MISSING PLOT TECHNIQUES

by

CHING-LAN WU

B. A., National Taiwan University, 1969

A MASTER'S REPORT

submitted in partial fulfillment of the

requirements for the degree

MASTER OF SCIENCE

Department of Statistics

KANSAS STATE UNIVERSITY
Manhattan, Kansas

1973

Approved by:

Major Professor

THIS BOOK CONTAINS NUMEROUS PAGES WITH THE ORIGINAL PRINTING ON THE PAGE BEING CROOKED.

THIS IS THE BEST IMAGE AVAILABLE.

TABLE OF CONTENTS

I. INTRODUCTION

The occurrence of missing data in a statistically designed experiment requires some modification of the usual statistical techniques because the orthogonality, or balance, of the design is destroyed.

One of the first papers on the subject of estimating the yield of a missing unit in the field experimental work is stated by Anderson (1) to have been published by Allan and Wishart. They derived formulae and illustrated their uses for a single missing plot in a randomized block and in a latin square experiment. These estimation methods were expanded by Yates (16) to cover several missing units in a given experiment. The procedure he used is to minimize the error variance obtained when unknowns are substituted for the missing yields. The formula given by Yates for estimating the yield of a single missing unit in a randomized complete block experiment is

$$y = \frac{rB + tT - G}{(r - 1)\ (t - 1)}$$

where r = the number of blocks, t = the number of treatments in the experiment, B = the total yield of the remaining units in the block where the missing unit appears, T = the total of the yields of the treatment with missing unit, and G = the grand total. Similarly for a single missing unit in a latin squares,

$$y = \frac{r(R + C + T) - 2G}{(r - 1)\ (r - 2)}$$

where r = the number of rows, columns, or treatments; and R, C, and T represent the total yields of the remaining units in the row, column and treatment containing the missing unit. G is the grand total. He used these formulae for several missing units by means of iterative methods.

Yates also showed that in a complete analysis of variance of the augmented data, the treatment sum of squares is overestimated but may be corrected by subtracting a correction for bias. The formulas for correction for bias in two designs with one missing unit are given. For a randomized block experiment, the correction for bias is

$$\frac{\left[B - (t - 1)y\right]^2}{t(t - 1)} ,$$

which is subtracted from the treatment sum of squares.

The correction for bias in a latin square experiment is to subtract

$$\frac{\left[G - R - C - (r - 1)T\right]^2}{(r - 1)^2(r - 2)^2}$$

from the treatment sum of squares.

Anderson (1) derived some formulas for missing plots in split-plot experiments by minimizing the error variance. His covariance methods are used in the derivations which follow in order to furnish an easy means for correcting the bias in the treatment sum of squares, and of estimating the missing data. We assume that we have a split-plot experiment with r replications, a whole-plot, and b sub-plot treatments so that the total number of units is N = rab. Let the single missing sub-unit be that for the whole-plot treatment a_1, sub-plot treatment b_1 and replication r_1. Also let A_1 be the total yield of all existing units with treatment a_1, B_1 the total yield of all existing units with treatment b_1, R_1 the total yield in replication r_1, (A_1B_1) the total yield of all existing units with both a_1 and b_1, (R_1A_1) the total yield of all existing units with both r_1 and a_1 and G the grand total. Set x=0 and y= the actual yield for the existing units and x=-1 and y=0 for the missing unit. In the analysis of covariance table, any sum of squares, $S(x^2)$, equals its degrees of freedom divided by N, in all cases. The best estimate of the yield of the missing unit in order to minimize the sub-plot error mean square is simply the error b) regression coefficient,

$$y = \frac{r(R_1A_1) + b(A_1B_1) - A_1}{(r - 1)(b - 1)} .$$

As this yield is used for the missing unit, all sums of squares except that for error b) are over-estimated. The unbiased estimate of any sum of squares is found by computing a new $S(x_1^2)$ and $S(x_1y)$. Where $S(x_1^2)$ and $S(x_1y)$ are the $S(x^2)$ and $S(xy)$ plus error b) respectively. Then the new regression coefficient is

$$y_1 = \frac{S(x_1y)}{S(x_1^2)} .$$

The bias in estimating the sum of squares under consideration is: $(y - y_1)^2 S(x_1^2)$. The bias is always positive; that is, the sum of squares is always over-estimated in the analysis of variance.

Thus, for the treatment B,

$$y_1 = \frac{ra(R_1A_1) + ab(A_1B_1) - aA_1 - bB_1 + G}{(b - 1)(ra - a + 1)} ,$$

and $S(x_1^2) = (b - 1)(ra - a - 1)/rab$.

For the interation AB,

$$y_1 = \frac{ra(R_1A_1) + bB_1 - G}{(ra - 1)(b - 1)} ,$$

and $S(x_1^2) = (ra - 1)(b - 1)/rab$.