Serdica

Bulgariacae mathematicae publicationes

Сердика

Българско математическо списание

ESTIMATION OF A SINGLE MISSING VALUE WITH FIXED EFFECTS EXPERIMENTAL DESIGNS

SHAWNM ABDUL-KADIR

The paper describes a method of estimating a single missing value using covariance analysis (ANCOVA) with the help of a dummy variable. Thus the estimate found coincides with the one obtained by the minimizing the residual sum of squares.

Introduction. One of the important statistical tasks of the design of experiments is to identify the sources of variation. The investigator develops some methods of estimating the variance components as well as methods of testing some hypotheses concerning them. In most experiments, some variation is imposed by the investigator in form of different treatments. Inferences are desired concerning only the particular treatment used in the experiment. It is a common practice to call such sources of variation fixed effects. In some cases it is assumed that the variability among the sources is inherent to the experiment and thus it represents the variability in a larger population. These sources are usually called random effects.

Carrying out the experiments is often accompanied by spoiling or omitting some of the observations. The statistical analysis of a Latin-Square design, Cross-Over design and Graeco-Latin Square are sometimes complicated by the fact that some observation is missing. The usual method of analysis is based on the assumption that the design is balanced, but the missing value destroys this balance property.

Most of the papers treating the missing observation problem for Latin-Square design have been written by the authors mentioned below. Cochran & Cox [3] present a formula for a single missing value in Latin-Square design. Al-Rawl [1] and Kramer & Glass [5] give a formula of estimating several missing values in a Latin-Square design by minimizing the residual sum of squares. Haseman & Gaylor [4] and Rubin [6] introduce a non-iterative procedure for estimating missing values for a $(r\times r)$ Latin-Square design. The procedure involves a direct solution of a system of linear equations. Cochran & Cox [3] present also a formula of estimating a single missing value in a Cross-Over design by minimizing the residual sum of squares.

The present paper describes a procedure of estimating a single missing value in a Graeco-Latin square design by minimizing the residual sum of squares (MSSR). Then we study the single missing value case for a Latin-Square design, a Cross-Over design and a Graeco-Latin square design by using another technique (i. e. covariance analysis with a dummy variable, (ANCOVA method)) and show that the estimator is the same. Several examples are given by illustrating the implementation of these two methods.

Further θ stands for the missing value estimator with indices either MSSR or COVA indicating the method. Also the errors ϵ_{ijk} in the linear models are assumed to be independent and normally distributed with mean 0 and variance σ^2.

 SERDICA Bulgaricae mathematical publications, Vol. 17, 1991, p. 177-184.

1. MSSR method for Graeco-Latin square design. The mathematical model for Graeco-Latin square design is set as follows:

$$Y_{ijkl}=\mu+T_i+R_j+C_k+G_l+\epsilon_{ijkl}, \tag{1.1}$$

for $1\leq i\leq t$, $1\leq j\leq r$, $1\leq k\leq c$, $1\leq l\leq s$ and $t=r=c=s$, $r>3$, where μ is the general mean, T_i, R_j, C_k and G_l indicate treatment, row, column and Greek-Letter effects respectively assuming that their sums are zero and the errors are denoted by ϵ_{ijk}.

Let Y_{ijkl} denote the missing value for the i-th treatment in the j-th row, the k-th column and for the l-th Greek-Letter and usually

$$Y_{ijkl}=0.$$

The marginal sums of squares are given as follows:

$$\begin{aligned}
\mathrm{SSR}&=\frac{1}{r}\sum_j R_j^2+\frac{1}{r}[R_j'+\theta_{ijkl}]^2-\frac{G^2}{r^2};\\
\mathrm{SSC}&=\frac{1}{c}\sum_k C_k^2+\frac{1}{c}[C_k'+\theta_{ijkl}]^2-\frac{G^2}{r^2};\\
\mathrm{SST}&=\frac{1}{t}\sum_i T_i^2+\frac{1}{t}[T_i'+\theta_{ijkl}]^2-\frac{G^2}{r^2};\\
\mathrm{SSGL}&=\frac{1}{s}\sum_l G_l^2+\frac{1}{s}[G_l'+\theta_{ijkl}]^2-\frac{G^2}{r^2};\\
\mathrm{SSG}&=\sum_i\sum_j\sum_k\sum_l Y_{ijkl}^2+\theta_{ijkl}^2-\frac{G^2}{r^2},
\end{aligned}\tag{1.2}$$

and T_i', R_j', C_k', G_l' stand for the i-th treatment, j-th row, k-th column and l-th Greek-Letter totals.

Theorem 1. *For the Graeço-Latin square design, model (1.1), the missing value estimator obtained by minimizing the residual sum of squares is given by*

$$\theta_{ijkl}=\frac{r(T_i'+R_j'+C_k'+G_l')-3G}{(r-1)(r-3)}=\theta_{\mathrm{MSSR}},\quad \text{for } r>3. \tag{1.3}$$

Proof. The residual sum of squares, denoted by SSE, is determined from the equation

$$\mathrm{SSE}=\mathrm{SSG}-\mathrm{SST}-\mathrm{SSR}-\mathrm{SSC}-\mathrm{SSGL}.$$

Substituting SSG, SST, SSR, SSC and SSGL by (1.2), we obtain that

$$\begin{aligned}
\mathrm{SSE}=&\sum_i\sum_j\sum_k\sum_l Y_{ijkl}^2+\theta_{ijkl}^2-\frac{G^2}{r^2}\\
&-\frac{1}{r}\sum_j R_j^2-\frac{1}{r}[R_j'+\theta_{ijkl}]^2+\frac{G^2}{r^2}\\
&-\frac{1}{c}\sum_k C_k^2-\frac{1}{c}[C_k'+\theta_{ijkl}]^2+\frac{G^2}{r^2}\\
&-\frac{1}{t}\sum_i T_i^2-\frac{1}{t}[T_i'+\theta_{ijkl}]^2+\frac{G^2}{r^2}\\
&-\frac{1}{s}\sum_l G_l^2-\frac{1}{s}[G_l'+\theta_{ijkl}]^2+\frac{G^2}{r^2}.
\end{aligned}$$

The value of θ_{ijkl} which minimizes SSE is found from the equation

$$\frac{\partial\,\text{SEE}}{\partial\,\theta_{ijkl}}=0$$

which results in:

$$\frac{\partial\,\text{SSE}}{\partial\,\theta_{ijkl}}=2\theta_{ijkl}-\frac{2}{r}\,[R'_j+\theta_{ijkl}]-\frac{2}{c}\,[C'_k+\theta_{ijkl}]$$

$$-\frac{2}{t}\,[T'_i+\theta_{ijkl}]-\frac{2}{l}\,[G'_l+\theta_{ijkl}]+\frac{6}{r^2}\,(G'+\theta_{ijkl})=0.$$

Briefly we obtain

$$(r-1)(r-3)\,\theta_{ijkl}=r(T'_i+R'_j+C'_k+G'_l)-3G.$$

Therefore

$$\theta_{ijkl}=\frac{r(T'_i+R'_j+C'_k+G'_l)-3G}{(r-1)(r-3)}=\theta_{\text{MSSR}}\,.$$

Corollary. *Fo, a* (3×3) *Graeco-Latin square design the missing value is always zero, i. e.* $\theta_{ijkl}=0$.

Proof. Since the grand total is

$$G=(T'_i+R'_j+C'_k+G'_l).$$

For $r=3$ by using (1.3), we obtain:

$$\frac{3G-3G}{(r-1)(r-3)}=\frac{0}{0}\,. \tag{1.4}$$

Thus the missing value is zero.

2. **ANCOVA method for fixed effects experimental designs.** The mathematical model for fixed effects experimental designs is, as follows:

$$Y_{ijk}=\mu+T_i+R_j+C_k+\epsilon_{ijk} \tag{2.1}$$

$$1\leqq i\leqq t;\quad 1\leqq j\leqq r;\quad 1\leqq k\leqq c,$$

where μ is the general mean, and T_i, R_j and C_k are respectively the treatment, the row and the column effects. Their sums are zero and the errors ϵ_{ijk} are assumed to be independent and normally distributed with mean 0 and variance σ^2.

Using ANCOVA method for estimating the single missing value, we obtain the following theorem:

Theorem 2. *The estimator for a single missing value obtained by ANCOVA method for the fixed effects experimental designs given by model (2.1) coincides with that obtained by minimizing the residual sum of squares, i. e.* $\theta_{\text{COVA}}=\theta_{\text{MSSR}}$.

Proof. According to ANCOVA method we have:

$$Y_{ijk}=\mu+T_i+R_j+C_k+\beta X_{ijk}+\epsilon_{ijk}, \tag{2.1a}$$

where

$$X_{ijk}=\begin{vmatrix} 0, \text{ if } Y_{ijk} \text{ is observed;} \\ 1 \text{ (or } -1) \text{ if } Y_{ijk} \text{ is missing.}\end{vmatrix}$$

We assume that the single missing value will appear at the i-th treatment in the i-th row and k-th column, i. e. $X_{ijk}=1$. Obviously, considering (2.1a) as a simple regression model, θ_{COVA} is the negative regression coefficient estimator $\widehat{\beta}$, i. e. $\theta_{COVA}=-\widehat{\beta}$.

Next we are going to discuss the three kinds of fixed effects designs one by one. The ANCOVA Table and the corresponding estimator are obtained for each of the three cases by applying Theorem 2.

2.1. Latin-Square design. The mathematical model of a Latin-Squares design is, as follows

$$Y_{ijk}=\mu+T_i+R_j+C_k+\epsilon_{ijk}$$

$$1\leq i\leq t;\quad 1\leq j\leq r;\quad 1\leq k\leq c \quad \text{and} \quad t=r=c,\ r>2.$$

Table 1. *ANCOVA Table*

S. O. V.	dt	XX	XY	YY
Row	$r-1$	$1/r-1/r^2$	$R_j'/r-G'/r^2$	
Column	$c-1$	$1/c-1/r^2$	$C_k'/r-G'/r^2$	
Treatment	$t-1$	$1/t-1/r^2$	$T_i'/r-G'/r^2$	
Error	$(r-1)(r-2)$	$(r-1)(r-2)/r^2$	$-[r(R_j'+C_k'+T_i')-2G]/r^2$	
Total	r^2-1	$1-1/r^2$	$-G'/r^2$	

Then

$$-\widehat{\beta}=\frac{\text{Residual sum of product of } X \text{ and } Y}{\text{Residual sum of squares of } X}$$

and

$$-\widehat{\beta}=\frac{E(XY)}{E(XX)}=\frac{r(T_i'+R_j'+C_k'-2G'}{(r-1)(r-2)}\,.$$

Thus we obtian that $\theta_{COVA}=\theta_{MSSR}$, i. e. θ_{COVA} coincides with the estimator given in [3].

2.2. Cross-Over design. The mathematical model for a Cross-Over design is the following:

$$Y_{ijk}=\mu+T_i+R_j+C_k+\epsilon_{ijk}$$

for $i, j=1, 2, \dots, t$, $k=1, 2, \dots, c$, and $c\geq t$.

Therefore

$$-\widehat{\beta}=\frac{t(R_j'+T_i')+rC_k-2G'}{(r-1)(r-2)} \quad \text{for } r>2.$$

Thus we have that $\theta_{COVA}=\theta_{MSSR}$, i. e. θ_{COVA} coincides with the estimator given in [3].

2.3. Graeko-Latin square design. The mathematical model of a Graeco-Latin squares design is given by (1.1).

Table 2. *ANCOVA Table*

S. O. V.	df	XX	XY	YY
Row	$r-1$	$1/r-1/r^2$	$R'_j/r-G'/r^2$	
Column	$c-1$	$1/t-1/r^2$	$C'_k/r-G'/r^2$	
Treatment	$t-1$	$1/r-1/r^2$	$T'_i/r-G'/r^2$	
Error	$(r-1)\ (r-2)$	$(r-1)(r-2)/tr$	$-[t\,(R'_j+T'_i)+rC'_k-2G']/tr$	
Total	$tr-1$	$1-1/tr$	$-G'/tr$	

Table 3. *ANCOVA Table*

S. O. V.	df	XX	XY	YY
Row	$r-1$	$1/r-1/r^2$	$R'_j/r-G'/r^2$	
Column	$c-1$	$1/c-1/r^2$	$C'_k/r-G'/r^2$	
Treatment	$t-1$	$1/t-1/r^2$	$T'_i/r-G'/r^2$	
Greek-Letter	$t-1$	$1/l-1/r^2$	$G_i/r-G'/r^2$	
Error	$(r-1)\ (t-3)$	$(r-1)(r-3)/r^2$	$-[r\,(R'_j+C'_k+T'_i+G'_i)-2G']/r^2$	
Total	r^2-1	$1-1/r^2$	$-G'/r^2$	

Hence

$$-\widehat{\beta}=\frac{r(R'_j+C'_k+T'_i+G'_l)-3G'}{(r-1)(r-3)}\quad \text{for} \quad r>3,$$

which coincides with the estimator given in (1.3).

Thus $\theta_{\mathrm{COVA}}=\theta_{\mathrm{MSSR}}$ for all the three models under investigation. Q. E. D.

3. Examples. No reference for a missing value estimator in case of a Graeco-Latin Squares design is known to the author. The ANOVA method avoids the direct calculation of the bias in the treatment sum of squares arising from the analysis of augmented data (i. e. including estimates of missing values) by the standard MSSR method application. The approximate test of significance for the treatment effects is carried out on the augmented data with degrees of freedom for both total and error sums reduced by one.

3.1. MSSR method for Graeco-Latin square design.

Example 1 (see [2, p. 315]). Let us have a (5×5) Graeco-Latin square for estimating the response of Hylotrupes larvae to four nutrients, each at five equally spaced log concentrations; y = average log weight after 103 ± 5 days on the experimental diet; rows are yeast extract, columns are poptone levels, treatments are cholesterol levels and Greek-Letters are riboflavin levels.

Let observion Ar in column 4 and row 5 be missing. Then from the data $R'_5=2.96$, $C'_4=3.34$, $T'_3=1.77$, $G'_1=2.34$, $G'=13.76$ are calculated and we replace them in (1.3). We obtain that $\theta_{ijkl}=1.34625$. Hence $\theta_{\text{MSSR}}=1.35$. The estimator replaces the missing value in the table and then the sums of squares are calculated by taking into account that $\frac{G^2}{r^2}$ is denoted by cF. Thus we have:

Correction $=cF=9.1204$,

Row SS $=\frac{1}{r}\ \Sigma\ R_j^2-cF=0.4546$,

Column SS $=\frac{1}{c}\ \Sigma\ C_k^2-cF=3.221$,

Treatment SS $=\frac{1}{t}\ \Sigma\ T_i^2-cF=0.058$,

Greek-Letter SS $=\frac{1}{s}\ \Sigma\ G_l^2-cF=0.170$,

Total SS $=\Sigma\ Y_{ijk}^2-cF=3.919$,

Error SS = Total SS — Row SS — Column SS — Treatment SS — Greek-Letter SS = 0.0154.

Next by using MSSR method we obtain:

Table 4. *Approximate ANOVA Table*

S. O. V.	df	SS	MS	F
Row	4	0.4546	0.11	
Column	4	3.221	0.805	
Treatment	4	0.058	0.0145	6.59*
Greek-Letter	4	0.170	0.042	19.09**
Error	7	0.0154	0.0022	
Total	23			

The approximate test indicates significance at 5% level, and the theoretical table value is $F_{4,7(0.05)}=4.12$. We have denoted by * the significant value and by ** the high significant value. Thus one may choose another significance level taking into consideration that $F_{(4,\ 7)}(0.01)=7.85$, $F_{(4,\ 7)}(0.005)=10.05$, $F_{(4,\ 7)}(0.001)=17.19$, and finally we have $F_{(4,\ 7)}(0.0005)=21.4$.

Example 2. Let us consider an industrial experiment when the three processes are compared and one makes a guess at systematic changes in external conditions from day to day and also between different times of day. Suppose that the observations are done by three observers A, B, C and that each experimental unit is to be measured by an observer. As a further classification we consider these three observers, requiring that each observer is measuring once for each treatment, once a day and once at each time of day. These requirements are satisfied when the observations form a Graeco-Latin square with row parameter equal to 3, i. e. $t=r=c=s=3$.

Let us suppose that the value in row 3, column 3, treatment B and Greek-Letter α is missing. Then $R_3'=5$, $C_3'=6$, $T_2'=6$, $G_1'=7, G'=24$ and

Table 5

	Time			Total
Day	Aα	Bβ	Cτ	
	2	3	2	7
	Bτ	Cα	Aβ	
	3	5	4	12
	Cβ	Aτ	Bα	
	1	4	—	5
Total	6	12	6	21

$$\theta_{ijkl}=\frac{3(7+6+5+6)-3\times 24}{2\times 0}=0 \quad \text{(see Corollary).}$$

3.2. ANCOVA method for Graeco-Latin square design. The ANCOVA method explores a dummy variable constructed for as follows $X=0$ for an observed Y and $X=\pm 1$ for $Y=0$, i. e. when the value is missing. (The data used in this example are the same as those used in the first example where the first method is applied).

Example 1. (see [2, p. 315]).

Table 6 *ANCOVA Table*

S. O. V.	df	XX	XY	YY
Row	$5-1$	$1/5-1/5^2=4/25$	$R_5'/5-G'/5^2=$	0.0416
		1 1 4	C' G'	
Column	$5-1$	$1/5-1/5^2=4/25$	$C_4'/5-G'/5^2=$	0.1176
		1 1 4	T' G'	
Treatment	$5-1$	$1/5-1/5^2=4/25$	$T_3'/5-G'/5^2=$	-0.1964
		1 1 4	G' G'	
Greek-Letter	$5-1$	$1/5-1/5^2=4/25$	$G_1'/5-G'/5^2=$	-0.0824
Error	$(5-1)(5-3)$	$(5-1)(5-3)/5^2=8/25$	$-[5(R_5'+C_4'+T_3'+G_1')-3G']/5^2=$	0.4308
Total	5^2-1	$1-1/5^2=24/25$	$-G'/5^2=$ -0.5504	

where

$$-b=\frac{5(R_5'+C_4'+T_3'+G_1')-3G'}{(5-1)(5-3)}=\frac{0.4308}{8/25}=1.34625.$$

Thus $\theta_{\mathrm{COVA}}=1.35$.

Comments. Two methods have been used in this paper so as to estimate the missing value. Both of them can be recommended but their application would depend on the situation. The first method is preferred since it allows a direct estimation of the missing value without using the analysis of variance. It also enables us to calculate approximate standard errors. The second method is more labour-consuming but it gives rise to unbiased estimates of standard errors and exact tests of significance. This may be of some importance when the two methods mentioned above have to be compared.

We wish to draw attention to the coincidence of the estimators obtained by both methods in this particular case of a single missing value. One can realize that this is due to the fact that once we carry out the ANCOVA method the original mode is replaced by a simple regression model involving a dummy variable indicatingl the missing value. Then the estimator is the negative regression coefficient obtained by minimizing the correspondent residual sum of squares. One can say that such a coincidence is not "unexpected". But is it really "expected"? In a forthcoming paper such an effect of coincidence of both the types of estimators will be shown to hold also for the multiple missing values case.

Acknowledgements. I wish to thank my research superviser Professor D. Vandev for our discussions. I am much obliged to Professor L. Boneva and Dr. I. Tzankova for their recommendations. The referees' comments are most gratefully acknowledged.

REFERENCES

1. K. W. Al-Rawl. Design and analysis of agricultural experiments. 1980.
2. G. I. Bliss. Statistics in Biology, New York, 1968.
3. W. G. Cochran, G. M. Cox. Experimental design. New York, 1957.
4. J. K. Haseman, D. W. Gaylom. An algorithm for non-iterative estimation of multipie missing values for crossed classifications. *Technometrics*, 15, 1973, 631-636.
5. C. Y. Kramer, S. Glass. Analysis of variance of a Latin-Square design with missing observation. *Appl. statist.*, 9, 1960, 43-50.
6. D. B. Rubin. A non-iterative algorithm for least squares estimation of missing values in any analysis of variance design. *Appl. Statist.*, 21, 1972, 136-141.

Institute of Mathematics
P. O. Box 373
Sofia 1090

Received 21. 01. 1991
Revised 05. 06. 1991