

12.6 Additional notes on implementing control charts

We offer here a few additional notes regarding the implementation of control charts in real applications. The control-chart dialog in PRIMER 8 offers many options. It is especially important to pay close attention to all of the choices that can affect the null hypothesis and/or the decision criterion, i.e., the upper control-chart limit (U_{CL}). We offer below some comments on these topics.

Start with a decent sample size

Control charts have historically arisen from industrial settings, where sample sizes, particularly for establishing baseline information, are typically very large. It is important to recognise that we are trying here to characterise the entire distribution's shape (for the in-control samples), and not just to estimate a centroid. Therefore, we should always apply the control-chart tool with a view to including as many 'in-control' (reference) samples as we possibly can. Mathematically, there are lower limits on the number of in-control points we need in order to run the analysis (i.e., $n_c = 4$ points), but as a general rule, we should typically aim to run the control-chart routine on no fewer than $n_c = 10$ sample points, and having more ($n_c = 20$ or 30) would certainly be preferable.

If your total sample size is $N \geq 11$, then the default for the **Control Chart** routine in P8 for the minimum number of in-control samples is $n_c = 10$. If $N < 11$, then the default is $n_c = N - 1$, but with a strict lower bound of $n_c = 4$.

A further practical point is that the **Control Chart** routine in P8 cannot handle missing values, so these will need to be removed prior to running the routine.

Be aware of H_0 for different types of control chart

The null hypothesis (H_0) for the specific test done at each time point in a given control chart depends critically on the **type of control chart** you are running: *progressive*, *fixed baseline* or *moving window*. You need to carefully consider which type of control chart is appropriate for your particular application (there may be more than one).

The default in PRIMER 8 is to run the control-chart by reference to a fixed baseline set of $n_c = 10$ samples. However, the number of 'in-control' samples clearly needs to be thought about carefully and set to something appropriate for each specific dataset, driven by the null hypothesis of interest.

It is also important to consider how each type of control chart plays out in the specific tests it performs through time. For example, the 'progressive' type of control-chart may not produce output that 'makes sense' after an 'out-of-control' sample has been identified. For example, suppose you are looking at a progressive control-chart and an 'out-of-control' point has been

identified at time-point t . The progressive chart will subsequently include that point at time t as part of the 'in control' distribution of samples when it goes on to test subsequent time-points $(t+1)$, $(t+2)$, etc. This might not be appropriate. One might consider removing the out-of-control sample before proceeding with the subsequent tests. These sorts of decisions will depend on the specific hypotheses to be examined for any particular dataset.

Be aware of important settings affecting U_{CL}

The upper control chart limit U_{CL} and hence the assessment of whether a point is in control or out of control will clearly be critically affected by the following choices:

- choice of parametric vs non-parametric approach
- choice of α -level (e.g., 0.05)
- choice to apply shrinkage (or not) in estimating the variance-covariance matrix
- choice of ordination method (PCO, mMDS or tmMDS)
- choice of m , the dimensionality of the ordination

The *defaults* for the **Control chart** routine in PRIMER 8 will be quite sensible for a pretty wide variety of cases. These defaults are:

- non-parametric
- $\alpha = 0.05$
- apply shrinkage
- use threshold metric MDS (tmMDS)
- choose m so that the matrix correlation is $r_{e,d} = 0.99$.

However, thinking carefully about each of these choices is almost always warranted. For example, it is useful to observe that the default choice of 'non-parametric' may not be particularly sensible if the sample size n_c is quite small (less than 10).

Revision #15

Created 5 July 2025 01:16:34 by Marti

Updated 26 April 2026 23:50:49 by Marti