14th IMEKO TC10 Workshop Technical Diagnostics
New Perspectives in Measurements, Tools and Techniques
for system's reliability, maintainability and safety
Milan, Italy, June 27-28, 2016

A MATLAB Framework for Measurement System Analysis based on ISO 5725 Standard

Giuseppe Maria D'Aucelli, Nicola Giaquinto, Sabino Mannatrizio[1] and Mario Savino

DEI, Politecnico di Bari, Via Orabona 4, 70125 Bari, Italy - giuseppemaria.daucelli@poliba.it

Abstract **– In this paper, a MATLAB framework for automated Measurement System Analysis (MSA) procedures according to ISO 5725 standard is presented. Data consistency tests and their relations with the statistical model used in the standard are illustrated, with sample output from the corresponding MATLAB functions. Moreover, the ANOVA framework is introduced in the MSA workflow as recommended in the ISO/TR 22971 standard. Measurement system performance assessment is, finally, performed using the framework and given in terms of repeatability, interlaboratory and reproducibility variances.**

I. INTRODUCTION

The ISO 5725 standard "Accuracy (trueness and precision) of measurement methods and results" defines a procedure of statistical testing for accuracy and precision assessment in interlaboratory measurement campaigns. It consists of six parts:

1. ISO 5725-1 "General principles and definitions";
2. ISO 5725-2 "Basic method for the determination of repeatability and reproducibility of a standard measurement method";
3. ISO 5725-3 "Intermediate measures of the precision of a standard measurement method";
4. ISO 5725-4 "Basic methods for the determination of the trueness of a standard measurement method";
5. ISO 5725-5 "Alternative methods for the determination of the precision of a standard measurement method";
6. ISO 5725-6 "Use in practice of accuracy values".

The standardization involves both the technical aspects of measurement campaigns and the validation of their results when such campaigns are performed under repeatability and reproducibility conditions (R&R). In the latter, such factors as the environmental conditions, the operator and the calibration of the equipment contribute to the variability of test results.

Therefore, the need to simplify the treatment of interlaboratory measurement data according to the ISO standard and possibly to automate this process arises. The automated routine should be able to produce a set of statistical outputs which need to be easy to interpret and should provide:

- Overall performance assessment for the measurement system under R&R conditions;
- Specific alerts about data consistency (e.g. outliers)

From a technical point of view, the proposed approach to statistical testing and, therefore, to certification of measurement systems (Measurement System Analysis, MSA) aims to validate such systems to show their suitability for the characterization of production apparata operating in particularly sensible contexts (e.g. security related or other applications which are explicitly regulated by the law [1]).

It should also be noted that the approach followed in ISO 5725 standard straightforwardly applies to the problem of quantifying the contribution of systematic errors to the overall measurement uncertainty, also in contexts different from interlaboratory experiments [2] [3]. It is a "frequentist" and operational approach to uncertainty evaluation, which can be hardly considered compatible with the Bayesian view prescribed by the Supplement 1 of the GUM, but is not prone to the paradoxes often arising in that context [4] [5].

II. MATLAB FRAMEWORK FOR ISO 5725

ISO 5725 standard proposes [6] a detailed processing workflow for statistical processing, which starts from a standard data entry layout and provides guidelines for outlier or other nonconsistency detection. After the elimination or substitution of inconsistent data, Reproducibility and Repeatability outputs can be produced.

Such testing procedure has been implemented in MATLAB environment. The program accepts a spreadsheet as input containing a set of q sheets, one for each measurement level. Each sheet must contain, by column, the measurement data coming from every laboratory involved in the test.

The output of the program consists in two spreadsheets:

- Consistency test results and R&R outputs;
- Log file outlining detected anomalies (if any)

[1] Sabino Mannatrizio is currently Verification & Validation Engineer at Mer Mec S.p.A., www.mermecgroup.com

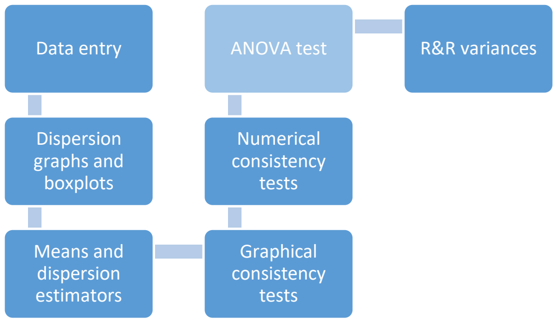

Figure 1 - Block diagram for ISO 5727 procedure with ANOVA test (ISO/TR 22971)

Moreover, boxplots and dispersion graphs are produced for each measurement level, along with h and k Mandel's statistics plots. Some samples of these graphical outputs are presented in Section III-A.

The main script is a wrapper for the functions which independently perform statistical tests. A comprehensive framework has been developed, in which each function is capable of performing a single test. The functions make use of statistical tables in the public domain [7], [8], **Error! Reference source not found.**, which are the same used in [10]. Currently available functions are:

- *SCboxplot, SCdispersionplot*: provide tools for boxplots and dispersion graphs
- *SChMandel, SCkMandel*: compute h and k Mandel's statistics and compare results with the critical values reported in the ISO standard
- *SCcochran, SCgrubbs*: perform Cochran and Grubbs consistency tests
- *SCreprip*: produces repeatability and reproducibility outputs
- *SCanova*: performs "one-way between groups ANOVA test"

The ANOVA test routine has been added although it is not directly related to ISO 5725 testing procedure to strenghten consistency tests. In fact, the ANOVA finds straightforward application in MSA [11], and further details will be given in Section III-E.

In other words, the testing workflow schematized in Figure 1 is implemented by the aforementioned functions.

With the perspective of collaborative software development, a Git repository for the framework has been created and is available online[2].

A. Interlaboratory experiments and statistical model

Before examining in depth the tests implemented by the MATLAB functions, the statistical model for the interlaboratory measurement campaigns proposed by the ISO 5725 standard must be outlined. The k-th measurement performed by the i-th laboratory on the same unknown quantity x_j is modeled by the following equation:

$$y_{ik}(x_j) = m(x_j) + B_i(x_j) + e_{ik}(x_j) \quad (1)$$

where:

- m is the grand mean between the laboratories
- B_i is the (systematic) bias of the measurements performed at the i-th laboratory
- e_{ik} is the random error affecting each measurement under repeatability conditions

Starting from the error components in Eq. (1) the estimators for repeatability and reproducibility variances are determined:

- Interlaboratory variance $\sigma_L^2 = var(B_i)$
- Intralaboratory variance $\sigma_{Wi}^2 = var_k(e_{ik})$
- Repeatability variance $\sigma_r^2 = \overline{var_k(e_{ik})} = \overline{\sigma_W^2}$
- Repeatability standard deviation $\sigma_r = \sqrt{\overline{var_k(e_{ik})}}$
- Reproducibility standard deviation $\sigma_R = \sqrt{\sigma_L^2 + \sigma_r^2}$

It should be noted that the intralaboratory variance σ_{Wi}^2 refers to the variability with respect to the k index and is, in general, different between laboratories. The repeatability variance σ_r^2 is, on the other hand, the arithmetic mean of the intralaboratory variances and is therefore a global statistic for the measurement campaign.

III. SOFTWARE ROUTINES

A. Means and dispersion estimators

The first issue feature of the framework which needs to be outlined is the data entry model. Measurement data need to be put on a suitably formatted spreadsheet so they can be automatically imported in the MATLAB environment. In a typical scenario, data collected by each of the p laboratories involved in the MSA process are divided into n repetitions of measurement performed for each of the q levels of the measurand. The grand total of measurement is therefore $N = pqn$ which must be organized as in Table 1.

The cell size n has (i, j) subscripts which reminds that however it is advisable to collect samples of the same size n_{ij} for each cell, this condition is neither guaranteed nor required.

[2] https://bitbucket.org/misurepoliba/iso5725test/src

Table 1. Data entry model.

	x_1	x_2	$\cdots$	x_j	$\cdots$	x_q
Lab 1	$y_1^{(1,1)}$ $y_2^{(1,1)}$ $\vdots$ $y_{n_{1,1}}^{(1,1)}$	$y_1^{(1,2)}$ $y_2^{(1,2)}$ $\vdots$ $y_{n_{1,2}}^{(1,2)}$	$\cdots$	$y_1^{(1,j)}$ $y_2^{(1,j)}$ $\vdots$ $y_{n_{1,j}}^{(1,j)}$	$\cdots$	$y_1^{(1,q)}$ $y_2^{(1,q)}$ $\vdots$ $y_{n_{1,q}}^{(1,q)}$
Lab 2	$y_1^{(2,1)}$ $y_2^{(2,1)}$ $\vdots$ $y_{n_{2,1}}^{(2,1)}$	$\ddots$				$y_1^{(1,q)}$ $y_2^{(1,q)}$ $\vdots$ $y_{n_{1,q}}^{(1,q)}$
$\vdots$	$\vdots$		$\ddots$			$\vdots$
Lab i	$y_1^{(i,1)}$ $y_2^{(i,1)}$ $\vdots$ $y_{n_{i,1}}^{(i,1)}$			$y_1^{(i,j)}$ $y_2^{(i,j)}$ $\vdots$ $y_{n_{i,j}}^{(i,j)}$		
$\vdots$	$\vdots$				$\ddots$	$\vdots$
Lab p	$y_1^{(p,1)}$ $y_2^{(p,1)}$ $\vdots$ $y_{n_{p,1}}^{(p,1)}$	$\cdots$				$y_1^{(p,q)}$ $y_2^{(p,q)}$ $\vdots$ $y_{n_{p,q}}^{(p,q)}$

The statistical analysis of interlaboratory data begins by computing the mean and variance for each laboratory and measurand level, using their classic estimators:

- Cell mean: $\overline{y_{ij}} = \frac{1}{n_{ij}} \sum_{k=1}^{n_{ij}} y_k^{(i,j)}$
- Cell variance: $s_{ij}^2 = \frac{1}{n_{i,j}-1} \sum_{k=1}^{n_{ij}} \left(y_k^{(i,j)} - \overline{y_{ij}}\right)^2$

It is also advisable to perform a preliminary graphical analysis on data dispersion, using well known tools like dispersion graphs and boxplots (Figure 2).

B. Graphical consistency tests

At this stage, single outliers or outlying data sets need to be detected and treated [12]. In this Section a graphical approach based on h and k Mandel statistics is presented. The framework functions SC*hMandel* and SC*kMandel* automatically perform such tests and produce graphical outputs.

Mandel's h statistic for the i-th laboratory and the j-th level is computed as in Eq. (2).

$$h_{ij} = \frac{\overline{y_{ij}} - \overline{\overline{y_j}}}{\sqrt{\frac{1}{p_j - 1} \sum_{i=1}^{p_j} \left(\overline{y_{ij}} - \overline{\overline{y_j}}\right)^2}} \qquad (2)$$

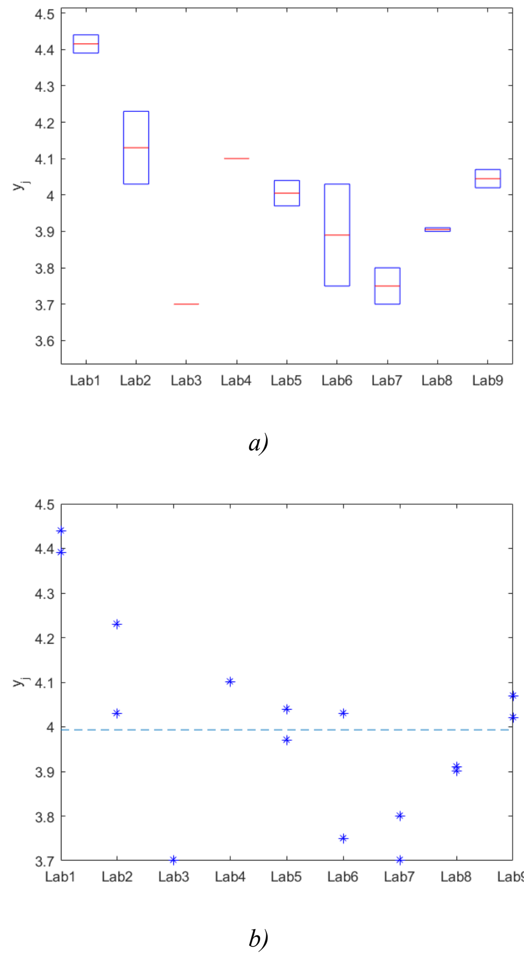

Figure 2 - Boxplot (a) and dispersion graph (b)

Where $\overline{\overline{y_j}} = \frac{\sum_{i=1}^{p} n_{ij} \overline{y_{ij}}}{\sum_{i=1}^{p} n_{ij}}$ is the grand mean for level j or, in other words, an estimator of m in Eq. (1), and p_j is the number of laboratories which, after the preliminary data analysis, have kept at least a valid measurement value for the j-th level. The h_{ij}, ordered by level and grouped by laboratory, must be compared with the critical values corresponding to 1% and 5% significance levels, as shown in Figure 2.

Mandel's h statistic depends on the difference between the intralaboratory mean and the grand mean. It can be observed that the numerator in Eq (2), $\overline{y_{ij}} - \overline{\overline{y_j}}$, is an estimator of B_i in Eq. (1) for the j-th measurement level.

The sign of Mandel's h statistic can be either positive or negative, and determine each laboratory's trend to overestimate or underestimate the measurements with respect to the grand mean.

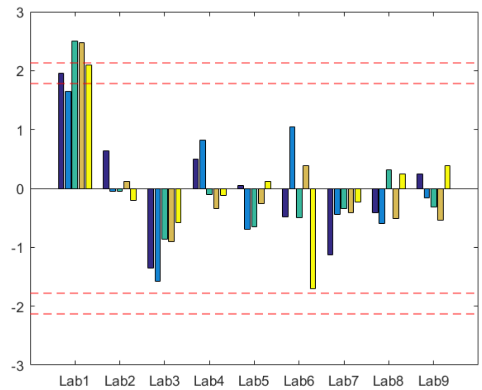

Figure 3 - Mandel's h statistic

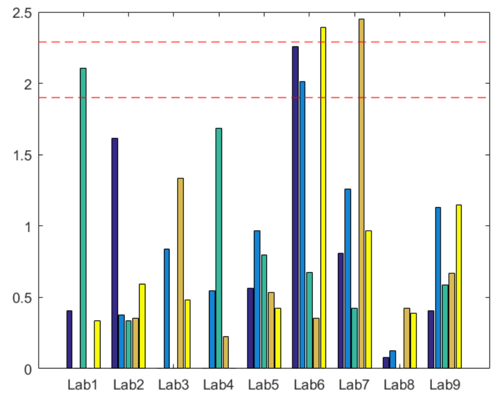

Figure 4 - Mandel's k statistic

Although no result alone can be considered abnormal, some specific scenarios need particular attention. One of those arises, e.g., when individual laboratories over/underestimate x_j for all the q levels.

Mandel's k statistic is, on the other hand, computed as follows:

$$k_{ij} = \frac{s_{ij}}{\sqrt{\frac{1}{p_j}\sum_{i=1}^{p_j} s_{ij}^2}} \tag{3}$$

This statistic represents the variability of measurements taken by the i-th laboratory with respect to the average interlaboratory variance. Therefore, k_{ij} is always positive and can be interpreted as a figure of merit of the i-th laboratory with respect to global measurement dispersion.

The numerator of Eq. (3) is the natural estimator of $var_k[e_{ik}]$ for the j-th level in Eq. (1).

High values of k_{ij} indicate poor measurement repeatability in the i-th laboratory with respect to grand mean repeatability. On the contrary, low values may depend on a number of factors such as too rough measurement rounding or poor sensitivity.

When h or k graphs, grouped by laboratory, exhibits many values close to the critical values, it is advisable to examine the corresponding graph grouped by level. Such analysis might help detecting inconsistent data coming from one or more laboratories. These laboratories need to be directly contacted for further details.

C. Numerical consistency tests

ISO 5725 standard also requires more synthetic numerical consistency tests (i.e. single numerical output for each measurand level) [6]. Cochran and Grubbs tests, similarly to Mandel's statistics, provide respectively intralaboratory and interlaboratory results.

These tests are implemented by *SCcochran* and *SCgrubbs* framework functions.

Cochran's homoscedasticity test studies, for each j-th level, the degree of heterogeneity between cell variances. The C_j statistic simply considers the greatest cell variance, as shown in Eq. (4).

$$C_j = \frac{\max_i\left[s_{ij}^2\right]}{\sum_{i=1}^{p_j} s_{ij}^2} \tag{4}$$

This test is suitable to MSA because it one-sidedly detects data classes with excessively high variance. Any other choice would somehow disadvantage laboratories whose results appear to be more precise.

C_j has straightforward application in triggering further investigation on the laboratory which possibly produced the abnormal value, but needs more complex interpretation in scenarios where more than a few laboratories produce irregular values.

Finally, although Cochran's test loses its significance when variances are estimated on samples of different size, ISO 5725 standard postulates its applicability in two specific conditions:

- $n_{mj} = n_{nj}\ \forall(m,n)$
- $n_{mj} \simeq n_{nj}\ \forall(m,n)$

The latter occurs when all n_{ij} sizes are great in comparison with the number of measurements discarded in the previous steps.

D. Repeatability and reproducibility variances

After data consistency tests have been performed, repeatability and reproducibility variances must be computed [13]. Such task is performed by the *SCreprip* function.

Repeatability variance, defined in Eq. (5), provides an estimate of the measurement system's suitability for the j-th measurand level among all p laboratories, assuming all the data coming from a single laboratory. In must be noted, obviously, that cell variances are weighted with the number of consistent data left in the (i, j)-th cell after consistency tests.

$$s_{rj}^2 = \frac{\sum_{i=1}^{p_j}(n_{ij}-1)s_{ij}^2}{\sum_{i=1}^{p_j}(n_{ij}-1)} \qquad (5)$$

The interlaboratory variance is computed with a similar approach, taking into account the variance between cell means and the grand mean. The repeatability variance is subtracted from this results, as shown in Eq. (6).

$$s_{Lj}^2 = \frac{s_{dj}^2 - s_{rj}^2}{\bar{\bar{n}}_j} \qquad (6)$$

where:

$$s_{dj}^2 = \frac{1}{p_j - 1}\sum_{i=1}^{p_j} n_{ij}\left(\overline{y_{ij}} - \bar{\bar{y}}_j\right)^2$$
$$\bar{\bar{n}}_j = \frac{1}{p_j - 1}\left[\sum_{i=1}^{p_j} n_{ij} - \frac{\sum_{i=1}^{p_j} n_{ij}^2}{\sum_{i=1}^{p_j} n_{ij}}\right] \qquad (7)$$

The last output of *SCreprip* is the reproducibility variance. This is the higher bound of the MSA and therefore it is computed as the sum of repeatability and interlaboratory variances.

$$s_{Rj}^2 = s_{rj}^2 + s_{Lj}^2 \qquad (8)$$

The reproducibility variance takes into account every error component defined in Eq. (1) and is therefore a general quantification of the measurement system's performance.

E. ANOVA Test

The repeatability and reproducibility variances are the final outputs of ISO 5725 MSA protocol. However, in many inferential statistics applications data homogeneity is tested using the ANOVA (Analysis Of VAriance) framework. This set of techniques can prove to be an efficient alternative to Cochran and Grubbs tests. As regards the null hypothesis, however, ANOVA behaves more strictly, since homoscedasticity is replaced by the hypothesis according which data come from the same stochastic distribution.

It is however important to provide guidelines about the application of the ANOVA framework to MSA, given the evident similarity between this approach to variance analysis and the consistency tests presented in the previous Sections. Moreover, the ANOVA techniques are recommended in ISO/TR 22971 standard [14]. Table 2 outlines the ANOVA table with a formalism adapted to the MSA context.

Table 2. ANOVA table for MSA

Source of variation	Sum of squares	Degrees of freedom
Between	$SS_{between} = \sum_{i=1}^{p_j} n_j\left(\overline{y_{ij}} - \bar{\bar{y}}_j\right)^2$	$p_j - 1$
Within	$SS_{within} = \sum_{i=1}^{p_j}\sum_{k=1}^{n_{ij}}\left(y_{ij} - \overline{y_{ij}}\right)^2$	$\sum_{i=1}^{p_j} n_{ij} - p_j$
Total	$SS_{total} = \sum_{i=1}^{p_j}\sum_{k=1}^{n_{ij}}\left(y_{ij} - \bar{\bar{y}}_j\right)^2$	$\sum_{i=1}^{p_j} n_{ij} - 1$
	Mean squares	F-ratio
Between	$MS_{between} = \frac{SS_{between}}{p_j - 1}$	$\frac{MS_{between}}{MS_{within}}$
Within	$MS_{within} = \frac{SS_{within}}{\sum_{i=1}^{p_j} n_{ij} - p_j}$	

This table must be applied for each measurand level and each significance level must be evaluated according to its p-value. The data entry model outlined in Section II-A suggests that the most suitable ANOVA model for MSA applications is the "one-way between groups" analysis [15]. It stresses the differences between individual cells and helps detecting abnormal measurement groups.

IV. CONCLUSIONS

In this paper a MATLAB framework for automated MSA procedures according to ISO 5725 standard has been presented. The testing workflow follows the exact sequence and protocol required by the standard. Along with graphical and numerical consistency tests, the ANOVA framework is introduced to achieve a complete overview of data consistency, as required by the ISO/TR 22971 standard. The proposed framework's source code is available in a public Git repository to encourage collaborative development.

REFERENCES

[1] C. Feng, H. Janssen, Y. Feng, and Q. Meng, "Hygric properties of porous building materials: Analysis of measurement repeatability and reproducibility," Building and Environment, vol. 85, pp. 160–172, Feb. 2015.

[2] F. Attivissimo, A. Cataldo, N. Giaquinto, and M. Savino, "Assessment of the uncertainty associated

with systematic errors in digital instruments: an experimental study on offset errors," Measurement Science and Technology, vol. 23, no. 3, p. 035004, 2012.

[3] F. Attivissimo, A. Cataldo, L. Fabbiano, and N. Giaquinto, "Systematic errors and measurement uncertainty: an experimental approach," Measurement, vol. 44, no. 9, pp. 1781–1789, Nov. 2011.

[4] F. Attivissimo, N. Giaquinto, and M. Savino, "A Bayesian paradox and its impact on the GUM approach to uncertainty," Measurement, vol. 45, no. 9, pp. 2194–2202, Nov. 2012.

[5] N. Giaquinto and L. Fabbiano, "Examples of S1 coverage intervals with very good and very bad long-run success rate," Metrologia, vol. 53, no. 2, pp. S65–S73, Apr. 2016.

[6] ISO 5725-2:1994 - Accuracy (trueness and precision) of measurement methods and results – Part 2: Basic method for the determination of repeatability and reproducibility of a standard measurement method. International Organization for Standardization, 1994.

[7] "Indicators for Mandel's h and k statistics at 1% significance level," Applications of statistics in Measurement & Testing, 06-Sep-2009. [Online]. Available: https://goo.gl/wrCNQP [Accessed: 05-May-2016].

[8] "Cochran test critical values," *University of Washington*. [Online]. Available: http://goo.gl/hRCzjY. [Accessed: 05-May-2016].

[9] "Grubbs test critical values." [Online]. Available: http://goo.gl/wI7zp8. [Accessed: 05-May-2016].

[10] ISO 5725-1:1994 - Accuracy (trueness and precision) of measurement methods and results – Part 1: General principles and definitions. International Organization for Standardization, 1994.

[11] A. Zanobini, B. Sereni, M. Catelani, and L. Ciani, "Repeatability and Reproducibility techniques for the analysis of measurement systems," Measurement, vol. 86, pp. 125–132, May 2016.

[12] EA, EUROLAB, and EURACHEM, "Use of proficiency testing as a tool for accreditation in testing." Aug. 2001, pub. ref. EA-3/04 G:2001.

[13] EA Expert group on uncertainty of measurement, "EA guidelines on the expression of uncertainty in quantitative testing." Dec. 2003, pub. ref. EA-4/16 G:2003.

[14] ISO/TR 22971:2005 - Accuracy (trueness and precision) of measurement methods and results — Practical guidance for the use of ISO 5725-2:1994 in designing, implementing and statistically analysing interlaboratory repeatability and reproducibility results. International Organization for Standardization, 2005.

[15] A. M. Kazerouni, "Design and analysis of gauge R&R studies: Making decisions based on ANOVA method," World Academy of Science, Engineering and Technology, vol. 52, pp. 31–35, 2009.