# Homework 1: Intro to T-Tests, Paired, and Unpaired data

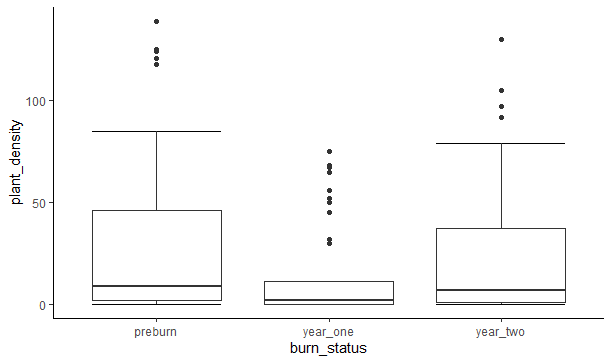
**Names: Malachi Miller**

**Grade: \_\_\_ / 10**

**Question set 1: Badland’s Fire Ecology**

**About the data:** A controlled or prescribed burn, also known as hazard reduction burning, is a wildfire set intentionally for purposes of forest management, farming, and prairie restoration. These data are individual abundance measures of certain guilds of plant life (shrubs, forbs, grasses, and the like) from Badlands National Park before, one year after, and two years after a controlled burn in a block design. ***Important note: the data are paired!***

1. Create a figure that properly demonstrate the difference between each time period (hint: use the *gather( )* function to turn data from wide to long)



1. Based on your answer above, what are your initial thoughts on the relationship between the *Pre-burn* density and the years following a controlled burn?

Based on the data, we can see that the pre-burn density is very high, the density 1 year after the burn is relatively low, and that the density increases quite a bit by year 2.

1. Do a proper statistical analysis comparing the abundance measures of *Pre-burn­* vs *1 Year Post-Burn* and *Pre-burn­* vs *2 Year Post-Burn*. Be sure to report the estimated ***difference in means***, ***95% confidence interval*** for that difference, the ***interpretation of that 95% C.I***., the ***p-value with interpretation*** for each test.

Pre-burn vs. 1 year

data: badlands$preburn and badlands$year\_one

t = 3.8695, df = 41, p-value = 0.0003832

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

7.854266 25.002877

sample estimates:

mean of the differences

16.42857

Based on this data, we have 95% confidence that the difference in means is between 7.854 and 25.003. The p-value is less than 0.05, so we can conclude that there is a meaningful difference between the two.

Pre-burn vs. 2 Year

data: badlands$preburn and badlands$year\_two

t = 2.0671, df = 41, p-value = 0.04507

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

0.1496336 12.8503664

sample estimates:

mean of the differences

6.5

Based on this data, we have 95% confidence that the difference in means is between 0.150 and 12.85. The p-value is less that .05, so we can conclude that there is a meaningful difference between the pre-burn and year 2 plant density.

1. Let’s say that we want to round the p-values to the second decimal place (i.e. 0.035 🡪 0.04). Does this change your interpretation of your analysis?

Yes. If we round the p-values to the second decimal place, then we can conclude that there is no significant difference between the pre-burn and the year 2 plant densities.

1. Lastly, complete a 1-sample t-test version of your above analysis. Only report any difference you see between the output from question 1C and this output (hint: create new variables based on the difference between *post-burn(s)* and *pre-burn*, subtraction is your friend here)

Year 1 Difference

One Sample t-test

data: badlands$diff\_yr1

t = -3.8695, df = 41, p-value = 0.0003832

alternative hypothesis: true mean is not equal to 0

95 percent confidence interval:

-25.002877 -7.854266

sample estimates:

mean of x

-16.42857

Year 2 Difference

One Sample t-test

data: badlands$diff\_yr2

t = -2.0671, df = 41, p-value = 0.04507

alternative hypothesis: true mean is not equal to 0

95 percent confidence interval:

-12.8503664 -0.1496336

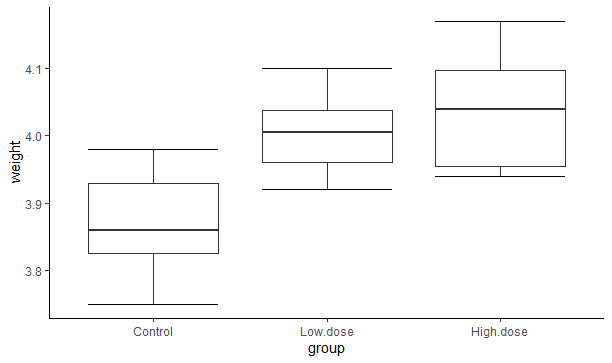
sample estimates:

mean of x

-6.5

**Question set 2: Chicken Weight**

**About the data:** The weight of chickens has increased greatly within the last century, mainly due to changes in nutrition and pharmaceuticals. The purpose of this study was to investigate a drug added to the feed of chicks in an attempt to promote growth. Chickens were divided into *dosage (g)* and *weighed in pounds* (average per block. Read: each block represents some number of chickens, rather than just one. Also, data are **unpaired**).

1. Create a figure that properly demonstrate the difference between each dosage (hint: use the *gather( )* function to turn data from wide to long)
2. Based on your answer above, what are your initial thoughts on the relationship between the control and the two dosage types?

The control chickens had a lower weight, while thoe chickens that received doses had a higher weight in the end. Those chickens that had higher doses ended up with higher average weights.

1. Do a proper statistical analysis comparing the abundance measures of *Control* vs *Low Dose* and *Control* vs *High Dose*. Be sure to report the estimated ***difference in means***, ***95% confidence interval*** for that difference, the ***interpretation of that 95% C.I***., the ***p-value with interpretation*** for each test.

Control vs. Low Dose

data: chicky$Control and chicky$Low.dose

t = -4.567, df = 7, p-value = 0.002583

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

-0.2105891 -0.0669109

sample estimates:

mean of the differences

-0.13875

Based on the data, we have 95% confidence that the difference in means is between 0.21 and 0.06. Because the p-value is less than .05, we can conclude that there is a statistically significant difference between the control and the low dose

Control vs. High Dose

data: chicky$Control and chicky$High.dose

t = -4.8969, df = 7, p-value = 0.001759

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

-0.25394396 -0.08855604

sample estimates:

mean of the differences

-0.17125

Based on the data, we can say that we have 95% confidence that the difference in means is between 0.25 and 0.08. Because the p-value is less than .05, we can conclude that there is a statistically significant difference between the control and the high dose.

1. Let’s say that we have a hunch that the weight on the control group was going to be *less than* the two other groups *before* doing our analysis. Run a *one-sided t-test* with respect to our *alternative* hypothesis (hint: alternative = “less”). Report the relationship between the p-values from these tests and the previous two-sided tests.

**Question set 3: Concepts**

1. What are the assumptions for t-tests?

The assumptions for t-tests are that they are normally distributed, they have a large enough sample size, and that there is homogeneity of variance.

1. How does changing the confidence level change the confidence interval (i.e. how is a 95% C.I. different from a 99% C.I. and 80% C.I.)? Feel free to try this on either t-test above (in the t.test function, change the ‘conf.level’ argument to ‘conf.level = 0.99’ etc etc). If you decide to try this on the t-tests above, there’s no need to report the actual interval, just how it behaves as you change the confidence level.

As the confidence level goes up, the values get broader so that we can have confidence that the mean is actually between the two.

1. Let’s say you were looking at the *difference* between two samples of *paired* data. You then calculate the 95% confidence interval to be (-1.45, 2.54). Will the p-value be greater than or less than 0.05? Explain your reasoning.

It will be less than .05 because there is a good chance that there will be a significant difference between the two

**Please answer the following questions for your homework to be graded:**

1. How did you and your partner(s) split the work on this assignment?

We met and each worked on it individually and asked each other for help if we didn’t understand something.

1. What did you find difficult in this assignment?

The hardest part for me was running the different t-tests and interpreting them.