Our Statistics – What Do They Mean?

The five statistics we have added to our reports are designed to quantify confidence in results, and to help you decide how many spur or cane samples are needed to produce your desired level of confidence.

Statistics on bunches per bud

30	Number of buds in sample
0.033	Precision (change from subtracting 1 bunch from total)
0.691	Standard Deviation of bunches per bud
0.21	Margin of Error with 90% confidence
129	Bud sample size needed for 0.1 Margin of Error

All of these statistics apply to the average or mean number of bunches per bud.

In this example, a grower sent us 10 spurs of 3 buds each, for a total of 30 buds in the sample. We counted a total of 32 bunches, so the mean was 32/30 = 1.067 bunches per bud.

Each of the 32 bunches added 1/30 or 0.033 to the total. If the total was 1 bunch lower, or 31 bunches, the mean would change to 1.033 bunches per bud. It is impossible to have a mean between 1.033 and 1.067. So 0.033 is the **Precision** of the sample. Increasing the sample size makes the estimate more precise (smaller number).

The **Standard Deviation** (SD) is a measure of how much the results vary among buds. If all the buds are the same, such as 1 bunch in every bud, then the SD is 0. If the sample has many buds with 0, 1, and 2 bunches, the SD increases. The SD of 0.691 in this example is high when compared to an average of 1.067.

The SD of the sample is an estimate of the SD of the underlying population. In other words, if we counted bunches in every bud which will remain on the vines after you prune, we could calculate the population SD. If our sample of 30 buds is perfectly unbiased, then the population SD will be equal to the sample SD of 0.691. Of course, a sample is never unbiased, so the two SDs will almost certainly be different.

Increasing the sample size will improve the accuracy of the estimated SD, just as it will improve the accuracy of the estimated mean number of bunches. Accuracy means how close the sample statistics are to the underlying population statistics. But we cannot guess whether the estimated

SD will increase or decrease. In theory, increasing the sample size does not change the SD.

The statistic we want to reduce is the **Margin of Error** (E). E is calculated using the standard deviation and the confidence level.

The **Confidence Level** is the percentage of correct estimates. Correct means close to the mean, or within the margin of error. In the example, E is 0.21 for a mean of 1.067. The confidence level is 90%, so if we repeat our sampling 10 times, then we expect that only 1 sample will have a mean below 0.857 or above 1.277. 9 out of 10 or 90% will have means within that margin of error.

Increasing the confidence level also increases the margin of error. For example, if the confidence level is 95%, we expect only 1 sample out of 20 to have a mean outside the margin of error. For our grower, the margin of error for 95% confidence increases to 0.25. For 99% confidence, or 99 out of 100, the margin of error is 0.33.

A large standard deviation also increases the margin of error. If most of the buds are the same, then it will not matter very much which ones are sampled. The SD will be small, and the sample means will all be close to the population mean. If the buds have different numbers of bunches, then random samples will produce different estimated means, and both the SD and the margin of error will be greater.

The only way to reduce the margin of error at a given confidence level is to increase the sample size. Consider this formula for calculating the margin of error (E) from the standard deviation (s) and the number of buds sampled (n):

$$\mathbf{E = z*s/\sqrt{n}}$$

z is a factor that represents the confidence level, taken from a table:

Confidence	**z**
99%	**2.58**
95%	**1.96**
90%	**1.64**
80%	**1.28**

So the margin of error is z times the SD divided by the square root of the sample size. We used this formula to calculate E as 0.21, 0.25, and 0.33 for 90%, 95% and 99% confidence, respectively.

If we rearrange the formula, we can calculate the number of buds we need to sample to achieve a desired margin of error.

$n = (z*s/E)^2$

So the number of buds needed is z times the standard deviation divided by the desired margin of error, all squared. For our example grower, we decided we want a margin of error of plus or minus 0.1 with 90% confidence. This means that 9 out of 10 sample means will fall between 0.967 and 1.167 (if we pretend that 1.067 is the actual population mean). Plugging in the numbers:

$129 = (1.64 * 0.691/ 0.1)^2$

The grower would need to sample 129 buds (43 spurs) to achieve this small margin of error of plus or minus 0.1.

It is important to remember that all of these statistics assume that the samples are random and unbiased. This ideal is almost impossible for growers to achieve with their sampling. Sampling bias is not factored into any of these statistics, and we have no way of estimating that bias.