**10-30 notes**

Hypothesis testing and Statistical Tests

Goals:

By the end of this lesson, you will:

* Make progress toward completing your project, finding and analyzing a dataset of your choice.
* Be able to apply ANOVA to compare the means of three or more groups.

Explain the difference between a null hypothesis and an alternate hypothesis.

* Null hypothesis - Statement or claim being made (which we are trying to disprove)
* Alternative hypothesis - Hypothesis that we are trying to prove and which is accepted if we have sufficient evidence to reject the null hypothesis

Apply a one-sample t-test to determine if sample and population data are significantly different.

* Goodness of fit test

Apply a two-sample t-test to determine if two groups are significantly different.

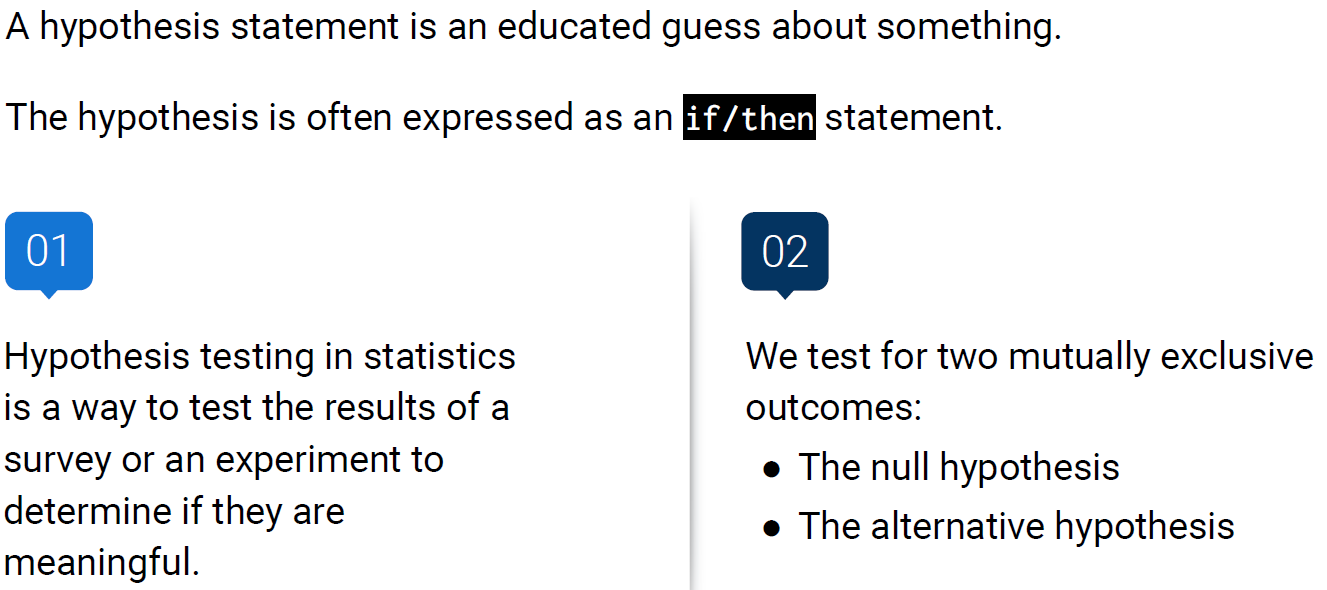
Apply ANOVA to compare the means of three or more groups.

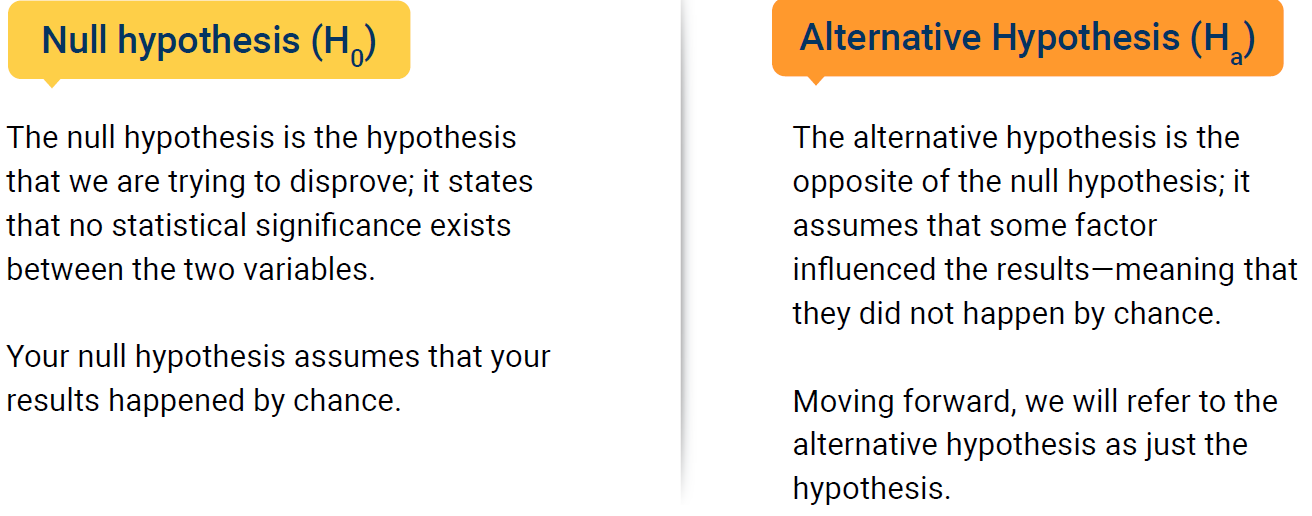
* Continuous data (ie: Age)

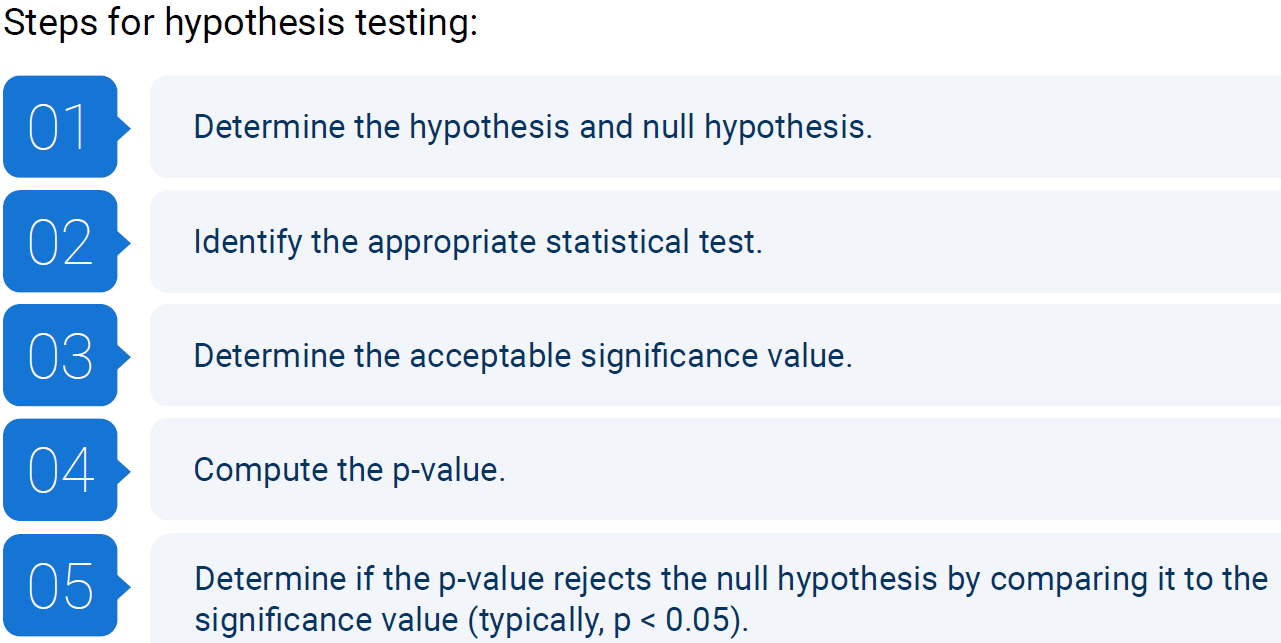
Perform a chi-square test to compare the distribution of categorical data.

* Categorical data (ie:)

**MUST CREATE NULL AND ALTERNATIVE HYPOTHESES BEFORE PERFORMING ANALYSIS**







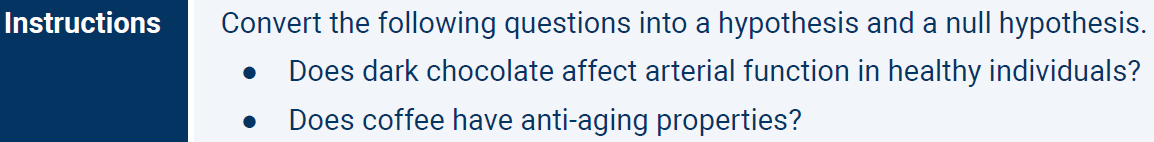
Step 2- is it continuous or categorical data

Step 3- confidence level that’s acceptable (ie: 90% is 0.90, 95% is 0.95)

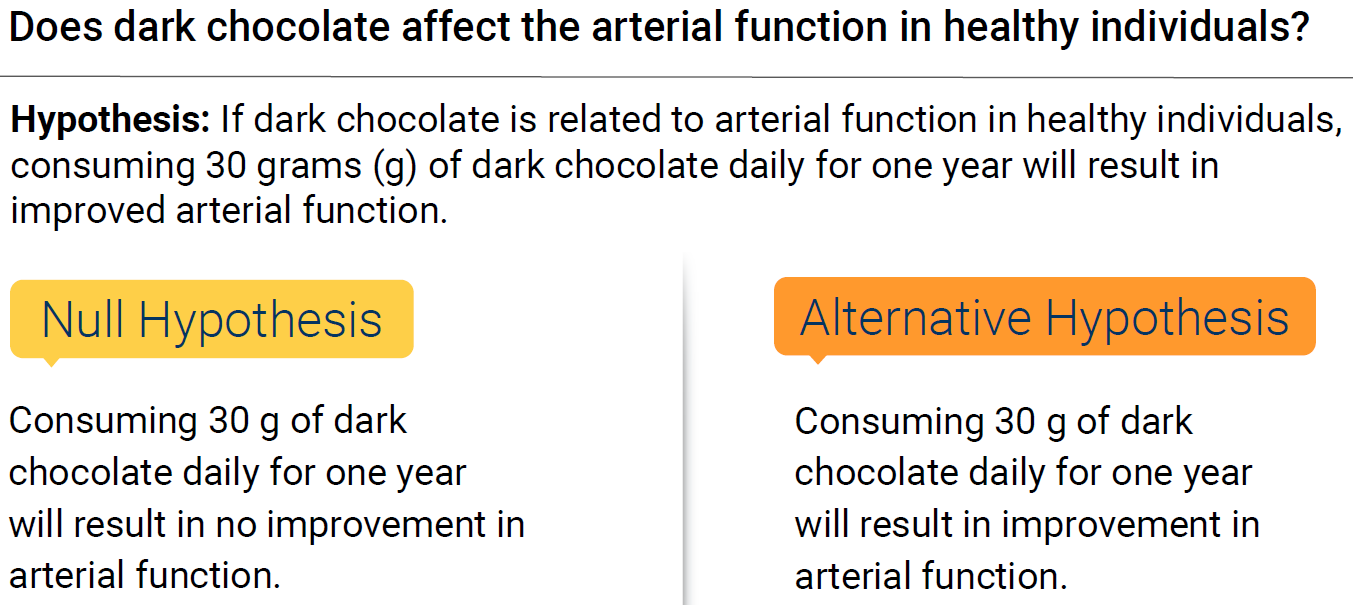
Step 5- If p is less than 0.05, you reject the null hypothesis.

In an analysis, we test the hypothesis and decide whether we can reject the null hypothesis. If we can, the only option left is the alternative hypothesis!

Activity:

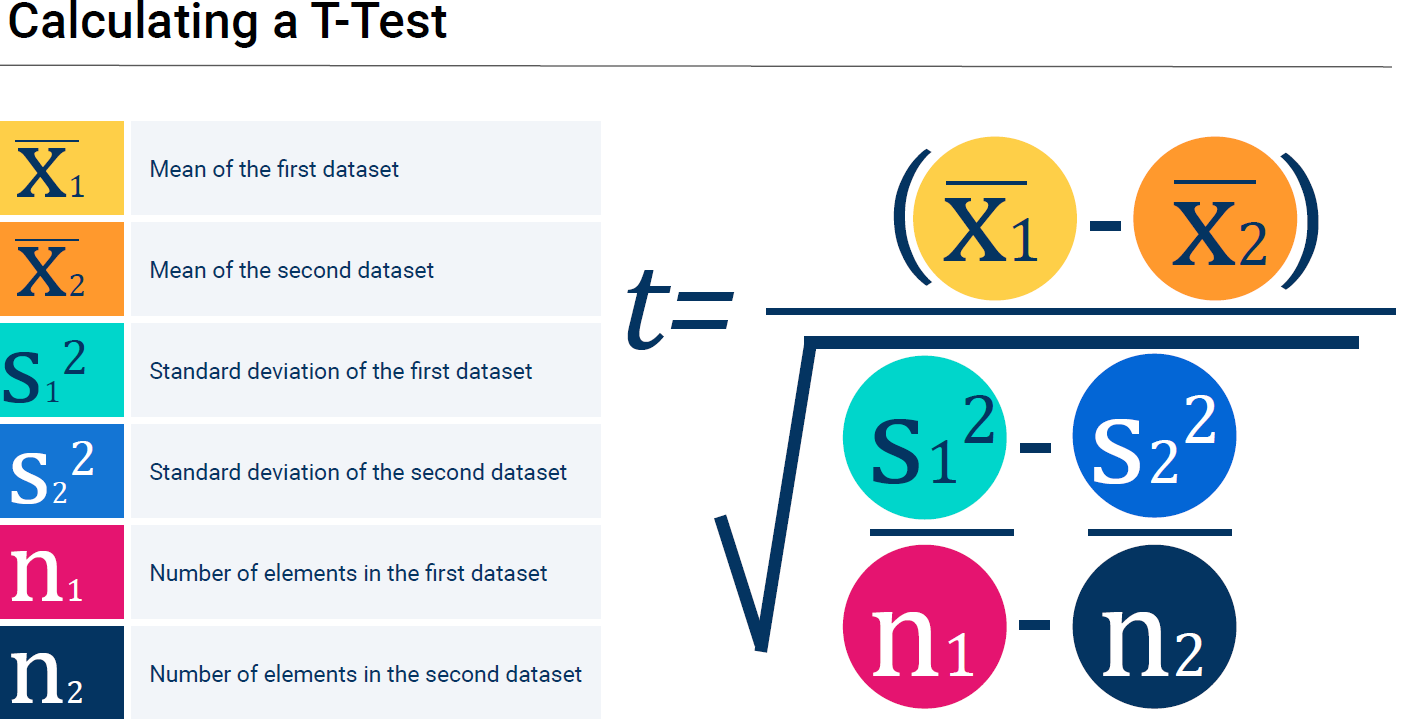


If you wanted to prove that chocolate does impact arterial function, the null is that it doesn’t impact arterial function, and the alternative is that it does.

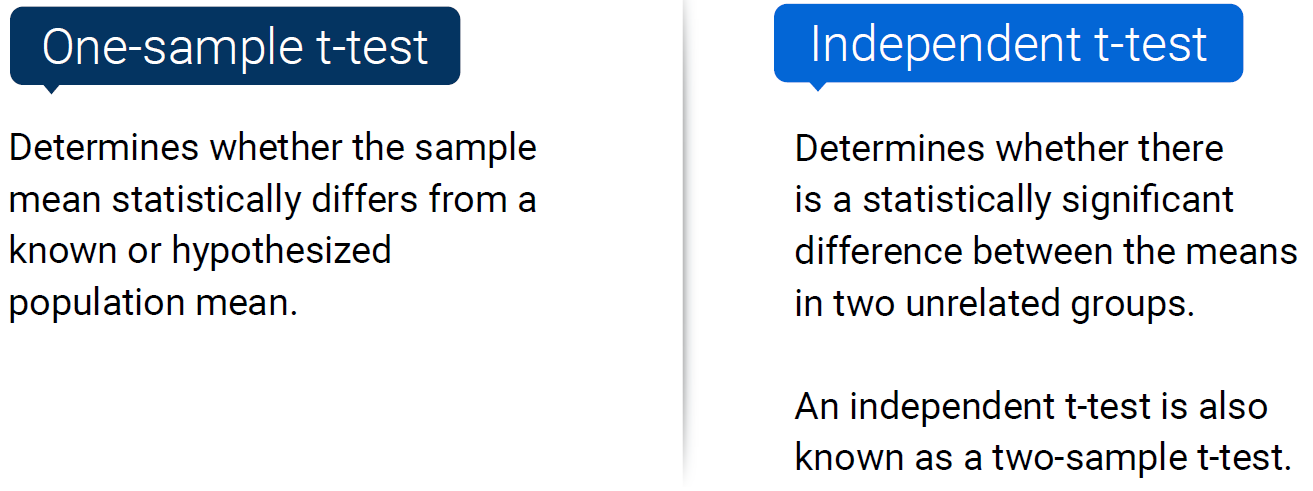


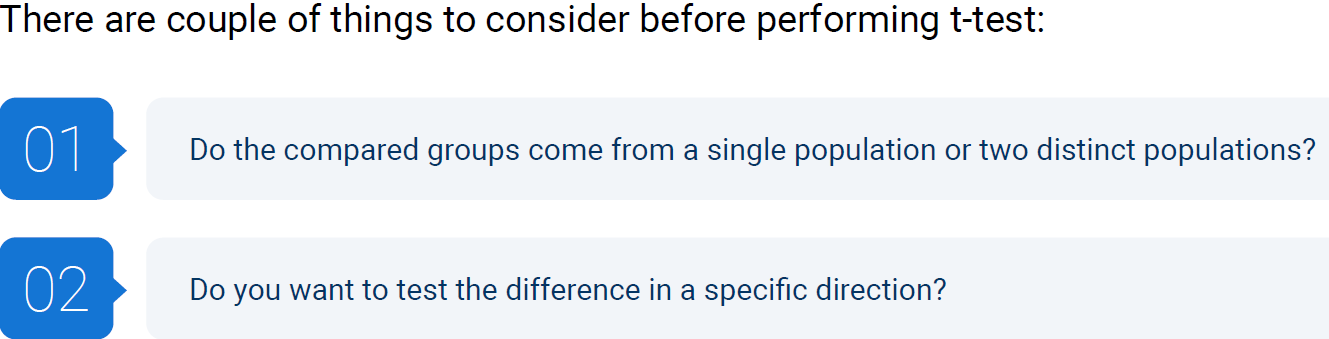
A t-test tells you how significant the differences between groups are.

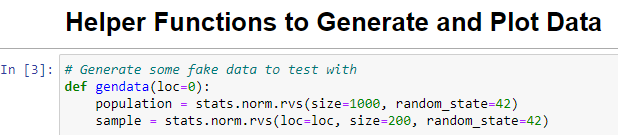
It lets you know if the differences, measured in means (averages), could have happened by chance.



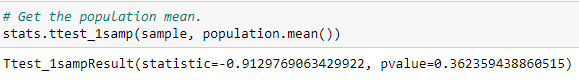
^ N1 and N2 are the number of records for each data set.







* Stats.norm generates a sample with a normal distribution (even, non-slanted bell curve)

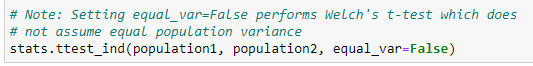




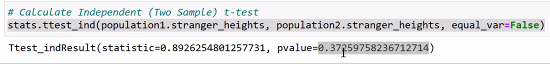
^ In example above, P-Value is less than 0.5, so the sample’s mean is close enough to the mean of the population to continue with confidence.

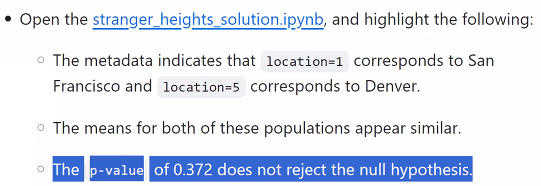
Goodness of fitness test:

* Did you prove that your data is a good fit for the model.
* It assumes that Data is:
  + Normally distributed
  + Independent
  + Homogenous (Standard deviations are roughly equal).



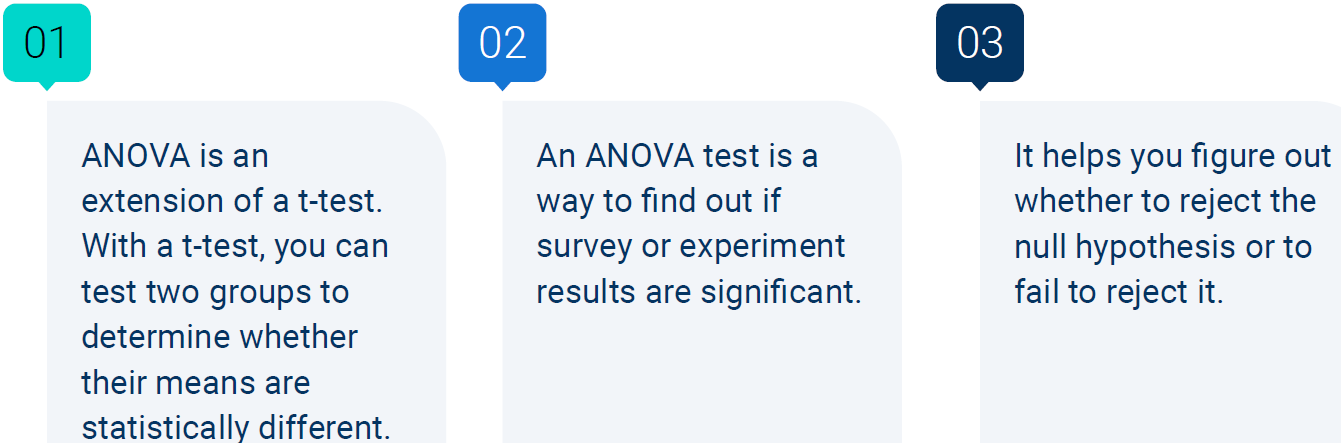
**Stranger Heights solution:**





0.372 is much greater than 0.05, therefore you cannot reject the null hypothesis

**ANOVA (Analysis of Variance):**



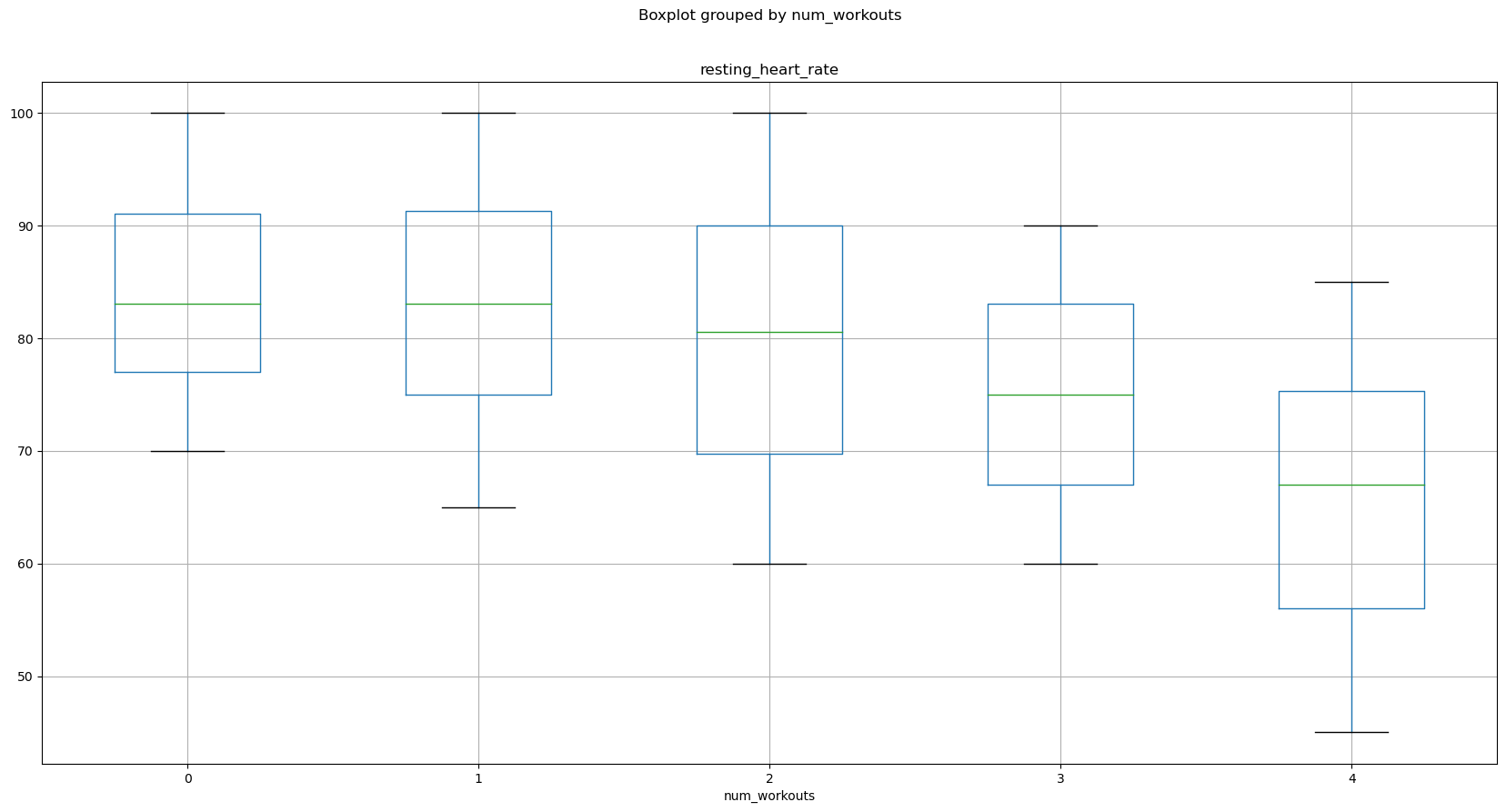
Any mean that is significantly different from the rest will result in low p-value.

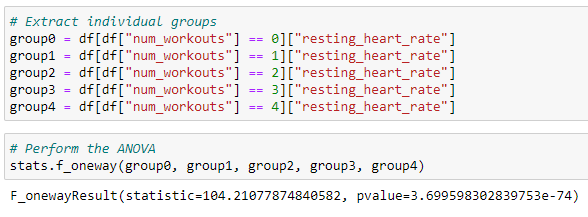
**Dataset below: resting\_heart\_rate.csv**

Source: Internally generated data.

Description: Comparison of people who work out during the week and average resting heart rate.

Number of Workouts per week: 0=zero workouts, 1=one 30 minute workout, 2=two 30 minute workouts, 3=three 30 minute workouts, 4=four or more 30 minute workouts





F\_oneway says that you want to do the test in one direction. We aren’t going in to bi-directional analysis

**Chi-Squared: (8:40pm)**

* 1. Determine Chi-squared value
  2. Determine Degree of freedom
     + Categories minus 1
  3. Choose P-value
  4. Determine critical value
     + Refer to: <https://www.medcalc.org/manual/chi-square-table.php> to get the number
     + Define it manually in the code
  5. Make a decision

