INDIAN STATISTICAL INSTITUTE

# QUESTION PAPERS

*for*

The Statistician's Diploma Examination

1954

*Price Re*

# INDIAN STATISTICAL INSTITUTE

## STATISTICIAN'S DIPLOMA EXAMINATION, 1954.

### PAPER I : THEORETICAL STATISTICS (GENERAL)

Time : 4 hours  Full marks : 100

N.B. (a) Answers to the different groups are to be given in separate books.
(b) Attempt any three questions from each group.
(c) All questions carry equal marks.

### GROUP A.

1. (a) If $W_1 = \sum_{i=0}^{4} u_{3i}$, $W_0 = \sum_{i=-2}^{2} u_{3i}$ and

$$W_{-1} = \sum_{i=-7}^{-3} u_{3i}$$

obtain the value of $u_0$ in terms of $W$'s.

(b) Obtain the approximate formula

$$\int_{-1}^{1} u_x dx = \frac{13(u_1+u_{-1})-(u_3+u_{-3})}{12}$$

and indicate the order of difference to which it holds.

2. (a) If for $x \geq 20$ the probability that a man aged $x$ years marries at time $t$ from the present date is $\frac{15\,dt}{(t+x)^2}$ find the probability that $A$ marries before $B$ assuming the ages of $A$ and $B$ as 20 years and 30 years respectively.

(b) On the average how many times must a coin be tossed in order that a succession of three tails may be obtained ?

3. (a) If in a population of $N$ units a particular characteristic is present in $m$ of them and if a sample of $n$ units is drawn without replacement, obtain an expression for the standard deviation of the number of units in the sample with the particular characteristic. Comment on the utility of this result.

(b) State the Weak Law of Large Numbers and discuss its utility in statistics.

4. (a) Show that if a normal distribution is grouped in intervals with total frequency $N$ and $S$ is the sum of the squares of the frequencies, an estimate of population standard deviation is given by

$$0.282\,\frac{N^2}{S} \qquad \text{approximately.}$$

(b) In a Poisson distribution obtain the following recurrence formula for the moments $\mu_r'$, $\lambda$ being the usual expected value.

$$\mu_{r+1} = r\lambda\, \mu_{r-1} + \lambda\, \frac{d\mu_r'}{d\lambda}$$

## Group B.

5. (a) Briefly state the different types of problems for which Student's $t$ provides the appropriate test of significance. Also state the assumptions under which the test is valid in each case.

(b) Write a critical note on the following statement of R. A. Fisher:

"The confidence to be placed in a result depends not only on the magnitude of the mean value obtained but equally on the agreement between parallel experiments. Thus, if in an agricultural experiment a first trial shows an apparent advantage of 8 bushels to the acre and a duplicate experiment shows an advantage of 9 bushels, we have $n = 1$, $t = 17$ and the results would justify some confidence that a real effect had been observed. But if the second experiment had shown an advantage of 18 bushels, although the mean is now higher, we should place no more but less confidence in the conclusion that the treatment was beneficial."

6. (a) Explain why the entries in the last line (corresponding to $n = \infty$) of the $t$-table are the same as those in the table of the standardised normal deviate? Can you supply a proof for it? How would you make use of this fact in actual practice?

(b) Under usual assumptions (to be stated) how do you test for equality of two variances? What is the reason for placing the larger variance in the numerator of the Statistic when a test is applied?

7. (a) Distinguish between correlation and regression. In what sense is the Correlation Coefficient a measure of association between two variates? Comment on the following statement:

> "In a certain country it is observed over a large number of years that there is strong correlation between the number of sun stroke cases and the amount of wheat produced. This shows that the two variates are influenced by each other."

(b) A number of persons are measured for their height ($x$), weight ($y$) and chest expansion ($z$) and the product-moment correlation coefficients are calculated. Prove that

$$r_{xy} + r_{xz} + r_{yz} \geq -\tfrac{1}{2}$$

8. Give a method of obtaining the large sample standard errors of functions of sample moments. Apply this method to find the standard error of the coefficient of variation in random samples of size $n$ from a normal population with mean $m$ and variance $\sigma^2$. What is its value when the population coefficient of variation is zero?

---

STATISTICIAN'S DIPLOMA EXAMINATION, 1954.

PAPER II : APPLIED STATISTICS (GENERAL)

Time : 4 hours Full marks : 100

*N.B.* (a) Answers to the different groups are to be given in separate books.

(b) Attempt any three questions from each group.

(c) All questions carry equal marks.

GROUP A.

1. Explain multistage sampling. Under what circumstances would you adopt it and with what advantages ?

Outline a design for estimating the yield per acre of paddy in the State of West Bengal. How would you obtain the sampling error of the estimate ?

2. Indicate the various stages in the planning, execution and tabulation of results, of a socio-economic sample survey.

What are the various types of error in such a survey, and how would you study the differences due to investigators ?

3. Describe the official sources for statistics of employment and unemployment and state their main limitations and gaps.

What steps should you suggest for obtaining the size, composition and the extent of unemployment of the entire labour force in India ?

4. Define National Income of a country. Describe briefly the method of estimation of the National Income of India employed by the National Income Committee.

What are the sectors which are particularly weak from the point of view of reliability ? Indicate how far the National Sample Survey could be utilised to improve them.

5. Describe the various methods of constructing Index Number of Prices. Describe in detail the official index of wholesale prices in India, the method of computation and its defects. Would you use it for deflation of national income with respect to prices ?

6. Explain the various components of a time series and show how you would estimate seasonal variation, if any.

If you are given annual rainfall data at a particular place over a long period, explain how you would analyse them for any cycles.

## Group B.

7. Describe the direct and indirect methods of standardisation of death rates and give an example of a situation where the latter, and not the former, could be used. Discuss the possible reasons for inaccuracies in the computation of the age specific death rates.

8. Describe any method of standardisation of death rates wherein the weights for the age specific death rates are not arbitrary but are functions of the specific death rates themselves.

Define the net reproduction rate and show how it is related to the growth of population.

9. Layout the design of an experiment to assess the effect in growth of body weight of children due to the addition of a particular proteinous food to the diet of children classified as vegetarian and on-vegetarian.

How can the efficiency of the design be enhanced by taking account of the relationship between rate of growth of body weight and variables like sex and age?

10. Explain the principle of confounding and its advantages in factorial experiments. Derive a confounded design for 3 factors each at two levels in blocks of size 4. Give the analysis of variance table for the design.

11. State Mendel's law of inheritance and calculate the expected phenotype frequency ratios in the $F_2$ generation for two independent genetic factors, each having a dominant gene.

If the two factors are linked in one chromosome calculate the phenotype frequency ratios resulting from the crossing of an $F_1$ female with a double recessive male assuming that 20% of the gametes of $F_1$ were cross-overs.

12. Describe Spearman's two-factor theory and show how the criterion of proportionality of the inter correlations between different tests in a battery can be used to test this theory.

In a battery consisting of three tests it is known that there are two independent common abilities involved. Give a graphic representation of the factor pattern and establish therefrom that an infinite number of factor patterns can give rise to the same correlation matrix.

## STATISTICIAN'S DIPLOMA EXAMINATION, 1934.

### PAPER III : STATISTICAL INFERENCE

Time : 4 hours Full marks : 100

*N.B.* (a) Answers to the different groups are to be given in separate books.
(b) Attempt any three questions from each group.
(c) All questions carry equal marks.

### GROUP A.

1. Describe the method of Maximum Likelihood for estimating a parameter. What useful properties, if any, does such an estimator have ?

Give an example where the Maximum Likelihood estimator is identical with the estimator obtained by Least Squares method.

Find the Maximum Likelihood estimators for the limits of the range of a rectangular distribution.

2. Define the concepts of unbiasedness, consistency and sufficiency for an estimator of a parameter.

Show that the average number of successes in $n$ independent trials, each having probability $p$ of success, is both unbiased and sufficient for estimating $p$.

Find the necessary and sufficient condition that a distribution admits a sufficient statistic for a single parameter $\theta$.

3. (a) Show that if $T_1$ and $T_2$ are two unbiased estimators of a population parameter, both with minimum variance, then $T_1$ and $T_2$ are essentially identical.

(b) If an unbiased minimum variance estimator $T$ of a population parameter $\theta$ exists and if a sufficient statistic $T'$ exists for $\theta$, then $T$ is a function of $T'$ alone.

4. Discuss clearly the theory of estimation of a parameter by confidence intervals. Give a method of obtaining confidence intervals and use it to obtain the same for the mean $\mu$ of a normal population with unknown variance.

How will you find the number of observations necessary to get a confidence interval for $\mu$ of a specified length $l$, the confidence coefficient being $\alpha$ ?

### GROUP B.

5. Define 'similar regions' and explain how a best critical region could be constructed for testing a composite hypothesis. Illustrate with the test for equality of variances in two normal populations.

Explain how the test of a linear hypothesis is equivalent to the above test.

6. What is meant by bias in statistical tests? When there are two tests of the same statistical hypothesis, show how you can compare them with respect to bias and say that one is less biased than the other.

Describe the different types of unbiased critical regions suggested in the literature. Find a Type $A$ unbiased test of the hypothesis that the variance of a normal population is $\sigma_0^2$. Show that it is also of the $A_1$ Type.

7. Write down the sampling distribution of the partial correlation coefficient $r_{x_1 x_2 . x_3 \ldots x_p}$ when $x_1, x_2, \ldots, x_p$ are normally distributed and obtain the test of significance of a sample partial correlation coefficient when its population value is (i) zero, (ii) equal to $\rho_{x_1 x_2 . x_3 \ldots x_p}$.

Indicate how the test of significance of the multiple correlation coefficient $R$ between $x_1$ and $x_2, x_3, \ldots, x_p$ is similar to that of the correlation ratio $\eta$. If this test shows that $R$ is not significant at a specific level, can you infer that the individual correlations between pairs of $x$'s are also insignificant?

8. Explain carefully, stating any assumptions, the uses of the $D^2$ statistic as a measure of group distance and as providing a means of testing whether two observed samples belong to the same multivariate normal population.

If we have $k$ $p$-variate normal populations having the same dispersion matrix obtain Wilk's criterion for testing for equality of means of the $p$ variates. Give a large sample test of this hypothesis.

---

## STATISTICIAN'S DIPLOMA EXAMINATION, 1954.

### Paper VI : Practical

Time : 6 hours — Full marks : 100

*N.B.* (a) Figures in the margin indicate full marks.
(b) Use of calculating machines is permitted.

1. From the following table of values of the functions *Cos x* with $x$ in radians at intervals of 0.1, calculate *Cos x* for (i) $x = 5.138$ and (ii) $x = 5.350$. (15)

| $x$ | *Cos x* | $x$ | *Cos x* |
|---|---|---|---|
| 5.0 | 0.283662185 | 5.4 | 0.634692876 |
| 5.1 | 0.377977743 | 5.5 | 0.708669774 |
| 5.2 | 0.468516671 | 5.6 | 0.775565879 |
| 5.3 | 0.554374336 | 5.7 | 0.834712785 |

2. (a) Below are given $\mu_2, \mu_3, \mu_4$ for four frequency distributions. Indicate which type of Pearson's system of curves is suitable for each of these distributions. (8)

| | I | II | III | IV |
|---|---|---|---|---|
| $\mu_2$ | 1.44 | 2.25 | 4.00 | 1.00 |
| $\mu_3$ | 0.95 | 3.22 | 1.36 | 2.19 |
| $\mu_4$ | 8.07 | 17.27 | 66.20 | 13.02 |

(b) A frequency distribution of weights of male students in a university is given below. It is found that Pearson's type VII curve of the form $y = y_0\left(1 + \frac{x^2}{a^2}\right)^{-m}$ with origin at mean fits the distribution and the constants in this case are $a^2 = 1607.5$, $m = 5.1$, $y_0 = 35.8\frac{\Gamma(5.1)}{\Gamma(.5)\ \Gamma(4.6)}$. Calculate the expected frequencies using Incomplete $\beta$-Function tables or otherwise and test for goodness of fit. (Take $\pi$ correct to two places of decimals only). (18)

| Class interval (lb.) | frequency | |
|---|---|---|
| below 77.5 | 63 | |
| 77.5— 92.5 | 270 | |
| 92.5—107.5 | 593 | Mean = 102.67 lb. |
| 107.5—122.5 | 397 | |
| 122.5—137.5 | 91 | |
| 137.5 and above | 23 | |
| Total | 1437 | |

3. (a) Data relating to the acreage of land possessed by households were collected from two random samples of households in two districts as given below. Test whether the average size of land possessed per household is the same in the two districts. (10)

sample 1 0.01, 0.02, 0.12, 0.23, 0.58, 0.89, 1.12, 1.30, 1.28, 1.40, 1.52, 2.11, 2.59, 4.00, 7.19.

sample 2 0.16, 0.32, 0.41, 0.83, 0.57, 0.76, 0.62, 1.25, 3.79, 7.48, 10.56, 12.19.

(b) In a public opinion survey on a political proposal, the attitude of the interviewees was recorded under three groups : "Favour", "oppose", "undecided". A total number of 3759 individuals were interviewed of whom 1872 were men and the rest women. 2237 individuals were in favour of the proposal and 917 were opposed to it. 243 men were undecided and 442 women were opposed to the proposal. Do you justify or contradict the hypothesis that there is no association between sex and attitude? (12)

4. Correlation Coefficients were calculated from 100 sets of observations of $x_1, x_2, x_3$. If $\gamma_{12} = 0.761$, $\gamma_{13} = 0.036$, $\gamma_{23} = 0.782$ calculate $R_{1 \cdot 23}$ and $\gamma_{12 \cdot 3}$ and test their significance. (12)

5. The following data relate to the yields in suitable units of six varieties of potatoes in a randomized block experiment. Test whether the varieties are significantly different among themselves. (22)

| Variety \ Block | I | II | III | IV | V | VI |
|---|---|---|---|---|---|---|
| A | 390 | 282 | 384 | 323 | 432 | 259 |
| B | 527 | 489 | 422 | 384 | 411 | 326 |
| C | 504 | 475 | 483 | 439 | 466 | 500 |
| D | 416 | 415 | 334 | 448 | 452 | 366 |
| E | 633 | 488 | 481 | 505 | 594 | 420 |
| F | 652 | 571 | 646 | 620 | 617 | 505 |

STATISTICIAN'S DIPLOMA EXAMINATION, 1954.

PAPER VII : PRACTICAL

Time : 6 hours Full marks : 100

*N.B.* (a) Figures in the margin indicate full marks.
(b) Use of calculating machines is permitted.

1. The following table shows the production of wheat in U.S.A. during the period 1900–1938. Fit a polynomial trend of suitable degree to the data. Represent the data and the trend values on a graph.

| Year | Wheat production in U.S.A. (million bushels) | Year | Wheat production in U.S.A. (million bushels) |
|---|---|---|---|
| 1900 | 603 | 1920 | 833 |
| 01 | 789 | 21 | 815 |
| 02 | 725 | 22 | 868 |
| 03 | 664 | 23 | 797 |
| 04 | 597 | 24 | 863 |
| 05 | 727 | 25 | 677 |
| 06 | 757 | 26 | 834 |
| 07 | 638 | 27 | 875 |
| 08 | 645 | 28 | 913 |
| 09 | 700 | 29 | 822 |
| 1910 | 635 | 1930 | 890 |
| 11 | 621 | 31 | 932 |
| 12 | 730 | 32 | 740 |
| 13 | 763 | 33 | 529 |
| 14 | 891 | 34 | 526 |
| 15 | 1026 | 35 | 626 |
| 16 | 636 | 36 | 627 |
| 17 | 630 | 37 | 876 |
| 18 | 637 | 38 | 932 |
| 19 | 968 | | |

(20)

2. A factorial experiment involving three fertilizers N, P, K each at two levels was carried out in 12 blocks, each of 4 plots, the layout being a confounded one. Analyse the data and write a non-technical report explaining the conclusions drawn from the results of the experiments (20)

Layout and Yields for a Confounded 2×2×2 Factorial Experiment.

| Block 1 | | Block 2 | | Block 3 | | Block 4 | |
|---|---|---|---|---|---|---|---|
| NK | 159 | P | 153 | O | 145 | K | 189 |
| O | 179 | NPK | 202 | PK | 191 | P | 272 |
| PK | 135 | N | 153 | NK | 300 | N | 160 |
| NP | 130 | K | 182 | NP | 240 | NPK | 305 |

| Block 5 | | Block 6 | | Block 7 | | Block 8 | |
|---|---|---|---|---|---|---|---|
| NPK | 155 | O | 159 | P | 266 | NK | 300 |
| N | 191 | PK | 210 | NPK | 240 | PK | 233 |
| K | 151 | NK | 138 | K | 159 | O | 226 |
| P | 188 | NP | 129 | N | 278 | NP | 182 |

| Block 9 | | Block 10 | | Block 11 | | Block 12 | |
|---|---|---|---|---|---|---|---|
| NP | 119 | NPK | 128 | NPK | 173 | NK | 213 |
| PK | 179 | N | 115 | K | 93 | O | 170 |
| NK | 77 | P | 154 | N | 209 | PK | 224 |
| O | 92 | K | 143 | P | 186 | NP | 245 |

3. A producer of electric bulbs, in his desire for putting only good bulbs for sale, rejects all bulbs for which a certain quality characteristic $x$ of the filament is below 65 units. Assuming that the quality characteristic $x$ and the life of the bulb in hours $y$ are jointly normally distributed with (15)

| | $x$ | $y$ |
|---|---|---|
| mean | 80 | 1100 |
| standard deviation | 10 | 100 |
| correlation coefficient | 0.60 | |

find:

(1) the proportion of bulbs produced that will burn for less than 1000 hours.
(2) the proportion of bulbs produced that will be put for sale
(3) the average *extra* life of bulbs not burnt out before 950 hours of burning
(4) the average life of bulbs put for sale.

4. It is known that the quality characteristic of a manufactured product is normally distributed with an initial mean of 113.7 units and a standard deviation of 0.8 units. Samples of four are taken every hour and a three-sigma control-chart of the sample mean is maintained at this level. If because of tool-wear the mean continuously falls off at the steady rate of 0.085 units per hour, after how many hours will the probability of point on the $X$ control chart going below the lower control line be $\geq 95\%$? (15)

5. Fill up the missing entries (*) in the following life table : (15)

| $x$ | $l_x$ | $d_x$ | $1000q_x$ | $L_x$ | $T_x$ | $e^o_x$ |
|---|---|---|---|---|---|---|
| 25 | 78,046 | * | * | * | * | 39.60 |
| 26 | 77,614 | 440 | * | * | * | * |
| 27 | * | * | * | * | * | * |
| 28 | 76,723 | * | 6.06 | * | * | * |
| 29 | * | * | * | * | * | * |
| 30 | * | * | * | 75,532 | 2,705,310 | * |

6. The following table gives the intercorrelations of scores for a battery of 5 tests administered to a group of persons : (15)

| | | | | |
|---|---|---|---|---|
| 1.000 | .915 | .830 | .918 | .892 |
| | 1.000 | .909 | .951 | .930 |
| | | 1.000 | .970 | .922 |
| | | | 1.000 | .949 |
| | | | | 1.000 |

On the assumption that there exists only a single common factor estimate the factor loadings *either* by Hotelling's method of principal component *or* by Thurstone's centroid method.

Also find out the residual matrix.

---

## STATISTICIAN'S DIPLOMA EXAMINATION, 1954.

PAPERS IV & V (SPECIAL SUBJECT) : THEORIES OF INFERENCE (THEORETICAL)

Time : 4 hours Full marks : 100

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

1. Describe Neyman and Pearson's method of determining similar regions for testing a composite hypothesis and of selecting from amongst them a best critical region with regard to a family of alternative hypotheses. Following this method derive the best test for equality of means of two normal populations, assuming their variances to be the same.

Where uniformly most powerful tests do not exist, do you consider unbiased locally most powerful tests as next best ? Show that for a composite hypothesis $p_\theta$ with one unspecified parameter $\theta$ and for which $\phi' = A + B\phi$ $\phi$ being $\frac{\partial \log p_\theta}{\partial \theta}$ the unbiased locally most powerful critical region is formed by combining sections of the $\phi$ surfaces within which $\frac{\partial^2 p_\theta}{\partial \theta^2} - k \frac{\partial p_\theta}{\partial \theta} - l p_\theta \geq 0$ $k$ and $l$ being suitably chosen constants.

2. Define the information limit to variance for the estimation of $\psi(\theta)$ where $\theta$ is the single parameter occurring in the probability density function $f$.

Show that distributions admitting estimates with the information limit to the variance should satisfy

$$f = f_1 \exp(t\Theta_1 + \Theta_2)$$

where $f$ and $t$ are functions of the observations only and $\Theta_1, \Theta_2$ are functions of $\theta$ only. Illustrate on the distribution.

$$f(x) = \frac{a^p}{\Gamma(p)} e^{-ax} x^{p-1}$$

$p$ being the parameter to be estimated.

Prove that minimum variance estimates are necessarily functions of sufficient statistics.

3. Defining a sequential probability ratio test, show that the probability that the test terminates is 1. If in a normal population with unit variance, $H_0$ and $H_1$ specify $\mu_0$ and $\mu_1$ for the mean, write down the expression for the probability that the test for $H_0$ against $H_1$ will terminate after the $n$th observation. Explain how it is possible to know approximately the average number of observations required for terminating the test procedure.

Develop a sequential test of the composite hypothesis that the mean of a normal population is $\mu_0$.

4. If $x_1, x_2, \ldots, x_p$ are $p$ variates distributed in a multivariate normal form, obtain Fisher's linear discriminant function of these variates on the basis of two samples from two $p$-variate normal populations having the same dispersion matrix. Show that this is the same as Mahalanobis's $D^2$.

What criterion would you use for testing the hypothesis that there is no difference in the mean values of the $p$ variates in the two populations? Obtain the distribution of this statistic.

5. Describe Wald & Wolfowitz's run test for testing the hypothesis that the distribution functions of $x$ and $y$ are identical. Can you suggest any improvement by using the lengths of the $x$- and $y$-runs.

Outline the following non-parametric tests for the same hypothesis:

1. Wilcoxon's Test
2. Dixon's Test
3. Statistical Sign Test.

Explain how the concept of power in the Neyman-Pearson's sense could be applied to non-parametric tests. Can you use this concept to compare the efficiency of a non-parametric test with that of a parametric test of a hypothesis when a parametric test is possible. Define the property of consistency of a non-parametric test and state which of the above four tests are consistent.

6. How will you test whether a sample median differs significantly from a given population median $\rho$? If $\rho$ is not known, how would you find an interval estimate for it?

Describe briefly the median tests for a linear hypothesis developed by Brown and Mood.

If $x_1, x_2, \ldots, x_N$ be observations on $N$ chance variables $X_1, X_2, \ldots, X_N$. Obtain a test of the hypotheses that $X$'s are independent and indentically distributed.

7. Define a randomized decision function and show how it can be used to determine uniquely a procedure for making experimentation and selecting a terminal decision.

Define the risk function and show how Wald used this function to develop Bayes and Minimax solutions of the decision problem. Suppose $\Omega$ contains two elements $F_1$ and $F_2$ with probability density functions $f_1$ and $f_2$ and $D^t$ contains trio elements $d_1^t, d_2^t$ and that experimentation is performed in one stage with the observations on $X_1, X_2, \ldots, X_N$. Show that the following is the necessary and sufficient condition for $\delta$ to be a Bayes solution relative to an a priori distribution $\xi$ on $F$:

$$\delta(x_1, \ldots, x_N) = 0 \quad \text{whenever} \quad \frac{f_2(x_1, x_2, \ldots, x_N)}{f_1(x_1, \ldots, x_N)} > h$$

and $$\delta(x_1, \ldots, x_N) = 1 \quad \text{whenever} \quad \frac{f_2(x_1, \ldots, x_N)}{f_1(x_1, \ldots, x_N)} < h$$

$h$ being a function of $\xi_1, \xi_2$ the components of the vector $\xi$ and the weight functions $W(F_1, d_2^t)$, $W(F_2, d_1^t)$

---

## STATISTICIAN'S DIPLOMA EXAMINATION, 1954

### Papers IV & V (Special Subject) : Sample Surveys I (Theoretical)

Time : 4 hours Full marks : 100

*N.B.* (a) Attempt any four questions.
(b) All questions carry equal marks.
c) Treatment should be mathematical wherever possible.

1. Find the expectation of the sample variance for a simple random sample from a finite population.

Also obtain the coefficient of variation of the estimated variance (you may restrict to sampling with replacement only). Derive an approximate expression for the coefficient of variation of the estimated standard deviation. Are these results of any practical importance ?

2. Write a fairly detailed note on the problem of allocation of sample to strata in a stratified simple random sampling scheme. (Restrict to the case of the usual simple unbiased estimate of the population total).

Explain how you would evaluate the gain in precision due to stratification from the results of an actual sampling enquiry with a stratified random sampling design. Do stratification always give rise to gain in precision ?

3. From a random sample of $n$ elements a random sub-sample $n'$ elements are duplicated and added to the original sample. Prove that the mean based on these $(n+n')$ elements is an unbiased estimate of the population mean and its variance is greater than the variance of the mean based on the original $n$ elements by the approximate factor,

$$1+\frac{M-1}{(M+1)^2}$$

where $M=\frac{n}{n'}$. What is the effect of elimination instead of duplication of these $n'$ elements?

In the tabulation of the results of a multi-purpose survey with a stratified simple random sampling scheme it is desired to obviate the trouble of weighting separately the strata means (or totals) by having recourse to the principle of random substitution or elimination of sample units. give the necessary theory which will help to approach the optimum extent of substitution or elomination;—use of simple cost function. (Restrict to 2 strata only).

4. Obtain an approximate expression for the standard error or a ratio estimate (for the case of simple random sampling). Under what conditions is the ratio estimate more precise than the simple unbiased estimate? What evidence is there to show that the bias of the ratio estimate decreases faster than the standard error as the size of the sample increases?

Compare the two ratio estimates which are usually associated with a stratified simple random sampling scheme in respect of (a) bias, (b) sampling variance.

5. Discuss carefully various aspects of the problem of sampling on two successive occasions with particular reference to the question of estimating (1) means, (2) change and (3) means and change jointly in the following case:

A simple random sample (srs.) of $n$ units is selected on the first occasion and a srs. of $fn$ of these units are retained for the second occasion and these are supplemented by $(1-f)n$ independently selected units. Use linear unbiased estimates of the form $a_1\bar{x}_1+a'_1\bar{x}'_1+a_2\bar{x}_2+a'_2\bar{x}'_2$ where $\bar{x}_i$ is the mean per unit for the $i$-th occasion for the units common to the two samples and $\bar{x}'_i$ is the one corresponding to units which are not common ($i=1, 2$). Also for the sake of simplicity assume sampling to be with replacement.

6. Write a note on sampling without replacement with variable probabilities of selection. Explain various alternative methods of obtaining unbiased estimates of the population mean. Find the sampling variance of such estimates and derive methods of estimating these from the sample. Do you consider Yates and Grundy's estimator of the sampling variance to be better than that of Horvitz and Thompson? Restrict to the case of (unistage) stratified sampling where *two* units are selected from each stratum.

7. Find the variance of simple unbiased estimate of a total for two-stage sample design in which the units in both the stages are selected with equal probability, the primary sampling units (psu) being of unequal size. Discuss the effect of variation in size of the psu on the variance. Explain the role of selection with varying probability in "controlling" the size of psu and compare it with other methods available for the purpose.

STATISTICIAN'S DIPLOMA EXAMINATION, 1954.

PAPERS IV & V (SPECIAL SUBJECT) : ECONOMIC STATISTICS (THEORETICAL)

Time : 4 hours | Full marks : 100

*N.B.* (a) Attempt any six questions.
(b) All questions carry equal marks.

1. Describe the additional information that has been made available by the 1951 Census of population in India as compared with the 1941 Census. How far is this information useful in getting an occupational pattern of the rura and urban populations.

2. Describe the present official statistical organizations in India both at the Centre and in the States. Suggest improvements in them, if any.

3. Compare the Fixed-Base method with the chain method of calculating a series of index numbers. Does the Chain Index have a cumulative error ?

4. What are the difficulties raised by government expenditures and taxes in the estimation of national income ? How are these difficulties overcome ?

5. State some of the difficulties that have to be faced in constructing an index number of industrial production and explain how they may be solved. What special technique has to be adopted in the case of a seasonal industry such as sugar ?

6. Graduation by the least square method leads to the same results as by the method of weighted moving average. Illustrate this by fitting a cubic to five consecutive items in a time series and showing that this is in effect a moving average with weights $-3$, 12, 17, 12 and $-3$ respectively.

7. Discuss the general problem of business forecasting. Briefly describe some of the methods commonly employed and consider their validity.

8. In calculating demand elasticities from family budget data, state and justify your preference between the alternatives given below :

(i) quantity vs. expenditure as dependent variable
(ii) income vs. total expenditure as independent variable.

9. What are the assumptions that have to be made in deriving statistical demand curves from Marshall's demand curve for an individual specially with regard to (a) stability in "other conditions" and (b) transition from individual to market demand ? What expenditure have been adopted in order to make the assumptions valid ?

STATISTICIAN'S DIPLOMA EXAMINATION, 1954.

PAPERS IV & V (SPECIAL SUBJECT) : STATISTICAL QUALITY CONTROL (THEORETICAL)

Time : 4 hours Full marks : 100

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

1. (a) In a manufacturing process tolerances specified on technical grounds are found to be much wider than the $3\sigma$ control limits. Explain how you would modify the control chart for variables in this case so as to avoid taking unnecessary corrective action.

(b) Explain the applications of the theory of runs in (1) standard control charts and (2) control chart for moving averages.

2. (a) Discuss the relationship between the range and the standard deviation in samples from a Normal population. Examine the efficiency of the range of sample as an estimate of the population standard deviation; for what sample size is this efficiency a maximum.

(b) Explain the use of range in control charts. Why are the upper and lower limits for the range chart not equidistant from the central line ?

3. Give details of any variable inspection plan for acceptance inspection for percent defective. What are the advantages of such a plan over inspection by attributes ?

Discuss under what conditions would you reduce or tighten inspection under such a plan.

4. Prove that the probability is unity that the sequential probability ratio test procedure will eventually terminate.

Derive a sequential test procedure for discriminating between two values of the standard deviation, $\sigma_1$ and $\sigma_2$ of a normal population with a known mean. Draw a sketch of the OC curve for this test of strength $\alpha, \beta$

5. Explain how you would build your own sampling acceptance plans having resource to the standard sampling inspection tables.

Explain the use of the control chart for variables for acceptance purposes.

6. Examine critically the follwing statements :
(i) "The Control chart is a mechanical form of analysis of variance".
(ii) "Corresponding to any statistical test of significance one can set up a Control chart."

7. Write short notes on the following :
(a) Simon Charts
(b) Percentage sampling inspection
(c) Control chart for defects when sample size varies
(d) ASN curve.

---

STATISTICIAN'S DIPLOMA EXAMINATION, 1954.

PAPERS IV & V (SPECIAL SUBJECT) : MATHEMATICAL THEORY OF SAMPLING DISTRIBUTIONS (THEORETICAL)

Time : 4 hours Full marks : 10

*N.B.* (a) Attempt any five questions.

(b) All questions carry equal marks.

1. The distribution of the 2 dimensional random variable $(X, Y)$ is a bivariate normal. Find the distribution function of the ratio $\frac{X}{Y}$.

2. $\chi^2$ denotes the random variable with chi-square distribution with $n$ degrees of freedom. Show that for suitably chosen constants $a_n$ and $b_n$ $(b_n>0)$ the characteristic function of $\frac{\chi^2-a_n}{b_n}$ tends to the characteristic function of the normal distribution. From this what do you conclude about the behaviour for large $n$ of Prob $\left(\frac{\chi^2-a_n}{b_n} \leqq x\right)$? Give reasons for your statements.

3. $X_1, X_2, \ldots, X_n$ are independent random variables all having the same frequency function (probability density)

$$f(x,\theta) = \begin{cases} \frac{h(x)}{H(\theta)} \text{ for } 0<x<\theta \\ 0 \text{ otherwise} \end{cases}$$

Here $h(x) \geqq 0$ and $H(\theta)$ is a suitable constant.

Find the value of $\theta$ as a function of the observations $x_1, x_2, \ldots, x_n$ for which the product $f(x_1,\theta), \ldots, f(x_n,\theta)$ is a maximum. Denoting this value by $\hat{\theta}_n$ and taking $h(x)=1$ if $0 \leqq x \leqq \theta$ and 0 otherwise, find the limiting distribution of

$$n\left(1-\frac{\hat{\theta}_n}{\theta}\right) \quad \text{as } n \to \infty$$

4. Derive the distribution function of the correlation coefficient of a sample size $n$ from a bivariate normal population in which the correlation coefficient $\rho \neq 0$

5. $X$ is a random variable having the (conditional) Poisson distribution with probabilities $\frac{a^k e^{-a}}{k!}$ $(k=0,1,\ldots,)$ where the parameter a is itself a random variable with the frequency function $\frac{a\lambda}{\Gamma(\lambda)} a^{\lambda-1} e^{-a}$ $(a>0)$

Show that

$$\text{Prob } (X=k) = \left(\frac{a}{1+a}\right)^{\lambda} \binom{-\lambda}{k} \frac{(-1)^k}{1+a)^k} (k=0,1,2,\ldots\ldots)$$

Find the mean, variance and the generating function of $X$.

6. Show that the mean $\bar{x}$ and variance $S^2$ of a sample of size $n$ from a normal population are independent. State and prove the analogous property for a multivariate normal sample.

7. Obtain the distribution of the Studentized $D^2$-statistic for two samples from two $p$-variate normal populations with the same dispersion matrix.

---

## STATISTICIAN'S DIPLOMA EXAMINATION, 1954.

PAPERS IV & V (SPECIAL SUBJECT) : DESIGN OF EXPERIMENTS 2—CONSTRUCTION OF DESIGNS (THEORETICAL).

Time : 4 hours Full marks : 100

*N.B.* (a) Attempt any six questions.
(b) All questions carry equal marks.

1. Prove that if $S=p^n$ where $p$ is any prime and $n$ any positive integer, there exists a set of $(S-1)$ orthogonal squares of side $S$.

2. Show that the maximum number of factors that can be accommodated in a symmetrical factorial experiment in which each factor is at $S=p^n$ levels (where $p$ is a prime integer) there being $S^r$ plots in every book, subject to the condition that no main effect or two-factor interaction is confounded is $(S^r-1)/(S-1)$.

3. Explain the *Principle of Generalised Interaction* with reference to a symmetrical factorial experiment involving $m$ factors at $S$ levels each. Illustrate its use in constructing confounded designs for an experiment where $m=S=3$.

4. Define *arrays of strength* d *and hypercubes of strength* d. Discuss a general method of constructing the latter.

5. Using the theory of finite Abelian groups, give an account of the principles of fractional replication with particular reference to the $2^n$ and $3^n$ factorial experiments.

6. Show that it is impossible to construct a balanced incomplete block design with the parameters :—

$$v = 15, \quad k = 5, \quad r = 7, \quad b = 21, \quad \lambda = 2.$$

7. Define a Youden square. Show that such a square can *always* be constructed from a set of balanced incomplete blocks in which the number of varieties is equal to the number of blocks.

8. Show that a connected partially balanced incomplete block design with just two associated classes and two replications must belong to one or other of the following series :—

$$\text{I} : \quad v = p^2, \quad k = p$$

$$\text{II} : \quad v = \tfrac{1}{2}p(p-1) \quad k = (p-1)$$

$$\text{III} : \quad v = 3p \quad k = 2p$$

---

## STATISTICIAN'S DIPLOMA EXAMINATION, 1954.

PAPERS IV & V (SPECIAL SUBJECT) : SAMPLE SURVEY–APPLIED (THEORETICAL).

Time : 4 hours — Full marks : 100

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

1. Discuss how will you plan a pilot survey for collecting relevant information regarding cost and variance factors, which information will be used later in planning a large-scale sample survey for estimating the acreage under wheat in a particular state (or province) of India.

2. Describe a suitable plan for a sample survey for estimating the yield-rate of wheat in a state (or province) of India, and prepare a rough budget of the expenditure to be incurred. A brief note on the budget should also be given justifying the different heads of expenditure. (The total expenses, inclusive of the preparation of the sample designs, forms etc., the field-work, and the final processing of the data and submission of the report, should not exceed sixty thousand rupees).

3. Write a note on the observational errors in sample surveys, including the following points :—(i) general nature of such errors, (ii) any suitable probabilistic model approximately representing these errors, (iii) experimental evidence in support of the model in (ii), (iv) the effect of these errors on the estimate (say, of the average) and its standard error, and (v) methods of controlling such errors, if necessary.

4. Discuss the problem of "non-response" in a socio-economic survey, and the methods for solving the problem.

5. Write a note on the problem of constructing the strata for stratified sampling, including the following points :—(i) suitable characteristic for the construction of strata, (ii) determination of the *number* of strata, and (iii) determination of the boundaries between the strata.

Discuss also the method of stratification *after* sampling when it is not possible to stratify before

6. Describe and discuss any large-scale sample survey recently conducted in India, or in any part of India.

7. a) Write out an instruction sheet (in English) for an enumerator explaining the uses of "random sampling numbers" for selecting randomly 10 per cent of the families in a village, and to select randomly one member from each selected family.

b) There is a region consisting of 1000 villages and there are 100 families in each village. Suppose 20 families are selected from the region in either of the following four alternative ways :— i) 20 families are selected at random from all the 100,000 families (ii) a village is selected at random and a family is selected randomly from the village; this procedure is repeated 20 times; (iii) 10 villages are selected randomly, and 2 families are selected randomly from each selected village; and (iv) 10 villages are selected not randomly, but subjectively, and 2 families are selected randomly from each selected village.

Discuss and compare these four types of sampling. (The distinction between sampling with replacement and sampling without replacement may be ignored).

8. Try any *two* of the following :—

(a) "Usually the two major objectives of forestry survey (by sampling) are the estimation of timber volume and the construction of a contour map. It has been frequently urged that the latter (map construction) leave no practical alternative to the systematic pattern of the sampling units". Discuss the statement quoted above.

(b) Discuss how to estimate by sampling *either* the number of squirrels in a small closed forest, *or* the number of fish in a tank, using the capture-recapture method.

(c) Suppose you are required to sample from a number of villages with probabilities proportional to their geographical areas, and a map showing the boundaries of these villages is available. Discuss whether the location of "points" at random on the map (say, by random choices of rectangular cartesian co-ordinates, $x$, $y$) will be useful in this connection.

---

## STATISTICIAN'S DIPLOMA EXAMINATION, 1954.

### PAPERS IV & V (SPECIAL SUBJECT) : VITAL STATISTICS AND POPULATION STUDIES (THEORETICAL)

Time : 4 hours — Full marks : 100

*N.B.* (a) Attempt any four questions.

(b) All questions carry equal marks.

1. Describe in detail the method adopted in calculating the survival probabilities ${}_{10}p_x$ and $p_x$ in the different age ranges for the Census of India 1941-50 Life Tables : discuss its comparative merits in the context of the methods applied in construction of the earlier Census of India Life Tables.

If a curve of the type $l = A + H_x + Bc^x + \frac{m}{nx+1}$ is used to extrapolate for the life table values of survivors at ages $0 \leqq x \leqq 5$ utilising supplementary information about infant and child mortality, suggest how the constants should be determined.

2. Show that in a stable population with constant husband-wife age difference $d$ at marriage, the consistent female net reproduction rate $\bar{R}_{0F} = R_{0M} e^{-rd}$ where $R_{0M}$ is the male net reproduction rate and $r$ the rate of natural increase of the population. Also show that a first approximation to $r$ is given by $\frac{1}{G_F} \log_e R_{0F}$ where $G_F$ is the female mean generation interval.

Can Lotka's fundamental equation $\int_0^\infty e^{-rx} x p_0 b_x \, dx = 1$ be applied to sections of a population like population in working ages (14 – 59) or in old age (60 +) to obtain the natural rate of increase of the sections? Give illustrations and reasons.

3. Define infant mortality. Why is the infant mortality rate liable to be more inaccurate than either birth or death rates when derived from the same source of defective birth and death registration data?

The proportions of children dying in the first year of life out of every 1000 children born to couples classified in marriage cohort groups, taken from by a recent sample survey on an all-India frame, are given in the table below:

| type of couple | marriage period of couple | | | | |
|---|---|---|---|---|---|
| | before 1910 | 1910–1919 | 1920–1929 | 1930–1939 | 1940–1951 |
| 1. couples with total children born 1 or 2. | 94 | 126 | 91 | 118 | 127 |
| 2. all couples | 88 | 102 | 126 | 134 | 143 |

Information was collected by the interview method for all sample couples with at least one partner alive on the date of survey.

Comment on the salient features of the results; with what official statistics would you compare them and in what manner?

4. Set out the principles and facts on which the Registrar General relied to draw the line of 'improvement maternity' in his Report on Census of India, 1951. Keeping in view the general experience of demographic behaviour and probable developments, comment on the conclusion in the Report that by putting an end to improvident maternity, as defined, the excess of births over deaths will be reduced and a substantially stationary population achieved.

5. What statistics of international migrations are available for India? Indicate broadly the defects and the gaps. Give the dimensional magnitude of the total net migration from India during the present century and rank the principal receiving countries. List the broad standard categories into which international migrations have been classified, separately for arrivals and departures.

Study of internal migration in India with particular reference to urbanisation is under contemplation. Suggest a basic structure for an ad hoc internal migration survey to specified regions mentioning the principal classifications with respect to type and cause of migration that you would recommend.

STATISTICIAN'S DIPLOMA EXAMINATION, 1954.

PAPERS IV & V (SPECIAL SUBJECT) : DESIGN OF EXPERIMENTS I—APPLIED (THEORETICAL)

Time : 4 hours Full marks : 100

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

1. In the case of a randomised block experiment on $v$ variation in $r$ blocks (of $v$ plots each) repeated over $p$ places, construct a mathematical model to show how the yield data can be analysed into various components (the $p$ places being regarded as a random sample from a population of places). Exhibit also the expectation of the mean square (a) for error, (b) for variety and (c) for interaction between varieties and places.

2. (a) Discuss the occasions for and the advantages of the use of hierarchal classification in statistical practice, as applied to design of experiments.

(b) Explain the idea of local control and orthogonality in the design of experiments.

3. What is the main difference between the "switch-over design" and the "long-term experiment" in the case of treatments applied in sequence ? Consider the following pattern for testing 3 nutritional treatments on the milk yield of dairy cows in which the total experimental time of 3 months is divided into 3 periods each of one month :

| | | Cow | | |
|---|---|---|---|---|
| | | 1 | 2 | 3 |
| Period | I | a | b | c |
| | II | b | c | a |
| | III | c | a | d |

The treatments are denoted by a, b, c. If $r$ sets of 3 cows are used, with a random square for each set, and if the error variance is homogeneous over the sets, show the analysis of variance resulting from a reasonable mathematical model to be suggested by you.

4. (a) Put a case for the use of Factorial exerpiments as against comparable experiments in which all factors except one are kept constant, and a series of experiments is performed for investigating the effect of different factors.

(b) Give the plan and the method of analysis of a $3 \times 3 \times 3$ Factorial experiment in 4 replications in which a set of 2 d.f. of the triple interaction is totally confounded.

5. (a) "A suitably chosen sub-set of the full factorial set can provide worthwhile information"—Comment fully on the above statement.

(b) What are Affine Resolvable B.I.B. design ?

Show that in the designs $b = v + r$ , where $b$ is the number of blocks, $v$ the number of varieties, and $r$ the replication of each variety.

Examine whether a design with parameters $b = 18$, $v = 10$, $r = 9$, $k = 5$, $\lambda = 4$ can be a resolvable design, where $k$ is the number of varieties (all different) in a block, and $\lambda$ the number of times a pair of varieties occurs together in a block.

6. (a) In a balanced incomplete block design if the model for the intra-block comparisons as taken as $y_{ij} = \mu + b_i + v_j + e_{ij}$

where the suffixes $i$ and $j$ refer to blocks and varieties respectively; and $y$, $\mu$, $b$, $v$ and $e$ refer to the plot yield, mean yield, effect of block, effect of treatment, and random fluctuations (with zero mean and a constant s.d.), show that the equation for estimating treatment differences are :

$$\left(N_{oj} = \sum_i \frac{n_{ij}^2}{N_{io}}\right)\hat{v}_j - \sum_{j' \neq j}\left(\sum_i \frac{n_{ij}\, n_{ij'}}{N_{io}}\right)\hat{v}_{j'} = Q_j \quad (j = 1, 2, .., v)$$

where $n_{ij}$ is the number of times the treatment $j$ occurs in block $i$,

$$N_{oj} = \sum_i n_{ij}, \text{ and } Q_j = Y_{oj} - \sum \frac{n_j}{N_{io}}\, Y_{io}$$

Clearly state the assumptions underlying the mathematical treatment.

(b) If from a given symmetrical B.I.B. design one block is omitted as also the varieties contained in this block, show how the parameters of the original design are changed to form the new design. In the case of $\lambda = 2$, show how this new design is changed if it is "truncated" as above.

7. Write short notes on any *two* of the following :

(a) Analysis of mixed up yields in a randomised block experiment (two yields having been mixed, so that their sum only is known).

(b) Relative efficiency of an ungrouped randomised experiment, randomised block experiment, and a Latin square design under comparable circumstances.

(c) Analysis of non-orthogonal data resulting from disproportionate numbers of items in the classes in a randomised block experiment.

(d) Covariance analysis in a Latin square design.

---

## STATISTICIAN'S DIPLOMA EXAMINATION, 1954.

### PAPERS VIII & IX : THEORIES OF INFERENCE (PRACTICAL)

Time : 4 hours — Full marks : 50

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

✓ 1. Blood Plasma Chlorides in e.g. ($x_1$), Free Acid in c.c. ($x_2$) and Total Acid in c.c. ($x_3$) obtained from gastric analysis were recorded in the following table for 17 normal healthy persons, twentyfour hours before and after the Percoten Intra-muscular injection. Examine whether the injection has produced any effect on the characters, $x_1$, $x_2$, and $x_3$.

| sl. | 24 hours before injection | | | 24 hours after injection | | |
|---|---|---|---|---|---|---|
| no. | $x_1$ | $x_2$ | $x_3$ | $x_1$ | $x_2$ | $x_3$ |
| 1 | 56.0 | 43.0 | 56.5 | 60.0 | 64.0 | 78.5 |
| 2 | 58.0 | 22.0 | 42.0 | 60.0 | 26.0 | 44.0 |
| 3 | 58.0 | 44.0 | 69.0 | 62.0 | 41.5 | 58.5 |
| 4 | 52.0 | 00.0 | 18.5 | 60.0 | 26.0 | 41.5 |
| 5 | 54.0 | 00.0 | 22.5 | 58.0 | 00.0 | 19.5 |
| 6 | 58.0 | 60.0 | 86.0 | 56.0 | 79.0 | 97.5 |
| 7 | 58.0 | 11.0 | 29.5 | 60.0 | 20.0 | 41.0 |
| 8 | 56.0 | 17.5 | 35.0 | 62.0 | 38.5 | 53.5 |
| 9 | 56.0 | 67.5 | 85.0 | 60.0 | 00.0 | 6.5 |
| 10 | 55.0 | 29.0 | 47.0 | 58.0 | 6.3 | 9.0 |
| 11 | 56.0 | 44.5 | 62.5 | 58.0 | 45.0 | 60.5 |
| 12 | 56.0 | 00.0 | 10.0 | 56.0 | 35.5 | 52.0 |
| 13 | 52.0 | 00.0 | 7.5 | 56.0 | 62.5 | 78.5 |
| 14 | 58.0 | 00.0 | 14.0 | 60.0 | 00.0 | 7.5 |
| 15 | 48.0 | 0.0 | 29.5 | 62.0 | 44.0 | 66.5 |
| 16 | 58.0 | 32.5 | 52.5 | 64.0 | 00.0 | 23.0 |
| 17 | 56.0 | 00.0 | 15.0 | 62.0 | 85.0 | 102.5 |

2. The following figures relate to observations in two independent samples :—

Sample 1 : 25, 30, 28, 34, 24, 20, 13, 32, 22, 38, 31, 35, 7, 10, 19

Sample 2 : 40, 34, 22, 8, 47, 31, 40, 30, 23, 36, 17, 21, 27, 29, 26

(a) Test whether the two samples have come from the same population by applying the following non-parametric tests.

1. Wald-Wolfowitz Run Test
2. Mann & Whitney's Test
3. Statistical Sign Test
4. Wilcoxon's Test.

(b) Test the hypothesis that the medians of the two populations from which the above samples are drawn are equal. If the hypothesis is not rejected at 5% level obtain point and interval estimates of the common median. If it is rejected, obtain similar estimates for each median.

*Given*

Wald-Wolfowitz Test : $m=n=15$

| $U$ | 10 | 11 | 12 | 13 |
|---|---|---|---|---|
| $P\{U \leqslant U_0\}$ | .0190 | .0457 | .0974 | .1749 |

Wilcoxon's Test : Number of paired comparisons $n=15$

| | | | |
|---|---|---|---|
| Sum of rank numbers + or −, whichever is less. | 16 | 19 | 23 |
| Prob. of this total or less: | .0103 | .019 | .054 |

3. Construct the *OC* curve for the sequential probability ratio test for testing $H_0(p=.1)$ against $H_1(p=.25)$ where $H$ specifies

$$f(x,\ p)=p \quad \text{for} \quad x=0$$
$$=1-p \quad \text{for} \quad x=1.$$

Take $\alpha=\beta=.05$

Also construct the ASN curve for the same test. Compare the average size of sample required by the above sequential test when the true value of $p$ is .1 with the sample size required for the corresponding non-sequential test with each of the two types of error equal to 0.05. (You may use the normal approximation to the binomial for finding the sample size in the non-sequential case).

---

## STATISTICIAN'S DIPLOMA EXAMINATION, 1954.

### Papers VIII & IX : Sample Survey–Theory (Practical)

Time : 4 hours — Full marks : 50

*N.B.* Use of calculating machines is permitted.

1. You are given the Census (1951) Handbook of a district.

(1) Draw a random sample of 30 villages.

(2) Also draw a sample of 30 villages with probability proportional to population (with replacement).

*N.B.* Give sufficient details so that the correctness of the procedure followed by you may be verified.

(3) Estimate the proportion of "cultivators of land wholly or mainly owned and their dependents" (i.e., those belonging to the livelihood class I) to the total population of the district by using (a) the ratio method for sample (1) above, and (b) an unbiased method for sample (2).

(4) Calculate the standard errors of your estimates.

(5) Compare your estimates with the actual value and make any useful comments which you consider worth mentioning.

---

STATISTICIAN'S DIPLOMA EXAMINATION, 1954.

PAPERS VIII & IX : ECONOMIC STATISTICS (PRACTICAL)

Time : 4 hours | Full marks : 50

*N.B.* (a) Answer question No. 1 and any two of the rest.
(b) Use of calculating machines is permitted.

1. The following table gives the prices and quantities consumed of food articles per family of 3.27 units at Detroit (U.S.A.) and Copenhagen (Denmark).

| food articles | quantity consumed per family of 3.27 units (Kg.) | | price per unit (Ore) | |
|---|---|---|---|---|
| | Detroit | Copenhagen | Detroit | Copenhagen |
| (1) | (2) | (3) | (4) | (5) |
| 1. Beef fresh (steak) | 17.4 | 37.0 | 322 | 181 |
| 2. Beef fresh (roast) | 22.5 | | 273 | |
| | 39.9 | | 298 | |
| 3. Veal | 7.4 | 29.2 | 276 | 194 |
| 4. Pork (fresh) | 30.2 | 15.9 | 221 | 194 |
| 5. Pork (salt) | 24.7 | 18.5 | 247 | 200 |
| 6. Cooked meat | 7.6 | 12.5 | 258 | 248 |
| 7. Fish (fresh) | 7.9 | 20.2 | 236 | 115 |
| 8. Fish (salt) | 1.8 | 1.4 | 220 | 88 |
| 9. Milk | 507.0 | 389.5 | 52 | 32 |
| 10. Butter | 30.0 | 30.3 | 414 | 294 |
| 11. Cheese | 7.3 | 19.0 | 279 | 248 |
| 12. Lard | 22.1 | 10.4 | 135 | 156 |
| 13. Eggs | 55.0 | 29.7 | 248 | 176 |
| 14. Flour (wheat) | 87.0 | 51.3 | 40 | 34 |
| 15. Bread | 236.1 | 67.0 | 68 | 77 |
| 16. Sugar | 88.6 | 114.5 | 54 | 42 |
| 17. Potatoes | 268.3 | 275.6 | 20 | 13 |
| 18. Tea | 2.3 | 1.2 | 603 | 900 |
| 19. Coffee | 14.8 | 15.8 | 322 | 428 |

Calculate any index you think suitable for the comparison of cost of living at the two places. Write a note on the limitations of the formula you use.

2. Estimate the constants of the Pareto curve $n = Ax^{-a}$ which fit the data below and discuss to what extent the Pareto Law is verified :—

Number of incomes $n$ more than Rs. $x$ during 1948-49.

| Income ($x$) | Number ($n$) |
|---|---|
| 15,000 | 9,100 |
| 10,000 | 21,700 |
| 5,000 | 62,500 |
| 1,500 | 1,76,100 |

3. (a) From the following data test whether infant mortality is significantly different from one State to another.

| State | No. of infants | Deaths per 1,000 |
|---|---|---|
| I | 65,457 | 66 |
| II | 33,471 | 72 |
| III | 66,544 | 70 |
| IV | 40,477 | 61 |
| V | 62,041 | 58 |
| VI | 57,185 | 58 |

(b) Draw up a Lorenz curve from the following data of 1948-49 incomes in India and comment on the inequality of the distribution :—

| Grade of total income (Rupees) | No. of assessee (In hundreds) | Total income assessed (In Rs. lakhs) |
|---|---|---|
| 0— 1,500 | 32 | 28 |
| 1,501— 3,000 | 555 | 1341 |
| 3,001— 3,500 | 256 | 823 |
| 3,501— 5,000 | 325 | 1349 |
| 5,001— 7,500 | 261 | 1605 |
| 7,501—10,000 | 147 | 1296 |
| 10,001—12,500 | 79 | 879 |
| 12,501—15,000 | 47 | 642 |
| 15,001—20,000 | 42 | 710 |
| 20,001—25,000 | 25 | 561 |
| 25,001—30,000 | 11 | 287 |
| 30,001—40,000 | 8 | 257 |
| 40,001—50,000 | 3 | 126 |
| 50,001 and over | 2 | 137 |
| Total | 1793 | 10041 |

4. In a certain budget inquiry of working class families in Bombay, the following Engel equations were obtained, $E$ being the total expenditure :—

| | |
|---|---|
| Food ($e_1$) | $0.459\ E + 0.31$ |
| Housing ($e_2$) | $0.069\ E + 2.70$ |
| Clothing ($e_3$) | $0.062\ E + 0.72$ |
| Fuel & lighting ($e_4$) | $0.055\ E + 0.76$ |

The price relatives for all the constituents viz., Food, Housing Clothing, Fuel & Lighting and Miscellaneous were respectively 131, 100, 120, 112 and 115. Find the index numbers of cost of living when the income is (a) the average income Rs. 45.98. (b) Rs. 27.60 and (c) Rs. 88.75.

## STATISTICIAN'S DIPLOMA EXAMINATION, 1954.

### PAPERS VIII & IX : STATISTICAL QUALITY CONTROL (PRACTICAL)

Time : 4 hours — Full marks : 50

*N.B.* (a) Attempt any three questions.

(b) Figures in the margin indicate full marks.

(c) Use of calculating machines is permitted.

1. The following table gives averages and ranges for sub-groups of size 4. To facilitate arithmetic, deviations from the specified standard are given. Specified tolerances on individual items were ± 8 units from the standard. Plot a suitable control chart and give your comments. Is the machine working according to specifications ? (17)

| Sub-group No. | $\bar{X}$ | $R$ | Sub-group No. | $\bar{X}$ | $R$ | Sub-group No. | $\bar{X}$ | $R$ | Sub-group No. | $\bar{X}$ | $R$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | +4.5 | 2 | 11 | +3.5 | 4 | 21 | +4.5 | 4 | 31 | −7.0 | 14 |
| 2 | +6.5 | 12 | 12 | −0.5 | 14 | 22 | +4.5 | 6 | 32 | −16.0 | 17 |
| 3 | 0.0 | 4 | 13 | −2.0 | 6 | 23 | +0.5 | 8 | 33 | −14.0 | 15 |
| 4 | +6.0 | 6 | 14 | −2.0 | 6 | 24 | −3.0 | 16 | 34 | −1.0 | 15 |
| 5 | −2.5 | 8 | 15 | −7.0 | 5 | 25 | −1.5 | 13 | 35 | −9.5 | 8 |
| 6 | −3.0 | 5 | 16 | −10.0 | 15 | 26 | +0.5 | 14 | 36 | −7.5 | 30 |
| 7 | +3.5 | 5 | 17 | +7.0 | 10 | 27 | +6.0 | 26 | 37 | −8.5 | 12 |
| 8 | +15.0 | 8 | 18 | −6.5 | 10 | 28 | −8.0 | 26 | 38 | −10.0 | 10 |
| 9 | +7.0 | 11 | 19 | +3.0 | 10 | 29 | −2.0 | 10 | 39 | −8.5 | 12 |
| 10 | −4.0 | 10 | 20 | −4.5 | 8 | 30 | −2.5 | 11 | 40 | −10.0 | 13 |

2. The standard deviation of the specific gravity of bricks being manufactured is 0.0132 units. The following sampling inspection scheme is suggested: (17)

Test 4 bricks at random from each lot; if the average specific gravity of these 4 bricks is less than 2.365, accept the lot, otherwise reject it.

Under such a scheme, consider the chances of accepting lots with true mean as (i) 2.355, (ii) 2.365 and (iii) 2.373.

Draw a rough sketch of the OC curve.

3. The following table gives the daily number inspected, number defective and percent defective for two machines working on the same product. By plotting suitable control charts, state whether the two machines are producing the same level of defectives. Give reasons. Comment on the control charts. (15)

| Sample No. | Machine A | | | Machine B | | |
|---|---|---|---|---|---|---|
| | number inspected | number defective | percent defective | number inspected | number defective | percent defective |
| 1 | 140 | 6 | 4.3 | 144 | 17 | 11.8 |
| 2 | 149 | 8 | 5.3 | 132 | 13 | 10.0 |
| 3 | 153 | 9 | 5.9 | 139 | 13 | 9.3 |
| 4 | 151 | 1 | 0.7 | 113 | 6 | 5.3 |
| 5 | 149 | 9 | 6.0 | 136 | 12 | 8.8 |
| 6 | 151 | 3 | 2.0 | 127 | 12 | 9.4 |
| 7 | 110 | 0 | 0 | 127 | 9 | 7.1 |
| 8 | 120 | 5 | 4.2 | 133 | 10 | 7.5 |
| 9 | 153 | 6 | 3.9 | 142 | 6 | 4.2 |
| 10 | 150 | 3 | 2.0 | 144 | 10 | 6.9 |
| 11 | 124 | 2 | 1.6 | 133 | 7 | 5.3 |
| 12 | 90 | 8 | 8.9 | 111 | 10 | 9.0 |
| 13 | 107 | 6 | 5.6 | 135 | 4 | 3.0 |
| 14 | 141 | 4 | 2.8 | 161 | 10 | 6.2 |
| 15 | 150 | 4 | 2.7 | 148 | 7 | 4.7 |
| 16 | 124 | 4 | 3.2 | 122 | 10 | 8.2 |
| 17 | 166 | 13 | 7.8 | 139 | 7 | 5.0 |
| 18 | 155 | 7 | 4.5 | 160 | 12 | 7.5 |
| 19 | 142 | 5 | 3.5 | 130 | 8 | 5.7 |
| 20 | 143 | 5 | 3.5 | 140 | 16 | 11.4 |
| 21 | 69 | 1 | 1.4 | 128 | 3 | 2.3 |
| 22 | 109 | 4 | 3.6 | 144 | 12 | 7.5 |
| Total | 2945 | 113 | | 2997 | 214 | |

4. A standard gun is to be compared with an experimental gun on the basis of hits on a target under specified conditions. If the ratio of hits to misses for the experimental gun is 3 or more times as large as for the standard gun, it is important to decide in favour of the experimental gun; but if the ratio is 1.2 or less times as large for the experimental as for the standard gun, it is important to decide in favour of the standard. Give $\alpha = .02$ and $\beta = .05$, set up graphically a suitable sequential inspection procedure. Draw a rough sketch of the OC curve.

## STATISTICIAN'S DIPLOMA EXAMINATION, 1954.

PAPERS VIII & IX : MATHEMATICAL THEORY OF SAMPLING DISTRIBUTIONS (PRACTICAL)

Time : 4 hours Full marks : 50

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

1. Draw 20 random samples of size 5 from the Cauchy population

$$dF = \frac{dx}{\pi[(x-1)^2+1]}$$

and show that the sample means follow the same distribution.

2. Draw 25 random samples of size 4 from a normal population with mean zero and standard deviation unity and calculate $u = \frac{\bar{x}}{w}$ for each sample, $\bar{x}$ and $w$ being the mean and the range. Find the moments and determine what Pearson Type curve could be fitted to the sampling distribution of $u$.

3. The following measurements were taken on two tri-variate normal samples :—

| Sample 1 | | | Sample 2 | | |
|---|---|---|---|---|---|
| $x_1$ | $x_2$ | $x_3$ | $x_1$ | $x_2$ | $x_3$ |
| 5.1 | 3.5 | 1.4 | 7.0 | 3.2 | 4.7 |
| 4.9 | 3.0 | 1.4 | 6.4 | 3.2 | 4.5 |
| 4.7 | 3.2 | 1.3 | 6.9 | 3.1 | 4.9 |
| 4.6 | 3.1 | 1.5 | 5.5 | 2.3 | 4.0 |
| 5.0 | 3.6 | 1.4 | 6.5 | 2.8 | 4.6 |
| 5.4 | 3.9 | 1.7 | 5.7 | 2.8 | 4.5 |
| 4.6 | 3.4 | 1.4 | 6.3 | 3.3 | 4.7 |
| 5.0 | 3.4 | 1.5 | 4.9 | 2.4 | 3.3 |
| 4.4 | 2.9 | 1.4 | 6.6 | 2.9 | 4.6 |
| 4.9 | 3.1 | 1.5 | 5.2 | 2.7 | 3.9 |
| 5.4 | 3.7 | 1.5 | 5.0 | 2.0 | 3.5 |
| 4.8 | 3.4 | 1.6 | 5.9 | 3.0 | 4.2 |
| | | | 6.0 | 2.2 | 4.0 |
| | | | 6.1 | 2.9 | 4.7 |
| | | | 5.6 | 2.9 | 3.6 |

Calculate the value of the statistic

$$D^2 = \sum_{i}^{3}\sum_{j=1}^{3} w^{ij}(\bar{x}_{i1}-\bar{x}_{i2})\,(\bar{x}_{j1}-\bar{x}_{j2}),$$

where $(w^{ij})$ is the matrix reciprocal to $(w_{ij})$, the dispersion matrix estimated from the two samples together.

Under the hypothesis that corresponding population means of the variates are equal, find the probability of getting a value of $D^2$ equal to or larger than the above value.

## STATISTICIAN'S DIPLOMA EXAMINATION, 1954.

PAPERS VIII & IX : DESIGN OF EXPERIMENTS 2—CONSTRUCTION OF DESIGNS (PRACTICAL)

Time : 4 hours  Full marks : 50

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

1. (a) Using Galois field, construct a complete set of orthogonal squares of side 9.
(b) Construct the arrays (25, 6, 5, 2) and (12, 11, 2, 2).

2. Illustrate on a $3^3$ experiment in blocks of 3 plots the punched card technique developed by Rao, using the power cycle of a Galois field, for the construction of confounded designs.

3. Write down the 16 possible alias sub-groups, such as

$$I = A^3BC^2DE = AB^2CD^2E^2$$

for which no pair of degrees of freedom has more than one alias that is a main effect or two-factor interaction in the one-third replicate of a $3^5$ factorial experiment.

Construct an arrangement of the 1/3-replicate in nine blocks of nine plots in which only two degrees of freedom from one two-factor interaction are lost under any of their aliases.

4. Construct a $8 \times 8$ Quasi-Latin Square for a $2^3$ experiment in which it is desired to confound the following effects in the rows and columns :

| | | | | |
|---|---|---|---|---|
| ABC; | ADE; | BCDE | in rows | 1—4 |
| ABD; | BCE; | ACDE | in rows | 5—8 |
| ACE; | BCD; | ABDE | in columns | 1—4 |
| ACD; | BDE; | ABCE | in columns | 5—8 |

5. Construct the balanced incomplete block design :

$$v = b = 16, \qquad r = k = 6, \qquad \lambda = 2.$$

and the corresponding Youden square.

6. Construct the $n$-ple rectangular lattice for which

$$v = p(p-1), \quad k = (p-1), \quad r = n, \quad b = np \qquad 2 \leqq n \leqq p$$

for the following values of $p$ and $n$

$$p = 5; \quad n = 2, 3 \text{ and } 4.$$

Show that the lattices for $n = 2$ and 4 *are partially balanced incomplete block designs* (in the sense of Bose and Nair) with 4 and 3 associate classes respectively and that the lattice for $n = 3$ does not belong to this class of design.

## STATISTICIAN'S DIPLOMA EXAMINATION, 1954.

### PAPERS VIII & IX : SAMPLE SURVEYS—APPLIED (PRACTICAL)

Time : 4 hours Full marks : 50

*N.B.* (a) Answer question No. 1 and two others from the remaining questions.
(b) All questions carry equal marks.
(c) Use of calculating machines is permitted.

1. Draw up a questionnaire form for a family budget enquiry in any large city of India and write out instructions for the enumerators in this connection. (The form and the instructions are to be given in English).

2. For a particular type of two-stage sampling survey it is stated that the variance, $V$, for the estimate 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 units selected in each first stage unit in the sample, and $\sigma_1^2$, $\sigma_2^2$ are the two stage variances in the population. Further, the total cost, $W$, of the survey is given by

$$W = a + bm + cmn$$

The following numerical values are also supplied :—

$\sigma_1^2 = 4.2$, and $\sigma_2^2 = 12.1$ in some convenient units and $a$ = Rs. 500, $b$ = Rs. 17, and $c$ = Rs. 3.

(a) For a total cost, $W$, of Rs. 5000, work out the optimum values of $m$ and $n$ and the corresponding value of $V$; and

(b) For any other pair of values for $m$ and $n$ (i.e., different from their optimum values) with the same total cost, $W$, (i.e., Rs. 5000) work out the value of $V$.

3. To estimate the yield-rate of a crop in a region consisting of six villages only, the following sampling plan was adopted : all the villages were sampled, three plots being selected randomly in each village, and two "cuts" being made randomly inside each selected plot. The data for the yield-rate, measured in a particular unit, are shown in the Table below. (Simplifying assumptions of homoscedasticity, equal-sizes in the population, etc., may be made for both the stages, and no 'finite population corrections" need be made in any of the two stages).

TABLE SHOWING THE YIELD-RATES.

| Pl. No. / Cut No. | 1st cut | 2nd cut | 1st cut | 2nd cut | 1st cut | 2nd cut |
|---|---|---|---|---|---|---|
| | Village no. 1 | | Village no. 2 | | Village no. 3 | |
| 1st Plot | 167 | 87 | 67 | 57 | 187 | 87 |
| 2nd Plot | 267 | 317 | 57 | 167 | 37 | 107 |
| 3rd Plot | 117 | 137 | 167 | 57 | 167 | 207 |
| | Village no. 4 | | Village no. 5 | | Village no. 6 | |
| 1st Plot | 137 | 137 | 177 | 157 | 127 | 157 |
| 2nd Plot | 77 | 67 | 117 | 217 | 117 | 177 |
| 3rd Plot | 77 | 27 | 177 | 257 | 87 | 107 |

(a) Work out the analysis of variance of the given data, and find out the standard error of the sample-mean; and

(b) Instead of the sampling plan actually adopted, would you recommend an alternative plan in which two plots are taken per village, and four "cuts" from each selected plot ? Assume that the total cost involved, $W$, in Rupees, is given by

$$W = 500 + 6 \times (20 + 4m + 2mn),$$

where $m$ is the number of plots taken per village, and $n$ is the number of "cuts" taken per selected plot.

4. (a) There is a circular region of radius 10 miles and you are required to locate in it 50 "points" randomly with the help of polar co-ordinates $\gamma$ and $\theta$ ($0 < \gamma < 10$ miles, and $0° < \theta < 360°$). The $\gamma$ co-ordinates are to be chosen randomly from equal steps of $\gamma^2$ (nor $\gamma$) of 1 sq. mile, i.e., from the values of $\gamma^2 = 1, 2, 3, 4, \ldots, 99, 100$ sq. miles; $\theta$ is to be chosen randomly correct to a degree. (You are allowed the use of "random sampling numbers"; the method of random selection is to be clearly indicated).

(b) Divide the circular region into eight parts of equal area by means of the circumferences of three suitably chosen circles, concentric with the circular region, and any (one) arbitrarily chosen diameter of the circular region. Test now the hypothesis of equal probability of occurrence of the "points", chosen by you in (a) above, in all the eight parts.

*N.B.* You need not draw any diagram for either (a) or (b)

---

## STATISTICIAN'S DIPLOMA EXAMINATION, 1954.

PAPERS VIII & IX : VITAL STATISTICS AND POPULATION STUDIES (PRACTICAL)

Time : 4 hours — Full marks : 50

*N.B.* (a) Attempt question No. 1 and any one of the remaining questions.
(b) Use of calculating machines is permitted.
(c) Census of India 1931 Report, Vol. I, Part 1, will be provided for use at the examination.

1. (a) The age distribution of the population of a certain sector of India on 1-3-1952 is given below :

| Age group | Population | Age group | Population |
|---|---|---|---|
| 5— 9 | 2,634 | 40—44 | 966 |
| 10—14 | 2,337 | 45—49 | 746 |
| 15—19 | 1,706 | 50—54 | 703 |
| 20—24 | 1,705 | 55—59 | 434 |
| 25—29 | 1,489 | 60—64 | 469 |
| 30—34 | 1,324 | 65—69 | 224 |
| 35—39 | 1,063 | | |

Assuming that net migration effects in any of the age groups are negligible and the 1931 Census all-India Female Life Table mortality will be experienced by the population during the decade, estimate the total population in working age (defined as 14—58) that the sector will contribute to 1961 Census enumeration.

(b) If the mortality actually experienced during the whole of the decade by the population in the 10—19 age range be 10% lower than the 1931 Census all-India Female Life Table mortality, the experience in other age range being as expected (i.e., 1931 Census all-India Female Life Table mortality), estimate the eventual increase in the working age population at 1961 Census enumeration. State the assumptions that you have to make.

2. For the population distribution given in question no. 1 above, find out the graduated pivotal values of the population at individual ages 17, 22, etc., so far as the data permit, using the osculatory interpolation formula

$$10{,}000u_n = 66w_{n-10} - 344w_{n-5} + 2556w_n - 344w_{n+5} + 66w_{n+10}$$

where $w_n$ represents the population in quinary age group $n-2$ to $n+2$.

Develop a formula to obtain the graduated pivotal value of the population at age 12 and compute the value.

3. Assuming that the incidence of illegitimate births is negligible, work out the paternal and maternal net reproduction rates and generation intervals for the population (assumed stable) particulars of which are given in the table below:

| Wife's age at marriage groups $x$ | Husband's mean age at marriage $\bar{y}$ | Number of marriages in the generation of 10,000 | | Number of maternities per marriage of women aged $x$ $F_0(x, y)$ | Mean duration of marriage at maternity |
|---|---|---|---|---|---|
| | | men $10^4 \times u_m(x,y)$ | women $10^4 \times u_f(x,y)$ | | |
| 15—19 | 23.4 | 527 | 528 | 3.383 | 7.3 |
| 20—24 | 25.5 | 2,080 | 2,102 | 2.236 | 6.8 |
| 25—29 | 28.7 | 1,203 | 1,208 | 1.512 | 5.5 |
| 30—34 | 34.0 | 307 | 309 | 1.049 | 4.2 |
| 35—39 | 40.8 | 188 | 190 | .510 | 3.2 |
| 40—44 | 48.0 | 113 | 115 | .128 | 2.0 |
| 45—49 | 53.0 | 82 | 84 | .025 | 2.5 |

The net reproduction rates envisaged are those derived from formula of the type

$$\int_0^\infty \int_0^\infty u(x, y)\, F_0(x, y)\, dx\, dy.$$

## STATISTICIAN'S DIPLOMA EXAMINATION, 1954.

### PAPERS VIII & IX : DESIGN OF EXPERIMENTS I—APPLIED (PRACTICAL)

Time : 4 hours Full marks : 50

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

1. A $3 \times 3$ factorial experiment with two factors A and B was laid out in two replications according to the following plan :

Replicate 1 : (00, 11, 22); (10, 21, 02); (20, 01, 12).

Replicate 2 : (00, 21, 12); (10, 01, 22); (20, 11, 02).

(The first figure represents the level of A, and the second figure the levels of B).

The yield data in the two replications are shown, classified, in the following table :

Replicate 1

| | | B | | |
|---|---|---|---|---|
| | | 0 | 1 | 2 |
| A | 0 | 18 | 18 | 13 |
| | 1 | 16 | 16 | 17 |
| | 2 | 12 | 14 | 14 |

Replicate 2

| | | B | | |
|---|---|---|---|---|
| | | 0 | 1 | 2 |
| A | 0 | 19 | 13 | 19 |
| | 1 | 18 | 17 | 15 |
| | 2 | 13 | 16 | 18 |

Analyse the data, construct the yield table corrected for block effects, and examine the significance of the difference between the treatments (00 and 11), (00 and 02), (01 and 02).

2. In the table below (1) represents the lay out of an incomplete block design and (2) represents the corresponding yields :

| | | | | |
|---|---|---|---|---|
| (1) | 1,3,7,8 | 1,4,6,10 | 2,4,8,9 | 2,5,6,7 | 3,5,9,10 |
| (2) | 10,7,11,13 | 8,6,9,15 | 14,9,12,8 | 18,10,14,6 | 8,7,10,12 |

Analyse the data and calculate the efficiency of the design, clearly indicating how "efficiency" is defined.

3. In a randomised block experiment with 6 treatments and 4 blocks 3 yields ($m_1$, $m_2$, $m_3$) are missing as shown in the yield-table below :

| | | Treatments | | | | | |
|---|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 | 6 |
| Blocks | 1 | 18.5 | 15.7 | 16.2 | 14.1 | 13.0 | 13.6 |
| | 2 | 11.7 | $m_1$ | 12.9 | $m_2$ | 16.9 | 12.5 |
| | 3 | 15.4 | 16.6 | 15.5 | 20.3 | 18.4 | 21.6 |
| | 4 | $m_3$ | 18.6 | 12.7 | 15.7 | 16.5 | 18.0 |

Analyse the data and compare the treatments (6 and 3); (4 and 2); (3 and 1).

4. The table below indicates yields of tea in pounds of each yield-figure being a total for five plots.

| Year | Treatment | | | | | |
|---|---|---|---|---|---|---|
| | $N_0K_0$ | $N_0K_1$ | $N_1K_0$ | $N_1K_1$ | $N_2K_0$ | $N_2K_1$ |
| 1933 | 254 | 251 | 251 | 257 | 259 | 263 |
| 1934 | 466 | 439 | 471 | 485 | 476 | 512 |
| 1935 | 286 | 291 | 326 | 355 | 396 | 381 |

The experiment was designed to test the effect on the yield of tea of sulphate of ammonia ($N$) and muriate of potash ($K$) alone and in combination. The former was applied at the rate, 0,1,2 cwts and the latter at 0,1 cwt per acre per annum. The lay out was on the randomised block principle with 3 blocks (of 6 plots each). The yields for individual plots are not recorded, but it was found that the total S.S. (sum of squares) for the individual readings was 36,406.3; the Block S.S. was 1,630.0; and the Block × Year Interaction S.S. was 909.1.

Analyse and interpret the data.

---

## STATISTICIAN'S DIPLOMA EXAMINATION, 1954.

### Papers VIII & IX : Probit Analysis (Practical)

Time : 4 hours — Full marks : 50

*N.B.* (a) Attempt both questions.

(b) Use of calculating machines is permitted.

1. The following table shows the results of an experiment in which rotenone, a deguelin-concentrate and a 1 : 4 mixture of rotenone and deguelin-concentrate were tested.

$n$ = total number of insects subjected to poison.

$r$ = number affected.

| concentration | $n$ | $r$ | concentration | $n$ | $r$ | concentration | $n$ | $r$ |
|---|---|---|---|---|---|---|---|---|
| Rotenone | | | Deguelin concentrate | | | Mixture | | |
| 10.2 | 50 | 44 | 50.2 | 48 | 48 | 25.1 | 50 | 48 |
| 7.7 | 49 | 42 | 40.8 | 50 | 47 | 20.4 | 46 | 43 |
| 5.1 | 46 | 24 | 30.2 | 49 | 47 | 15.1 | 48 | 38 |
| 3.8 | 48 | 16 | 20.4 | 48 | 34 | 10.0 | 46 | 27 |
| 2.6 | 50 | 6 | 10.0 | 48 | 18 | | | |
| 0 | 49 | 0 | | | | | | |

(a) Estimate log LD 50 and its standard error for rotenone.

(b) Fit the probit regression lines in the three cases.

(c) Test the hypothesis that the three regression lines are parallel.

2. The following table shows the total number of insects ($n$) and number killed ($r$) under different concentrations of Pyrethrin with different weights of spray deposit (mg./sq. cm.); each combination of concentration and deposit being used, on separate batches of insects, both as a direct spray and as a film on which the insects were after wards placed.

| Pyrethrin concentration | Deposit (mg./sq. cm.) | $n$ | $r$ |
|---|---|---|---|
| Exposure to direct spray | | | |
| 0.5 | 0.29 | 27 | 1 |
| 1.0 | 0.29 | 29 | 15 |
| 2.0 | 0.29 | 30 | 27 |
| 4.0 | 0.29 | 28 | 28 |
| 0.5 | 0.57 | 29 | 4 |
| 1.0 | 0.57 | 29 | 19 |
| 2.0 | 0.57 | 27 | 26 |
| 4.0 | 0.57 | 30 | 30 |
| 0.5 | 1.08 | 30 | 6 |
| 1.0 | 1.08 | 24 | 15 |
| 2.0 | 1.08 | 31 | 31 |
| 4.0 | 1.08 | 19 | 19 |

| Pyrethrin concentration | Deposit (mg./sq. cm.) | $n$ | $r$ |
|---|---|---|---|
| Exposure to film | | | |
| 0.5 | 0.29 | 29 | 3 |
| 1.0 | 0.29 | 30 | 10 |
| 2.0 | 0.29 | 29 | 24 |
| 4.0 | 0.29 | 29 | 29 |
| 0.5 | 0.57 | 27 | 4 |
| 1.0 | 0.57 | 28 | 14 |
| 2.0 | 0.57 | 28 | 27 |
| 4.0 | 0.57 | 29 | 29 |
| 0.5 | 1.08 | 28 | 8 |
| 1.0 | 1.08 | 28 | 17 |
| 2.0 | 1.08 | 28 | 26 |
| 4.0 | 1.08 | 17 | 17 |

It is known that 4% of test-subjects would die even without any poison.

(a) Find out the equations of the two probit planes for direct spray and film

(b) Test for parallelism of the two probit-planes.