Jour. Ind. Soc. Ag. Statistics
***43** (3), 1991 : 310-320*

Analysis Of RBD With Missing Observations

Sheela S. Deo
University of Poona, Pune.
(Received : November, 1989)

Summary

Consider a randomised block design (RBD) with S missing observations belonging to a different treatment - block combination. Generally, these are substituted by the estimates of the missing observations which make the error sum of squares (s.s) minimum. Using the substitutes for missing observations we get error s.s. correctly but the hypothesis s.s. is not correct. we are interested in testing equality of treatment effects in RBD. The distribution of the treatment s.s. using these substitutes is worked out in this paper. It is a mixture of chi- square variables. Using this exact distribution, an approximate F test is suggested with illustrative example.

Key Words: RBD, C matrix, Bias, ch. roots.

Introduction

In & RBD with v treatments and b blocks let $s \geq 1$ observations be missing, such that each belongs to a different block and a different treatment. Let the missing observations be denoted by the unknowns $f_1, f_2, \ldots, f_s$ written in some convenient order. Let T_i and B_i denote the total of the treatment and block to which i-th missing observation belongs and be obtained by taking zero for the missing observations. Let G denotes the grand total of all the observations taking zero for the missing ones. The error s.s.,

$$SSE(f) = C_i + R \sum_i f_i^2 - 2 \sum_i f_i M_i + (2/bv) \sum_{i>j} f_i f_j \qquad (1.1)$$

where C_i does not contain any of the f_i's and

$$R = 1-(1/b)-(1/v)+(1/bv), \; M_i = (T_i/b)+(B_i/v)-(G/bv) \qquad (1.2)$$

The missing values are estimated by minimising the error s.s.

with respect to f_i's. They are given in Das and Giri [2]

$$f_i = \frac{bv}{(bvR-1)}\left[M_i - \frac{\sum_i M_i}{bvR-1+s}\right] \tag{1.3}$$

Let SSEC denote the conditional error s.s. under the hypothesis

$$H_o : \tau_1 = \tau_2 = \ldots = \tau_v.$$

$$SSEC = C_2 + R_1 \sum_i f_{ic}^2 - 2\sum_i B_i\ f_{ic}/v \tag{1.4}$$

where C_2 is a constant, independent of f_{ic}'s, and $R_1 = (v-1)/v$, and f_{ic} ($i = 1, 2, \ldots, s$) denote the unknowns substituted for the missing values under the hypothesis H_o. Minimising SSEC, we get

$$f_{ic} = B_i/(v-1),\ i = 1, 2, \ldots, s. \tag{1.5}$$

2. *Distribution of* $f - f_c$

Theorem 1: Expected values of $f_i - f_{ic}$ is a treatment contrast for every i, $i = 1, 2, \ldots, s$.

Proof: Without loss of generality, we can renumber blocks and treatments such that f_i is the missing observation in i-th treatment and i-th block. Making use of the model of RBD, result (1.3) and (1.5), we get,

$$E(f_i) = \mu + \tau_i + \beta_i \ \text{ and } E(f_{ic}) = \mu + \beta_i + \sum_{j \neq i}^{v} \tau_j/(v-1)$$

Here μ is the average effect, τ_i is the effect of i-th treatment and β_i is the effect of i-th block as usual. Hence

$$E(f_i - f_{ic}) = \frac{(v-1)\,\tau_i - \sum_{j \neq i}^{v} \tau_j}{(v-1)} \tag{2.1}$$

for every i, $i = 1, 2, \ldots, s$. It is a treatment contrast.

Let us denote by f and f_c, the vector of substitutes for missing observations, minimising SSE and SSEC respectively.

Theorem 2: The dispersion matrix of $f - f_c$ is $\sigma^2 \sum$

where

$$\sum = [(\lambda_1 - \lambda_2)\, I_s + \lambda_2\, E_{ss}] \tag{2.2}$$

$$\lambda_1 - \lambda_2 = \frac{v^2}{(v-1)\,(bv-v-b)} \tag{2.3}$$

$$\text{and } \lambda_1 + (s-1)\,\lambda_2 = \frac{v(v-s)}{(v-1)\,(bv-b-v+s)} \tag{2.4}$$

Proof: We get the following results for i = 1, 2, . . ., s.

(i) $\quad \text{Var } f_{ic} = \dfrac{\sigma^2}{(v-1)}$, (ii) $\text{Cov}(M_i, B_i) = R\sigma^2$

(iii) $\quad \text{Cov}\left(B_i, \sum\limits_i M_i\right) = \dfrac{(bvR - 1 + s)\,\sigma^2}{bv}$

(iv) $\quad \text{Var}(M_i) = \dfrac{[(b-1)(v^2-2v) + (v-1)\,(b^2-2b) + (bv-s)]\,\sigma^2}{b^2\,v^2}$

(v) $\quad \text{Cov}(M_i, M_j) = \dfrac{(2v + 2b - bv - s)\,\sigma^2}{b^2\,v^2}$, for $i \neq j$

From (iv) and (v) we get

$$\text{Var}(f_i) = \frac{b^2\,v^2}{(bvR-1)^2}\left(\text{Var}(M_i) + \frac{\text{Var}\left(\sum\limits_i M_i\right)}{(bvR-1+s)^2} - \frac{2\,\text{Cov}\left(M_i, \sum M_i\right)}{(bvR-1+s)}\right)$$

Now $\quad \text{Var}\left(\sum\limits_i^s M_i\right) = S(\text{Var } M_i) + \sum\limits_{i \neq j}\sum \text{Cov}(M_i, M_j)$ and

$\text{Cov}\left(M_i, \sum\limits_i M_i\right) = \text{Var } M_i + (s-1)\,\text{Cov}(M_i, M_j)$ gives

$$\text{Var}(f_i) = \frac{b^2 v^2}{(bvR-1)^2}\left[\text{Var}(M_i)\left\{1+\frac{s}{(bvR-1+s)^2}-\frac{2}{(bvR-1+s)}\right\}\right.$$

$$\left.+\frac{\text{Cov}(M_i, M_j)}{(bvR-1)}\left\{\frac{s(s-1)}{bvR-1+s}-2(s-1)\right\}\right]$$

We note that $\text{Var}(M_i)$ and $\text{Cov}(M_i, M_j)$ are independent of i and j from (iv) and (v).

Hence Var $f_i = \sigma^2 g$, where g is a constant for every i

var f_{ic} = cov $(f_i, f_{ic}) = \sigma^2/(v-1)$, so that

$$\text{var}(f_i - f_{ic}) = \sigma^2\left[g - \frac{1}{(v-1)}\right] = \lambda_1\sigma^2 \text{ (say)} \qquad (2.6)$$

Now consider

$$\text{cov}[(f_i-f_{ic}), (f_j-f_{jc})] = \text{cov}(f_i, f_j) + \text{cov}(f_{ic}, f_{jc}) - \text{cov}(f_i, f_{jc}) - \text{cov}(f_j, f_{ic})$$

Using the following results,

(a) $\text{cov}(f_i, f_{jc}) = \text{cov}\left\{f_i, \frac{B_j}{(v-1)}\right\}$,

(b) $\text{cov}(M_i, B_j) = \frac{\sigma^2}{bv}$,

(c) $\text{cov}(B_j, T_i) = \sigma^2$,

(d) $\text{cov}(B_j, B_i) = 0$

(e) $\text{cov}(B_j, G) = (v-1)\,\sigma^2$

we get $\text{cov}\left(B_j, \sum_i M_i\right) = \frac{(bvR-1+s)\,s^2}{bv}$

$$\text{cov}(f_i, f_{jc}) = 0 = \text{cov}(f_{ic}, f_{jc})$$

Hence

$$\text{cov}[(f_j, f_{jc}), (f_i - f_{ic})] = \text{cov}(f_i, f_j) \qquad (2.7)$$

$$\text{Cov}(f_i, f_j) = \frac{b^2 v^2}{(bvR-1)^2}\left[\text{Cov}(M_i, M_j)\left\{1 - \frac{2(s-1)}{bvR-1+s} + \frac{s(s-1)}{(bvR-1+s)^2}\right\} + \text{Var } M_i \left\{\frac{s-2\,(bvR-1+s)}{(bvR-1+s)^2}\right\}\right] = \lambda_2 \sigma^2 \quad (2.8)$$

Using (2.5), (2.6), (2.7) and (2.8)

$$(\lambda_1 - \lambda_2)\,\sigma^2 = \frac{b^2 v^2}{(bvR-1+s)^2}\left[\text{var } M_i - \text{Cov}(M_i, M_j)\right] - \left(\frac{\sigma^2}{(v-1)}\right) \quad (2.9)$$

and

$$(\lambda_1 + (s-1)\lambda_2)\,\sigma^2 = \frac{b^2 v^2}{(bvR-1)^2}\left\{\text{cov}(M_i, M_j)\,\frac{(s-1)\,(bv-b-v)^2}{(bvR-1+s)^2} + \text{Var}(M_i)\,\frac{(bv-b-v)^2}{(bvR-1+s)^2}\right\} - \frac{\sigma^2}{(v-1)} \quad (2.10)$$

After substituting the expressions for var M_i and Cov (M_i, M_j) from (iv) and (v) in (2.9) and (2.10) we get the required results.

Theorem 3: The distribution of $(f-f_c)' \sum^{-1} (f-f_c)$ is $\sigma^2 \chi_s^2$ when the hypothesis of equality of treatments is true.

Proof: Using Theorem 1, we get $E(f - f_c) = 0$, when the hypothesis $H_o : \tau_1 = \tau_2 = \ldots = \tau_v$ is true. We note that the characteristic roots of Σ [dispersion matrix of $\frac{(f-f_c)}{\sigma^2}$ are $\lambda_1 - \lambda_2$ with multiplicity s-1 and $\lambda_1 + (s-1)\lambda_2$ with multimplicity one. If we make an orthogonal transformation $\underline{Y} = D(\underline{f} - \underline{f}_c)$ such that $D \Sigma D' = \text{diag}(\theta_1, \ldots, \theta_s)$ where θ_i's are the ch. roots of Σ,

$$\theta_i = \lambda_1 - \lambda_2, \; i = 1, 2, \ldots, s-1 \quad (2.11)$$

and $$\theta_s = \lambda_1 + (s-1)\lambda_2 \quad (2.12)$$

then $$(f-f_c)' \Sigma^{-1} (f-f_c) = \sum_{i=1}^{s} \frac{y_i^2}{\theta_i} \quad (2.13)$$

Here each $\frac{y_i^2}{\theta_i}$ is distributed independently of the others, E(y) = $\underline{O}$ under H_o and the observations are independently normally distributed in RBD together implies that $\frac{y_i^2}{\theta_i}$ is $\sigma^2 \chi^2$ with 1 d.f. for every i = 1, 2, . . ., s when the hypothesis H_o is true.

3. *The Distribution of the Treatment S.S.*

The treatment s.s. using substitutes f for missing observations is denoted by SST(f), without loss of generality.

$$SST(f) = \frac{\sum_{i=1}^{s} (T_i + f_i)^2 + \sum_{i=s+1}^{v} T_i^2}{b} - \frac{(G + \Sigma f_i)^2}{bv} \qquad (3.1)$$

The correct hypothesis s.s. is, SST, where

$$SST = \min_f \ [SSE(f) + SST(f)] - \min_f [SSE(f)]$$

$$SST = SSE(f_c) + SST(f_c) - SSE(f) \qquad (3.2)$$

The bias in the hypothesis s.s. for $H_o : \tau_1 = \tau_2 = \ldots = \tau_v$ is defined as SST(f) - SST, which is always positive and is given by

$$\text{bias} = (f - f_c)' \operatorname{diag}\left\{\frac{(v-1)}{v}, \ldots, \frac{(v-1)}{v}\right\}(f - f_c) \qquad (3.3)$$

We note that

$$SST(f) = SST + \text{bias}$$

$$= SST + (f - f_c)' \operatorname{diag}\left\{\frac{(v-1)}{v}, \ldots, \frac{(v-1)}{v}\right\}(f - f_c) \qquad (3.4)$$

It is well known that SST has $\sigma^2 \chi^2$ distribution with v-1 d.f. with complete data. The rank of C matrix of the design is v-1. $C=R-NK^{-1}N$. Here N = (nij) = Evb : incidence matrix.

$R = bI_v$, $K = vI_b$, nij denotes the number of observations on i-th treatment and j-th block. For complete data, nij = 1 for RBD.

The C matrix gives v-1 linearly independent estimable treatment contrasts. We are considering the case of s missing observations. This implies that s linearly independent estimable treatment contrasts are affected. These are given in (2.1). The remaining v-1-s treatment contrasts are unaffected by the missing values. The distribution of SST is $\sigma^2 \chi^2_{v-1-s} + \sigma^2 \chi^2_s$, where these two chi-squares are independent.

The exact s.s. for s linearly independent affected treatment contrasts and bias using Theorem 2, with s missing observations is

$$(f-f_c)' \Sigma^{-1} (f-f_c) + (f-f_c)' \varphi (f-f_c)$$

Here $\varphi = \frac{v-1}{v} I_s$, combining these two terms we get

$$(f-f_c)' (\varphi + \Sigma^{-1}) (f-f_c)$$

Using the orthogonal transformation in Theorem 2 and 3, it becomes

$$\sum_{i=1}^{s} y_i^2 \left[\frac{v-1}{v} + \left(\frac{1}{\theta_i}\right)\right] \tag{3.5}$$

The θ_i's are given in (2.11), (2.12). Substituting the values from (2.3) and (2.4), we get

$$\frac{b(v-1)}{bv-b-v} \sum_{i=1}^{s-1} Z_i^2 + \frac{b(v-1)}{bv-v-b+s} Z_s^2 \tag{3.6}$$

Here $Z_i^2 = \frac{y_i^2}{\theta_i}$, i = 1, 2, . . ., s.

Hence using Theorem 3, SST(f) in (3.4) has

$$\sigma^2 \chi^2_{v-s-1} + \frac{b(v-1)}{bv-b-v} \sigma^2 \chi^2_{s-1} + \frac{b(v-1)}{bv-v-b+s} \sigma^2 \chi^2_1 \tag{3.7}$$

Comment : Let us consider the C matrix of available observations as C^*. Here nij = 0 for s missing observations. The ch. roots of C^* are b with multiplicity v-s-1, $b-1-\frac{1}{(v-1)}$ with multiplicity s-1 and

$\frac{(bv-b-v+s)}{(v-1)}$ with multiplicity one. We know that C matrix of RBD with v treatments and b blocks has v-1 non zero ch. roots as b. In this paper we are considering s missing observations in a different treatment block combination. Here s ch. roots of C^* matrix are different than C matrix of RBD with complete data. The coefficients of χ^2 in (3.7) are b times reciprocals of the ch. roots of C^* matrix.

When two observations are missing, we do not have enough observations for analysis in the following cases.

(i) v = 2, b = 2, (ii) v = 2, b = 3 and (iii) v = 3, b =2.

The rest of the cases are partitioned into three situations.

Case I: v = 2, b ≥ 4, s = 2.

In this case missing observations are estimable. The ch. roots of C matrix are 0, b while ch. roots of C^* are 0 and b-2. Hence SST(f) has $\left(\frac{b}{(b-2)}\right)\sigma^2\chi_1^2$ distribution.

Case II: v ≤ b, v ≥ 3, s = v

The design is balanced in this case, C^* has (v-1) ch. roots equal and different from zero. Hence SST(f) has $\frac{b(v-1)}{(bv-b-v)}\sigma^2\chi^2_{v-1}$

Case III: When b ≥ 2, v ≥ 4, v > b and s ≤ min (v, b)

The distribution of SST(f) is given by (3.7)

4. *Testing of Hypotheses*

For $H_0 : \tau_1 = \tau_2 = \ldots = \tau_v$, SST(f) can be used. The mixture of independent chi-squares can be approximated using Patnaik's approximation [9] given in Johnson and Kotz [10].

Let $E\left[\frac{SST(f)}{\sigma^2}\right] = a_1$, say and $var\left[\frac{SST(f)}{\sigma^2}\right] = 2a_2$, then $\frac{a_1\,SST(f)}{a_2}$ follows $\sigma^2\chi^2$ distribution with $\frac{a_1^2}{a_2}$ d.f. When the hypothesis H_0 is true, using (3.7)

$$a_1 = v-s-1+\frac{b(v-1)(s-1)}{bv-b-v}+\frac{b(v-1)}{bv-b-v+s} \qquad (4.1)$$

and $$a_2 = v - s - 1 + \frac{b^2(v-1)^2(s-1)}{(bv-b-v)^2} + \frac{b^2(v-1)^2}{(bv-b-v+s)^2} \tag{4.2}$$

Hence the following test is proposed.

$$F = \frac{SST(f)}{s_s^2} \times \frac{(b-1)(v-1)-S}{a_1} \tag{4.3}$$

Deo and Kharshikar [3] [5] [6] have proved that under H_0, this test statistics, follows F distribution with

$$\left[\frac{a_1^2}{a_2}, (b-1)(v-1)-s\right] \text{d.f.} \tag{4.4}$$

5. *Application*

The illustrative example is taken from Das, and Giri [2] (page 65).

A problem was posed to estimate the petrol consumption rates of the four different makes of cars for suitable average speed and compare them. The following experiment was conducted.

Five different cars of each of four makes were chosen at random. The five cars of each make were put on road on 5 different days. The cars of a make ran with different speeds on different days, which car was to put on the road on which day and what speed it should have was determined through a chance mechanism subject to the above condition of the experiment.

For each car the number of miles covered per gallon of petrol was observed. The observations are presented below.

Table 1. Miles per gallon of Petrol

Makes of Car	Speed of the cars in miles per hour (mph)				
	25	35	50	60	70
A	(20.6)	19.5	18.1	17.9	16.0
B	19.5	(19.0)	15.6	16.7	14.1
C	20.5	18.5	(16.3)	15.2	13.7
D	16.2	16.5	15.7	(14.8)	12.7

The analysis of complete data gives.

Table 2. Analysis of Variance Table

Source	d.f.	s.s.	m.s.s.	F
Speeds (block)	4	63.77	15.94	
Makes (treatments)	3	26.35	8.78	14.18**
Error	12	7.43	0.62	
Total	19	97.55		

There is highly significant difference among the different makes. If some observations are missing then this F value decreases. So significant value of F may become insignificant due to missing values and conclusion will be changed.

(1) We will consider first that four diagonal observations which are bracketed are missing then analysis is done with substitutions by (1.3).

$f_1 = 20.8667$, $f_2 = 18.1576$, $f_3 = 16.5667$, $f_4 = 14.2757$

ANOVA **Table 3.** (Analysis with four missing)

Source	d.f.	s.s.	m.s.s.	F
Speeds (block)	4	63.72	15.93	
Makes (treatments)	3	29.25	9.75	11.676
Error	8	6.74	0.84	
Total	15	99.70		

The F ratio with biased hyp. s.s. = 11.676.

Usual way of doing the exact analysis.

$$\text{The amount of bias} = (v-1)\sum_{1}^{4}\frac{(f_i - f_{ic})^2}{v} \text{ using (3.3)}$$

$$= 3(10.4)/4 = 7.8$$

$$SST = SST(f) - \text{bias} = 29.25 - 7.8 = 21.45.$$

The F ratio of the mean exact treatment s.s. to mean error s.s. is 8.489. The F statistics suggested in this paper is based on SST(f) using (4.3) and (4.4).

Since $\frac{SST(F)}{a_1} = 7.15$ where $a_1 = 4.0909$ and $a_2 = 3$. F = 8.489

To get a test of level α, the critical point of F distribution with $\left(\frac{a_1^2}{a_2}, 8\right)$ is $\frac{F_0(v-1)}{a_1}$. Here F_0 is $(1-\alpha)\%$ cumulative point of F distribution with (3, 8) d.f.

We note that the ch. roots of C^* matrix with four missing observations are $b-1-\frac{1}{v-1}=\frac{11}{3}$ with multiplicity 3. The distribution of the hypothesis s.s. is

$$\frac{b(v-1)}{bv-b-v}\sigma^2\chi^2_{v-1} = \frac{15\sigma^2\chi^2_3}{11}$$

REFERENCES

[1] Das M.N., 1954. Missing plots and randomised block designs with balanced incompleteness. *Jour. Ind. Soc. Agrl. Stat.*, **6**, 54-76.

[2] Das, M.N. and Giri N.C., 1966. *Design and Analysis of Experiments.* Wiley Eastern Ltd.

[3] Deo, S.S. and Kharshikar, A.V., 1985. On consequences of one missing observation in a balanced design. *Commun. Statist.*, **14**, No. 12, 3091-3100.

[4] Deo, Sheela, S., 1986. *Study of Linear Models with Missing Observations*, Ph.D. thesis, Poona Univ. Pune, India.

[5] Deo, S.S. and Kharshikar, A.V., 1988. Effect of two mixed up yields in a randomised block design, Calcutta Statistical Association Bulletin, **37**, 145-146, 105-109.

[6] Deo, S.S. and Kshirsagar, A.M., 1989. Distribution of biased hypothesis sum of squares in linear models with missing observations, *Commun. Statist.* **18(8)**, 2747-2754.

[7] Deo, S.S. and Kharshikar, A.V., 1986. Analysis of covariance with one missing observation in a balanced block design. To appear in the Journal of Indian Society for Probability and Statistics, **8.**

[8] Kshirsagar, A.M., 1971. Bias due to missing plots. *American Statistician*, **25**, 47-50.

[9] Patnaik, P.B., 1949. The non-central χ^2 and F distributions and the applications. *Biometrika*, **36**, 202-232.

[10] Johnson, N.L. and Kotz, S., 1970. *Continuous Univariate Distributions.* Houghton Mifflin Company Bonston.