XVIII IMEKO WORLD CONGRESS
Metrology for a Sustainable Development
September, 17 – 22, 2006, Rio de Janeiro, Brazil

STATISTICAL ANALYSIS OF CORRELATED RESULTS IN INTERLABORATORY COMPARISONS

San Andrés MM [1], Becerra L. O. [2], Vallejos G. [3], Quiroga A. [4], Chanchay R.[5], Bermúdez A. [6], Salazar E.[7]

[1] CEM, Madrid, Spain, mmsanandres@cem.es
[2] CENAM, Queretalo, Mexico, lbecerra@cenam.mx
[3] IBMETRO, La Paz, Bolivia, gvallejos@ibmetro.org
[4]INDECOPI, Lima, Peru, aquiroga@indecopi.gob.pe
[5]INEN, Quito, Ecuador, rchanchay@inen.gov.ec
[6]SIC, Bogota, Colombia, abermudez@correo.sic.gov.co
[7]SENCAMER, Caracas, Venezuela, elias.salazar@sencamer.gov.ve

Abstract: The purpose of this report is to show different ways for the analysis of Interlaboratory Comparisons results, evaluated from real results obtained in the Comparison of mass standards within the Program of Co-operation and Technical Attendance UE-CAN in Quality, NºAECR/B7-31/IB/96/0188.
Keywords: Weighted average, Monte Carlo simulation, correlations, least squares, normalized deviation.

1. INTRODUCTION

The present study is based on the analysis of the results of a Comparison of mass standards made in the nominal values of 1 kg, 100 g, 20 g, 5 g and 100 mg. The Comparison was made within the Program of Cooperation and Technical Attendance in quality matters UE-CAN, NºAECR/B7-31/IB/96/0188. This comparison has been piloted by the Centro Español de Metrología (CEM), Co-piloted by the Centro Nacional de Metrología (CENAM) and co-ordinated within the Andean Community (CAN) by the Instituto Boliviano de Metrología (IBMETRO). A total of seven NMIs have participated in the comparison: CEM, CENAM, IBMETRO, INEN, SIC, INDECOPI and SENCAMER.

The measurements were made in CEM at the beginning and at the end of the comparison. The drift of the mass standards, estimated using the difference between the initial and final CEM measurements, is negligible compared with the associated uncertainty.

The travelling standards used were a set of weights Class OIML E2 belonging to SIM, with the following nominal values of 1 kg, 100 g, 20 g, 5 g, and 100 mg.

The density and volume of the 1 kg, 100 g, 20 g, and 5 g standards were measured by the Density Laboratory of CEM. The density and volume for 100 mg weight were provided by SIM.

CEM and CENAM are signatories of the MRA, and their measurements are supported by their declared CMCs.

The intention is to evaluate the Comparison results by different estimators and mathematical algorithms so that the possible influences of correlations between the input values, and the inconsistency in the values that generate the reference value (RV) can be analysed [1].

In this work, some of the models already studied by recognised authors are applied, giving a comparative analysis of results.

In order to verify the consistency of the RV estimated from the weighted mean between the values of CEM and CENAM, the reference value was also estimated using several methods. Procedures such as the arithmetic mean of CEM and CENAM values, the arithmetic mean of both CEM values, and also the least squared method, were applied. The difference between the estimated RV was not significant in most of the cases. Great consistency between the values evaluated by the different methods and the simulated values has been obtained.

2. METHODS

All the standards were circulated among the NMIs. Each NMI measured the mass and its uncertainty for each weight using their own procedures and methods. The measurements were carried out from September 2004 to June 2005.

The measurement method used by all the laboratories was the substitution method. CEM used a least squares procedure to obtain the results by the Gauss-Marcov estimation that uses weights of the same nominal value. CENAM used the subdivision procedure, starting from one kilogram and generating a set of independent measurements modelled by a system of linear equations whose solution was found by least squares, applying the Lagrange multipliers approximation.

Several estimators for the calculation of the RV were used, as the arithmetic average, the weighted average and the least squares, analyzing their uncertainty contributions. The RV was calculated from both CEM values (at the beginning and at the end of the comparison), and from the CEM-CENAM values. The influence of the correlation between the CEM values was studied, analyzing as much the case of independent variables as correlated variables.

These methods consider the correlation between the values determined by CEM, but do not consider the correlation between the different countries with common traceability.

2.1. Weighted average

The reference value was also calculated from a weighted average between both values measured in CEM and the value measured in CENAM. The mathematic model used for the calculation of the reference value is given in the equation (1):

$$e_{ref} = \frac{\left[\frac{e_{CEM1}}{u^2 e_{CEM1}} + \frac{e_{CEM2}}{u^2 e_{CEM2}} + \frac{e_{CENAM}}{u^2 e_{CENAM}}\right]}{\left[\frac{1}{u^2 e_{CEM1}} + \frac{1}{u^2 e_{CEM2}} + \frac{1}{u^2 e_{CENAM}}\right]} + \delta\Delta e_{CEM} \qquad (1)$$

Where Δe_{CEM} is the difference between the values measured at CEM at the beginning and at the end of the comparison. This value is considered as a possible drift, its estimated value, $\delta\Delta e_{CEM}$, will be null but not its uncertainty contribution.

Equation (2) represents the combined standard uncertainty of the reference value without considering the covariance between e_{CEM1} and e_{CEM2}:

$$u^2(e_{ref}) = 1/\left[\left[\frac{1}{u(e_{CEM1})}\right]^2 + \left[\frac{1}{u(e_{CEM2})}\right]^2 + \left[\frac{1}{u(e_{CENAM})}\right]^2\right] + \left[\frac{\Delta e_{CEM}}{\sqrt{3}}\right]^2 \qquad (2)$$

Equation (3) represents the combined standard uncertainty of the reference value considering a whole correlation between e_{CEM1} and e_{CEM2}:

$$u^2(e_{ref}) = \left[\frac{z}{u(e_{CEM1})} + \frac{z}{u(e_{CEM2})}\right]^2 + \left[\frac{z}{u(e_{CENAM})}\right]^2 + \left[\frac{\Delta e_{CEM}}{\sqrt{3}}\right]^2 \qquad (3)$$

Where $$z = 1/\left[\frac{1}{u^2 e_{CEM1}} + \frac{1}{u^2 e_{CEM2}} + \frac{1}{u^2 e_{CENAM}}\right] \qquad (4)$$

In all the cases it has been considered the maximum uncertainty value of $\delta\Delta e_{CEM}$.

2.2. **Average** *between the values of CEM*

An analysis of the results was made using other estimates in order to evaluate the influence of the estimate chosen for the calculation of the RV.

The arithmetic average between the CEM values was calculated considering the correlation of both CEM values and not taken into account the correlation, as it appears in equation (5).

$$e_{ref} = \frac{e_{CEM1} + e_{CEM2}}{2} + \delta\Delta e_{CEM} \qquad (5)$$

Equation (6) represents the combined standard uncertainty of the reference value without considering the covariance between e_{CEM1} and e_{CEM2}:

$$u^2(e_{ref}) = \left[\frac{u(e_{CEM1})}{2}\right]^2 + \left[\frac{u(e_{CEM2})}{2}\right]^2 + \left[\frac{\Delta e_{CEM}}{\sqrt{3}}\right]^2 \qquad (6)$$

Equation (7) represents the combined standard uncertainty of the reference value where a total correlation between e_{CEM1} and e_{CEM2} is considered:

$$u^2(e_{ref}) = \left[\frac{u(e_{CEM1})}{2} + \frac{u(e_{CEM2})}{2}\right]^2 + \left[\frac{\Delta e_{CEM}}{\sqrt{3}}\right]^2 \qquad (7)$$

2.3. **Average** *between the values of CEM and CENAM*

Due to the consistency between the results of CEM and CENAM, the arithmetic average between them was calculated considering the correlation of two CEM values and without correlation.

The reference value calculated is given in the equation (8):

$$e_{ref} = \frac{e_{CEM1} + e_{CEM2} + e_{CENAM}}{3} + \delta\Delta e_{CEM} \qquad (8)$$

Equation (9) represents the combined standard uncertainty of the reference value without considering the covariance between e_{CEM1} and e_{CEM2}:

$$u^2(e_{ref}) = \left[\frac{u(e_{CEM1})}{3}\right]^2 + \left[\frac{u(e_{CEM2})}{3}\right]^2 + \left[\frac{u(e_{CENAM})}{3}\right]^2 + \left[\frac{\Delta e_{CEM}}{\sqrt{3}}\right]^2 \qquad (9)$$

Equation (10) represents the combined standard uncertainty of the reference value considering a whole correlation between e_{CEM1} and e_{CEM2}:

$$u^2(e_{ref}) = \left[\frac{u(e_{CEM1})}{3} + \frac{u(e_{CEM2})}{3}\right]^2 + \left[\frac{u(e_{CENAM})}{3}\right]^2 + \left[\frac{\Delta e_{CEM}}{\sqrt{3}}\right]^2 \qquad (10)$$

2.4. *Monte Carlo Simulation*

The approach of the real design to the simulated design and the consistency of the presented results were evaluated by the Monte Carlo simulation.

Monte Carlo simulation technique allows all the possible results of a situation to be seen. The method consists of replacing the uncertain values by probability distributions. These functions simply represent a set of possible values, instead of limiting itself to a single value. By means of a simulation it is possible to calculate the model hundreds or thousands of times.

In our case a Monte Carlo simulation of more than 10000 iterations was made. In each simulation, samples were taken from random values of the introduced functions. They were included in the model and the obtained results were registered. The result is a view of the range of possible results, including the probability that they take place. The advantage of the Monte Carlo simulation is that the model not only represents a single result, but provides thousands of possible results.

2.5. *Least squares method*

The well known least squares method was applied in order to estimate the RV and to test consistency between measurements performed by the participating countries [2].

The weighted least squares analysis is able to take into account known covariances between measurement results. It is based on the design of a matrix mathematical model by means of the equation:

$$y = \boldsymbol{X}\beta \tag{11}$$

Where $\boldsymbol{X}$ is the well-known matrix design of the intercomparison. The elements of the matrix are known a priori, in principle with zero uncertainty. y are the input parameters (the measurement results provided by the participants), and β are the unknown factors and are estimated from the n measurement results and the associated covariance matrix Φ.

$$\Phi = V(y) \tag{12}$$

The solution of the system is obtained with the expression:

$$\hat{\beta} = (\mathbf{X}^{\mathrm{T}} \cdot \Phi^{-1} \cdot \mathbf{X})^{-1} \cdot \mathbf{X}^{\mathrm{T}} \cdot \Phi^{-1} \cdot y \tag{13}$$

Being $\hat{\beta}$ the estimated value of β.

The variance-covariance matrix (14) provides the values of the associated uncertainty to each input variables (diagonal elements) and the covariances between the variables (not diagonal elements).

$$\mathrm{V}(\hat{\beta}) = (\mathbf{X}^{\mathrm{T}} \cdot \Phi^{-1} \cdot \mathbf{X})^{-1} \tag{14}$$

Using the estimates $\hat{\beta}$, the predicted values $\hat{y}$ corresponding to the measured values y are calculated in accordance to the model given in (11):

$$\hat{y} = \mathbf{X}\hat{\beta} \tag{15}$$

The associated covariance matrix is:

$$V(\hat{y}) = \mathbf{X} \cdot V(\hat{\beta}) \cdot \mathbf{X}^{\mathrm{T}} \tag{16}$$

The predicted values $\hat{y}$ are defined as the reference values of the intercomparison measurements.

3. RESULTS

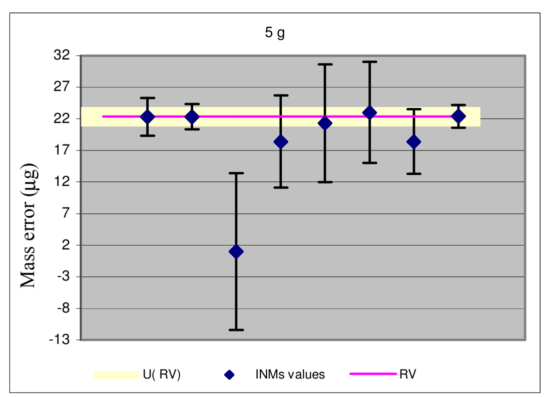
The reference value was calculated as a weighted mean of the values measured by CEM at the beginning and at the end of the comparison and the value measured by CENAM. The inverse of the squares of the declared uncertainty was considered as the ponder factor and the correlation between CEM values was taken into account.

Although in some cases there were anomalous values of some laboratories with respect to the others. They have not been eliminated because the reference value was not calculated by a general weighting.

We will show as an example the results obtained for the weight of 5 g.

Graph 1 shows the results measured by each participant with its uncertainty contribution and the RV considered for the weight of 5 g.

Graph 1: Results weight of 5g.



The expanded uncertainty of the RV is given for a confidence level of 95,45 % according to the equation (17).

$$U(e_{ref}) = 2 \cdot u(e_{ref}) \tag{17}$$

The results obtained with each statistic are represented in table 1:

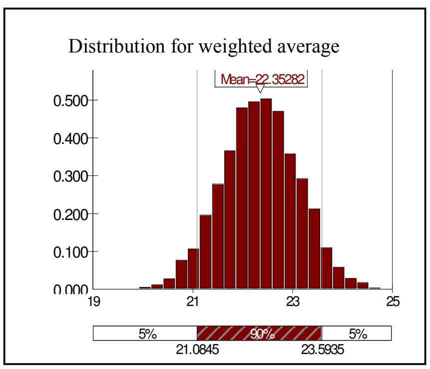
Table 1: RV obtained with each statistic.

	100 mg Error µg	100 mg U(k=1) µg	5 g Error µg	5 g U(k=1) µg	20 g Error mg	20 g U(k=1) mg	100 g Error mg	100 g U(k=1) mg	1 kg Error mg	1 kg U(k=1) mg
A	20,50	0,68	22,34	0,76	0,0553	0,0020	0,2041	0,0031	1,565	0,014
B	20,50	0,21	22,34	0,65	0,0553	0,0011	0,2041	0,0025	1,565	0,009
C	20,62	0,61	22,33	1,16	0,0551	0,0024	0,2043	0,0049	1,580	0,032
D	20,34	0,92	22,33	1,84	0,0552	0,0044	0,2029	0,0077	1,576	0,056
E	21,18	0,66	22,35	1,24	0,0550	0,0028	0,2073	0,0052	1,590	0,038
F	21,18	0,89	22,35	1,70	0,0550	0,0043	0,2073	0,0071	1,590	0,055
G	20,50	0,68	22,35	0,77	0,0553	0,0020	0,2041	0,0031	1,565	0,015
H	20,56	0,81	22,34	0,85	0,0552	0,0374	0,2042	0,0544	1,571	0,082
I	19,72	0,89	22,34	0,94	0,0554	0,0528	0,1986	0,0904	1,56	0,10

A: CEM-CENAM weighted average with correlation between the CEM values
B: CEM-CENAM weighted average without correlation between the CEM values
C: CEM-CENAM average without correlation between the CEM values
D: CEM-CENAM average with correlation between the CEM values
E: CEM average without correlation between the CEM values
F: CEM average with correlation between the CEM values
G: CEM-CENAM Monte Carlo Simulation
H: CEM-CENAM least square without correlation between the CEM values
I: CEM-CENAM least square with correlation between the CEM values

Graph 2 shows the Monte Carlo simulation for the weight of 5 g.

Graph 2: Distribution for weighted average of 5g.



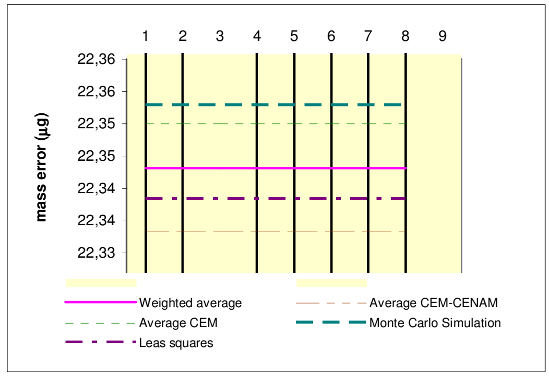
The summary statistic of the Monte Carlo Simulation obtained for the weight of 5 g is in table 2.

Table 2: Summary statistic for the weight of 5g.

Statistic	Value
Minimum	19,444
Maximum	24,779
Mean	22,353
Std Dev	0,767
Variance	0,588687344
Skewness	-0,034759338
Kurtosis	2,907764958
Median	22,355
Mode	21,496
Left X	21,085

Graph 3 shows the RV obtained with each statistic for the weight of 5 g.

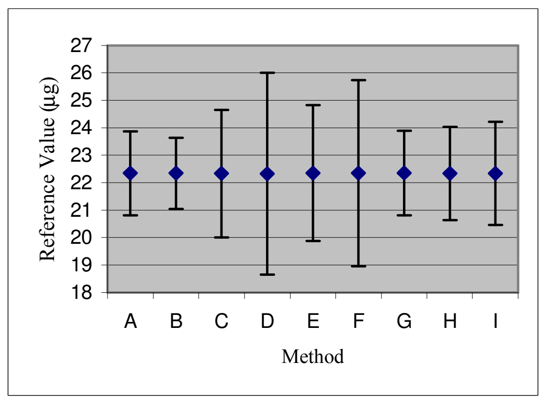
Graph 3: Different RV for the weight of 5g.



The difference of results of the different statistics is 0,028 µg, and it is negligible in comparison with the uncertainty assigned to the considered RV.

Graph 4 shows the RV with its uncertainty (k=1) obtained with each statistic for the weight of 5 g.

Graph 4: Different RV with its uncertainty for the weight of 5g.



In order to quantify the quality of the results given by the participating laboratories, it was necessary to calculate the degree of equivalence in relation to the RV and its associated uncertainty.

The degrees of equivalence of each laboratory i are:

$$d_i = e_i - e_{\text{ref}} \qquad (18)$$

And its associated uncertainty is $U(d_i) = 2 \cdot u(d_i)$ for a 95 % coverage factor under assumption of normality, where:

$$u^2(d_i) = u^2(e_i) - u^2(e_{\text{ref}}) \qquad (19)$$

In the equation (19) the covariance between e_i and e_{ref} was considered. When the RV has been evaluated from several results, there will be a correlation between RV and each input value that must be considered when the uncertainty is calculated.

The correlation effects between e_i and e_{ref} may be avoided by the use of an 'exclusive' definition of the RV that takes into account all results except the one that is being considered [3].

The results of the inclusive (equation 20) and exclusive (equation 21) approaches are closely related and the results given by the exclusive approach are used in order to calculate without using covariances.

$$e_{ref} = \frac{\sum_{i=1}^{N} \left(e_i / u^2(e_i)\right)}{\sum_{i=1}^{N} \left(1/u^2(e_i)\right)} \qquad (20)$$

$$e_{ref} = \frac{\sum_{i=1;i\neq j}^{N} \left(e_i / u^2(e_i)\right)}{\sum_{i=1}^{N} \left(1/u^2(e_i)\right)} = \frac{\sum_{i=1}^{N} \left(e_i / u^2(e_i)\right) - \left(e_j / u^2(e_j)\right)}{\sum_{i=1}^{N} \left(1/u^2(e_i)\right) - \left(1/u^2(e_j)\right)} \qquad (21)$$

The 'exclusive' form is useful to check whether there is any significant influence of the correlation between the RV and the individual results.

There was no quantitative difference between the RV values calculated by means of inclusive and exclusive statistics. The exclusive effect is only significant when the weight $1/u^2(e_j)$ assigned to the laboratory under consideration is significant compared with the sum $\sum_{i=1}^{N} \left(1/u^2(e_i)\right)$.

In order to compare the deviations with their associated standard uncertainties, the normalized deviations D_i of each laboratory were calculated using equation (22), and represented in table 3:

$$D_i = \frac{e_i - e_{\mathrm{ref}}}{\sqrt{u^2(e_i) - u^2(e_{\mathrm{ref}})}} \qquad (22)$$

Table 3: normalized deviations of each laboratory

INM	100 mg	5 g	20 g	100 g	1 kg
1	0,15	0,03	0,86	0,53	0,42
2	1,34	0,07	0,06	2,05	0,21
3	1,81	3,42	2,06	2,32	1,53
4	1,41	1,06	0,18	4,91	9,60
5	4,14	0,22	4,20	0,41	0,05
6	1,81	0,16	0,40	0,23	0,23
7	4,66	1,48	1,25	0,12	0,98
8	1,61	0,10	0,55	1,82	1,08

If the absolute value of the normalized D_i deviation is larger than the coverage factor k=2, that is:

$$|D_i| > 2 \qquad (23)$$

Then the measured value e_i is classified as an outlying at a 5% level of significance.

Consistency of the results can be evaluated with the normalized D_{ij} deviations, where the degrees of equivalence between the laboratories are clearly represented.

$$D_{ij} = \frac{e_i - e_{\mathrm{j}}}{\sqrt{u^2(e_i) + u^2(e_{\mathrm{j}})}} \qquad (24)$$

Table 4: Normalized deviations between laboratories for the weight of 5g.

INM	1	2	3	4	5	6	7	8
1	0,00	0,01	3,34	0,99	0,21	0,16	1,32	0,05
2	0,01	0,00	3,39	1,03	0,21	0,17	1,42	0,07
3	3,34	3,39	0,00	2,42	2,62	2,98	2,60	3,41
4	0,99	1,03	2,42	0,00	0,49	0,85	0,00	1,06
5	0,21	0,21	2,62	0,49	0,00	0,28	0,55	0,23
6	0,16	0,17	2,98	0,85	0,28	0,00	0,97	0,15
7	1,32	1,42	2,60	0,00	0,55	0,97	0,00	1,48
8	0,05	0,07	3,41	1,06	0,23	0,15	1,48	0,00

If the absolute value of the normalized D_{ij} deviation is larger than the coverage factor k=2, that is:

$$|D_{ij}| > 2 \qquad (25)$$

Then the measured pair (e_i, e_j) is classified as inconsistent at a 5% level of significance.

4. DISCUSSION

In view of the calculations presented, the advantages and disadvantages of each method can be quantitatively evaluated.

The disadvantage of the mean is that the standard uncertainties and covariances of measurements are neglected and the changes in the travelling standards cannot be taken into account.

Although, in the weighted mean, the measurement uncertainties are taken into account but the disadvantage is that the results are assumed to be independent and mutually consistent and only normal distributions are assigned to results. Furthermore, the travelling standards have been assumed to be stable.

The advantage of the least squares estimate is that covariances can be taken into account in the calculation of RV and also travelling standard drifts can be modeled.

The Monte Carlo Simulation advantage is that the normal distribution assumption is not required.

5. CONCLUSION

At the end great consistency between the values evaluated by the different methods and the considered RV has been obtained, observing the clear difference in the graphical representations.

Table 5 shows the difference between the calculated RV using each statistic and the weighted average. Generally this difference is lower than the uncertainty associated to each RV (it was only bigger in the data marked in italic).

Table 5. Differences with respect to the weighted average

	100 mg µg	5 g µg	20 g mg	100 g mg	1 kg mg
C	0,12	0,010	0,00015	0,0002	*0,016*
D	0,16	0,018	0,00007	0,001	0,011
E	0,68	0,007	0,00030	0,003	0,025
F	0,68	0,007	0,00030	0,003	0,025
G	0,002	0,010	0,00001	0,00007	0,0002
H	0,07	0,005	0,00007	0,00004	0,0066
I	0,78	0,006	0,00015	0,0055	0,0032

It should be kept in mind, however, that each method has contributes advantages and disadvantages in comparison to the others.

ACKNOWLEDGMENTS

The authors would like to thank the technicians from the different laboratories for their careful work. They would also like to thank to the European Union for the funds enabling the cooperation in technical projects. We are grateful for the PTB and its collaboration in the project.

REFERENCES

[1] M. Cox1, “The evaluation of key comparison data”, Metrologia 39, 589-595, (2002)

[2] L. Nielsen, “Evaluation of measurement intercomparisons by the method of least squares”, DFM-99-R39, (2000)

[3] M J T Milton, M G Cox, “Evaluating degrees of equivalence using ‘exclusive’ statistics”, Metrologia 40, L1-L2, April 2003.