Hypothesis tests, t-test review

Your Name Here

Tue Jan 19 16:07:50 2021

## Importing the data

Note that this comes after the summary command in the online instructions, but we will put it here for the sake of repeatability of the analysis:

## Summarizing a data set

The onesample data set has just a single variable, called Body.temperature. Using the summary() command on the onesample data set gives you a numerical summary of Body.temperature.

**Question: which of the reported summary statistics are measures of central tendency? (write your answer below the questions)**

## One sample t-tests

One-sample t-tests test hypotheses about the population mean for a single sample of data. We have data on body temperature, and we want to know if it is consistent with the typical value used for normal human body temperature (98.6 degrees F), or not.

**Question: what is the null hypothesis for a test of our data having a mean body temperature of 98.6 degrees F? Use symbolic notation (i.e. Ho: …)**

The one-sample t-test assumes normality. Test normality here:

**Question: do you pass the normality test? How do you know?**

Whenever you are analyzing data, it’s important to look at its distributed so that you can see whether the data is clearly not normally distributed, spot data entry errors, and look for “outliers”, which are data points that could have a large effect on the outcome of your analysis. First do a strip chart:

Then a boxplot:

**Question: are there any extreme values in your plots? How do you know?**

Now conduct the t-test:

**Question: how are the degrees of freedom calculated?**

**Question: what does the t-value tell you?**

**Question: do you reject or retain the null hypothsis? How do you know?**

**Question: what does your result tell you about mean body temperature? Do the results pertain to the sample mean, or the population mean?**

**Question: is there a homogeneity of variances assumption for a one-sample t-test? Why or why not?**

## Two-sample t-tests

We have data comparing heights of plants that were either self-pollinated, or pollinated with pollen from another individual. We wish to compare the mean heights between these two treatment groups.

**Question: what is the null hypothesis for this test?**

Import the data from worksheet “twosample” of file “review\_data.xls”:

Before analysis look at your data with a stripchart:

and a boxplot:

**Question: based on the boxplots, are the data skewed (and if so, are they positively or negatively skewed)?**

Test for normality:

**Question: do we pass normality for both groups? Why did we have to split the data into the two groups to test for normality?**

Test for HOV:

**Question: do we pass the HOV test? How do you know?**

Now do the t-test:

**Question: is the difference between the two group means statistically significant? How do you know?**

**Question: based on these results, does outcrossing increase plant height? How can you tell (is the p-value enough to know, or do you need to look at additional information)?**

## Paired t-test

You have data on CO2 uptake by a single set of plants exposed to two different ambient CO2 concentrations. Since the same plants are used for each treatment you need to use a paired t-test to account for the lack of independence of measurements for the same plant at each treatment level.

Import the data from sheet “paired” in the file “review\_data.xls”:

Stack the data and rename the columns so that you can plot the raw data values:

Look at the data using a strip chart:

and using a boxplot:

**Question: is there a clear separation in the data values between the groups, or is there a lot of overlap in uptake amounts between them?**

The raw data are good to look at, but for a paired t-test it is the differences between paired of values that are used. Calculate the differences:

Make a boxplot of the differences:

**Question: are the differences mostly positive or mostly negative? Or are they a roughly even mix of positive and negative?**

Check normality of the differences:

**Question: do you pass the normality test for differences between the pairs of measurements?**

Now do the paired t-test:

**Question: what do you conclude about the effect of ambient concentration on uptake of CO2? Is the difference statistically significant (how do you know)? And, how do you know if increased ambient CO2 increases uptake or decreases it?**

**Question: explain how it was possible for the two sets of data values that were so broadly overlapping to produce differences that were so clearly, consistently positive? How did using an analysis of the differences help detect the effect of ambient CO2 on uptake in this experiment?**