Assessing the quality of an estimated value

Author: Nick Ridler, NPL, Malvern

Introduction

It is often helpful, when reporting an estimated value, to give some indication of the reliability of the estimate so that those using the estimate can assess its usefulness. Producing a figure for the reliability of an estimate often involves interpreting the dispersion of a series of scattered values. The scatter may have been produced by a single measurement system (e.g., from multiple, repeated measurements) or from multiple systems (e.g., from inter-laboratory test studies).

In ANAlyse 7, two methods of estimating a value were investigated; the arithmetic mean and the median. This Note, which is a companion to ANAlyse 7, discusses the choice of method for quantifying dispersion. Two methods are examined; the standard deviation and the inter-quartile range. The standard deviation is calculated from the estimated mean value, while the inter-quartile range is associated with the median estimate.

The data used to examine the properties of these two methods is the same as that used in ANAlyse 7 and was obtained during the second ANAMET measurement comparison exercise.

Standard deviation

The calculation for the standard deviation, \( s(x) \), is given by the following expression;

\[
s(x) = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2}
\]

where \( x_i \) represents the \( i \)th individual result (of which there are a total number \( n \)), and \( \bar{x} \) the estimated mean value.

To demonstrate the use of this equation, we shall calculate the standard deviation for seven \( S_{11} \) phase measurement results for the 3 dB attenuator at 1 GHz obtained during the second ANAMET measurement comparison exercise. The results were as follows; +118.2°, +120.6°, +115.2°, +108.9°, +112.9°, +116.3° and +114.0°.

The ANAlyse series of Technical Notes are produced by, and for, the members of ANAMET. They are intended for fast dissemination of technical information for discussion purposes and do not necessarily represent an official viewpoint. No responsibility is accepted by the author(s) or ANAMET for any use made of the information contained in this Note. This note has been approved by the ANAMET Steering Committee.
First, calculate the mean;

\[
\overline{\phi} = \frac{118.2 + 120.6 + 115.2 + 108.9 + 112.9 + 116.3 + 114.0}{7} = 115.157 \approx 115.2^\circ
\]

then the sum of squared deviations;

\[
\sum_{i=1}^{n} (\phi_i - \overline{\phi})^2 = (118.2 - 115.2)^2 + (120.6 - 115.2)^2 + (115.2 - 115.2)^2
\]

\[
+ (108.9 - 115.2)^2 + (112.9 - 115.2)^2 + (116.3 - 115.2)^2 + (114.0 - 115.2)^2
\]

\[
= 85.79
\]

and finally the standard deviation;

\[
s(\phi) = \sqrt{\frac{85.79}{6}} = 3.781 \approx 3.8^\circ
\]

Under normal circumstances, the standard deviation gives a useful measure of the dispersion of the results about the mean.

However, if we include the measurement made by an eighth participant (who measured the phase to be \(-58.3^\circ\)) and re-calculate the standard deviation, we get the following.

First, calculate the new mean value;

\[
\overline{\phi} = \frac{118.2 + 120.6 + 115.2 + 108.9 + 112.9 + 116.3 + 114.0 - 58.3}{8} = 93.475 \approx 93.5^\circ
\]

then the modified sum of the squared deviations;

\[
\sum_{i=1}^{n} (\phi_i - \overline{\phi})^2 = (118.2 - 93.5)^2 + (120.6 - 93.5)^2 + (115.2 - 93.5)^2 + (108.9 - 93.5)^2
\]

\[
+ (112.9 - 93.5)^2 + (116.3 - 93.5)^2 + (114.0 - 93.5)^2 + (-58.3 - 93.5)^2
\]

\[
= 26412.24
\]

and finally a new value for the standard deviation;

\[
s(\phi) = \sqrt{\frac{26412.24}{7}} = 61.426 = 61.4^\circ
\]

Including this extra result has caused a dramatic increase in the standard deviation (from 3.8° to 61.4°). This indicates that the standard deviation is giving a very erratic estimate for the dispersion of the results. The erratic behaviour has been caused by the inclusion of a result which is far removed from the other seven results. Such behaviour makes the standard deviation of limited use under these circumstances.

**Inter-quartile range**

Another method of quantifying dispersion is given by the inter-quartile range, which is the difference between the values for the upper and lower quartiles. Quartiles (like the median, discussed in
ANALyse 7) are evaluated by arranging the data to be analysed in ascending numerical order of size. The lower quartile divides the data so that 25% of the observations are below its value, the median has 50% below and 50% above, and the upper quartile has 25% above its value.

However, the ordered data cannot always be split exactly into four equal parts. This calls for more precise definitions for these terms, which are given in terms of a calculated 'depth', i.e., how far to go through the data to find the required value. For \( n \) observations, the depth for the median and quartiles are as follows:

**Median:**

\[
\text{Depth} = \frac{n + 1}{2}
\]

**Quartiles:**

\[
\text{Depth} = \frac{\text{Depth of median} + 1}{2}
\]

The depth for the lower quartile is found by entering the data from the low end (i.e., starting with the lowest value) and the depth for the upper quartile is found by entering at the high end (starting with the highest value).

The inter-quartile range is then simply the difference between these values. We will use the data already analysed (in terms of standard deviations) to illustrate finding the inter-quartile range.

Firstly, we will use our seven results which, when organised in ascending order of size, are:

\(+108.9^\circ \quad +112.9^\circ \quad +114.0^\circ \quad +115.2^\circ \quad +116.3^\circ \quad +118.2^\circ \quad +120.6^\circ\)

The depth for the median is \((7 + 1)/2 = 4\), so the depth for the quartiles is \((4 + 1)/2 = 2.5\), i.e., the quartiles are midway between the second and third values (when entering the data from either end).

\[
\text{Lower quartile} = \frac{112.9 + 114.0}{2} = 113.45^\circ
\]

\[
\text{Upper quartile} = \frac{116.3 + 118.2}{2} = 117.25^\circ
\]

\[
\text{Inter-quartile range} = 117.25 - 113.45 = 3.8^\circ
\]

Now, if we repeat the calculations using all eight results, organised in ascending order of size, we have:

\(-58.3^\circ \quad +108.9^\circ \quad +112.9^\circ \quad +114.0^\circ \quad +115.2^\circ \quad +116.3^\circ \quad +118.2^\circ \quad +120.6^\circ\)

The depth for the median is \((8 + 1)/2 = 4.5\), and the depth for the quartiles is \((4.5 + 1)/2 = 2.75\), i.e., the quartiles are three-quarters of the way between the second and third values (when entering the data from either direction).

For the lower quartile, the second and third values are \(+108.9^\circ\) and \(+112.9^\circ\), and three-quarters of the way between these two values (in an ascending direction) is \(+111.9^\circ\).

For the upper quartile, the second and third values from the top are \(+118.2^\circ\) and \(+116.3^\circ\), and three-quarters of the way between these two values (in a descending direction) is \(+116.775^\circ\).

---

1 Some textbooks give slightly different expressions for finding the depth for quartiles. However, the one used here is the same as that in the software used to analyse the data in the second ANAMET measurement comparison exercise.
The inter-quartile range is therefore 116.775 - 111.9 = 4.875° ≈ 4.9°

The change in the inter-quartile range when a very different result has been included in the results being analysed is relatively small (it has increased from 3.8° to 4.9°). This indicates that the inter-quartile range estimate is less vulnerable to variation due to the inclusion of a result which is very different from the other, majority, values. Such resistant behaviour makes the inter-quartile range a useful estimator of dispersion when a relatively small number of observations, containing a contaminated value, are being analysed.

Discussion

The Table below summarises the calculations for the standard deviation and inter-quartile range for the phase measurement results given in this Note.

<table>
<thead>
<tr>
<th></th>
<th>Calculation based on seven results.</th>
<th>Calculation based on eight results (with one contaminated value).</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard deviation</td>
<td>3.8</td>
<td>61.4</td>
</tr>
<tr>
<td>Inter-quartile range</td>
<td>3.8</td>
<td>4.9</td>
</tr>
</tbody>
</table>

It is worth remembering that the two intervals produced by the standard deviation and the inter-quartile range have different meanings. For measurement results from a normal (or Gaussian) distribution, the standard deviation defines an interval about the mean which will contain approximately 68% of the values. Whereas the inter-quartile range defines an interval, about the median, which contains half the total number of measurement results - one quarter on either side of the median. An important distinction between the two intervals is that, for the inter-quartile range, no assumptions need to be made about the nature of the underlying distribution (e.g., normal, rectangular etc). Assuming that a set of results comes from a normal distribution can be difficult to justify when the sample of results is relatively small.

A second consideration in the use of such statistics is their ease of calculation. Most pocket calculators have a facility for calculating the standard deviation for a series of measurement results, so that the calculation can be made fairly easily. The inter-quartile range requires the measurement results to be sorted before calculations can begin. This usually results in a more lengthy calculation time. However, such calculations can often be made using software spreadsheet packages on computers.

Conclusions

Statistics can be used to gain insight into the reliability of an estimated value. However, it is important to choose an appropriate statistic for the particular set of measurement results being analysed. In particular, a set of results containing a value exhibiting a gross difference from the other values is particularly problematical. Under such circumstances, a more robust method than standard deviation is desirable. There are many such methods of estimating the reliability of a value which fall into this category, but perhaps the inter-quartile range is one of the easier to compute and interpret.