

How to spot an outlier

Analytical method validation can be a challenging beast to handle at the best of times, and after all mistakes happen – we are only human after all.

A single erroneous data point can be enough to skew results and cause a validation to fail.

Just one simple mistake in sample preparation for example can lead to hours of wasted sample preparation and instrument run time – a nightmare scenario in any high throughput lab.

Fortunately, this can be easily avoided through careful planning of a method validation. The inclusion of an additional linearity or quality control (QC) sample than the regulated minimum requirement allows for the rejection of the spurious data point with the requirements still met.

But, how exactly do we determine what is and isn't an outlier? How do we justify the rejection of a suspected spurious data point?

Fear not for an unlikely hero exists – statistics!

There are several statistical tests used throughout the world of analytical chemistry to determine the presence of a single outlier such as the Grubbs^[1] test and Dixons Q^[2] test.

In the following example we see how the Dixons Q test can be used to identify and reject an outlier from a set of quality control (QC) samples.

Example of Outlier Rejection Using Dixons Q Test

The following is an example data set of quality control (QC) samples taken from a method validation for an impurity assay. This assay forms part of a submission to EU regulatory bodies for the import of a pesticide or plant protection product (PPP).

These QC samples have been prepared by spiking a known amount of the impurity (20.02 µg/mL) onto 6 independently weighed technical samples. This is equivalent to 0.1% w/w level of the impurity with respect to the active ingredient content. According to the guidance document for pesticide submission in the EU SANCO/3030/99 rev.5^[3] these samples can be used to determine the repeatability and limit of quantification (LOQ) of the impurity.

Within this document (section 4.1.2.) it defines that a minimum of 5 independent weighing's are required for determining the precision of a method.

In this example 6 samples have been prepared in order to allow for the rejection of a single outlier and still meet the requirements of the validation.

Results for the 6 samples are shown in Table 1, sample number 6 appears to show an obvious spurious result.

Table 1 Initial Results of the QC Samples (n=6)

Sample No.	Recovered Conc (µg/mL)	Recovery (%)	Average Recovery	Relative Standard Deviation (RSD)
1	20.15	100.6	105.93	11.99
2	20.06	100.2		
3	20.01	100.0		
4	19.95	99.7		
5	20.18	100.8		
6	26.89	134.3		

According to the guidance document for this result to be considered acceptable the RSD should be ≤ 5.36 or ≤ 10.72 with a suitable explanation provided. The result of 11.99 would lead to the validation failing.

So, how do we determine if sample 6 is indeed an outlier, and save the day?

The Dixons Q test, first of all we need to find $Q_{\text{experimental}}$ which is calculated as follows:

$$Q_{\text{exp}} = \frac{(\text{Outlier value} - \text{nearest value})}{(\text{Largest value} - \text{smallest value})}$$

Once we have this value, we then need to compare it to the Q_{critical} value which is determined using a Dixons Q table based on the number of points within the data set.

Table 2 Dixons Q Table

Number of data points	Q_{crit} (CL:90%)	Q_{crit} (CL:95%)	Q_{crit} (CL:99%)
3	0.941	0.970	0.994
4	0.765	0.829	0.926
5	0.642	0.710	0.821
6	0.560	0.625	0.740
7	0.507	0.568	0.680
8	0.468	0.526	0.634
9	0.437	0.493	0.598
10	0.412	0.466	0.568

CL refers to confidence limit – that is to say that at CL:90% we can say that the chance of an erroneous rejection of a data point would be 10%.

The most commonly used Q_{crit} value is at 95% CL, however a lab should define which limit is acceptable for use within their standard operating procedures (SOP).

So, for our example data set the Q_{exp} value would be calculated as follows:

$$Q_{exp} = \frac{(\text{Outlier value} - \text{nearest value})}{(\text{Largest value} - \text{smallest value})} = \frac{(26.89 - 20.18)}{(26.89 - 19.95)}$$

This calculation gives a Q_{exp} value of 0.967.

The number of data points in our data set is $n=6$ therefore using the Dixons Table our Q_{crit} value, with a confidence value of 95% is 0.625.

In order for a result to be rejected the test demands that $Q_{exp} > Q_{crit}$.

In this example we have calculated a Q_{exp} value of 0.967 which is greater than the Q_{crit} value of 0.625, therefore we can confidently state that sample number 6 is indeed an outlier and can reject the results from this sample.

With this in mind we can now recalculate our QC sample results with the data points adjusted to $n = 5$, excluding sample 6.

The recalculated results are shown in Table 3.

Table 3 Final Results of the QC Samples ($n=5$) – excluding sample 6 as outlier according to Dixons Q

Sample No.	Recovered Conc ($\mu\text{g/mL}$)	Recovery (%)	Average Recovery	RSD
1	20.15	100.6	100.25	0.43
2	20.06	100.2		
3	20.01	100.0		
4	19.95	99.7		
5	20.18	100.8		
6	26.89	134.3		

The results of the QC samples at $n=5$ show a massive improvement on the original calculation, now that the outlier has been removed.

The QC samples now meet the requirements of the guidance document and so the validation will pass.

Conclusion

Statistical tests such as Dixons Q and the Grubbs test are very useful tools which can be easily implemented to determine whether or not a single suspect data point is an outlier.

These tools, along with careful planning of a method validation, can lead to a reduction in the number of repeat assays required, helping to ensure the high throughput your lab requires.

References

- <https://www.itl.nist.gov/div898/handbook/eda/section3/eda35h1.htm> Grubbs Test
- <https://www.statisticshowto.com/dixons-q-test/> Dixons Q Test
- https://food.ec.europa.eu/system/files/2019-03/pesticides_ppp_app-proc_guide_phys-chem-ana_3030.pdf SANCO/3030/99 rev.5 European Commission, Pesticides, Plant Protection Products, 22 March 2019