Lab Assignment: Two-Sample t-Procedures

Paul Isaacson

Feb. 2, 2017

## Exercise 1

An experiment was conducted to evaluate the effectiveness of a drug treatment for tapeworm in the stomachs of sheep. A random sample of 24 worm-infected lambs of approximately the same age was divided into two groups. Twelve of the lambs were injected with the drug and the remaining twelve were left untreated. After a 6-month period the lambs were slaughtered and the worm counts recorded.

### Part 1a

Load the dataset WormSheep from the DS705data package. Note that the package is loaded above at line 18. You just need the data() command. To see all the datasets in the package, type data(package='DS705data').

### -|-|-|-|-|-|-|-|-|-|-|- Answer 1a -|-|-|-|-|-|-|-|-|-|-|-

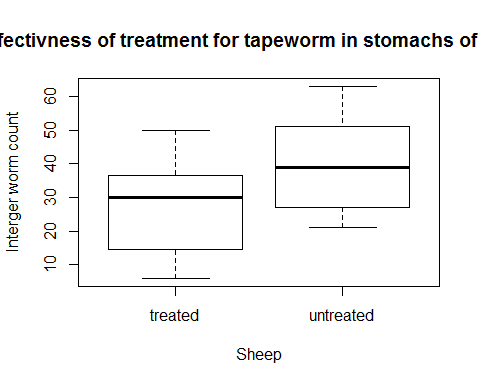
data(WormSheep)

### Part 1b

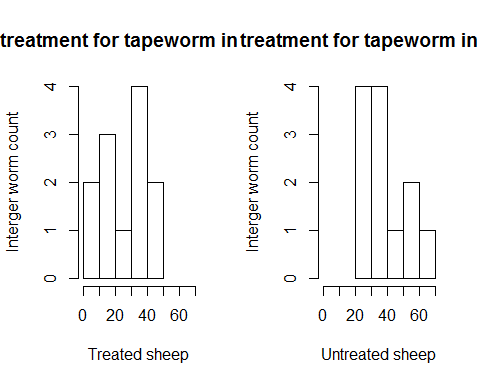
Create boxplots and histograms for each group (treated vs. untreated). Be sure that each plot is labeled appropriately.

### -|-|-|-|-|-|-|-|-|-|-|- Answer 1b -|-|-|-|-|-|-|-|-|-|-|-

boxplot(worms ~ treated, data=WormSheep, xlab='Sheep', ylab='Interger worm count', main='Effectivness of treatment for tapeworm in stomachs of sheep')



par(mfrow=c(1,2))  
with(WormSheep, hist(worms[treated=='treated'], xlab='Treated sheep', ylab='Interger worm count', main='Effectivness of treatment for tapeworm in stomachs of sheep', xlim=c(0,70)))  
with(WormSheep, hist(worms[treated=='untreated'], xlab='Untreated sheep', ylab='Interger worm count', main='Effectivness of treatment for tapeworm in stomachs of sheep', xlim=c(0,70)))



### Part 1c

Do the boxplots show any outliers?

### -|-|-|-|-|-|-|-|-|-|-|- Answer 1c -|-|-|-|-|-|-|-|-|-|-|-

No.

### Part 1d

Describe the shapes of the histograms for the sample data for each sample.

### -|-|-|-|-|-|-|-|-|-|-|- Answer 1d -|-|-|-|-|-|-|-|-|-|-|-

The shape of each has a peak in the middle.

### Part 1e

Conduct an appropriate test to determine if the worm counts in each population can be considered as normally distributed. Provide the p-value and the conclusion of the test.

### -|-|-|-|-|-|-|-|-|-|-|- Answer 1e -|-|-|-|-|-|-|-|-|-|-|-

with(WormSheep, shapiro.test(worms[treated=='treated']))

##   
## Shapiro-Wilk normality test  
##   
## data: worms[treated == "treated"]  
## W = 0.95092, p-value = 0.6504

with(WormSheep, shapiro.test(worms[treated=='untreated']))

##   
## Shapiro-Wilk normality test  
##   
## data: worms[treated == "untreated"]  
## W = 0.94382, p-value = 0.5491

At a .05 significance level, it is plausible that both the treated (p-value 0.6504) and the untreated (p-value 0.5491) came from a normal distribution.

### Part 1f

Conduct an appropriate test to determine if the worm counts in each population can be considered to have equal variances. Provide the p-value and the conclusion of the test.

### -|-|-|-|-|-|-|-|-|-|-|- Answer 1f -|-|-|-|-|-|-|-|-|-|-|-

library(car)  
leveneTest(worms~treated, data=WormSheep)

## Levene's Test for Homogeneity of Variance (center = median)  
## Df F value Pr(>F)  
## group 1 0.1415 0.7104  
## 22

At .05 significance level, we fail to reject the null hypothesis (p-value 0.7104). In other words, it is entirely plausible that the variables are equal.

### Part 1g

Conduct the test of your choice to compare the population mean worm count for all sheep treated with the drug to the mean worm count for the population of untreated sheep. Let .

#### Step 1

Define the parameters in words in the context of the problem.

#### -|-|-|-|-|-|-|-|-|-|-|- Answer 1g.step1 -|-|-|-|-|-|-|-|-|-|-|-

Let u1 represent the population mean worms per sheep that are is sheep with a treatment. Let u2 represent the population mean worms per sheep that are is sheep without a treatment.

#### Step 2

State the null and alternative hypotheses for the test.

#### -|-|-|-|-|-|-|-|-|-|-|- Answer 1g.step2 -|-|-|-|-|-|-|-|-|-|-|-

#### Step 3

Use R to generate the output for the test you selected.

#### -|-|-|-|-|-|-|-|-|-|-|- Answer 1g.step3 -|-|-|-|-|-|-|-|-|-|-|-

t.test(worms~treated, data=WormSheep, var.equal=TRUE)

##   
## Two Sample t-test  
##   
## data: worms by treated  
## t = -2.2709, df = 22, p-value = 0.03329  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## -25.031761 -1.134906  
## sample estimates:  
## mean in group treated mean in group untreated   
## 26.58333 39.66667

#### Step 4

State both a statistical conclusion at and interpret it in the context of the problem.

#### -|-|-|-|-|-|-|-|-|-|-|- Answer 1g.step4 -|-|-|-|-|-|-|-|-|-|-|-

At a .05 significance level, we are able to reject the null hypothesis. That is to say, there is sufficient evidence to claim that the population mean of worms in sheep for the treated group is different than the population mean of worms in sheep for the untreated group.

### Part 1h

Write an interpretation in the context of the problem for the 95% CI.

### -|-|-|-|-|-|-|-|-|-|-|- Answer 1h -|-|-|-|-|-|-|-|-|-|-|-

We are 95% certain that the population mean of worms in sheep for the treated group is between 25.032642 and 1.134025 less than the population mean of worms in sheep for the untreated group.

### Part 1i

Did you use the separate-variance t-procedures or the pooled t-procedures? Justify your choice, including some discussion of how well the conditions for the hypothesis test and confidence interval procedures were met.

### -|-|-|-|-|-|-|-|-|-|-|- Answer 1i -|-|-|-|-|-|-|-|-|-|-|-

I used the pooled t-procedure. This is due to the fact that the results of the levene test showed that the variances of each group was equal.

## Exercise 2

Data was collected for a sample of college students at a university in the Midwest. One variable measured was the number of words per minute that they could type and also whether or not they had previously taken a typing course (Method 1) or if they were self-taught (Method 2).

### Part 2a

Load the dataset Typing from the DS705data package.

### -|-|-|-|-|-|-|-|-|-|-|- Answer 2a -|-|-|-|-|-|-|-|-|-|-|-

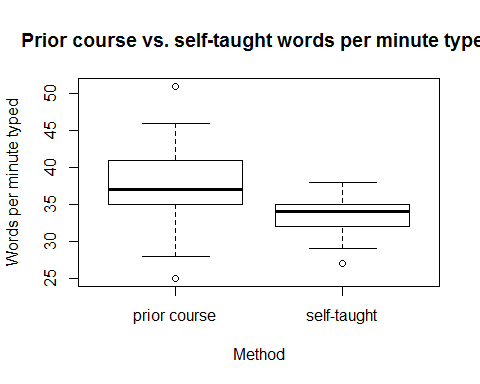
data(Typing)

### Part 2b

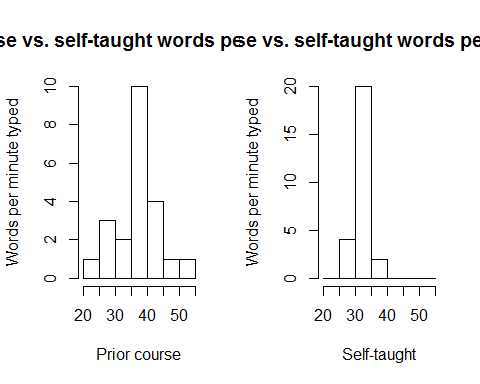
Create boxplots and histograms for each group (previous typing class vs. self-taught). Be sure that each plot is labeled appropriately.

### -|-|-|-|-|-|-|-|-|-|-|- Answer 2b -|-|-|-|-|-|-|-|-|-|-|-

boxplot(Words ~ Method, data=Typing, xlab='Method', ylab='Words per minute typed', main='Prior course vs. self-taught words per minute typed')



par(mfrow=c(1,2))  
with(Typing, hist(Words[Method=='prior course'], xlab='Prior course', ylab='Words per minute typed', main='Prior course vs. self-taught words per minute typed', xlim=c(20,55), breaks=c(20,25,30,35,40,45,50,55)))  
with(Typing, hist(Words[Method=='self-taught'], xlab='Self-taught', ylab='Words per minute typed', main='Prior course vs. self-taught words per minute typed', xlim=c(20,55), breaks=c(20,25,30,35,40,45,50,55)))



### Part 2c

Do the boxplots show any outliers?

### -|-|-|-|-|-|-|-|-|-|-|- Answer 2c -|-|-|-|-|-|-|-|-|-|-|-

Yes, both have outliers.

### Part 2d

Describe the shapes of the histograms for the sample data for each sample.

### -|-|-|-|-|-|-|-|-|-|-|- Answer 2d -|-|-|-|-|-|-|-|-|-|-|-

The histogram of the students who had a prior course appears to be more normally distributed. There is greater variation in the words per minute typed. However, the self taught students appear to be centered more closely to one mean. This mean does appear to be lower than the students with a prior course.

### Part 2e

Conduct an appropriate test to determine if typing speed in each population can be considered as normally distributed. Provide the p-value and the conclusion of the test.

### -|-|-|-|-|-|-|-|-|-|-|- Answer 2e -|-|-|-|-|-|-|-|-|-|-|-

with(Typing, shapiro.test(Words[Method=='prior course']))

##   
## Shapiro-Wilk normality test  
##   
## data: Words[Method == "prior course"]  
## W = 0.96676, p-value = 0.6362

with(Typing, shapiro.test(Words[Method=='self-taught']))

##   
## Shapiro-Wilk normality test  
##   
## data: Words[Method == "self-taught"]  
## W = 0.93509, p-value = 0.1025

It appears as though the students with a prior course is from a normal distribution (p-value 0.6362). It is less clear whether the students who were self-taught is from a normal distribution (p-value 0.1025) However, I am willing to accept that it was using a significance level of 0.5.

### Part 2f

Conduct an appropriate test to determine if typing speed in each population can be considered to have equal variances. Provide the p-value and the conclusion of the test.

### -|-|-|-|-|-|-|-|-|-|-|- Answer 2f -|-|-|-|-|-|-|-|-|-|-|-

with(Typing, leveneTest(Words~Method))

## Levene's Test for Homogeneity of Variance (center = median)  
## Df F value Pr(>F)   
## group 1 9.0646 0.004222 \*\*  
## 46   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

The variances are not equal (p-value 0.004222)

### Part 2g

Conduct the test of your choice to test that the population mean typing speed for all college students with a previous course in typing is higher than for those who were self-taught. Let .

#### Step 1

Define the parameters in words in the context of the problem.

#### -|-|-|-|-|-|-|-|-|-|-|- Answer 2g.step1 -|-|-|-|-|-|-|-|-|-|-|-

Let u1 represent the population mean words per minute typed for students with a prior course. Let u2 represent the population mean words per minute typed for students who are self-taught.

#### Step 2

State the null and alternative hypotheses for the test.

#### -|-|-|-|-|-|-|-|-|-|-|- Answer 2g.step2 -|-|-|-|-|-|-|-|-|-|-|-

#### Step 3

Use R to generate the output for the test you selected.

#### -|-|-|-|-|-|-|-|-|-|-|- Answer 2g.step3 -|-|-|-|-|-|-|-|-|-|-|-

t.test(Words~Method, data=Typing)

##   
## Welch Two Sample t-test  
##   
## data: Words by Method  
## t = 2.6505, df = 26.876, p-value = 0.0133  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## 0.8333886 6.5512267  
## sample estimates:  
## mean in group prior course mean in group self-taught   
## 37.00000 33.30769

#### Step 4

State both a statistical conclusion at and interpret it in the context of the problem.

#### -|-|-|-|-|-|-|-|-|-|-|- Answer 2g.step4 -|-|-|-|-|-|-|-|-|-|-|-

At a .05 significance level, we are able to reject the null hypothesis. That is to say, the population mean of all students words per minute typed for prior course students is different than the population mean of all students words per minute typed for self-taught students.

### Part 2h

Write an interpretation in the context of the problem for a 90% confidence interval.

### -|-|-|-|-|-|-|-|-|-|-|- Answer 2h -|-|-|-|-|-|-|-|-|-|-|-

We are 90% confident that the mean of the population of students who took a prior course is between 0.8333886 and 6.5512267 words per minute higher than the population of students who took are self-taught.

### Part 2i

Did you use the separate-variance t-procedures or the pooled t-procedures? Justify your choice, including some discussion of how well the conditions for the hypothesis test and confidence interval procedures were met.

### -|-|-|-|-|-|-|-|-|-|-|- Answer 2i -|-|-|-|-|-|-|-|-|-|-|-

I used the separate-variance t-procedure. This was due to the fact that the levene test showed the variances of the samples are not equal (p-value 0.004222). Homogeneity of variance is a requirement in order to use the pooled t-procedure.