AN INVESTIGATION INTO MISSING OBSERVATIONS IN A RANDOMIZED BLOCK EXPERIMENT

A THESIS

PRESENTED TO

THE FACULTY OF GRADUATE STUDIES

UNIVERSITY OF MANITOBA

IN PARTIAL FULFILLMENT

OF THE REQUIREMENTS OF THE DEGREE

MASTER OF SCIENCE

BY

MARY LYNN FOLLIOTT

APRIL 1970

ABSTRACT

In experimental research, it is not always possible to obtain the desired information from each experimental unit. This can be due to such things as death, disease or human errors. This paper investigates the estimation and analysis of these experiments. The investigation is done in reference to the randomized block design.

The procedure of estimating the missing data is divided into two cases namely, one missing observation and more than one missing value. For the case of one missing value, either a least squares analysis or a covariance analysis is used to calculate an estimate. When there are more than one missing observation, a procedure presented in 1959 by Biggers is used to obtain the estimates. Using an estimate in an analysis of variance, introduces a bias in the treatment sum of squares. The formulation of this bias is derived for the most general case.

Experimental results generated by a computer are used to obtain the emperical powers of the F values associated with the data with estimates for missing observations. These experimental results are compared with the theoretical values of the power.

TABLE OF CONTENTS

Chapter		Page
I.	INTRODUCTION	1
II.	ONE MISSING VALUE	4
III.	MORE THAN ONE MISSING VALUE	17
IV.	ANALYSIS OF DATA	28
V.	SUMMARY AND CONCLUSIONS	33
APPENDIX A: COMPUTER PROGRAM		37
APPENDIX B: TABLES		61
BIBLIOGRAPHY		68

CHAPTER I

INTRODUCTION

A statistical analysis of experimental data obtained from a population with an underlying linear model is carried out under the following assumptions, 1) randomness of the errors, 2) their mutual independence, 3) normality and 4) homogeneity of variance. In most of these situations, it is necessary to have the observed values for each experimental unit. In those cases when one or more observations are missing, a value must be supplied to these experimental units so that the standard analysis can be performed. The methods of estimating values for the missing data, the effect of these estimates on the analysis of the experiment, and the power of the statistical test used in the analysis will be investigated in this paper. This will be done in reference to the randomized block model.

In the case of one missing observation, least squares analysis and covariance analysis are used to calculate the required estimate. Both methods are theoretically identical, but differ in ease of application. For manual calculations, the least squares analysis is used but when the calculations are done by a computer, a covariance analysis is performed. The covariance method lends itself more easily to programming. When more than one observation is missing, a procedure developed by Biggers (1959) is used to obtain the essential estimates.

After supplying the required estimates, an analysis of variance performed on the data will result in an upward bias in the treatment sum of squares. The formula given by Kenney and Keeping (1951) is valid only for the case when there is at most one missing value in a treatment or block group. The general formula for which this bias is a special case will be derived in chapter III.

Emperical results, generated by an I.B.M. 360/65, are used to obtain the experimental or Monte Carlo power of the F-test associated with the treatment sum of squares. The power is found by counting the proportion of F values greater than the critical value, in a set of K independent identically distributed F values. This estimate of the power can be compared with the theoretical power obtained from tables given by Tiku (1967).

Allan and Wishart (1930) were the first to develop a formula which estimated a value for a missing value. In 1933, Yates demonstrated that Allan and Wishart had obtained their formula by minimizing the sum of squares due to error with respect to the missing observation. Yates also presented an iterative procedure for calculating estimates when there are several missing values.

Tables for the power of the F-test were published by Tang (1938). He developed a set of tables which depended on the degrees of freedom of the F-test and a function ϕ of the non-centrality parameter λ. Pearson and Hartley (1951) and Fox (1956) charted the function ϕ dependent on the degrees of freedom and for several values of α the probability of the Type I error and β the probability of the Type II error. Tiku (1967) extended Tang's tables and also gave a method of interpolation to obtain a value of the power not given in the tables.

CHAPTER II

ONE MISSING VALUE

It sometimes happens that one or more observations in an experiment are missing. In such cases, an estimate for each missing value must be obtained. If an analysis is performed with an experimental unit missing, the sum of squares for treatments, blocks and error do not add up to the total sum of squares. In other words, the model is not additive. This can be shown in an example as follows. Given a randomized block design with 3 treatments and 3 blocks, and a missing observation in treatment 2 block 3,

BLOCK	TREATMENT			TOTAL	MEAN
	1	2	3		
1	9	3	9	21	7
2	8	5	2	15	5
3	4		10	14	7
TOTAL	21	8	21	50	$6\frac{1}{4}$
MEAN	7	4	7	$6\frac{1}{4}$	

The analysis is made omitting observation x_{32}.

$$TSS = 9^2 + 3^2 + \ldots + 4^2 + 10^2 - \frac{(50)^2}{8} = 67\tfrac{1}{2}$$

$$SST = \frac{(21)^2}{3} + \frac{(8)^2}{2} + \frac{(21)^2}{3} - \frac{(50)^2}{8} = 7\tfrac{1}{2}$$

$$SSB = \frac{(21)^2}{3} + \frac{(15)^2}{3} + \frac{(14)^2}{2} - \frac{(50)^2}{8} = 13\tfrac{1}{2}$$

$$SSE = TSS - SST - SSB = 47\tfrac{1}{2}$$

The calculation of the error sum of squares is given by.

$$SSE = (9-7-7+6\tfrac{1}{4})^2 + (3-7-4+6\tfrac{1}{4})^2 + \ldots + (10-7-7+6\tfrac{1}{4})^2 = 49\tfrac{1}{2}$$

SSE (subtraction) $\neq$ SSE (calculated).

Two familiar methods of calculating a missing value are,

1) least squares analysis,

2) covariance analysis.

In least squares analysis, we begin with the mathematical model of the randomized block design.

$$x_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij} \qquad \begin{array}{l} i=1,2,\ldots,t \\ j=1,2,\ldots,b \end{array}$$

where x_{ij} is defined in terms of a general mean μ, a treatment effect α_i, a block effect β_j and a random error ε_{ij}. The values of μ, α_i and β_j are determined so as to minimize the sum of squares due to error. The sum of squares due to error if found to be equal to,

$$\Sigma_i \Sigma_j x_{ij}^2 - \frac{\Sigma_i(\Sigma_j x_{ij})^2}{b} - \frac{\Sigma_j(\Sigma_i x_{ij})^2}{t} + \frac{(\Sigma_i \Sigma_j x_{ij})^2}{tb}$$

When a value is missing an estimate can be found by minimizing the sum of squares due to error with respect to this value. By obtaining an estimate in this way, a value is found which contributes least to the error sum of squares.

To obtain the value which is to be assigned to the missing experimental unit consider the following in which x_{gh} is missing.

$$SSE = \Sigma_i \Sigma_j x_{ij}^2 - \frac{\Sigma_i (\Sigma_j x_{ij})^2}{b} - \frac{\Sigma_j (\Sigma_i x_{ij})^2}{t} + \frac{(\Sigma_i \Sigma_j x_{ij})^2}{tb}$$

$$\frac{\partial}{\partial x_{gh}} SSE = x_{gh} - \frac{\Sigma_j x_{gj}}{b} - \frac{\Sigma_i x_{ih}}{t} + \frac{\Sigma_i \Sigma_j x_{ij}}{tb}$$

Set this equal to zero and solve for x_{gh}.

$$x_{gh}(1 - \frac{1}{b} - \frac{1}{t} + \frac{1}{tb}) + (-\frac{\Sigma_{j \neq h} x_{gj}}{b} - \frac{\Sigma_{i \neq g} x_{ih}}{t} + \frac{\Sigma_{i \neq g} \Sigma_{j \neq h} x_{ij}}{tb} = 0$$

$$x_{gh}(\frac{(b-1)(t-1)}{tb}) = \frac{(tT'_g + bB'_h - G')}{tb}$$

$$x_{gh} = \frac{(tT'_g + bB'_h - G')}{(t-1)(b-1)}$$

where

$$T'_g = \Sigma_j x_{gj} - x_{gh}$$

$$B'_h = \Sigma_i x_{ih} - x_{gh}$$

$$G' = \Sigma_i \Sigma_j x_{ij} - x_{gh}$$

Having an estimate of x_{gh}, the analysis of variance can be carried out in the usual manner. However, the treatment sum of squares when calculated using this estimate is biased. By this we mean that the expected value of the treatment mean squares is equal to

$$\sigma^2 + \frac{b}{(t-1)} \Sigma_i \alpha_i^2 + q.$$

Under the null hypothesis H_o: $\alpha_i=0$, this value becomes

$$\sigma^2 + q', \text{ where } q'>0.$$

The usual F-test gives rise to an exaggeration of the treatment effect.

The amount of bias can be found by considering the hypothesis that there are no differences between treatments ie. H_o: $\alpha_i=0$ for all i. By combining the treatment and error sums of squares and minimizing this conditional error denoted by S_c with respect to the missing value, a different estimate for the missing observation will be obtained. This is demonstrated by,

$$S_c = \Sigma_i \Sigma_j x_{ij}^{*2} - \frac{\Sigma_j (\Sigma_i x_{ij}^*)^2}{t}$$

$$\frac{\partial}{\partial x_{gh}^*} S_c = x_{gh}^* - \frac{\Sigma_i x_{ih}^*}{t}$$

Set equal to zero and solve for x_{gh}^*.

$$x_{gh}^*(1 - \frac{1}{t}) = \frac{B_h'}{t}$$

$$x^*_{gh} = \frac{B'_h}{(t-1)} = A$$

The bias is introduced by using

$$x_{gh} = \frac{tT'_g + bB'_h - G'}{(t-1)(b-1)}$$

as an estimate instead of

$$x_{gh} = \frac{B'_h}{(t-1)}$$

in testing the null hypothesis. The former should be used in order to obtain an unbiased estimate of the missing value. The amount of bias is,

$$x_{gh}^2 - A^2 - \frac{(x_{gh} - B'_h)^2}{t} + \frac{(A - B'_h)^2}{t}$$

$$= \frac{(t-1)}{t}(x_{gh}^2 - A^2) - \frac{2B'_h}{t}(x_{gh} - A)$$

$$= \frac{(t-1)}{t}(x_{gh} - A)(x_{gh} + A - 2A)$$

$$= \frac{(t-1)}{t}(x_{gh} - A)^2$$

$$= \frac{(t-1)}{t}(x_{gh} - \frac{B'_h}{(t-1)})^2$$

Considering the degrees of freedom associated with the error sum of squares, each observation x_{ij} has an associated ε_{ij}. It is assumed that $\Sigma_i \varepsilon_{ij} = \Sigma_j \varepsilon_{ij} = 0$. Besides this, it is known that $\varepsilon_{gh}=0$ since observation x_{gh}

is missing. Evaluating the number of dependent ε_{ij}'s, there are t treatments and b blocks, hence there are t+b-1 dependent ε_{ij}'s. Besides this there is one due to ε_{gh} being equal to zero. As a result, there will be tb-(t+b-1+1) or tb-t-b = (t-1)(b-1)-1 independent ε_{ij}'s and the degrees of freedom associated with the sum of squares due to error is one less than the case with no missing values.

Consider estimating the value of the observation missing in the example at the beginning of this chapter ie. x_{32}.

$$x_{32} = \frac{3(8) + 3(14) - 50}{(2)(2)} = 4$$

ANOVA

SOURCE OF VARIATION	D.F.	SS	MS	F
Treatments(adj.)	2	18-6=12	6	0.375
Blocks	2	6	3	
Error	4-1=3	48	16	
Total	8-1=7	72-6=66		

$$\text{bias} = \frac{2}{3}\left(4 - \frac{14}{2}\right)^2 = 6$$

In covariance analysis the mathematical model for a randomized block design is given by,

$$y_{ij} = \mu + \alpha_i + \rho_j + \beta x_{ij} + \varepsilon_{ij}$$
$$i=1,2,\ldots,t$$
$$j=i,2,\ldots,p$$

where y_{ij} is the dependent variable given in terms of a general mean μ, a treatment effect α_i, a block effect ρ_j, a β-multiple of the independent variable x_{ij}, and a random error ε_{ij}. The underlying assumptions of the model are,

1) the x's are fixed variables and measured without error.

2) the ε's are independent normally distributed variables with mean zero and common variance σ^2.

3) the regression of y on x after the removal of the treatment and block effects is linear and independent of treatments and blocks.

When an observation is missing, an estimate of its value is found by performing a covariance with the dependent y-variates being the existing observations with zero for the missing value, and the independent x-variates consisting of zeros corresponding to the existing observations and a minus one for the missing observation. The value of the regression coefficient b is the estimate used for the the missing value. This can be shown by the following, where x_{gh} is assumed to be missing.

$$\hat{y}_{ij} = \bar{y}_{i.} + \bar{y}_{.j} - \bar{\bar{y}}_{..} + b(x_{ij} - \bar{x}_{i.} - \bar{x}_{.j} + \bar{\bar{x}}_{..})$$

hence

$$y_{gh} = \bar{y}_{g.} + \bar{y}_{.h} - \bar{\bar{y}}_{..} + b(x_{gh} - \bar{x}_{g.} - \bar{x}_{.h} + \bar{\bar{x}}_{..})$$

now

$$\bar{y}_{g.} = \frac{T'_g}{p} \qquad \bar{y}_{.h} = \frac{B'_h}{t} \qquad \bar{\bar{y}}_{..} = \frac{G'}{tp}$$

$$\hat{y}_{gh} = \frac{T'_g}{p} + \frac{B'_h}{t} - \frac{G'}{tp} + b(x_{gh} - \bar{x}_{g.} - \bar{x}_{.h} + \bar{\bar{x}}_{..})$$

also

$$b = \frac{tT'_g + pB'_h - G'}{(t-1)(p-1)}$$

for $x_{gh} = -1$,

$$\hat{y}_{gh} = \frac{tT'_g + pB'_h - G'}{tp} + \left(\frac{tT'_g + pB'_h - G'}{(t-1)(p-1)}\right)\left(-1 + \frac{1}{p} + \frac{1}{t} - \frac{1}{tp}\right)$$

$$= \frac{tT'_g + pB'_h - G'}{tp} + \left(\frac{tT'_g + pB'_h - G'}{(t-1)(p-1)}\right)\left(-\frac{(t-1)(p-1)}{tp}\right)$$

$$= 0$$

for $x_{gh} = 0$,

$$\hat{y}_{gh} = \frac{tT'_g + pB'_h - G'}{tp} + \left(\frac{tT'_g + pB'_h - G'}{(t-1)(p-1)}\right)\left(0 + \frac{1}{p} + \frac{1}{t} - \frac{1}{tp}\right)$$

$$= \frac{tT'_g + pB'_h - G'}{tp} + \left(\frac{tT'_g + pB'_h - G'}{(t-1)(p-1)}\right)\left(-1 + \frac{1}{p} + \frac{1}{t} - \frac{1}{tp} + 1\right)$$

$$= 0 + \frac{tT'_g + pB'_h - G'}{(t-1)(p-1)}$$

$$= b$$

This shows that for x_{gh} equal to minus one, the corresponding y-variate will be zero. If instead of minus one, the value of x_{gh} had been zero (as was the case for the other observations), the corresponding y value would be the regression coefficient b. Hence the estimate used for the missing y_{gh} is b.

A complete analysis of covairiance table for the design is given below.

ANCOVA

SOURCE	D.F.	SUMS OF PROD. XX	XY	YY	D.F.	ADJ. ΣYY
TOTAL	tp-1	Σxx	Σxy	Σyy		
BLOCKS	p-1	B_{xx}	B_{xy}	B_{yy}		
TREAT.	t-1	T_{xx}	T_{xy}	T_{yy}		
ERROR	(t-1)(p-1)	E_{xx}	E_{xy}	E_{yy}	(t-1)(p-1)-1	$E_{yy}-\frac{E_{xy}^2}{E_{xx}}$
T + E	p(t-1)	S_{xx}	S_{xy}	S_{yy}	p(t-1)-1	$S_{yy}-\frac{S_{xy}^2}{S_{xx}}$
T(ADJ)					t-1	$S_{yy}-\frac{S_{xy}^2}{S_{xx}} - E_{yy}+\frac{E_{xy}^2}{E_{xx}}$

Notation:

$$\Sigma xx = \Sigma_i \Sigma_j (x_{ij} - \bar{\bar{x}}_{..})^2 = \frac{tp-1}{tp} \qquad \Sigma yy = \Sigma_i \Sigma_j (y_{ij} - \bar{\bar{y}}_{..})^2$$

$$\Sigma xy = \Sigma_i \Sigma_j (x_{ij} - \bar{\bar{x}}_{..})(y_{ij} - \bar{\bar{y}}_{..}) = \frac{G'}{tp}$$

$$B_{xx} = \frac{\Sigma_j (\bar{x}_{.j} - \bar{\bar{x}}_{..})^2}{t} = \frac{p-1}{tp} \qquad B_{yy} = \frac{\Sigma_j (\bar{y}_{.j} - \bar{\bar{y}}_{..})^2}{t}$$

$$B_{xy} = \frac{\Sigma_j (\bar{x}_{.j} - \bar{\bar{x}}_{..})(\bar{y}_{.j} - \bar{\bar{y}}_{..})}{t} = \frac{G' - pB'_h}{tp}$$

$$T_{xx} = \frac{\Sigma_i (\bar{x}_{i.} - \bar{\bar{x}}_{..})^2}{p} = \frac{t-1}{tp} \qquad T_{yy} = \frac{\Sigma_i (\bar{y}_{i.} - \bar{\bar{y}}_{..})^2}{p}$$

$$T_{xy} = \frac{\Sigma_i (\bar{x}_{i.} - \bar{\bar{x}}_{..})(\bar{y}_{i.} - \bar{\bar{y}}_{..})}{p} = \frac{G' - tT'_g}{tp}$$

$$E_{xx} = \Sigma xx - B_{xx} - T_{xx} = \frac{(t-1)(p-1)}{tp} \qquad E_{yy} = \Sigma yy - B_{yy} - T_{yy}$$

$$E_{xy} = \Sigma xy - B_{xy} - T_{xy} = \frac{tT'_g + pB'_h - G'}{tp}$$

$$S_{xx} = T_{xx} + E_{xx} = \frac{t-1}{t} \qquad S_{yy} = T_{yy} + E_{yy}$$

$$S_{xy} = T_{xy} + E_{xy} = \frac{B'_h}{t}$$

Comparing the two methods of estimating a missing value, it is noticed that both methods obtain the same estimates for the missing observation. Both analyses decrease the degrees of freedom associated with the error sum of squares by one. The adjustment to the treatment sum of squares can be seen by considering the following,

Notation:

$$SSE = E_{yy} - \frac{E_{xy}^2}{E_{xx}}$$

$$SS(T+E) = S_{yy} - \frac{S_{xy}^2}{S_{xx}}$$

$$SST(ADJ) = SS(T+E) - SSE$$

$$T_{yy} = \frac{\Sigma_i y_{i.}^2}{p} - \frac{(\Sigma_i \Sigma_j y_{ij})^2}{tp}$$

T^*_{yy} is the sum of squares for treatments with the missing y observation having its value of zero replaced by the regression coefficient b.

$$T_{yy}^{*} = \frac{(\Sigma_{i\neq g}\bar{y}_{i.}^{2} + (\bar{y}_{g.} + b)^{2})}{p} - \frac{(\Sigma_i\Sigma_j(y_{ij} + b))^{2}}{tp}$$

$$SS(T+E) = \Sigma yy - B_{yy} - \frac{(\Sigma xy - B_{xy})^{2}}{(\Sigma xx - B_{xx})}$$

$$= \Sigma yy - B_{yy} - \frac{(\frac{G'}{tp} - \frac{G'}{tp} + \frac{pB_h'}{tp})^{2}}{\frac{tp-1}{tp} - \frac{t-1}{tp}}$$

$$= \Sigma yy - B_{yy} - \frac{(B_h')^{2}}{t(t-1)}$$

$$SSE = \Sigma yy - B_{yy} - T_{yy} - \frac{(\Sigma xy - B_{xy} - T_{xy})^{2}}{\Sigma xx - B_{xx} - T_{xx}}$$

$$= \Sigma yy - B_{yy} - T_{yy} - \frac{(tT_g' + pB_h' - G')^{2}}{tp(t-1)(p-1)}$$

$$SST(ADJ) = T_{yy} - \frac{(B_h')^{2}}{t(t-1)} + \frac{b(tT_g' + pB_h' - G')}{tp}$$

$$T_{yy}^{*} = \frac{\Sigma_{i\neq g}\bar{y}_{i.}^{2}}{p} + \frac{\bar{y}_{g.}^{2}}{p} + 2\frac{\bar{y}_{g.}b}{p} + \frac{b^{2}}{p} - \frac{\Sigma_i\Sigma_j y_{ij}^{2}}{tp}$$

$$- 2\frac{b\Sigma_i\Sigma_j y_{ij}^{2}}{tp} - \frac{b^{2}}{tp}$$

$$= \frac{\Sigma_i y_{i.}^{2}}{p} - \frac{\Sigma_i\Sigma_j y_{ij}^{2}}{tp} + 2\frac{bT_g'}{p} - 2\frac{bG'}{tp} + \frac{b^{2}(t-1)}{tp}$$

$$= T_{yy} + 2\frac{bT_g'}{p} - 2\frac{bG'}{tp} + \frac{b^{2}(t-1)}{tp}$$

hence,

$$T_{yy} = T_{yy}^{*} - 2\frac{bT'_g}{p} + 2\frac{bG'}{tp} - \frac{b^2(t-1)}{tp}$$

$$\begin{aligned}
SST(ADJ) &= T_{yy}^{*} - 2\frac{bT'_g}{p} + 2\frac{bG'}{tp} - \frac{b^2(t-1)}{tp} \\
&\quad - \frac{(B'_h)^2}{t(t-1)} + \frac{b(tT'_g + pB'_h - G')}{tp} \\
&= T_{yy}^{*} - 2\frac{bT'_g}{p} + \frac{bT'_g}{p} + 2\frac{bG'}{tp} - \frac{bG'}{tp} \\
&\quad - \frac{bB'_h}{t} - \frac{b^2(t-1)}{tp} - \frac{(B'_h)^2}{t(t-1)} + 2\frac{bB'_h}{t} \\
&= T_{yy}^{*} - \frac{b(tT'_g + pB'_h - G')}{tp} - \frac{b^2(t-1)}{tp} \\
&\quad + 2\frac{bB'_h}{t} - \frac{(B'_h)^2}{t(t-1)} \\
&= T_{yy}^{*} - \frac{b^2(t-1)(p-1)}{tp} - \frac{b^2(t-1)}{tp} + 2\frac{bB'_h}{t} - \frac{(B'_h)^2}{t(t-1)} \\
&= T_{yy}^{*} - \frac{b^2(t-1)}{t} + 2\frac{bB'_h}{t} - \left(\frac{B'_h}{t-1}\right)^2\frac{(t-1)}{t} \\
&= T_{yy}^{*} - \frac{(t-1)}{t}\left(b - \frac{B'_h}{t-1}\right)^2
\end{aligned}$$

Hence the bias is equal to

$$\frac{(t-1)}{t}\left(b - \frac{B'_h}{t-1}\right)^2.$$

This is identical to the bias obtained by least squares analysis. The two methods, therefore yeild the same results.

Consider the example with observation x_{32} missing. Estimate its value using covariance analysis.

ANCOVA

SOURCE	D.F.	SUMS OF PROD XX	XY	YY	b	ADJUSTED YY D.F.	SS	MS	F
TOTAL	8	$\frac{8}{9}$	$\frac{50}{9}$	$\frac{920}{9}$					
BLOCKS	2	$\frac{2}{9}$	$\frac{8}{9}$	$\frac{86}{9}$					
TREAT.	2	$\frac{2}{9}$	$\frac{26}{9}$	$\frac{338}{9}$					
ERROR	4	$\frac{4}{9}$	$\frac{16}{9}$	$\frac{96}{9}$	4	3	48	16	
T + E	6	$\frac{6}{9}$	$\frac{42}{9}$	$\frac{434}{9}$		5	60		
T(ADJ)						2	12	6	$\frac{3}{8}$

CHAPTER III

MORE THAN ONE MISSING VALUE

When the number of missing observations is greater than one, estimates may be obtained by an iterative procedure suggested by Yates that is equivalent to solving a set of simultaneous equations in K unknowns for K missing values. An alternative procedure is to complete a multiple covariance analysis to obtain a set of regression coefficients which are the estimates for the missing observations. Both of the methods become more and more difficult as the number of missing observations increases. Biggers (1959) presented a procedure which simplifies the calculatio of estimates for the cases with more than one missing value.

The procedure developed by Biggers assumes that there are p independent missing values and the error sum of squares is expressed as a quadratic of these p values. The vector of partial derivatives of this function is given by,

$$2\frac{(\underline{x}'A - \underline{q}')}{N}$$

where,

$\underline{x}$ is a column vector of the p missing values,

A is a p×p symmetric matrix determined by the experimental design and the missing value,

$\underline{q}$ is a column vector calculated from the available data,

N is a constant determined by the experimental design.

Equating this function to zero and solving for $\underline{x}$.

$$A\underline{x} = \underline{q}$$

$$\underline{x} = A^{-1}\underline{q}$$

Consider a randomized block design with t treatments and b blocks. The sum of squares due to error is given by,

$$\Sigma_i \Sigma_j (x_{ij} - \bar{x}_{i.} - \bar{x}_{.j} + \bar{\bar{x}}_{..})^2$$

$$i=1,2,\ldots,t$$
$$j=1,2,\ldots,b.$$

Evaluating the error sum of squares in terms of summations over the set of missing observations, one obtains

$$SSE = \Sigma_{(i)} \Sigma_{(j)} (x_{ij} - \bar{x}_{i.} - \bar{x}_{.j} + \bar{\bar{x}}_{..})^2 + C$$

$$= \Sigma_{(i)} \Sigma_{(j)} x_{ij}^2 - \Sigma_{(i)} \Sigma_{(j)} \bar{x}_{i.}^2 - \Sigma_{(i)} \Sigma_{(j)} \bar{x}_{.j}^2$$

$$+ \Sigma_{(i)} \Sigma_{(j)} \bar{\bar{x}}_{..}^2 + C$$

$$= \Sigma_{(i)} \Sigma_{(j)} x_{ij}^2 - \frac{1}{b} \Sigma_{(i)} (T_i + \Sigma_{(j)} x_{ij})^2$$

$$- \frac{1}{t} \Sigma_{(j)} (B_j + \Sigma_{(i)} x_{ij})^2 + \frac{1}{tb} (G + \Sigma_{(i)} \Sigma_{(j)} x_{ij})^2 + C$$

where,

$$T_i = \Sigma_j x_{ij} \qquad B_j = \Sigma_i x_{ij} \qquad G = \Sigma_i \Sigma_j x_{ij}$$

and C is a constant made up of those terms that do not contain the missing value. Differentiating the sum of squares due to error with respect to the missing value x_{gh} and equating to zero, we obtain the following,

$$\frac{\partial}{\partial x_{gh}} SSE = x_{gh} - \frac{1}{b}(T_g + \Sigma_{(j)} x_{gj}) - \frac{1}{t}(B_h + \Sigma_{(i)} x_{ih})$$

$$+ \frac{1}{tb}(G + \Sigma_{(i)} \Sigma_{(j)} x_{ij})$$

$$x_{gh} - \frac{1}{b}\Sigma_{(j)} x_{gj} - \frac{1}{t}\Sigma_{(i)} x_{ih} + \frac{1}{tb}\Sigma_{(i)} \Sigma_{(j)} x_{ij}$$

$$= \frac{T_g}{b} + \frac{B_h}{t} - \frac{G}{tb}$$

Separating into associates, namely those terms with common i and/or j subscripts,

$$tbx_{gh} - tx_{gh} - bx_{gh} + x_{gh} + (1-t)\Sigma_{(j)\, j\neq h} x_{gj}$$

$$+ (1-b)\Sigma_{(i)\, i\neq g} x_{ih} + \Sigma_{(i)\, i\neq g} \Sigma_{(j)\, j\neq h} x_{ij}$$

$$= tT_g + bB_h - G$$

$$(t-1)(b-1)x_{gh} + (1-t)\Sigma_{(j)\, j\neq h} x_{gj}$$

$$+ (1-b)\Sigma_{(i)\, i\neq g} x_{ih} + \Sigma_{(i)\, i\neq g} \Sigma_{(j)\, j\neq h} x_{ij}$$

$$= tT_g + bB_h - G$$

where,

(1-t) is the coefficient associated with the i association ie. treatment associates,

(1-b) corresponds to the j association ie. block associates, 1 is associated with the zero associates,

$(t-1)(b-1)$ is the coefficient associated with the missing observation.

The association matrix is composed of the associations between the missing observation's block and treatment numbers. The q-vector is evaluated as the value of

$$tT_g + bB_h - G$$

for the x_{gh} missing value. Consider the example of a 3×3 randomized block design given in chapter II. Assume observations x_{31} and x [illegible] are missing. The estimates are found in the following manner.

BLOCK	TREATMENT			TOTAL	MEAN
	1	2	3		
1	9	3		12	6
2	8	5	2	15	5
3	4		10	14	7
TOTAL	21	8	12	41	$5\frac{6}{7}$
MEAN	7	4	6	$5\frac{6}{7}$	

The association matrix A is given by,

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

A is obtained in the following manner. The block and treatment numbers for each of the missing values is compared with the block and treatment numbers of the complete set of missing observations. In general, assume that x_{ij} and x_{kl} are the missing values. The comparisons between these values will lead to one of the following four values,

if $i = k$ and $j = l$ then the value is $(1-t)(1-b)$,

if $i = k$ and $j \neq l$ then the value is $(1-t)$,

if $i \neq k$ and $j = l$ then the value is $(1-b)$,

if $i \neq k$ and $j \neq l$ then the value is 1.

This value is placed in the association matrix opposite x_{ij} and x_{kl}. In the example under consideration,

x_{31} and x_{31} give the value 4 to the A-matrix,

x_{31} and x_{23} give the value 1 to the A-matrix,

and x_{23} and x_{23} give the value 4 to the A-matrix.

If instead of observations x_{31} and x_{23} being missing, observations x_{21} and x_{22} were the the missing values. The corresponding association matrix would have been,

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

since x_{21} and x_{22}would have given the value $(1-t) = 1 - 3 = -2$

Returning to the example under discussion,

	31	23	T_i	B_j	G
31	4	1	12	12	41
23	1	4	8	14	41
			t=3	b=3	-1

$$\underline{x} = A^{-1}\underline{q}$$

$$\begin{pmatrix} x_{31} \\ x_{23} \end{pmatrix} = \begin{pmatrix} \frac{4}{15} & \frac{-1}{15} \\ \frac{-1}{15} & \frac{4}{15} \end{pmatrix} \times \begin{pmatrix} 31 \\ 25 \end{pmatrix} = \begin{pmatrix} 6\frac{3}{5} \\ 4\frac{3}{5} \end{pmatrix}$$

$$\text{bias} = \tfrac{2}{3}[\ (6\tfrac{3}{5} - 6)^2 + (4\tfrac{3}{5} - 4)^2\ ]$$

The corrected analysis of variance (CANOVA) is given by,

CANOVA

SOURCES	D.F.	SS	MS	F
TREATS.(ADJ)	2	2.88-0.48=2.40	1.20	0.054
BLOCKS	2	12.58	6.29	
ERROR	4-2=2	44.50	22.25	
TOTAL	8-2=6	59.96-0.48=59.48		

Consider the case where there are more than one missing value in a block. The method of calculating the bias derived in chapter II fails. This can be shown by going back to first principles.

The adjusted sum of squares is found by subtracting the sums of squares due to a combined error and the the error with the estimates substituted in for the missing observations. This combined error term is obtained by finding the difference between the total sum of squares and the sum of squares due to blocks under the assumption that the design has unequal treatment effects in each block. The missing treatment effects are those treatments which are applied to the missing observations.

The bias that results from one missing observation per block was shown to be equal to

$$\frac{(t-1)}{t}\left(E_j - \frac{B'_j}{t-1}\right)^2$$

where the j^{th} block has a missing observation. The bias is subtracted from the treatment sum of squares for each missing observation. The bias which occurs when there are more than one missing observation per block does not reduce to as compact a formula as in the case of one missing observation.

Notation:

$\{Z_{ij}\} = \{X_{ij}\}$ with missing observations

$\{Y_{ij}\} = \{X_{ij}\}$ with estimates for missing data

$\{E_{ij}\}$ = estimate for a missing value

= 0 for an observed value

N is the number of observations

M is the number of missing observations

K is the number of observed observations $\rightarrow K = N - M$

K_j is the number of observed values in the j^{th} block

$\rightarrow \Sigma_j K_j = K$

t is the number of treatments

b is the number of blocks

$$TSS_z = \Sigma_i \Sigma_j z_{ij}^2 - \frac{(\Sigma_i \Sigma_j z_{ij})^2}{K}$$

$$SSB_z = \Sigma_j \frac{(\Sigma_i z_{ij})^2}{K_j} - \frac{(\Sigma_i \Sigma_j z_{ij})^2}{K}$$

$$SSE_z = \Sigma_i\Sigma_j z_{ij}^2 - \Sigma_j\frac{(\Sigma_i z_{ij})^2}{K_j}$$

$$TSS_y = \Sigma_i\Sigma_j y_{ij}^2 - \frac{(\Sigma_i\Sigma_j y_{ij})^2}{N}$$

$$SSB_y = \Sigma_j\frac{(\Sigma_i y_{ij})^2}{t} - \frac{(\Sigma_i\Sigma_j y_{ij})^2}{N}$$

$$SSE_y = \Sigma_i\Sigma_j y_{ij}^2 - \Sigma_j\frac{(\Sigma_i y_{ij})^2}{t} - SST_y$$

$$SST(ADJ)_y = SSE_z - SSE_y$$

$$BIAS = SST_y - SST(ADJ)_y$$

$$= \Sigma_i\Sigma_j y_{ij}^2 - \Sigma_j\frac{(\Sigma_i y_{ij})^2}{t} - \Sigma_i\Sigma_j z_{ij}^2 + \Sigma_j\frac{(\Sigma_i z_{ij})^2}{K_j}$$

$$\Sigma_i\Sigma_j y_{ij}^2 - \Sigma_i\Sigma_j z_{ij}^2 = \Sigma_i\Sigma_j e_{ij}^2$$

$$\Sigma_j\frac{(\Sigma_i y_{ij})^2}{t} - \Sigma_j\frac{(\Sigma_i z_{ij})^2}{K_j}$$

$$= \Sigma_j\left(\frac{(\Sigma_i(z_{ij} + e_{ij}))^2}{t} - \frac{(\Sigma_i z_{ij})^2}{K_j}\right)$$

$$= \Sigma_j\left(\frac{(\Sigma_i z_{ij})^2}{t} + 2\frac{(\Sigma_i z_{ij})(\Sigma_i e_{ij})}{t} + \frac{(\Sigma_i e_{ij})^2}{t}\right.$$

$$\left. - \frac{(\Sigma_i z_{ij})^2}{K_j}\right)$$

$$= \Sigma_j\left(\frac{(K_j-t)(\Sigma_i z_{ij})^2}{K_j t} + 2\frac{(\Sigma_i z_{ij})(\Sigma_i e_{ij})}{t} + \frac{(\Sigma_i e_{ij})^2}{t}\right)$$

$$= \Sigma_j \Big(-\frac{(t-K_j)(1+K_j-1)(t-1-(t-2))}{t}\frac{(\Sigma_i z_{ij})^2}{K_j^2}$$

$$+ 2\frac{(1+K_j-1)(t-1-(t-2))}{t}\left(\frac{\Sigma_i z_{ij}}{K_j}\right)(\Sigma_i e_{ij})$$

$$+ \frac{\Sigma_i e_{ij}^2}{t} + \Sigma_i \Sigma_{r\neq i} e_{ij} e_{rj}\Big)$$

$$= \Sigma_j \Big(-\frac{(t-K_j)(t-1)}{t}\left(\frac{\Sigma_i z_{ij}}{K_j}\right)^2 - \frac{(t-K_j)(K_j-1)}{t}\left(\frac{\Sigma_i z_{ij}}{K_j}\right)^2$$

$$+ \frac{(t-K_j)(t-2)}{t}\left(\frac{\Sigma_i z_{ij}}{K_j}\right)^2 + 2\frac{(t-1)}{t}\left(\frac{\Sigma_i z_{ij}}{K_j}\right)(\Sigma_i e_{ij})$$

$$- 2\frac{(t-2)}{t}\left(\frac{\Sigma_i z_{ij}}{K_j}\right)(\Sigma_i e_{ij}) + 2\frac{(K_j-1)}{t}\left(\frac{\Sigma_i z_{ij}}{K_j}\right)(\Sigma_i e_{ij})$$

$$+ \frac{\Sigma_i e_{ij}^2}{t} + \Sigma_i \Sigma_{r\neq i} e_{ij} e_{rj}\Big)$$

To simplify this formula, use the following

$$\text{FACTOR} = -\Sigma_j \Big(\frac{(t-K_j)(t-K_j-1)}{t}\left(\frac{\Sigma_i z_{ij}}{K_j}\right)^2 - 2\frac{(t-K_j-1)}{t}\left(\frac{\Sigma_i z_{ij}}{K_j}\right)$$

$$\cdot(\Sigma_i e_{ij}) + \Sigma_i \Sigma_{r\neq i} e_{ij} e_{rj}\Big)$$

$$\text{BIAS} = \Sigma_i \Sigma_j e_{ij}^2 - \Sigma_j \Big(-\frac{(t-K_j)(t-1)}{t}\left(\frac{\Sigma_i z_{ij}}{K_j}\right)^2$$

$$+ 2\frac{(t-1)}{t}\left(\frac{\Sigma_i z_{ij}}{K_j}\right)(\Sigma_i e_{ij}) + \frac{\Sigma_i e_{ij}^2}{t}\Big) + \text{FACTOR}$$

$$= \Sigma_j \Big(\frac{(t-1)}{t}\Sigma_i e_{ij}^2 - 2\frac{(t-1)}{t}\left(\frac{\Sigma_i z_{ij}}{K_j}\right)(\Sigma_i e_{ij})$$

$$+ \frac{(t-K_j)(t-1)}{t} (\frac{\Sigma_i z_{ij}}{K_j})^2) + \text{FACTOR}$$

$$= \Sigma_j \frac{(t-1)}{t} (\Sigma_i e_{ij}^2 - 2(\frac{\Sigma_i z_{ij}}{K_j})(\Sigma_i e_{ij})$$

$$(t-K_j)(\frac{\Sigma_i z_{ij}}{K_j})^2) + \text{FACTOR}$$

In the p^{th} block there are $(t-K_p)$ missing values and hence $(t-K_p)$ estimates. These estimates make up the set $\{E'_{gp}\}_{g=1}^{t-K_p}$. Hence by the deffinition of $\{E_{ij}\}$, K_p of the E_{ip}'s are zero. Define B'_j as $\Sigma_i x_{ij} - \Sigma_{(i)} x_{ij}$, where $\Sigma_{(i)} x_{ij}$ is the sum of the missing values in the j^{th} block. The bias can now be written as,

$$\text{BIAS} = \Sigma_j \frac{(t-1)}{t} (\Sigma_{r=1}^{t-K_j} (e'_{rj} - \frac{B'_j}{t-K_j})^2) + \text{FACTOR}$$

or

$$\text{BIAS} = \Sigma_j \frac{(t-1)}{t} (\Sigma_{r=1}^{t-K_j} (e'_{rj} - \frac{B'_j}{t-K_j})^2) - \Sigma_j (\frac{(t-K_j)(t-K_j-1)}{t} (B'_j)^2$$

$$- 2 \frac{(t-K_j-1)}{t} (B'_j)(\Sigma_i e_{ij}) + \Sigma_i \Sigma_{r \neq i} e_{ij} e_{rj})$$

The bias for the cases with only one observation missing per block is a special case of the bias derived above.

$K_j = t - 1$ or $t - K_j = 1$ hence $t - K_j - 1 = 0$

also $\Sigma_i \Sigma_{r \neq i} e_{ij} e_{rj} = 0$, since e_{rj} does not exist for all r. Therefore the trem FACTOR equals zero. The bias becomes,

$$\text{BIAS} = \Sigma_j \frac{(t-1)}{t} (e'_{j1} - \frac{B'_j}{t-1})^2 .$$

CHAPTER IV

ANALYSIS OF DATA

In this chapter the estimation of values for missing observations and the subsequent analysis of the experimental results are dicussed. The computer program that was used to analyse the data is given in Appendix A.

A randomized block design with 3 treatments and 5 blocks is used to illustrate the procedure in calculating missing values, in adjusting for bias and in estimating power. The block effects are represented by the set [-3 -2 0 2 3] and the treatment effects are represented by [-1 0 1]. Random errors were chosen from a normal population having mean equal to zero and a variance of two. Using this information, a set of fifteen observations was generated.

The structural part of the model underlying the observations is

$$X_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij}$$
$$i=1,2,\ldots,t$$
$$j=1,2,\ldots,b$$

An analysis of variance is completed on the data and the F value obtained is stored for future use in the calculation of the power of the F-test.

In the investigation of the problem of missing values, the following procedure was used

1) select one of the fifteen observations at random and delete it from the experiment.
2) estimate a value for the missing observation by a covariance analtsis.
3) obtain the F value associated with the adjusted treatment effect and store this value for future use for the purpose of calculating the power of the F-test.

The two analyses, the complete case and the one with an estimated value, are repeated K times. Experimental powers are obtained from these two sets of F values.

The procedure and analysis for two missing values is completed in a similar manner. The estimates for the missing observations are obtained by using a method developed by Biggers (1959). This procedure is continued for three, four, five and six missing values, respectively.

The limit on the number of missing values allowed in an experiment is dependent on the degrees of freedom associated with the sum of squares due to error. For each missing observation, the degrees of freedom for error loses one degree of freedom. In order to be able to test for treatment differences, we need at least one degree of freedom for error, but in order to apply the power tables given by Tiku (1967), the error sum of squares must have at least two degrees of freedom. Hence, in the present example, there can be at most six missing observations.

In the calculation of power, a procedure outlined by Tang (1938) and dicussed in Scheffé (1951) is used. Given that

$$\alpha = Pr(\text{reject } H_o/H_o)$$

and

$$\beta = Pr(\text{accept } H_o/H_1)$$

then

$$\text{power} = 1 - \beta = Pr(\text{reject } H_o/H_1).$$

The F-test used in testing for differences among treatment effects in an analysis of variance is given by

$$F_{t-1,(t-1)(b-1)} = \frac{\frac{MST}{E(MST)}}{\frac{MSE}{E(MSE)}}$$

where,

$$MST = \frac{1}{t-1}(\Sigma_i \Sigma_j (\bar{x}_{i.} - \bar{\bar{x}}_{..})^2)$$

$$= \frac{1}{t-1}(\Sigma_i \Sigma_j (\alpha_i + \bar{\varepsilon}_{i.} - \bar{\bar{\varepsilon}}_{..})^2)$$

$$MSE = \frac{1}{(t-1)(b-1)}(\Sigma_i \Sigma_j (x_{ij} - \bar{x}_{i.} - \bar{x}_{.j} + \bar{\bar{x}}_{..})^2)$$

$$= \frac{1}{(t-1)(b-1)}(\Sigma_i \Sigma_j (\varepsilon_{ij} - \bar{\varepsilon}_{i.} - \bar{\varepsilon}_{.j} + \bar{\bar{\varepsilon}}_{..})^2)$$

$$E(MST) = \frac{1}{t-1}(\Sigma_i \Sigma_j \alpha_i^2 + \Sigma_i b\frac{\sigma^2}{b} - tb\frac{\sigma^2}{tb})$$

$$= \frac{b}{t-1}\Sigma_i \alpha^2 + \sigma^2$$

$$E(MSE) = \frac{1}{(t-1)(b-1)}(tb\sigma^2 - tb\frac{\sigma^2}{b} - tb\frac{\sigma^2}{t} + tb\frac{\sigma^2}{tb})$$
$$= \sigma^2$$

therefore

$$F_{t-1,(t-1)(b-1)} = \frac{\dfrac{MST}{(\frac{b}{t-1}\Sigma_i\alpha_i^2 + \sigma^2}}{\dfrac{MSE}{\sigma^2}} .$$

Under the null hypothesis H_o: $\alpha_i=0$ for all i, this F is a central F. But, under the alternate hypothesis H_1: $\alpha_i \neq 0$ for some i, the F is a non-central F. The non-centrality parameter λ is given by,

$$\lambda = \frac{b}{2\sigma^2}\Sigma_i\alpha_i^2 .$$

Working with E(MST), it follows that

$$\lambda = \frac{b}{2\sigma^2}\Sigma_i\alpha_i^2 = \frac{SST}{2} - \frac{t-1}{2} .$$

The power of the F-test is therefore dependent on the non-centrality parameter λ and the degrees of freedom associated with the F-test.

Tables given by Tang (1938) and Tiku (1967) give the power in terms of the degrees of freedom and a function ϕ of the non-centrality parameter. This function is given by

$$\phi = \sqrt{\frac{2\lambda}{t}}$$

In the example treated in this chapter, t=3, b=5,

$\sigma^2=2$ hence

$$\lambda = \frac{5}{2(2)}(2) = \frac{5}{2}$$

$$\phi = \sqrt{\frac{5}{3}} = 1.29$$

Tiku (1967) included a method of iteration for intermediate values of β and gave also, illustrations of the accuracy of the procedures.

Tables I, II and III in appendix B, containing the experimental powers were constructed from data collected from one hundred, two hundred and fifty and five hundred trials respectively. Each table consists of ten sets of powers. Each set was initiated by an independent random number. The ideal value of this number is unknown and for this reason ten different numbers were selected to generate the data.

CHAPTER V

SUMMARY AND CONCLUSIONS

Least squares analysis and covariance analysis lead to identical results when calculating an estimate for a missing observation. For the case when the number of missing values exceeds one, the iterative procedure developed by Yates (1933), multiple covariance analysis and a technique developed by Biggers (1959) also give the same results. When these procedures are applied to a set of data, the sum of squares due to treatments is biased upwards. This bias is removed by subtracting the quantity

$$\frac{(t-1)}{t}\left(e_{gh} - \frac{B'_h}{(t-1)}\right)^2$$

for each missing value x_{gh}. This quantity is modified slightly when there are more than one missing value per block. Covariance analysis is specifically suited for this problem since it allows for the correction of the bias automatically in the analysis.

A useful biproduct of this investigation is the computer program in appendix A. This program can be used in part or in whole. Subroutine ANOVA and subroutine ANCOVA can be applied to any randomized block experiment to obtain an analysis of variance and an analysis of covariance, respectively. The procedure devised by Biggers (1959) is performed by using subroutine BIGGER. The correction of the treatment sum of squares is obtained by using subroutine CANOVA. When using the whole program, the user can output any of the following,

1) the data to be analysed,
2) the missing observation value number, ranging from 1 to N by blocks,
3) the estimate for the missing data,
4) an analysis of variance table or any value in the table,
5) an analysis of covariance table or any value in the table,
6) an analysis of variance with the adjustment to the treatment effect or any value in the table.

The values of the experimental powers are given in appendix B. Tables I, II and III in this appendix were found for the F values corresponding to the ANOVA, ANCOVA and CANOVA tables. The proportion of these F values greater than the critical value is an estimate of the theoretical power. Any discrepancies between the emperical and theoretical results may be the result of some flaws in the emperical sampling procedure. For example, if the data generated was not random, the sum of squares due to error would tend to be larger than expected. This in turn would cause the F value associated with the treatment effect to be smaller. The experimental power would thus be less than the actual emperical power.

The areas where the emperical sampling procedure could have been incorrect, are in subroutines RANDU and GAUS. RANDU generates uniformly distributed real random number between zero and one. GAUS calculates a normal variate with mean AM and standard deviation SD. Subroutine RANDU erquires a positive odd integer with at most nine digits. This value is multiplied by the number 65539 which may result in an exponent overflow (a value larger than 2147483647). If this does not occur, the value is scaled down to a number between zero and one. If an exponent overflow does occur, a value unknown to the user would be the result of the multiplication. This value is modified to be a value between zero and one. There is some question as to the 'randomness' of this variable. The second area where an erreo may arise is in subroutine GAUS. GAUS uses twelve uniformly distributed variables to obtain a normal variate. This is done by calculating the sum of the twelve variables, subtracting their expected mean six, this variable approximates a normal zero one variate. A normal variable having mean AM and standard deviation SD is obtained by multiplying by SD and then adding AM. The error which arises by approximating the normal distribution in this way is given by Esseen (1945) to be

$$\left| F_n(x) - \Phi(x) \right| \leq c \frac{1}{\sqrt{n}}$$

In the case where n=12, this error is of the order of 0.29.

Further investigation into the procedure of obtaining random numbers and independent normal variates could lead to more explicit conclusions about the experimental results.

APPENDIX A

COMPUTER PROGRAM

The program which follows, consists of a main program and four subroutines. The main program generates data coming from a population having an underlying randomized block model (4)†. This data is analysed to test for differences among the treatments (7). The data is then modified by deleting an observation (9). An estimate for the missing data is obtained and a second analysis is performed (10). This whole procedure is repeated K times. The number of missing values is increased by one and the complete procedure is again repeated K times. The F values, calculated from the initial and the subsequent analyses, are collected in seperate vectors.

The subroutines used by the main program are ANOVA, ANCOVA, BIGGER and CANOVA. Subroutine ANOVA is used to perform an analysis of variance on the data with no missing observations. An estimate for the case of one missing value and the analysis of the data with this estimate included are obtained by using subroutine ANCOVA.

† the number refers to the block in the flowchart where the procedure is carried out.

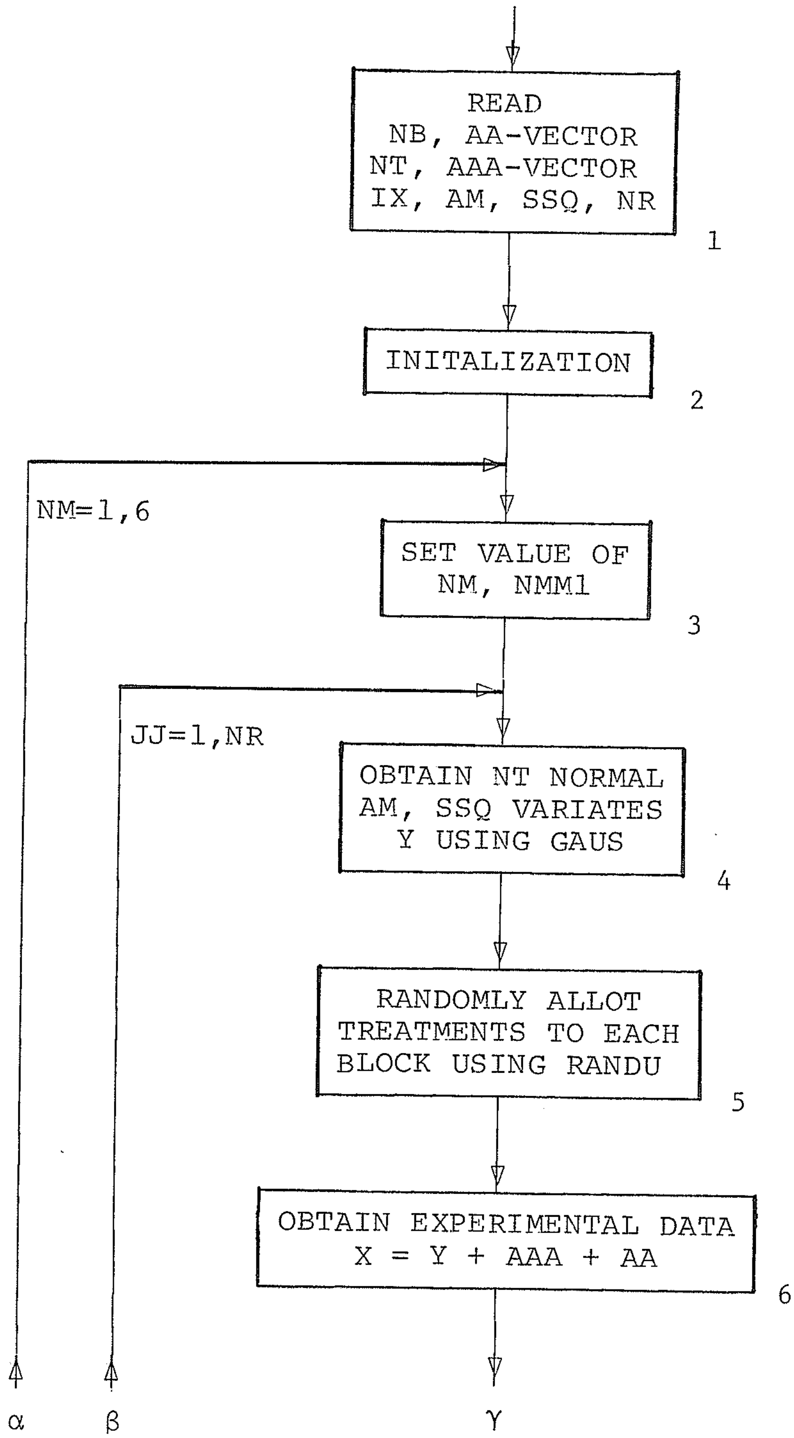
For the case of more than one missing value, subroutine BIGGER is used to calculate the estimates. An analysis, correcting for the bias in the treatment sum of squares, is done by subroutine CANOVA.

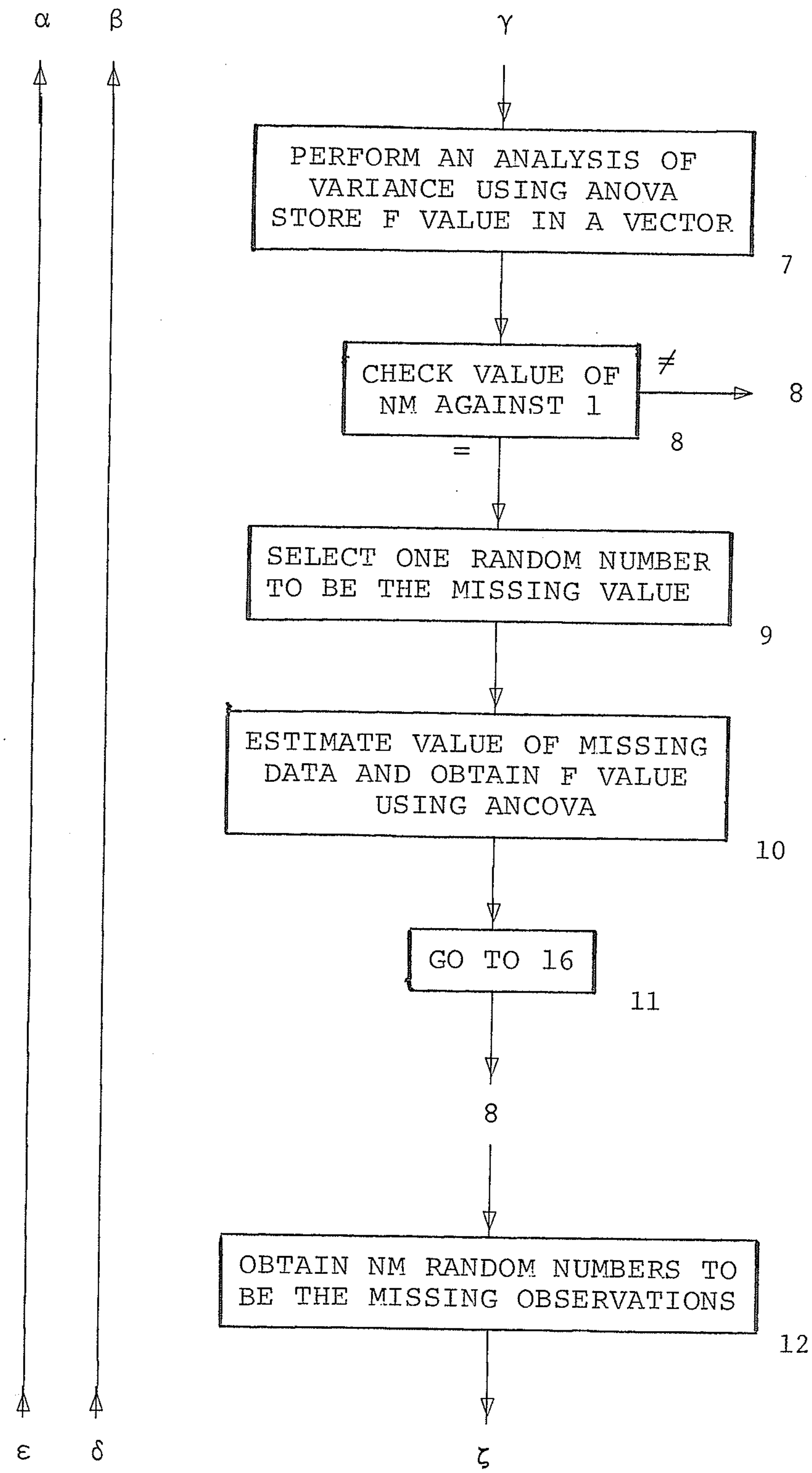
Three other subroutines RANDU, GAUS and MINV are used by the program. These subroutines were written for I.B.M. and are included in the Scientific Subroutine Package. These subroutines can be found in manual number H 20-0205-3. RANDU generates random numbers from a uniform zero one population. Normal variables having mean AM and standard deviation SD are generated by using GAUS. Subroutine MINV inverts a matrix using the standard Gauss-Jordan method.

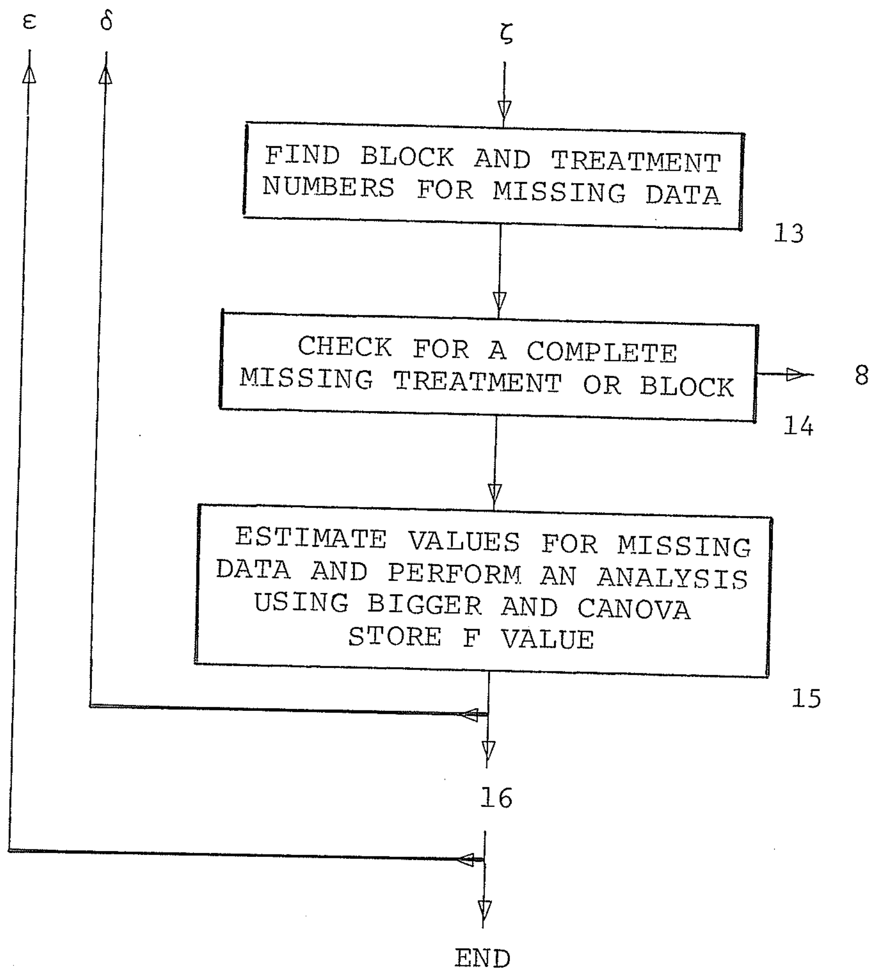
The CPU time required for running the program with an initial value of 13165 and 500 replications is approximately 2½ minutes.

FLOWCHART FOR

"MAIN"

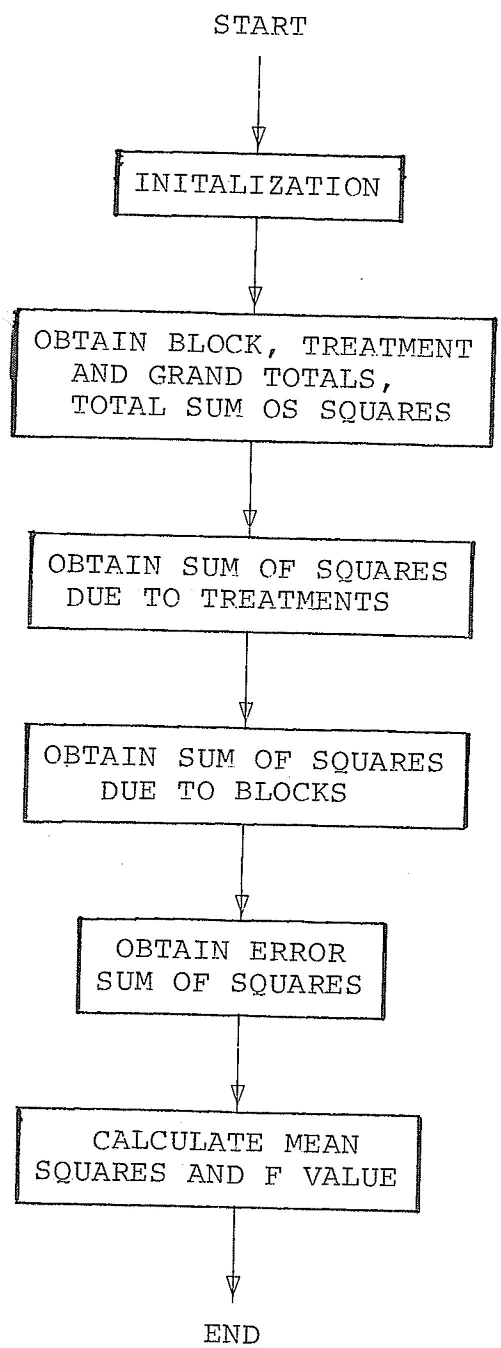



α
β
γ
PERFORM AN ANALYSIS OF VARIANCE USING ANOVA STORE F VALUE IN A VECTOR
7
CHECK VALUE OF NM AGAINST 1
≠
8
8
=
SELECT ONE RANDOM NUMBER TO BE THE MISSING VALUE
9
ESTIMATE VALUE OF MISSING DATA AND OBTAIN F VALUE USING ANCOVA
10
GO TO 16
11
8
OBTAIN NM RANDOM NUMBERS TO BE THE MISSING OBSERVATIONS
12
ε
δ
ζ


ε
δ
ζ
FIND BLOCK AND TREATMENT NUMBERS FOR MISSING DATA
13
CHECK FOR A COMPLETE MISSING TREATMENT OR BLOCK
8
14
ESTIMATE VALUES FOR MISSING DATA AND PERFORM AN ANALYSIS USING BIGGER AND CANOVA
STORE F VALUE
15
16
END

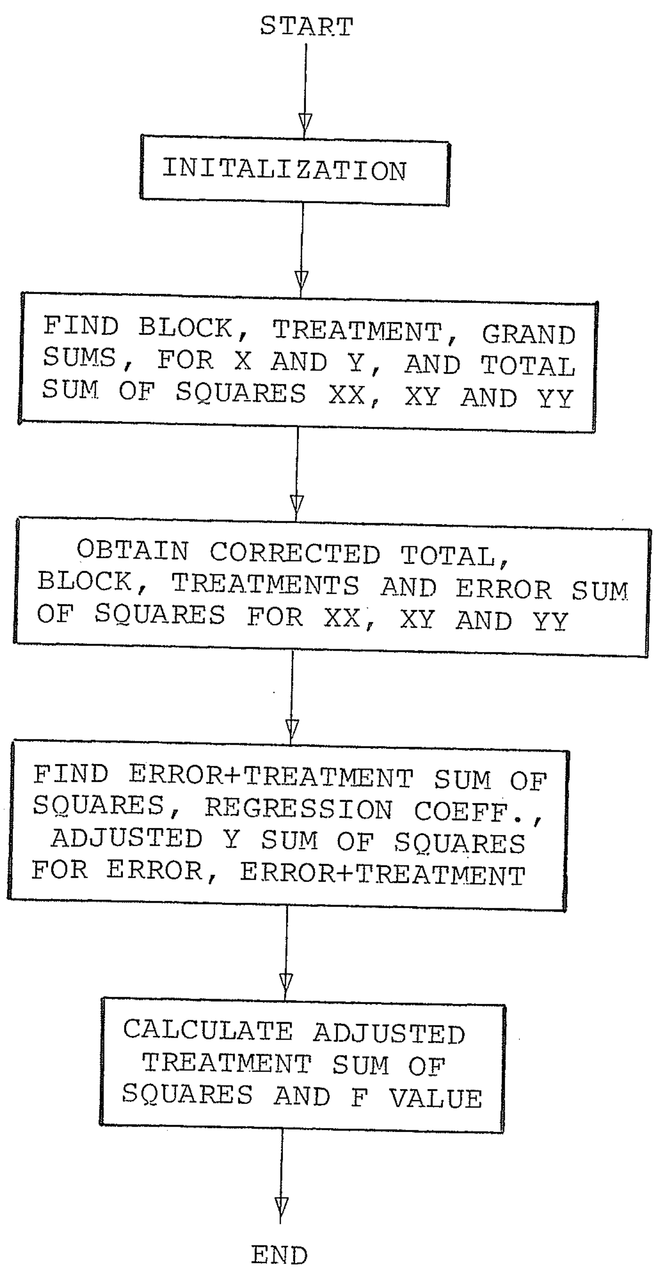
FLOWCHART FOR

SUBROUTINE ANOVA



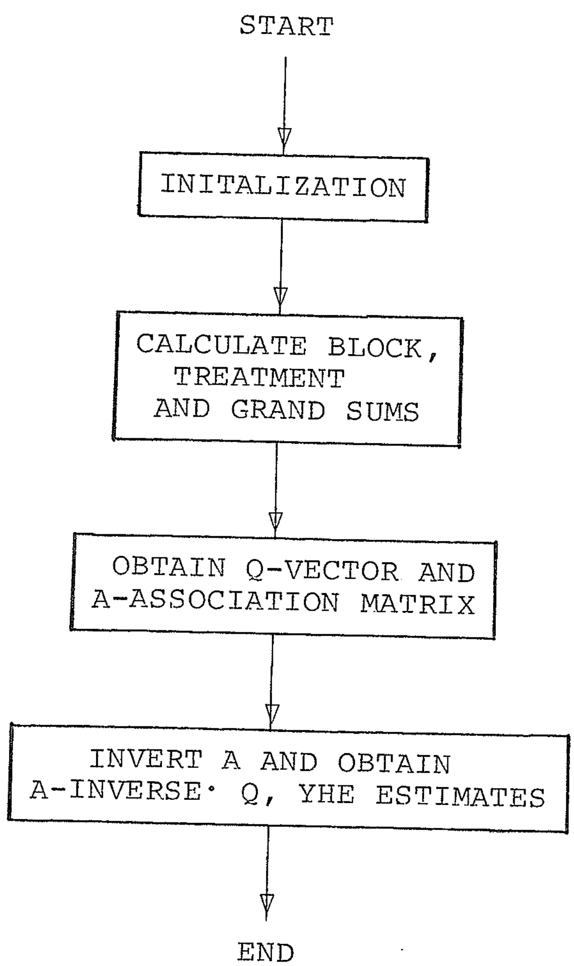
FLOWCHART FOR

SUBROUTINE ANCOVA



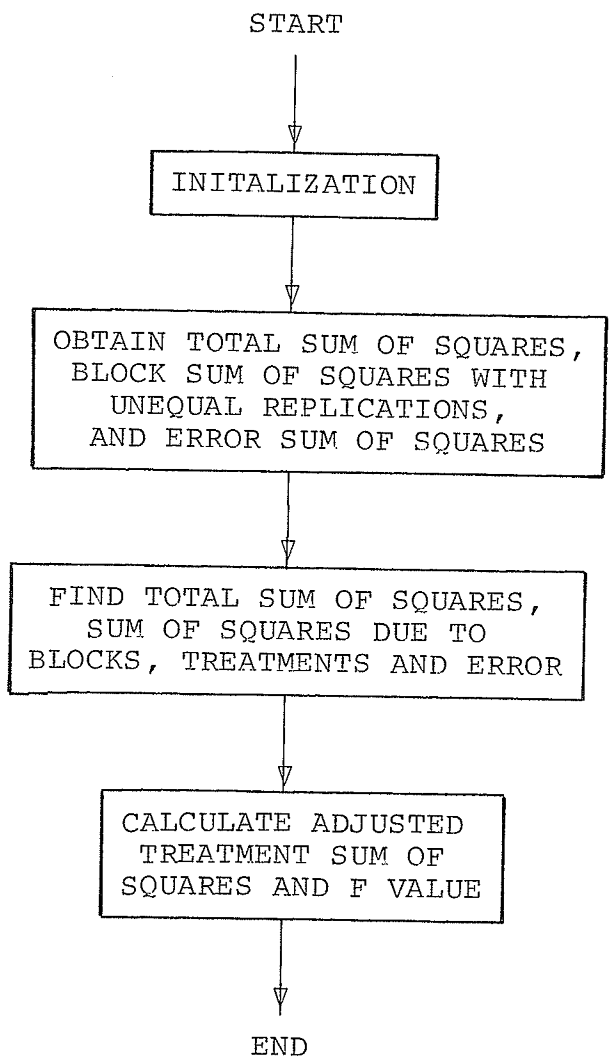
FLOWCHART FOR

SUBROUTINE BIGGER



FLOWCHART FOR

SUBROUTINE CANOVA



```
 DIMENSION X(15), Y(15), IT(15) ,T(3), B(5), S(20), TX(3), TY(3),
1BX(5), BY(5), DF(500), DIF(500), MI(6), MJ(6), M(6), MV(3000),
2Q(6), A(36), AA(15), AAA(15), E(6), LW(6), MW(6)

    SET VALUE OF UNIT CARD

 DATA IR,IW/5,6/

    NB - NUMBER OF BLOCKS
    AA - VECTOR OF LENGTH NB CONTAINING BLOCK EFFECT

 READ(IR,  ) NB, (AA(I),I=1,NB)

    NT  - NUMBER OF TREATMENTS
    AAA - VECTOR OF LENGTH NT CONTAINING TREATMENT EFFECT

 READ(IR,  ) NT, (AAA(I),I=1,NT)

    IX  - ODD NUMBER USED IN GAUSS TO GENERATE RANDOM NUMBER
    AM  - POPULATION MEAN
    SSQ - POPULATION VARIANCE

 READ(IR,  ) IX, AM, SSQ

    NR - NUMBER OF REPLICATIONS

 READ(IR,  ) NR
 SD = SQRT(SSQ)
 N = NT*NB
 L = N - 1
 LB = NB - 1
 LT = NT - 1
 LE = LT*LB
 LTPE = LT + LE
 NRM1 = NR - 1

    SET VALUE OF NM

 DO 18 NM=1,6
 NMM1 = NM - 1

    REPLICATE PROCEDURE NR TIMES

 DO 17 JJ=1,NR
 DO 6 J=1,NB

    INITIALIZE IT VECTOR EQUAL TO I
    IT VECTOR CONTAINS THE TREATMENT NUMBERS
```

```
      GENERATE NT RANDOM NUMBERS FROM N(AM,SSQ) USING GAUSS

   DO 2 I=1,NT
   IT(I) = I
   ABC = 0.0
   DO 1 II=1,12
   IY = IX*65539
   IF(IY.LT.0)IY=IY+2147483647+1
   YFL = IY
   YFL = YFL*0.4656613E-9
   IX = IY
   ABC = ABC + YFL
 1 CONTINUE
   V = (ABC - 6.0)*SD + AM
   Y(I) = V
 2 CONTINUE

      RANDOMLY ORDER IT VECTOR
       - SELECT A RANDOM NUMBER U(0,1):YFL
       - OBTAIN NI, SET SIZE, WHICH INITIALLY IS NT BUT DECREASES BY
         ONE EACH TIME A RANDOM NUMBER IS SELECTED
       - EXPAND TO U(1,NI):K
       - EXCHANGE VALUE OF IT(K) WITH VALUE OF IT(NI)

   DO 3 I=1,LT
   CALL RANDU(IX,IY,YFL)
   IX = IY
   NI = NT + 1 - I
   K = NI*YFL + 1.0
   ITEMP = IT(NI)
   IT(NI) = IT(K)
   IT(K) = ITEMP
 3 CONTINUE

      ORDER IT VECTOR IN INCREASING ORDER, WITH THEIR RESPECTIVE
      N(AM,SSQ) VARIATES

   DO 5 K=1,LT
   IF(IT(K).EQ.K)GO TO 5
   KP1 = K + 1
   DO 4 KK=KP1,NT
   IF(IT(KK).NE.K)GO TO 4
   ITEMP = IT(K)
   IT(K) = IT(KK)
   IT(KK) = ITEMP
   TEMP = Y(K)
   Y(K) = Y(KK)
   Y(KK) = TEMP
```

```
    4 CONTINUE
    5 CONTINUE

          ADD APPROPRIATE BLOCK AND TREATMENT EFFECTS TO EACH NORMAL
          VARIATE

      DO 6 I=1,NT
      IJ = I + (J - 1)*NT
      X(IJ) = Y(I) + AAA(I) + AA(J)
    6 CONTINUE

          PREFORM ANOVA
          STORE F VALUE IN VECTOR DIF

      CALL ANOVA(X,NB,NT,S,F,B,T,G)
      DIF(JJ) = F

          CHECK VALUE OF NM
           - IF NM EQUALS 1 CONTINUE TO ESTIMATE BY COVARIANCE METHOD
           - IF NM GREATER THAN 1 GO TO STATEMENT NUMBER 8 AND ESTIMATE
             BY BIGGER'S METHOD

      IF(NM.NE.1)GO TO 8

          OBTAIN A RANDOM NUMBER FROM U(1,N) TO BE USED AS MISSING
          OBSERVATION NUMBER

      CALL RANDU(IX,IY,YFL)
      IX = IY
      I = N*YFL + 1
      MV(JJ) = I

          PERFORM ANCOVA TO OBTAIN REGRESSION COEFFICIENT WHICH ESTIMATES
          THE VALUE OF THE MISSING OBSERVATION
          STORE F-VALUE IN VECTOR DF

      DO 7 J=1,N
      Y(J) = 0.0
    7 CONTINUE
      X(I) = 0.0
      Y(I) = -1.0
      NDF1 = N
      NDF2 = NB
      NDF3 = NT
      CALL ANCOVA(X,Y,NDF1,NDF2,NDF3,REGCF,S,F,BX,BY,TX,TY,GX,GY)
      DF(JJ) = F
      GO TO 17
```

```
      RANDOMLY SELECT NM DISTINCT NUMBERS FROM U(1,N) TO BE USED AS
      MISSING VALUE SAMPLE NUMBER
      NM IS THE NUMBER OF MISSING VALUES
      N IS SAMPLE SIZE

    8 DO 9 I=1,N
      IT(I) = I
    9 CONTINUE
      DO 10 I=1,NM
      CALL RANDU(IX,IY,YFL)
      IX = IY
      NI = N - I + 1
      K = NI*YFL + 1.0
      M(I) = IT(K)
      IT(K) = IT(NI)
   10 CONTINUE

      ORDER THE NM RANDOM NUMBER IN INCREASING ORDER

      DO 11 I=1,NMM1
      IP1 = I + 1
      DO 11 J=IP1,NM
      IF(M(I).LT.M(J))GO TO 11
      ITEMP = M(I)
      M(I) = M(J)
      M(J) = ITEMP
   11 CONTINUE

      OBTAIN BLOCK AND TREATMENT NUMBERS FOR MISSING VALUE
      PLACE BLOCK AND TREATMENT NUMBERS IN VECTORS MJ AND MI
      RESPECTIVELY

      DO 12 I=1,NM
      J = M(I)
      J = J - 1
      MJ(I) = J/NT + 1
      MI(I) = MOD(J,NT) + 1
   12 CONTINUE

      CHECK FOR COMPLETE MISSING BLOCK OR TREATMENT

      DO 15 I=1,NMM1
      IC = 1
      JC = 1
      IP1 = I + 1
      DO 13 J=IP1,NM
      IF(MI(I).EQ.MI(J))IC=IC+1
      IF(MJ(I).EQ.MJ(J))JC=JC+1
```

```
13 CONTINUE
   IF(IC.LT.NB)GO TO 14
   GO TO 8
14 IF(JC.LT.NT)GO TO 15
   GO TO 8
15 CONTINUE

      SET THE TRUE VALUE OF THE MISSING UNIT EQUAL TO ZERO

   DO 16 I=1,NM
   J = M(I)
   KK = I + NM*(JJ - 1)
   MV(KK) = J
   X(J) = 0.0
16 CONTINUE

      ESTIMATE VALUE FOR THE MISSING OBSERVATION USING BIGGER
      OBTAIN CORRECTED ANOVA USING CANOVA
      STORE F VALUE IN VECTOR DF

   CALL BIGGER(X,NB,NT,NM,MJ,MI,E,B,T,Q,A,LW,MW)
   CALL CANOVA(X,NB,NT,NM,M,E,S,F,B,T)
   DF(JJ) = F
17 CONTINUE
18 CONTINUE
   CALL EXIT
   END
```

```
      SUBROUTINE ANOVA

   PURPOSE
      TO PREFORM AN ANALYSIS OF VARIANCE ON A RANDOMIZED BLOCK DESIGN

   DESCRIPTION OF PARAMETERS
      X  - INPUT VECTOR OF LENGTH NB*NT CONTAINING DATA TO BE
           ANALYSED
      NB - NUMBER OF BLOCKS
      NT - NUMBER OF TREATMENTS
      S  - OUTPUT VECTOR OF LENGTH 7 CONTAINING THE FOLLOWING
           INFORMATION:
           S(1) - SUM OF SQUARES DUE TO TREATMENTS
           S(2) - TREATMENT MEAN SQUARE
           S(3) - SUM OF SQUARES DUE TO BLOCKS
           S(4) - BLOCK MEAN SQUARE
           S(5) - SUM OF SQUARES DUE TO ERROR
           S(6) - ERROR MEAN SQUARE
           S(7) - TOTAL SUM OF SQUARES
      F  - F VALUE ASSOCIATED WITH TREATMENT EFFECT
      B  - VECTOR OF LENGTH NB CONTAINING BLOCK SUMS
      T  - VECTOR OF LENGTH NT CONTAINING TREATMENT SUMS
      G  - GRAND SUM

   INPUT PARAMETERS
      VECTOE X, NB, NT

   OUTPUT PARAMETERS
      VECTOR S, F

      SUBROUTINE ANOVA(X,NB,NT,S,F,B,T,G)
      DIMENSION X(1), S(1), B(1), T(1)
      G = 0.0
      DO 1 I=1,NT
      T(I) = 0.0
    1 CONTINUE
      DO 2 I=1,7
      S(I) = 0.0
    2 CONTINUE
      DO 3 J=1,NB
      B(J) = 0.0
      DO 3 I=1,NT
      IJ = I + (J - 1)*NT
      Z = X(IJ)
      B(J) = B(J) + Z
      T(I) = T(I) + Z
      G = G + Z
```

```
      S(7) = S(7) + Z*Z
    3 CONTINUE
      Z = NB*NT
      G = G*G/Z
      S(7) = S(7) - G
      DO 4 I=1,NT
      S(1) = S(1) + T(I)*T(I)
    4 CONTINUE
      S(1) = S(1)/NB - G
      Z = NT - 1
      S(2) = S(1)/Z
      DO 5 I=1,NB
      S(3) = S(3) + B(I)*B(I)
    5 CONTINUE
      S(3) = S(3)/NT - G
      G = NB - 1
      S(4) = S(3)/G
      S(5) = S(7) - S(1) - S(3)
      G = Z*G
      S(6) = S(5)/G
      F = S(2)/S(6)
      RETURN
      END
```

```
      OUTPUT PARAMETERS
         VECTOR S, F

      SUBROUTINE ANCOVA(X,Y,NDF1,NDF2,NDF3,B,S,F,BX,BY,TX,TY,GX,GY)
      DIMENSION X(1), Y(1), BX(1), BY(1), TX(1), TY(1), S(1)
      GY = 0.0
      GX = 0.0
      DO 1 I=1,15
      S(I) = 0.0
    1 CONTINUE
      DO 2 J=1,NDF3
      TY(J) = 0.0
      TX(J) = 0.0
    2 CONTINUE
      DO 3 I=1,NDF2
      BY(I) = 0.0
      BX(I) = 0.0
      DO 3 J=1,NDF3
      K = J + NDF3*(I - 1)
      YI = Y(K)
      XI = X(K)
      TY(J) = TY(J) + YI
      TX(J) = TX(J) + XI
      BY(I) = BY(I) + YI
      BX(I) = BX(I) + XI
      GY = GY + YI
      GX = GX + XI
      S(1) = S(1) + YI*YI
      S(2) = S(2) + YI*XI
      S(3) = S(3) + XI*XI
    3 CONTINUE
      CX = GX*GX/NDF1
      B = GY*GX/NDF1
      CY = GY*GY/NDF1
      S(1) = S(1) - CY
      S(2) = S(2) - B
      S(3) = S(3) - CX
      DO 4 I=1,NDF2
      YI = BY(I)
      XI = BX(I)
      S(4) = S(4) + YI*YI
      S(5) = S(5) + YI*XI
      S(6) = S(6) + XI*XI
    4 CONTINUE
      S(4) = S(4)/NDF3 - CY
      S(5) = S(5)/NDF3 - B
      S(6) = S(6)/NDF3 - CX
```

```
      SUBROUTINE ANCOVA

   PURPOSE
      TO PREFORM AN ANALYSIS OF COVARIANCE ON A RANDOMIZED BLOCK
      DESIGN

   DESCRIPTION OF PARAMETERS
      X    - INPUT VECTOR OF LENGTH NDF1 CONTAINING X-VARIATE
      Y    - INPUT VECTOR OF LENGTH NDF1 CONTAINING Y-VARIATE
      NDF1 - NUMBER OF OBSERVATIONS
      NDF2 - NUMBER OF BLOCKS
      NDF2 - NUMBER OF TREATMENTS
      B    - REGRESSION COEFFICIENT
      S    - OUTPUT VECTOR OF LENGTH 20 CONTAINING THE FOLLOWING
             INFORMATION:
             S(1)  - YY-SUM OF PRODUCTS
             S(2)  - XY-SUM OF PRODUCTS
             S(3)  - XX-SUM OF PRODUCTS
             S(4)  - YY-SUM OF PRODUCTS DUE TO BLOCKS
             S(5)  - XY-SUM OF PRODUCTS DUE TO BLOCKS
             S(6)  - XX-SUM OF PRODUCTS DUE TO BLOCKS
             S(7)  - YY-SUM OF PRODUCTS DUE TO TREATMENTS
             S(8)  - XY-SUM OF PRODUCTS DUE TO TREATMENTS
             S(9)  - XX-SUM OF PRODUCTS DUE TO TREATMENTS
             S(10) - YY-SUM OF PRODUCTS DUE TO ERROR
             S(11) - XY-SUM OF PRODUCTS DUE TO ERROR
             S(12) - XX-SUM OF PRODUCTS DUE TO ERROR
             S(13) - YY-SUM OF PRODUCTS DUE TO TREATMENT + ERROR
             S(14) - XY-SUM OF PRODUCTS DUE TO TREATMENT + ERROR
             S(15) - XX-SUM OF PRODUCTS DUE TO TREATMENT + ERROR
             S(16) - SUM OF SQUARES X ADJUSTED FOR Y DUE TO ERROR
             S(17) - SUM OF SQUARES X ADJUSTED FOR Y DUE TO ERROR +
                     TREATMENTS
             S(18) - SUM OF SQUARES X ADJUSTED FOR Y DUE TO TREATMENTS
             S(19) - ERROR MEAN SQUARE
             S(20) - TREATMENT MEAN SQUARE
     F     - F VALUE ASSOCIATED WITH ADJUSTED TREATMENT EFFECT
     BX    - VECTOR OF LENGTH NDF2 CONTAINING X-BLOCK SUMS
     BY    - VECTOR OF LENGTH NDF2 CONTAINING Y-BLOCK SUMS
     TX    - VECTOR OF LENGTH NDF3 CONTAINING X-TREATMENT SUMS
     TY    - VECTOR OF LENGTH NDF3 CONTAINING Y-TREATMENT SUMS
     GX    - X-GRAND SUM
     GY    - Y-GRAND SUM

  INPUT PARAMETERS
     VECTOR X, VECTOR Y, NDF1, NDF2, NDF3
```

```
      DO 5 I=1,NDF3
      YI = TY(I)
      XI = TX(I)
      S(7) = S(7) + YI*YI
      S(8) = S(8) + YI*XI
      S(9) = S(9) + XI*XI
    5 CONTINUE
      S(7) = S(7)/NDF2 - CY
      S(8) = S(8)/NDF2 - B
      S(9) = S(9)/NDF2 - CX
      S(10) = S(1) - S(4) - S(7)
      S(11) = S(2) - S(5) - S(8)
      S(12) = S(3) - S(6) - S(9)
      S(13) = S(10) + S(7)
      S(14) = S(11) + S(8)
      S(15) = S(12) + S(9)
      B = S(11)/S(10)
      S(16) = S(12) - S(11)*S(11)/S(10)
      S(17) = S(15) - S(14)*S(14)/S(13)
      S(18) = S(17) - S(16)
      NDF2 = NDF2 - 1
      NDF3 = NDF3 - 1
      NDF1 = NDF2*NDF3 - 1
      NDF2 = NDF1 + NDF3
      S(19) = S(16)/NDF1
      S(20) = S(18)/NDF3
      F = S(20)/S(19)
      RETURN
      END
```

```
      SUBROUTINE BIGGER

   PURPOSE
      TO ESTIMATE THE VALUES FOR NM MISSING OBSERVATIONS IN A
      RANDOMIZED BLOCK DESIGN

   DESCRIPTION OF PARAMETERS
      X   - INPUT VECTOR OF LENGTH NB*NT CONTAINING DATA TO BE
            ANALYSED
      NB  - NUMBER OF BLOCKS
      NT  - NUMBER OF TREATMENTS
      NM  - NUMBER OF MISSING OBSERVATIONS
      MJ  - INPUT VECTOR OF LENGTH NM CONTAINING MISSING VALUE BLOCK
            NUMBERS
      MI  - INPUT VECTOR OF LENGTH NM CONTAINING MISSING VALUE
            TREATMENT NUMBERS
      E   - OUTPUT VECTOR OF LENGTH NM CONTAINING THE ESTIMATES
      B   - VECTOR OF LENGTH NB CONTAINING BLOCK SUMS
      T   - VECTOR OF LENGTN NT CONTAINING TREATMENT SUMS
      Q   - PRODUCT VECTOR OF LENGTH NM
      A   - ASSOCIATION MATRIX OF ORDER NM, IN VECTOR FORM AND LATER
            ITS INVERSE
      LW  - WORKING VECTOR OF LENGTH NM
      MW  - WORKING VECTOR OF LENGTH NM

   INPUT PARAMETERS
      VECTOR X, NB, NT, NM, VECTOR MJ, VECTOR MI

   OUTPUT PARAMETERS
      VECTOR E

   REMARKS
      NM MUST BE GREATER THAN ONE
      THIS METHOD WAS PRESENTED IN A PAPER BY J. D. BIGGERS

   SUBROUTINES USED
      MINV

   SUBROUTINE BIGGER(X,NB,NT,NM,MJ,MI,E,B,T,Q,A,LW,MW)
   DIMENSION X(1), E(1), Q(1), A(1), MI(1), MJ(1), LW(1), MW(1), B(1)
  1, T(1)
   G = 0.0
   JAS = 1 - NB
   IAS = 1 - NT
   IJAS = IAS*JAS
   DO 2 I=1,NT
   T(I) = 0.0
```

```
 2 CONTINUE
   DO 3 J=1,NB
   B(J) = 0.0
   DO 3 I=1,NT
   IJ = I + NT*(J - 1)
   Z = X(IJ)
   G = G + Z
   B(J) = B(J) + Z
   T(I) = T(I) + Z
 3 CONTINUE
   DO 4 K=1,NM
   I = MI(K)
   J = MJ(K)
   Q(K) = NB*B(J) + NT*T(I) - G
   KK = NM*(K - 1) + K
   A(KK) = IJAS
   IF(K.EQ.NM)GO TO 5
   KM1 = K - 1
   KP1 = K + 1
   DO 4 KKK=KP1,NM
   IC = MI(KKK)
   JC = MJ(KKK)
   KK1 = NM*KM1 + KKK
   KK2 = NM*(KKK - 1) + K
   Z = 1
   IF(I.EQ.IC)Z=IAS
   IF(J.EQ.JC)Z=JAS
   A(KK1) = Z
   A(KK2) = Z
 4 CONTINUE
 5 N = NM*NM
   CALL MINV(A,NM,Z,LW,MW)
   DO 6 I=1,NM
   E(I) = 0.0
   DO 6 J=1,NM
   IJ = NM*(I-1) + J
   E(I) = E(I) + A(IJ)*Q(J)
 6 CONTINUE
   RETURN
   END
```

```
      SUBROUTINE CANOVA

   PURPOSE
      TO PREFORM AN ANALYSIS OF VARIANCE ON A RAMDOMIZED BLOCK DESIGN
      WHERE NM OBSERVATIONS ARE MISSING AND ESTIMATES HAVE BEEN
      CALCULATED FOR THESE MISSING VALUES

   DESCRIPTION OF PARAMETERS
      X  - INPUT VECTOR OF LENGTH NB*NT CONTAINING DATA TO BE
           ANALYSED
      NB - NUMBER OF BLOCKS
      NT - NUMBER OF TREATMENTS
      NM - NUMBER OF MISSING OBSERVATIONS
      M  - INPUT VECTOR OF LENGHT NM CONTAINING MISSING VALUE
           OBSERVATION NUMBERS
      E  - INPUT VECTOR OF LENGTH NM CONTAINING ESTIMATES OF
           MISSING VALUES
      S  - OUTPUT VECTOR OF LENGTH 7 CONTAINING THE FOLLOWING
           INFORMATION:
           S(1) - ADJUSTED TREATMENT SUM OF SQUARES
           S(2) - TREATMENT MEAN SQUARE
           S(3) - BLOCK SUM OF SQUARES
           S(4) - BLOCK MEAN SQUARE
           S(5) - ERROR SUM OF SQUARES
           S(6) - ERROR MEAN SQUARE
           S(7) - TOTAL SUM OF SQUARES
      F  - F VALUE ASSOCIATED WITH TREATMENT EFFECT
      B  - VECTOR OF LENGTH NB CONTAINING BLOCK SUMS
      T  - VECTOR OF LENGTH NT CONTAINING TREATMENT SUMS

   INPUT PARAMETERS
      VECTOR X, NB, NT, NM, VECTOR M, VECTOR E

   OUTPUT PARAMETERS
      VECTOR S, F

   SUBROUTINE CANOVA(X,NB,NT,NM,M,E,S,F,B,T)
   DIMENSION X(1), M(1), E(1), S(1), B(1), T(1)
   N = NT*NB
   L = N - 1
   LT = NT -1
   LB = NB - 1
   LE = LT*LB - NM
   S(1) = 0.0
   S(2) = 0.0
   DO 1 I=1,NB
   B(I) = 0.0
```

```
      DO 1 J=1,NT
      IJ = J + NT*(I - 1)
      Z = X(IJ)
      S(1) = S(1) + Z*Z
      B(I) = B(I) + Z
    1 CONTINUE
      DO 3 I=1,NB
      D = NT
      DO 2 J=1,NM
      MV = (M(J) - 1)/NT + 1
      IF(I.EQ.MV)D=D-1.0
    2 CONTINUE
      Z =B(I)/D
      S(2) = S(2) + D*Z*Z
    3 CONTINUE
      C = S(1) - S(2)
      DO 4 I=1,NM
      MV = M(I)
      X(MV) = E(I)
    4 CONTINUE
      DO 5 I=1,7
      S(I) = 0.0
    5 CONTINUE
      G = 0.0
      DO 9 J=1,NT
      T(J) = 0.0
    9 CONTINUE
      DO 6 I=1,NB
      B(I) = 0.0
      DO 6 J=1,NT
      IJ = J + NT*(I - 1)
      Z = X(IJ)
      G = G + Z
      S(7) = S(7) + Z*Z
      T(J) = T(J) + Z
      B(I) = B(I) + Z
    6 CONTINUE
      G = G*G/N
      S(7) = S(7) - G
      DO 7 I=1,NT
      Z = T(I)
      S(1) = S(1) + Z*Z
    7 CONTINUE
      S(1) = S(1)/NB - G
      DO 8 I=1,NB
      Z = B(I)
      S(3) = S(3) + Z*Z
    8 CONTINUE
```

```
S(3) = S(3)/NT - G
S(4) = S(3)/LB
S(5) = S(7) - S(1) - S(3)
S(1) = C - S(5)
S(2) = S(1)/LT
S(6) = S(5)/LE
S(7) = S(1) + S(3) + S(5)
F = S(2)/S(6)
RETURN
END
```

APPENDIX B

TABLES

TABLE I

POWER OF THE F-TEST - 100 REPLICATIONS

NUMBER MISSING	RANDOM INITIAL VALUE										THEORETICAL POWER†
	13161		35065		46173		59575		65539		
	ANOVA	CANOVA	ANOVA	CANOVA	ANOVA	CANOVA	ANOVA	CANOVA	ANOVA	CANOVA	0.3679*
1	0.32	0.26	0.28	0.26	0.32	0.30	0.31	0.34	0.41	0.32	0.3518
2	0.28	0.24	0.33	0.27	0.28	0.16	0.46	0.30	0.34	0.22	0.3323
3	0.31	0.23	0.38	0.27	0.35	0.19	0.36	0.22	0.42	0.24	0.3015
4	0.42	0.20	0.37	0.17	0.42	0.20	0.44	0.21	0.38	0.21	0.2706
5	0.34	0.11	0.37	0.13	0.42	0.22	0.29	0.08	0.32	0.15	0.2175
6	0.31	0.08	0.29	0.07	0.32	0.06	0.34	0.09	0.31	0.05	0.1644

† Evaluated using Tiku (1967).

* Power in the case of no missing observations, ie. ANOVA F-test.

TABLE I (CONTINUED)

POWER OF THE F-TEST - 100 REPLICATIONS

NUMBER MISSING	RANDOM INITIAL VALUE										THEORETICAL POWER†
	69725		73441		76333		84765		91841		
	ANOVA	CANOVA	ANOVA	CANOVA	ANOVA	CANOVA	ANOVA	CANOVA	ANOVA	CANOVA	0.3679*
1	0.40	0.33	0.44	0.37	0.42	0.36	0.27	0.30	0.49	0.37	0.3518
2	0.44	0.26	0.33	0.22	0.26	0.19	0.33	0.21	0.41	0.32	0.3323
3	0.34	0.19	0.33	0.19	0.44	0.27	0.32	0.20	0.37	0.17	0.3015
4	0.41	0.21	0.35	0.18	0.37	0.20	0.34	0.12	0.39	0.13	0.2706
5	0.40	0.21	0.31	0.09	0.36	0.13	0.36	0.15	0.36	0.08	0.2175
6	0.36	0.16	0.27	0.08	0.33	0.10	0.40	0.06	0.44	0.06	0.1644

† Evaluated using Tiku (1967).

* Power in the case of no missing observations, ie. ANOVA F-test.

TABLE II

POWER OF THE F-TEST - 250 REPLICATIONS

NUMBER MISSING	RANDOM INITIAL VALUE										THEORETICAL POWER†
	13163		35067		46175		59577		65541		
	ANOVA	CANOVA	ANOVA	CANOVA	ANOVA	CANOVA	ANOVA	CANOVA	ANOVA	CANOVA	0.3679*
1	0.348	0.276	0.368	0.304	0.312	0.292	0.424	0.348	0.344	0.288	0.3518
2	0.360	0.300	0.360	0.256	0.412	0.300	0.320	0.268	0.376	0.288	0.3323
3	0.368	0.204	0.332	0.192	0.384	0.208	0.356	0.208	0.352	0.213	0.3015
4	0.360	0.196	0.356	0.172	0.300	0.172	0.328	0.176	0.400	0.184	0.2706
5	0.360	0.088	0.348	0.112	0.356	0.160	0.288	0.100	0.352	0.122	0.2175
6	0.372	0.088	0.356	0.092	0.368	0.088	0.364	0.068	0.336	0.060	0.1644

† Evaluated using Tiku (1967).

* Power in the case of no missing observations, ie. ANOVA F-test.

TABLE II (CONTINUED)

POWER OF THE F-TEST - 250 REPLICATIONS

NUMBER MISSING	RANDOM INITIAL VALUE										THEORETICAL POWER†
	69727		73443		76335		84767		91843		
	ANOVA	CANOVA	ANOVA	CANOVA	ANOVA	CANOVA	ANOVA	CANOVA	ANOVA	CANOVA	0.3679*
1	0.364	0.312	0.344	0.284	0.372	0.312	0.336	0.276	0.316	0.300	0.3518
2	0.404	0.260	0.348	0.276	0.332	0.240	0.340	0.272	0.364	0.252	0.3323
3	0.328	0.200	0.356	0.156	0.336	0.232	0.352	0.204	0.372	0.248	0.3015
4	0.348	0.144	0.344	0.160	0.368	0.156	0.384	0.156	0.416	0.136	0.2706
5	0.288	0.128	0.380	0.160	0.380	0.156	0.316	0.120	0.324	0.152	0.2175
6	0.356	0.136	0.372	0.128	0.368	0.112	0.380	0.088	0.376	0.076	0.1644

† Evaluated using Tiku (1967).

* Power in the case of no missing observations, ie. ANOVA F-test.

TABLE III

POWER OF THE F-TEST - 500 REPLICATIONS

NUMBER MISSING	RANDOM INITIAL VALUE										THEORETICAL POWER†
	13165		35069		46177		59579		65543		
	ANOVA	CANOVA	ANOVA	CANOVA	ANOVA	CANOVA	ANOVA	CANOVA	ANOVA	CANOVA	0.3679*
1	0.346	0.300	0.374	0.318	0.356	0.312	0.378	0.332	0.408	0.346	0.3518
2	0.348	0.264	0.358	0.282	0.360	0.268	0.386	0.280	0.378	0.268	0.3323
3	0.358	0.234	0.340	0.206	0.316	0.222	0.388	0.264	0.372	0.232	0.3015
4	0.368	0.196	0.340	0.162	0.390	0.174	0.390	0.186	0.370	0.154	0.2706
5	0.356	0.120	0.354	0.118	0.344	0.120	0.338	0.136	0.340	0.124	0.2175
6	0.354	0.090	0.352	0.078	0.332	0.104	0.366	0.092	0.356	0.098	0.1644

† Evaluated using Tiku (1967).

* Power in the case of no missing observations, ie. ANOVA F-test.

TABLE III (CONTINUED)

POWER OF THE F-TEST - 500 REPLICATIONS

NUMBER MISSING	RANDOM INITIAL VALUE										THEORETICAL POWER†
	69729		73445		76337		84769		91845		
	ANOVA	CANOVA	ANOVA	CANOVA	ANOVA	CANOVA	ANOVA	CANOVA	ANOVA	CANOVA	0.3679*
1	0.372	0.320	0.394	0.348	0.336	0.278	0.382	0.324	0.332	0.314	0.3518
2	0.340	0.242	0.358	0.256	0.362	0.248	0.358	0.268	0.340	0.262	0.3323
3	0.368	0.212	0.358	0.212	0.328	0.196	0.366	0.262	0.368	0.230	0.3015
4	0.362	0.158	0.366	0.166	0.342	0.178	0.346	0.182	0.356	0.164	0.2706
5	0.338	0.136	0.334	0.130	0.364	0.130	0.370	0.128	0.340	0.126	0.2175
6	0.334	0.068	0.390	0.070	0.346	0.088	0.402	0.094	0.348	0.098	0.1644

† Evaluated using Tiku (1967).

* Power in the case of no missing observations, ie. ANOVA F-test.

BIBLIOGRAPHY

Allan, F. E. and Wishart, J., (1930) "A Method of Estimating Yield of a Missing Plot in Field Experimental Work," Journal of Agricultural Science, 20:399-406.

Biggers, J. D., (1959) "The Estimation of Missing and Mixed-Up Observations in Several Experimental Designs," Biometrika, 46:91-105.

Esseen, Carl-Gustav, (1945) "Fourier Analysis of Distribution Functions. A Mathematical Study of the Laplace-Gaussian Law," Acta Mathematica, 77:71.

Fox, M., (1956) "Charts of the Power of the F-test," Annals of Mathematical Statistics, 27:484-497.

Hogg, R. V. and Craig, A. T., (1965) Introduction to Mathematical Statistics. New York (Macmillan & Co.).

Kenney, J. F. and Keeping, E. S., (1951) Mathematics of Statistics, Part Two, New York (D. Van Nostrand Co. Ltd.).

Pearson, E. S. and Hartley, H. O., (1951) "Charts of the Power Function for Analysis of Variance Tests, Derived from the Non-Central F-Distribution," Biometrika, 38:112-130.

Scheffé, H., (1959) The Analysis of Variance, New York (John Wiley).

Steel, R. G. D. and Torrie, J. H., (1960) Principles and Procedures of Statistics, New York (McGraw-Hill Book Co., Inc.).

Tang, P. C., (1938) "The Power Function of the Analysis of Variance Tests with Tables and Illustrations of Their Use," Statistical Research Memoirs, 2:126-149 + Tables.

Tiku, M. L., (1967) "Tables of the Power of the F-test," Journal of the American Statistical Association, 62:525-539.

Yates, F., (1933) "The Analysis of Replicated Experiments When the Field Results are Incomplete," Empire Journal in Experimental Agriculture, 1:129-142