Khan Inan

Transcriptomics 7633

Exercise #2

M3_1.1) Let's say your goal is to determine the average size of an apple of a specific variety. You go to an apple farm and randomly pick several apples. What are some factors that may influence the size of the apple other than the variety of apple it is?

Soil quality, age of the apple and location of the tree(in regards to sunlight)

M3_1.2) Why is it helpful to know if the data is normally distributed?

Because when data is normally distributed it is easier to predict values in relation to other values

M3_2.1) What is standard deviation?

In simple terms standard deviation represents the spread or variation in values

M3_2.2) What is the standard error?

Standard error is the calculated accuracy of an estimate

M3_2.3) What is the confidence interval?

The chance that a population parameter will fall between two specific values

M3_2.4) What is a null hypothesis?

A null hypothesis is the notion that no statistical relationship exists between two variables

M3_2.5) What is a p-value?

The p-value is the calculated value of how likely you are to have found an observation if the null hypothesis is true

M3_2.6) What is a T-test and what assumptions have to be true for it to be used?

The T-test is the statistical test for finding if there is a difference between two groups and their means

M3_3.1) What do the rnorm, pnorm, dnorm, and qnorm provide?

Rnorm generates random deviates, pnorm gives the distribution function, dnorm is the density and qnorm is the quantile function

M3_4.1) Now that you have your apple size data, you run a t-test and get a p-value of 0.3. What is your null hypothesis? Would you accept or reject the null hypothesis?

My null hypothesis is that there is a relationship between apple size and something else besides apple variety. I would accept the null hypothesis with that p-value

M3_5.1) What is the difference between a paired and unpaired t-test?

Paired means that you are pairing two variables together, while un-paired means you are testing them seperately

M3_5.2) How can you test if the variance of the two samples are the same?

You would use an f-test to test if variables are the same

M3_6.1) What happens to the standard error as you reduce the sample size from 10 to 5.

The standard error would increase since there are less sample size

M3_7.1) What are the steps for the shuffle test?

The shuffle test is a test where the values are constantly shuffled to the point where the plot is a representation of the mean of a set of values