The **one-way analysis of variance (ANOVA)** is used to determine whether there are any statistically significant differences between the means of two or more independent (unrelated) groups (although you tend to only see it used when there are a minimum of three, rather than two groups). For example, you could use a one-way ANOVA to understand whether exam performance differed based on test anxiety levels amongst students, dividing students into three independent groups (e.g., low, medium and high-stressed students). Also, it is important to realize that the one-way ANOVA is an **omnibus** test statistic and cannot tell you which specific groups were statistically significantly different from each other; it only tells you that at least two groups were different. Since you may have three, four, five or more groups in your study design, determining which of these groups differ from each other is important. You can do this using a post hoc test (N.B., we discuss post hoc tests later in this guide).

The **Kruskal-Wallis test** is a nonparametric (distribution free) test, and is used when the assumptions of one-way ANOVA are not met. Both the Kruskal-Wallis test and one-way ANOVA assess for significant differences on a continuous dependent variable by a categorical independent variable (with two or more groups). In the ANOVA, we assume that the dependent variable is normally distributed and there is approximately equal variance on the scores across groups. However, when using the Kruskal-Wallis Test, we do not have to make any of these assumptions. Therefore, the Kruskal-Wallis test can be used for both continuous and ordinal-level dependent variables. However, like most non-parametric tests, the Kruskal-Wallis Test is not as powerful as the ANOVA.

Linear regression is a basic and commonly used type of predictive analysis. The overall idea of regression is to examine two things: (1) does a set of predictor variables do a good job in predicting an outcome (dependent) variable? (2) Which variables in particular are significant predictors of the outcome variable, and in what way do they–indicated by the magnitude and sign of the beta estimates–impact the outcome variable? These regression estimates are used to explain the relationship between one dependent variable and one or more independent variables. The simplest form of the regression equation with one dependent and one independent variable is defined by the formula $y = c + b*x$, where y = estimated dependent variable score, c = constant, b = regression coefficient, and x = score on the independent variable.

- Simple linear regression
  1 dependent variable (interval or ratio), 1 independent variable (interval or ratio or dichotomous)
- Multiple linear regression
  1 dependent variable (interval or ratio) , 2+ independent variables (interval or ratio or dichotomous)
- Logistic regression
  1 dependent variable (dichotomous), 2+ independent variable(s) (interval or ratio or dichotomous)
- Ordinal regression
  1 dependent variable (ordinal), 1+ independent variable(s) (nominal or dichotomous)

- Multinominal regression
  1 dependent variable (nominal), 1+ independent variable(s) (interval or ratio or dichotomous)

  A **t-test** is an analysis framework used to determine the difference between two sample means from two normally distributed populations with unknown variances. A t-test is an analysis of two populations means through the use of statistical examination; a t-test with two samples is commonly used with small sample sizes, testing the difference between the samples when the variances of two normal distributions are not known.

  A t-test looks at the t-statistic, the t-distribution and degrees of freedom to determine the probability of difference between populations; the test statistic in the test is known as the t-statistic. To conduct a test with three or more variables, an analysis of variance (ANOVA) must be used.

  The **Wilcoxon signed-rank te**st is the nonparametric test equivalent to the dependent t-test. As the Wilcoxon signed-rank test does not assume normality in the data, it can be used when this assumption has been violated and the use of the dependent t-test is inappropriate. It is used to compare two sets of scores that come from the same participants. This can occur when we wish to investigate any change in scores from one time point to another, or when individuals are subjected to more than one condition.

  F1 score = 2 * (Precision * Recall) / (Precision + Recall)

  Z-score normalization: To calculate a z-score, **subtract the mean from the raw score and divide that answer by the standard deviation**. (i.e., raw score =15, mean = 10, standard deviation = 4. Therefore 15 minus 10 equals 5. 5 divided by 4 equals 1.25.