Thailand Statistician
January 2012; 10(1) : 121-128
http://statassoc.or.th
Contributed paper

Expected Mean Squares for the Random Effects One-Way ANOVA Model when Sampling from a Finite Population

Teerawat Simmachan [a] John J. Borkowski [b] and Kamon Budsaba* [a]

[a] Department of Mathematics and Statistics, Faculty of Science and Technology, Thammasat University, Pathum Thani 12120, Thailand.

[b] Department of Mathematical Sciences, Montana State University, Bozeman, MT 5971, USA.

* Author for correspondence; e-mail : kamon@mathstat.sci.tu.ac.th

Received: 15 February 2012

Accepted: 16 March 2012

Abstract

The purpose of this work is to determine the expected value of the mean square error (MS_E) and the expected value of the treatment mean square (MS_{trt}) for the random effects one-way ANOVA model assuming a finite population.

For the case of balanced data (equal sample sizes), both the expected value of the mean square error and the expected value of the treatment mean square for the finite population are the same as that for the infinite population.

For the case of unbalanced data, the expected value of MS_E for the finite population is equal to that for the infinite population which is also the same as the expected value for the balanced case. On the other hand, the expected value of MS_{trt} for the finite population is different from that for the infinite population because of the different multiplier values of the population variance (σ_τ^2).

Keywords: expected mean squares, finite population, finite population correction, random effects model, variance components.

1. Introduction

Analysis of variance (ANOVA) is probably the most frequently applied of all statistical analyses. ANOVA is extensively applied in many fields of research, such as anthropology, biology, commerce, economics, education, industry, medicine, political science, psychology, sociology, and etc. One reason for the popularity of ANOVA is its suitability for many different types of study design. ANOVA places no restriction on the number of groups or conditions that may be compared. Another reason of the frequency of ANOVA applications is that it is suitable for most effect comparisons by testing for differences between means [1].

One of the most basic designed experiments is the one-factor Completely Randomized Design (CRD). Data from the CRD is typically analyzed using a one-way ANOVA model associated with either a fixed effects model or a random effects model [2]. In this study, we especially consider the random effects model having a finite population of an effect in the model having been sampled at random without replacement.

Finite populations for variance components models have been considered in various cases; for example, Bennett and Franklin [3] discussed finite nested populations, but only for balanced data, Cornfield and Tukey [4] and Tukey [5] discussed them for balanced data. Nevertheless, for unbalanced data, Tukey [6] studied the variances of variance components for single classification, Gaylor and Hartwell [7] and Mahamunulu [8] dealt in detail with the 3-way nested classification. Searle and Fawcett [9] presented expected mean squares in variance components models having populations of finite size. They developed a rule for converting expectations under infinite models into expectations under finite models. It was assumed that the levels of each factor are finite.

Note that all authors mentioned above have only considered that population sizes of all effects are finite. They even assumed that the population of error terms is finite. However, we are not assuming this about the errors which are a random sample from a $N(\mu+\tau_i,\sigma^2)$ distribution. For example, the k machines are selected from a finite population but the n replications taken within each machine are normally distributed (not finite).

In this paper we focus on the random effects one-way ANOVA model. Throughout the subsequent sections, we consider the case where the population of treatment effect (τ_i) is finite. We use the similar approach of Searle and Fawcett for the finite population of the random effect, but we have to adjust the results for a normal error distribution, and the finite population correction (fpc) should be used.

Finally, we investigate the expected values of the mean square error (MS_E) and the treatment mean square (MS_{trt}) for the random effects one-way ANOVA model when sampling from a finite population. Section 2 presented materials and methods used to find the expected mean squares. The results were shown in Section 3 and we discussed the results in Section 4.

2. Materials & Methods

In this section, we describe the research methodology of this study in detail as follows:

2.1 Random Effects Model

For random effects, in theory, we assume the population is infinite [10]. In practice, it is acceptable if the number of randomly selected factor levels (k) is small relative to the number of levels in the population (N), that is the sampling fraction k/N should not exceed 5% [11].

The equation for the random effects model for a single factor CRD is:

$$y_{ij} = \mu + \tau_i + \varepsilon_{ij} \tag{1}$$

where $i = 1,...,k$ are levels of a factor, and *j* is the replication ($j = 1,2,...,n_i$) for the i^{th} factor level, and both τ_i and ε_{ij} are random variables whose distributions we have to specify.

2.2 Derivation of the Expected Mean Squares Assuming a Finite Population

The expectations of the mean squares are obtained by substituting the model (1) in mean squares and taking expected values. However, when sampling from a finite population, some assumptions are no longer true. The assumptions can be replaced with:

- The treatment effects τ_i are selected from a finite population of size *N* with mean 0 and population variance σ_τ^2.
- The covariance between every different pair of random effects is nonzero. That is, $\operatorname{cov}(\tau_i, \tau_{i'}) \neq 0$ for the random τ_i and $\tau_{i'}$ terms.

We still assume the homogeneity (equal) variance assumption within the ith level.

In Gaylor and Hartwell [7] and Searle and Fawcett [9], it is assumed that the mean of each population is zero, so that

$$\sum_{i=1}^{N}\tau_i = 0 \text{ and } \sigma_\tau^2 = \sum_{i=1}^{N}\tau_i^2 /(N-1) \tag{2}$$

is defined as the population variance.

Consequently, $(\sum_{i=1}^{N}\tau_i)^2 = \sum_{i=1}^{N}\tau_i^2 + \sum_{i\neq i'}^{N}\sum_{i\neq i'}^{N}\tau_i\tau_{i'} = 0$;

$$\sum_{(i,i'\in T)}\tau_i\tau_{i'} = -(N-1)\sigma_\tau^2, \tag{3}$$

where set $T=\{(i,i'): i = 1,\ldots,N,\ i' = 1,\ldots,N,\ i \neq i'\}$. Note there are $N(N-1)$ ordered pairs in set T. We assume that the finite population of effects in the model has been sampled at random without replacement. If τ_r is a random sampled value of the effects then, because of (2),

$$\text{mean}(\tau_r) = E(\tau_r) = \sum_{i=1}^{N}\tau_i P(\tau_i) = \sum_{i=1}^{N}\tau_i(\frac{1}{N}) = \frac{1}{N}\sum_{i=1}^{N}\tau_i = 0;$$

and

$$\text{var}(\tau_r) = E(\tau_r^2) = \frac{1}{N}\sum_{i=1}^{N}\tau_i^2 = (\frac{N-1}{N})\sigma_\tau^2 \tag{4}$$

and by (3), for two sample values τ_r and τ_s,

$$\text{cov}(\tau_r\tau_s) = E(\tau_r\tau_s) = \frac{1}{N(N-1)}\sum_{i\neq i'}^{N}\sum_{i\neq i'}^{N}\tau_i\tau_{i'} = -\frac{\sigma_\tau^2}{N}. \tag{5}$$

Since the corresponding values of (4) and (5) for an infinite population are σ_τ^2 and 0, respectively, the expected values of mean squares in finite population models are not the same as with the infinite population. In either case the expected values are linear functions of the variance components; the coefficients are determined for the finite population models in accord with (4) and (5).

2.3 Some Useful Results Involving Expectations for Balanced Data

Result 1: $E(\tau_i \varepsilon_{ij}) = 0$ because τ_i and ε_{ij} are independent for all *i* and *j*.

Result 2: $E\left[\left(\sum_{j=1}^{n}\varepsilon_{ij}\right)^2\right] = n\sigma^2$.

Proof: $$E\left[\left(\sum_{j=1}^{n}\varepsilon_{ij}\right)^2\right] = \text{var}\left(\sum_{j=1}^{n}\varepsilon_{ij}\right) + \left[E\left(\sum_{j=1}^{n}\varepsilon_{ij}\right)\right]^2$$

$$= \sum_{j=1}^{n}\text{var}(\varepsilon_{ij}) + \left(\sum_{j=1}^{n}E(\varepsilon_{ij})\right)^2 = \sum_{j=1}^{n}\sigma^2 + \left(\sum_{j=1}^{n}(0)\right)^2 = n\sigma^2 \;\square$$

Result 3: $E\left[\left(\sum_{i=1}^{k}\sum_{j=1}^{n}\varepsilon_{ij}\right)^2\right] = kn\sigma^2 = n_T\sigma^2$.

Proof: $$E\left[\left(\sum_{i=1}^{k}\sum_{j=1}^{n}\varepsilon_{ij}\right)^2\right] = \text{var}\left(\sum_{i=1}^{k}\sum_{j=1}^{n}\varepsilon_{ij}\right) + \left[E\left(\sum_{i=1}^{k}\sum_{j=1}^{n}\varepsilon_{ij}\right)\right]^2$$

$$= \sum_{i=1}^{k}\sum_{j=1}^{n}\text{var}(\varepsilon_{ij}) + \left(\sum_{i=1}^{k}\sum_{j=1}^{n}E(\varepsilon_{ij})\right)^2$$

$$= \sum_{i=1}^{k}\sum_{j=1}^{n}\sigma^2 + \left(\sum_{i=1}^{k}\sum_{j=1}^{n}(0)\right)^2 = kn\sigma^2 = n_T\sigma^2 \qquad \square$$

Result 4: $E\left[\left(\sum_{i=1}^{k}\tau_i\right)^2\right] = \frac{k(N-k)}{N}\sigma_\tau^2$.

Proof: $$E\left[\left(\sum_{i=1}^{k}\tau_i\right)^2\right] = \text{var}\left(\sum_{i=1}^{k}\tau_i\right) + \left[E\left(\sum_{i=1}^{k}\tau_i\right)\right]^2$$

$$= \sum_{i=1}^{k}\text{var}(\tau_i) + \sum_{(i,i'\in K),}\text{cov}(\tau_i, \tau_{i'}) + \left(\sum_{i=1}^{k}E(\tau_i)\right)^2,$$

where set $K = \{(i,i'): i = 1,\ldots,k,\ i' = 1,\ldots,k,\ i \neq i'\}$,

$$= \sum_{i=1}^{k} \frac{N-1}{N}\sigma_\tau^2 + \sum_{ii' \in K,} (-\frac{\sigma_\tau^2}{N}) + \left(\sum_{i=1}^{k} (0) \right)^2$$

$$= \frac{k(N-1)\sigma_\tau^2}{N} + k(k-1)(-\frac{\sigma_\tau^2}{N})$$

$$= \frac{k}{N}\left[(N-1)-(k-1)\right]\sigma_\tau^2$$

$$= \frac{k(N-k)}{N}\sigma_\tau^2 \qquad \square$$

2.4 Some Useful Results Involving Expectations for Unbalanced Data

The following useful results involving expectations are an extension of the results from section 2.3. Result 1 is the same as it appeared in section 2.3 and it is restated below.

Result 1: $E(\tau_i \varepsilon_{ij}) = 0$ because τ_i and ε_{ij} are independent for all *i* and *j*.

Result 2: $E\left[\left(\sum_{j=1}^{n_i} \varepsilon_{ij} \right)^2 \right] = n_i \sigma^2 .$

Result 3: $E\left[\left(\sum_{i=1}^{k} \sum_{j=1}^{n_i} \varepsilon_{ij} \right)^2 \right] = \sum_{i=1}^{k} n_i \sigma^2 = n_T \sigma^2 .$

Result 4: $E\left[\left(\sum_{i=1}^{k} n_i \tau_i \right)^2 \right]^2 = \frac{(N-1)}{N} \sigma_\tau^2 \left[\sum_{i=1}^{k} n_i^2 - \frac{1}{N-1} \sum_{(i,i' \in K)} n_i n_{i'} \right].$

3. Results

For the case of balanced data (equal sample sizes), the theoretical results are shown in Table 1 that both the expected values of MS_E and MS_{trt} for the finite population are the same as that for the infinite population.

For the case of unbalanced data (not all sample sizes are equal), the expected value of MS_E for the finite population is equal to that for the infinite population which is also the same as the expected value for the balanced case. On the other hand, the expected value of MS_{trt} for the finite population is different from that of the infinite population because of the different multiplier values of C_I and C_F. The results are shown in Table 2.

Table 1. Expectations of Mean Squares under both the Finite Population and the Infinite Population Assumptions for a Balanced CRD

Expectations	Infinite Population	Finite Population
MS_E	σ^2	σ^2
MS_{trt}	$\sigma^2 + n\sigma_\tau^2$	$\sigma^2 + n\sigma_\tau^2$

Table 2. Expectations of Mean Squares under both the Finite Population and the Infinite Population Assumptions for an Unbalanced CRD

Expectations	Infinite Population	Finite Population
MS_E	σ^2	σ^2
MS_{trt}	$\sigma^2 + C_I\sigma_\tau^2$	$\sigma^2 + C_F\sigma_\tau^2$

$$C_I = \frac{1}{k-1}\left(n_T - \frac{1}{n_T}\sum_{i=1}^{k} n_i^2\right), C_F = \frac{(N-1)}{(k-1)N}\left(n_T - \frac{1}{n_T}\left[\sum_{i=1}^{k} n_i^2 - \frac{1}{N-1}\sum_{(i,i'\in k)} n_i n_{i'}\right]\right).$$

The n and n_i are the numbers of replications per factor level for balanced and unbalanced data, respectively. The n_T is the total number of observations in the experiment.

4. Discussion

In this article, we have determined the expected mean squares for the random effects one-way ANOVA model assuming a finite population.

For the case of balanced data, both the expected value of the mean square error and the expected value of the treatment mean square for the finite population are the same as that for the infinite population. However, σ_τ^2 represents two different variances. In the infinite case, σ_τ^2 is the variance of a normally distributed random variable. In the finite case, σ_τ^2 is the variance of a finite population.

For the case of unbalanced data, the expected value of the mean square error for the finite population is equal to that for the infinite population which is also the same as the expected value for the balanced case. On the other hand, the expected value of the treatment mean square for the finite population is different from that for the infinite population because of the different multiplier values of the population variance (C_I and C_F). Also, for the infinite case, σ_τ^2 is the variance of a normally distributed random variable. For the finite case, σ_τ^2 is the variance of a finite population.

Note that the expected values of the mean squares are different from the results of Searle and Fawcett because we did not assume a finite population of the errors.

References

[1] Rutherford, A., Introducing ANOVA and ANCOVA: a GLM approach, London, SAGE Publications, Inc., 2001.

[2] Montgomery, D.C., Design and Analysis of Experiments (5th ed.), New York, John Wiley & Sons, Inc., 2001.

[3] Bennett, C.A. and Flanklin, N.L., Statistical Analysis in Chemistry and the Chemical Industry, New York, John Wiley & Sons, Inc., 1954.

[4] Cornfield, J. and Tukey, J.W., Average values of mean squares in factorials, Ann, Math, Statist, 1956; 27: 907-949.

[5] Tukey, J.W., Variances of variance components: I Balanced designs. Ann, Math, Statist, 1956; 27:722-736.

[6] Tukey, J.W., Variances of variance components: II The unbalanced single Classification, Ann, Math, Statist, 1957; 28:43-56.

[7] Gaylor, D.W. and Hartwell, T.D. Expected mean squares for nested Classifications, Biometrics, 1969; 25:427-430.

[8] Mahamunulu, D.M., Sampling variances of the estimates of variance components in the unbalanced 3-way nested classification, Ann, Math, Statist, 1963; 34:521-527.

[9] Searle, S.R. and Fawcett, R.F., Expected mean squares in variance components models having finite populations. Biometrics, 1970; 26:243-254.

[10] Searle, S.R., Casella, G. and Mcculloch, C.E., Variance Components, New Jersey, John Wiley & Sons, Inc., 1992.

[11] Cochran, W.G., Sampling Techniques, (3rd ed.), New York, John Wiley & Sons, Inc., 1977.