Two-Sample Inference Procedures for Population Central Values

Matt Allen

06/12/2018

## 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 randomly 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 data sets in the package, type data(package=‘DS705data’).

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

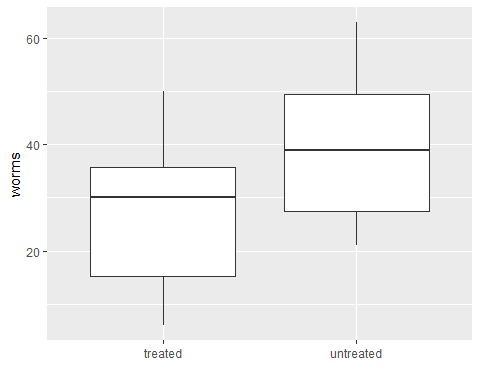
# Load the dataset WormSheep  
data("WormSheep")

### Part 1b

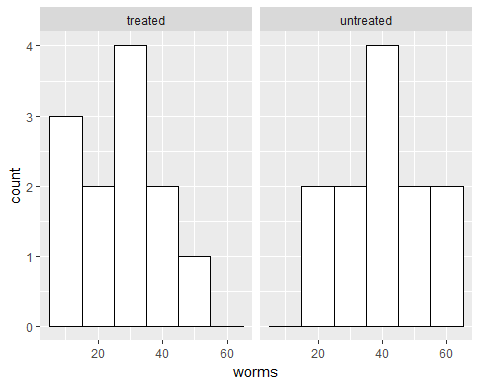
Create boxplots and histograms for each group (treated vs. untreated). Be sure that each plot is labeled appropriately.

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

# Create Boxplots and Histograms for each group treated vs. untreated.  
ggplot(data = WormSheep, aes(x = treated, y = worms)) + geom\_boxplot() + labs(x = "")



ggplot(WormSheep, aes(x=worms)) +   
 geom\_histogram(binwidth=10, colour="black", fill="white") +   
 facet\_grid(. ~ treated)



### Part 1c

Do the boxplots show any outliers?

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

Outliers are represented by dots above and below the whiskers. The boxplots do not show any outliers.

### Part 1d

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

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

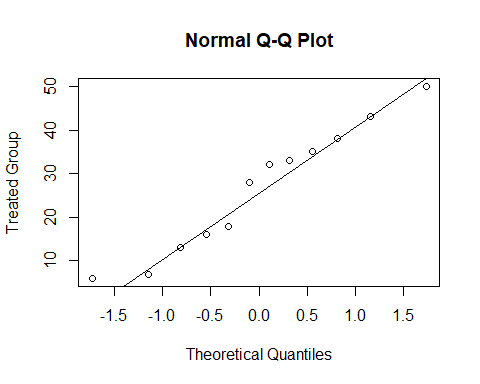
It is difficult to describe the shapes, but there does appear to be some mounding between 25 and 35 for the treated group and between 35 and 45 for the Untreated group.

### Part 1e

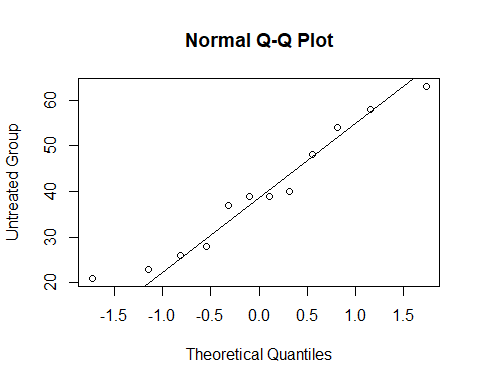
Do either of the normal probability plots show evidence of non-normality? Explain.

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

# Make normal probability plot of treated group and add fit line.  
  
treatedForWorms <- WormSheep[WormSheep$treated=="treated","worms"]  
{  
 qqnorm(treatedForWorms, ylab="Treated Group")   
 qqline(treatedForWorms)  
}



untreatedForWorms <- WormSheep[WormSheep$treated=="untreated","worms"]  
{  
 qqnorm(untreatedForWorms, ylab="Untreated Group")   
 qqline(untreatedForWorms)  
}



The normal probability plots for both treated and untreated appear normal.

### Part 1f

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 at a 5% level of significance.

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

# Use Shapiro-Wilk normality test  
shapiro.test(treatedForWorms)

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

shapiro.test(untreatedForWorms)

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

With p-value= 0.6504, we fail to reject the null hypothesis that the treated group comes from a normal distribution at a 5% level of significance.

With p-value= 0.5491, we fail to reject the null hypothesis that the untreated group comes from a normal distribution at a 5% level of significance.

### Part 1g

Conduct the test of your choice to determine if the population mean worm count for all sheep treated with the drug is less than the mean worm count for the population of untreated sheep. Let .

#### Step 1

State the null and alternative hypotheses for the test. If you use symbols, please define them.

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

where

is the mean worm count with treatment and

is the mean worm count without treatment.

#### Step 2

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

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

# use Welch t-test.  
t.test(treatedForWorms,untreatedForWorms,alternative="less",conf.level=.95)

##   
## Welch Two Sample t-test  
##   
## data: treatedForWorms and untreatedForWorms  
## t = -2.2709, df = 21.972, p-value = 0.01665  
## alternative hypothesis: true difference in means is less than 0  
## 95 percent confidence interval:  
## -Inf -3.189613  
## sample estimates:  
## mean of x mean of y   
## 26.58333 39.66667

#### Step 3

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

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

At a 5% level of significance, we reject the null hypothesis that the treated group has the same or more number of worms than the untreated group in favor of the alternative that the treated group has fewer worms than the untreated group.

Based on our statistical test, there is evidence that using the treatment leads to fewer worms in sheep than not treating at all.

### Part 1h

Compute a 95% confidence interval for the difference in population means. Write an interpretation in the context of the problem for this interval.

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

# Insert your R code here.  
t.test(treatedForWorms,untreatedForWorms,conf.level=.95)$conf.int

## [1] -25.032642 -1.134025  
## attr(,"conf.level")  
## [1] 0.95

At a 95% confidence level, using the treatment leads to between 1 and 25 fewer worms than not being treated at all.

### Part 1i

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

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

Based on the Shapiro-Wilk test for normality, the distributions are approximately normal. Since the Welch t-test is more powerful for approximately normally distributed data, it was used versus the Wilcoxon rank sum test.

We are also assuming independent and random samples.

## Problem 2

Consider two independent, random samples of delay times (in hour) for two airlines over the past 10 years. The task here is to compare the population central values by conducting a hypothesis test and constructing a 95% confidence interval for the difference in average delay time between the airlines for the population of all delays in the past 10 years.

### Part 2a

Load the dataset delays from the DS705data package.

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

# Load the data set delays  
data("delays")

### Part 2b

Begin by exploring the sample means, medians, and standard deviations for each airline. That is, find the sample means, medians, and standard deviations.

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

# Find sample mean, median and standard deviation for the airlines Skybird and Omega  
skybird\_delay <- delays[delays$airline == "Skybird","delay"]  
#skybird\_delay mean  
mean(skybird\_delay)

## [1] 1.421333

#skybird\_delay median  
median(skybird\_delay)

## [1] 1.26

#skybird\_delay standard deviation  
sd(skybird\_delay)

## [1] 0.9461924

omega\_delay <- delays[delays$airline == "Omega","delay"]  
#omega\_delay mean  
mean(omega\_delay)

## [1] 1.4084

#omega\_delay median  
median(omega\_delay)

## [1] 1.085

#omega\_delay standard deviation  
sd(omega\_delay)

## [1] 1.058213

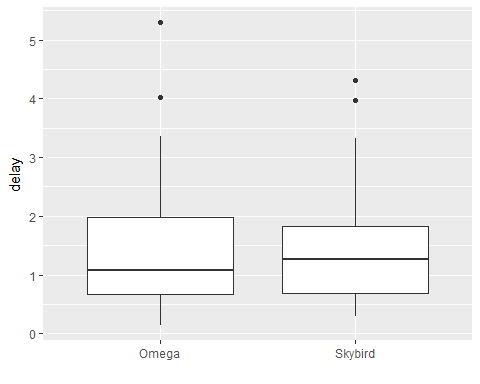
### Part 2b

Produce boxplots, histograms, and normal probability plots for the delay times of both airlines.

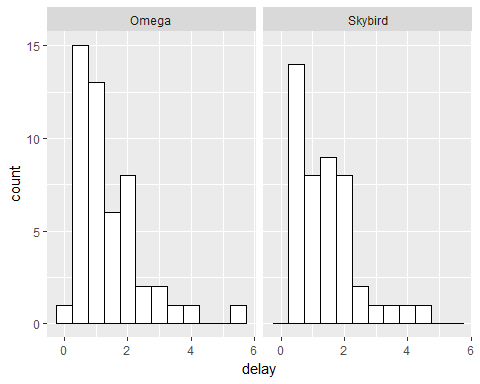
Discuss the aspects of the plots that are relevant to the construction of the best type of test and confidence interval to compare the population central values.

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

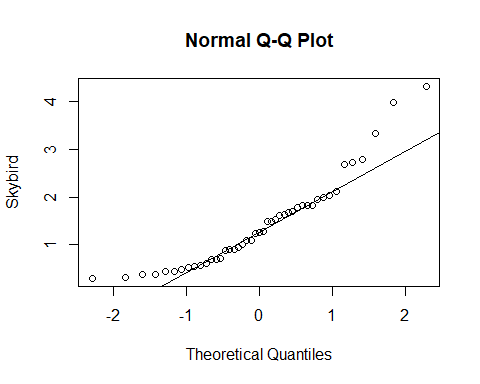
# Create Boxplots and Histograms for delay times of both airlines.  
ggplot(data = delays, aes(x = airline, y = delay)) + geom\_boxplot() + labs(x = "")



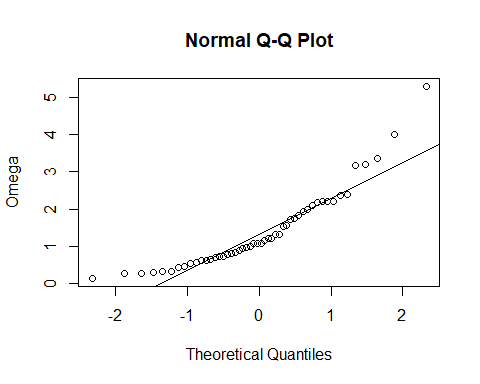
ggplot(delays, aes(x=delay)) +   
 geom\_histogram(binwidth=.5, colour="black", fill="white") +   
 facet\_grid(. ~ airline)



# Make normal probability plot of Skybird airline and add fit line.  
{  
 qqnorm(skybird\_delay, ylab="Skybird")   
 qqline(skybird\_delay)  
}



# Make normal probability plot of Omega airline and add fit line.  
{  
 qqnorm(omega\_delay, ylab="Omega")   
 qqline(omega\_delay)  
}



Based on the boxplots, histograms, and normal probability plots both distributions are right skewed. The Wilcoxon test works best for data that is clearly non normal. Also the distributions look similar to each other. We just need to check whether or not they are shifted from each other.

### Part 2c

Conduct hypothesis tests for the normality of the delay times for both Skybird and Omega using a 5% level of significance in each case.

Use R to compute the test statistic and P-value.

State the hypotheses and conclusion of the test. Include the P-value in your conclusion.

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

# Use Shapiro-Wilk normality test  
shapiro.test(skybird\_delay)

##   
## Shapiro-Wilk normality test  
##   
## data: skybird\_delay  
## W = 0.89362, p-value = 0.0006031

shapiro.test(omega\_delay)

##   
## Shapiro-Wilk normality test  
##   
## data: omega\_delay  
## W = 0.87099, p-value = 6.027e-05

The null hypothesis is that skybird\_delay and omega\_delay come from normal probability distributions. The alternative hypothesis is that they are non-normal. Based on a p-value = 6.027e-05 at a 5% level of significance, we reject the hypothesis that the distributions are normal in favor of the alternative that they are nonmormal.

### Part 2d

Perform the most appropriate test to compare population central values for all Skybird and Omega delay times for the past 10 years. Use a 10% level of significance.

#### Step 1

State the null and alternative hypotheses for the test (in words).

#### -|-|-|-|-|-|-|-|-|-|-|- Answer 2d.step1 -|-|-|-|-|-|-|-|-|-|-|-

The null hypothesis is that the distribution of delay times of Skybird and Omega are the same against the alternative that the delay times are shifted from each other.

#### Step 2

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

#### -|-|-|-|-|-|-|-|-|-|-|- Answer 2d.step2 -|-|-|-|-|-|-|-|-|-|-|-

# Use Wilcox test to find if  
wilcox.test(delay~airline,data=delays,conf.int=TRUE,conf.level=.90)

##   
## Wilcoxon rank sum test with continuity correction  
##   
## data: delay by airline  
## W = 1093.5, p-value = 0.8173  
## alternative hypothesis: true location shift is not equal to 0  
## 90 percent confidence interval:  
## -0.3299311 0.2399673  
## sample estimates:  
## difference in location   
## -0.04997005

#### Step 3

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

#### -|-|-|-|-|-|-|-|-|-|-|- Answer 2d.step3 -|-|-|-|-|-|-|-|-|-|-|-

At a 10% level of significance, we fail to reject the null hypothesis that the distribution of delay times are the same for Skybird and Omega.

In other words, both airlines are delayed about the same amount.

### Part 2e

Construct the most appropriate 95% confidence interval to compare population central values. Write a sentence to interpret the interval in the context of the problem.

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

# Create a 95% confidence interval to compare population central values.  
wilcox.test(delay~airline,data=delays,conf.int=TRUE,conf.level=.95)$conf.int

## [1] -0.3900065 0.2800410  
## attr(,"conf.level")  
## [1] 0.95

The difference in flight delays between Omega and Skybird is between 16.8 minutes and 23.4 minutes.

### Part 2f

Did you use the Welch t-test and interval or the Wilcoxon rank sum test with the corresponding interval? Justify your choice, including some discussion of how well the conditions for the hypothesis test and confidence interval procedures were met.

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

I used the Wilcoxon rank sum test. The problem statement stated that the delay times were independent and random. Due to the right skewness of the delay times, the Wilcoxon rank sum test that uses median as a measure of central tendency works better than Welch t-test which assumes the distributions are approximately normal. Also the distributions looked similar, and Wilcoxon rank test is a good choice to determine if they are shifted from each other.

### Part 2g

What type of hypothesis testing error could have been made here? Describe it in the context of the problem.

### -|-|-|-|-|-|-|-|-|-|-|- Answer 2g -|-|-|-|-|-|-|-|-|-|-|-

It is possible that we could have made a Type II error. We failed to find a difference in delay times when there may actually be a significant difference in delay times between Omega and Skybird airlines.

### Part 2g

Construct a 95% bootstrap confidence interval to compare population medians for the delay times over the past 10 years for the two airlines. Use the Bca method and a bootstrap sample size of 2000. Compare it to the interval produced by the wilcox.test function.

### -|-|-|-|-|-|-|-|-|-|-|- Answer 2g -|-|-|-|-|-|-|-|-|-|-|-

# Construct 95% confidence interval for median delay times between Omega and Skybird  
library(boot)  
bootPctDiff <- function(d,i,p=.5){  
 pct <- tapply(d[i,1],d[,2],quantile,prob=p)  
 pct[1]-pct[2]  
}  
boot.object <- boot(delays, bootPctDiff, R = 2000,p=.5,  
strata=delays$airline)  
boot.ci(boot.object,type='bca')$bca[4:5]

## [1] -0.66 0.32

The wilcox.test function produced the interval [-0.3900065, 0.2800410] and the bootstrap confidence intervals [-0.665, 0.320]. They agree with each other that 0 is in the interval. The bootstrap interval is wider than the Wilcox interval at the same level of significance.