A Note on Degrees of Freedom

In a general sense, **Degrees of Freedom** are the number of observations in a sample that are free to vary while estimating statistical parameters. **Typically, the degrees of freedom equal your sample size minus the number of parameters you need to calculate during an analysis**.

In other word, degrees of freedom of an estimate are the number of independent pieces of information (observations) in a sample that went into calculating the estimate.

Suppose we collect the random sample of observations shown below. Now, imagine that we know the mean, but we don't know the value of an observation—the X in the table below.

6,8,5,9, 6, 8, 4, 11, 7, X   subject to constraint **mean = 6.9**

Now, the mean is 6.9, and it is based on 10 values. So, we know that the values must sum to 69 based on the equation for the mean.

Sum = mean* number of observation = 69

Using Algebra we can find

$6 + 8 + 5 + 9 + 6 + 8 + 4 + 11 + 7 + X = 69$

$X = 69 - 65 = 5$

As you can see, that **last number has no freedom to vary**. **It is not an independent piece of information because it cannot be any other value.**

The degrees of freedom in case of sample are n-1. WHY?

For example, the population variance measures the average of the sum of squared deviation from the population mean which we obtain by dividing sum of squared deviation by number of observation N.

However, in most cases, population mean or population variance are not known. Hence we have to rely on sample mean and sample variance. Remember, sample variance is a random variable that changes from sample to sample, but the population variance is a constant.

When we calculate the sample standard deviation from a sample of n values, we are using the sample mean already from that same sample of n values. We are not using population mean since we are don't know population mean. Sample mean acts as a constraint in the calculation of sample variance. The calculated sample mean has already "used up" one of the "degrees of freedom" **to which the n sample values are "free to vary" around the mean**. Hence only n-1 degrees of freedom of variability are left for the calculation of the sample standard deviation.

2. Standard deviation

If we know the population mean, the standard deviation is given by

$$\sigma = \sqrt{\frac{\sum(x-\mu)^2}{n}} \quad (2)$$

If we don't know the population mean, we can use sample mean to calculate the standard deviation.

$$s = \sqrt{\frac{\sum(x-\bar{x})^2}{n-1}} \quad (3)$$

Why different degree of freedom for the sample standard deviation and the population standard deviation?

By any chance, if we know the population mean, then one can take the sum of squares of differences from the population mean and divide it by n to get an unbiased estimate of the population variance.