Journal of the Nigerian Statistical Association Vol. **28**, 2016, 121– 125

Sample Size Effect on Variance of Treatment Mean and Relative Efficiency in a Split-Plot Design

A. V. Oladugba[a], E. O. Ossai[b] and E. E. Umanah[c]

[a,b]*Department of Statistics, University of Nigeria, Nsukka*; [c]*Ministry of Economic Development, Idongesit Nkanga Secretariat Complex, Uyo, Akwa Ibom State*

In this paper, we derived an expression for the variance of treatment means in a split plot design with sampling using the variance components associated with the whole plot, split plot and sampling errors. The result from the data analysis shows that relative efficiency increases with increase in sample size while variance of treatment means decreases with increase in sample size.

Keywords: Sample size; variance of treatment mean; relative efficiency; sampling error; variance components; sampling size.

1. Introduction

Sample size is an imperative feature of any empirical study in which the goal is to make inferences about a population from a sample. It is important that a study has an adequate sample size. It must be large enough that an effect of such magnitude as to be of scientific significance will also be statistically significant. It is just as important, however, that the study not be too large where an effect of little scientific importance is nevertheless statistically detectable (Russell, 2001). This is necessary to ensure that the study has a good chance of detecting a statistically significant result and also to ensure that adequate resources are allocated. A study that has an inadequate sample size will have a low probability of detecting a statistically significant result and therefore represents a waste of valuable resources and add nothing to scientific knowledge. In practice, the sample size used in a study is determined based on the expense of data collection, and the need to have sufficient statistical power (Cohen, 1977). Increasing sample size leads to increase in statistical power and decrease in error rate (Caimiao et al., 2004). A sufficiently large sample size leads to more precise information, avoids the bias obtained by choosing a non-representative subset of samples, and also increases the reliability of conclusions, thus reducing uncertainty in results (Russell, 2001). On the other hand, an insufficient sample size will give a result which may not be sufficient to detect a difference between the groups and the study may own out to be incorrect leading to a type II error (Jonathan et al., 2005). It also wastes time and money as the result will be variably inconclusive. The sample size is decided arbitrarily based on the researchers convenience, available time, and resources, resulting in lack of precision due to insufficient number of subjects studied (Barun, 2010).

The variance of treatment means or error is the average of the square variations of each population mean from the grand mean. Braun (2012) described it the variation which exists between treatment means. When populations are grouped, it helps to decide if variability between and within each populations are significantly different as evaluated by Martin and Naomi (2014), and Anwesha (2012). Efficiency is a term used in the comparison of various statistical procedures and, in particular, it refers to a measure of the optimality of an estimator of an experimental design, or of a hypothesis testing procedure. The relative efficiency of two experimental designs is the ratio of their efficiency statistics, although this term is often used where the comparison is made between a given design and a notional 'best possible' design (Nikulin, 2001). The relative efficiency of two designs theoretically depends on the available sample size. This implies that there is a significant effect of sample size on the relative efficiency of experimental designs. The objectives of this paper are to derive an expression for the variance of treatment means in a split plot design, and also to determine the effect of sample size on variance of treatment means and relative efficiency.

2. Definition of Split-Plot Design with Sampling

Suppose an experiment was carried out with r replications of a whole-plot treatments (A) and another set of b treatments referred to as the split-plot treatments (B) were randomly assigned to the ar whole-plot and n random samples are taken from each of the abr plots. The model for the observation denoted by Y_{ijkl} of the lth sample from the ith whole-plot and jth split-plot treatments on the kth replication is given by:

Model:

$$Y_{ijkl} = \mu + \rho_k + A_i + E_{ik} + B_j + (AB)_{ij} + E_{ijk} + E_{ijkl} \tag{1}$$

where Y_{ijkl} = observed response from the lth sampling of the jth split plot treatment, ith whole plot treatment and the kth replication; μ = overall mean; ρ_k effect of the kth replication; A_i = ith level of the whole plot treatment; E_{ik} = whole plot error; B_j = jth level of the split plot treatment; $(AB)_{ij}$ = interaction between whole plot and split-plot treatments; E_{ijk} = split plot error and E_{ijkl} = the sampling error. The components E_{ik}, E_{ijk} and E_{ijkl} in (1) are random variable with means equal to zero and variances equal to σ_a^2, σ_b^2 and σ_n^2 respectively, known as variance components.

The Analysis of Variance (ANOVA) table resulting from this experiment is presented in Table 1; where E_1, E_2 and E_3, are mean square for the whole plot, split plot and sampling, respectively, and a denotes the whole-plot level, r denotes the number of replication, n denotes the size of the sampling and b is the split plot level.

Table 1. Analysis of variance table for split-plot design with sampling

Source of Variation	Degrees of Freedom	Mean Square	Expected Mean Square
Replications	$r-1$		
Whole plot (A) treatment	$a-1$		
Whole plot Error (E_1)	$(r-1)(a-1)$	E_1	$bn\sigma_a^2 + n\sigma_b^2 + \sigma_n^2$
Split plot (B) treatment	$(b-1)$		
Interaction between whole plot and split plot treatments (AB)	$(a-1)(b-1)$		
Split plot Error (E_2)	$a(b-1)(r-1)$	E_2	$n\sigma_b^2 + \sigma_n^2$
Sampling errors (E_3)	$abr(n-1)$	E_3	σ_n^2
Total	$abrn-1$		

3. Estimation of Variance Components, Variance of Treatment Mean and Relative Efficiency

3.1 *Variance components*

The variance components σ_a^2, σ_b^2 and σ_n^2 are estimated using the ANOVA method. The ANOVA method of estimating variance components consists of equating mean squares to their respective expected values and solving the resultant equations for the variance components (Sahai and Ojeda, 2003). From Table 1, equating mean squares to their respective expected mean squares we have: $E_1 = \hat{\sigma}_n^2 + n\hat{\sigma}_b^2 + bn\hat{\sigma}_a^2$; $E_2 = \hat{\sigma}_n^2 + n\hat{\sigma}_b^2$ and $E_3 = \hat{\sigma}_n^2$. These give the following estimators:

$$\hat{\sigma}_n^2 = E_3 \tag{2}$$

$$\hat{\sigma}_b^2 = \frac{E_2 - E_3}{n} \tag{3}$$

$$\hat{\sigma}_a^2 = \frac{E_1 - E_2}{bn} \tag{4}$$

where $\hat{\sigma}_n^2$, $\hat{\sigma}_b^2$ and $\hat{\sigma}_a^2$, respectively are the estimate of the variance component for sampling, split plot and whole plot.

3.2 *Variance of treatment means*

Based on the assumptions made for the random variables, E_{ik}, E_{ijk} and E_{ikjl}, in (1), we obtain the estimate for variance of treatment means denoted by V as:

$$V = \frac{\hat{\sigma}_a^2}{ar} + \frac{\hat{\sigma}_b^2}{ar} + \frac{\hat{\sigma}_n^2}{arn} \tag{5}$$

substituting (2), (3), and (4) into (5), the estimated variance of treatment mean per sample basis will be:

$$V = \frac{E_1 + E_2(b-1)}{abrn} \tag{6}$$

Similarly, by varying the values of n sample per plot the estimated variance of treatment mean denoted by V_n is:

$$V_n = \frac{\hat{\sigma}_a^2}{ar} + \frac{\hat{\sigma}_b^2}{ar} + \frac{\hat{\sigma}_n^2}{arn'} = \frac{E_1 + E_2(b-1)}{abrn} + \frac{E_3}{ar}\left[\frac{1}{n} - \frac{1}{n'}\right] \tag{7}$$

where n' is the altered values of n.

3.3 *Relative efficiency*

The relative efficiency (E) is the amount of gain in information by sampling (V_n) relative to complete observation (V), that when n is one. This is obtained by taking the ratio of the amount of information by sampling $(V_n)^{-1}$ to the amount of information from complete observation $(V)^{-1}$; this results to:

$$E = \frac{V}{V_n} \tag{8}$$

where V is the variance of treatment mean (complete observation) and V_n is the sampling variance of treatment mean.

4. Results and Discussion

The data used in this work were adopted from (Montgomery, 2001). The experiment was carried out in a split plot design with three replicates to study the effect of three different pulp preparation methods (whole-plot treatment) and four different cooking temperature (split plot treatment) for the pulp on the tensile strength of paper. The data are shown in Table 2 (see Appendix), where the number of sampling $n = 1$, $a = 3$, $b = 4$ and $r = 3$. In order to determine the effect of sampling size on variance of treatment means and relative efficiency, we use data in Table 2, by increasing the values of n. Using (2), (3), (4), (6), (7) and (8), the results in Table 3 (see Appendix) were obtained.

The results in Table 3 show that when the size of the sampling unit equals one (1), the variance component of the sampling unit becomes zero (0), which makes the effect of the sample size on the variance of treatment mean almost insignificant. When the size of the sampling unit is greater or equal to two (2), the effect of the sample size becomes clear, because as the size of the sampling unit increases from two (2), and above, the variance of treatment mean decreases significantly. Furthermore, the relative efficiency which is obtained by taking the ratio of various combinations of variance of treatment mean increases significantly as the sample size increases. This implies that there is a significant effect of sample size on the relative efficiency of experimental designs.

5. Conclusion

We conclude that in order to reduce variation among treatment means, and increase relative efficiency which the amount of gain in information in any given experiment, the sample size should be increased.

References

Anwesha, B. (2012). Cross cultural variance of Schizophrenia in symptoms, diagnosis and treatments. *George Town Journal of Health Science*, 2: 18–24.

Barun, K. N. (2010). Understanding the relevance of sample size calculation. *Indian J. Ophthalmology*, 58(6): 469–470.

Braun, W. J. (2012). Naive analysis of variance. *Journal of Statistical Education*, 20(2):1–17.

Caimiao, W., Jiangning L. and Roger E. B. (2004). Sample size for detecting differentially expressed genes in microarray experiments. *BMC Genomics*, 5: 87.

Cohen, J. (1977). *Statistical Power Analysis for the Behavioral Sciences*. New York: Academic press.

Jonathan, J. D., Petra, M. and Les, I. (2005). The performance of tests of publication bias and other sample size effects in systematic reviews of diagnostic test accuracy was assessed. *Journal of Clinical Epidemiology*, 58(9) : 882-893.

Montgomery, D. C. (2001). *Design and Analysis of Experiment* (Fifth Edition). John Wiley and Sons, Inc. New York.

Martin, K. and Naomi A. (2014). Point of significance: analysis of variance and blocking. *Nature Methods*, 11, 699-700.

Russell, V. (2001). Some practical guidelines for effective sample size determination. *American Statistician*, 55(3), 187–193

Nikulin, M.S. (2001). *Encyclopedia of Mathematics.* Springer (Place of publication).

Sahai, H. and Ojede, M. M. (2003). *Analysis of Variance for Random Models.* Volume 1: Balanced Data. Springer Science+Business Media, LLC.

Appendix

Table 2: The experiment on the tensile strength of paper

	Replicate 1			Replicate 2			Replicate 3		
Preparation method	1	2	3	1	2	3	1	2	3
Temperature (°F)									
200	30	34	29	28	31	31	31	35	32
225	35	41	26	32	36	30	37	40	34
250	37	38	33	40	42	32	41	39	39
275	36	42	36	41	40	40	40	44	45

Table 3: Estimated variances of treatment mean and relative efficiency for specified sampling size

Sampling size	Total observation	Variance Component			Variance of treatment mean	Relative efficiency (E)
n	N	σ_a^2	σ_b^2	σ_n^2	V_n	$E = \frac{V}{V_n}$
1	36	1.561	2.824	0	0.4872	1
2	72	6.8699	7.8525	4.031	1.642	0.2967
3	108	1.3	4.39	4.116	0.6365	0.7654
4	144	0.4034	0.5526	7.2285	0.1118	4.3578
5	180	0.324	0.1536	5.7661	0.0206	23.65
6	216	0.2753	0.1331	5.4726	0.0103	47.301
7	252	0.1449	0.0763	6.9271	0.0052	93.692
8	288	0.1386	0.0583	7.9633	0.0028	174
9	324	0.0962	0.0322	8.4272	0.0013	374.77
10	360	0.0954	0.0196	8.7973	0.0007	696