ANALYSIS OF VARIANCE
AND MISSING OBSERVATIONS
IN COMPLETELY RANDOMIZED, RANDOMIZED BLOCKS
AND LATIN SQUARE DESIGNS

A Thesis

Presented to

The Department of Mathematics

Kansas State Teachers College, Emporia, Kansas

In Partial Fulfillment

of the Requirements for the Degree

Master of Science

---

by

Kiritkumar K. Talati

May 1972

Approved by the Department of Mathematics

John M. Burger

Approved by the Graduate Division

[illegible]

ACKNOWLEDGEMENTS

Sincere thanks and gratitude is expressed to Dr. John Burger for his assistance, patience, and his prompt attention in all directions in preparing this paper. A special note of appreciation goes to Dr. Marion Emerson, Dr. Thomas Davis, Dr. George Poole, Dr. Darrell Wood, who rendered assistance in the research of this paper.

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

1.1. Preliminary Considerations. The objective of this paper is to study the applications of the analysis of variance in experimental design and to develop the formula for missing values in the analysis of variance. According to Sir R.A. Fisher the analysis of variance is a convenient and powerful method of analysis for the research worker in the planning, design, and analysis of research in a variety of disciplines.[1]

The other powerful technique is the analysis of co-variance which is now common in experimental design. Analysis of co-variance is a technique which combines the features of linear regression and the analysis of variance.[2]

Before studying the material of the following chapters, a reader should refresh his background on t, χ^2, and F distributions, testing of hypotheses, regression analysis, and the analysis of variance.

1.2. Purposes and Assumptions of the Analysis of Variance. The main purposes of analysis of variance are:

1. To estimate certain treatment effects which are of interest.

---

[1]A.L. Edwards, Experimental Design in Education and Psychology (New York, Holt, Rinehart and Winston, 1965), pp. 117-18.

[2]G.W. Snedecor and W.G. Cochran, Statistical Methods (Ames, Iowa State University Press, 1967), pp. 419.

2. To determine the accuracy of the estimates such that estimated variances are unbiased.

3. To perform a test of significance by testing the null hypothesis that a treatment difference is zero or has some known value.[3]

In setting up an analysis of variance, one generally recognizes three types of effects, one is the treatments which are of interest, second is the experimental material in which one applies the treatments, and last is the experimental errors which consider the variability during the experiments.

The main assumptions of the analysis of variance are: the treatment and environment effects must be additive, have common variances, and should be normally distributed.[4]

It is important before one applies the analysis of variance that one knows whether the assumptions required for the analysis of variance are satisfied or not.

1.3. Analysis of Co-variance. The analysis of co-variance is another technique which may be used as an alternative or supplement to blocking devices of local control.[5] In considering a problem of 20 replicates and

---

[3]M.G. Kendall and A. Stuart, The Advanced Theory of Statistics (New York, Hafner Publishing Co. 1963-67), vol. 3, pp. 88-118.

[4]W.G. Cochran, "Some Consequences When the Assumptions for the Analysis of Variance Are Not Satisfied," Biometric III (1947), pp. 22-38.

[5]E.F. Lindquist, Design and Analysis of Experiments in Psychology and Education (Boston, Houghton Mifflin Co., 1956), pp. 317-20.

4 treatments, it is quite possible that after arranging blocking, variability may still exist, and such variability is not controlled by the experimental design. A remedy for such problems in an experiment is obtained by using the initial measurement (con-comitant-variable) to determine net value or effect of the treatment, and by using analysis of co-variance.[6] Another example is that of determining which of five teaching methods is "best", in which the criterion is the examination scores obtained by the students. However, before one judges the various teaching methods it might be well to consider the I-Q ratings of the individual students and thus to use the analysis of co-variance. There are some experimental situations, however, in which it is impossible or impracticable to control a con-comitant variable so in that situation it is good to use the analysis of co-variance.

1.4. Definitions.

Definition 1. The subject of "design of experiments" whether in Science, Agriculture, Industry, or sample surveys, tries to lay down the basic principles and design for collecting the data in the most economical and useful form.

Definition 2. Random selection is a method of selecting sample units such that each possible sample has the probability of being drawn; that is, in the population each case has an equal chance of being included in a

[6] D.B. Delury, "The Analysis of Co-variance," Biometrics (September, 1948), pp. 153-57.

particular sample. Ready random tables, which are given in most of the statistics books may be used to select random samples.

1.5. <u>Organization of paper</u>. Chapter II will represent the completely randomized design, and the problem which shows the idea of the homogeneity of variance, and the use of the analysis of variance.

Chapter III represents the randomized block design. The illustrative problem shows how the assumptions of additivity and homogeneity of variance are satisfied. It also illustrates the methods for developing the formulas for missing observations and the analysis of variance.

Finally, Chapter IV is most important because it covers the most important design, the Latin square design. It contains an example illustrating the method of Latin square design. It also illustrates how to develop formulas for missing values, and how to apply the analysis of variance.

Chapter V is the conclusion and summation of the three designs presented in this paper.

CHAPTER II

COMPLETELY RANDOMIZED DESIGN

2.1. Description. The completely randomized design is the basic design, and all other randomized designs stem from it by placing restrictions upon the allocation of the treatments within the experimental area. The completely randomized design is suited only for small numbers of treatments, and for homogeneous experimental material. The main advantage of the completely randomized design is that the analysis of variance remains simple with missing observations. When one can find a homogeneous experimental material and has a small number of treatments, the completely randomized design allows more degrees of freedom and so increases the sensitivity of the experiment.

2.2. Randomization. One may be interested in the effect of medicine on the relief of headaches, say, Excedrin, Aspirin, Bufferin, and Anacin. It is assumed that a group of 20 people are approximately similar in average percentage of headaches. After selecting the 20 people four random samples of five people each would be drawn from the group of 20 people. The ready random tables may be used to select the random samples. Then the treatments are applied randomly to every group. Random numbers could be: 9, 16, 18, 15; 12, 17, 14, 19; 6, 7, 3, 4; 10, 18, 15, 2; and 13, 11, 1, 5.

Similar procedures follow for any number of experimental units and treatments. One should keep in mind that in the completely randomized design, different number of replicates can be used for the different treatments.

2.3. Problem and Computations. For example, the experimenter is interested to see the effect of four medicines on the relief of headaches. The randomization procedure is that given in section 2.2. The artificial data given in Table 2.31, are the percentage of relief of headaches after taking the medicines.

Table 2.31 The Problem of Testing the Medicines on Headache Relief (Percentage of relief of headaches).

Medicines	Excedrin	Aspirin	Bufferin	Anacin
Person 1	30	71	39	50
	40	82	22	23
	53	59	91	69
	61	63	67	91
	65	81	89	49
Total $y_{i.}$	249	356	308	282
$\bar{y}_{i.}$	49.8	71.2	61.6	50.4

The statistical procedures are identical with the one way classification of analysis of variance. Here table 2.31 is calculated in which y_{ij} is the observation (pain release after taking the medicine) of ith medicine of the jth person, where $i = 1, 2, 3, 4$ and $j = 1, 2, \ldots 5$; $\bar{y}_{i.}$ is the treatment mean of the ith medicine (mean of the ith column).

From the analysis of variance of one way classification, the total sum of squares (total S.S.) can be partitioned into treatment sum of squares and error sum of squares (error S.S.);

Total SS = Treatment S.S. + Error S.S.

This follows from the algebraic relation

$$\sum_i \sum_j (y_{ij} - \bar{y}_{..})^2 = \sum_i \sum_j (y_{ij} - \bar{y}_{i.} + \bar{y}_{i.} - \bar{y}_{..})^2$$

$$= \sum_i \sum_j (y_{ij} - \bar{y}_{i.})^2 +$$

$$\sum_i \sum_j (\bar{y}_{i.} - \bar{y}_{..})^2 +$$

$$2 \sum_i \sum_j (y_{ij} - \bar{y}_{i.}) (\bar{y}_{i.} - \bar{y}_{..})$$

$$= \sum_i \sum_j (y_{ij} - \bar{y}_{i.})^2 + k \sum_i (\bar{y}_{i.} - \bar{y}_{..})^2$$

$$+ 2k \sum_i (\bar{y}_{i.} - \bar{y}_{i.}) (\bar{y}_{i.} - \bar{y}_{..})$$

$$= \sum_i \sum_j (y_{ij} - \bar{y}_{i.})^2 + k \sum_i (\bar{y}_{i.} - \bar{y}_{..})^2$$

The sum of squares on the left hand side is called total S.S., and right hand side is the combination of the treatment S.S. and the error S.S.

The computations and statistical analysis for all completely randomized designs are similar to those of this problem. From the table 2.31 the total S.S. and treatment S.S. can be obtained and the error S.S. can be calculated by subtraction, from equation above;

Total S.S. = Treatment S.S. + Error S.S. Hence,

Error S.S. = Total S.S. - Treatment S.S.

The computations of problem 2.31 are:

$$\text{Total S.S.} = (30)^2 + (40)^2 + \ldots + (49)^2 - \frac{(1195)^2}{20}$$

$$= 8797.75$$

$$\text{Treatment S.S.} = \frac{(249)^2}{5} + \frac{(356)^2}{5} + \ldots + \frac{(282)^2}{5} - \frac{(1195)^2}{20}$$

$$= 1223.75$$

$$\text{Error S.S.} = \text{Total S.S.} - \text{Treatment S.S.}$$

$$= 8797.75 - 1223.75$$

$$= 7574.00$$

The analysis of variance table can be prepared similar to the analysis of variance table of one way classification. Table 2.32 is the table of analysis of variance of the present problem.

Table 2.32 ANOVT for Data of Table 2.31

Source of Variation	d.f.	S.S.	M.S.	F
Among medicines	3	1223.75	407.91	.8617
Among people within each group of medicines	16	7574.0	473.375	
Total	19	11,097.75		

Before pooling the individual within lot sums of squares into a sum of squares and obtaining the error S.S.

such as 473.34, one should know about the homogeneity of the individual variances. Bartlett's test may be used to test the homogeneity of variance.

The formula for Bartlett's test is:[7]

$$\chi^2_{(v-1)\text{d.f.}} = \sum_{i}^{v} df_i \log_e s^2 - \sum_{i}^{v} df_i \log_e s_i^2$$

where v is the total number of treatments, and s^2 is from error S.S., while the s_i^2 are the individual variances.

$$s_i^2 = \frac{\sum_j (y_{ij} - \bar{y}_{i.})^2}{(n - 1)}$$

From the table 2.31 the s_i^2 is computed for $i = 1, 2, \ldots, 4$, and the s_i^2's are: 213.7, 107.20, 930.8 and 641.75. Hence,

$$\begin{aligned}\chi^2_{(3\text{ d.f.})} &= 2.3026 \{16 \log 473.375 - 4(\log 213.7 + \log 107.20 + \log 930.8 + \log 641.75)\} \\ &= 2.3026 \{16(2.6749) - 4(2.3298 + 2.0302 + 2.9689 + 2.8074)\} \\ &= 5.9867.\end{aligned}$$

The value of $\chi^2_{(.95,\ 3df)}$ is 7.81.[8]

Hence the calculated value of $\chi^2_{(3df)}$ is less than the tabulated $\chi^2_{(3df)}$ at 5% significance level and so the variances are not different.

---

[7] Snedecor and Cochran, op. cit., pp. 296-98.

[8] W.J. Dixon and F.J. Massey, Introduction to Statistical Analysis (New York, McGraw-Hill, 1967), p. 465.

As individual s_i^2's are considered to be estimates of the same population variance σ^2, the generalized error S.S. equal to 473.375 should be used to compare treatment means $\overline{y}_{i.}$.

Snedecor's F test can be used to test the hypothesis: all means are equal.[9]

$$F_{(3,16\ df)} = \frac{407.91}{473.375} = .8617$$

The value .8617 is less than the tabulated value of $(F_{.01,(3,16)\,df})$ which is 5.29 and so it is highly non-significant. Hence one should accept the hypothesis that all medicines are equally effective on headaches.[10]

2.4. Conclusion and Further Applications. In the above problem the analysis of variance shows that all four medicines would give approximately the same relief on the headaches. This is not the specific conclusion about the medicines (four different tablets), on the contrary this conclusion was made from the assumed data.

The best advantage of completely randomized design is that the analysis of variance remains the same even if some observations are missing.

In Analysis of Variance for unequal replication of treatments, the computations are slightly different. Below is the symbolical table 2.33 of the analysis of variance

---

[9] Ibid., pp. 150-84.

[10] Ibid.

for a completely randomized design with unequal number of observations per treatment.

Table 2.33 The Analysis of Variance with Unequal Number of Replications in Completely Randomized Design.

Source of Variation	d.f.	S.S.
Among treatments	$s-1$	$\sum_{t_i} y_{i.}^2 - \frac{y^2_{..}}{\Sigma t_i}$
Within treatments	$\sum_{i}^{s} t_i - s$	$\sum_{i}^{s} \left\{ \sum_{j}^{t_i} y^2_{ij} - \frac{y^2_{i.}}{t_i} \right\}$
Total	$\sum t_i - 1$	$\sum_{i}^{s} \sum_{j}^{t_i} y^2_{ij} - \frac{y^2_{..}}{st_i}$

For the other situation, where an experimental unit may be subsampled or several readings per experimental unit have been made, the Analysis of Variance is slightly different, and is discussed by Snedecor[11], Cochran and Cox[12], and many authors.

[11] G.W. Snedecor and W.G. Cochran, Statistical Methods (Ames, Iowa State College Press, 1956) pp. 14-127.

[12] Cochran and Cox, op. cit., pp. 73-345.

CHAPTER III

RANDOMIZED BLOCK DESIGN

3.1. Description. The randomized block design has wider applications in agricultural field experiments. The randomized block design is useful whenever it is thought that a small set of trials under restricted conditions, believed to be uniform, will within the small set give better indications of differences due to treatments, than the completely randomized design because of great non-uniformity of uncontrolled or uncontrollable conditions.

If the experimental material is not homogeneous (i.e. heterogeneous), then due to more variability in the experimental material one cannot apply the completely randomized design. It may be possible to group the material into homogeneous subgroups. If the treatments are applied to the relatively homogeneous material in each homogeneous subgroup, the design is randomized block design. In completely randomized design no stratification of experimental material is made. The treatments are randomly allotted to the experimental units. In randomized block design the treatments are randomly allotted within each stratum, i.e. randomization is restricted. Therefore, if it is desired to control one source of variation by stratification the experimenter should select a randomized block design rather than a completely randomized design.

3.2. Randomization. In an experiment with 25 replicates to which one wishes to apply five treatments, if the replicates can be divided into five subgroups of five replicates each such that each subgroup is homogeneous, then one would use randomized block design. If one has the five blocks,

Block I	10, 25, 1, 24, 23
Block II	5, 7, 9, 13, 12
Block III	16, 18, 22, 2, 5
Block IV	3, 6, 8, 15, 19
Block V	4, 11, 14, 17, 20

then in each block one should randomly select one replicate for each of the treatments. The ready random tables may be used to select the replicates randomly for every block.

3.3. Problem and Statistical Analysis. A problem has 25 replicates which are to be used in an experiment with 5 replicates assigned to each of five treatments. For example, the firm in a particular town wants some students of mathematics in a field of research work. It is common that before they select the students from the university, the manager of the firm will want to perform an experiment which tests the general intellectual capacity of the students in the university. So the manager selects 25 students randomly from the mathematics department, and he prepares five different tests to measure their intellectual capacities. In table 3.31 the score of each student is recorded, and the analysis of variance is performed on the scores. The students are grouped according to their previous rank, so there are five groups of people (or if one prefers blocks of five different tests) to be applied. The computations are similar to the analysis of variance with two way classifications.

From Table 3.31 the total S.S. and the treatment S.S. can be obtained similar to those for the completely randomized design.

$$\text{Total S.S.} = (25)^2 + (10)^2 + \ldots + (13)^2 - \frac{(332)^2}{25}$$

$$= 575.04$$

$$\text{Treatment S.S.} = \sum_j y^2_{\cdot j} - \frac{\left(\sum \sum y_{ij}\right)^2}{25}$$

$$= (60)^2 + \ldots + (86.0)^2 - \frac{(332)^2}{25}$$

$$== 161.06$$

$$\text{Block S.S.} = \sum_{i=1}^{5} \frac{y^2_{i\cdot}}{5} - \frac{\left(\sum_i \sum_j y_{ij}\right)^2}{25}$$

$$= (66)^2 + \ldots + (63)^2 - \frac{(332)^2}{25}$$

$$= 23.66$$

Error S.S. can be obtained by subtraction.

Error S.S. = Total S.S. - Treatment S.S. - Block S.S.

$$= 575.04 - 161.06 - 23.66$$

$$= 390.32$$

Table 3.31 The Problem of the Randomized Block Design (Scores obtained out of 25 maximum possible scores)

	Test I	Test II	Test III	Test IV	Test V	$y_{i.}$	$\bar{y}_{i.}$
Block I	15	12	20	3	16	66	13.20
Block II	10	1	13	19	17	60	12.00
Block III	10	15	15	15	19	74	14.08
Block IV	12	16	12	8	21	69	13.8
Block V	13	17	13	7	13	63	12.6
$y_{.j}$	60	61	73	52	86		
$\bar{y}_{.j}$	12	12.2	14.6	10.4	17.2		

Table 3.32 is the table of analysis of variance for Table 3.31.

Table 3.32 ANOVT Table for Table 3.31

Source of Variation	d.f.	S.S.	M.S.
Treatment	4	161.06	35.25
Blocks	4	23.66	5.86
Error	16	390.32	24.40
Total	24		

Before one can use the t or F test, it is necessary to test the additivity and homogeneity of variance. If one has the additivity then the treatment effect will add a constant to the basic or control yield of each replicate. In the randomized block design with additivity, generally error variance will be constant for all normalized treatment comparisons. It seems that the additivity is much more important than homogeneity of error, because non-additivity will commonly produce heterogeneity of error and without additivity the estimates of treatment effects and differences is obscure.[13]

Tukey's non-additivity test can be used to see whether treatment effects are additive or not.

---

[13] O. Kempthorne and W.G. Barclay, "The partion of error in randomized blocks," American Statistical Association Journal, (September, 1953), pp. 610-15.

According to Tukey's test, the sum of squares for non-additivity is given by[14]

$$\text{Non-additivity} = \frac{\left[\sum_j x_{ij}\,(x_{i\cdot} - \bar{x}_{\cdot\cdot})\,(\bar{x}_{\cdot j} - x_{\cdot\cdot})\right]^2}{\sum_i (x_{i\cdot} - x_{\cdot\cdot})^2 \;\sum_j (x_{\cdot j} - x_{\cdot\cdot})^2}$$

With 10 degrees of freedom for this example, the computations are given in Table 3.33.

$$\text{Non-additivity S.S.} = \frac{(23.5168)^2}{(28.12)(4.69)}$$

$$= 4.18$$

To test the significance, the remainder is obtained by subtraction.

Remainder = Error S.S. - Non-additivity S.S.

$$= 390.32 - 4.18$$

$$= 386.14$$

The mean square for non-additivity may be tested for significance by dividing it by the mean square for remainder.

Thus $F = \frac{\text{non-additivity}}{\text{remainder}}$ with (1,15) d.f.

$$= \frac{4.18}{24.40}$$

$$= .1713$$

Table 3.34 is the ANOVT for the non-additivity test for the example.

---

[14]Snedecor and Cochran, op. cit., pp. 330-33.

Table 3.33 Calculations for the Test of Non-additivity for Table 3.31

	Test I	Test II	Test III	Test IV	Test V	$\overline{y}_{i\cdot} - \overline{y}_{\cdot\cdot}$
Block I	15	12	20	3	11	- .08
Block II	10	1	13	19	17	-1.28
Block III	10	15	15	15	19	1.52
Block IV	12	16	12	8	21	.52
Block V	13	17	13	7	13	- .68
$y_{\cdot j}$	60	61	73	52	86	
$\overline{y}_{\cdot j}$	12	12.2	14.6	10.4	17.2	
$\overline{y}_{\cdot j} - \overline{y}_{\cdot\cdot}$	-1.28	-1.08	1.32	-2.58	3.92	

Table 3.34 ANOVT for Non-additivity

Source of Variation	d.f.	S.S.	M.S.
Non-additivity	1	4.18	4.18
Remainder	15	386.14	24.13
Error S.S.	16	390.32	

For the present example, the calculation value of $F_{(1,15)}$ is less than the tabulated value of $F_{(1,15)}$, which is equal to 4.54 at 5% significance level. So the conclusion is made that non-additivity is not effected in the experiment or there is additivity of treatments in the experiment.

One may be interested to see if homogeneity of variance is also present in the experiment or not. The same Bartlet's Test as used in section 2.3 can be performed.

In the present example the calculated value of $\chi^2_{(4df)}$ should be found less than the tabulated values at $\chi^2_{(.95,\ 4df)}$, which is non-significant, and so the variances have the homogeneity of variance.

Now one should apply the F test. The value of F from 3.32 is

$$F_{(4,16)} = \frac{\text{treatment M.S.}}{\text{error M.S.}}$$

$$= \frac{35.25}{24.40}$$

$$= 1.44$$

Here the calculated value of $F_{(4,16)}$ is less than the tabulated value of $F_{(4,16)}$ at 5% significance level (3.26), so one may accept the hypothesis that all 25 students have the same I.Q. or intellectual capacity. The manager may conclude that all 25 students are equally capable of work in five different branches of mathematics.

3.4. Efficiency of Randomized Block Design as compared to Completely Randomized Design. Since the block mean square is less than the error mean square, the present design is less efficient than a completely randomized design would have been. The efficiency of randomized block design as compared to completely randomized design can be computed by using the formula given by Federer.[15]

Federer's formula is,

$$E' = \frac{[(v-1) + (r-1)(v-1)]E + [r-1]B}{[(v-1) + (r-1)(v-1) + (r-1)]} \qquad (1)$$

Here v = total number of treatments and r = total number of replicates or blocks. E and B are the error mean sum of squares and block mean sum of squares respectively.

The efficiency of the randomized block design to what it would be had a completely randomized design been used is the ratio

$$\frac{(rv-r-v+2)(rv-v+3)\ E'}{(rv-r-v+4)(rv-v+1)\ E} \qquad (2)$$

where E' is defined in equation (1). In this example,

---

[15]W.T. Federer, Experimental Design (New York, The Macmillan Company, 1955) pp. 116-17.

$$E' = \frac{6 + 16\ (24.40) + 4(5.86)}{24}$$

$$= 23.3433$$

and the ratio equal to

$$\frac{(25 - 5 - 5 + 2)\ (25 - 5 + 3)\ 23.3433}{(25 - 5 - 5 + 4)\ (25 - 5 + 1)\ 24.40}$$

$$= \frac{(17)\ (23)\ (23.3433)}{(19)\ (21)\ (24.40)}$$

$$= .9373$$

The relative efficiency has been estimated to be .9373 or 93.73%.

3.5. Missing Observations in Randomized Block Design. To illustrate the missing observations effect on analysis of variance, and how to minimize the error S.S., one may take the general form of randomized block design given in table 3.51.

Table 3.51 Missing Values in Randomized Block Design

Block / Treatment	1	2	. . .	n	Total
1	$\hat{y}_{11}$	y_{12}	. . .	y_{1n}	$y_{1.} + \hat{y}_{11}$
2	y_{21}	y_{22}	. . .	y_{2n}	$y_{2.}$
.	.	.	. . .	.	.
.	.	.	. . .	.	.
.	.	.	. . .	.	.
m	y_{m1}	y_{m2}	. . .	y_{mn}	$y_{m.}$
Total	$y_{.1} + \hat{y}_{11}$	$y_{.2}$	. . .	$y_{.n}$	$y_{..} + \hat{y}_{11}$

If one supposes in the example that the y_{11} observation is missing then there is no loss of generality in placing y_{11} in the first row and first column.[16]

If the analysis of variance on the above value is calculated, the error S.S. is

$$\text{Error S.S.} = \hat{y}^2_{11} + \sum\sum y^2_{ij} - \frac{\left[(y_{i\cdot} + \hat{y}_{11})^2 + \sum_{i=2}^{m} y^2_{i\cdot}\right]}{n}$$

$$- \frac{\left[(y_{\cdot 1} + \hat{y}_{11})^2 + \sum_{j=2}^{n} y^2_{\cdot j}\right]}{m} + \frac{(y_{\cdot\cdot} + \hat{y}_{11})^2}{nm} \qquad (3)$$

The best estimate value of the missing observation can be found by minimizing the error S.S., i.e. by taking a first derivative with respect to the missing observation and equating to zero. Hence,

$$\frac{\partial \text{Error S.S.}}{\partial y_{ii}} = 2\hat{y}_{11} - \frac{2(y_{1\cdot} + \hat{y}_{11})}{n} - \frac{2(y_{\cdot 1} + \hat{y}_{11})}{m}$$

$$+ \frac{2(y_{\cdot\cdot} + \hat{y}_{11})}{mn} = 0 \qquad (4)$$

by solving equation (4) the value of $\hat{y}_{11}$ is

$$\hat{y}_{11} = \frac{my_{1\cdot} + ny_{\cdot 1} - y_{\cdot\cdot}}{(n-1)(m-1)}$$

If more than one observation is missing a similar procedure proposed by Yates[17] can be used.

For example, if in table 3.51 the two values, y_{22} and y_{44} are missing, then the computed error S.S. is

---

[16] Ibid, pp. 133-34.

[17] F. Yates, "The analysis of replicated experiments when the field results are incomplete," Empire Journal of Experimental Agriculture, I (1933), 129-42.

$$\text{Error S.S.} = \sum_i \sum_j y_{ij} + \hat{y}^2_{22} + \hat{y}^2_{44} - \frac{\left[(y_{2\cdot} + \hat{y}_{22})^2 + (y_{4\cdot} + \hat{y}_{44})^2 + \sum y^2_{i\cdot}\right]}{n} - \frac{\left[(y_{\cdot 2} + \hat{y}_{22})^2 + (y_{\cdot 4} + \hat{y}_{44})^2 + \sum_{j \neq 2,4} y^2_{\cdot j}\right]}{m} + \frac{(y_{\cdot\cdot} + \hat{y}_{22} + \hat{y}_{44})^2}{nm}$$

A first derivative with respect to $\hat{y}_{22}$ is

$$\frac{\partial \text{ Error S.S.}}{\partial \hat{y}_{22}} = 2\hat{y}_{22} - \frac{2(y_{2\cdot} + \hat{y}_{22})}{n} - \frac{2(y_{\cdot 2} + \hat{y}_{22})}{m} + \frac{2(y_{\cdot\cdot} + \hat{y}_{22} + \hat{y}_{44})}{nm} = 0 \qquad (5)$$

By solving equation (5) the value of $\hat{y}_{22}$ is

$$\hat{y}_{22} = \frac{my_{2\cdot} + ny_{\cdot 2} - y_{\cdot\cdot} - \hat{y}_{44}}{(nm - n - m + 1)} \qquad (6)$$

Similarly by taking the derivative with respect to $\hat{y}_{44}$ the equation (7) can be obtained.

$$\hat{y}_{44} = \frac{my_{4\cdot} + ny_{\cdot 4} - y_{\cdot\cdot} - \hat{y}_{22}}{(nm - n - m + 1)} \qquad (7)$$

By solving equations 6 and 7 simultaneously the value of $\hat{y}_{44}$ and $\hat{y}_{22}$ can be obtained as given in equations (8) and (9) which are:

$$\hat{y}_{22} = \frac{(n-1)(m-1)(ny_{\cdot 2} + my_{2\cdot}) - my_{4\cdot} - ny_{\cdot 4} - (nm-n-m)y_{\cdot\cdot}}{(nm - m - n)\ (nm - n - m + 2)} \qquad (8)$$

$$\hat{y}_{44} = \frac{(n-1)(m-1)(ny_{\cdot 4} + my_{4\cdot}) - my_{2\cdot} - ny_{\cdot 2} - (nm-n-m)y_{\cdot\cdot}}{(nm - n - m)\ (nm - n - m + 2)} \qquad (9)$$

By using equations (8) and (9) the missing value from the table 3.31 can be calculated.

$$\hat{y}_{22} = \frac{16(59 \times 5 + 5 \times 60) - (5 \times 61) - (5 \times 44) - 55 \times 323)}{255}$$

$$= \frac{9520 - 5370}{255}$$

$$= 16.278$$

Similarly y_{44} can be obtained.

$$\hat{y}_{44} = \frac{16(5 \times 61 \times 5 \times 44) - (5 \times 60) - (5 \times 59) - (15 \times 323)}{255}$$

$$= 11.6078$$

So the $\hat{y}_{22}$ and $\hat{y}_{44}$ can be replaced by the values 16.2 and 11.6 respectively.

Computations can be made after replacing the missing values by 16.2 and 11.6. The analysis of variance should be done similarly to the randomized block design, except the degrees of freedom for total and error each being reduced to two less than in the original design.

If one lost the data, one should also lose the degree of freedom and indirectly the sensitivity of the experiment.

3.6. Further Applications. It is not necessary that one must have equal number of observations per treatment. If in experiments where the observations are more than one, or readings more than one the analysis of variance differs, and would be appropriate like table 3.61.

Table 3.61. Analysis for More-than-one Observation per Experimental Unit.

Source of Variation	d.f.	S.S.
Blocks	s - 1	$\sum_{j}^{s} \frac{y^2_{.j.}}{tu} - \frac{y^2_{...}}{stu}$
Treatments	t - 1	$\sum_{i}^{t} \frac{y^2_{i..}}{su} - \frac{y^2_{...}}{stu}$
Error	(s-1) (t-1)	$\sum_{i} \sum_{j} \frac{y^2_{ij.}}{u} - \sum_{i} \frac{y^2_{i..}}{su} - \sum \frac{y^2_{.j.}}{tu} + \frac{y^2_{...}}{stu}$
Sampling Error	st(u-1)	$\sum_{i} \sum_{j} \sum_{k}^{u} \left(y^2_{ijk} - \frac{y_{ij.}}{u}\right)$
Total	stu - 1	$\sum_{i} \sum_{j} \sum_{k} y^2_{ijk} - \frac{y^2_{...}}{stu}$

where s is equal to the total number of blocks, t equals the total number of treatments and u equals the total number of observations per experimental unit.

The use of sampling error is dependent upon the hypothesis tested and upon the assumptions made about the data. More detailed description and an example is given by Federer.[18]

[18]Federer, op. cit., pp. 120-25.

The method of analysis of variance when the results consist of data classified into two classes, has been given by Cochran and Cox.[19]

---

[19]Cochran and Cox, *op. cit.*, pp. 115-17.

CHAPTER IV

LATIN SQUARE DESIGN

4.1. Description. The Latin square design has applications in industry, biological sciences, agriculture, medicine and a variety of other experimentations. By two way stratification the Latin square design controls more of the variation than the completely randomized design, and the randomized block design. The experiment in which variability cannot be controlled by randomized block design, like the agriculture experiment in which the gradient of fertility may run from either side or in different direction the most useful design for such an experiment is the Latin square design.

4.2. Construction and Model of Latin Square. To construct the Latin square for Latin square design one must have the number of rows equal to number of columns equal to number of treatments. To construct the Latin square the first step is to write down a systematic arrangement of letters, then arrange rows and columns at random, and assign treatments at random to the letters. There are many methods developed for randomization of Latin squares; Fisher and Yates[20] has given a detailed description of how to construct a Latin square.

---

[20]R.A. Fisher and F. Yates, Statistical Tables for Biological, Agriculture, and Medicine Research (New York, Hafner, 1948), pp. 47-52, 137-43.

For a 5 x 5 Latin square one may select the Latin square according to Fisher and Yates.[21] In agriculture one may be interested to see the effects of five fertilizers on the production of wheat. As one does not know the gradient of fertility of land, the best design is the Latin square. It may be constructed like table 4.21.

Table 4.21 5 x 5 Latin Square

B	E	D	C	A
C	A	B	E	D
D	B	C	A	E
E	C	A	D	B
A	D	E	B	C

In table 4.21 it may be noticed that a treatment occurs once in each row and column.

The model of Latin square is:

$$y_{ijk} = \mu + r_i + \theta_j + \tau_k + \varepsilon_{ijk}$$

where i, j, k = 1, 2, ... , n, is the general effect for all observations, y_{ijk} is the observed value of the ith row, jth column and kth treatment. r_i, θ_j, and τ_k are the row effect of the ith observation, column effect of jth observation, and treatment effect of kth treatment respectively. It is assumed that all assumptions of analysis of variance are satisfied.

4.3. <u>Problem and Statistical Analysis</u>. The problem of an agriculture experiment in which one is interested

[21]Ibid.

actually in finding the effects of the five fertilizers by looking at the effects on wheat production may be used as an illustration.

It is a fact that sometimes the gradient in fertility may run in different directions in an agricultural field. One should keep this in mind, and if he does not know whether the agricultural field is homogeneous or not, he may use a Latin square design. One may construct the Latin square according to section 4.2. With five fertilizers, A, B, C, D, and E, it may be like this:

B	E	D	C	A
C	A	B	E	D
D	B	C	A	E
E	C	A	D	B
A	D	E	B	C

In the experiment described, the observations corresponding to all entries are assumed to be those shown in the table 4.31.

Table 4.31 The Experiment of Five Fertilizers (Quantities in Tons)

	Col. 1	Col. 2	Col. 3	Col. 4	Col. 5
Row 1	B (7)	E (17)	D (26)	C (7)	A (18)
Row 2	C (27)	A (27)	B (5)	E (9)	D (13)
Row 3	D (5)	B (15)	C (16)	A (39)	E (10)
Row 4	E (30)	C (14)	A (13)	D (22)	B (9)
Row 5	A (9)	D (10)	E (15)	B (24)	C (8)

The computations are similar to that of randomized block design, to compute row S.S. and column S.S., while treatment S.S. is slightly different. Error S.S. can be obtained by subtraction.

Error S.S. = Total S.S. - Row S.S. - Column S.S. - Treatment S.S.

As one assumed that the assumptions of analysis of variance are satisfied, it may be of interest to check that the non-additivity is affected or not.

The presence of correlation between the variances and means of the treatments is one indication of departure from normality and this is likely associated with hetrogeneity of variance.[22]

One should keep this in mind before he starts his computations and then apply the analysis of variance. In the present table 4.31 the means and standard deviations are computed and are equal to:

Means: 21.2, 12.00, 14.04, 15.2, 16.2

S.D.'s: 10.74, 6.87, 7.17, 7.73, 7.52

which seems, approximately, the means are proportional to the standard deviations. There are mathematical reasons why this type of relation between standard deviations and mean is likely to be found when the treatment effects are proportional rather than additive. For such a situation

---

[22]M.S. Bartlett, "The Use of Transformations," Biometrics, (3, 1947), pp. 39-52.

the logarithm transformation can be used to make the effect additive rather than proportional.[23]

Table 4.32 is the "transformed to logs" of table 4.31.

Table 4.32 The Transformation to Logs of Table 4.31

Rows \ Columns	1	2	3	4	5	$y_{i..}$
1	.85	1.23	1.42	.85	1.25	5.60
2	1.43	1.43	.70	.95	1.11	5.62
3	.70	1.18	1.20	1.51	1.00	5.70
4	1.47	1.15	1.11	1.37	.95	6.10
5	.95	1.00	1.18	1.38	.90	5.41
$y_{.j.}$	5.40	5.99	5.61	6.14	5.21	$y_{...}$ 28.35
$y_{..k}$	6.33	5.06	5.53	5.60	5.83	

According to Tukey's equation the non-additivity test is performed. The Table 4.33 is computed according to Snedecor and Cochran; and the test is applied to see that the non-additivity is effected on experiment or not.[24]

The computation can be made according to table 4.33. In that table $\hat{y}_{ijk}$ is calculated by, $\hat{y}_{ijk} = \overline{y}_{i..} + \overline{y}_{.j.} + \overline{y}_{..k} - 2\overline{y}_{...}$. The residual $d_{ijk} = y_{ijk} - \hat{y}_{ijk}$ as shown,

---

[23] Ibid.

[24] Snedecor and Cochran, op. cit., pp. 330-37.

adjusting such that the sums are zero over row, column and treatment. In the present problem the sums are zero over row and column while A, C, and E are slightly different which can be adjusted by trial and error methods. Adjusting 3 values by .03 would effect the total terms of non-additivity little, but if the effect is considerably high the adjustment is necessary. For the 25 values of V_{ijk} equal to $m_1(\hat{y}_{ijk} - m_2)^2$ where m_1 and m_2 are any two convenient constants, one may take $m_2 = y_{\cdots} = 1.33$ and m_1 equal 1000 so that the values are between 1 and 100.

Now the value of N is calculated by the equation (A)

$$N = \sum d_{ijk} V_{ijk} \tag{A}$$

$$= (-.09)(37) + \ldots + (-.07)(26)$$

$$= -2.7$$

Now non-additivity is tested by: $\frac{N^2}{D}$

where D is the error sum of squares of the V_{ijk} which in the present problem is equal to 3041.67. Hence,

$$\frac{N^2}{D} = \frac{(-2.7)^2}{3041.67}$$

$$= .0023$$

To perform the non-additivity test the remainder terms can be obtained by subtracting error S.S. of transformed data and non-additivity. Hence:

Remainder = Error S.S. - Non-additivity

$$= 1.03 - .0023$$

$$= 1.0277$$

The error S.S. is found to be 1.03 as is shown in the next few pages.

Table 4.33 Test of Additivity in a Latin Square Design

Row \ Column		1	2	3	4	5	$\bar{y}_{i\cdot\cdot}$
1	y_{ijk}	.85	1.23	1.42	.85	1.25	1.12
	$\hat{y}_{ijk}$	.94	1.21	1.09	1.19	1.17	
	d_{ijk}	-.09	.02	.33	-.34	-.08	
	V_{ijk}	(37)	(6)	(2)	(3)	(1)	
2		1.43	1.43	.70	.95	1.11	1.12
		1.05	1.33	.99	1.24	1.01	
		.38	.10	-.29	-.29	.10	
		(7)	(39)	(20)	(11)	(15)	
3		.70	1.18	1.20	1.59	1.00	1.13
		1.06	1.08	1.10	1.36	1.07	
		-.36	.10	.10	.23	-.07	
		(5)	(3)	(1)	(52)	(4)	
4		1.47	1.15	1.11	1.37	.95	1.20
		1.18	1.25	1.33	1.27	.99	
		.29	-.10	-.22	.07	-.04	
		(2)	(14)	(39)	(19)	(20)	
5		.95	1.00	1.18	1.38	.90	1.08
		1.17	1.12	1.10	1.05	.97	
		-.22	-.12	.08	.33	-.07	
		(1.3)	(0)	(1)	(7)	(26)	
$\bar{y}_{\cdot j\cdot}$		1.08	1.20	1.12	1.22	1.04	$\bar{y}_{\cdot\cdot\cdot}$ 1.33
$\bar{y}_{\cdot\cdot k}$		1.27	1.01	1.11	1.11	1.16	
		(A)	(B)	(C)	(D)	(E)	

The analysis of variance is performed which is given in Table 4.34.

Table 4.34 The ANOVT Table for Non-additivity Test

Source of Variation	d.f.	S.S.	M.S.
Error S.S.	12	1.03	
Non-additivity	1	.0023	.0023
Remainder	11	1.0277	.0938

From the above table the value of $F_{(1,11)}$ is equal to .0245; which is less than the tabulated value of $F_{(1,11)}$ equal to 4.84 at 5% significance level. The non-additivity should not affect the experiments.

The analysis of variance should apply to the transformed data, i.e. all computations of Latin square design is made from the Table 4.31.

$$\text{Total S.S.} = (.85)^2 + (1.43)^2 + \ldots + (.90)^2 - \frac{(28.35)^2}{25}$$

$$= 33.61 - 32.15$$

$$= 1.44$$

$$\text{Column S.S.} = \frac{(5.40)^2}{5} + \ldots \frac{(5.21)^2}{5} - \frac{(28.35)^2}{25}$$

$$= .13$$

$$\text{Row S.S.} = \frac{(5.60)^2}{5} + \ldots + \frac{(5.41)^2}{5} - \frac{(28.35)^2}{25}$$

$$= .14$$

$$\text{Treatment S.S.} = \frac{(6.33)^2}{5} + \frac{(5.06)^2}{5} + \frac{(5.53)^2}{5} + \frac{(5.60)^2}{5} + \frac{(5.83)^2}{5} - \frac{(28.35)^2}{25}$$

$$= .14$$

Hence, error S.S. is equal to: $1.44 - .13 - .14 - .14$

$$= 1.03$$

The analysis of variance is given in table 4.35.

Table 4.35 The ANOVT Table for the data given in Table 4.31.

Source of Variation	d.f.	S.S.	M.S.
Row	4	.14	.035
Column	4	.13	.033
Treatment	4	.14	.035
Error	12	1.03	.086
Total	24	1.44	

The value of $F_{(4,12)}$ from table 4.35 is .41 and is less than the $F_{(4,12)}$ at 5% significance level and so the conclusion is that all fertilizers are equally good in quality in wheat production.

4.4. <u>Efficiency of Latin Square Design</u>. Since the row S.S. and column S.S. are smaller than the error S.S. the Latin square design is less efficient than the completely randomized design or the randomized block design. The efficiency of Latin square design as compared to completely randomized design is:[25]

$$\left(\frac{(k-1)(k-2)+1}{(k-1)(k-2)+3}\right)\left(\frac{k(k-1)+3}{k(k-1)+1}\right)\left(\frac{R+C+(k-1)E}{(k+1)E}\right)$$

where k is equal to the total number of treatments, R, C, and E are the total mean sums of squares of row, column and error respectively. In this problem it is:

$$\left(\frac{(5-1)(5-2)+1}{(5-1)(5-2)+3}\right)\left(\frac{5(5-1)+3}{5(5-1)+1}\right)\left(\frac{.035+.033+(5-1).086}{(6 \times .086)}\right)$$

$$= \left(\frac{13}{15}\right)\left(\frac{23}{21}\right)\left(\frac{.035+.033+(5-1).086}{(6 \times .086)}\right)$$

= .7578 or 75.78%

As row S.S. is higher than column S.S. the efficiency of Latin square design as compared to randomized block design (row as block) is:

$$\left(\frac{(k-1)(k-2)+1}{(k-1)(k-2)+3}\right)\left(\frac{(k-1)^2+3}{(k-1)^2+1}\right)\left(\frac{R+(k-1)E}{kE}\right)$$

which for this problem is:

$$\left(\frac{(5-1)(5-2)+1}{(5-1)(5-2)+3}\right)\left(\frac{(5-1)^2+3}{(5-1)^2+1}\right)\left(\frac{.035+(5-1).086}{5 \times .086}\right)$$

$$= \left(\frac{13}{15}\right)\left(\frac{19}{17}\right)\left(\frac{.035+4(.086)}{5(.086)}\right)$$

= .86 or 86.1

Approximately 4 replicates of the completely randomized design or the randomized block design is equivalent to the 5 replicates of Latin square design.

---

[25]Federer, op. cit., p. 163.

4.5. Development of Formula for Missing Experimental Units. Accidents often result in the damage of an experiment, by loss of data. Animals may die, crops may be destroyed or one may miss the reading. The missing data destroy the sensitivity and simplicity of the analysis. Fortunately, missing observations can be determined by least square methods and replaced.

The 5 x 5 Latin square design as given in table 4.51 may be considered. If one supposes that the observation is missing in the first row, in the first column, and for treatment one, there is no loss of generality if it is replaced by $\hat{y}_{111}$.

Table 4.51 Latin Square Design (5 x 5) to Develop the Missing Values

	Col. 1	Col. 2	Col. 3	Col. 4	Col. 5	Total
Row 1	(A) $\hat{y}_{111}$	(E) y_{125}	(D) y_{134}	(C) y_{143}	(B) y_{152}	$y_{1\cdot\cdot} + \hat{y}_{111}$
Row 2	(C) y_{213}	(B) y_{222}	(A) y_{231}	(E) y_{245}	(D) y_{254}	$y_{2\cdot\cdot}$
Row 3	(D) y_{314}	(A) y_{321}	(C) y_{333}	(B) y_{342}	(E) y_{355}	$y_{3\cdot\cdot}$
Row 4	(E) y_{415}	(C) y_{423}	(B) y_{432}	(D) y_{444}	(A) y_{451}	$y_{4\cdot\cdot}$
Row 5	(B) y_{512}	(D) y_{524}	(E) y_{535}	(A) y_{541}	(C) y_{553}	$y_{5\cdot\cdot}$
Total	$y_{\cdot 1\cdot} + \hat{y}_{111}$	$y_{\cdot 2\cdot}$	$y_{\cdot 3\cdot}$	$y_{\cdot 4\cdot}$	$y_{\cdot 5\cdot}$	$y_{\cdot\cdot\cdot}$

If y_{111} is missing, then residual S.S. is equal to:

$$R = \hat{y}^2_{111} + {}_{ijk}\, y^2_{ijk} - \frac{1}{5}\left\{(y_{1..} + \hat{y}_{111})^2 + \sum_{2}^{5} y^2_{i..} + (y_{.1.} + \hat{y}_{111})^2 + \sum_{2}^{5} y^2_{.j.} + (y_{..1} + \hat{y}_{111})^2 + \sum_{2}^{5} y^2_{..k}\right\} + \frac{2(y_{...} + \hat{y}_{111})^2}{(5)^2} \quad (10)$$

Now, according to the least square method, one can take the partial derivative of r with respect to y_{111} and equate to zero to obtain the value of y_{111}. From equation (10)

$$\frac{\partial R}{\partial(\hat{y}_{111})} = 2\hat{y}_{111} - \frac{2}{5}\left\{y_{1..} + \hat{y}_{111} + y_{.1.} + \hat{y}_{111} + y_{..1} + \hat{y}_{111}\right\} + \frac{4(y_{...} + \hat{y}_{111})}{(5)^2} = 0 \quad (11)$$

Hence by solving equation (11)

$$\hat{y}_{111} = \frac{5(y_{1..} + y_{.1.} + y_{..1}) - 2y_{...}}{(5-1)\ (5-2)}$$

One may generalize the formula for any missing value as: $\hat{y}_{ijk} = \dfrac{k(y_{i..} + y_{.j.} + y_{..k}) - 2y_{...}}{(k-1)\ (k-2)}$

where k is the total number of treatments.

If more than one value is missing the same method of least square has been used. If y_{333} and y_{221} are missing in table 4.51, the error sum of squares is equal to:

$$\text{Error S.S.} = \hat{y}^2_{333} + \hat{y}^2_{221} + y^2_{ijk} - \frac{1}{5}\left\{(y_{3..} + \hat{y}_{333})^2 + (y_{22.} + \hat{y}_{221})^2 + y^2_{1..} + y^2_{4..} + y^2_{5..} + (y_{.3.} + \hat{y}_{333})^2 +\right.$$

$$(y_{\cdot 2 \cdot} + \hat{y}_{221})^2 + y^2_{\cdot 1 \cdot} + y^2_{\cdot 4 \cdot} + y^2_{\cdot 5 \cdot} + (y_{\cdot\cdot 3} + \hat{y}_{333})^2 + (y_{\cdot\cdot 1} + \hat{y}_{221})^2 + y^2_{\cdot\cdot 2} + y^2_{\cdot\cdot 4} + y^2_{\cdot\cdot 5}\Big\} + \frac{2}{(5)^2}(y_{\cdot\cdot\cdot} + \hat{y}_{333} + \hat{y}_{221})^2$$

By taking partial derivative with respect to $\hat{y}_{333}$ and $\hat{y}_{221}$ the equations (13) and (14) would be obtained.

$$\frac{\partial \text{ Error S.S.}}{\partial(\hat{y}_{333})} = 2\hat{y}_{333} - \frac{2}{5}\left\{y_{3\cdot\cdot} + y_{\cdot 3 \cdot} + y_{\cdot\cdot 3} + 3\hat{y}_{333} + \hat{y}_{221}\right\} + \frac{4}{(5)^2}\left\{y_{\cdot\cdot\cdot} + \hat{y}_{333} + \hat{y}_{221}\right\} \quad (13)$$

$$\frac{\partial \text{ Error S.S.}}{\partial(\hat{y}_{221})} = 2\hat{y}_{221} - \frac{2}{5}\left\{y_{2\cdot\cdot} + y_{\cdot 2 \cdot} + y_{\cdot\cdot 3} + y_{333} + 3\hat{y}_{221}\right\} + \frac{4}{(5)^2}\left\{y_{\cdot\cdot\cdot} + \hat{y}_{333} + \hat{y}_{221}\right\} \quad (14)$$

The equations (13) and (14) are equated to zero. By solving equations (13) and (14) simultaneously for $\hat{y}_{333}$ and $\hat{y}_{221}$ the equations (15) and (16) have been obtained.

$$\hat{y}_{333} = \frac{5(y_{3\cdot\cdot} + y_{\cdot 3 \cdot} + y_{\cdot\cdot 3}) - 2y_{\cdot\cdot\cdot}}{\frac{5(5-1)^2}{(5-1)}} + \frac{5(y_{2\cdot\cdot} + y_{\cdot 2 \cdot} + y_{\cdot\cdot 1} - 2(y_{\cdot\cdot\cdot})}{5(5-1)^2} \quad (15)$$

$$\hat{y}_{221} = \frac{5(y_{2\cdot\cdot} + y_{\cdot 2 \cdot} + y_{\cdot\cdot 1}) - 2y_{\cdot\cdot\cdot}}{\frac{5(5-1)^2}{(5-1)}} + \frac{5(y_{3\cdot\cdot} + y_{\cdot 3 \cdot} + y_{\cdot\cdot 3}) - 2y_{\cdot\cdot\cdot}}{5(5-1)^2} \quad (16)$$

Similarly the formula can be developed for any missing observations by the least square methods. Bartlett[26] suggests the procedure of inserting one for the missing values and zeros, otherwise, and performing a covariance analysis with the zeros and one as an independent variate. If more than one observation is missing the same procedure is followed except a multiple covariance is performed. Yates[27] has given the iterative method for estimating the yields for several missing values in any Latin square design.

Analysis of variance is performed similar to Latin square design except for each missing datum computed, one degree of freedom is subtracted from errors degree of freedom.

Now if one assumed that the y_{221} and y_{333} were missing in the table 4.33, the missing values could be computed from the developed formulas (15) and (16).

$$\hat{y}_{333} = \frac{5(5.61 + 5.53 + 5.70) - 2(28.35)}{\frac{5(5-1)^2}{(5-1)}} + \frac{5(5.99 + 5.62 + 6.33) - 2(28.35)}{5(5-1)^2}$$

$$= \frac{27.5}{20} + \frac{33}{80}$$

$$= 1.78$$

---

[26]M.S. Bartlett, "Some Examples of Statistical Methods of Research in Agriculture and Applied Biology," Journal of Psycological Statistical Society, Suppl. IV, (1937), pp. 137-83.

[27]F. Yates, "The Analysis of Replicated Experiments When the Field Results Are Incomplete," Empire Journal of Experimental Agriculture, I, (1933), pp. 129-42.

$$\hat{y}_{221} = \frac{5(5.61 + 5.53 + 5.70) - 2(28.35)}{5(5-1)^2} + \frac{5(5.99 + 5.62 + 6.33) - 2(28.35)}{\frac{5(5-1)^2}{(5-1)}}$$

$$= \frac{33}{20} + \frac{27.5}{80}$$

$$= 1.99$$

The missing values y_{333} and y_{221} should be replaced by 1.78 and 1.99, and the analysis of variance is performed similar to Latin square design except two degrees of freedoms are subtracted from total and error degrees of freedom.

4.6. Summary. In the Latin square design sometimes experiments present more than one observation per experimental unit. Federer has given an example and the analysis of variance for such a problem.[28] A typical problem of the "Bliss and Rose" experiment in Latin square has been given by Edward[29] in his psychological research work.

---

[28] Federer, op. cit., pp. 153-57.

[29] A.L. Edward, Experimental Design in Psychological Research (New York, Holt, Rinehart and Winston, 1965), pp. 259-64.

CHAPTER V

CONCLUSION

5.1. Summary. This paper has dealt with the applications of the analysis of variance in three major randomized designs. It was shown that the Latin square design and the randomized block design were less efficient, but in some experiments they controlled more variability and so reduced the error variance. For example Cochran and Cox found the efficiency of the Latin square design for a specific example, relative to the completely randomized design was 222% and relative to the randomized block design was 137%. This implies that 10 replicates of a completely randomized design or 6 replicates of a randomized block design are roughly equivalent to 4 or 5 replicates of a Latin square.

It seems that the completely randomized design is the simplest one and the analysis of variance remains the same with unequal number of replicates for different treatments. So if observations are missing, one does not worry about it.

It was shown in Chapter III and IV, how to use the analysis of variance and select the appropriate experimental design when experimental materials are not homogeneous. It was also shown how to develop the missing values and to carry the analysis of variance. The problem of the Latin square design illustrates that the experiment has proportional treatment effects rather than additive, and so log transformation was used on the original data and the analysis of variance was performed on transformed data.

5.2. Suggestions for Further Study. Statistical procedures and concepts are useful in the analysis of data and in the interpretation of the results from an experiment. A number of useful statistical tools--test of significance for comparisons among a set of ranked means, transformations for experimental data might merit further study.

Also the use of the analysis of variance in other randomized group designs might prove to be an interesting research.

Also the other powerful technique, the analysis of co-variance has been more popular in psychological research work, and the further study of its applications can lead to an interesting research.

BIBLIOGRAPHY

BIBLIOGRAPHY

A. BOOKS

1. Anderson, R.L., and Bancraft, T.A. Statistical Theory in Research. New York: McGraw Hill, 1952.

2. Brownlee, K.A. Industrial Experimentation. Brooklyn: Chemical Publishing Co., 1948.

3. Cochran, W.G. and Cox, G.M. Experimental Designs. New York: Wiley Publications in Statistics, 1957.

4. Cramer, H. Mathematical Methods of Statistics, Princeton: Princeton University Press, 1946.

5. Davies, O.L. Design and Analysis of Industrial Experiments. London: Oliver and Boyd, 1954.

6. Dixon, W.J. and Massey, F.J. Introduction to Statistical Analysis, New York: McGraw Hill, 1967.

7. Deming, W.E. Statistical Adjustment of Data. New York: Wiley, 1946.

8. Edwards, A.L. Experimental Design in Psychological Research. New York: Holt, Rinehart and Winston, 1965.

9. Eisenhart, C., Hastay, and Wallis. Selected Techniques of Statistical Analysis. New York: McGraw Hill, 1947.

10. Federer, W.T. Experimental Design. New York: The MacMillan Co., 1955.

11. Finney, D.J. Experimental Design and its Statistical Basis. Chicago: University of Chicago Press, 1955.

12. Fisher, R.A. The Design of Experiments. New York: Hafner Publishing Co., Inc., 1960.

13. Fisher, R.A. and Yates, F. Statistical Tables for Biological, Agriculture, and Medicine Research. New York: Hafner, 1948.

14. Kempthorne, O. The Design and Analysis of Experiments. New York: John Wiley and Sons, 1952.

15. Kendall, M.G. and Stuart, A. *The Advanced Theory of Statistics*. New York: Hafner Publishing Co., 1963-67.

16. Lindquist, E.F. *Design and Analysis of Experiments in Psychology and Education*. Boston: Houghton Mifflin Co., 1956.

17. Mann, H.B. *Analysis and Design of Experiments*. New York: Dover Publications, 1949.

18. Rao, C.R. *Advanced Statistical Methods in Biometric Research*. New York: John Wiley and Sons, 1952.

19. Snedecor, G.W. and Cochran, W.G. *Statistical Methods*. Ames: The Iowa State University Press, 1967.

20. __________, *Statistical Methods*. Ames: Iowa State College Press, 1956.

B. PERIODICALS

1. Bartlett, M.S., "The square root transformation in analysis of variance," Journal of Royal Statistical Society, Suppl. III, 1936.

2. ________, "A note on the analysis of variance," Journal of Agricultural Science, 26, 1936.

3. ________, "Some examples of statistical methods of research in agriculture and applied biology," Journal of Psycological Statistical Society, Suppl. IV, 1937.

4. ________, "The use of transformations," Biometrics, 3, 1947.

5. Cochran, W.G., "Recent work on the analysis of variance," Journal of Royal Statistical Society, 101, 1936.

6. Cochran, W.G., "Some consequences when the assumptions of analysis of variance are not satisfied," Biometrics 3, 1947.

7. Delury, D.B., "The analysis of latin square when some observations are missing," Journal of American Statistical Association, 41, 1946.

8. ________, "The analysis of covariance," Biometrics 4, 1948.

9. Duncan, D.B., "Multiple range and multiple F test," Biometrics, March, 1955.

10. Kemphorne, O. and Barclay, W.G., "The partition of error in randomized blocks," American Statistical Association Journal, September, 1953.

11. Kramer, C.Y., "Extension of multiple range test to group correlated adjusted means," Biometrics, March, 1957.

12. ________, "Extension of multiple range tests to group means with unequal numbers of replicates," Biometrics, September, 1956.

13. Yates, F., "The analysis of replicated experiments when the field results are incomplete," Empire Journal of Experimental Agriculture, I, 1933.