		(0) 61	
	(2) 126	V_2	(2) 121	V_3	(1) 91	V_1
	(0) 70		(0) 96		(2) 97	
Block 4	(2) 109		(0) 80		(1) 90	
	(0) 63	V_3	(2) 94	V_2	(2) 100	V_1
	(1) 70		(1) 82		(0) 62	

Analyse the data to test for varietal differences. Do these varietal differe depend upon the level at which the fertiliser is applied? Is it possible to make s recommendations regarding the optimum level at which the fertiliser should be appli Make specific suggestions.

4. You are given below the data from a 2^3 factorial experiment on soyabeans in 4×4 Quasi-latin squares. The factors were limestone (A), phosphorus (B) and pot (C) and in each case the lower level implies no application of the treatment.

Plan and yields of soyabean (in bushels per acre).

Square I				Square II			
C 58	(1) 62	abc 60	ab 59	c 62	a 65	abc 69	b 63
a 66	ac 61	b 60	bc 53	(1) 62	ac 63	bc 59	ab 69
abc 64	bc 61	(1) 63	a 59	abc 66	b 57	(1) 64	ac 67
b 58	ab 62	ac 67	c 62	ab 66	bc 53	a 58	c 58

(i) Identify the effects that are confounded with the rows and columns of squares.

(ii) Analyse the data and test the individual main effects and the interactions.

(iii) Write a report on your findings suggesting the best factor combination ar the ones studied, if any.

PAPER VIII AND IX : MATHEMATICAL THEORY OF SAMPLING DISTRIBUTION (Practical)

Time : 4 Hours Full marks : 100

(a) Attempt *all* questions.

(b) All questions carry equal marks.

1. The following table gives the mean values, pooled estimates of standard deviations and the correlation coefficient computed from observations on lengths of Femur and Humerus obtained on 27 Anglo Indians and 20 Indians.

	sample size	mean length of		pooled standard deviation (45 d.f.)
		Femur	Humerus	
Anglo Indians	27	460.4	335.1	23.7
Indians	20	444.3	323.2	18.2
Product moment correlation = 0.4777.				

(i) Test for differences in the individual measurements separately.

(ii) Test by using an appropriate criterion, the joint hypothesis that the mean values for Femur and Humerus are same for both the communities.

(iii) What inference can you draw on the differences between the communities based on the tests (i) and (ii) ?

2. From the table of random numbers supplied, generate a sample of size 10 from bivariate normal population having the mean values 20 and 30 and the variance covariance matrix

$$\begin{pmatrix} 2 & 1 \\ 1 & 4 \end{pmatrix}$$

state the procedure you adopt clearly and indicate all the necessary computations on your answer sheet. Examine on the basis of the sample you have generated whether such a sample could reasonably arise from the given bivariate population by testing the significance of the deviations of the observed mean values, variances and covariance from the assigned values. You may, if you like, carry out separate tests for mean values, variances and co-variance.

Explain if there is any difficulty in interpreting the results of multiplicity of tests you will be carrying out.

PAPER VIII AND IX : VITAL STATISTICS AND POPULATION STUDIES (Practical)

Time : 4 Hours Full marks : 100

(a) Attempt *all* questions.
(b) Use of calculating machines is permitted.
(c) Logarithm Tables will be supplied.

1. The number of children in the age-group 0-4 and the age distribution of women in the child-bearing period according to the 1951 census are given below for each of the following two districts in Bombay State.

	Greater Bombay	Ratnagiri
(a) Number of children in the age group 0-4	28026	22243

(b) Age distribution of women in the child-bearing period

age-group	number of women (in thundreds)	
	Greater Bombay	Ratangiri
15–19	12302	7924
20–24	11100	7900
25–29	10023	7388
30–34	9394	6923
35–39	7324	5786
40–44	4180	5537
all child-bearing ages	54323	41458

Calculate the child-women ratio of the two districts and estimate how much of the difference between the two ratios is accountable by the difference in the age distribution of women in the child-bearing period. You may assume that the specific fertility rates in the age groups 15-19, 20-24....40-44 are for both the districts identical and proportional to 1 : 7 : 7 : 6 : 4 : 1.

2. Assuming that

$$\frac{1}{P_t} = \frac{a}{P_{t-1}} + b, \text{ where } a \text{ and } b \text{ are}$$

constants and P_t stands for population at time t, estimate the population of India in 1961, given that the population of India in the last three censuses were

1931	276 million
1941	313 million
1951	357 million

3. The age-specific death rates of an area during a particular years are given below.

age group x to $x+n$	age-specific death rate $_nm_x$
0– 1	.038857
1– 5	.001533
5–10	.000732
10–15	.000734
15–20	.001421
20–24	.001937
25–29	.001981

Estimate the following : (a) the probability that a person exactly 5 years of age will survive the next ten years, (b) the probability that a person in the age-group 5-9 will survive to age-group 15-19.

4. The following are the age-specific fertility rates (female births only) and life table population (females) for Japan and Taiwan. The age distribution of women in the child-bearing period is also shown for Japan only.

		Japan		Taiwan	
age-group	number of women	age-specific fertility rates (female births)	life-table population (females)	age-specific fertility rates (female births)	life-table population (females)
15–19	4250101	6.62	451996	30.45	415090
20–24	3889727	80.35	446257	122.50	402807
25–29	3363222	118.12	438345	147.65	388127
30–34	2841997	87.34	429707	133.45	374303
35–39	2671968	52.19	420647	95.75	361177
40–44	2284025	17.95	410550	55.45	347835
45–49	1985701	1.06	398390	14.70	333295

Calculate the Net Reproduction Rates for these two countries. Estimate the Net Reproduction Rate for Japan assuming that the age-specific fertility rates for Japan are unknown but are proportional to those in Taiwan.

PAPER VIII AND IX : PROBIT ANALYSIS (Practical)

Time : 4 Hours Full marks : 100

(a) Attempt *any two* questions.
(b) All questions carry equal marks.
(c) Credit will be given for neat, brief answers, tabular arrangement of computation and graphical presentation of data.
(d) Use of calculating machines is permitted.

1. If the proportion expected to respond at dosage x is given by

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

What tables of weights etc. would you require to facilitate computation of maximum like hood estimates of α and β from a quantal type bioassay? Prepare specimens of these tabl by computing, say ten entries of each such table. Also write down instructions for usi these tables.

2. At each concentration of a lethal preparation, two samples of 100 each were test to determine the percentage kill. Using this information, test whether assumptions need for computing the probit-regression line are valid or not.

NUMBER KILLED OUT OF 100 TESTED AT EACH DOSE

log concentration	sample 1	sample 2
2.16	93	89
1.98	76	80
1.67	52	50
1.09	41	45

3. The table below gives the results of an insulin assay for a standard preparati and a test preparation where n is the number of rats in batch and r the number affecte Using the logarithm of the dose as dose-metameter , fit parallel probit regression lin Obtain the relative potency of the test preparation and its standard error.

standard preparation		
dose	n	r
3.4	33	0
7.0	38	11
10.5	40	18
18.0	31	23
28.0	30	27

test preparation		
dose	n	r
6.5	40	2
10.0	30	10
14.0	40	18
21.5	35	21
29.0	37	27

4. The table below shows the total number of insects (n) and number killed (r) und different concentrations of Pyrethrin. It is known that some test subjects would die ev without any poison.

Using this data obtain the maximum likelihood estimate of the dosage requir to kill 80 per cent of an insect population in which natural mortality rate is 5 percent.

concentration	n	r
0.5	30	1
1.0	30	15
2.0	30	27
4.0	30	28
control	30	2

Use log concentration as dose metameter.

PAPER VIII AND IX : PSYCHOLOGY AND EDUCATION (Practical)

Time : 4 Hours | Full marks : 100

(a) Attempt *any three* questions.
(b) All questions carry equal marks.
(c) Use of statistical tables and calculating machines is permitted.

1. Convert the raw scores given in the distribution below to percentile scores. Express each raw score as a percentile score.

score	frequency	score	frequency
0	1	12	168
1	3	13	153
2	5	14	133
3	12	15	125
4	20	16	106
5	35	17	69
6	52	18	45
7	59	19	39
8	89	20	19
9	109	21	12
10	114	22	6
11	147	23	2
		24	1

2. Scores are given for 16 persons on two tests, A and B, below. Rank each set of scores and compute the rank-order correlation coefficient.

serial number	score on test A	score on test B	serial number	score on test A	score on test B
1	15	12	9	14	12
2	26	25	10	13	13
3	9	1	11	15	14
4	8	18	12	19	25
5	12	16	13	13	15
6	22	26	14	17	10
7	10	11	15	21	22
8	16	16	16	23	24

3. Proportion passing and item-test correlation are given below for 35 test items. Plot the items on a graph with proportions passing as the *x*-axis (abscissa) and item-test correlation as the *x*-axis (ordinate). Two forms of the test are to be made up consisting of 10 items each. The mean proportion passing and the mean item-test correlation of the two forms should be as equal as possible. From the plot of items, select 10 pairs of items with similar

values on the x- and y-axes. Write the serial numbers of the selected items for each for and compute mean proportion passing and mean item-test correlation for each form.

serial number	proportion passing	item-test correlation	serial number	proportion passing	item-test correlation
1	.78	.33	19	.73	.39
2	.67	.23	20	.64	.41
3	.72	.29	21	.58	.48
4	.70	.45	22	.41	.57
5	.70	.17	23	.65	.56
6	.32	.36	24	.73	.56
7	.47	.52	25	.40	.59
8	.36	.41	26	.64	.49
9	.70	.33	27	.57	.50
10	.81	.65	28	.71	.65
11	.14	.45	29	.62	.40
12	.78	.33	30	.57	.34
13	.72	.54	31	.59	.36
14	.70	.55	32	.52	.48
15	.37	.53	33	.50	.62
16	.42	.43	34	.52	.56
17	.66	.43	35	.42	.15
18	.81	.34			

4. The proportion passing each item of a 36 item test is given below. Using th·lity proportions, compute the reliability of the whole test consisting of 36 items using Kud·the Richardson formula (20).

item number	proportion passing	item number	proportion passing
1	.96	20	.42
2	.61	21	.80
3	.86	22	.76
4	.77	23	.85
5	.92	24	.77
6	.90	25	.86
7	.84	26	.87
8	.44	27	.86
9	.72	28	.71
10	.69	29	.62
11	.86	30	.63
12	.68	31	.56
13	.69	32	.55
14	.79	33	.39
15	.66	34	.43
16	.42	35	.56
17	.83	36	.52
18	.81		
19	.63		

STATISTICIAN'S DIPLOMA EXAMINATION, SEPTEMBER 1960

PAPER I : THEORETICAL STATISTICS, GENERAL

Time : 4 Hours | Full marks : 100

(a) Attempt *any five* questions.
(b) All questions carry equal marks
(c) Use of calculating machines is not permitted.
(d) Figures in the margin indicate full marks.

1. (a) Obtain an interpolation formula when arguments are known to be at unequal intervals. (10)

(b) Apply Lagrange's formula to find $f(4)$, given that

$$f(1) = 3, f(2) = 9, \quad f(3) = 27, \quad f(5) = 243. \qquad (10)$$

and explain why the result differs from that obtained by completing the series of powers of 3.

2. (a) A and B play a match to be decided as soon as either has won two games. The probability of either winning a game is [illegible], and that of a game being drawn is [illegible]. What is the probability that the match is finished in 5 or less games ? (10)

(b) From a population of n elements a sample of size r is taken. Find the probability that none of N prescribed elements out of n will be included in the sample, assuming the sampling to be (i) without replacement, (ii) with replacement. Compare the numerical values for the two methods when (i) $n = 100$ $r = N = 3$, and (ii) $n = 100$, $r = N = 10$. (10)

3. (a) Define Student's t. Explain how t can be used to test

(i) that the mean has a given value, (10)

(ii) two means are equal.

(b) Describe some of the essential features of the normal probability distribution.

A sample of 900 values has mean 3.47. Can it be reasonably regarded as a sample from a normal population with mean 3 and standard deviation 2? (10)

4. Define the partial correlation $r_{12.3}$ and the multiple correlation $R_{1(23)}$. Show that $R_{1(23)}$ is not numerically less than any of the correlation coefficents r_{12} and r_{13}. What is the significance of $R_{1(23)}$ being (i) zero and (ii) unity ? (10)

If $r_{23} = 0$ prove that $R^2_{1(23)} = r^2_{12} + r^2_{13}$. (10)

5. (a) Show that any linear function of a set of independent normal variates is normally distributed. Hence obtain the distribution of the mean in random samples of size n from a univariate normal population with mean $= \mu$ and S.D. $= \sigma$. (10)

(b) $\log_{10} x$ is normally distributed with mean 7 and variance 3. $\log_{10} y$ is normally distributed with mean 3 and unit variance. If the distributions of x and y are independent, find the probability of

$$1.202 < \frac{x}{y} < 83180000$$

(Given $\log_{10} 1202 = 3.08$, $\log_{10} 8318 = 3.92$). (10)

6. (a) Explain the terms (i) probability density function (ii) distribution function, (iii) moment generating function, (iv) cumulants. (8)

(b) Find the distribution function, moment generating function, the mr variance and the rth cumulant of the distribution with probability density function

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

PAPER II : APPLIED STATISTICS, GENERAL

Time : 4 Hours Full marks :

(a) Attempt *any five* questions.
(b) All questions carry equal marks

1. *Either*,

It has been stated that the official estimates of agricultural production tend to under-estimates. Comment on this statement in relation to the statistics of foodgr production in India.

You are required to design a sample survey for the estimation of foodgrains product in India. Give adequate details of the sample design and the related estimation proced which you would recommend. It is known that non-sampling errors are likely to be v serious in the execution of such type of surveys. List carefully the points on which y would concentrate to minimise these errors.

Or,

A company manufacturing toilet soap is interested in estimating the volume of sa in a big city of four different brands (A, B, C and D) of toilet soap. Give a plan for sample survey to collect the information in which you should discuss the following points :

i) the frame and the sampling unit,
ii) the questionnaire or schedule,
iii) the sampling design,
iv) control of non-sampling errors,
(v) estimation formulae,
vi) sampling errors of estimates,
vii) a rough estimate of cost.

2. How will you design an enquiry for the construction of cost of living index numbe Give your views on reference period and method of averaging for the determination weighting pattern. In such type of enquiries either tenement sampling or pay r sampling is used. Do you have any preference ?

You have made a purposive selection of a number of centres in the country for t construction of an all India index. There are several ways of combining the centre-w indices. Examine the relative merits of these methods.

3. (a) Discuss critically, giving examples, the various techniques for increasing t precision of an experiment without narrowing down the scope of the experiment.

(b) Your advice has been sought in planning an experiment on maize to be so in lines 3′ × 1′. The treatments to be tried are three dates of sowing in conjunction w four levels of nitrogen. Specify the design and the treatments which you are going to recommend and prepare a sketch plan for a single replication of the recommended design. Choose your scale to indicate the shape and size of the experimental plot and the block in the sketch plan.

4. (a) What is an Engel Curve and how will you estimate it on the basis of family budget data ?

(b) Elasticities of demand and supply are worked out both from cross sectional data and from time series data. How are these elasticities estimated?

5. (a) Describe the 0—A—B—AB system of classification of human blood and the genetics of its inheritance.

(b) Discuss the uses of this kind of blood-grouping in problems of (i) anthropology, (ii) blood transfusion and (iii) doubtful paternity.

6. (a) Give your critical comments on FACTOR-ANALYSIS as a psychometric method.

(b) Derive the fundamental equation of factor analysis :

$$R = F F'$$

where R is the reduced correlation matrix of test scores and F the matrix of factor loadings.

(c) Describe the centroid method of factoring and show how the above equation is used in this method.

7. (a) Describe the method you would adopt to forecast the population of India in 1970 and its distribution by sex and age.

(b) What statistics do you need for this purpose and from what sources would you collect them ?

(c) Give your critical comments on the reliability of such statistics.

PAPER III : STATISTICAL INFERENCE

Time : 4 Hours　　　　Full marks : 100

(a) Attempt *any four* questions.

(b) All questions carry equal marks.

1. (a) In a series of Bernoullian trials the probability π of a success has one of three alternative values, $\frac{1}{4}$, $\frac{1}{2}$, and $\frac{3}{4}$, the a priori probabilities of these values being in the ratio 1 : 2 : 1 respectively.

In an actual experiment with 10 trials, 7 successes were observed.

Use Bayes' theorem to select the best value of the probability π.

(b) Discuss briefly why this method cannot be used in general for the estimation of unknown parameters.

(c) The experimental result being the same as in (a), and if nothing is known a priori about π, obtain its maximum likelihood estimate and show that this estimate has the minimum attainable variance.

2. State precisely and give an outline of the proof of Rao-Blackwell's theorem wh tells us how to form an improved estimator given an unbiased estimator and a suffic estimator for a population parameter.

Let $Y_1 < Y_2 < Y_3 < Y_4 < Y_5$ be the order statistics of size 5 from the unif distribution having p.d.f. $f(x, \theta) = 1/\theta$ for $0 < x < \theta$, $0 < \theta < \infty$, zero elsewhere. S that Y_5 is a sufficient statistic and $2Y_3$ is an unbiased statistic for θ. Determine the jl p.d.f. of Y_3 and Y_5. Compare the variances of Y_5, $2Y_3$ and $E[2Y_3/Y_5]$.

3. (a) Outline Neyman and Pearson's theory of construction of most powerful (M test of a simple hypothesis H_0 against a simple alternative hypothesis H_1. Show that ev MP test is unbiased.

(b) Obtain the most powerful critical region to test the hypothesis $\sigma = \sigma_0$ ag an alternative $\sigma = \sigma_1$ regarding the standard deviation of a normal population $N(0,$ Examine if the critical region is also uniformly most powerful.

4. Define a confidence interval for a parameter. How can the best interval selected ?

Obtain the expression for a 95 percent confidence interval of

(a) the logarithm of the variance of a normal population

(b) the ratio of variances of two correlated variables obeying a bivariate nor distribution.

5.(a) Define the 'consistency' of estimates.

(b) Show that T_n is a consistent estimate of if $E(T_n) \rightarrow \theta$ and Var $(T_n) \rightarrow$ $n \rightarrow \infty$. Examine the consistency of

(i) mean and median for the estimation of the population mean of a symmetr distribution

(ii) the sample variance $s^2 = \sum_{i=1}^{n} (x_i - \bar{x})^2/n$ for the estimation of the popu tion variance.

(c) State a set of conditions under which the likelihood equation for the estimat of the parameter θ of continuous frequency function $f(x, \theta)$ has a consistent solution.

6. Show (without giving detailed proofs) how the β-distribution can be utilized n making the following tests of hypotheses regarding multivariate normal distribution :-

(a) If $m_1, \ldots, m_p$ denote the population means of the p variables, they have spe fied values $m_{10}, \ldots, m_{p0}$, irrespective of the dispersion matrix.

(b) The variable x_1 is independent of the variables $x_2, \ldots, x_p$.

PAPER VI : (Practical)

Time : 6 Hours

Full marks :

(a) Attempt *any five* questions.

(b) Figures in the margin indicate full marks.

(c) Use of calculating machines is permitted.

1. The following table shows the two-way distribution of 411 informants of a ran sample survey according to their educational attainments and their first preference to ferent kinds of medical treatment.

education	kind of medical treatment				
	allopathic	homeopathic	ayurvedic	indigenous	total
(1)	(2)	(3)	(4)	(5)	(6)
iterate	74	16	3	1	94
to primary standard	76	16	7	3	102
n-matric	87	17	10	-	114
der-graduate	56	18	3	1	78
aduate and above	15	6	2	-	23
Total	308	73	25	5	411

Is there any significant association between standard of education and first prefer-ce to kind of medical treatment ? (20)

2. *Either,*

The following table gives the results of a completely randomised experiment for the yield of wheat (mds. per acre) under uniform conditions in 1957 and under three treatments T_1, T_2, T_3 in 1958, the treatments being applied to groups of three plots chosen at random.

treatments	yields		
T_1	17.5	19.3	20.5
	(16.7)	(18.0)	(19.4)
T_2	21.6	19.5	20.7
	(19.7)	(20.2)	(18.8)
T_3	18.6	19.5	18.3
	(19.1)	(18.5)	(17.7)

The figures in brackets are yield of the field under uniform conditions in 1957.

From the data examine if there are any treatment differences eliminating variability in the normal yield rates if possible. (20)

Or,

With a view to using regular steps by investigators as a unit of measure of length, h experiment was conducted to standardise the steps. In this experiment 6 investigators ere asked to pace a distance of 575 ft. five times . The mean length of steps in ft. by ch investigator in each of the five replications are given below :-

worker number	average length of steps in feet for each traverse				
	1	2	3	4	5
1	2.68	2.59	2.67	2.68	2.53
2	2.44	2.40	2.45	2.54	2.46
3	2.71	2.70	2.66	2.71	2.71
4	2.99	2.95	3.00	2.99	2.94
5	2.52	2.37	2.50	2.48	2.47
6	2.80	2.70	2.76	2.69	2.67

Test whether there is any significant difference between workers in respect of len
of steps. (

3. (a) In an examination out of 50 candidates, the percentage of students who
more than 35 percent and 60 percent of the total marks are respectively 60 percent a
8 percent. Assuming that the distribution of marks is normal, estimate the minimum p
marks if on the average 75 percent of the students are to pass. (1

(b) Considering the 50 students examined as a sample from a population of studen
examine whether the observed percentages of 60 percent and 36 percent are in agr
ment with the hypothesis that the marks in the population have a normal distribution w
mean 40 percent and standard deviation 12 per cent. (1

4. In an experiment carried out for purposes of standardising the garment si
of women, the following correlation coefficients were obtained from measurements tak
on 50 women.

x_1 = bust girth, x_2 = back width, x_3 = hip girth

$r_{12} = 0.78$, $r_{13} = 0.87$, $r_{23} = 0.71$

(i) Calculate $r_{12.3}$ and $R_{1.23}$ and test for their significance. (1

(ii) Test whether inclusion of x_3 in the multiple linear regression equation of
on x_2 and x_3 improves the prediction of x_1. (1

5. The following table shows the estimated rural population (in millions) of a count
on 1st March each year.

x = time in years lapsed since 1st March 1951,

$y = p_x$ = population at time x.

x	1	2	3	4	5	6	7	8
y	301.5	305.2	309.0	313.1	317.3	321.8	326.4	331.3

(a) Fit a curve of the form $P_x = ab^x$. (1

(b) Estimate the population on 1st September 1956 and 1st March 1961. (

6. The following table shows the cumulative percentage expenditure on medici
(y) against cumulative percentage of population (x) obtained from a survey after arrangi
the households according to increasing order of per capita household expenditure.

x	0	10	20	30	40	50	60	70	80	90	100
y	0.0	0.8	1.8	2.8	4.3	7.0	11.0	17.0	26.6	51.4	100.0

(a) Find by suitable method of interpolation the value of y for $x = 39$ using di
ferences upto 3rd order. What are the assumptions involved in the procedure you adop
(1

(b) Find by Simpson's $\frac{1}{3}$ rule of numerical quadrature the value of $\int_0^{100} y\, d$

at are the assumptions involved in this procedure ? (

PAPER VII : (Practical)

Time : 6 Hours | Full marks : 100

(a) Attempt *any four* questions.

(b) All questions carry equal marks

(c) Use of calculating machine is permitted.

1. For 20 subgroups of 5 measures each, the values of $\bar{x}$ and R are given below with respect to a quality characteristic on a product having a specification of 0.5025±0.0010. [The values given are the last two figures of the measurement, i.e. 24.0 means 0.50240].

sub-group	$\bar{x}$	R
1	24.0	5
2	21.6	5
3	20.8	3
4	23.0	4
5	25.0	6
6	22.2	2
7	23.0	5
8	22.6	12
9	23.8	17
10	25.8	7

sub-group	$\bar{x}$	R
11	25.8	5
12	25.7	5
13	24.0	15
14	24.5	5
15	24.8	8
16	25.1	6
17	25.6	6
18	26.0	4
19	26.5	9
20	26.9	7

(a) Set up appropriate control charts.

(b) Is the production process in control ?

(c) If not, (i) what evident is there of lack of control ?

and (ii) suggest values for the two control limits to be used for succeeding sub-groups.

(d) If the process is in control with respect to standards (parameters) used under (c) (ii) above, will the process be able to meet the specification ?

2. (a) The following information has been taken form the Life Table for white females in the U.S.A. You are required to fill in the blanks in the table.

x	l_x	d_x		m_x	L_x	T_x	e_x
0	100000	—	—	0.038620	—	6728965	—
1	—	—	—	0.004329	—	6632143	
2	—	—	—	0.002202	—	6536177	
3	—	—	—	0.001611	—	6440493	—
4	—	—	—	0.001281	.95367	6344988	

(b) Under a certain mortality table the mortality between ages 25 and 40 is such that the probability that a life aged 25 will die in the n-th year after the attainment of age 25 increases as n increases from 1 to 15 in arithmetical progression. Given that l_{25} = 930440, l_{38} = 838244 and l_{40} = 822770, find the average age at death of persons dying between ages 25 and 38.

3. *Either,*

The layout and yields of a cacao manurial experiment using three fertilizers :

A : No manure

B : 1½ lb. superphosphate per tree

C : 3 lb. superphosphate per tree

in three 3×3 Latin squares are given below.

Analyse the data and recommend the best dose of the fertilizer.

LAYOUT AND YIELDS IN PODS PER TREE FROM $\frac{1}{15}$ ACRE PLOTS

square 1			square 2			square 3		
B 41	C 25	A 51	C 27	B 28	A 3	A 11	C 15	B 17
A 20	B 32	C 24	A 4	C 17	B 9	B 24	A 14	C 33
C 22	A 12	B 21	B 22	A 4	C 17	C 22	B 20	A 15

Or,

An experiment was conducted to test whether hormone treatment on seeds bef planting has effect on yield. Six pots were taken and in each eight seeds were planted random each of the four pairs of the seeds in a pot having been subjected to a differ dose of hormone, out of four doses. The yields from each plant in grammes are given bel

Analyse the data with a view to finding the best dose, if any, of hormone for seed

Pot \ Dose	D_1		D_2		D_3		D_4	
	P_1	P_2	P_1	P_2	P_1	P_2	P_1	P_2
1	17.7	19.2	22.5	23.2	20.9	23.2	16.5	18.2
2	21.8	21.3	27.2	26.8	26.1	27.2	23.6	21.1
3	16.9	16.6	20.4	20.2	19.3	20.0	17.4	15.3
4	21.5	20.9	25.8	23.7	22.2	22.9	22.8	18.0
5	22.6	24.5	29.5	28.1	29.3	28.4	24.2	26.6
6	17.8	19.3	26.5	24.2	20.7	23.1	20.6	20.5

4. *Either,*

For a two stage sample with replacemeno the variance of an estimate of t population average is given by

$$V = \frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{mn}$$

where m is the number of first stage units selected and n is the number of second stage un selected in each first stage unit in the sample and σ_1^2 and σ_2^2 are the two stage varian in the population. Results of a pilot survey are available in the form of an analysis of variance table. 20 first stage units with equal number of second stage units in each first stage units are sampled.

Analysis of variance table

	D.F.	S.S.
between first-stage units	19	467.4
within first-stage between second-stage units	40	484.5
total	59	951.9

The total cost for carrying out the survey is given by

$$T = a+bm+cmn$$

the numerical values of the parameters being a = Rs. 500, b = Rs. 17 and c = Rs. 3.

Estimate the values of m and n which minimises V for the total cost of Rs. 4000 for the survey.

Suggest the graphs which will enable one to read the minimum attainable variance V and the corresponding optimum values of m and n for a given cost T in a specified range.

Or,

(a) In a paddy field, the grain (y) and the grain plus straw (x) were weighed for each of a large number of sampling units selected at random over the field. The total produce (grain plus straw) of the entire field was also weighed. The following values were obtained

$$C_{yy} = 1.24 \qquad C_{yx} = 0.81 \qquad C_{xx} = 1.20$$

(where C_{yy} and C_{xx} are the squares of the coefficients of variation of y and x respectively and C_{yx} is the analogous relative covariance).

(i) Compute the gain in precision obtained by estimating the grain yield of the field from the ratio of grain to total produce instead of from the mean yield of grain per unit.

(b) A population 3000 primary units is divided into 10 strata each containing 60 larger units consisting of 5 primary units. The analysis of variance of the population for a certain characteristic is as follows :-

	d.f.	m.s.
between strata	9	31.2
between larger units within strata	590	3.1
between primary units within larger units	2400	1.5

Ignoring finite population correction, is the relative precision of the large to the small unit greater with simple random sampling than with stratified random sampling with proportional allocation ?

5. (a) The following table gives the tonnage of shipping entered with cargoes . U.K. ports. Analyse the figures of 1931 into parts due to the influence of trend, seaso and random variation respectively

Tonnage of shipping entered with cargoes at U.K. ports (00,000 tons)

year	1st quarter	2nd quarter	3rd quarter	4th quarter
1927	134	153	163	151
1928	135	154	159	155
1929	132	159	177	159
1930	139	166	176	156
1931	133	153	167	150

(b) The following table gives interim index of industrial production (Avera 1946-100).

group	weights	group indices	
		June 1947	June 1948
China and earthenware	4	122	149
glass	6	113	126
brick, cement etc.	21	138	160
metal : ferrous	38	110	120
: non-ferrous	18	125	119
precision instruments	8	120	123
leather goods etc.	6	97	89
manufacture of wood and cork	25	104	102
paper and printing	39	117	106
other manufacturing industries	19	137	150
Total	184		

(i) Calculate index numbers of production for the above groups combined 1947 and 1948.

(ii) There are 21 groups in the complete index with total weights 1000. If t complete index for June 1947 was 114, what was the index at that date for the remaining groups ?

6. The inter-correlation of four characters ascribed to success in examinations statistics, based on test results on 142 candidates scoring more than 60 per cent marks statistics, are given below.

	1	2	3	4
1. previous academic record	1.00	.	.	.
2. reasoning ability	0.85	1.00	.	.
3. mathematics knowledge	0.61	0.62	1.00	.
4. ability to express	0.57	0.35	0.46	1.00

Using Hotelling's method of principal components, estimate the factor whi accounts for the greatest variability

Also find out what percentage of the total variability is accounted by this factor.

PAPER IV AND V : ECONOMIC STATISTICS (Theoretical)

Time : 4 Hours Full marks : 100

(a) Attempt *any five* questions.
(b) All questions carry equal marks.

1. Discuss methods of measuring capital formation and the relevance of such measures in the developmental planning of an under-developed economy like India.

2. Discuss the method of construction and uses of an inter-industry table.

3. Discuss the method of construction and use of an index number of agricultural production in the context of the various improvements in agricultural statistics that have taken place in India.

4. Describe an econometric growth model appropriate for planning the development of an under-developed economy in a country like India.

5. Explain what is meant by Linear Programming. Illustrate by means of an example applications of this technique to Indian conditions.

6. Discuss the various improvements in statistics that have taken place recently and their effect in the computation of national income, discussing in particular how statistical increases in national income are to be seggregated from real increases in national income.

7. Describe some important techniques of demand analysis and show what data are needed and are available for an analysis of demand and price situation of commerical crops in India.

8. Describe methods of construction of the index of consumers' price for an industrial centre in India and indicate the appropriateness of its use in wage regulation.

PAPER IV AND V : STATISTICAL QUALITY CONTROL (Theoretical)

Time : 4 Hours Full marks : 100

(a) Attempt *any four* questions
(b) Figures in the margin indicate full marks

1. (a) Explain the following terms with reference to an acceptance sampling plan

OC curve, ASN, AQL, AOQL, LTPD, Indifference Quality, Producer's risk, Consumer's risk. (16)

(b) Comment on the salient points of Dodge Romig, Mil-Std and Philips sampling plans. (9)

2. (a) Describe the sequential probability ratio test for a simple hypothesis against a simple alternative, and its uses in acceptance sampling. (10)

(b) Develop a sequential test procedure for controlling the mean of a process with known standard deviation when the quality characteristic is Normally distributed.

Devise a simple graphical procedure for use when inspection is on a routine basis. (15)

3. (a) The average number of visual defects in a specified size of cloth has b found from past records to be 2.8. It is desired to establish control at this level so t there is 1 in 100 chance of the control limit being reached or exceeded. Determine t control limit stating clearly your assumptions.

What is the chance of this limit being reached or exceeded when the average numb of defects increases to 3.6.

[Do not try numerical evaluation. You should give simply the formulae from wh these can be obtained.] (

(b) Samples of size n are drawn from each of 3 machines at regular intervals study a measurable characteristic. After k samples are drawn it is desired to test whet the variabilities of the 3 machines are the same. Suggest a quick and ready method using ranges.

(c) Explain how the 'sign' and 'run' tests help in interpreting the pattern of poi on a control chart. (

4. It is known that the life of electric lamps is distributed exponentially with t density function

$$C \cdot e^{-x/\theta} \quad \theta > 0; \quad 0 < x < \infty .$$

A sample of n bulbs were tested for life. The test had to be terminated just af m bulbs had burnt out. The lengths of the lives of these m bulbs were recorded. Obta by the method of maximum likelihood or otherwise, an expression for estimating the avera life and its variance. (

5. Write an essay on the exploration and exploitation of a response surface for det mining the optimum operating conditions of a production process. (

6. It is desired to instal control over the utilisation of machines in a machine sh How would you plan collection of information? Explain with a suitable layout the relev data to be collected and the analysis to be carried out. What precautions would you ta to ensure validity of results? (

PAPER IV AND V : SAMPLE SURVEYS, THEORY (Theoretical)

Time : 4 Hours Full marks : 1

(a) Attempt *any four* questions.

(b) All questions carry equal marks.

1. Discuss the situations under which systematic samples are preferred to other ty of samples in censuses and surveys.

Show that if the population consists solely of a linear trend then, for the purp of estimating the mean, a systematic sample is more efficient than a simple random samp but less efficient than a stratified random sample. Discuss the effect of suitable 'end co rections' to the systematic sample estimate.

2. What is stratification? How is it useful in sample surveys? Write a note on Neyman's allocation in stratified sampling.

There are two strata in an actual sampling enquiry, and if 'r' is the ratio of the actual value of $\frac{n_1}{n_2}$ to what it would have been under Neyman allocation for the same sample size; show that whatever be N_1, N_2, S_1 and S_2, the relative efficiency of the actual allocation to that of Neyman allocation is never less than $\frac{4r}{(1+r)^2}$.

[N_1 and N_2 are the number of units in the two strata, n_1 and n_2 are the respective numbers of sample units and S_1^2, S_2^2 are the variances within stratum].

3. From a population consisting of N units in existence over a period of time, a simple random sample of n units is selected on a first occasion. [N is large compared to n]. On a second occasion a simple random sample of 'pn' of these n units is retained and are supplemented by 'qn' independently selected units obtained by simple random sampling, where $p+q = 1$. Under the assumption that the variance between units remains constant on both the occasions, and that the value of ρ, the coefficient of correlation between the observations on the two occasions on the same unit, is known, obtain the best linear unbiased estimates for :-

(a) the mean for the second occasion, (b) the change in the population mean. Find the variances of these estimates.

Also show that the optimum replacement fraction 'q' for estimating the mean on the second occasion is given by

$$q = \frac{1-\sqrt{1-\rho^2}}{\rho^2}$$

What is the optimum replacement fraction for estimating the change in the population mean?

4. What are non-sampling errors? Write a note on control and measurement of these errors.

Show how the technique of interpenetrating sub-sample can be used in this connection. Indicate its limitations.

5. A population consists of MN elements which can be grouped into 'M' non-overlapping clusters of N elements each. The variance within clusters (S_{W}^2) follows the empirical law $S_{W}^2 = AN^g$ where A and g are positive constants. How will you obtain the optimum size of the cluster for estimating the mean per element if for a simple random sampling of m clusters the cost function is of the form

$$C_1 mN + C_2 \sqrt{m} \text{ ?}$$

If unequal sized clusters are used, show that the sample mean per element reduces to a ratio estimate. Compare its precision with that of a simple random sample of the same average size (in terms of elements).

6. Write critical notes on :-

(a) Quota sampling

(b) Regression estimate

(c) Sampling with probability proportional to size.

PAPER IV AND V : SAMPLE SURVEYS, APPLIED (Theoretical)

Time : 4 Hours | Full marks : 100

(a) Attempt *any four* questions.

(b) All questions carry equal marks.

1. (a) Discuss possible defects in a sampling frame. How will you scrutinise an actual frame to detect and assess the magnitude of such errors ?

(b) Suggest possible frames for conducting a working class family-budget enquiry in a big city for determining weights for cost of living indexes. How will you make a choice among the alternatives available ?

(c) A frame contains some overlapping area sampling units. The object of the survey is to estimate population total of a particular characteristic. It is suggested that the defect of the frame can be rectified by either of the following two methods :-

(i) Boundaries of all the overlapping sampling units may be arbitrarily defined before drawing the sample.

(ii) As the boundaries of the sampled units only are necessary for purposes of estimation, it is sufficient if boundaries of these units only are arbitrarily defined.

Give your critical comments on these two methods.

2. West Bengal is divided into a large number of small geographical units called 'mauzas'. There is a list of mauzas under each police station. The area of each of the mouzas is shown in the list. The map of each mouza shows the boundaries of different plots of land. Give the structure of a suitable sample survey design for estimating the total production of 'aman'-paddy in West Bengal.

Clearly specify the details of the preparatory steps with special reference to the organisation of field work and selection of sample units.

3. Describe in brief the important points that are to be considered in writing a report on a sample survey. Explain why you consider these points to be important.

With reference to any report on a sample survey (conducted in India or abroad) with which you are faimiliar, discuss in which respects the report is not quite satisfactory.

4. Prepare a process-chart showing in detail the various stages in the processing of large-scale survey results from transference of information to punched cards to preparation of final tables.

What steps do you recommend for control of numerical errors in the processing ?

5. A newspaper printed on its pages daily for a period of one week the following form :

PUBLIC OPINION POLL

I am in favour of / *against* / *undecided about* divorce in the case of Hindu marriages

Signature..........................

Address..........................

---

and invited its readers to fill in the form and return to the newspaper office.

About ten thousand filled in forms were received by the newspaper and 85 percent of these were in favour of divorce on the basis of which the newspaper declared that public opinion was in favour of divorce by an overwhelming majority.

Give your critical comments on the validity of the conclusion drawn, examining carefully possible sources of bias.

How would you plan such a survey ? Explain how in your plan the above sources of bias are eliminated. Devise a suitable simple questionnaire and explain how the information is to be collected

---

PAPER IV AND V : DESIGN OF EXPERIMENTS, APPLIED (Theoretical)

Time : 4 Hours Full marks : 100

(a) Attempt *any four* questions.

(b) All questions carry equal marks.

1. An exploratory trial with seven factors A, B, C, D, E, F and G each of two levels is to be laid out to investigate their effect on the wheat crop. A field divided into 8 blocks of 8 plots each, a plot being of the size of 1/80-th of an acre, is available for conducting the experiment. Prepare a layout plan for the experiment which, in your opinion, would be the most suitable to adopt in this situation, stating what degrees of freedom have been confounded. Also, explain in detail the method of analysis you would employ and give the structure of analysis of variance of the design.

2. Four roasts can be cut from each of 81 animals. The experimenter wishes to study the effect of 36 non-interacting treatments on the tenderness of roasts. Give a layout which would be appropriate in this case. Write down the structure of the analysis of variance of the design and give in detail the method of analysis you would adopt. What is the efficiency factor of the design ?

3. Combinations of two factors, A and B each at three levels are given to 81 wholeplots arranged in a 9×9 Latin-square. Each wholeplot is then divided into 8 subplots to which the combinations of three factors C, D and E each at two levels are allocated at random. Give the structure of the analysis of variance for this design. Explain in detail the method of analysis you would adopt and give the standard error to be used for the various comparisons.

4. (a) State briefly the principles of analysing the data of similar experiments at a number of places.

(b) At each of four experimental stations situated in different parts of a state an experiment was laid out in five replications of a complete randomized block to test the differences in five treatments. How would you analyse the data received from all the stations and what conditions should be satisfied if valid conclusions are to be drawn about treatment differences for the state as a whole ?

5. Give an exact method of analysis of a Latin square design with some plots missing when number of plots missing is less than the number of treatments and each belongs to a different row, column and treatment.

Derive the results for a single missing plot as a special case.

6. Write short notes on the following :

a) Analysis of non-orthogonal data.

b) Linear and quadratic responses.

c) Asymmetrical factorial designs.

PAPER IV AND V : MATHEMATICAL THEORY OF SAMPLING DISTRIBUTION (Theory)

Time : 4 Hours — Full marks : 100

(a) Attempt *any four* questions.

(b) All questions carry equal marks—

1. Define Mahalanobis' Studentized D^2-statistic for testing equality of the mean-vectors of two p-variate Normal populations with a common dispersion matrix.

Derive its distribution in the null case, and describe the computational procedure for evaluating the statistic given the means and the matrices of corrected sums of squares and products of the two samples.

2. Obtain the joint distribution of the means and second order central moments in a sample of size n from bivariate Normal population. Derive the distribution of the sample correlation coefficient.

Write a note on the nature of the frequency curves of the correlation coefficient when $n \leqslant 4$.

3. (a) If x_1 and x_2 are independent random observations from a population mean with probability density function.

$$f(x) = \frac{\sqrt{2}}{1+x^4}, \quad -\infty < x < +\infty$$

find the distribution of the function

$$\frac{x_1 + x_2}{|x_1 - x_2|}$$

(b) Let X take the value 0 or 1 with probability $1-p$ and p respectively. If n independent observations $x_1, x_2, \ldots, x_n$ are taken and r denotes the number of observations which are as large as x_1, find the probability distribution of r.

4. A set of n independent normal variates $x_1, x_2, \ldots, x_n$ each with mean zero and variance σ^2 are given.

Let $y_1, y_2, \ldots, y_p$ $(p < n)$ be a set of linear functions in x's defined as

$$y_i = \sum_{j=1}^{n} C_{ij} x_j \quad i = 1, 2, \ldots, p$$

such that

$$\sum_{j=1}^{n} C_{ij} C_{i'j} = \partial_{ii'}$$

where $\partial_{ii'}$ is Kronecker delta, $\partial_{ii} = 1$, $\partial_{ii'} = 0$ if $i \neq i'$

If $Q_1 = \sum_{i=1}^{p} y_i^2$ and $Q_2 = \sum_{i=1}^{n} x_i^2 - \sum_{i=1}^{p} y_i^2$

derive the distribution of $U = Q_1/Q_2$ and deduce the distribution of Fisher's t and z from the distribution of U.

5. Obtain the characteristic function of the distribution

$$dF = \tfrac{1}{2}k\, e^{-x}(1+x)^2\, dx \qquad -1 < x < \infty$$

and hence find the third and fourth moments.

If $f_1(t) = \frac{2}{\pi}\int_0^\infty \frac{1-\cos x}{x^2} \cos tx\, dx$

$$f_2(t) = \tfrac{1}{2} + \frac{4}{\pi^2}\sum_{n=0}^{\infty} \frac{\cos(2n+1)\pi t}{(2n+1)^2}$$

Indicate the forms of the distributions of which $f_1(t)$ and $f_2(t)$ are the characteristic functions.

6. Three independent observations are made on a normal distribution with mean m and variance σ^2. Find the distribution of (a) the range, (b) the mean of the two closest observations.

7. On the basis of a random sample of size n, develop the likelihood-ratio criterion for mutual independence of k sets of normal variates

$$(x_1, x_2, \ldots, x_{p_1});\quad (x_{p_1+1}, x_{p_1+2}, \ldots, x_{p_1+p_2});$$

$$\ldots\ldots;\ (x_{p_1+p_2+\ldots+p_{k-1}+1}, \ldots, x_{p_1+p_2+\ldots+p_k})$$

and express it in terms of correlation coefficients. Find the moments of the criterion. Deduce the distribution of the sample multiple correlation coefficient when in the parent distribution, the multiple correlation coefficient is zero.

PAPER IV AND V : PROBIT ANALYSIS (Theoretical)

Time : 4 Hours　　　　　　　　　　　　　　　Full marks : 100

Attempt *any four* questions

1. Explain the term *tolerance* as used in quantal-type bio-assays. Describe the general features of a tolerance-distribution. Show how by a proper choice of dose-and response-metameters, the problem of estimating the parameters of tolerance-distribution may be reduced to that of fitting a regression line. Why is it necessary in some situations to use different weights for different points on the above line ?

2. Define the term median effective dose as used in quantal-type bio-assays.
If tolerance (z) has the distribution

$$\beta\, e^{-\beta z}\, dz \quad (0 < z < \infty, \beta > 0)$$

how would you obtain the maximum likelihood estimate of the median effective dose from quantal response data ?

Describe the computational procedure in detail. Make a list of special mathematical functions which if tabulated would be useful in this computation.

3. Write notes on :

(a) Abbotts' formula for adjustment for natural mortality.

(b) Method of extreme effective doses for estimation of the median effective dose.

(c) Use of a probit regression plane in bio-assays.

4. You are given a test and a standard preparation of drug and a very rough estimate of the relative potency of the test preparation. Describe as fully as you can the plan of an assay for the estimation of the relative potency when the response is quantal and you can have at the most N subjects. Describe the different precautions to be taken. How would you choose the doses to increase the precision of the estimate of relative potency ?

5. If the average quantity Q of cereals consumed by households with income x is given by

$$Q = K.\phi\ (\alpha + \beta \log x)$$

where $\phi\ (u) = \int_{-\infty}^{u} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2}\, dt$ and K, α, β are unknown parameters, describe how you would use the probit-technique to estimate the parameters K, α, β when observations (Q_i, x_i) $i = 1, 2, \ldots, n$ are available for a sample of n households.

PAPER VIII AND IX : ECONOMIC STATISTICS (Practical)

Time : 4 Hours | Full marks : 100

(a) Attempt *any two* questions.

(b) All questions carry equal marks.

1. Fit a production function of the Cobb-Douglas type to the following data and determine the production elasticities.

serial number	value of output (Rs.)	land (acres)	labour (Rs.)	capital (Rs.)
1	3764	11	2426	1859
2	2784	11	1353	1551
3	1537	12	1052	613
4	1620	10	1034	629
5	1883	13	1288	726
6	851	10	757	389
7	3966	17	1595	1744
8	1489	12	483	416
9	1384	15	1113	630
10	1797	16	713	1464

2. Comment on the nature of cyclical variations in the following series of index numbers of Jute prices from 1920 to 1956 after eliminating the trend by means of a 9-year moving average.

year	index	year	index
1920	235	1939	123
1921	245	1940	141
1922	249	1941	124
1923	237	1942	131
1924	199	1943	201
1925	268	1944	204
1926	332	1945	194
1927	207	1946	259
1928	222	1947	419
1929	202	1948	492
1930	159	1949	491
1931	98	1950	486
1932	97	1951	815
1933	86	1952	497
1934	80	1953	378
1935	108	1954	398
1936	199	1955	442
1937	111	1956	448
1938	97		

3. From the following data estimate the number of acres of land to be put under cotton in 1966 :-

(i) overall increase in consumer expenditure between 1956 and 1966 .. .. 30 percent

(ii) population in 1956 ... .. 391.4 millions

(iii) rate of change of population .. .. 2.6 percent per annum

(iv) expenditure elasticity for clothing .. 1.6

(v) demand for cloth in 1956 .. .. 6,275 million yards

(vi) assume that 1500 yards of cloth require one bale of cotton[one bale = 392 lbs. of cotton].

(vii) the yield rate of cotton is 78 lbs. per acre in 1956 and is expected to increase at a rate of 3 percent per annum.

A. From (i) to (v) calculate the demand for cloth in 1966 due to changes in population and changes in expenditure.

B. Obtain the export demand for cotton and cotton cloth by using the following figures and projecting by fitting a linear trend for the year 1966.

year	cotton cloth exports (mill. yards)	exports of raw cotton ('000 bales of 392 lbs. each)
1950	1209.9	330
1951	383.7	84
1952	560.9	132
1953	702.5	405
1954	761.3	200
1955	680.0	171
1956	788.8	690

PAPER VIII AND IX : STATISTICAL QUALITY CONTROL (Practical)

Time : 4 Hours Full marks : 100

(a) Attempt *any three* questions.
(b) All questions carry equal marks.
(c) Use of calculating machines is permitted.

1. The following data relates to the deviation from the specified standard for the inside length of polythene bags in units of one sixtyfourth of an inch.

sample number	observations 1	2	3	4
		Bag size S_1		
1	17	15	11	14
2	5	13	15	7
3	12	11	8	9
4	15	13	11	8
		Bag size S_2		
5	4	0	12	9
6	4	8	14	12
7	4	0	6	14
8	0	2	8	6
9	4	0	12	9
10	11	8	4	12
11	7	12	10	15
		Bag size S_3		
12	15	10	8	10
13	14	14	12	18
14	12	7	11	10
15	9	12	10	8
16	9	9	13	9
17	4	6	13	13
		Bag size S_4		
18	0	10	9	12
19	13	13	7	5
20	15	9	5	3

Samples of 4 bags were collected from the machine at regular intervals and data have been obtained on 4 bag sizes S_1, S_2, S_3, S_4. Apriori, there is no reason for either the mean excess length or the variability to be different for different sizes.

Analyse the data to examine this aspect and obtain the process capability.

2. (a) The tolerances specified for the inside diameter of a component are 0.500 $\pm 0.001''$. The components having a lower diameter than the specified could be reworked whereas those with larger diameters would be rejected. The loss of a rejected item is equal to the cost of reworking 4 items. Where should the process be centred so that the total cost of rework and rejection is minimum ?

You may assume that the standard deviation of the diameters is 0.0005".

(b) Past records indicate that the standard deviation of the weight of a cigarette is 0.04 gm. The mean weight of a cigaratte is to be controlled at 1.20 gms. A ten percent increase in the mean weight is desired to be detected with a probability of 0.95.

On the other hand, the probability of wrongly deciding that the mean weight has increased, should be 0.01. Determine the size of the sample to ensure the above probabilities mentioning the decision rule based on the sample.

3. The minimum requirements on the weight of the contents of a certain container is 80.0 oz. It is desired to instal a suitable sampling scheme with the following properties.

AQL = 3 percent and producer's risk = 0.01

LTPD = 8 percent and consumer's risk = 0.05

Specify a suitable single sampling variable plan and the corresponding attribute plan. Clearly set down the plan of action in each case. Under what conditions would you prefer to use the variable plan in preference to the attribute plan ?

4. An experiment was conducted in a rubber factory with a view to reduce buckling in hose pipes, the factors being : coil length (C), sheet length (S) and degree of tension (T) each at three levels. For each combination 4 coils were observed and the total number of buckles in the four coils is given below :-

		S_1	S_2	S_3
T_1	C_1	2	4	1
	C_2	6	4	4
	C_3	8	5	2
T_2	C_1	2	0	3
	C_2	1	0	0
	C_3	1	2	0
T_3	C_1	5	5	0
	C_2	6	0	5
	C_3	2	0	1

Comment upon the experiment and the results making any analysis that you deem appropriate.

PAPER VIII AND IX : SAMPLE SURVEYS, THEORY (Practical)

Time : 4 Hours Full marks : 100

(a) Attempt *all* questions.
(b) All questions carry equal marks.
(c) Use of calculating machines is permitted.

1. The following data relating to the yield of paddy were obtained from crop cutting experiments conducted in a certain tract. There were 40 random cuts of size 1/100-th of an acre in all. The weight of fresh paddy was observed for all the 40 cuts. Out of these,

on 20 cuts selected at random observations on dry yield were also taken. The data are given below :-

serial number of cut	weight of fresh paddy in lbs.	dry yield in lbs.	serial number of cut	weight of fresh paddy in lbs.
1	16.8	15.2	21	8.7
2	12.7	11.8	22	11.6
3	18.8	17.5	23	11.5
4	13.9	12.5	24	14.4
5	11.3	10.4	25	17.8
6	10.9	10.1	26	8.4
7	12.5	11.2	27	8.7
8	17.4	15.8	28	14.6
9	14.1	13.0	29	12.1
10	11.9	10.8	30	7.9
11	13.4	12.3	31	8.9
12	13.5	12.4	32	11.1
13	8.3	7.6	33	13.0
14	13.7	12.5	34	10.5
15	14.6	13.3	35	14.2
16	14.4	13.5	36	12.7
17	17.1	16.2	37	11.9
18	14.5	13.5	38	15.5
19	11.4	10.3	39	17.1
20	14.0	12.8	40	10.9

(a) Estimate the average yield of dry paddy per acre along with its sampling error utilising the information available both on fresh weight and on dry yields.

(b) What would have been the loss in precision had we neglected the additional information on fresh weight and estimated the average yield from the subsample of only 20 cuts for which dry yield is available ?

(c) If the cost of obtaining the dry weight from a cut is 1½ times that of obtaining the fresh weight of the cut, determine the optimum proportion of cuts for which dry weight is to be observed.

2. Data extracted from a sample survey partaining to the number of fruit trees obtained by complete enumeration of all orchards in each of the 48 selected villages in a district are given below. The district was divided into six strata and the sample villages were allotted to the six strata in proportion to the area under orchards in each of them. From each of the first five strata, the villages were selected with replacement and with probability proportional to the total area under orchards in the village. In the sixth stratum, the required villages were selected simply at random without replacement.

From the data estimate the total number of fruit trees in the district and its sampling error.

If the average cost of enumeration of trees per selected village in the six strata are taken as Rs. 30,20,25,25, 100 and 30 respectively and if the total cost on enumeration in the survey is not to exceed Rs. 2,500 determine the optimum size of the sample to be allocated.

stratum no.	no. of villages in the stratum	no. of villages selected	total number of fruit trees in each selected village (in tens); figures in brackets give the ratio (area under orchards in the whole stratum/area under orchards in the sample village)							
(1)	(2)	(3)	(4)							
I	63	4	48	46	13	175				
			(41.3)	(24.8)	(24.8)	(13.8)				
II	192	8	80	36	32	110	63	36	34	33
			(42.3)	(29.3)	(69.6)	(34.6)	(63.5)	(95.2)	(95.2)	(381.0)
III	74	8	16	158	19	18	89	45	69	24
			(102.7)	(20.5)	(154.0)	(38.5)	(25.7)	(51.3)	(7.3)	(7.3)
IV	97	4	75	75	165	115				
			(14.6)	(14.6)	(21.3)	(19.9)				
V	153	16	19	75	73	29	169	99	112	105
			(170.9)	(293.0)	(293.0)	(157.8)	(293.0)	(93.2)	(30.2)	(157.8)
			129	1093	1093	480	431	1025	1025	8
			(64.1)	(16.2)	(16.2)	(66.1)	(76.0)	(20.7)	(20.7)	(512.7)
VI	302	8	39	82	83	32	34	7	24	0

PAPER VIII AND IX : SAMPLE SURVEYS, APPLIED (Practical)

Time : 4 Hours Full marks : 100

(a) Attempt *all* questions

(b) Use of calculating machines is permitted.

1. The city of Calcutta has been divided into 3200 mutually exclusive and exhaustive blocks. A sample of 320 blocks will be chosen at random and completely enumerated by 80 investigators for estimating the number of employed and unemployed persons of different categories, reference period being a week.

The following estimates will be prepared from the data :-

(i) the number of unemployed persons of different age-sex groups

(ii) distribution of persons in the labour force by sex and number of days employed during the reference period.

(iii) distribution of unemployed persons seeking employment for the first time by educational qualification and employment preference.

In order to bring out more fully the significance of the numbers shown in the above tables, these are to be supplemented by 2 or 3 auxiliary tables (to be shown by the examinee).

(a) Prepare a suitable schedule for the survey and write down necessary instructions for filling up the schedule.

(b) Prepare forms for tabular presentation of the estimates (i) (ii) and (iii) (and the auxiliary information) with necessary headings and other relevant particulars.

2. For purposes of estimating the 1960 population of a State, it is proposed to draw from every district a sample of specified number of villages with probability proportional to 1951-population and with replacement.

(i) Indicate different stages of computation work necessary for drawing the sample and write out necessary instructions to the computers for this purpose.

(ii) On the basis of the data obtained from the sample drawn as per plan above, explain in writing how to obtain a suitable estimate for the total population of 1960 and estimate the variance of the estimate.

(iii) Use the information in the 'enclosure' for illustrative purposes and estimate the 1960 population of the district 'A' from a sample of ten villages drawn this way. Find the standard error of this estimate.

3. (i) From the 'enclosure' mentioned in question 2 draw a simple random sample of 10 villages with replacement.

(ii) From this sample estimate the 1960-population of district

(a) by the ratio-method, using the 1951-population figures : (ratio-estimate),

(b) from the mean of the 1960 population figures for the sampled villages (ignoring 1951 population figures),

(iii) Find the standard errors of the estimate (a) and (b) above.

(iv) Compare the efficiencies of the estimates (a), (b) and that obtained in question 2(iii).

ENCLOSURE

List of villages in the district 'A'

serial number of village	population in 1951	population in 1960
(1	(2)	(3)
01	294	325
02	262	328
03	284	404
04	453	437
05	401	441
06	112	121
07	116	140
08	168	165
09	83	103
10	93	147
11	335	386
12	231	293
13	178	196
14	140	158
15	118	174
16	457	514
17	383	401
18	701	817
19	594	467
20	555	817
21	551	480
22	723	774
23	1105	993
24	1161	1922
25	4111	4836

PAPER VIII AND IX : DESIGN OF EXPERIMENTS, APPLIED (Practical)

Time : 4 Hours | Full marks : 100

(a) Attempt *any two* questions.
(b) All questions carry equal marks.
(c) Use of calculating machines is permitted.

1. The table below presents the design and the yields of air-dried soybean in grams per pot in an experiment carried out to investigate the effect of 6 trace elements A, B, C, D, E, and F, each at two levels, on the growth of soybean.

Block I		Block II		Block III		Block IV	
(1)	1.18	f	2.48	d	3.73	a	4.18
ade	2.50	adef	2.83	ae	3.43	de	2.73
acf	3.18	ac	2.98	acdf	3.50	cf	3.98
bdf	2.73	bd	4.50	bf	2.33	abdf	3.88
bce	3.83	bcef	2.33	bcde	2.33	abce	4.48
abcd	4.25	abcdf	4.18	abc	4.38	bcd	2.58
abef	5.65	abe	4.48	abdef	2.55	bef	2.03
cdef	2.88	cde	2.98	cef	4.05	acdef	3.90

Block V		Block IV		Block VII		Block VIII	
c	3.58	b	2.65	e	5.45	ab	5.10
acde	3.50	abde	2.78	ad	5.35	bde	2.60
af	3.80	abcf	4.45	acef	3.70	bcf	3.33
bcdf	3.60	df	3.45	bdef	2.35	adf	3.85
be	3.13	ce	3.90	bc	3.20	ace	4.75
abd	3.03	acd	3.38	abcde	4.30	cd	4.80
abcef	2.68	aef	3.05	abf	3.20	ef	2.75
def	2.60	bcdef	3.98	cdf	3.93	abcdef	4.58

Identify the confounded degrees of freedom. Analyse the data and interpret the results of analysis.

2. Given below are the yields of an experiment on wheat laid out in a randomized block design with 6 replications, the nine treatment combinations tested being :-

irrigation		*seed rate per acre*
9″		35 lbs.
12″	×	50 lbs.
15″		65 lbs.

The figures in bracket correspond to the yields obtained from a uniformity trial carried out on the same place in the previous year.

	Yield in lbs. per plot of 1/40th of an acre								
	35 lbs.			50 lbs			65 lbs.		
	9″	12″	15″	9″	12″	15″	9″	12″	15″
I	15 (18)	10 (20)	17 (15)	8 (6)	22 (24)	29 (21)	8 (9)	22 (21)	29 (27)
II	21 (20)	30 (26)	30 (26)	20 (16)	16 (18)	13 (15)	10 (14)	16 (15)	13 (9)
III	11 (21)	31 (24)	24 (21)	14 (12)	24 (21)	18 (16)	17 (16)	24 (25)	18 (13)
IV	17 (13)	19 (16)	14 (11)	8 (6)	19 (16)	24 (21)	19 (18)	19 (16)	24 (21)
V	31 (26)	20 (21)	25 (21)	14 (12)	10 (13)	21 (17)	14 (17)	16 (14)	22 (22)
VI	13 (9)	11 (12)	20 (16)	6 (9)	23 (25)	17 (9)	19 (17)	23·(20)	17 (20)

Perform the analysis of variance and covariance. Calculate the gain in precision due to the use of uniformity trial yields in analysis of covariance. Also calculate the standard error of difference of two adjusted means.

3. The table below presents in summary form data in respect of weights and numbers of lambs from 3 breeds of sheep over a period of 7 years. The figures outside bracket are the sub-totals for each cell and the figures inside bracket the corresponding numbers on which these are based.

weight of female lambs from 3 breeds for 7 years

years	Breeds			Total
	A	B	C	
1	464 (8)	1361 (22)	369 (5)	2194 (35)
2	845 (14)	2758 (40)	3778 (52)	7381 (106)
3	707 (13)	2310 (36)	4281 (57)	7298 (106)
4	411 (6)	1595 (23)	5465 (66)	7471 (95)
5	756 (11)	1618 (24)	4003 (47)	6377 (82)
6	489 (9)	1016 (20)	3915 (55)	5420 (84)
7	1227 (18)	1798 (27)	6813 (83)	9838 (128)
	4899 (79)	12456 (192)	28624 (365)	45979 (636)

The total sum of squares for the data is 113 232.0. Analyse the data for testing the effects of breed and season and decide as to the suitablity of the adjusted means for representing these effects. If the adjusted means are not satisfactory, state the action that should be taken.

PAPER VIII AND IX : MATHEMATICAL THEORY OF SAMPLING DISTRIBUTION (Practical)

Time : 4 Hours Full marks : 100

(a) Attempt *any two* questions

(b) Use of calculating machines is permitted.

1. Using a table of random sampling numbers, draw a random sample of size 10 from a tri-variate Normal population with the following parameters :

means : $\mu_1 = 4.2$ $\mu_2 = 2.5$ $\mu_3 = 3.7$

variances and covariances : $\sigma_{11} = 4.7$ $\sigma_{12} = 1.1$ $\sigma_{13} = 1.9$

$\sigma_{22} = 1.2$ $\sigma_{23} = 2.1$

$\sigma_{33} = 8.6$

Estimate the means from the sample. Test whether the estimated values of the mean are in statistical agreement with their actual values in the population. [You may make use of the values of the variances and covariances in the population.]

2. For a random variable X with mean 0 and variance 1, the Gram-charlier Type A series expansion of the probability density function $f(x)$ of X (to first four terms) is

$$f(x) \sim [1 + \tfrac{1}{6}\mu_3 (x^3 - 3x) + \tfrac{1}{24}(\mu_4 - 3)(x^4 - 6x^2 + 3)]\, e^{-\frac{1}{2}x^2}$$

Using this approximation, find the upper 5 percent point of the distribution of the mean of a sample of size 4 from a rectangular population with range (0, 1).

3. Let X_i ($i = 1, 2, \ldots$) be a sequence of independent random variables with a common probability distribution given by

$$\text{Prob } (X_i = 0) = 0.4, \quad \text{Prob } (X_i = 1) = 0.6$$

Let $S_n = X_1 + X_2 \ldots + X_n$ and let N be the value of n for which $S_n > \frac{1}{2}n$ for the first time.

Estimate $E(N)$ and $V(N)$ by carrying out a model sampling experiment to obtain 10 observations on N.

Estimate on the basis of these results, how many such observations are required to estimate $E(N)$ with a coefficient of variation of 0.1 percent.

PAPER VII AND IX : VITAL STATISTICS AND POPULATION STUDIES (Practical)

Time : 4 Hours Full marks : 100

(a) Attempt *all* questions.

(b) Use of calculating machines is permitted.

(c) Logarithmic tables will be supplied.

1. The following table gives the number of women aged 10 to 44 years by age group and the number out of them with at least one child born to them. Estimate the age at which a woman may be expected to give birth to her first child. State clearly the assumptions made in this connection.

age group	number of women	number of women with at least one child
10–14	32,204	501
15–19	27,382	4,467
20–24	22,632	12,814
25–29	20 807	17,402
30–34	17,057	15,450
35–39	15,623	14,346
40–44	13,137	12,597

2. Assuming that $\frac{1}{l_x} \frac{dl_x}{dx}$ = constant in the range $x = 10$ to 60 years, calculate the expectation of life at ages 45 and 50 years, it being given that $l_{10} = 67,167$; $l_{60} = 24,347$; $e_{60} = 11.50$ years.

[l_x is the number of survivors at age x in a life table population and e_x is the expectation of life at x years.]

3. From the data given below, calculate the percentage of persons aged 15 years and above in the labour force in U.S.A. in 1920 and 1940. Standardise the 1940 value with respect to sex and colour-nativity taking the distribution of 1920 as standard, using both the 'direct' and 'indirect' methods.

Explain the utility of this standardisation.

(a) Distribution of persons (15 years and above) according to sex and colour-nativity, in 000's.

colour-nativity	1920		1940	
	male	female	male	female
native whtie	30,650	30,254	40,036	44,313
foreign white	7,416	6 090	5,994	5,374
non-white	4,218	4,159	5,240	5,501

(b) Distribution of persons (15 years and above) in the labour force according to sex and colour-nativity in 000's.

colour nativity	1920		1940	
	male	female	male	female
native white	22,497	5,718	31,662	10,192
foreign white	6,467	1,084	4,837	979
non-white	3,341	1,547	3,783	1,843

4. The age distribution of males in Bombay State in the 1941 and 1951 censuses are given below. Estimate the effect of mortality and net-migration during 1941-51 on the males in the five-year age groups 5-9, 10-14, etc. to 35-39 in the 1941 census, assuming that the life-table reproduced below is applicable to Bombay state males during 1941-51.

(a) Age distribution of males in Bombay State in 1941 and 1951 in 00's.

age group	1941	1951
0– 4	19907	24954
5– 9	20458	23765
10–14	17414	23284
15–19	11783	16847
20–24	12288	16270
25–29	13547	16070
30–34	12037	13926
35–39	9940	12116
40–44	8703	10321
45–49	6893	7913

(b) Life table survivors in five year age-groups

age group	survivors
5– 9	329,868
10–14	318,340
15–19	307,698
20–24	293,312
25–29	276,725
30–34	259,585
35–39	241,430
40–44	221,395
45–49	198,805