Robust Statistics in BD FACSDiva™ Software

Robust statistics defined

Robust statistics provide an alternative approach to classical statistical estimators such as mean, standard deviation (SD), and percent coefficient of variation (%CV). These alternative procedures are more resistant to the statistical influences of outlying events in a sample population—hence the term "robust." Real data sets often contain gross outliers, and it is impractical to systematically attempt to remove all outliers by gating procedures or other rule sets.

The robust equivalent of the mean statistic is the median. The robust SD is designated rSD and the percent robust CV is designated %rCV. For perfectly normal distributions, classical and robust statistics give the same results.

How robust statistics are calculated in BD FACSDiva™ software

Median
The mean, or average, is the sum of all the values divided by the number of values. For example, the mean of the values of [13, 10, 11, 12, 114] is 160 ÷ 5 = 32. If the outlier value of 114 is excluded, the mean is 11.5.

The median is defined as the midpoint, or the 50th percentile of the values. It is the statistical center of the population. Half the values should be above the median and half below. In the previous example, the median is 12 (13 and 114 are above; 10 and 11 are below). Note that this is close to the mean with the outlier excluded.

Robust standard deviation (rSD)
The classical SD is a function of the deviation of individual data points to the mean of the population. Similarly, the robust SD is based upon the deviation of individual data points to the median of the population. It is calculated as:

$$rSD = \left(\text{Median of}\left\{\left|X_i - Median_x\right|\right\}\right) \times 1.4826$$

The value 1.4826 is a constant factor that adjusts the resulting robust value to the equivalent of a normal population distribution. Thus, for a normally distributed population, the SD and the rSD are equal.

Robust coefficient of variation (rCV and %rCV)
The classical CV is equal to the population SD divided by the population mean.

Similarly, the robust CV and percent robust CV are calculated as:

$$\text{rCV} = \frac{rSD}{Median_x} \qquad \%\,\text{rCV} = \frac{rSD}{Median_x} \times 100\%$$

Population	%CV	%rCV
Total	128	37.9
Gated [P2]	38	37.9

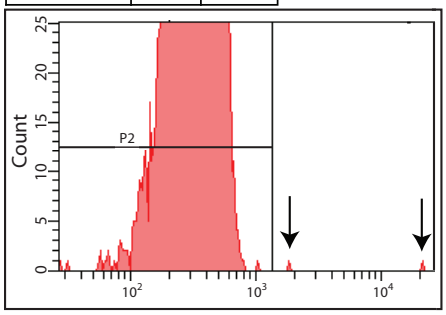

Figure 1. Impact of outlying events on %CV and %rCV

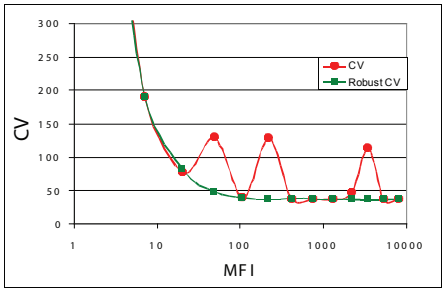

Figure 2. Comparison of data analysis using %CV vs %rCV

How robust statistics are used in BD FACSDiva software

Robust statistics are used within the BD™ Cytometer Setup & Tracking (CS&T) module of BD FACSDiva software and are available as selectable statistics on worksheets. The CS&T module calculates and presents various performance parameters including Qr (detector efficiency), Br (optical background), and SD_{EN} (standard deviation of electronic noise), which are based upon estimates of the %CVs of the BD FACSDiva bright, mid, and dim setup beads. The accuracy of these performance parameters is dependent upon the accuracy of the statistical estimators. Typically, obtaining good statistical measurements on uniform bead populations requires placing a tight gate around the population of interest. Figure 1 shows the %CV of a population of CS&T dim beads with (Gated [P2]) and without (Total) a tight gate. Data in Figure 1 are from the MFI = 200 sample in Figure 2.

There is a large discrepancy in the %CVs (128 vs 38) of these two populations due to the presence of only two outlying events (beads, indicated by the arrows) out of 4,500 total analyzed. However, the calculated %rCV is equivalent for both the gated and ungated populations (37.9 vs 37.9). Thus, using %rCV facilitates obtaining relevant statistical information without having to set individual gates on every fluorescence parameter. This is why the BD FACSDiva software's CS&T baseline and performance reports show only robust statistics.

Figure 2 shows a plot similar to the PMTV (photomultiplier tube voltage) optimization plots displayed in the CS&T baseline report. The dim beads were analyzed at different PMTVs with the resulting median fluorescence intensities plotted against the %CV and %rCV. Three of the samples show significant differences between %CV and %rCV. In each case, the difference is due to the presence of two to four outlying events (data points).

In summary, robust statistics are new analytical tools in BD FACSDiva software. They minimize the impact of small numbers of outlying data points on statistical estimates. When used in cytometer setup and tracking, robust statistics ensure the accuracy of the performance parameters. Robust statistics can also reduce the need to set tight gates around populations of interest.

References

1. Huber PJ. *Robust Statistics*. Wiley. 1981.

23-9609-01