

# Lab 10

April 7-8, 2020

## Objectives:

1. Review one-sample continuous data using a t-test
2. Perform a paired t-test
3. Practice power and sample size calculations in R
4. Perform Fisher's Exact test and Chi-squared test

## Load data

Today we will be using three datasets:

The first dataset has various lipid levels of the 3026 adults in a study.

The second dataset includes weights of young female anorexia patients before and after treatment (paired data). You must install the package 'PairedData' to access the dataset.

The last dataset (lister) compares the survival of surgery patients and whether sterilization practices were used in 1867.

```
install.packages("PairedData")

#Part 1
lipids <- read.delim("http://myweb.uiowa.edu/pbreheny/data/lipids.txt")

#Part 2 & 3
library(PairedData)
data("Anorexia")
data("Corn")

#Part 4
lister <- read.delim("http://myweb.uiowa.edu/pbreheny/data/lister.txt")
```

## Part 1: One-sample continuous data and review of the t.test function

In the lipids dataset, the variable TRG is continuous. Suppose the national average for TRG is said to be 120. Is our data significantly different than the national average?

We can do this in R using the t.test() function as shown below. For one-sample continuous data insert the continuous observations into the first argument. The default tests  $\hat{\mu}$  against  $\mu=0$ . We want to test whether  $\hat{\mu}$  is equal to 120 in this case. Therefore, you must specify your null hypothesis by saying  $\mu=120$ .

```
t.test(lipids$TRG, mu = 120)
```

*Interpret your results. Compare the confidence interval to the  $\hat{\mu} = 120$ . The upper bound of the confidence interval is barely less than the null. Why do you think the p-value is still fairly low?*

## Part 2: Paired t-test

If we have data that is matched or paired and we want to determine if there is a difference between the two groups, we can use a paired t-test.

**For the anorexia dataset, we want to determine whether the patient’s weight improved post-treatment compared to pre-treatment.**

First, we want to calculate the difference in weight for each patient.

```
#create a new column  
Anorexia$diff<- Anorexia$Post-Anorexia$Prior
```

*Take a look at the new column “diff”. Are most of the values positive or negative? What does this indicate?*

Now calculate the standard error:

```
SE<- sd(Anorexia$diff)/sqrt(17)
```

Find the t-statistic. Note that the numerator is simply the mean of the differences calculated.

```
t<- mean(Anorexia$diff)/SE  
  
2*(1-pt(t, df=16))
```

```
## [1] 0.0007002531
```

A “paired t-test” does the same thing as the t-test above however it does it on the difference between the paired observations. To run a paired t-test in R we will need to use the code below:

```
t.test(Anorexia$Post, Anorexia$Prior, paired = TRUE)
```

The `t.test()` function has other arguments. We have shown you the `paired = TRUE` input for when you want to run a paired t-test. If you do not put `paired = TRUE` for a paired t-test it will assume that the two sets of numbers are two separate samples and will run a two-sample t-test. There is also the alternative = “two.sided” or “less”, or “greater” in that it will change the alternative hypothesis to find the p-value you are searching for. We can also change the `conf.level = .95` to other values to get confidence intervals other than a 95% CI.

## Practice Problem: Paired T-test

15 types of corn are measured in the “Corn” dataset and their heights are recorded. We are interested in whether the heights of corn are higher when they were produced by cross-fertilization instead of self-fertilization. Conduct a paired t-test in R to find if there is a significant difference in the effectiveness of these two methods.

## Part 3: Power and sample size calculations in R

In R there is a `power.t.test()` function that is useful. It takes the parameters of delta (expected difference in means), sd, n, and power. Let’s say we want to run a more powerful study using the anorexia dataset.

**How many patients would we need in our NEW study to achieve a power of .99?**

When using the `power.t.test()` function, the parameter you do not specify will be the parameter the function will calculate. But you must specify 3 of the 4 parameters (difference, sd, n, power).

In this example we would use the following code to find the **sample size**:

```
#n is not specified
power.t.test(delta = mean(Anorexia$Post-Anorexia$Prior), power = .99,
             sd = sd(Anorexia$Post-Anorexia$Prior), type = "paired")
```

The results give us the estimated number of pairs needed to achieve a power of at least .99 (given the parameters are accurate). We need at least 20 pairs.

*Why do you think it only takes 20 pairs to achieve a power of 0.99?*

### Find power

This function can also be used to find power. So lets say we have an n value of 50 in this study what would the power be? Well we can use the function to find:

```
power.t.test(delta = mean(Anorexia$Post-Anorexia$Prior), n = 50,
             sd = sd(Anorexia$Post-Anorexia$Prior), type = "paired")
```

## Part 4: Fisher's Exact test and Chi-squared test

In class you have learned how to do calculate the chi-squared statistic by hand. In R we can perform a Fisher's Exact and Chi-squared test quickly by putting the data into a table and running the `chisq.test()` function or the `fisher.test()` function.

For this example create a table from the `lister` dataset:

```
tab <- table(lister)
tab
```

```
##           Outcome
## Group      Died Survived
## Control    16      19
## Sterile     6       34
```

The table above saved as "tab" can be placed directly into the functions to receive the results of a chi-squared test and fisher's exact test. **NOTE:** if you do not put `correct=FALSE` the statistic will be different from the calculation you do by hand.

```
chisq.test(tab, correct = FALSE)
```

```
##
## Pearson's Chi-squared test
##
## data:  tab
## X-squared = 8.4952, df = 1, p-value = 0.003561
```

```
fisher.test(tab)
```

```
##
## Fisher's Exact Test for Count Data
##
## data:  tab
## p-value = 0.005018
## alternative hypothesis: true odds ratio is not equal to 1
## 95 percent confidence interval:
##  1.437621 17.166416
## sample estimates:
## odds ratio
```

```
## 4.666849
```

Looking at these two outputs we can see that the chi-squared test gives the test statistic, DF, and p-value where as the fisher's exact test gives us the p-value, CI for the odds ratio, and the odds ratio. The odds ratio is something we will be talking about within the next two weeks.

### Practice problem:

a) (By hand) Consider a problem in which we observe students to see if they took notes by hand or by laptop and recorded whether they passed or failed. Of the 67 that used a laptop 44 passed. Of the 76 that wrote by hand 56 of them passed. Construct an observed and expected table. Compute the chi-squared statistic and find the p-value from the chi-squared table. Interpret your results.

b) (Using R) Do this same problem using R. To start, use the following code:

```
chidata <- data.frame("Method" = c(rep("Laptop", 67), rep("Hand", 76)), "Result" = c(rep("Passed", 44),  
table(chidata)
```

```
##           Result  
## Method   Failed Passed  
##   Hand         20    56  
##   Laptop        23    44
```