International Journal of Statistics and Applied Mathematics 2024; 9(1): 52-57

International Journal of Statistics and Applied Mathematics

ISSN: 2456-1452
Maths 2024; 9(1): 52-57

https://www.mathsjournal.com
Received: 07-11-2023
Accepted: 08-12-2023

Nitiraj Shete
Research Scholar,
Department of Statistics, Sardar Patel University, Gujarat, India

Dr. Ashok Shanubhogue
Retired Professor,
Department of Statistics, Sardar Patel University, Gujarat, India

Corresponding Author:
Nitiraj Shete
Research Scholar,
Department of Statistics, Sardar Patel University, Gujarat, India

A frailty model approach to the randomized block design

Nitiraj Shete and Dr. Ashok Shanubhogue

DOI: https://dx.doi.org/10.22271/maths.2024.v9.i1a.1600

Abstract
In this research paper we discussed how to predict some part of variability of unexplained part of variation using frailty random variable method. In this paper, the discussion done on the frailty model approach to the randomized block design. We conduct an experiment in which we can visually see the difference between the different treatments used but statistically we are not able to show the significant difference among these treatments. So this paper used so far to tackle this type of situation or problem. We develop a theory for this situation and used same for the conducted experiment.

Keywords: RBD, frailty, Bartlett's and Leven's Test, K-S test, AD test

Introduction

In survival analysis the problem of heterogeneity is dealt by incorporating frailty random variable. The first univariate frailty model was suggested by Beard (1959) [2], considering different mortality models. The same model was independently suggested by Vaupel (1979) [4] and Lancaster (1979) [4]. Beard (1959) [2] used longevity factor instead of the term frailty and later on the term frailty was introduced by Vaupel (1979) [4] in the univariate case. We observe that same concept can be incorporated in other statistical studies suitable to solve some seemingly mysterious problems. Also Ashok Shanubhogue and Nitiraj Shete (2018) [1], discussed the analysis of frailty model approach completely randomized design. In this research paper is we have used frailty model approach to analyze the data based on RBD.

We have found many situations where the experimental units within the blocks, though homogeneous will have hidden source of variability due to environment, genetic factor etc. Thus it becomes necessary to incorporate these hidden sources of variability in the model. One way is to consider the hidden variability is explained by a common unknown random variable Z which has expected value one. The random variable Z which is used as variance modifier is known frailty random variable.

Proposed model for randomized block design (RBD)

In RBD each treatment replicated exactly once in each block. Therefore, for RBD no. of blocks (b) equals to no. of replications (r) and no. of treatments (v) equals to size of the block (k). Let ith treatment be replicated r (=b) times (i=1, 2, 3…, v); n=vr the total no. of observation. The linear model assuming various effects to be additive becomes.

$$y_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij} \text{ for all } i= 1,2,\ldots, v \text{ and } j=1,2,\ldots,b \qquad (1)$$

Where, y_{ij} be the yield or response in j^{th} block receiving i^{th} treatment
μ be the general mean effect.
α_i be the effect due to i^{th} treatment.
β_j be the effect due to j^{th} block.
ε_{ij} be the error effect due to chance.

We assume that
i. The various effects are additive in nature
ii. ε_{ij} are i.i.d. N(0, σ_e^2)

Let us consider i.i.d continuous frailty random variable Z_{ij} associated with (i, j)th experimental unit. We assume that $\varepsilon_{ij}|z_{ij} \sim N(0, \frac{\sigma^2}{z_{ij}})$ for all i= 1,2,…., ν and j=1,2,…,b. Consequently, $Y_{ij}|Z_{ij}$ follows normal with mean $\mu + \alpha_i + \beta_j$ and variance $\frac{\sigma^2}{z_{ij}}$ for all i= 1,2,…., ν and j=1,2,…,b with the density function.

$$f(y_{ij}|z_{ij}) = \frac{(z_{ij})^{\frac{1}{2}}}{\sigma\sqrt{2\pi}} exp\left\{-\frac{z_{ij}(y_{ij}-\mu-\alpha_i+\beta_j)^2}{2\sigma^2}\right\} \tag{2}$$

We further assume that the distribution of Zij as standard exponential. That is,

$$g(z_{ij}) = exp\{-z_{ij}\} \quad \forall \quad i,j \tag{3}$$

Then, using (2) and (3), the joint distribution of Y_{ij} and Z_{ij} is,

$$f(y_{ij}, z_{ij}) = \frac{(z_{ij})^{\frac{1}{2}}}{\sigma\sqrt{2\pi}} exp\left\{-\frac{z_{ij}\left((y_{ij}-\mu-\alpha_i-\beta_j)^2+2\sigma^2\right)}{2\sigma^2}\right\} \tag{4}$$

Integrating above with respect to Z_{ij} , we get,

$$f(y_{ij}) = \frac{\sigma^2}{\left((y_{ij}-\mu-\alpha_i-\beta_j)^2+2\sigma^2\right)^{\frac{3}{2}}} \tag{5}$$

Using (4) and (5) we get the following conditional distribution of Z_{ij} given Y_{ij}

$$f(z_{ij}|y_{ij}) = \frac{(z_{ij})^{\frac{1}{2}}\left((y_{ij}-\mu-\alpha_i-\beta_j)^2+2\sigma^2\right)^{\frac{3}{2}}}{\sigma^3\sqrt{2\pi}} exp\left\{-\frac{z_{ij}\left((y_{ij}-\mu-\alpha_i-\beta_j)^2+2\sigma^2\right)}{2\sigma^2}\right\} \tag{6}$$

Therefore,

$$E(z_{ij}|y_{ij}) = \int_0^\infty z_{ij}\frac{(z_{ij})^{\frac{1}{2}}\left((y_{ij}-\mu-\alpha_i-\beta_j)^2+2\sigma^2\right)^{\frac{3}{2}}}{\sigma^3\sqrt{2\pi}} exp\left\{-\frac{z_{ij}\left((y_{ij}-\mu-\alpha_i-\beta_j)^2+2\sigma^2\right)}{2\sigma^2}\right\} dz_{ij}$$

By solving above integral,

$$E(z_{ij}|y_{ij}) = \left(\frac{3\sigma^2}{(y_{ij}-\mu-\alpha_i-\beta_j)^2+2\sigma^2}\right) \tag{7}$$

It can be easily seen that $E\left(E(z_{ij}|y_{ij})\right) = 1$

Maximum Likelihood Estimates
From the joint distribution given in equation (4), the likelihood function is given by,

$$L(\mu, \alpha_i, \sigma|y_{ij}, z_{ij}) = \frac{(z_{ij})^{\frac{n}{2}}}{(\sigma\sqrt{2\pi})^n} exp\left\{-\frac{\sum_{i,j}\left(z_{ij}\left((y_{ij}-\mu-\alpha_i-\beta_j)^2+2\sigma^2\right)\right)}{2\sigma^2}\right\} \tag{8}$$

$$\frac{\partial(logL)}{\partial\alpha_i} = 0 \quad \Rightarrow \quad \widehat{\alpha_i} = \frac{\sum_{j=1}^{b}(z_{ij}Y_{ij})}{z_{i.}} - \hat{\mu} \quad \forall\, i = 1,2,\dots,v \tag{9}$$

$$\frac{\partial(logL)}{\partial\beta_j} = 0 \quad \Rightarrow \quad \widehat{\beta_j} = \frac{\sum_{i=1}^{v}(z_{ij}Y_{ij})}{z_{.j}} - \hat{\mu} \quad \forall\, j = 1,2,\dots,b \tag{10}$$

$$\frac{\partial(logL)}{\partial\mu} = 0 \quad \Rightarrow \quad \hat{\mu} = \frac{\sum_{i,j}(z_{ij}Y_{ij})}{z..} \; provied\ that \sum_i(\alpha_i z_{i.}) = \sum_i(\beta_j z_{.j}) = \sum_{i,j}(\alpha_i\beta_j z_{ij}) = 0 \tag{11}$$

$$\frac{\partial(logL)}{\partial\sigma^2} = 0 \quad \Rightarrow \quad \widehat{\sigma^2} = \frac{\sum z_{ij}\left(y_{ij} - \hat{\mu} - \widehat{\alpha_i} - \widehat{\beta_j}\right)^2}{n} \tag{12}$$

Where

$$z_{i.} = sumof\ all\ z_{ij}\ receiving\ i^{th} tratement = \sum_j z_{ij} \quad \forall i = 1,2,3,\dots,v$$

$$z_{.j} = sumof\ all\ z_{ij}\ in\ j^{th} block = \sum_i z_{ij} \quad \forall\ j = 1,2,3,\dots,b$$

$$z.. = \sum_{i,j} z_{ij} = Sum\ of\ all\ z_{ij}$$

Algorithm to Compute $E(Z_{ij}|y_{ij})$

1. Enter the values of y_{ij} for all i=1,2,3,…,v and j=1,2,3,…,b
2. Initially consider all $z_{ij} = 1$ for all (i, j). Also compute $z_{i.}$ and $z_{j.}$ as the i^{th} treatment and j^{th} block total for all (i, j) and z as the sum of all z_{ij}.
3. Use values of y_{ij} and z_{ij} to obtain maximum likelihood estimates of the model parameters α, μ, β and σ by using the equations given in (9), (10), (11) and (12).
4. Use the given y_{ij} and estimated values of model parameter viz., α, μ, β and σ and find $E(Z_{ij}|y_{ij})$ by using relation given in equation (7) and substituting the unknown parameter by their estimates.
5. Again use the $E(Z_{ij}|y_{ij})$ and compute the maximum likelihood estimates of the model parameters α, β, μ and σ for given y_{ij}.
6. Repeat the (vi) until mean of all values of Z_{ij} is 1.
7. Once these $E(Z_{ij}|Y_{ij})$ are predicted then form new Z_{ij} corresponding to each Y_{ij} by using the relation $Z_{ij} = Z_{i.} * Z_{.j} * (Z..)^{-1}$
8. Use these new values of Zij to construct ANOVA table.

Construction of ANOVA Table

Let us consider the linear model assumed in equation (1); $y_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij}$ for all i= 1,2,…., v and j=1,2,…,b

As considered earlier, for given z_{ij} (obtained from algorithm) for all i= 1,2,…., v and j=1,2,…,b,

$$\epsilon_{ij}|Z_{ij} = \left(Y_{ij} - \mu - \alpha_i - \beta_j\right) \sim N\left(0, \frac{\sigma^2}{Z_{ij}}\right) \; \forall\ i = 1,2,\dots,v\ and\ j = 1,2,\dots,b$$

Our derivation, matches with the approach followed in general least square theory discussed in Rao (2001) [6].

$$\left(\frac{\epsilon_{ij}\ z_{ij}}{\sigma^2}\right) = \left(\frac{Z_{ij}(y_{ij} - \mu - \alpha_i - \beta_j)}{\sigma^2}\right) \sim N(0,1)$$

$$\left(\frac{Z_{ij}\ \epsilon_{ij}^2}{\sigma^2}\right) \sim \chi^2_{(1)} \quad and \quad \sum_{(i,j)}\left(\frac{Z_{ij}\ \epsilon_{ij}^2}{\sigma^2}\right) \sim \chi^2_{(n-1)}$$

Therefore, the sum of squares due to error (SSE) is given by,

$$SSE = \sum_{(i,j)} z_{ij}\left(y_{ij} - \hat{\mu} - \widehat{\alpha_i} - \widehat{\beta_j}\right)^2 = \sum_{(i,j)} Z_{ij}\left(y_{ij} - \hat{\mu} - \left(\frac{\sum_{j=1}^{r_i}(Z_{ij}Y_{ij})}{Z_{i.}} - \hat{\mu}\right) - \frac{\sum_{i=1}^{v}(z_{ij}Y_{ij})}{z_{.j}} - \hat{\mu}\right)^2 \tag{13}$$

Let us consider,

$$\bar{Y}_{..}^{w} = \hat{\mu} = \frac{\sum_{i,j}(z_{ij}Y_{ij})}{Z_{..}} , \quad \bar{Y}_{i.}^{w} = \frac{\sum_{j=1}^{b}(z_{ij}Y_{ij})}{Z_{i.}} \ \forall\, i, \quad \bar{Y}_{.j}^{w} = \frac{\sum_{i=1}^{v}(z_{ij}Y_{ij})}{Z_{.j}} \ \forall\, j \tag{14}$$

Using (14) in (13) and simplifying, we get,

Using (14) in (13) and simplifying, we get, $SSE = \sum_{(i,j)} z_{ij}\left(y_{ij} - \bar{Y}_{..}^{w}\right)^2 - \sum_i z_{i.}(\bar{Y}_{i.}^{w} - \bar{Y}_{..}^{w})^2$

Therefore,

$$TSS = \sum_{(i,j)} z_{ij}\left(y_{ij} - \bar{Y}_{..}^{w}\right)^2 \quad SSB = \sum_{j} z_{.j}\left(\bar{Y}_{.j}^{w} - \bar{Y}_{..}^{w}\right)^2 \quad and \quad SST = \sum_{i} z_{i.}(\bar{Y}_{i.}^{w} - \bar{Y}_{..}^{w})^2 \tag{15}$$

For algebraic computation, we simplify the different sum of squares given in equation (14) as follow,

$$TSS = \sum_{(i,j)} z_{ij}y_{ij}^2 - (CF)_w \quad SST = \sum_i\left(\frac{\left(\sum_j z_{ij}Y_{ij}\right)^2}{Z_{i.}}\right) - (CF)_w \quad and\ SST = \sum_j\left(\frac{\left(\sum_i z_{ij}Y_{ij}\right)^2}{Z_{.j}}\right) - (CF)_w \tag{15}$$

Where,

$$CF_w = \frac{G_w{}^2}{z_{..}} = \frac{\left(\sum_{(i,j)} z_{ij}Y_{ij}\right)^2}{z_{..}} \tag{16}$$

Example: A study was conducted to know effect of four drugs on the size of the prostate (in CC). A group of 16 BPH (Benign Prostatic Hyperplasia) patients were selected for the study whose prostate size is in between (60-65 CC). Further they were divided in to the four groups (blocks) according to their BMI value, (< 20, 20-25, 25-30 and >30). After 15 days of treatment again the prostate size was measured as given in the following Table (1).

Table 1: Prostate Size (In CC) of BPH

Treatment	**Gr 1**	**Gr 2**	**Gr 3**	**Gr 4**	**Mean**
Drug 1	43.5	40	45	43.5	43.00
Drug 2	33.5	47.5	41	45	41.75
Drug 3	36.5	41.5	40.5	43	40.37
Drug 4	39	34.5	36.5	37.5	36.87
Mean	38.125	40.875	40.75	42.25	42.25

To know the significant difference between treatment effects (Effect of Drug), we need to conduct analysis randomized block design. To answer the above questions, we need to carryout ANOVA provided following assumptions are valid.

1. Homogeneity of variance between the groups
2. Error must be normally distributed.

As we know that Bartlett test is the commonly used test for the testing homogeneity of variance when errors are normal and the Leven test for any distribution and one sample Kolmogorov-Smirnov test (KS-test) or Anderson Darling test (AD test) is used for normality. Using Minitab statistical software, we carry out these two tests. Following are the Minitab output.

Output: (1)				
Test for Equal Variances: 95% Bonferroni confidence intervals for standard deviations				
Treatment	N	Lower	StDev	Upper
Drug1	4	1.04515	2.12132	12.7213
Drug2	4	3.01374	6.11692	36.6825
Drug3	4	0.92993	1.88746	11.3189
Drug4	4	1.36974	2.78014	16.6722

Bartlett's Test (Normal Distribution)
Test statistic = 5.03, p-value = 0.170

Levene's Test (Any Continuous Distribution)
Test statistic = 1.87, p-value = 0.188

Output: (2)					
Test for equality of means					
Source	DF	SS	MS	F	P
Drug	3	83.875	27.95	2.03	0.180
Group	3	35.625	11.87	0.86	0.495
Error	9	124.0	13.78		
Total	15	243.50			

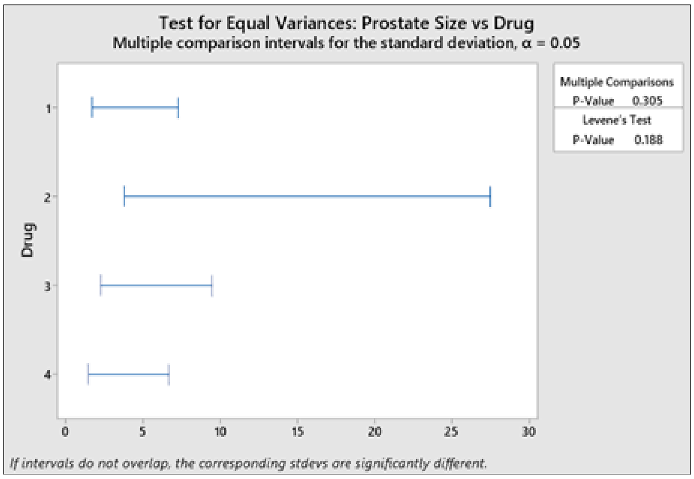

Fig 1: Test for Equal Variance

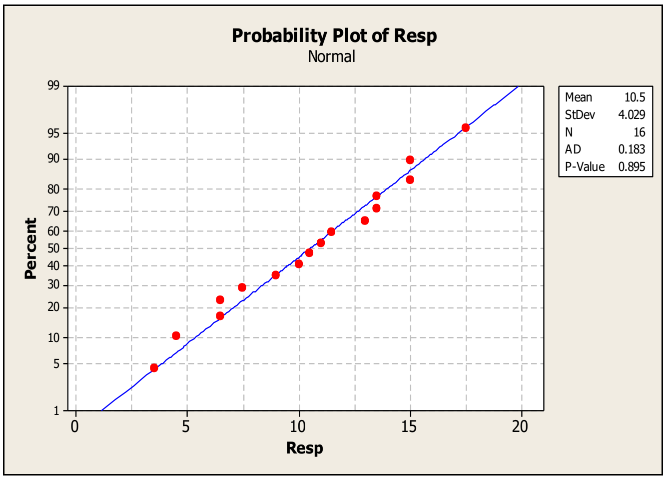

Fig 2: Test for Normality

So from the above Minitab output; p-value of Bartllet's Test and Leven's Test is 0.170 and 0.188 (output: (1) and Fig. (1) ensures the homogeneity and p-value of AD test is 0.895 (Fig 2) ensures the normality of the data. Also from the output (2) of Minitab; test for equality of means p-value 0.180 which is greater than the 0.05 significance level, we do not reject the null hypothesis that the average prostate size for the different drugs are all equal.

In this study, we can say that there is hidden source of variability related to each unit under study. This hidden variability may be the genetic factor, intellectual level (not graded is one point), background they came (place, religion), environmental condition, eating habits etc. These hidden sources of variability cause not to detect significant difference among four drugs. This needs to be extracted, so that detect signal can be rightly detected. The proposed model addresses the above situation.

The estimated values for the Z_{ij} using algorithm of $E(Z_{ij}|Y_{ij})$ is given in the following Table (2).

Table 2: Estimated $Z_{ij}|Y_{ij}$ for example

Drug	**$E(Z_{ij}\| Y_{ij})^*$**				**$Z_{i.}$**
	Gr 1	**Gr 2**	**Gr 3**	**Gr 4**	
Drug1	1.474853	1.459228	0.421587	1.285535	4.641203
Drug2	0.052956	0.148828	0.887098	1.464548	2.553429
Drug3	0.17402	0.686193	1.492298	1.475795	3.828305
Drug4	0.885607	1.496653	1.336204	1.258597	4.977062
$Z_{.j}$	2.587436	3.790902	4.137186	5.484475	$Z_{..}$ = 16
α_i	2.51791	2.157621	0.40633	-3.76749	
β_j	-1.7113	-0.69573	0.100176	1.21266	
μ	10.54069				

$E(Z_{ij}|Y_{ij})$ obtained on 116 iteration

Using equation (15) and (16), and estimated frailty random variable Z_{ij}, we can compute different sum of squares. Therefore, constructed ANOVA according to new criterion is given as follow.

Table 3: ANOVA based on frailty model approach

SV	**SS**	**D.F.**	**MSS**	**F-Value**	**P-Value**
Drug	112.588	3	37.52933	4.0938	0.04348
Group	17.51878	3	5.839593	0.6369987	0.6098
Error	82.50621	9	9.167357	---	---
Total	212.613	15	---	---	---

Since the p-value for drug effect in ANOVA based on frailty model approach (Table (3) is 0.04348 which is less than 0.05 (LOS), we reject the null hypothesis. Therefore, there is significant difference between the average prostate vol. (CC) obtained from four different drugs.

Conclusion

We observed from the regular approach of conducting ANOVA for the randomized block design, to know the significance of treatment effects is not able to defect the apparent differences because of heterogeneity and carrying hidden source of variability with individual observation. This fails to detect this smaller differences which lead to very important conclusions and infer a lot about drug comparisons. Developed frailty approach to the randomized block design able to detect this difference clearer and lead to conclusive statement for use of drug to reduce the prostate volume (CC). This will be very useful for the practitioners to come up with the correct idea about drug in BPH patients.

Conflict of Interest: None

Acknowledgement: I would like to acknowledge Muljibhai Patel Urological hospital, Nadiad, Gujarat, India for continuous support during the research study.

References

1. Shanubhogue A, Shete N. A Frailty Model Approach to the Completely Random Design, *PRAJÑĀ* - Journal of Pure and Applied Sciences. 2018;26:47-52.
2. Beard RE. Note on some mathematical mortality models. The Lifespan of Animals. G.E.W. Wolstenholme, M.O' Conner (Ed., Ciba Foundation Colloquium on Ageing, Little, Brown, Boston; c1959. p. 302-311.
3. Lancaster T. Econometric methods for the duration of unemployment, Econometrica. 1979;47:939-956.
4. Vaupel JW, Manton KG, Stallard E. The impact of heterogeneity in individual frailty on the dynamics of mortality, Demography. 1979;16:439-454.
5. Douglas C, Montogomery. Design and Analysis of Experiments, 7th Ed, John Wiley; c2012.
6. Rao CR. Linear Statistical Inference and it's Applications, Wiley; c2001.