Quim. Nova, Vol. 26, No. 3, 417-419, 2003

Nota Técnica

ANALYSIS OF LABORATORY INTERCOMPARISON DATA. A MATTER OF INDEPENDENCE

Mauro F. Rebelo*
Instituto de Biofísica Carlos Chagas Filho, Centro de Ciências da Saúde, Universidade Federal do Rio de Janeiro, 21949-900 Rio de Janeiro - RJ
José M. Monserrat
Departamento de Ciências Fisiológicas, Universidade do Rio Grande, CP 474, 96201-900 Rio Grande - RS
Wanderley G. Bastos
Departamento de Medicina, Universidade Federal de Rondônia, Br364 Km 9,5, 78900-970 Porto Velho - RO

Recebido em 3/1/01; aceito em 19/12/02

When laboratory intercomparison exercises are conducted, there is no *a priori* dependence of the concentration of a certain compound determined in one laboratory to that determined by another(s). The same applies when comparing different methodologies. A existing data set of total mercury readings in fish muscle samples involved in a Brazilian intercomparison exercise was used to show that correlation analysis is the most effective statistical tool in this kind of experiments. Problems associated with alternative analytical tools such as mean or paired 't'-test comparison and regression analysis are discussed.

Keywords: laboratory intercomparison; mercury; fish.

INTRODUCTION

There is a clear need for intercomparison exercises when studies involving more than one laboratory are conducted[1,2]. When several species and compounds are analyzed, such as mercury[3-5] and cholinesterase activity[6], this provides confidence results as verifies the accuracy of the measurements.

However, it must be recognize that there is no *a priori* dependence of the concentration of a certain compound determined in one laboratory to that determined by others. A mercury concentration datum, for example, is dependent only on the mercury content of a tissue or blood or whatever the sample matrix is. On the other hand, the readings data supplied by one or more laboratories could be related or not, depending on the degree of intercalibration.

Regression analysis, correlation tests and the t-test, are commonly used to evaluate the degree of likelihood between two data sets, as in intercomparison exercises.

REGRESSION ANALYSIS

Whenever there are two data sets, in which one is dependent (Y variable) on the other (X variable), regression analysis is used to estimate the percentage variability of Y explained by its relationship with X. This quantity is expressed by the r^2 value (the closer to 1, the better the estimation). If there are two data sets from two laboratories to be intercalibrated, the regression model is not suitable, since it is impossible to say *a priori* which set of data is the response variable and which independent one. In this way, none of the equations can represent better the relation between the readings (Figure 1).

As an example, an existing data set of total mercury readings in fish muscle samples (Table 1) analyzed by the Laboratory of Radioisotopes from Rio de Janeiro Federal University (LREPF) and the Analytical Chemistry Laboratory from Rondonia University (UNIR), has been used.

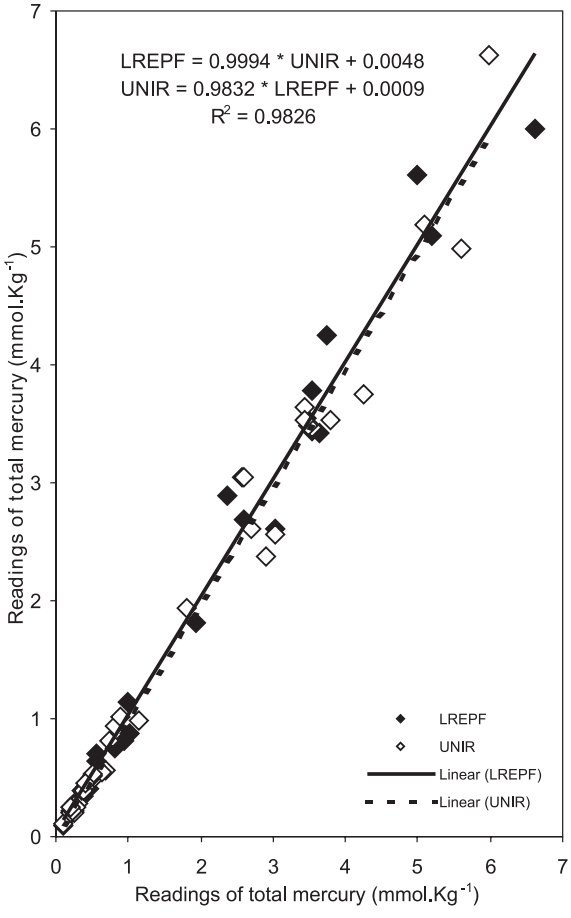

Figure 1. *Estimated regression functions using UNIR or LREPF as response variables. Since is not possible to establishes which is the independent and the response variable, either equation could be chosen to represent the relationship. But there is no right criteria to do so, showing that this model is not suitable*

*e-mail: mrebelo@biof.ufrj.br

Table 1. Readings of total mercury (mmol.Kg^{-1}) on fish muscle samples used for the intercalibration laboratory exercise. UNIR: Analytical Chemistry Laboratory from Rondonia University. LREPF: Laboratory of Radioisotopes from Rio de Janeiro Federal University

Reading	UNIR	LREPF	Reading	UNIR	LREPF
1	3.744	4.252	20	0.204	0.199
2	3.470	3.540	21	0.259	0.249
3	0.279	0.299	22	0.429	0.399
4	0.110	0.100	23	1.939	1.810
5	3.535	3.789	24	0.563	0.698
6	0.224	0.199	25	0.244	0.274
7	0.813	0.748	26	0.219	0.229
8	0.484	0.499	27	0.548	0.648
9	6.620	5.992	28	0.518	0.499
10	2.562	3.041	29	1.012	0.882
11	0.508	0.499	30	0.249	0.214
12	4.990	5.603	31	3.644	3.425
13	3.041	2.602	32	3.480	3.485
14	0.384	0.349	33	0.389	0.379
15	0.937	0.818	34	0.523	0.513
16	5.195	5.090	35	2.602	2.692
17	0.204	0.249	36	0.449	0.399
18	0.254	0.249	37	2.373	2.891
19	0.992	1.147	38	3.530	3.440

MEAN COMPARISONS TEST

The method most commonly used to evaluate the differences in the means of two groups is the t-test. However, it is clear that two means can be statistically similar despite there being no correspondence between the two data sets. Therefore this approach does not apply.

Alternatively, if each laboratory analyzed *n* samples, there will be *n* pairs coming from the same sample. In this case, it is common practice to analyze these data by the means of paired t-test[7,8]. Still, this is not the appropriate tool. It must be pointed out that, if this analysis is conducted over the original data set (Table 1), no significant difference is detected between the readings that came from UNIR and those from the LREPF (p= 0.567). However, this result must be considered as proof of absence of a consistent higher (or lower) reading from one laboratory, with respect to the other. It is not possible to imply that there is a linear relationship between the readings of the two laboratories, which would be the expected result if the two laboratories were intercalibrated.

CORRELATION COEFFICIENT

When the objective is to show a linear relationship between two data sets, correlation coefficients are the correct tool. Note that this kind of relationship is what we would expect if the two laboratories were equally calibrated. The Pearson's correlation coefficient, r, (or a non parametric equivalent) will quantify how similar the laboratories readings are. The parametric correlation coefficient assumes that the two data sets belong to a bivariate normal population and, if this is true, the distribution of each variable must be normal, although the converse does not apply[7].

As can be seen in Figure 2, the distribution for both UNIR and LREPF variables are not normal. Formal statistical test like Shapiro-Wilk's W test[5] rejected the null hypothesis of normality for both variables (p< 0.0001). So, the non-parametric Kendall Tau coefficient was calculated[9] and estimated as 0.931.

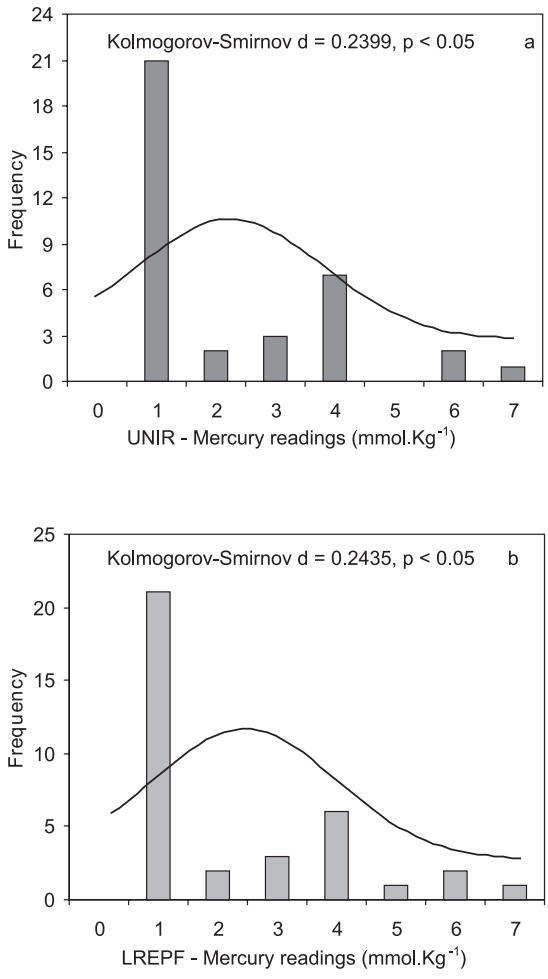

Figure 2. *Distribution of (a) UNIR and (b) LREPF mercury readings. The expected normal distribution is also shown*

The null hypothesis of absence of correlation between the two variables was tested using the normal approximation (Z= 8.22; p< 0.0001) and rejected. As shown in Figure 3, the conclusion is that the mercury readings from the two laboratories are linearly related.

Although our data exhibits linearity, this is not the only criterion need to demonstrate that the laboratories are intercalibrated. Is also necessary to prove that the slope (β) of the line is not significantly different from 1 and that the intercept (α) is not significantly different from 0. Note that if these conditions are fulfilled, the same regression line will be obtained, whatever which variable (UNIR or LREPF) was used as response variable. The hypothesis mentioned above can be tested using a t-test[7,8]. The two hypotheses ($\beta = 1$ and $\alpha = 0$) were not rejected when using the data shown in Table 1, suggesting that the two laboratories are, in fact, intercalibrated.

CONCLUDING REMARKS

Using a real data set, it has been shown that mistaken conclusions can be drawn if intercomparison data are analyzed by means of paired test comparisons, and also when regression analysis is used to estimate functions. Correlation analysis and tests of hypotheses about slope and intercept are more desirable statistical tools in the analysis of laboratory intercomparison data, since it is possible to test if the laboratories involved are intercarlibrated (ie: if a correlation

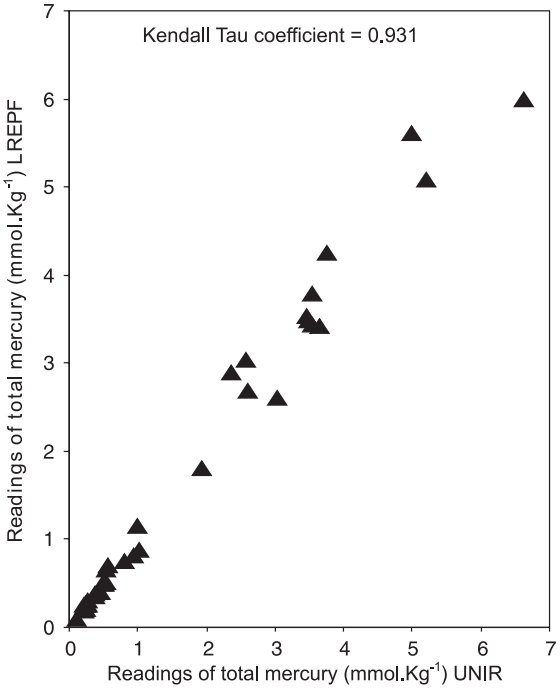

Figure 3. *Relationships between UNIR and LREPF mercury readings, using the original data set showed in Table 1. The Kendall Tau coefficient was estimated as 0.931*

coefficient is statistically different from zero). This would imply a linear relationship between the readings of different samples.

ACKNOWLEDGEMENT

The authors would like thanking J. Valentin, O. Malm and B. Howard for the careful reading of this manuscript.

REFERENCES

1. Cofino, W. P.; Baarsma, J. P.; Peerboom, R. A. L.; *Eur. Water Pollut. Control* **1995**, *5*, 8.
2. Arri, E.; Cabiati, F.; D'Emilio, S.; Gonella, L.; *Meas.* **1995**, *16*, 51.
3. Bloom, N. S.; Horvat, M.; Watras, C.J.; *Water, Air, Soil Pollut.* **1995**, *80*, 1314.
4. Kehrig, H. A.; Malm, O.; Akagi, H.; *Water, Air, Soil Pollut.* **1997**, *97*, 29.
5. Nott, B. R.; *Water, Air, Soil Pollut.* **1995**, *80*, 1314.
6. Marden, B. T.; Fairborther, A.; Bennet, J. K.; *Environ. Toxicol. Chem.* **1994**, *13*, 1768.
7. Zar, J, H.; *Biostatistical analysis*, Prentice Hall: New Jersey, 1984.
8. Snedecor, G. W.; Cochran, W. G.; *Statistical methods*, Iowa State University Press: Iowa. 1980.
9. Conover, W. J.; *Practical Nonparametric Statistics*, John Wiley & Sons: New York, 1980.