Prosiding Konperensi Nasional X Matematika
ITB, 17 – 20 Juli 2000

Testing Random Effects in the Balanced Incomplete Block Design

Sigit Nugroho
University of Bengkulu, Indonesia

Abstract

We study the statistical testing of random effects in the Balanced Incomplete Block Design by using the Locally Most Powerful Rank Test and some assumptions. The Logistic case having linear relationship with the statistic derived by Durbin for testing no means/medians differences among treatments are also derived.

1 Introduction

If one wishes to investigate a set of c treatments but blocks of size c are not available, i.e. there exist homogeneous blocks of size $k < c$, or can be constructed, the an incomplete block design is compulsory.

Definition 1 *A Balanced Incomplete Block Design (BIBD) is an experimental setup where c treatments are to be allocated to b blocks in such a way that: (i) Each block contains $k < c$ treatments. (ii) Each treatment appears in exactly r blocks. (iii) Every pair of treatments occurs together in λ blocks; that is, every pair of treatments appears together equally often.*

It follows, for every BIBD, that $\lambda < r < b$. The BIBD can be characterized by five numerical values: λ, r, b, k, and c. Another inequality for the BIBD, as stated in the definition, is $k < c$. Two important equalities are $rc = bk$ and $\lambda(c-1) = r(k-1)$.

The number rc represents the number of times all treatments appear in the experiment which is equal to the total number of experimental units (bk). Consider the blocks where a specific treatment occurs, say treatment j. Besides this treatment there are $\lambda(c-1)$ other treatment units in these blocks because each of the other $c-1$ treatments must occur λ times with treatment j. But treatment j must appear in r blocks each containing $k-1$ units other than the one on which treatment j occurs. The product, $r(k-1)$, represents the number of units available for the $\lambda(c-1)$ 'other treatments'.

BIBD always exists for any values of c and k. One can continue by adding blocks of size k until the conditions in the definition hold, but BIBD obtained in this manner may get prohibitively large. Possible values of λ, b and r for BIBD can be obtained for any given values of k and c; but not all values obtained in this way lead to BIBD; multiples of λ, b and r may be required. Block compositions could also be formed to construct the BIBD by trial and error, eventhough in general it is a difficult task.

Durbin (1951) proposed a statistic for testing the null hypothesis of no differences among treatments (means or medians) in the balanced incomplete block design as follows:

$$D = \frac{12(c-1)}{rc(k-1)(k+1)} \sum_{j=1}^{c} \left[\sum_{i=1}^{b} R_{ij} \right]^2 - \frac{3r(c-1)(k+1)}{k-1}$$

A statistic for testing the hypothesis of no difference between c treatments in the two-way fixed effects model by ranks was derived by Friedman; while Anderson proposed a nonparametric statistic for testing the consistency in randomized block experiment. The parametric statistic for testing similar hypothesis of no treatment effect

The parametric approach in data analyses sometimes causes problems. The assumption that the random errors are normally distributed is not always true, and the normality of the random variables of interest can be more questionable, because these variables are not directly observed.

In this paper we would like to study a statistic for testing the similar hypothesis in BIBD when we have fixed block effects and random treatment effects based on rank. Nugroho and Govindarajulu (1999) has derived the statistic for testing the random effects in two-way experiment with one observation per cell.

2 Notations and Assumptions

Randomization is done within each block. Thus, since each block having size k, then the toal number of observation in each block is k. Therefore, the rank in each block ranges from 1 to k. Let us denote $X_{i(j)}$ the observation in ith block receiving jth treatment and $W_1^{(i)} < W_2^{(i)} < \ldots < W_k^{(i)}$ the order sample of the variables $X_{i(j)}$, $j = 1, 2, ..., c$ $i = 1, 2, ...b$. Also denote the k sample rank order for the ith block as follows:

$$\underline{Z}^{(i)} = \left(Z_1^{(i)}, Z_2^{(i)}, ..., Z_k^{(i)} \right) \text{ where } Z_m^{(i)} = j \text{ jika } W_m^{(i)} = X_{i(j)}$$

So, for each block there are $k!$ possible rank orders. Now consider $\underline{Z} = \left(\underline{Z}^{(1)'}, \underline{Z}^{(2)'}, ..., \underline{Z}^{(b)'} \right)$. Let z denote any possible realization of the $(k!)^b$ possible rank orders. The model for the scheme is

$$X_{i(j)} = \mu + \beta_i + Y_j + \varepsilon_{ij} \quad i = 1, 2, ..., b \quad j = 1, 2, ..., c$$

Y_j has distribution G and ε_{ij} has distribution F; the Y_j and the ε_{ij} are mutually independent random variables; β_i is fixed block effect and additive. No interaction between block and treatment effects is assumed. The additional assumptions are $\int_{-\infty}^{\infty} y \, dG(y) = 0$ and $\int_{-\infty}^{\infty} y^2 \, dG(y) = \sigma^2 < \infty$, where the first does not cause any loss of generality. Then, we are interested in testing the hypotheses $H_0 : G(y) = \begin{cases} 0 & , \quad y < 0 \\ 1 & , \quad y \geq 0 \end{cases}$ against $H_1 : G(y)$ is member of the class of nontrivial distribution functions.

Let $\tilde{G}(y) = G(y/\Delta)$ be a class of nontrivial distribution functions for some small and positive Δ. We note that the distribution is degenerate at zero is equivalent to the statement that for this class, $\Delta = 0$. To derive a locally most powerful rank test for this design, we consider the following hypotheses: $H_0 : \Delta = 0$ against $H_\Delta : \Delta > 0$.

3 The LMP Rank Test Statistic

The locally most powerful rank test statistic for the balanced incomplete block design is derived in this section.

Theorem 2 *For the Balanced Incomplete Block Design Locally Most Powerful Rank Test for H_0 against H_Δ is given by : Reject H_0 if $\Lambda \geq K_\alpha$ where*

$$\Lambda = \sum_{i=1}^{b} \sum_{j=1}^{c} \sum_{m=1}^{k} E_0 \left[\frac{f'(W_m^{(i)})}{f(W_m^{(i)})} \right] \delta_{jZ_m^{(i)}}$$
$$+ \sum_{j=1}^{c} \sum_{i \neq l}^{b} \sum^{b} \left[\sum_{m=1}^{k} E_0 \left[\frac{f'(W_m^{(i)})}{f(W_m^{(i)})} \right] \right] \left[\sum_{t=1}^{k} E_0 \left[\frac{f'(W_t^{(l)})}{f(W_t^{(l)})} \right] \right]$$

and K_α is determined by the level of significance α and $\delta_{ij} = 1$ if $i = j$ and 0 otherwise, provided the following conditions are satisfied: (i) the density f has a derivative that is absolutely continuous over finite intervals, (ii) $f''(x)$ is continuous almost everywhere, (iii) $\int y^2 \, dG(y) < \infty$, and (iv) $E \left| (\delta^2/\delta X^2) \ln f(X) \right| < \infty$

Proof. Consider the following quantity $\lim_{\Delta \to 0} \left\{ \frac{P(\underline{Z}=\underline{z}|H_\Delta) - P(\underline{Z}=\underline{z}|H_0)}{\Delta^2} \right\}$. Using the conditional values of the $Y_j's$, we have $P(\underline{Z} = \underline{z}|H_\Delta) - P(\underline{Z} = \underline{z}|H_0) =$

$$\int_{\tilde{y}_1=-\infty}^{\infty} \cdots \int_{\tilde{y}_c=-\infty}^{\infty} [P(\underline{Z} = \underline{z}|H_\Delta, Y_1 = \tilde{y}_1, ..., Y_c = \tilde{y}_c)$$

$- P(\underline{Z} = \underline{z}|H_0, Y_1 = \tilde{y}_1, ..., Y_c = \tilde{y}_c)] \prod_{j=1}^{c} dG(\tilde{y}_j/\Delta)$. By letting $y_j = \tilde{y}_j/\Delta$, $j = 1, ..., c$. Then $P(\underline{Z} = \underline{z}|H_\Delta) - P(\underline{Z} = \underline{z}|H_0) =$

$$\int_{y_1=-\infty}^{\infty} \cdots \int_{y_c=-\infty}^{\infty} [P(\underline{Z} = \underline{z}|H_\Delta, Y_1 = \Delta y_1, ..., Y_1 = \Delta y_c)$$

$$- P(\underline{Z} = \underline{z}|H_\Delta, Y_1 = \Delta y_1, ..., Y_1 = \Delta y_c)] \prod_{j=1}^{c} dG(y_j).$$

Under the null hypothesis, we know that the treatment effects are constant and equal; therefore, conditioning of the $Y_j's$ is irrelevant for the term. We also have the independence from block to block. Thus, the last integrand above can be written as

$$\prod_{i=1}^{b} \left[P\left(\underline{Z}^{(i)} = \underline{z}^{(i)} | H_\Delta, Y_1 = \Delta y_1, ..., Y_c = \Delta y_c \right) \right]$$
$$- \prod_{i=1}^{b} \left[P\left(\underline{Z}^{(i)} = \underline{z}^{(i)} | H_0 \right) \right]$$

Since $P\left(\underline{Z}^{(i)} = \underline{z}^{(i)} | H_\Delta, Y_1 = \Delta y_1, ..., Y_c = \Delta y_c\right) =$

$$\int_{-\infty < w_1^{(i)} < ... < w_c^{(i)} < \infty} \cdots \int \prod_{m=1}^{k} \left[f\left(w_m^{(i)} - \Delta \sum_{j=1}^{c} y_j \delta_{jZ_m^{(i)}} \right) dw_m^{(i)} \right].,$$

then letting $\theta_m^{(i)} = \sum_{j=1}^{c} y_j \delta_{jZ_m^{(i)}}$ and $h_i(\Delta) = \prod_{m=1}^{k} f\left(w_m^{(i)} - \Delta\theta_m^{(i)}\right)$, we can find that

$$\prod_{i=1}^{b} \left[P\left(\underline{Z}^{(i)} = \underline{z}^{(i)} | H_\Delta, Y_1 = \Delta y_1, ..., Y_c = \Delta y_c\right) \right]. =$$

$$\prod_{i=1}^{b} \left[\int_{-\infty < w_1^{(i)} < ... < w_c^{(i)} < \infty} \cdots \int h_i(\Delta) \prod_{m=1}^{k} dw_m^{(i)} \right].$$

Using Taylor expansion $h_i(\Delta) = h_i(0) + \Delta h_i'(0) + \frac{1}{2}\Delta^2 h_i''(0) + \frac{1}{2}\Delta^2[h_i''(\Delta^*) - h_i''(0)]$, evaluating the integrand back into the form at the beginning of the proof, applying the assumptions, and finally applying the limit as $\Delta \to 0$, after dividing the result by Δ^2 we have $\frac{1}{2}\sigma^2\left(\frac{1}{k!}\right)^b \Lambda$ where Λ is the statistic in the Theorem. We just simply ignore the term $\frac{1}{2}\sigma^2\left(\frac{1}{k!}\right)^b$ since it is constant for any given b and k. ■

4 Test for Logistic Scores

Let R_{ij} be the ranks associated with observation getting j-th treatment in i-th block, i.e. $R_{ij} = \sum_{m=1}^{k} m\delta_{jZ_m^{(i)}}$. In the case of logistic distribution, we have

$$\sum_{m=1}^{k} E_0\left[\frac{f'(W_m^{(i)})}{f(W_m^{(i)})}\right] \delta_{jZ_m^{(i)}} = \sum_{m=1}^{k} E_0\left[1 - 2U_m^{(i)}\right] \delta_{jZ_m^{(i)}} = 1 - \frac{2}{k+1} R_{ij}$$

where $U_m^{(i)}$ is the m-th order statistic from a standard uniform distribution in the sample of size k for the i-th block. One can easily verify that

$$E_0\left(U_m^{(i)}\right) = \frac{m}{k+1} \text{ and } E_0\left(U_m^{(i)}\right)^2 = \frac{m(m+1)}{(k+1)(k+2)}$$

It can be shown that the first term of Λ is equal to zero. Then we have the following Lemma.

Lemma 3 *In the case of Logistic distribution for the errors, the test statistic becomes*

$$\Lambda_L = \frac{4}{(k+1)^2} \sum_{j=1}^{c} \left[\sum_{i=1}^{b} R_{ij}\right]^2 - b^2 k \cdot \frac{bk(k-1)}{3(k+1)}$$

Proof. As a special case of Λ and using the relation given in the beginning of this section we find that $\Lambda_L = \frac{4}{(k+1)^2} \sum_{j=1}^{c} \sum_{i \neq l}^{b} \sum^{b} R_{ij} R_{lj} - kb(b-1)$ or $\Lambda_L = \frac{4}{(k+1)^2} \sum_{j=1}^{c} \left[\left(\sum_{i=1}^{b} R_{ij} \right)^2 - \sum_{i=1}^{b} R_{ij}^2 \right] - kb(b-1)$ Using the fact that $\sum_{j=1}^{c} \sum_{i=1}^{b} R_{ij}^2 = \sum_{i=1}^{b} \sum_{j=1}^{c} R_{ij}^2 = \frac{bk(k+1)(2k+1)}{6}$ the result follows. ■

The following theorems can be easily verified or proven.

Theorem 4 *Λ_L and D are linearly related, and this can be described as $\Lambda_L = \frac{bk(k-1)}{3(c-1)(k+1)}[D - c + 1] + (r-b)bk$*

Theorem 5 *Since Λ_L is a linear combination of D, it has the same asymptotic properties as D when H_0 is true.*

Theorem 6 *Durbin's test criterion is locally most powerful for H_0 against alternatives in the balanced incomplete block design.*

5 Asymptotic Distribution of Λ_L Under $\mathbf{H}_0$

Let R_{ij} be the rank of $X_{i(j)}$. Note that R_{ij} is distributed uniformly under the null hypothesis. Also note that $R_{1j}, R_{2j}, ..., R_{bj}$ are independent and identically distributed on $(1, ..., k)$. Hence, $E(R_{ij}|H_0) = (k+1)/2$ and $Var(R_{ij}|H_0) = (k^2-1)/12$. Consequently, $\sum_{i=1}^{b} \left(R_{ij} - \frac{k+1}{2}\right) \quad \xrightarrow{D} \quad N\left(0, \frac{r(k^2-1)}{12}\right)$ for large b. That is also $\sqrt{\frac{3(k+1)}{b(k-1)}} \frac{2}{k+1} \sum_{i=1}^{b} \left(R_{ij} - \frac{k+1}{2}\right) \quad \xrightarrow{D} \quad N(0,1)$.

Because of one linear constraint, namely $\sum_{j=1}^{c} \sqrt{\frac{12}{r(k^2-1)}} \sum_{i=1}^{b} \left(R_{ij} - \frac{k+1}{2}\right) = 0$, $\sum_{j=1}^{c} \frac{12}{r(k^2-1)} \left\{ \sum_{i=1}^{b} \left(R_{ij} - \frac{k+1}{2}\right) \right\}^2 \quad \xrightarrow{D} \quad \chi^2_{c-1}$. Here we apply the theorem in Hajek and Sidak or Cochran's theorem.

And as b and c get large $\sum_{j=1}^{c} \frac{12}{r(k^2-1)} \left\{ \sum_{i=1}^{b} \left(R_{ij} - \frac{k+1}{2}\right) \right\}^2$ will have a standard normal distribution when suitably standardized.

References

[1] Anderson, R.L. (1959). Use of contingency tables in the analysis of consumer preferences studies. *Biometrics* **15**. 582-290.

[2] Durbin, J. (1951). Incomplete blocks in ranking experiments. *The British Journal of Psychology*, Statistical Section **IV** (II), 85-90

[3] Friedman, M. (1937). The use of ranks to avoid the assumption of normality implicit in the analysis of variance. *J. of Amer. Statist. Assoc.* **32**, 675-701.

[4] Govindarajulu, Z. and J.V. Deshpande. (1972). Random effects model: Nonparametric case. *Ann. Inst. Statist. Math.* **24**, 165-170.

[5] Govindarajulu, Z. (1975). Locally most powerful rank order tests for one-way random effects model. *Studia Scientarium Mathematicarum Hungarica.* **10**, 47-60.

[6] Hájek, J. and Z. Sidák. (1967). *Theory of Rank Tests.* Academic Press, New York.

[7] Nugroho, S. and Z. Govindarajulu (1999). Tests for Random Effects in Two-way Experiment with One Observation per Cell. *Indian Journal of Mathematics.* B.N. Prasad Birth Centenary Commemoration Volume. **41** (1) 55-66.

[8] Odeh, R.E. (1977). Extended tables of the distribution of Friedman's S-statistic in two-way lay-out. *Commun. Statist.* B6: 29-48.

[9] Skilling, J.H. and G.A. Mack. (1981). On the use of a Friedman-type statistic in Balanced and Unbalanced Block Designs. *Technometrics* **23**. 171-177.