Charlee Cobb - Transcriptomics, Exercise 3

netID: cac9995

Charlee Cobb

2023-01-17

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? There are many factors. Some include: The soil the apple trees grow in, nutrients or pesticides added by the farmer, the time of when an apple is picked off the tree, and the location of the trees on the farm in terms of sunlight and wind impact.

M3_1.2) Why is it helpful to know if the data is normally distributed? It's helpful because you can understand the potential level of biases that will go into the data. We can also make assumptions on variation in the data.

M3_2.1) What is standard deviation? Standard deviation explains the distribution of the given data. This tells us the variability of the data.

M3_2.2) What is the standard error? This is derived from standard deviation by dividing the square root of number of samples. The smaller the standard error, the more confident you can be in that your data values are true.

M3_2.3) What is the confidence interval? We can use standard error to obtain the confidence interval. This can be used to set boundaries in data and compartmentalize the values to say where the estimated mean should appear.

M3_2.4) What is a null hypothesis? The null hypothesis is the thing you want to test based on the given data.

M3_2.5) What is a p-value? The p-value is used to test the null hypothesis against data. If the null hypothesis was true, the p-value indicates how many data observations validate the null hypothesis.

M3_2.6) What is a T-test and what assumptions have to be true for it to be used? A t-test determines if two variables are significantly different from each other. The assumptions include independence of the two variables, normal distribution, and random sampling.

M3_3.1) What do the rnorm, pnorm, dnorm, and qnorm provide? These refer to distributions. pnorm is probaility, qnorm is quantile, dnorm is density, and rnorm is random number generator.

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 apples have an average lenght of size of X metric. Because my p-value is 0.3, I will accept the null hypothesis because the cutoff to reject the null hypothesis is 0.05.

M3_5.1) What is the difference between a paired and unpaired t-test? Unpaired means the data is completely independent, and paired means the values are related.

M3_5.2) How can you test if the variance of the two samples are the same? F-test will tell you the variation between two samples that will indicate if they are the same or different samples.

M3_6.1) What happens to the standard error as you reduce the sample size from 10 to 5. The standard error will become larger because n is smaller.

M3_7.1) What are the steps for the shuffle test? Combine vectors of data and create a factor of association. Then it calculate the average, and repeat the steps of sampling and associating without replacing them.