EXACT INVESTIGATION OF ALL EFFECTS FOR EXTENSIONS OF ONE-WAY ANOVA MODEL WITH FIXED EFFECTS

John E. WALSH
Southern Methodist University, Dallas, U.S.A.

ABSTRACT. — Consider the standard one-way analysis of variance model with fixed effects. The customary assumptions of zero mean, no correlation, and equal variance are made for the random « error » terms in this balanced model. The usual normality assumption is also made when tests or confidence regions are desired. Extensions are made of this one-way model by addition of further « error » terms of one or two kinds. The extended models apply to much more general situations than does the standard model. However, exact procedures are obtained for investigating all the effects that appear in the standard model, and for investigating subsets of these effects. The generality level of an extension depends on what effects are investigated. For several extensions, the customary results for the standard model remain applicable. Some procedures differing from the customary ones are used for the other extensions. An application of the standard model to rejection of outlying observations is described and shown to be usable for some extensions.

Introduction.

The balanced fixed effects model for one-way analysis of variance is

$$y_{jk} = \mu + \alpha_j + e_{jk} \tag{1}$$

where $j = 1, \ldots, J$ and $k = 1, \ldots, K$, with $J, K \geqslant 2$. Here, y_{jk} is an observed random variable, μ is a parameter, α_j is a parameter such that $\alpha_1 + \ldots + \alpha_J = 0$, and e_{jk} is an unobserved random variable. The e_{jk} are assumed to be uncorrelated with zero expectation and the same positive variance σ^2. They are also assumed to be from a normal distribution when something other than a point estimate is desired.

Model (1) provides a basis for investigating μ, σ^2, and the α_j by point estimates, significance tests, and confidence regions. Moreover, as shown next, it furnishes a basis for deciding on rejection of outlying observations. An attractive feature is that significance levels, confidence coefficients, unbiasedness of estimates, etc. are determined exactly.

* Research partially supported by Air Force Contract AFOSR F33615-71-C-1178 and by Mobil Research and Development Corporation. Also associated with ONR Contract N00014-68-A-0515 and NASA Grant NGR 44-007-028 and Dept. of Labor Grant 31-46-70-06.

Now, consider the outline of a basis for rejection of outlying observations when model (1) holds. Suppose that y_{jk*} is selected to be investigated as a possible outlier (without information about the observed values). That is, the possibility that e_{jk*} does not have the same distribution as the other e_{jk} is investigated. First, without knowledge of the observation values, divide the observations with this value of j into sets of size two, and zero or one set of size three (with y_{jk*} not in a set of size three). Any set of size three is converted to a set of size two, by adding two of its observations and dividing this sum by $\sqrt{2}$ (to yield one « observation »). The resulting « observation » is denoted by $y_{jk'}$, where k' is the smaller of the values for k in the two observations summed. For each set (now all of size two), a statistic of the type

$$y_{jk(1)} - y_{jk(2)}$$

is formed, where $k(1) = k^*$ for the set containing y_{jk*}. The resulting statistics are uncorrelated with zero mean and the same variance. Using the normality assumption, the statistic containing y_{jk*} can be investigated by a procedure for examining whether a specified (chosen without information about the observation values) observation, supposedly in a sample from a normal population with zero mean, is an outlier.

This paper gives seven extensions of the standard model (1). An extension is made by adding more « error » terms, of one or two kinds, to the standard model for y_{jk}. An extension occurs such that μ, σ^2, and one or more of the α_j can all be investigated by exact procedures. Some of the statistics for these investigations differ from those customarily used for model (1). Other extensions are developed for investigating σ^2, for investigating μ and σ^2, and for investigating σ^2 and one or more of the α_j. Only the customary results need be used for some of these extensions but some different procedures occur for the others. These extensions are based on an approach given in [1].

The extended models are usable for much more general situations than is the standard model. However, some extended models are much more generally applicable than others. The level of generality for an extended model depends on what is investigated. For example, the extended model for investigating all the types of effects is less generally applicable than models for investigating subsets of the types of effects (when there is no restriction on the eligibile procedures). Also, the generality level of an extended model depends on whether, or not, all the investigation procedures are to be those customarily used for model (1). The generality level decreases when the eligible procedures are limited to the customary ones.

The extensions of model (1) are stated and discussed in the next section. Investigation procedures for use with these various extensions are outlined in the final section.

Extended models.

Seven extensions of model (1) are given. These depend on which effects are to be investigated and on whether the statistical procedures are restricted to those customarily used for model (1).

First, consider the extended model when μ, σ^2, and one or more of the α_j are all to be investigated. Limitation to the results customarily used does not apply to this case. The model is

$$y_{jk} = \mu + \alpha_j + e_{jk} + e'_k \tag{2}$$

where μ, α_j, and e_{jk} have the same properties as for model (1). The additional random variables $e'_1, \ldots, e'_K$ must sum to zero but otherwise can have an arbitrary joint distribution. Also, $e'_1, \ldots, e'_K$ can have any allowable dependences with the e_{jk} (the dependence can be different for each combination of an $e'_{k'}$ with an e_{jk}). The values of the e'_k are contributions imposed on the observations by the experimental circumstances, with different e_{jk} possibly having different influence on the random values that occur for $e'_1, \ldots, e'_K$.

Second, consider the case where the eligible procedures are not restricted and investigation of σ^2 and one or more of the α_j is to occur. The extended model for this case is

$$y_{jk} = \mu + \alpha_j + e_{jk} + e_k^* \tag{3}$$

where, as throughout the remaining material, μ, α_j, and e_{jk} have the same properties as in model (1). The additional random variables $e_1^*, \ldots, e_K^*$ do not necessarily sum to zero and can have an arbitrary joint distribution. Also, they can have any permissible dependences with the e_{jk}. Model (3) is an extension of model (2).

Third, consider the case where the procedures are restricted to these customarily used for model (1) and investigation of σ^2 and one or more of the α_j is to occur. The extension for this case is

$$y_{jk} = \mu + \alpha_j + e_{jk} + e' \tag{4}$$

where e' can have an arbitrary distribution and any permissible dependences with the e_{jk} (the dependence can be different for each e_{jk}). This is the least general of the extensions considered and, for $K > 2$, is much less general than model (2).

Fourth, consider the case where the eligible procedures are not limited and both μ and σ^2 are investigated. The extended model is

$$y_{jk} = \mu + \alpha_j + e_{jk} + e'_k + e''_j \tag{5}$$

where $e'_1 + \ldots + e'_K = 0$ and $e''_1 + \ldots + e''_J = 0$ but otherwise the additional random variables can have an arbitrary joint distribution. Also they can have any allowable dependences with the e_{jk}. Model (5), which extends models (2) and (4), has a high level of generality.

Fifth, consider the case where the procedures are restricted to the customary ones for model (1) and investigation of both μ and σ^2 occurs. The extension is

$$y_{jk} = \mu + \alpha_j + e_{jk} + e''_j \tag{6}$$

where $e''_1 + \ldots + e''_J = 0$ but otherwise these random variables can have an arbitrary joint distribution. Also, $e''_1, \ldots, e''_J$ can have any permissible dependences with the e_{jk}. Model (5) also is a substantial extension of model (6).

Only σ^2 is investigated for the final two extensions. When the eligible procedures are not restricted, the extension is

$$y_{jk} = \mu + \alpha_j + e_{jk} + e_k^* + e_j^{**} \tag{7}$$

where the additional J + K random variables can have an arbitrary joint distribution. Also, they can have allowable dependences with the e_{jk}. This is the most general model considered and is an extension of model (5).

Finally, suppose that the procedures are limited to those customarily used and σ^2 is investigated. The extended model is

$$y_{jk} = \mu + \alpha_j + e_{jk} + e_j^{**} \tag{8}$$

where $e_1^{**}, \ldots, e_J^{**}$ can have an arbitrary joint distribution, and any allowable dependences with the e_{jk}. Model (8) has substantially less generality than model (7) but is the most general extension given that is usable with the method for rejection of outliers (as is easily seen).

As noted in the discussions of the extensions, the generality of a model is strongly reduced when the eligible procedures are limited to those customarily used for model (1). Actually, the only statistic encountered that is not customarily used for model (1) is the statistic for investigating σ^2. This statistic also occurs in tests and confidence intervals for investigating the other effects.

Basis for investigations.

The forms of the statistics used for an investigation identify the extended model that is appropriate for this investigation. That is, the extension should be as general as possible, subject to all statistics used being such that the additional random variables for the extension do not occur (cancel out or sum out).

This section states the statistics considered for possible use along with the effects they investigate, the extended model(s) for which the additional terms do not occur, and pertinent properties. For a given investigation, at least one statistic is introduced for each type of effect (μ, σ^2, one or more of the α_j) that is to be investagted. A statistic for investigating σ^2 is always included, since this statistic occurs in the tests and confidence intervals for any type of effect, and also occurs in estimates of variances for point estimates of effects.

Some of the more elementary probability properties of the statistics are stated without verification. However, proofs are easily obtained from considerations such as those given in [2]. Also, the customary results are obtained from material in [2].

Some further notation is introduced for stating the statistics that are considered for possible use.

$$y_{j.} = \sum_{k=1}^{K} y_{jk}/K, \qquad y_{.k} = \sum_{j=1}^{J} y_{jk}/J$$

$$y_{..} = \sum_{j=1}^{J}\sum_{k=1}^{K} y_{jk}/JK = \hat{\mu}, \qquad \hat{\alpha}_j = y_{j.} - y_{..}$$

$$s_a^2 = \sum_{j=1}^{J} \hat{\alpha}_j^2/(J-1), \qquad s_I^2 = \sum_{j=1}^{J}\sum_{k=1}^{K} (y_{jk} - y_{j.})^2/J(K-1)$$

$$s_{II}^2 = \sum_{j=1}^{J}\sum_{k=1}^{K} (y_{jk} - y_{j.} - y_{.k} + y_{..})^2/(J-1)(K-1)$$

$$F_{aI} = K\, s_a^2/s_I^2, \qquad F_{aII} = K\, s_a^2/s_{II}^2$$

The statistic s_I^2 is the customary unbiased estimate of σ^2 for model (1) and is free of the additional random terms for extended models (4), (6), and (8). The statistic s_{II}^2 is an unbiased estimate of σ^2 and is free of the additional random terms for all the extended models. With models (4), (6), (8) and normality, $J(K-1)\, s_I^2/\sigma^2$ has a χ^2-distribution with $J(K-1)$

degrees of freedom. For all the models and normality, $(J-1)(K-1)\,s_{II}^2/\sigma^2$ has a χ^2-distribution with $(J-1)(K-1)$ degrees of freedom.

The statistic s_a^2 is free of the additional random terms for models (2), (3), and (4). When the normality assumption also holds for the e_{jk}, and any of models (2), (3), or (4) applies, the statistic F_{aII} has an F-distribution with $J-1$ and $(J-1)(K-1)$ degrees of freedom under the null hypothesis that the α_j are all zero. This is readily verified by showing that the $y_{j.} - y_{..}$ are uncorrelated with the $y_{jk} - y_{j.} - y_{.k} + y_{..}$ and that, under the null hypothesis, $K(J-1)\,s_a^2/\sigma^2$ has a χ^2-distribution with $J-1$ degrees of freedom and $(J-1)(K-1)\,s_{II}^2/\sigma^2$ has a χ^2-distribution with $(J-1)(K-1)$ degrees of freedom. For normality, the α_j all zero, and model (4), the statistic F_{aI} has an F-distribution with $J-1$ and $J(K-1)$ degrees of freedom (the customary result when model (1) applies).

The statistic $\hat{\alpha}_j$ is the customary unbiased estimate of α_j for model (1) and is free of the additional random terms for models (2), (3), and (4). For these extended models, $(J-1)\,s_{II}^2/JK$ is an unbiased estimate of the variance of $\hat{\alpha}_j$ and, when the normality assumption also holds for the e_{jk}, is independent of $\hat{\alpha}_j$ (since the $y_{j.} - y_{..}$ are uncorrelated with the $y_{jk} - y_{j.} - y_{.k} + y_{..}$). The statistic $(J-1)\,s_{II}^2/JK$ is an unbiased estimate of the variance of $\hat{\alpha}_j$ when model (4) applies and, if the normality assumption also holds, is independent of $\hat{\alpha}_j$ (the customary results when model (1) applies). The distribution of $\hat{\alpha}_j$ is normal with mean μ and variance $(J-1)\,\sigma^2/JK$ when the normality assumption applies and any of models (2), (3), or (4) holds. These properties can be used to construct *t*-statistics for investigating linear combinations of the α_j.

Finally, $\hat{\mu}$ is the customary unbiased estimate of μ for model (1) and is free of the additional random terms for models (2), (5), and (6). For these extended models, s_{II}^2/JK is an unbiased estimate of the variance of $\hat{\mu}$ and, when the normality assumption also holds for the e_{jk}, is independent of $\hat{\mu}$ (since $y_{..}$ is uncorrelated with the $y_{jk} - y_{j.} - y_{.k} + y_{..}$). The statistic s_I^2/JK is an unbiased estimate of the variance of $\hat{\mu}$ when model (6) holds and, if the normality assumption also applies, is independent of $\hat{\mu}$ (the customary model (1) results). The distribution of $\hat{\mu}$ is normal with mean μ and variance σ^2/JK when the normality assumption holds and any of models (2), (5), (6) applies. These properties can be used to construct a t-statistic for investigating μ.

REFERENCES

[1] John E. WALSH, *Handbook of Nonparametric Statistics, III : Analysis of Variance*, D. Van Nostrand Co., Inc., Princeton, N.J., 1968.

[2] Franklin A. GRAYBILL, *An Introduction to Linear Statistical Models, Volume 1*, McGraw-Hill, New York, 1961.