Effects of Intraclass Correlation on Weighted Averages

SHAYLE R. SEARLE and FRIEDRICH PUKELSHEIM*

Weighted averages of class means using different sets of weighting factors are compared in terms of sampling variances and of relative weights given to the class means. Details are given for the 1-way classification, and extensions to other models are indicated.

1. INTRODUCTION

When subclasses of data have unequal numbers of observations, averages of the subclass means can be defined in a variety of ways, depending on the weights used for (linearly) combining the subclass means. At least three different weighting systems are often used: (I) weighting by the number of observations, which leads to the grand mean; (II) weighting equally, which yields the simple average of the subclass means; and (III) weighting inversely according to variances of the observed subclass means. In the 1-way classification, with the fixed effects model, III is the same as II; but with the random effects model (which we call the mixed model, see Sec. 3.1) in which the class effects are taken as random, I and II are special cases of III corresponding to an intraclass correlation of 0 and 1, respectively.

Variances of these weighted averages are compared in each model, and the manner in which changes in the intraclass correlation affects the relative weights given to the class means is described. Extensions to 2-way classifications are suggested.

2. FIXED EFFECTS MODELS

2.1 A Model

Suppose that y_{ij} is the jth observation of the ith class of a 1-way classification, with $i = 1, \ldots, a$ and $j = 1, \ldots, n_i$; that is, a classes and n_i observations in the ith class. Then the model equation for y_{ij} can be taken as

$$y_{ij} = \mu + \alpha_i + e_{ij} = \mu_i + e_{ij}, \qquad (1)$$

in which $\mu_i = \mu + \alpha_i$ is the population mean of the ith class and the e_{ij} terms are random variables, identically distributed with zero mean, variance σ_e^2, and zero covariances. Under these conditions the best linear unbiased estimator (BLUE) of μ_i and the sampling variance of that estimator are

$$\hat{\mu}_i = \bar{y}_i = \sum_{j=1}^{n_i} y_{ij}/n_i \qquad \text{and} \qquad v_F(\bar{y}_i) = \sigma^2/n_i, \qquad (2)$$

respectively, similar to Searle (1971, p. 235 and 339). In (2) the subscript F in $v_F(\bar{y}_i)$ emphasizes that the variance is based on the fixed effects model.

2.2 Weighted Averages

We begin with weighted averages I and II of the introduction. The first is denoted by μ_n, in which weights proportional to the numbers of observations are used:

$$\mu_n = \sum n_i\mu_i/\sum n_i.$$

(All summations are with respect to i, over the range $i = 1, 2, \ldots, a$.) The second weighted average is denoted by μ_e and is based on equal weights:

$$\mu_e = \sum \mu_i/a.$$

The third average mentioned in the introduction uses weights inversely proportional to $v(\bar{y}_i)$ and so, on using (2), is the same as μ_n:

$$\sum(\mu_i n_i/\sigma_e^2)/\sum(n_i/\sigma_e^2) = \sum n_i\mu_i/\sum n_i = \mu_n.$$

A general form of weighted average is to use arbitrary, (usually) positive weights w_i:

$$\mu_w = \sum w_i\mu_i/\sum w_i.$$

Then μ_n and μ_e are special cases of μ_w, since $w_i = n_i$ gives $\mu_w = \mu_n$, and $w_i = 1$ gives $\mu_w = \mu_e$. The BLUE's of these three averages and their sampling variances are

$$\hat{\mu}_n = \sum n_i\bar{y}_i/\sum n_i = \bar{y}_{..}, \text{ with } v_F(\hat{\mu}_n) = \sigma_e^2/\sum n_i, \qquad (3)$$

$$\hat{\mu}_e = \sum \bar{y}_i/a, \text{ with } v_F(\hat{\mu}_e) = \sigma_e^2\left(\sum 1/n_i\right)\bigg/a^2, \qquad (4)$$

$$\hat{\mu}_w = \sum w_i\bar{y}_i/\sum w_i,$$

$$\text{with } v_F(\hat{\mu}_w) = \sigma_e^2\left(\sum w_i^2/n_i\right)\bigg/\left(\sum w_i\right)^2. \qquad (5)$$

Clearly, $\hat{\mu}_n$ is the grand mean $\bar{y}_{..}$, whereas $\hat{\mu}_e$ is the average of observed class means, $\Sigma\bar{y}_i/a$.

2.3 Discussion

Estimators (3), (4), and (5) are BLUE's of different parametric functions, so comparing their sampling variances does not seem, a priori, to be beneficial. In Section 3, where we are interested in the case in which the subclass means μ_i are all the same, namely μ, the three estimators then all estimate μ and comparing variances of those estimators is then of some interest. As a prelude, the variances in (3)–(5) are compared, beginning with those of $\hat{\mu}_n$ and $\hat{\mu}_w$.

From applying the Cauchy–Schwartz inequality, $\Sigma p^2\Sigma q^2 \geq (\Sigma pq)^2$, we have

$$\sum n_i \sum w_i^2/n_i \geq \left(\sum\sqrt{n_i w_i^2/n_i}\right)^2 = \left(\sum w_i\right)^2.$$

*Shayle R. Searle is Professor, Biometrics Unit, 337 Warren Hall, Cornell University, Ithaca, NY 14853. Friedrich Pukelsheim is Professor, Lehrstuhl für Stochastik, Institut für Mathematik, Universität Augsburg, Memminger Strasse 6, D-8900 Augsburg, Federal Republic of Germany. The work of the first author was done while on leave from Cornell University and supported by a Senior U.S. Scientist Award from the Alexander von Humboldt-Stiftung. This is Paper BU-477 in the Biometrics Unit, Cornell University.

Hence

$$1/\sum n_i \le \left(\sum w_i^2/n_i\right) \Big/ \left(\sum w_i\right)^2,$$

and so from (3) and (5),

$$v_F(\hat{\mu}_n) \le v_F(\hat{\mu}_w). \qquad (6)$$

Therefore, in the fixed effects model, no weighted average of the μ_i's has a BLUE with variance smaller than that of $\hat{\mu}_n$. This is an attractive property for $\hat{\mu}_n$. In particular, it applies for $w_i = 1$, giving

$$v_F(\hat{\mu}_n) \le v_F(\hat{\mu}_e).$$

This is perhaps a little surprising, since defining an overall mean as μ_e seems more natural than does μ_n because of the dependence of μ_n on the numbers of observations in the classes.

In applications, μ_w for a particular set of w_i values can well be a parameter of interest; for example, if three varieties of wheat are grown in a county in acreages proportional to $w_1{:}w_2{:}w_3$, the county's mean wheat yield per acre is μ_w. Therefore, if in some experiment designed to measure yield the areas in which the three varieties are grown are proportional to $n_1{:}n_2{:}n_3$, different from $w_1{:}w_2{:}w_3$, then $\mu_n \ne \mu_w$, and $\hat{\mu}_w$ will be the estimated mean of interest. Nevertheless, (6) shows that $\hat{\mu}_w$ always has variance never less than that of $\hat{\mu}_n$. This suggests one reason for having subclass sizes in data proportional to subclass population sizes.

3. MIXED MODELS

3.1 A Model

Suppose with the model equation (1) that we take the α_i's as uncorrelated random effects with zero means and variance σ_α^2, with the covariance between every α_i and every e_{hk} being zero. The e_{ij} terms retain the same mean, variance, and covariance properties as described following (1). With these properties, the model is usually called the random effects model, or random model, of the 1-way classification. But since μ is a fixed effect and the α_i's are random effects, it is strictly a mixed model, and we think of it in that manner for purposes of estimating μ in the presence of the random effects.

3.2 Weighted Averages and Estimators

In the preceding mixed model the BLUE of μ, to be denoted by $\hat{\mu}_r$, is similar to Searle (1971, p. 463):

$$\hat{\mu}_r = \sum \frac{n_i}{n_i\sigma_\alpha^2 + \sigma_e^2}\bar{y}_i \Big/ \sum \frac{n_i}{n_i\sigma_\alpha^2 + \sigma_e^2}$$

$$\text{with } v_M(\hat{\mu}_r) = 1 \Big/ \sum \frac{n_i}{n_i\sigma_\alpha^2 + \sigma_e^2}. \qquad (7)$$

The subscript M in v_M of (7) denotes variance based on the mixed model. The estimator $\hat{\mu}_r$ in (7) is, of course, a special case of $\hat{\mu}_w$ with $w_i = n_i/(n_i\sigma_\alpha^2 + \sigma_e^2)$; and if $\hat{\mu}_w$ for other values of w_i is to be used, its variance is

$$v_M(\hat{\mu}_w) = \sum w_i^2(\sigma_\alpha^2 + \sigma_e^2/n_i) \Big/ \left(\sum w_i\right)^2, \qquad (8)$$

derived by replacing σ_e^2/n_i in $v_F(\hat{\mu}_w)$ of (5) with $\sigma_\alpha^2 + \sigma_e^2/n_i$.

3.3 Comparing Variances of Estimated Averages

First, from (8) and (5) it is easily seen that

$$v_M(\hat{\mu}_w) = \sigma_\alpha^2 \frac{\sum w_i^2}{(\sum w_i)^2} + v_F(\hat{\mu}_w) > v_F(\hat{\mu}_w), \quad \text{for } \sigma_\alpha^2 > 0.$$

Thus every weighted average has variance in the mixed model that exceeds its variance in the fixed model, as one would expect. (When $\sigma_\alpha^2 = 0$ the variances are equal.) What is more interesting is that by applying the same reasoning to (7) and (8) as is used in deriving (6), it is easily shown that

$$v_M(\hat{\mu}_r) \le v_M(\hat{\mu}_w). \qquad (9)$$

This shows that in the mixed model no weighted average $\hat{\mu}_w$ has smaller variance than does $\hat{\mu}_r$ (as is to be expected because $\hat{\mu}_r$ is the BLUE of μ).

A special case of (9) is $v_M(\hat{\mu}_r) \le v_M(\hat{\mu}_n)$. Nevertheless, $v_F(\hat{\mu}_n)$ of (3) is less than $v_M(\hat{\mu}_r)$ of (7), as may be seen by observing that

$$1/v_F(\hat{\mu}_n) - 1/v_M(\hat{\mu}_r) = \sum n_i[1/\sigma_e^2 - 1/(n_i\sigma_\alpha^2 + \sigma_e^2)] > 0, \quad \text{for } \sigma_\alpha^2 > 0.$$

Hence

$$v_F(\hat{\mu}_n) \le v_M(\hat{\mu}_r) \le v_M(\hat{\mu}_n). \qquad (10)$$

Thus the variance of $\hat{\mu}_r$ in the mixed model is between that of $\hat{\mu}_n$ in the fixed and mixed models, with these variances being equal when $\sigma_\alpha^2 = 0$, for then $\hat{\mu}_n = \hat{\mu}_r$.

3.4 Relative Weights for Observed Subclass Means in $\hat{\mu}_r$

In $\hat{\mu}_n$ the observed subclass means, $\bar{y}_i$, are weighted in proportion to their n_i-values; in $\hat{\mu}_e$ they are weighted equally. In the mixed model with intraclass correlation $\rho = \sigma_\alpha^2/(\sigma_\alpha^2 + \sigma_e^2)$, it is interesting to see how the weights in $\hat{\mu}_r$ change from those of $\hat{\mu}_n$ when $\rho = 0$ to those of $\hat{\mu}_e$ when $\rho = 1$. To observe this, write $\hat{\mu}_r$ of (7) as

$$\hat{\mu}_{r,\rho} = \sum \frac{n_i}{n_i\rho + 1 - \rho}\bar{y}_i \Big/ \sum \frac{n_i}{n_i\rho + 1 - \rho}. \qquad (11)$$

Then $\rho = 0$ yields $\hat{\mu}_{r,0} = \hat{\mu}_n = \bar{y}$ of (3) and $\rho = 1$ gives $\hat{\mu}_{r,1} = \hat{\mu}_e$ of (4). This is not surprising. $\rho = 0$ is equivalent to $\sigma_\alpha^2 = 0$, which reduces the mixed model to being a fixed effects model $y_{ij} = \mu + e_{ij}$ and so $\hat{\mu}_{r,0} = \hat{\mu}_n$, the BLUE of μ in that model. And $\rho = 1$, although equivalent to $\sigma_e^2 = 0$, is more interestingly the case of observations within each class being perfectly correlated—in effect, identical. Hence no matter what the value of n_i is, $\bar{y}_i$ has variance σ_α^2, and so the linear combination of $\bar{y}_i$'s that has minimum variance is $\hat{\mu}_e = \Sigma\bar{y}_i/a$.

Despite these consequences of putting $\rho = 0$ and $\rho = 1$ in $\hat{\mu}_r$, it is nevertheless surprising how quickly the weights given to each $\bar{y}_i$ change from being proportional to n_i in $\hat{\mu}_{r,0} = \hat{\mu}_n$ to approaching being equal in $\hat{\mu}_{r,1} = \hat{\mu}_e$ as ρ increases from 0 to 1. Consider two classes, one described as having a large number of observations, n_L, and the other having a small number, n_S, with of course, $n_L > n_S$. In $\hat{\mu}_r$ the ratio of the weight given $\bar{y}_S$ to that given to $\bar{y}_L$ is τ_ρ, where, from (11),

$$\tau_\rho = \frac{\text{coefficient of } \bar{y}_S \text{ in } \hat{\mu}_{r,\rho}}{\text{coefficient of } \bar{y}_L \text{ in } \hat{\mu}_{r,\rho}}$$

$$= \frac{n_S(n_L\rho + 1 - \rho)}{n_L(n_S\rho + 1 - \rho)} = \frac{\rho + (1 - \rho)/n_L}{\rho + (1 - \rho)/n_S}. \quad (12)$$

Corresponding to $\rho = 0$ with $\hat{\mu}_{r,0} = \hat{\mu}_n$, we have $\tau_0 = n_S/n_L$; and as ρ increases from zero to unity, τ_ρ increases from $\tau_0 = n_S/n_L$ to $\tau_1 = 1$. Thus as $\rho \to 1$, we see that $\bar{y}_S$, the data mean of the smaller-sized class, gets increasingly larger weights in $\hat{\mu}_{r,\rho}$, relative to $\bar{y}_L$. It is interesting to see that this increase can, depending on the magnitudes of n_L and n_S, be quite appreciable, even for very small values of ρ. This is so because the first derivative of τ_ρ with respect to ρ is

$$\tau'_\rho = \partial\tau_\rho/\partial\rho = (1/n_S - 1/n_L)/(\rho + (1 - \rho)/n_S)^2, \quad (13)$$

and for small values of ρ and not-too-small values of n_S, this can be relatively large. In particular, for $\rho = 0$,

$$\tau'_0 = n_S(1 - n_S/n_L), \quad (14)$$

and so when n_S/n_L is small and n_S is not too small, τ'_0 can be relatively large [e.g., for $n_S = 20$ and $n_L = 100$, $\tau'_0 = 20(1 - .2) = 16$]. This is the slope at $\rho = 0$ of τ_ρ plotted against ρ. The value 16 represents an angle of 86.4° from the horizontal, which means that, for values of ρ near zero, τ_ρ increases very rapidly from $\tau_0 = n_S/n_L = 20/100 = .2$. This is evident in the second column of Table 1, which shows values of τ_ρ for three pairs of n_S, n_L values and a range of values of ρ.

3.5 Discussion

The BLUE of μ in the mixed model is $\hat{\mu}_r$; it reduces to $\hat{\mu}_n = \bar{y}_{..}$ in the fixed model wherein $\sigma^2_\alpha = 0$, and to $\hat{\mu}_e = \Sigma\bar{y}_i/a$ in the trivial case of $\sigma^2_e = 0$ when all observations in each class are then identical (and of course, if every n_i has the same value, then $\hat{\mu}_n = \hat{\mu}_e = \bar{y}_{..}$). Each of the estimators $\hat{\mu}_n$, $\hat{\mu}_e$, and $\hat{\mu}_r$ has variance in the mixed model that exceeds its variance in the fixed model, as is, of course, to be expected. In contrast, as in (9), in the mixed model $\hat{\mu}_r$ has the smallest variance of any (linearly) weighted average, although in the fixed model $\hat{\mu}_n$ has still smaller variance.

In $\hat{\mu}_r$ the weight given to $\bar{y}_S$ having n_S observations, relative to that given to y_L with $n_L > n_S$ observations, is τ_ρ given by (12). The value of τ_ρ is n_S/n_L for $\rho = 0$, that is, in $\hat{\mu}_n$; and it is 1.0 for $\rho = 1$, that is, in $\hat{\mu}_e$. The rate of increase in τ_ρ for ρ increasing from 0 to 1 is given by τ'_ρ of (13) with $\tau'_0 = n_S(1 - n_S/n_L)$. Thus for small values of ρ the rate of increase in τ_ρ depends not only on n_S/n_L but also on n_S; hence small changes in ρ can bring about big changes in τ_ρ. This is illustrated in Table 1, where, for the example having $n_S = 20$ and $n_L = 100$, changing ρ from 0 to .05 changes τ_ρ from .20 to .61. Thus not only can relative sizes of data subclasses be important in the contributions that observed subclass means make to $\hat{\mu}_r$, but absolute sizes are also important. This is also illustrated in Table 1, where in each of the first two examples $n_S/n_L = .2$: in the first of these, $n_S = 4$ and $\tau_{.05}$ is .33, whereas in the second, with $n_S = 20$ the value of $\tau_{.05}$ is .61, nearly double its value for $n_S = 4$.

Table 1. Dependence on Intraclass Correlation of the Relative Weights Given to Two Observed Subclass Means in the Estimator $\hat{\mu}_{r,\rho} = \Sigma[n_i/(n_i\rho + 1 - \rho)]\,\bar{y}_i/\Sigma[n_i/(n_i\rho + 1 - \rho)]$

Intraclass correlation, $\rho = \frac{\sigma^2_\alpha}{\sigma^2_\alpha + \sigma^2_e}$	$\tau_\rho = \frac{\text{coefficient of } \bar{y}_S \text{ in } \hat{\mu}_{r,\rho}}{\text{coefficient of } \bar{y}_L \text{ in } \hat{\mu}_{r,\rho}} = \frac{\rho + (1-\rho)/n_L}{\rho + (1-\rho)/n_S}$ for three (n_S, n_L) pairs		
	$n_S = 4$, $n_L = 20$	$n_S = 20$, $n_L = 100$	$n_S = 5$, $n_L = 100$
0 ($\hat{\mu}_{r,0} = \hat{\mu}_n$, $\tau_0 = n_S/n_L$)	.20	.20	.05
.05	.33	.61	.28
.1	.45	.75	.38
.3	.71	.92	.70
.5	.840	.962	.842
.7	.923	.983	.925
.9	.978	.996	.979
1.0 ($\hat{\mu}_{r,1} = \hat{\mu}_e$, $\tau_1 = 1$)	1.00	1.00	1.00

4. EXTENSIONS

Consider a 2-way nested classification in which the number of main classes is a, with the ith having b_i subclasses, in the jth of which there are n_{ij} observations y_{ijk} for $k = 1, \ldots, n_{ij}$, with $i = 1, \ldots, a$ and $j = 1, \ldots, b_i$. A mixed model for this situation can be taken as $y_{ijk} = \mu_i + \beta_{ij} + e_{ijk}$ with μ_i as a fixed effect and β_{ij} and e_{ijk} as random effects with zero means, variances σ^2_β and σ^2_e, respectively, and with all covariances zero. Then, similar to $\hat{\mu}_r$ of (7), the BLUE of μ_i is

$$\hat{\mu}_i = \sum_{j=1}^{b} \frac{n_{ij}}{n_{ij}\sigma^2_\beta + \sigma^2_e}\bar{y}_{ij\cdot} \Bigg/ \sum_{j=1}^{b} \frac{n_{ij}}{n_{ij}\sigma^2_\beta + \sigma^2_e}.$$

Discussion of this and of linear combinations of the $\hat{\mu}_i$'s can be made similar to those of Sections 2 and 3. Analogous extensions could also be made for a 2-way crossed classification for combining BLUE's $\hat{\mu}_{ij} = \bar{y}_{ij\cdot}$ in situations in which $v(\bar{y}_{ij\cdot}) = \sigma^2_\gamma + \sigma^2_e/n_{ij\cdot}$.

[*Received February 1985. Revised September 1985.*]

REFERENCE

Searle, S. R. (1971), *Linear Models*, New York: John Wiley.