Tatra Mt. Math. Publ. **39** (2008), 53–60

Tatra Mountains
Mathematical Publications

ESTIMATION OF THE COMMON MEAN AND DETERMINATION OF THE COMPARISON REFERENCE VALUE

Viktor Witkovský — Gejza Wimmer

ABSTRACT. We discuss the problem of evaluating measurement results from interlaboratory comparisons in metrology. Here we permit the laboratories to have systematic errors that are assumed to be uniformly distributed. Two approaches for modeling the measurement results are introduced and compared, and the associated problems connected with estimation of the quantity in question are analyzed. The first is a classical (frequentist) statistical approach resulting in a heteroscedastic one-way random effects model. The second approach is based on metrological methodology. In both cases the comparison reference value (an estimate of the unknown measured quantity) with the approximate interval estimators is proposed.

1. Introduction

Here we consider the metrological problem of combining measurements of an unknown quantity μ from several laboratories and/or measurement methods known in metrology as interlaboratory comparisons. The outcome of the interlaboratory comparisons is the comparison reference value, normally taken as a close approximation to the value μ, see [1]. In this paper, we formulate and illustrate two different approaches for modelling the measurement results from interlaboratory comparisons, see, e.g., the paper by Kacker, Datla and Parr [8]:

- the classical (frequentist) statistical approach resulting in a heteroscedastic one-way random effects model.

2000 Mathematics Subject Classification: Primary 62J05; Secondary 62E15.
Keywords: common mean problem, interlaboratory comparisons, comparison reference value.
This research was supported by grants No. 1/3016/06 and 2/7087/27 of the Scientific Grant Agency of the Slovak Republic, Project APVV RPEU-0008-06 and Project MŠMT ČR, LC 06024.

- the approach based on metrological methodology.

In both cases the comparison reference value (an estimate of the unknown measured quantity) is proposed together with the interval estimator for μ. Both approaches are numerically illustrated by an example of the key interlaboratory comparisons on charge sensitivity of the back-to-back accelerometer for 500 Hz, taken from the Final report on key comparison CCAUV.V-K1 [11].

2. Classical (frequentist) statistical approach

Let $k \geq 2$ be the number of independent laboratories or measurement methods. Assume that each laboratory repeats independently n_i times the measurement of the same object (quantity), μ being the true value, $n_i \geq 2$, $i = 1, \dots, k$.

For the interlaboratory comparisons we will consider the following heteroscedastic one-way random effects model:

$$Y_{ij} = \mu + B_i + \varepsilon_{ij}\,. \tag{1}$$

Y_{ij} denotes the jth measurement in the ith laboratory, μ represents the (unknown) common mean, $\varepsilon_{ij} \sim \mathcal{N}(0, \sigma^2_{(A),i})$ is the mutually independent normally distributed measurement error, $\sigma_{(A),i}$ being the unknown standard deviation. $B_i \sim \mathcal{U}_i(-\delta_i, \delta_i)$, $i = 1, \dots, k$ are the laboratory biases (systematic errors) which are independently uniformly distributed with $\delta_i = \sqrt{3}\,\sigma_{(B),i}$, where $\sigma_{(B),i}$ are the known standard deviations.

The model (1) with $B_i \sim \mathcal{N}(0, \sigma_0^2)$ and $\sigma^2_{(A),1} = \dots = \sigma^2_{(A),k}$ is basically the one-way random-effects ANOVA model, broadly studied by Searle [10], Graybill and Hultquist [4], Graybill [3], Harville [6], and others. The problem of deriving the confidence interval for the common mean μ in the model (1) with heteroscedastic errors was studied, e.g., by Rukhin and Vangel [9], Hartung, Böckenhoff and Knapp [5], Iyer, Wang and Mathew [7], and Wang and Iyer [12].

Here we suggest an approximate $(1-\alpha)\times 100\%$ confidence interval for the common mean μ based on an approach similar to that one used by Fairweather in [2]. The outcome of the experiment is given by the laboratory sample means (the estimators of $\mu_i = \mu + b_i$, where by b_i we denote the realization of random variable B_i) and sample variances (the estimators of $\sigma^2_{(A),i}$):

$$\bar{Y}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} Y_{ij}\,,$$

and

$$S_i^2 = \frac{1}{n_i - 1} \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_i)^2.$$

Note that $\bar{Y}_i$ and S_i^2, for $i = 1, \ldots, k$, are mutually independent random variables. Assuming the model (1) we have:

$$(\bar{Y}_i - \mu) \sim B_i + \bar{\varepsilon}_i, \quad \text{independent of} \quad \frac{(n_i - 1)S_i^2}{\sigma_{(A),i}^2} \sim \chi^2_{n_i - 1}\,, \tag{2}$$

with

$$\bar{\varepsilon}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} \varepsilon_{ij}\,,$$

$$\bar{\varepsilon}_i \sim \mathcal{N}(0, \sigma_{(A),i}^2 / n_i)\,,$$

where B_i and $\bar{\varepsilon}_i$ are independent random variables. Assuming further that $\delta_i = \sqrt{3}\,\gamma_i \sigma_{(A),i}$, where $\gamma_i = \sigma_{(B),i}/\sigma_{(A),i}$ is a known ratio of the standard deviations, we get

$$T_i^* = \frac{\bar{Y}_i - \mu}{\sqrt{\frac{S_i^2}{n_i}}} \sim \frac{\left(\sqrt{3n_i}\,\gamma_i\right) U_i + Z_i}{\sqrt{\frac{Q_i}{(n_i - 1)}}}, \tag{3}$$

where $U_i \sim \mathcal{U}(-1, 1)$, $Z_i \sim \mathcal{N}(0, 1)$, and $Q_i \sim \chi^2_{n_i - 1}$ are mutually independent random variables.

Let us denote

$$W^* = \sum_{i=1}^{k} \omega_i^* T_i^*,$$

where ω_i^* are non-stochastic coefficients, e.g., $\omega_i^* = 1/\mathrm{Var}(T_i^*)$, and let $q^*_{1-\alpha/2}$ denote the $(1 - \alpha/2)$-quantile of the distribution of W^*, such that

$$\Pr\left(|W^*| < q^*_{1-\alpha/2}\right) = 1 - \alpha\,. \tag{4}$$

From that, the exact $(1 - \alpha) \times 100\%$ confidence interval for μ is given by

$$\frac{\sum_{i=1}^{k} \sqrt{\frac{n_i}{S_i^2}}\, \omega_i^* \bar{Y}_i}{\sum_{i=1}^{k} \sqrt{\frac{n_i}{S_i^2}}\, \omega_i^*} \pm \frac{q^*_{1-\alpha/2}}{\sum_{i=1}^{k} \sqrt{\frac{n_i}{S_i^2}}\, \omega_i^*}\,. \tag{5}$$

The theoretical distribution function of $W^* = \sum_{i=1}^{k} \omega_i^* T_i^*$ is intractable as the probability density function $f(t)$ of T_i^* is expressed by non-standard integral

$$f(t) = \frac{2^{-\nu/2}}{\sqrt{\nu}\,\Gamma(\frac{\nu}{2})a} \int_0^\infty e^{-\frac{z}{2}} z^{\frac{\nu-1}{2}} \Big\{ \Phi\big(t\sqrt{z}\,\nu - a\big) - \Phi\big(-t\sqrt{z}\,\nu - a\big) \Big\}\, \mathrm{d}z\,, \qquad (6)$$

$a = \sqrt{3n_i}\,\gamma_i$, $\nu = n_i - 1$, and $\Phi(\cdot)$ being the standard normal cdf. However, for given γ_i, the critical value $q^*_{1-\alpha/2}$ could be approximated by Monte Carlo simulations.

The distribution of W^* depends on the parameters

$$\gamma_i = \sigma_{(B)\,,\,i}/\sigma_{(A)\,,\,i}\,, \quad i = 1, \dots, k$$

which are unknown. They could be naturally estimated by

$$\hat{\gamma}_i = \sigma_{(B),i}/\sqrt{S_i^2}\,.$$

In this situation, a reasonable guess of the $(1-\alpha)\times 100\%$ confidence interval for μ could be obtained by using $\hat{\gamma}_i$ instead of γ_i in (3) and (5).

3. Metrological approach

The outcome of the interlaboratory comparisons is the comparison reference value—an estimate of μ. The metrological approach combines the posterior information (in the form of the state-of-knowledge distributions, see, e.g., [8]) about the true value of the μ, given the observed data from each of the laboratories. Although this approach is closely related to the methods of the Bayesian statistical inference, it is not a fully Bayesian solution to the problem of estimation of the parameter μ. In this approach, the value of the comparison reference value is frequently given as a weighted mean or an arithmetic mean of the laboratory sample means $\bar{y}_i$.

Assuming the model (1), let $\mu_i = \mu + b_i$ denote the value of the measurand drifted by the systematic laboratory effect (b_i represents the realization of the random variable B_i). The value b_i is directly unobservable, and so, it remains to be an unknown constant. However, if we know the true value of the ith laboratory mean μ_i, then our knowledge about the true value of the measurand μ based on the full available information (i.e., the model (1) and the information on μ_i from the ith laboratory) is given by the probability distribution of the random variable

$$\tilde{\mu}_{(i)} = \mu_i - B_i\,. \qquad (7)$$

The value of the parameter μ_i is unknown and could be estimated by the ith laboratory sample mean $\bar{Y}_i$ together with its sample standard deviation, which is given by $\sqrt{S_i^2/n_i}$. Note that under the model assumptions (1) the random variable $T_i = (\bar{Y}_i - \mu_i)/\sqrt{S_i^2/n_i}$ has the Student's t distribution with $n_i - 1$ degrees of freedom.

Given the observed values of the sample statistics $\bar{y}_i$ and s_i^2, our knowledge about μ_i could be represented by the probability distribution of the random variable

$$\tilde{\mu}_i = \bar{y}_i - \sqrt{\frac{s_i^2}{n_i}}\, T_i\,, \tag{8}$$

Assuming $B_i \sim \mathcal{U}(-\delta_i, \delta_i)$ with $\delta_i = \sqrt{3}\,\sigma_{(B),i}$, and by combining (7) and (8), we can express our knowledge about the true value of the measurand μ (based on the information from the ith laboratory) by the probability distribution of the random variable

$$\tilde{\tilde{\mu}}_{(i)} = \tilde{\mu}_i - B_i = \bar{y}_i - \sqrt{\frac{s_i^2}{n_i}} T_i - B_i\,. \tag{9}$$

In the final step of this metrological approach the state-of-knowledge distribution about μ, based on information from all laboratories, is expressed by the probability distribution of a random variable $\tilde{\tilde{\mu}}$, which is a weighted mean of the random variables $\tilde{\tilde{\mu}}_{(i)}$

$$\tilde{\tilde{\mu}} = \sum_{i=1}^{k} w_i \tilde{\tilde{\mu}}_{(i)} = \sum_{i=1}^{k} w_i \bar{y}_i - \sum_{i=1}^{k} w_i \sqrt{\frac{s_i^2}{n_i}}\, T_i - \sum_{i=1}^{k} w_i \sqrt{3}\,\sigma_{(B)\,,\,i} U_i\,, \tag{10}$$

where w_i, $\sum_{i=1}^{k} w_i = 1$, are properly chosen weights, and $U_i \sim \mathcal{U}(-1, 1)$. The natural weights are those that are inversely proportional to $\mathrm{Var}\big(\tilde{\tilde{\mu}}_{(i)}\big)$, however, we suggest to use the weights w_i

$$w_i = \frac{\dfrac{1}{\left(\sqrt{\frac{s_i^2}{n_i}}\sqrt{\frac{s_0^2}{n_i}}\,\frac{n_i-1}{n_i-3} + \sigma^2_{(B),i}\right)}}{\sum_{l=1}^{k}\left(\dfrac{1}{\sqrt{\frac{s_l^2}{n_l}}\sqrt{\frac{s_0^2}{n_l}}\,\frac{n_l-1}{n_l-3} + \sigma^2_{(B),l}}\right)}\,, \tag{11}$$

where s_0^2 is the pooled variance estimate

$$s_0^2 = \sum_{i=1}^{k} (n_i - 1) s_i^2 \Big/ \left(\sum_{i=1}^{k} n_i - k\right).$$

Given the observed values of the sample statistics $\bar{y}_i$ and s_i^2, $i = 1, \dots, k$, the comparison reference value, say μ_{CRV}, is given as the expected value of the random variable $\tilde{\tilde{\mu}}$, i.e.,

$$\mu_{CRV} = \sum_{i=1}^{k} w_i \bar{y}_i \,. \tag{12}$$

The interval

$$\left\langle \mu_{CRV} + q_{\alpha/2},\ \mu_{CRV} + q_{1-\alpha/2} \right\rangle , \tag{13}$$

could be considered as a reasonable approximation of the $(1-\alpha) \times 100\%$ confidence interval estimate for μ, where $q_{\alpha/2}$ and $q_{1-\alpha/2}$ are the quantiles of the distribution $\tilde{\tilde{\mu}} - \mu_{CRV}$. These quantiles could be exactly evaluated by the algorithm `tdist`, for more details on an earlier version of the algorithm see W i t k o v s k ý [13].

4. Example

TABLE 1. Sample means, number of replications, and corresponding standard deviations s_i and $\sigma_{(B),i}$ of charge sensitivity measurements of the back-to-back accelerometer for 500 Hz.

No.	Laboratory	Country	$\bar{y}_i$	n_i	s_i	$\sigma_{(B),i}$
1	PTB	Germany	0.12662	9	0.0000429	0.0000617
2	BNM-CESTA	France	0.12690	5	0.0005477	0.0003164
3	CSIRO-NML	Australia	0.12670	5	0.0000837	0.0001864
4	CMI	Czech Republic	0.12670	5	0.0002321	0.0003260
5	CSIR-NML	South Africa	0.12710	5	0.0000837	0.0003795
6	CENAM	Mexico	0.12657	5	0.0000826	0.0003142
7	NRC	Canada	0.12650	5	0.0002688	0.0002650
8	KRISS	Korea	0.12659	6	0.0000361	0.0002274
9	NMIJ	Japan	0.12655	4	0.0000818	0.0003137
10	VNIIM	Russia	0.12694	5	0.0001140	0.0002746
11	NIST	United States	0.12640	5	0.0002000	0.0001954
12	Nmi-VSL	The Netherlands	0.12662	5	0.0001171	0.0001560

The data taken from the *Final report on key comparison CCAUV.V-K1* are presented in Table 1, see [11]. The key interlaboratory comparisons were taken by 12 national metrology institutes in the area of vibration (quantity of acceleration) on the measurements of the charge sensitivity of the accelerometer standards (back-to-back accelerometer) at different frequencies and acceleration amplitudes. For chosen $\alpha = 0.05$, the resulted interval estimate (5) for μ is $0.1266369 \pm 1.8238\text{e-}004$, the interval estimate (13) is $0.1266327 \pm 0.9628\text{e-}004$. The interval (5) was evaluated with estimated $\hat{\gamma}_i = \sigma_{(B),i}/\sqrt{s_i^2}$, the required variances $\mathrm{Var}(T_i^*)$ and the quantiles were calculated from simulated values of T_i^*, $i = 1, \dots, k$.

Further, 10000 realizations from the model (1) were simulated with $\mu = 0$, and the true parameters n_i, $\sigma_{(A),i}$, and $\sigma_{(B),i}$ were taken from the Table 1, i.e., we set $\sigma_{(A),i} = \sqrt{s_i^2}$.

The confidence interval (5) with $\gamma_i = \sigma_{(B),i}/\sigma_{(A),i}$ is an exact one with empirical coverage probability 0.9507 and average length 0.0003083. If we use $\hat{\gamma}_i$ instead of γ_i in (5), the interval seems to be too conservative, with the coverage probability 0.9962 and average length 0.0003744.

The approximate interval (13) with the weights (11) indicates good properties, with the coverage probability 0.9534 and average length 0.0001919. Further research is necessary for full characterization of the frequentist behaviour of the interval estimator (13) based on the metrological approach. However, based on our preliminary simulation studies, the metrological approach is superior to the above mentioned frequentist approach and might be applied to construct a confidence interval on the common mean in the one-way, heteroscedastic, random-effects model, in situation when the random effects are characterized by the completely known probability distributions.

REFERENCES

[1] CIPM—BIPM. : *Mutual recognition of national measurement standards and of calibration and measurement certificates issued by national metrology institutes with Technical Supplement revised in October 2003.* Comit international des poids et mesures (CIPM), Bureau international des poids et mesures (BIPM), Paris, 14 October, 1999.

[2] FAIRWEATHER, W. R.: *A method of obtaining an exact confidence interval for the common mean of several normal populations*, Appl. Stat. **21** (1972), 229–233.

[3] GRAYBILL, F. A.: *Theory and Application of the Linear Model*. Duxbury, North Scituate, Massachusetts, 1976.

[4] GRAYBILL, F. A.—HULTQUIST, R. A.: *Theorems concerning Eisenhart's Model II*, Ann. Math. Stat. **32** (1976), 261–269.

[5] HARTUNG, J.—BÖCKENHOFF, A.—KNAPP, G.: *Generalized Cochran-Wald statistics in combining of experiments*, J. Statist. Plann. Inference **113** (2003), 215–237.

[6] HARVILLE, D. A.: *Maximum-likelihood approaches to variance component estimation and to related problems*, J. Amer. Statist. Assoc. **72** (1977), 320–340.

[7] IYER, H. K.—WANG, C. M. J.—MATHEW, T.: *Models and confidence intervals for true values in interlaboratory trials*, J. Amer. Statist. Assoc. **99** (2004), 1060–1071.

[8] KACKER, R. N.—DATLA, R. U.—PARR, A. C.: *Statistical interpretation of key comparison reference value and degrees of equivalence*, J. Res. NIST **108** (2003), 439–446.

[9] RUKHIN, A. L.—VANGEL, M. G.: *Estimation of a common mean and weighted means statistics*, J. Amer. Statist. Assoc. **93** (1998), 303–308.

[10] SEARLE, S. R.: *Linear Models*. John Wiley & Sons, Inc., New York, 1971.

[11] VON MARTENS, H.-J.—ELSTER, C.—LINK, A.—TUBNER, A.—WABINSKI, W.: *Final report on key comparison CCAUV.V-K1*, Metrologia **40** (2003), Tech. Suppl. 09001.

[12] WANG, C. M. J.—IYER, H. K.: *A generalized confidence interval for a measurand in the presence of type-A and type-B uncertainties*, Measurement **39** (2006), 856–863.

[13] WITKOVSKÝ, V.: *Matlab algorithm TDIST: The distribution of a linear combination of Student's t random variables*. In: Proceedings in Computational Statistics––COMPSTAT '04 (J. Antoch, ed.), 16th Symposium, Prague, Czech Republic, Physica-Verlag, 2004, pp. 1995–2002.

Received October 16, 2006

Viktor Witkovský
Institute of Measurement Science
Slovak Academy of Sciences
Dúbravská cesta 9
841 01 Bratislava
SLOVAKIA
E-mail: witkovsky@savba.sk

Gejza Wimmer
Mathematical Institute
Slovak Academy of Sciences
Štefánikova 49
814 73 Bratislava
SLOVAKIA

Matej Bel University
Banská Bystrica
SLOVAKIA

Masaryk University Brno
CZECH REPUBLIC
E-mail: wimmer@mat.savba.sk