Possible advantages of a robust evaluation of comparisons*

by Jörg W. Müller

Bureau International des Poids et Mesures, F-92312 Sèvres Cedex

Abstract

Mean values, traditionally used as a location parameter in the analysis of intercomparisons, are known to lack stability against the effect of "outliers". It is therefore proposed to replace (or complement) them by the use of medians, which have better statistical "robustness". An estimate for the corresponding uncertainty is derived and the procedure is illustrated by a numerical example. The simplicity of the suggested robust approach should favour its practical use in a number of metrological applications.

1. Some generalities

The use of mean values as a location parameter has become for long a deeply entrenched habit which scarcely requires justification. Also, by applying the principle of "least squares", one can even prove that this procedure, in some specific sense, leads to the "best" choice that can be made.

For most users of elementary statistics it comes as a surprise, therefore, to learn that the very foundation of using mean values may come under question. However, a rapidly developing branch of modern statistics, that which studies "robust" methods of estimation, has concluded (for quite some time already) that interpreting measurement results in terms of mean values is actually not a "safe" procedure because such values have poor stability against the effect of "outliers" (for a general review, see e.g. [1]).

Outliers have been known for long, of course, but they were usually considered a nuisance in statistics - mainly because nobody really knew what to do with them. Yet, their incidental occurrence is a well-established fact. They pose a problem which seems to have no satisfactory solution. In principle, there are three ways to deal with outliers:

- leave them stay as they are,
- correct them, or
- delete them.

Unfortunately, all these procedures have unwanted features. Thus, if outliers are retained, they falsify both the mean value and its uncertainty, possibly to the point at which they become unacceptable. Deletion or correction, in practice often performed by applying some more or less obvious weighting procedure or rejection rule, would require a clear justification. Hence, whatever procedure is followed, it is easily criticized.

---

* This text is based on a conference held at the BIPM on 17 February 1995, entitled "Les valeurs moyennes - sont-elles fiables?", and a similar talk given at the IAEA in Vienna on 5 April 1995.

To this uncomfortable situation comes an unexpectly simple solution. It is offered by the expanding field of "robust statistics", and comes as a most welcome rescue.

Let us concentrate on a single way - certainly the simplest and no doubt one of the more efficient ones - to achieve a protection against the unwanted effects of outliers. It is simply to replace the mean value by the corresponding median (or central value). For a continuous variate x, the median $\tilde{m}$ is defined, using the (cumulative) distribution function F(x), by the condition

$$F(\tilde{m}) = \frac{1}{2} . \tag{1}$$

This means that one half of the observations are below and the other half above the median.

For a sample of n ordered variables x_1, x_2, ..., x_n, the sample median, denoted as $\tilde{m}$ = med $\{x_i\}$, is given by (with integer k)

$$\tilde{m} = \begin{cases} x_{k+1} , & \text{for } n = 2k + 1 , \\[2ex] \frac{1}{2} (x_k + x_{k+1}) , & \text{for } n = 2k. \end{cases} \tag{2}$$

As is well known, the median can also be obtained as solution from the condition that

$$\sum_{i=1}^{n} |x_i - \tilde{m}| = \text{minimum} . \tag{3}$$

This equation then takes the place of the traditional principle of the least mean squares (see Appendix).

2. Uncertainty of the median

While the replacement of the mean value by the corresponding median $\tilde{m}$ is a simple and straightforward procedure, the estimation of the uncertainty $s(\tilde{m})$ to be associated with $\tilde{m}$ requires some more thought.

In the spirit of our robust approach we base this estimate also on a quantity which involves medians. An obvious choice is to use the "median of the absolute deviations", often abbreviated by MAD (a rather unfortunate choice), and defined by

$$\text{MAD} = \text{med} \{|x_i - \tilde{m}|\} , \quad \text{for } i = 1, 2, ..., n . \tag{4}$$

The required estimate for the uncertainty of $\tilde{m}$ is then taken as

$$s(\tilde{m}) = C \text{ MAD} , \tag{5}$$

with a proportionality factor C which has to be evaluated.

The constant C is determined by requiring that, in the limit of large samples, the estimate coincides with what we would obtain for a sample taken from a normal

population. This is an arbitrary but reasonable normalization.

The goal is achieved in two steps. First, we establish a relation between MAD and the parameter σ of a normal distribution, and then we use the known ratio of the variances for the median and the mean, both for a sample of size n taken from a normal population.

For a normal distribution the density is

$$\varphi(x) = \frac{1}{\sigma\sqrt{2\pi}}\ e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}\ .$$

As for any symmetrical distribution, mean and median coincide, thus $\tilde{m} = \mu$. In addition, we can choose $\mu = 0$; thus $\mathrm{MAD} = \mathrm{med}\{|x|\}$. According to (1) we then have to evaluate the limits for

$$\int_{-\alpha}^{\alpha} \varphi(x)\ dx\ =\ 1/2\ ,$$

or

$$\int_{0}^{\alpha} \varphi(x)\ dx\ =\ 1/4\ .$$

Tables give the numerical solution (for $\sigma = 1$)

$$\alpha\ \cong\ 0{,}674\,5\ . \tag{6}$$

We thus find that MAD can be linked, for a normal distribution, with the average standard deviation $\sigma(x)$ of a single observation x, by

$$\mathrm{MAD}\ \cong\ \alpha\ \sigma(x)\ .$$

By increasing n, the precision of MAD is improved but its value remains essentially unchanged. Only for the special case of n = 1 do we always have MAD = 0. This can be taken into account by writing

$$\mathrm{MAD}\ =\ \sqrt{\tfrac{n-1}{n}}\ \alpha\ \sigma(x)\ . \tag{7}$$

However, we must not forget that the uncertainty to be determined is that of a median, not of a mean value. From the theory of order statistics it is known that, in the case of a normal distribution, the (asymptotic) variance of the median, based on a sample of n values, is given by (see e.g. [2] or [3])

$$s^2(\tilde{m})\ \cong\ \frac{\pi}{2n}\ \sigma^2(x)\ . \tag{8}$$

Therefore, the uncertainty to be associated with the sample median $\tilde{m}$ is

$$s(\tilde{m})\ =\ \sqrt{\tfrac{\pi}{2n}}\ \sqrt{\tfrac{n}{n-1}}\ \frac{\mathrm{MAD}}{\alpha}\ \cong\ \frac{1{,}858}{\sqrt{n-1}}\ \mathrm{MAD}\ . \tag{9}$$

In other words, the required proportionality factor in (5) can be taken as

$$C \cong \frac{1,9}{\sqrt{n-1}}.$$

Note that (9) disagrees with a corresponding result recently given in [4], where C is simply taken as $1/\alpha$.

3. An example

To illustrate with a numerical example, we choose the six half-life measurements for ^{125}I performed in the framework of a recent international comparison of activity measurements [5]. The results, obtained in different laboratories, are (in units of days and ignoring the stated uncertainties), when arranged in order of increasing values,

$$59,26 \quad 59,29 \quad 59,38 \quad 59,39 \quad 59,40 \quad 59,90.$$

This leads to the sample median

$$\tilde{m} = \frac{1}{2} (59,38 + 59,39) \, d \cong 59,38 \, d .$$

To determine the uncertainty of $\tilde{m}$, we list the absolute deviations of the values from their median, again in increasing order, i.e.

$$0,00 \quad 0,01 \quad 0,02 \quad 0,09 \quad 0,12 \quad 0,52.$$

This gives for their median, according to (4),

$$MAD = \frac{1}{2} (0,02 + 0,09) \, d \cong 0,06 \, d ,$$

thus with (9)

$$s(\tilde{m}) \cong \frac{1,9}{\sqrt{5}} \, MAD \cong 0,05 \, d .$$

The resulting estimate for the half life of ^{125}I is therefore

$$T_{\frac{1}{2}} = (59,38 \pm 0,05) \, d , \tag{10a}$$

which compares favourably with the latest adjusted value [6] of

$$T_{\frac{1}{2}} = (59,408 \pm 0,008) \, d . \tag{10b}$$

A traditional analysis (without weights) of the six values gives the mean value (59,44 $\pm$ 0,10) d, whereas, after deletion of the highest value as a possible outlier, one finds (59,34 $\pm$ 0,03) d.

It will be noted that the suggested robust estimation method is extremely simple to apply and, in our example, leads directly to a reasonable result. Obviously, there exist more sophisticated approaches with a somewhat higher efficiency (cf. [1] or [4]); their justification, however, is much less obvious and is not always free of subjective decisions. As a start in planned applications, the use of the simple method based on the median should be adequate.

4. Remarks on applications

An important task of the BIPM is to organize and analyse international comparisons in the various fields of its activity. Traditionally, an essential part of such an exercise is the evaluation of a mean value with its respective uncertainty. Experience shows that the occurrence of discrepant results (outliers) is a rather frequent nuisance for the analyzer. While it may be necessary to neglect some data to protect the majority of participants from a misinterpretation, it is an unpleasant task to inform a national laboratory that its result must be eliminated. Obviously, the BIPM would prefer to avoid such decisions which may cause problems to laboratories.

As we have seen above, an analysis based on the median is largely insensitive to the existence of outliers (and their position). This is why we suggest that the new technique be applied, perhaps simply as a complement to the traditional analysis, in all international comparisons organized by the BIPM. In situations without outliers, the additional result may serve as a welcome check.

Clearly, the analysis of an intercomparison largely depends on its purpose. While the determination of a consensus value is often the objective, in other cases the main interest is on discrepant data. Thus, for example in radiotherapy, all results within a given margin (for instance ± 2 %) may be equally acceptable, whereas those outside pose a serious problem, as such irradiations are either useless or dangerous. For such results, the aim is to find a reliable location with respect to a stable reference value, such as the median.

It will be obvious that the use of a robust analysis for data of heterogeneous origin has a much wider field of application than intercomparisons. Similar problems occur each time a compiler tries to determine a "best value" for application in physics, chemistry or technology. In particular, the technique should also prove useful in the analysis of data on fundamental constants.

If the data to be compared are not produced simultaneously (or "blindly", as in an intercomparison), but are assembled over a period of time, additional problems occur since it is unrealistic to assume that they remain independent. Strongly discrepant results are normally not published. While the resulting distortion may have a moderate effect on the adopted mean value, such "psychological" correlations inevitably lead to an underestimation of the uncertainty of input values for an adjustment, possibly by a factor of two.

An extension of the discussed robust approach to data with different statistical weights is possible, but not considered an urgent task since the process of selecting such weights is usually subjective in nature.

The above remarks are clearly of a personal nature and should not be taken as an official BIPM policy in these matters.

I thank Dr. T.J. Quinn, Director of the BIPM, for the kind interest he has shown in the approach presented here.

APPENDIX

The purpose of this appendix is to show in a simple and explicit way that the sample median $t = \tilde{m}$ is indeed the solution for the condition

$$Q = \sum_{i=1}^{n} |x_i - t| = \text{minimum} \tag{A1}$$
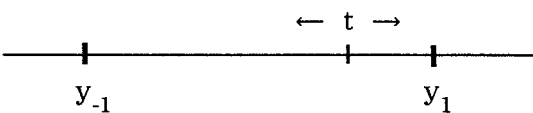
for a sample of n results x_i.

It is practical to consider for this purpose the measurements x_i in their ordered form, say

$$y_{-k} \leq y_{-k+1} \leq \cdots \leq y_0 \leq \cdots \leq y_{k-1} \leq y_k,$$

where $n = 2k+1$ for n odd, but without y_0 for an even number $n = 2k$ of results.

Let us consider the various possibilities.

- For <u>n = 2</u>:

$$\xleftarrow{\ \ } t \xrightarrow{\ \ }$$

If t is located between y_{-1} and y_1: $Q = y_1 - y_{-1} = Q_0$.

For t outside this region we have

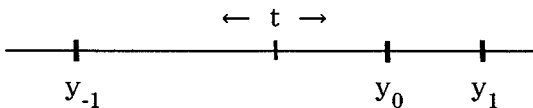$$Q = Q_0 + 2|t - y_1| \quad \text{or} \quad Q_0 + 2|t - y_{-1}| . \tag{A2}$$

The minimum Q_0 is reached for any t in the first configuration. One can choose $t = (y_1 + y_{-1})/2$.

In the more general case of n <u>even</u>, an equal number of measurements is added to the left $(y_{-2}, y_{-3}, ...)$ and to the right $(y_2, y_3, ...)$ of the interval considered above. Hence, the minimum

$$Q = Q_0 + \sum_{j=2}^{k} (y_j - y_{-j}) \tag{A3}$$

still applies for t between y_{-1} and y_1, as for n = 2.

- For <u>n = 3</u>:

$$\xleftarrow{\ \ } t \xrightarrow{\ \ }$$

In this case we obviously have (for t between y_{-1} and y_1)

$$Q = y_1 - y_{-1} + |t - y_0|$$

$$= y_1 - y_{-1} = Q_0 , \quad \text{if } t = y_0 . \tag{A4}$$

This feature remains if additional results of measurements are added symmetrically (as above). The minimum of Q thus corresponds to the choice $t = x_0$ for any <u>odd</u> value of n. The cases considered for n even or odd confirm the rule stated in (2).

References

[1] F.R. Hampel: "Introduction to 'Huber (1964) Robust estimation of a location parameter'", in "Breakthroughs in Statistics, Volume II" (Springer, New York, 1993), p. 479 ff.

[2] B.L. van der Waerden: "Mathematical Statistics" (Springer, New York, 1969), section 17

[3] S.S. Wilks: "Mathematical Statistics" (Wiley, New York, 1962), section 9.6

[4] D.F. Vecchia, J.D. Splett: "Outlier-resistant methods for estimation and model fitting", in "Advanced Mathematical Tools in Metrology" (World Scientific, Singapore, 1994), p. 143 ff.

[5] G. Ratel: "International comparison of activity measurements of a solution of ^{125}I (May 1988)", Rapport BIPM-90/3 (1990)

[6] J. Katakura, M. Oshima, K. Kitao, H. Iimura: "Nuclear data sheets for A = 125", *Nuclear Data Sheets* **70** (1993), 268

(April 1995)