# Statistical Analysis of Experimental Data from Different Analytical Techniques

**R.A. Tavares, L.K. Cordani**
IME-USP, Departamento de Estatistica, São Paulo, Brazil
**J.E. de Souza Sarkis, M. Hiromitu Kakazu**
IPEN-CNEN/SP, Grupo de Caracterização Isotópica, São Paulo, Brazil

## 1. Abstract

The determination of the true chemical composition of a given material is a very difficult task. Usually, the most probable value obtained by means of an already established analytical technique is accepted as the true value. However, in many routine situations in laboratories, or during material certification procedures, it is necessary to compare results obtained by means of different analytical techniques, an operation that may lead to wrong conclusions. This work will exhibit the results of a material certification exercise performed at the Instituto de Pesquisas Energéticas e Nucleares (IPEN) in cooperation with the Instituto de Matemática e Estatística da Universidade de São Paulo (IME - USP). Samples of $U_3O_8$ (Uranium Oxide) were analysed by different analytical techniques. A statistical analysis using ANOVA, as well as some robust and weighted methods, will be discussed.

## 2. Introduction

A chemical analysis of a given material is one of the oldest activities in science, however, in spite of the development of sofisticated analytical techniques, the determination of the true value of a chemical component is still an open question.

In a routine laboratory operation, the analyst needs to control many factors related with instruments and procedures that can affect in a significant way the quality of the results obtained, since every analytical technique has limitations. Many statistical methodologies are developed in order to assure the quality of the results, to provide an evaluation of the behavior of an analytical procedure and to establish a comparison pattern with the results obtained in other laboratories. The majority of these methods operates relative to standards, and makes use of highly pure and certified chemical materials.

In the choice of a material to be certified, several factors must be considered, among which: the future applications of the material, the methodologies used to obtain it, and the statistical methods used to investigate the numerical values. Taking all this into account, the program's protocols are prepared, the number and the quality of the essays to characterize the chemical and physical properties of the chosen material are defined, as well as the procedures for the statistical analysis of the data.

The aim of this work is to discuss different statistical approaches to analyse data from material certification programs.

## 3. Previous research exercises

With the purpose of developing a suitable statistical model, 15 analysis were carried out at the IPEN (Instituto de Pesquisas Energéticas e Nucleares) by 15 different analysts, allowing in this way to simulate two intercomparison exercises between 15 different laboratories employing the same analytical technique. The material used was $U_3O_8$ (Uranium Oxide) of nuclear grade produced at the IPEN.

Two statistical methodologies were used to analyse the data. In the first exercise, classical statistical techniques like ANOVA (Analysis of Variance) /1/ were used preceded by outlier detection tests to ensure that the data set fulfills the necessary conditions, such as variance equality and data with Normal distribution.

Data elimination may introduce tendency in the results, because the discarded values can be due to material variability, and not from measurement errors. In this sense, robust methods were used (Dod Estimators-Distribution of differences) keeping all values in the variability estimation of a data set derived from a Normal distribution. These estimators can be very useful because, to the contrary of classical methods, they are rather unsensitive in presence of outliers. Details about these methods can be found in Beyrich et al. /2/.

The results of the first exercise, found in Cordani and Yamamoto /3/, demonstrated that within the uncertainty of the same analytical technique, graphical methods of data analysis (Dotplot, Boxplot and Youden), as well as classical techniques with result eliminations can be used with the same efficiency when compared with robust methods.

In the second exercise (Cordani and Fukunaga /4/), time was included as perturbation factor, and an increase in the variability of the results was observed. In this case, the use of classical methods of analysis may distort the final results, and therefore the use of robust methods was more adequate.

## 4. The balanced method

In order to further develop the same research line, we investigated the case of an intercomparison of data from different laboratories, when obtained by different analytical techniques.

With this purpose the Balanced Method was employed, a robust statistical technique described in Analytical Methods Committee /5/. This technique consists in proposing estimators for the mean and for the standard deviation that are not sensitive to the presence of outliers.

Assuming the participation of L laboratories, each one with r measurements, a particular measurement from a given laboratory may be defined as in equation (1):

$$x_{ki} = \mu + \alpha_k + \varepsilon_{ki}, \qquad (1)$$

where:

$x_{ki}$: i-th measurement of the k-th laboratory
$\alpha_k$: k-th laboratory error.
$\varepsilon_{ki}$: associated error.
$VAR(\alpha_k)=\sigma_L^2 \quad k=1,...,L$
$VAR(\varepsilon_{ki})=\sigma_e^2 \quad k=1,...,r.$

In the same way, the mean from a given laboratory may be written as in equation (2):

$$\mu_k = \mu + \alpha_k \qquad (2)$$

At first, the internal variability $\sigma_e^2$ of each laboratory shall be estimated, as well as its respective mean value $\mu_k$. Secondly, the estimation of $\sigma_L^2$ shall be determined, representing the inter-laboratory variability.

Because all laboratories are equally treated by the method, the description below corresponds to an iterative calculation procedure applied to each laboratory.

It is here proposed that the estimation shall be done through a weighted mean of the observations $x_{ki}$ of each labora-

tory, where to each measurement will be given a weight $w_{ki}$, obtained according equation (3):

$$w_{ki} = \begin{cases} 1 & \text{if } |x_{ki} - \mu_k| \le c\sigma_e \\ \dfrac{c\sigma_e}{|x_{ki} - \mu_k|} & \text{otherwise} \end{cases} \quad (3)$$

where $\sigma_e$ is a robust standard deviation and c is a suitable constant previously determined by the user, with values between 1 and 2. In this work the more recommended value c = 1.5 will be used. Hence, according to equation (3), if a particular observation is dislocated from its laboratory mean more than $c\sigma_e$, its value will be "diminished" receiving a weight smaller than 1, that decreased with the increase of the distance between this observation and the mean intralaboratory value.

The estimation of $\sigma_e$ shall be made simultaneously with the estimation of $\mu_k$ at each step of this iterative process. Thus, a weighted observation may be defined by equation (4):

$$x_{ki}' = w_{ki} \cdot x_{ki} \quad (4)$$

The procedure to estimate $\sigma_e$ is similar to the one previously described, where in each step:

$$\sigma_e = \sqrt{\frac{\Sigma(x_{ki} - \mu_k)^2}{n\beta}} \quad (5)$$

where n is the number of measurements of each laboratory and ß is a constant chosen in order to obtain a consistent answer with normally distributed data. The values of ß are tabled in Analytical Methods Committee /5/.

The initial value for $\mu_k$ can be taken as the mean or median of the original data set, whereas the suggestion given in Analytical Methods Committee /5/ to the initial value of $\sigma_e$ is:

$$\text{median} \frac{\left(\left|x_{ki} - \mu_k\right|\right)}{0.6745} \quad (6)$$

The end of the process only occurs when the difference between two iterations would be smaller than a fixed value. In this exercise, $10^{-2}$ will be used as a stopping condition.

Next step corresponds to analyse the estimated means according to what was described to estimate the certified value μ and the material's variability $\sigma_1^2$ where:

$$\sigma_1^2 = \sigma_L^2 + \frac{\sigma_e^2}{r} \quad (7)$$

Once $\sigma_e^2$ is already estimated, only $\sigma_L^2$ remains to be estimated, according to equation (7). This estimation is done in analogy to the process described earlier, considering the L means estimated from all laboratories like observations from the same laboratory.

To compare this method with the others used in the previous exercises, purity percentages of $U_3O_8$ were measured in 6 IPEN's laboratories employing the following analytical techniques: Voltametry, Isotope Dilution Mass Spectrometry, Davies & Gray, High Performance Liquid Chromatography and Inductively Coupled Plasma Atomic Emission Spectrometry. In each laboratory 8 measurements were carried out, resulting in 48 observations presented in Table 1.

The values higher than 100% are a consequence of contamination in the measurements. Therefore, despite such results are not theorically expected, in practice they can occur. The data were analysed by ANOVA, Dod and by the balanced method.

## 5. Results

### 5.1. Analysis by ANOVA

At first a graphical analysis is presented of the data, through a Dotplot, by laboratory (Figure 1).

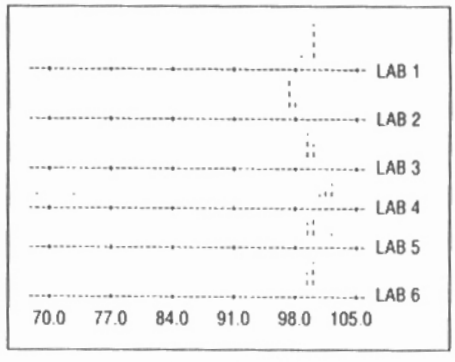


**Figure 1:** Dotplot of purity percentages of $U_3O_8$

It can be noted that, generally, data's dispersion is low. However, Laboratory 4 exhibits two observations located very distant from the others that may be considered outliers:

In Table 2, sample means and standard deviations by laboratory are presented.

From Table 2, the variability of Laboratory 4 resulted very large, because of the presence of extreme values affecting mean and standard deviation.

**Table 2:** Sample means and standard deviations

| Laboratory | Mean | Std. Deviation |
|---|---|---|
| 1 | 99.70% | 0.45% |
| 2 | 97.58% | 0.15% |
| 3 | 99.66% | 0.10% |
| 4 | 93.88% | **14.28%** |
| 5 | 100.19% | 0.92% |
| 6 | 99.74% | 0.18% |

Dixon's outliers detection test /6/ and Cochran's variance equality test /7/ were used in this classical approach with significance level of 5%. After the application of these tests, the two aberrant observations from Laboratory 4, as well as another one from Laboratory 1 and another from Laboratory 5 were eliminated. The two last ones present a relatively small displacement in the graph; moreover, if they originated for instance in Laboratory 4, they probably would not have been eliminated. The discarded observations are very visible in Table 1. All laboratories were kept by Cochran's test.

The new values for mean and standard deviation, after discarding the mentioned outliers, are presented in Table 3.

**Table 3:** Means and standard deviations after the elimination of the outliers

| Laboratory | Mean | Std. Deviation |
|---|---|---|
| 1 | 99.85% | 0.10% |
| 2 | 97.58% | 0.15% |
| 3 | 99.66% | 0.10% |
| 4 | 101.57% | **0.56%** |
| 5 | 99.89% | 0.38% |
| 6 | 99.74% | 0.18% |

A quick comparison of Tables 2 and 3 demonstrates the improvement, especially in the variability of Laboratory 4, resulting from the elimination of the extreme values.

**Table 1:** Purity percentages of $U_3O_8$

| Lab 1 | Lab 2 | Lab 3 | Lab 4 | Lab 5 | Lab 6 |
|---|---|---|---|---|---|
| **98.60%** | 97.52% | 99.65% | **68.65%** | 100.20% | 99.87% |
| 99.80% | 97.73% | 99.76% | **73.02%** | 99.60% | 99.49% |
| 99.80% | 97.83% | 99.65% | 101.88% | 100.20% | 99.84% |
| 100.00% | 97.36% | 99.76% | 101.66% | 99.50% | 99.89% |
| 99.80% | 97.62% | 99.53% | 100.50% | 100.40% | 99.62% |
| 99.80% | 97.46% | 99.76% | 101.46% | 99.50% | 99.79% |
| 99.80% | 97.62% | 99.53% | 101.97% | **102.30%** | 99.94% |
| 100.00% | 97.46% | 99.65% | 101.93% | 99.80% | 99.49% |

Finally, after the application of ANOVA for the 44 remaining values, the following estimatives were obtained:

- Average purity percentage: 99.63%
- StandardDeviation: 0.51%

### 5.2. Analysis by Dod

Using the original data set, the Dod robust technique was applied. This method do not eliminate observations and estimates only the variability of the measurements.

With the help of a computational program developed in Pascal language, the following variability estimative was obtained through the DodM estimator:

- Standard Deviation: 0.23%

### 5.3. Analysis by balanced method

In order to apply this technique, a computational program developed in Fortran language will be used (Analytical Methods Committee /5/). Table 4 presents means and standard deviations of the 6 laboratories obtained by the iterative process of balanced method. The estimatives of the mean and standard deviation for the whole data set are also presented.

- Average purity percentage: 99.69%
- Standard deviation: 0.18%

The results of Table 4 are very similar to the ones presented in Table 3, indicating that even such methods that do not eliminate observations can be good alternatives of analysis.

**Table 4:** Means and standard deviations of the final iteration of the balanced method

| Laboratory | Mean | Std. Deviation |
|---|---|---|
| 1 | 99.77% | 026% |
| 2 | 97.58% | 0.15% |
| 3 | 99.61% | 0.10% |
| 4 | 101.41% | 0.53% |
| 5 | 99.97% | 0.43% |
| 6 | 99.74% | 0.18% |

## 6. Discussion

For the purity average value, results of ANOVA and balanced method were very close, suggesting that robust methods allow similar conclusions to the classical, without eliminating data.

Purity's variability presented larger differences, specially when robust methods are used. It is important to remark that even when all observations are kept, the variability estimated by these methods was smaller than that one obtained by ANOVA. The reason is that robust estimators reduce the influence of data very far from the mean, introducing a significant effect in the standard deviation's estimative.

Another important point is that the balanced method assigned, in its final iteration, weights lower than 1 to 5 observations. Four among them are the same excluded by the outliers detection test, indicating that there is some coherence in this approach to detect suspicious observations.

It must be also noted that the variability estimated by ANOVA is acceptable, in view of the errors of the laboratorial techniques utilized. Moreover, the results of robust methods' variability were much smaller than the classical, due to the fact that the values measured by all laboratories, excepting some outliers, are very similar.

## 7. Conclusions

From all exercises carried out in the past 3 years, it can be concluded that methods that do not eliminate data are good alternative of analysis and that outliers must be considered as real circunstancial values, included within the material and technical variabilities.

It was also noticed that the balanced method reduces the influence of outliers in variability's estimation and that elimination of data may introduce tendency in the results. Moreover, discarding observations is nothing more than loosing information, diminishing the strength of the conclusions.

It became clear that the notion of outlier depends on a context, because a value can be considered outlier in a laboratory, but it may not be in another.

Therefore, alternative methods to the classical ones shall be more and more employed, in the future, with the aim of determining their advantages and limitations. We consider that what has been done so far is only a beginning for the exploration of alternative statistical methods. Moreover, a better quality control of the analytical performance of laboratories is needed, in order to improve every step of the materials certification processes.

## 8. References

/1/ NETER, J., WASSERMAN, W., KUTNER, M.H. (1990). Applied linear statistical models: regression, analysis of variance and experimental design. 3 ed. Homewood: Richard D. Irwin. 1181p.

/2/ BEYRICH, W., GOLLY, W. (1989). Evaluation of IDA-80 data by the Dod method. Relatório técnico KfK 4157, EUR 10533EN-PWA 64/89, Karlsruhe, Germany. 20p.

/3/ CORDANI, L.K., YAMAMOTO, W.H. (1992). Programa de certificação de materiais estratégicos. São Paulo, IME-USP, 42p. RAE-CEA-9215.

/4/ CORDANI, L. K., FUKUNAGA, E.T. (1993). Análise estatística de dados provenientes de programa de certificação química de compostos de uranio. São Paulo, IME-USP, 42p. RAE-CEA-9316.

/5/ Analytical Methods Committee (1989). Robust statistics - How not to reject outliers. Part 1 - Basic concepts. Part 2 - Intercomparative trials. Analyst, 114, 1693-1702.

/6/ DIXON, W.J. (1953). Processing data for outliers. Biometrics, 9, 74-89.

/7/ WINER, B.J. (1971). Statistical principles in experimental design. 2ed. New York: McGraw-Hill. 907p.