Performance metrics for testing statistical calculations in interlaboratory comparisons

Acko, B.[a,*], **Sluban, B.**[a], **Tasič, T.**[a], **Brezovnik, S.**[b]

[a]Faculty of Mechanical Engineering, University of Maribor, 2000 Maribor, Slovenia
[b]Gorenje gospodinjski aparati, d. d., 3503 Velenje, Slovenia

A B S T R A C T

Interlaboratory comparisons are the most powerful tools for determining the competences of laboratories performing calibrations and testing. Performance metrics is based on statistical analysis, which can be very complex in certain cases, especially for testing where transfer standards (samples) are prepared by the pilot laboratory. Statistical quantities are calculated using different kinds of software, from simple Excel applications to universal or specific commercial programmes. In order to ensure proper quality of such calculations, it is very important that all computational links are recognized explicitly and known to be operating correctly. In order to introduce a traceability chain into metrology computation, the European project EMRP NEW 06 TraCIM was agreed between the EC and the European Metrology Association (EURAMET). One of the tasks of the project was also to establish random datasets and validation algorithms for verifying software applications in regard to evaluating interlaboratory comparison results. The statistical backgrounds for resolving this task, and the basic concept of the data generator are presented in this paper. Background normative documents, calculated statistical parameters, boundary conditions for generating reference data sets are described, as well as customer interface.

References

[1] Cox, M.G. (2002). The evaluation of key comparison data, *Metrologia,* Vol. 39, No. 6, 589-595, doi: 10.1088/0026-1394/39/6/10.

[2] BIPM (2014). Guide for implementation of the CIPM MRA, CIPM MRA-G-01, v. 1.1, from *http://www.bipm.org/en/cipm-mra/documents*, accessed January 10, 2014.

[3] ISO 13528 (2005). *Statistical methods for use in proficiency testing by interlaboratory comparisons*, ISO copyright office, Geneva.

[4] Forbes, A. (2012). *NEW06 TraCIM – Traceability of computationally-intensive metrology, EMRP JRP Protocol*, NPL, Teddington.

[5] Cox, M.G. (2007). The evaluation of key comparison data: determining the largest consistent subset, *Metrologia*, Vol. 44 , No. 3, 187-200, doi: 10.1088/0026-1394/44/3/005.

[6] Cox, M.G., Harris, P.M. (2012). The evaluation of key comparison data using key comparison reference curves, *Metrologia*, Vol. 49, No. 4, 437-445, doi:10.1088/0026-1394/49/4/437.

[7] Acko, B., Brezovnik, S., Sluban, B. (2013). Verification of software applications for evaluating interlaboratory comparison results, In: *Annals of DAAAM International for 2013, Collection of working papers for 24th DAAAM International Symposium*, DAAAM International Vienna, Vienna, 9 pages.

[8] ISO 5725-1 (2001). *Accuracy (trueness and precision) of measurement methods and results – Intermediate measures of the precision of a standard measurement method*, ISO copyright office, Geneva.

[9] Härtig, F., Kniel, K. (2013). Critical observations on rules for comparing measurement results for key comparisons, *Measurement,* Vol. 46, No. 9, 3715-3719, doi: 10.1016/j.measurement.2013.04.079.

[10] Acko, B. (2012). Final report on EUROMET key comparison EUROMET L-K7: calibration of line scales, *Metrologia*, Vol. 49, No. 1A, Technical Supplement, doi: 10.1088/0026-1394/49/1A/04006.

[11] Raczynski, S. (2011). Uncertainty, dualism and inverse reachable sets, *International Journal of Simulation Modelling*, Vol. 10, No. 1, 38-45, doi: 10.2507/IJSIMM10(1)4.180.

Merila zmogljivosti za testiranje statističnih izračunov v medlaboratorijskih primerjavah

Acko, B.[a,*], **Sluban, B.**[a], **Tasič, T.**[a], **Brezovnik, S.**[b]

[a]Faculty of Mechanical Engineering, University of Maribor, 2000 Maribor, Slovenia
[b]Gorenje gospodinjski aparati, d. d., 3503 Velenje, Slovenia

POVZETEK

Medlaboratorijske primerjave so najučinkovitejše orodje za določanje usposobljenosti laboratorijev, ki izvajajo umerjanje in preskušanje. Merjenje zmogljivosti temelji na statistični analizi, ki je v nekaterih primerih zelo kompleksna. Posebej zahtevna je obdelava podatkov pri preskušanju tedaj, ko prenosne etalone (vzorce) pripravi pilotski laboratorij. Statistične veličine računamo z različnimi programskimi orodji, od preprostih aplikacij v Excelu do univerzalnih ali namenskih komercialnih programov. Za zagotovitev ustrezne kakovosti takšnih izračunov je zelo pomembno, da eksplicitno prepoznamo vse računske povezave in se prepričamo, da delujejo pravilno. Da bi omogočili vpeljevanje sledljivostnih verig v meroslovne izračune, sta se Evropska komisija in Evropska meroslovna organizacija EURAMET odločili za izvedbo evropskega projekta EMRP NEW 06 TraCIM. Ena od nalog tega projekta je tudi izdelava naključnih naborov podatkov in validacijskih algoritmov za preverjanje programskih aplikacij, ki se uporabljajo za vrednotenje rezultatov medlaboratorijskih primerjav. V članku so predstavljene statistične osnove za reševanje te naloge in osnovni koncept generatorja podatkov. Opisani so uporabljeni normativni dokumenti, izračunani statistični parametri, robni pogoji za generiranje referenčnih naborov podatkov in uporabniški vmesnik.