Lab Assignment: Two-Sample t-Procedures

Spencer Swartz

Feb. 7, 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(package='DS705data')  
data(WormSheep)  
head(WormSheep)

## worms treated  
## 1 18 treated  
## 2 43 treated  
## 3 28 treated  
## 4 50 treated  
## 5 16 treated  
## 6 32 treated

WormSheep$treated

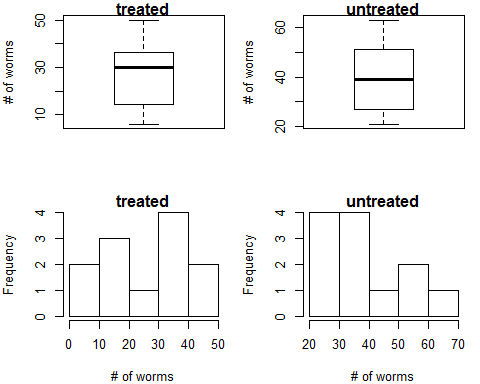
## [1] treated treated treated treated treated treated treated   
## [8] treated treated treated treated treated untreated untreated  
## [15] untreated untreated untreated untreated untreated untreated untreated  
## [22] untreated untreated untreated  
## Levels: treated untreated

### Part 1b

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

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

par(mfrow=c(2,2),mar=c(4,4,1,1))  
boxplot(WormSheep$worms[WormSheep$treated=="treated"],ylab="# of worms", main="treated")  
boxplot(WormSheep$worms[WormSheep$treated=="untreated"],ylab="# of worms", main="untreated")  
hist(WormSheep$worms[WormSheep$treated=="treated"],xlab="# of worms", main="treated")  
hist(WormSheep$worms[WormSheep$treated=="untreated"],xlab="# of worms", main="untreated")



# reset to one plot  
par(mfrow=c(1,1))

### Part 1c

Do the boxplots show any outliers?

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

There does not seem to be any outliers in the data based on these charts.

### Part 1d

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

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

Neither Box plots seem to be normally distributed as the treated seem to be low in the what would be the 50 %tile, and the untreated seem to be skewed to the left.

### 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 -|-|-|-|-|-|-|-|-|-|-|-

treated <- WormSheep$worms[WormSheep$treated=="treated"]  
untreated <- WormSheep$worms[WormSheep$treated=="untreated"]  
shapiro.test(treated)

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

shapiro.test(untreated)

##   
## Shapiro-Wilk normality test  
##   
## data: untreated  
## W = 0.94382, p-value = 0.5491

It is not unreasonable to assume that both the treated and the untreated samples come from a normally distributed population with there p-values being 0.65 and 0.54 respectively.

### 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)

## Warning: package 'car' was built under R version 3.3.2

leveneTest(WormSheep$worms~WormSheep$treated)

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

It is not unreasonable to believe that the population variance is equal (p-value=0.71).

### 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 represent the population mean worms per lamb that were in all treated sheep.

Let represent the population mean worms per lamb that were in all untreated sheep.

#### 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(WormSheep$worms~WormSheep$treated, var.equal = TRUE)

##   
## Two Sample t-test  
##   
## data: WormSheep$worms by WormSheep$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 -|-|-|-|-|-|-|-|-|-|-|-

Statistical conclusion: Reject at

Interpretation: There is sufficient evidence to claim that the population mean worms per lamb in all treated sheep is different form the population mean worms per lamb in all untreated sheep (P = 0.03329).

### Part 1h

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

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

With 95% confidence, the population mean worms per lamb in all treated sheep is 1.13 to 25.03 less than for all untreated sheep.

### 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-variance t-procedures as the variance of the two groups seemed to be equal and also seemed to be normally distributed.

## 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)  
head(Typing)

## Method Words  
## 1 prior course 28  
## 2 prior course 46  
## 3 prior course 36  
## 4 prior course 36  
## 5 prior course 28  
## 6 prior course 36

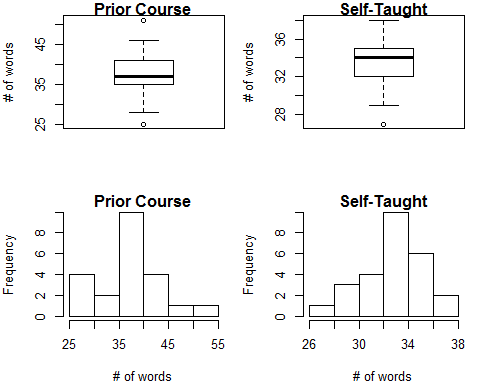
#Typing$Method

### 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 -|-|-|-|-|-|-|-|-|-|-|-

prior <- Typing$Words[Typing$Method=="prior course"]  
self <- Typing$Words[Typing$Method=="self-taught"]  
par(mfrow=c(2,2),mar=c(4,4,1,1))  
boxplot(prior,ylab="# of words", main="Prior Course")  
boxplot(self,ylab="# of words", main="Self-Taught")  
hist(prior,xlab="# of words", main="Prior Course")  
hist(self,xlab="# of words", main="Self-Taught")



# reset to one plot  
par(mfrow=c(1,1))

### Part 2c

Do the boxplots show any outliers?

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

Yes both samples have outliers shown.

### Part 2d

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

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

The Prior Course sample looks to be skewed to the left, and the Self-Taught sample looks as it may be normally distributed or maybe has a skew to the right.

### 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 -|-|-|-|-|-|-|-|-|-|-|-

shapiro.test(prior)

##   
## Shapiro-Wilk normality test  
##   
## data: prior  
## W = 0.96676, p-value = 0.6362

shapiro.test(self)

##   
## Shapiro-Wilk normality test  
##   
## data: self  
## W = 0.93509, p-value = 0.1025

It is not unreasonable to assume that both the prior course and the self taught samples come from a normally distributed population with there p-values being 0.63 and 0.10 respectively. ( )

### 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 -|-|-|-|-|-|-|-|-|-|-|-

leveneTest(Typing$Words~Typing$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

We can reject and can assume that the Population variances are not equal between the samples. (P = .0042)

### 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 represent the population mean words per min. typed by all who had a prior course.

Let represent the population mean words per min. typed by all who were 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 -|-|-|-|-|-|-|-|-|-|-|-

wilcox.test(Words~Method,data = Typing, conf.int=TRUE, conf.level=0.90)

## Warning in wilcox.test.default(x = c(28L, 46L, 36L, 36L, 28L, 36L, 36L, :  
## cannot compute exact p-value with ties

## Warning in wilcox.test.default(x = c(28L, 46L, 36L, 36L, 28L, 36L, 36L, :  
## cannot compute exact confidence intervals with ties

##   
## Wilcoxon rank sum test with continuity correction  
##   
## data: Words by Method  
## W = 434, p-value = 0.002181  
## alternative hypothesis: true location shift is not equal to 0  
## 90 percent confidence interval:  
## 1.999977 5.999975  
## sample estimates:  
## difference in location   
## 3.999964

#### Step 4

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

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

Statistical conclusion: Reject at

Interpretation: There is sufficient evidence to claim that the population mean words per min. in all students that took a prior course is different form the population mean words per min. in all students who were self taught (P = 0.002181).

### Part 2h

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

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

With 90% confidence, the population mean words per min. in all students that took a prior course is 2 to 6 more than for all students who were 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-procedures as the variance of the two groups seemed to be different. both populations looked to be normally distributed.