Confidence Intervals for Common Mean in One-way Classification Model with Fixed Effects[1]

Alexander Savin

Institute of Measurement Science, Slovak Academy of Sciences
Dúbravská cesta 9, 841 04 Bratislava, Slovak Republic
E-mail: savin@savba.sk

Abstract. *In this paper we consider the various methods for interval estimation of the common mean — the problem which appears in interlaboratory studies but is also closely related to the multicenter clinical trials and meta-analysis. Here we present the statistical properties (based on large simulation study) of the proposed confidence intervals.*

Keywords: *Interlaboratory study; Common mean; One-way classification model; Fixed effects.*

1. Introduction

In interlaboratory studies we suppose that the measurements on virtually the same object of interest are made by fixed (k) number of laboratories. The i-th laboratory repeats its measurements n_i times, $n_i \geq 2$. The laboratories may exhibit different within-laboratory variances (heteroscedasticity). Here we will assume that the measurements follow normal distribution. We consider the following fixed effects model:

$$Y_{ij} = \mu_i + \varepsilon_{ij}, \tag{1}$$

with mutually independent errors, $\varepsilon_{ij} \sim \mathcal{N}\left(0, \sigma_i^2\right)$, for $i = 1, \ldots, k$ and $j = 1, \ldots, n_i$. The (unknown) variance components σ_i^2 are the nuisance parameters: the within-laboratory variances. The parameter of interest is the (unknown) expected value $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_k)'$. We can make some inferences about the expected value $\boldsymbol{\mu}$. The most useful inference is testing about the homogenity of expected value, i.e. whether the expected values are equal in each class (laboratory).

$$H_0 : \; \mu_1 = \mu_2 = \cdots = \mu_k = \mu \text{ vs. } H_1 : \; \exists \text{ such } i, j \text{ that } \mu_i \neq \mu_j. \tag{2}$$

Interest is about the common mean μ, the common value of the laboratory measurements after not rejecting the hypothesis H_0. In this case we get model:

$$Y_{ij} = \mu + \varepsilon_{ij}, \tag{3}$$

where $\varepsilon_{ij} \sim \mathcal{N}\left(0, \sigma_i^2\right)$, for $i = 1, \ldots, k$ and $j = 1, \ldots, n_i$ are mutually independent.

2. Methods

The task is to make inference about the common mean μ, esp. confidence intervals for μ, so we need an estimator of μ. Consider an unbiased estimator $\hat{\mu}$ of the common mean μ with variance $\mathsf{Var}\left(\hat{\mu}\right) = \sum_{i=1}^{k} \lambda_i \sigma_i^2$, where $\lambda_i > 0$. If the variance components σ_i^2 are known then the pivot

$$Z = \frac{\hat{\mu} - \mu}{\sqrt{\mathsf{Var}\left(\hat{\mu}\right)}} \sim \mathcal{N}\left(0, 1\right) \tag{4}$$

[1] The paper was supported by grant from Scientific Grant Agency of the Slovak Republic VEGA 1/0264/03.

follows a normal distribution and its derived $(1-\alpha)\,100\%$ confidence interval is:

$$\hat{\mu} - u\,(1-\alpha/2)\,\sqrt{\mathsf{Var}\,(\hat{\mu})} \leq \mu \leq \hat{\mu} + u\,(1-\alpha/2)\,\sqrt{\mathsf{Var}\,(\hat{\mu})}, \tag{5}$$

where $u\,(.)$ is quantile function of normal distribution. If the variance components σ_i^2 are unknown then we don't know the exact distribution of Z.

So we want to compare some approximate confidence intervals for common mean derived from the simple t-statistic, the t-statistic with Satterthwaite's degrees of freedom, the t-statistic derived from Kenward-Roger method and by Welch's quantile approximation.

Interval derived from simple t-statistic.

The simple t-statistic T is given by

$$T = \frac{\bar{Y}_n - \mu}{\sqrt{\widehat{\mathsf{Var}\,(\bar{Y}_n)}}} \tag{6}$$

with $\widehat{\mathsf{Var}\,(\bar{Y}_n)} = S^2/N$, where $N = \sum_{i=1}^{k} n_i$, $\bar{Y}_{i.} = n_i^{-1}\sum_{j=1}^{n_i} Y_{ij}$, $\bar{Y}_n = N^{-1}\sum_{i=1}^{k} n_i \bar{Y}_{i.}$, $S_i^2 = (n_i-1)^{-1}\sum_{j=1}^{n_i}\left(Y_{ij}-\bar{Y}_{i.}\right)^2$, $S^2 = (N-k)^{-1}\sum_{i=1}^{k}(n_i-1)\,S_i^2$.

This statistic was derived under the assumption of the variance homogenity and has a t distribution with $N-k$ degrees of freedom. So the $(1-\alpha)\,100\%$ confidence interval is:

$$\bar{Y}_n - t_{N-k}\,(1-\alpha/2)\,\sqrt{\widehat{\mathsf{Var}\,(\bar{Y}_n)}} \leq \mu \leq \bar{Y}_n + t_{N-k}\,(1-\alpha/2)\,\sqrt{\widehat{\mathsf{Var}\,(\bar{Y}_n)}}, \tag{7}$$

where $t_{df}\,(.)$ is quantile function of Student's t distribution with df degrees of freedom. It is not correct to use this confidence interval with respect to heterogenity in the error variances.

Interval derived from t-statistic with Satterthwaite's degrees of freedom.

The t-test, T_S, is given by

$$T_S = \frac{\bar{Y}_n - \mu}{\sqrt{\widehat{\mathsf{Var}\,(\bar{Y}_n)}}} \tag{8}$$

with $\widehat{\mathsf{Var}\,(\bar{Y}_n)} = N^{-2}\sum_{i=1}^{k} n_i S_i^2$. Satterthwaite approximated in [2] the sum of χ^2 random variables to derive the null distribution of the statistic T_S as a t random variable with approximately ν degrees of freedom: $\hat{\nu} = \left(\sum_{i=1}^{k} n_i S_i^2\right)^2 / \left(\sum_{i=1}^{k}(n_i-1)^{-1} n_i^2 S_i^2\right)$. The $(1-\alpha)\,100\%$ confidence interval is

$$\bar{Y}_n - t_{\hat{\nu}}\,(1-\alpha/2)\,\sqrt{\widehat{\mathsf{Var}\,(\bar{Y}_n)}} \leq \mu \leq \bar{Y}_\omega + t_{\hat{\nu}}\,(1-\alpha/2)\,\sqrt{\widehat{\mathsf{Var}\,(\bar{Y}_n)}}. \tag{9}$$

Welch's quantile approximation.

Consider this probability equation:

$$\text{Pr}\left[\bar{Y}_n - \mu < u\,(\xi)\,\sqrt{\mathsf{Var}\,(\bar{Y}_n)}\right] = \xi. \tag{10}$$

If the variance components σ_i^2 are known then equation (10) holds true. If the variance components are unknown we have only estimators S_i^2. Welch's approach (see [3]) was to approximate the distribution, i.e. to find such a quantile function h

$$\text{Pr}\left[\bar{Y}_n - \mu < h\left(S_1^2,\ldots,S_k^2,\xi\right)\right] = \xi \tag{11}$$

that the equation (11) holds true. The $(1-\alpha)\,100\%$ confidence interval is:

$$\bar{Y}_n - h\left(S_1^2,\ldots,S_k^2,1-\alpha/2\right) \leq \mu \leq \bar{Y}_n + h\left(S_1^2,\ldots,S_k^2,1-\alpha/2\right), \tag{12}$$

where the approximated function h is

$$h\left(S_1^2,\dots,S_k^2,\xi\right)=u_\xi\sqrt{\sum_{i=1}^{k}\lambda_i S_i^2}\left[1+\frac{1+u_\xi^2}{4}\frac{\sum_{i=1}^{k}\lambda_i^2 S_i^4/f_i}{\left(\sum_{i=1}^{k}\lambda_i S_i^2\right)^2}-\frac{1+u_\xi^2}{2}\frac{\sum_{i=1}^{k}\lambda_i^2 S_i^4/f_i^2}{\left(\sum_{i=1}^{k}\lambda_i S_i^2\right)^2}\right.$$
$$\left.+\frac{3+5u_\xi^2+u_\xi^4}{3}\frac{\sum_{i=1}^{k}\lambda_i^3 S_i^6/f_i^2}{\left(\sum_{i=1}^{k}\lambda_i S_i^2\right)^3}-\frac{15+32u_\xi^2+9u_\xi^4}{32}\frac{\left(\sum_{i=1}^{k}\lambda_i^2 S_i^4/f_i\right)^2}{\left(\sum_{i=1}^{k}\lambda_i S_i^2\right)^4}\right], \quad (13)$$

where $f_i = n_i - 1, \qquad \lambda_i = \dfrac{n_i}{N^2}, \qquad \text{for } i = 1,\dots,k.$

Interval derived by Kenward Roger method

Kenward and Roger in [1] derived the method to estimate the variance of the generalized least square estimator (GLSE) and derived a test statistic for inference about expected values.

$$T_{KR}=\frac{\bar{Y}_{\hat{\omega}}-\mu}{\sqrt{\widehat{\mathsf{Var}\left(\bar{Y}_{\hat{\omega}}\right)}}} \quad (14)$$

with $\widehat{\mathsf{Var}\left(\bar{Y}_{\hat{\omega}}\right)}=\left(\sum_{i=1}^{k}\hat{\omega}_i\right)^{-1}+2\hat{\mathbf{\Lambda}}$, where $\hat{\omega}_i = n_i/S_i^2$, $\bar{Y}_{\hat{\omega}}=\left(\sum_{i=1}^{k}\hat{\omega}_i\right)^{-1}\sum_{i=1}^{k}\hat{\omega}_i\bar{Y}_{i.}$ and $\hat{\mathbf{\Lambda}}$ is penalty derived from Kenward-Roger method in [1]. The statistic T_{KR} has a t distribution with approximately $\hat{m}$ degrees of freedom, where degrees of freedom $\hat{m}$ are derived by Satterthwaite's method (see [1, 2]). The $(1-\alpha)\,100\%$ confidence interval is

$$\bar{Y}_{\hat{\omega}}-t_{\hat{m}}\left(1-\alpha/2\right)\sqrt{\widehat{\mathsf{Var}\left(\bar{Y}_{\hat{\omega}}\right)}}\le\mu\le\bar{Y}_{\hat{\omega}}+t_{\hat{m}}\left(1-\alpha/2\right)\sqrt{\widehat{\mathsf{Var}\left(\bar{Y}_{\hat{\omega}}\right)}}. \quad (15)$$

3. Simulation Study

In our simulation study we have examined the coverage probabilities of the confidence intervals for the common mean μ in the model (3). The design of the simulation study is similar to that presented in [4] by Savin, Wimmer and Witkovský. Assuming that model (3) is true, we have used the following values of the unknown parameters in the simulation study: $\mu = 0$, $k \in \{2, 5, 10, 21\}$. Given k, four patterns of n_i were used: $n_i = 5$, $i = 1,\dots,k$, further $n_i = \{5, 10, 5, 10, \dots\}$ for all i, $n_i = 10$ for all i and further $n_i = 30$ for all i. The within-laboratory variances were equally spaced values of $\left\{\sigma_1^2,\dots,\sigma_k^2\right\}$, where $\sigma_1^2 = 5$ and $\sigma_k^2 = 5$ in first (homoscedatic) pattern and $\sigma_1^2 = 1$ and $\sigma_k^2 = 10$ (or $\sigma_1^2 = 10$ and $\sigma_k^2 = 1$ only in case if n_i are not equal) in second (heteroscedastic) pattern. For each combination of parameters, 5,000 independent realizations of $\bar{y}_1,\dots,\bar{y}_k$ and $s_1^2,\dots,s_k^2$ were generated and the 95% confidence intervals for μ were calculated and compared. The relative frequency of cases such that the particular confidence interval contained the true value $\mu = 0$ was recorded and plotted in Figure 1.

4. Discussion and conclusions

In this paper we have suggested to use confidence intervals for the common mean in interlaboratory trials based on various methods. The presented simulation study shows that the suggested confidence intervals based on simple t-statistic and Welch's method have very good coverage properties for almost all cases. The method based on Satterthwaite's degrees of freedom has good coverage properties whenever the number of observations in one laboratory is sufficiently large (say 10 observations per one laboratory) or number of laboratories is increasing. The method based on Kenward and Roger [1] does not have good properties for this model with small number of observations in one laboratory.

The method based on Welch quantile approximation has the best properties, but also the simple t-statistic gives very good results, except for the case with small number of laboratories and small number of observations in one laboratory in heteroscedastic case ($k = 2$ and $n_i = 5$ for all i). In addition this method is very easily computable.

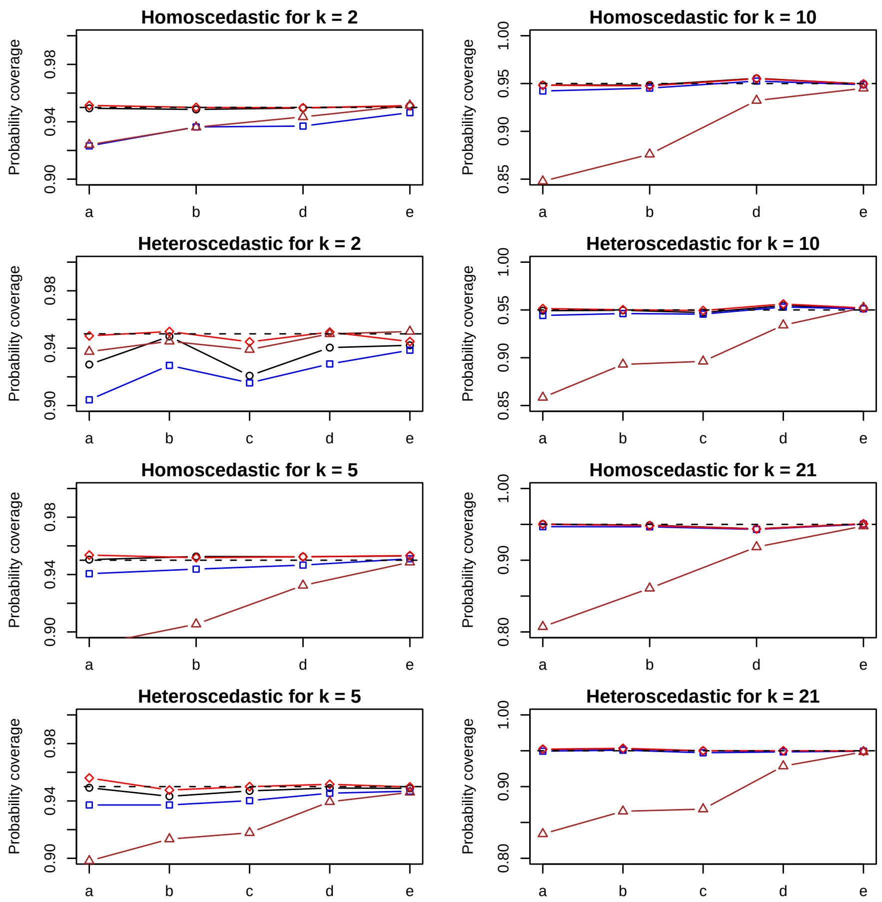

Fig. 1. The empirical coverage probabilities of the 95% *confidence intervals* based on 5,000 Monte Carlo runs for each specific design, where methods are simple t-statistic ○, Satterthwaite □, Welch ◇ and Kenward-Roger △. The letter "a" represents that number of laboratories is $n_i = 5$, "b" represents $n_i = 5, 10$, "c" represents $n_i = 5, 10$ with alternative pattern of variances only in heteroscedastic case, "d" represents $n_i = 10$ and "e" $n_i = 30$.

Reference

[1] Kenward, M. G. and Roger, J. H. (1997). Small sample inference for fixed effects from restricted maximum likelihood. *Biometrics*, 53, 983–997.

[2] Satterthwaite, E. F. (1941) Synthesis of variance. *Psychometrika*, 6, 309–316.

[3] Welch, L. B. (1947). The generalization of 'Student's'problem when several different population are involved. *Biometrics*, 34, 28–35.

[4] Savin, A., Wimmer, G. and Witkovský V. (2003). On Kenward–Roger confidence intervals for common mean in interlaboratory trials. *Measurement Science Review, Theoretical Problems Of Measurement, Vol.* 3, 53–56, http://www.measurement.sk.