**Degrees of Freedom**

 I asked my students to think of one number between 0 and 10 and then write it down on a slip of paper. Then I announced that we were going to create a set of 5 scores with a mean of 5. I called on four students, one at time, for the value of their number. Then I noted that the value of the fifth number was already determined because we had exhausted our degrees of freedom with the *N*–1 scores already selected. Putting a restriction on the process cost us that one degree of freedom. For example, suppose the four scores were 3, 4, 6, and 9. They sum to 22. For the mean to be five, the sum must be 25 and there is only one fifth score that makes the sum 25. The fifth score must be a 3.

**Calculating the Sample Variance**

 You start out with *df* = sample size, the number of independent pieces of information (scores) you have. When you are estimating a parameter, you lose one degree of freedom for every other parameter you estimate in the process. The population variance is defined as $\frac{\sum\_{}^{}\left(Y-μ\right)^{2}}{N}$. When we want to estimate the population variance from sample data we calculate $\frac{\sum\_{}^{}\left(Y-M\right)^{2}}{N-1}$. Since we do not know the population mean, we estimate it with the sample mean, and that reduces our degrees of freedom from *N* to *N*-1.

**Independent Samples *t* Test**

 The *t* is the ratio of the difference in sample means to the standard error. The standard error is . Each of the within groups sums of squares is $\sum\_{}^{}\left(Y-M\right)^{2}$. Since we have two groups, we have estimated two parameters, and that cost us two degrees of freedom. Our degrees of freedom are *N*-2.

**Correlated Samples *t* Test**

 The *t* is the ratio of the mean difference score (subject’s score in the one condition minus that in the other condition) to the standard error. The standard error is the standard deviation of the difference scores divided by the square root of n, where n is the number of difference scores. The standard deviation is the square root of $\frac{\sum\_{}^{}\left(Y-M\_{d}\right)^{2}}{N-1}$, where *Md* is the sample mean difference score, our estimate of the population mean difference score. By estimating the population mean difference score we lost one degree of freedom, so our degrees of freedom = *n*-1.

**Regression Analysis**

**Univariate Regression.** The model is Y = intercept. In other words, Y = *My*. You are estimating one parameter, so your model *df* = 1 and your residual *df* = *N*-1. In a one-sample *t* test, you test the model that Y = *Mnull*, with *df* = *N*-1.

**Bivariate Regression.** The model is Y = *a* + *b*X. You are estimating two parameters, the intercept and the slope. Your model *df* = 2 and your residual *df* = *N*-2. If X is group membership, you are doing a two independent samples *t* test and error *df* = *N*-2.

**Trivariate Regression.** The model is Y = *a* + *b1*X1 +*b2*X2. You are estimating three parameters, the intercept and two slopes. Your model *df* = 3 and your residual *df* = *N*-3. If X1 and X2 are dichotomous dummy variables (if X1 = 1 indicates the case is in Group 1 and X2 = 1 indicates the case is in Group 2, with cases in Group 3 having scores of 0 for both X1 and X2), then you are doing a one-way ANOVA, and the error *df* = *N*-3. And so on.



[Also see Degrees of Freedom by Jim](https://statisticsbyjim.com/hypothesis-testing/degrees-freedom-statistics/)

[Karl L. Wuensch](http://core.ecu.edu/psyc/wuenschk/klw.htm), December, 2019