MEASUREMENT SCIENCE REVIEW, Volume 1, Number 1, 2001

ON STATISTICAL MODELS FOR CONSENSUS VALUES

Viktor Witkovský and Gejza Wimmer
Institute of Measurement Science and
Mathematical Institute
Slovak Academy of Sciences
Dúbravská cesta 9
842 19 Bratislava, Slovakia
E-mail: umerwitk@savba.sk

Abstract

We consider the problem of measurements made by several laboratories (or methods) on virtually the same object of interest. In general, the number of measurements made at each laboratory may differ. Moreover, the laboratories may exhibit the between-laboratory variability caused by the systematic error due to each laboratory, as well as different within-laboratory variances caused by different within-laboratory precision of the used measurement method. In this paper we try to describe statistical models and methods that are appropriate for derivation of the consensus mean of the unknown (measured) value as well as related problems concerning the statistical inference on the unknown value.

1. Introduction

We consider that the measurements on virtually the same object of interest are made by $k \geq 2$ laboratories. The ith laboratory repeats its measurements n_i times, $n_i \geq 2$. The laboratories may exhibit the between laboratory variability, as well as different within-laboratory variances (heteroscedasticity). In this paper we will assume that the measurements follow normal distribution.

The results of a typical interlaboratory study (given in aggregated form) are presented in Table 1: In [2] the data on Selenium in non-fat milk powder were reported. The measurements are based on four independent measurement methods.

Table 1. Selenium in non-fat milk powder

Method	n_i	$\bar{y}_i$ (mean)	s_i^2 (variance)
A	8	105.00	85.711
B	12	109.75	20.748
C	14	109.50	2.729
D	8	113.25	33.640

As pointed in [7]: *A question of fundamental importance in the analysis of such data is how to form the best consensus mean, and what uncertainty to attach to this estimate.* This fundamental question is followed by a series of other questions regarding the statistical inference on the unknown common mean.

The problem, although not new in statistical literature, see e.g. [1, 4], is not completely solved. Many questions remain still open and unsolved. The problem of interlaboratory comparisons is of particular interest for applications that are looking for harmonization of industrial and scientific practice. The general problem covers many aspects of which the major ones are: (1) the choice of the appropriate model, (2) the choice of associated statistical methods and (3) the identification and verification of the model.

2. Subject and methods

Basic tool for analysis of interlaboratory measurements and to form the best consensus mean value will be the one-way classification (ANOVA) model, which may be both unbalanced and heteroscedastic, that is

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

with mutually independent errors, $\varepsilon_{ij} \sim N(0, \sigma_i^2)$, $i = 1, \ldots, k$ and $j = 1, \ldots, n_i$. The laboratory effects could be considered to be fixed effects or random effects. If the second case is true then we assume that $\alpha_i \sim N(0, \sigma^2)$ are mutually independent and independent with all ε_{ij}. The variance components σ_i^2 and σ^2 are the nuisance parameters: the within-laboratory and between-laboratory variances. The model could be written in matrix notation:

$$y = 1\mu + Z\alpha + \varepsilon, \tag{2}$$

where $y = (y_1', \ldots, y_k')'$ with $y_i = (y_{i1}, \ldots, y_{in_i})'$ is the n-vector of all measurements, $n = \sum n_i$. Further, $1 = (1_1' \ldots, 1_k')'$ with $1_i = (1, \ldots, 1)'$ (n_i-vector of ones); $\alpha = (\alpha_1, \ldots, \alpha_k)'$ is the (unknown) k-vector of laboratory effects; $Z = \mathit{Diag}\{1_i\}$ is known $n \times k$-matrix; and $\varepsilon = (\varepsilon_1', \ldots, \varepsilon_k')'$ with $\varepsilon_i = (\varepsilon_{i1}, \ldots, \varepsilon_{in_i})'$ represents the n-vector of within-laboratory errors. Moreover, we will use the following notation: $\bar{y}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} y_{ij}$, $s_i^2 = \frac{1}{n_i - 1} \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2$.

Notice that under given assumptions (random effects model) the following statistical properties hold true: $E(y) = \mu$, $\mathit{Var}(y) = \sigma^2 ZZ' + \mathit{Diag}\{\sigma_i^2 I_{n_i}\}$, where I_{n_i} denotes $n_i \times n_i$-identity matrix. If the variance components σ^2 and σ_i^2 would be known the optimal estimator for the unknown common mean μ would be the generalized least squares estimator (GLS estimator) which is (under the given assumptions) MVUE — minimum variance unbiased estimator, that is

$$\hat{\mu}_{GLS} = (1'\mathit{Var}(y)^{-1}1)^{-1} 1'\mathit{Var}(y)^{-1} y = \frac{\sum_{i=1}^k w_i \bar{y}_i}{\sum_{i=1}^k w_i}, \tag{3}$$

where $w_i = 1/\mathit{Var}(\bar{y}_i)$ with $\mathit{Var}(\bar{y}_i) = \sigma^2 + \frac{\sigma_i^2}{n_i}$, i.e. the optimal estimator of μ is the weighted average of k laboratory average values $\bar{y}_i$ with the weights w_i inversely proportional to the variances of the individual laboratory averages. Under given assumptions the exact distribution of the estimator is known: $\hat{\mu}_{GLS} \sim N(\mu, 1/\sum w_i)$. From that the standard statistical inference on μ could be performed.

If the variance components are unknown the situation becomes more complicated. Under normality assumptions the maximum likelihood estimator of μ is a reasonable choice to form the consensus mean value. The ML estimators of the (unknown) common mean μ and the between-laboratory variance σ^2 have the form

$$\hat{\mu}_{ML} = \frac{\sum_{i=1}^k \hat{w}_i^{ML} \bar{y}_i}{\sum_{i=1}^k \hat{w}_i^{ML}} = \frac{\sum_{i=1}^k \hat{\gamma}_i^{ML} \bar{y}_i}{\sum_{i=1}^k \hat{\gamma}_i^{ML}}, \tag{4}$$

$$\hat{\sigma}_{ML}^2 = \frac{1}{n} \sum_{i=1}^k \hat{\gamma}_i^{ML} \left((\bar{y}_i - \hat{\mu}_{ML})^2 + \frac{(n_i - 1)\frac{s_i^2}{n_i}}{1 - \hat{\gamma}_i^{ML}} \right), \tag{5}$$

where $\hat{w}_i^{ML}$ are the MLEs of the weights $w_i = 1/(\sigma^2 + \frac{\sigma_i^2}{n_i})$, $n = \sum_{i=1}^k n_i$, $\gamma_i = \sigma^2/(\sigma^2 + \frac{\sigma_i^2}{n_i})$ and according to [7], the MLEs $\hat{\gamma}_i^{ML}$ of γ_i, $i = 1, \ldots, k$, are found by minimizing

$$n \log \hat{\sigma}_{ML}^2 - \sum_{i=1}^k \log \hat{\gamma}_i^{ML} + \sum_{i=1}^k (n_i - 1) \log \frac{1 - \hat{\gamma}_i^{ML}}{\hat{\gamma}_i^{ML}}. \tag{6}$$

Based on $\hat{\sigma}^2_{ML}$ and $\hat{\gamma}_i^{ML}$ it is possible to derive MLEs of within-laboratory variances σ_i^2, $i = 1, \ldots, k$, that is

$$\hat{\sigma}^2_{i\,ML} = \frac{n_i \hat{\sigma}^2_{ML}(1 - \hat{\gamma}_i^{ML})}{\hat{\gamma}_i^{ML}}. \tag{7}$$

For more details see [7, 8]. Small sample distribution of $\hat{\mu}_{ML}$ is in general unknown. Under assumptions given in [5], the ML estimator is asymptotically normally distributed (as $n \to \infty$) with $Var(\hat{\mu}_{ML}) \to 1/\sum w_i = \sigma^2/\sum \gamma_i$. So, with z_α denoting the critical point of standard normal distribution, the interval

$$\hat{\mu} \pm z_{\alpha/2} \sqrt{\frac{\hat{\sigma}^2_{ML}}{\sum_{i=1}^k \hat{\gamma}_i^{ML}}} \tag{8}$$

provides an approximate $(1 - \alpha)\%$ confidence interval based on the asymptotic distribution.

Computationally simpler method, the Mandel-Paule estimator, was suggested in [6]. This method is widely used in applications and experience has shown that it often provides reasonable estimates and is recommended for use in the preparation of standard reference materials. The Mandel-Paule estimator (MP estimator) has the form

$$\hat{\mu}_{MP} = \frac{\sum_{i=1}^k \hat{w}_i^{MP} \bar{y}_i}{\sum_{i=1}^k \hat{w}_i^{MP}}, \tag{9}$$

where $\hat{w}_i^{MP} = 1/(\hat{\sigma}^2_{MP} + \frac{s_i^2}{n_i})$ are the MP estimators of the weights w_i. Here $\hat{\sigma}^2_{MP}$ estimates the between-laboratory variance σ^2 and could be derived iteratively from the equation

$$\sum_{i=1}^k \frac{(\bar{y}_i - \hat{\mu}_{MP})^2}{\hat{\sigma}^2_{MP} + \frac{s_i^2}{n_i}} = k - 1. \tag{10}$$

Notice, that the MP estimator (9) coincides with the Graybill-Deal estimator (GD estimator), another widely accepted estimator for the common mean μ, analyzed in [4], if the between-laboratory variance is zero, i.e. if the submodel $y_{ij} = \mu + \varepsilon_{ij}$ of the model (1) is true. The GD estimator has the form

$$\hat{\mu}_{GD} = \frac{\sum_{i=1}^k \hat{w}_i^{GD} \bar{y}_i}{\sum_{i=1}^k \hat{w}_i^{GD}} = \frac{\sum_{i=1}^k \frac{n_i}{s_i^2} \bar{y}_i}{\sum_{i=1}^k \frac{n_i}{s_i^2}}. \tag{11}$$

Rukhin and Vangel in [7] suggested the modified MP estimator to be as previously with k instead of $k - 1$ on the right hand side of (10). They showed that the MP estimator is well defined and the modified MP estimator is close to the ML estimator. They also derived the consistent estimator of the asymptotic variance of the consensus mean estimator $\hat{\mu}$ (as $k \to \infty$). So, the interval

$$\hat{\mu} \pm z_{\alpha/2} \frac{\sqrt{\sum_{i=1}^k \frac{(\bar{y}_i - \hat{\mu})^2}{(\hat{\sigma}^2 + \frac{s_i^2}{n_i})^2}}}{\sum_{i=1}^k \frac{1}{\hat{\sigma}^2 + \frac{s_i^2}{n_i}}} \tag{12}$$

provides an approximate $(1 - \alpha)\%$ confidence interval. Here $\hat{\mu}$ is the MP estimator or the modified MP estimator of μ and $\hat{\sigma}^2$ is the MP estimator or the modified MP estimator of σ^2.

Table 2 reports the estimated consensus mean values calculated by the four above mentioned methods together with their (approximate) 95% confidence intervals and together with the estimates of the between-laboratory

variance. The MLEs of the within-laboratory variances are: $\hat{\sigma}^2_{1\,ML} = 95.9274$, $\hat{\sigma}^2_{2\,ML} = 19.0497$, $\hat{\sigma}^2_{3\,ML} = 2.5397$ and $\hat{\sigma}^2_{4\,ML} = 42.9409$. The confidence interval reported with the GD estimate (denoted by $*$) is the exact 95% confidence interval estimate suggested by Fairweather in [3] and calculated according to [9]. The MATLAB code suggested for numerical evaluation of the consensus mean value is available from the authors.

Table 2. The consensus mean value
Selenium in non-fat milk powder

Estimator	$\hat{\mu}$	95% Confidence Interval	$\hat{\sigma}^2$
ML	109.5750	$\langle 108.8010; 110.3490 \rangle$	0.0000
MP	109.8214	$\langle 108.0596; 111.5832 \rangle$	4.1340
MMP	109.8184	$\langle 108.5439; 111.0928 \rangle$	1.5479
GD	109.6021	$\langle 108.5369; 110.7722 \rangle^*$	0

3. Discussion

Many problems concerning statistical inference in the statistical models with common mean are still open. If we consider one-way classification model (1) as the basic tool for such analysis, one important question is if the laboratory effects are fixed or random. Balancedness of the data (i.e. the situation when $n_1 = \cdots = n_k$) has special consequences for further statistical analysis as e.g.: point and interval estimation of the variance components; testing for homoscedasticity (i.e. testing the null hypothesis $H_0 : \sigma_1^2 = \cdots = \sigma_k^2$, and/or more specifically testing hypothesis $H_0 : \sigma_i^2 = \sigma_j^2$); testing statistical hypotheses on the laboratory effects (i.e. $H_0 : \alpha_1 = \cdots = \alpha_k$ or $H_0 : \sigma^2 = 0$, i.e. submodel testing, and/or more specifically testing hypothesis $H_0 : \alpha_i = \alpha_j$ or $H_0 : \sigma^2 \leq \sigma_0^2$); point estimation and interval estimation of the consensus value under those special situations.

4. Acknowledgement

The research has been supported by the Grant VEGA 1/7295/20 from the Science Grant Agency of the Slovak Republic.

References

[1] Cochran, W.G. (1937). Problems arising in the analysis of a series of similar experiments. *Journal of the Royal Statistical Society, Supplement*, 4, 102–118.

[2] Eberhardt, K.R., Reeve, C.P. and Spiegelman, C.H. (1989). A minimax approach to combining means, with practical examples. *Chemometrics Intell. Lab. Systems*, 5, 129–148.

[3] Fairweather, W.R. (1972). A method of obtaining an exact confidence interval for the common mean of several normal populations. *Applied Statistics*, 21, 229–233.

[4] Graybill, F.A. and Deal, R.D. (1959). Combining unbiased estimators. *Biometrics*, 3, 1–21.

[5] Miller, J.J. (1977). Asymptotic properties of maximum likelihood estimates in the mixed model of the analysis of variance. *Annals of Statistics*, 5(4), 746–762.

[6] Paule, R.C. and Mandel, J. (1982). Consensus values and weighting factors. *Journal of Research of the National Bureau of Standards*, 87(5), 377–385.

[7] Rukhin, A.L. and Vangel, M.G. (1998). Estimation of a common mean and weighted means statistics. *Journal of the American Statistical Association*, 93(441), 303–308.

[8] Searle, S.R., Casella G. and McCulloch, C.E. (1992). *Variance Components.* J. Wiley & Sons, New York.

[9] Witkovský, V. (2001). On the exact computation of the density and of the quantiles of linear combinations of t and F random variables. *Journal of Statistical Planning and Inference*, 94(1), 1–13.