Two-Sample Inference Procedures for Population Central Values

YJim Ryan

02/11/2020y

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

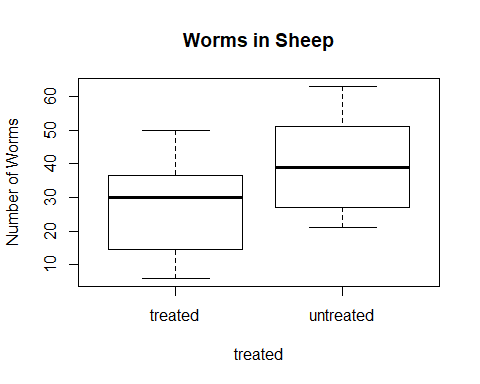
data("WormSheep")

### Part 1b

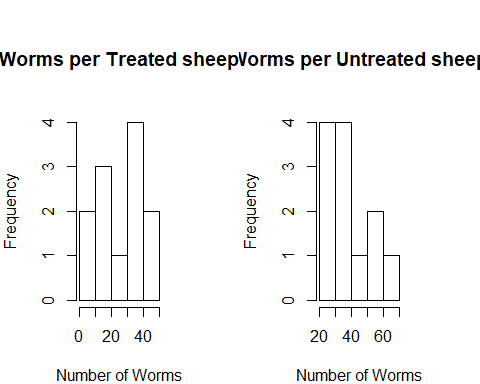
Create boxplot, histograms, and normal probability plots for each group (treated vs. untreated). Be sure that each plot is labeled appropriately.

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

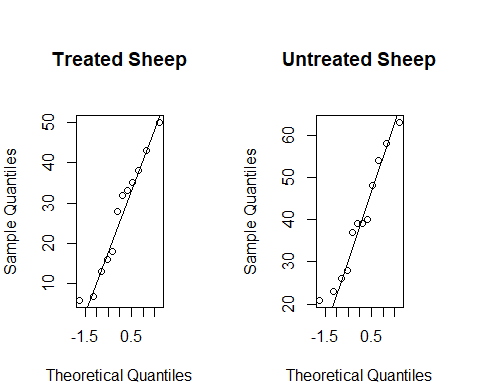
treated <- subset(WormSheep, treated=="treated",  
select=worms:treated)  
  
  
untreated<- subset(WormSheep, treated=="untreated",  
select=worms:treated)  
  
  
boxplot(WormSheep$worms~WormSheep$treated,names=c("treated","untreated"),ylab="Number of Worms",xlab="treated",main="Worms in Sheep")



par(mfrow=c(1,2),mar=c(4,4,6,4))  
  
hist(treated$worms,main="Worms per Treated sheep",xlab="Number of Worms")  
hist(untreated$worms,main="Worms per Untreated sheep",xlab="Number of Worms")



qqnorm(treated$worms,main = "Treated Sheep",  
 xlab = "Theoretical Quantiles", ylab = "Sample Quantiles")  
qqline(treated$worms,main = "Treated Sheep",  
 xlab = "Theoretical Quantiles", ylab = "Sample Quantiles")  
  
qqnorm(untreated$worms,main = "Untreated Sheep",  
 xlab = "Theoretical Quantiles", ylab = "Sample Quantiles")  
qqline(untreated$worms,main = "Untreated Worms",  
 xlab = "Theoretical Quantiles", ylab = "Sample Quantiles")



par(mfrow=c(1,1))

### Part 1c

Do the boxplots show any outliers?

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

No the boxplots do not show any outliers

### Part 1d

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

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

The shape of the Treated histogram is definitely not normalized. The lowest point is in the middle and right next to the highest point. The first and last bars appear to be about the same distribution but for the ones in-between are all over the place.

The shape of the Untreated histogram is definitely not normalized. The first 2 bars ahve the highest density, followed by the the lowest density. The last 3 bars have some normality to them as the middle bar is higher than the ones on either side but again the whole dsample is not normalized due the the first 2 bars being much higher than the remaining 3.

### Part 1e

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

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

Both of the normal probability plots show a little evidence of non-normality. For the Treated sheep The first point is way to the left of the beginning of the line. For the untreated sheep you have the starting point way to the left of the line as well. But I thnk they are both approximately distributed.

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

shapiro.test(untreated$worms)

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

shapiro.test(untreated$worms)

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

## The p-value of .5491 is large enough that the nornality of the number of worms would not be rejected at the 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 -|-|-|-|-|-|-|-|-|-|-|-

#### Step 2

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

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

t.test(WormSheep$worms~ WormSheep$treated, paired = TRUE, alternative = "less")

##   
## Paired t-test  
##   
## data: WormSheep$worms by WormSheep$treated  
## t = -3.5168, df = 11, p-value = 0.002413  
## alternative hypothesis: true difference in means is less than 0  
## 95 percent confidence interval:  
## -Inf -6.402189  
## sample estimates:  
## mean of the differences   
## -13.08333

#### Step 3

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

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

at the null hypothesis is rejected. and . There is not sufficient evidence to suggest that the mean of worms in treated sheep is equal to the mean of worms in untreated sheep.

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

t.test(treated$worms, untreated$worms)

##   
## Welch Two Sample t-test  
##   
## data: treated$worms and untreated$worms  
## t = -2.2709, df = 21.972, p-value = 0.03331  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## -25.032642 -1.134025  
## sample estimates:  
## mean of x mean of y   
## 26.58333 39.66667

at the null hypothesis is rejected. P-value = .003 and . There is sufficient evidence to suggest that the poplulation means of worms for treated sheep is different from the population mean of worms for un treated sheep. (P = .03). The confidence intervals indicate that the populations means of worms for treated sheep is 1.13 to 25.03 less than the population means for untreated sheep.

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

I used the Welch t-test. I used this test as our Shapiro-Wilk normality test indicated that the normality of both populations would not be rejected at 5% level of significance. The welch sample t-test indicated that the null hypothesis that the popuations means were equal could be rejected.

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

data("delays")  
summary(delays)

## delay airline   
## Min. :0.140 Omega :50   
## 1st Qu.:0.675 Skybird:45   
## Median :1.160   
## Mean :1.415   
## 3rd Qu.:1.880   
## Max. :5.300

Omega <- subset(delays, airline=="Omega",  
select=delay:airline)  
  
Skybird <- subset(delays, airline=="Skybird",  
select=delay:airline)

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

mean(Omega$delay)

## [1] 1.4084

sd(Omega$delay)

## [1] 1.058213

median(Omega$delay)

## [1] 1.085

mean(Skybird$delay)

## [1] 1.421333

sd(Skybird$delay)

## [1] 0.9461924

median(Skybird$delay)

## [1] 1.26

par(mfrow=c(1,1))

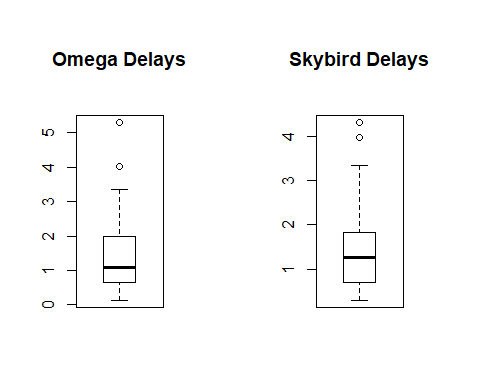
### Part 2b

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

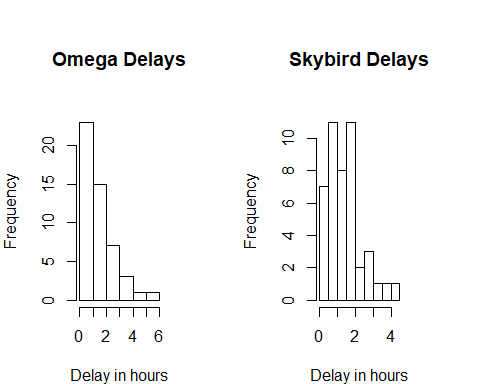
Discuss the aspects of the plots that are relevant to the construction of the best type of test and confidence interval to compare the population central values.

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

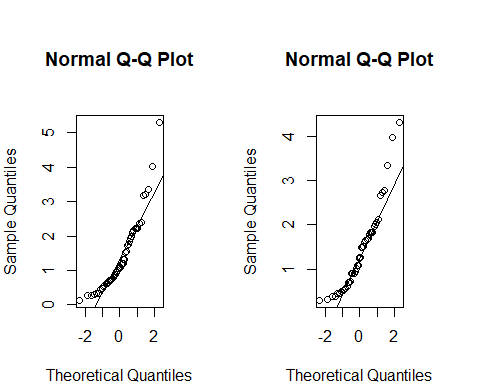
par(mfrow=c(1,2),mar=c(4,4,6,4))  
boxplot(Omega$delay,main='Omega Delays')  
boxplot(Skybird$delay,main='Skybird Delays')



hist(Omega$delay, main='Omega Delays',xlab='Delay in hours')  
hist(Skybird$delay,main='Skybird Delays',xlab='Delay in hours')



qqnorm(Omega$delay)  
qqline(Omega$delay)  
  
  
qqnorm(Skybird$delay)  
qqline(Skybird$delay)



The boxplots,hisatograms and probability plots show that both samples appear to be right-skewed and not normally distributed. The box plots show several outliers for each sample

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

shapiro.test(Omega$delay)

##   
## Shapiro-Wilk normality test  
##   
## data: Omega$delay  
## W = 0.87099, p-value = 6.027e-05

shapiro.test(Skybird$delay)

##   
## Shapiro-Wilk normality test  
##   
## data: Skybird$delay  
## W = 0.89362, p-value = 0.0006031

H\_0:Omega population is normally distributed H\_1:Omega population is NOT normally distributed

Conclusion - the p-value for the Omega dataset is 6.027e-05 which is less than .05 so we reject the null hypothesis.

H\_0:Skybird population is normally distributed H\_1:Skybird population is NOT normally distributed

Conclusion - the p-value for the Skybird dataset is 0.0006031 which is less than .05 so we reject the null hypothesis.

### 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 distributions for Omega and Skybird are identical for the past 10 years.

- The distributions for Omega and Skybird are Not identical for the past 10 years. #### Step 2

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

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

wilcox.test(delay~airline,data=delays,conf.int=TRUE,conf.level = 0.90)

##   
## 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

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

## Based upon the results of the Wilcoxon rank sum test we conclude that we have strong evidence to not reject . There is insufficient evidence to conclude that the distributions of the delay times for Omega and Starbird are shifted from each other.(P= .8173). With 90% confidence the shift in the distribution of delay times for Omega and Starbird is between -.33 and .24 hours.

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

wilcox.test(delay~airline,data=delays,conf.int=TRUE)

##   
## 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  
## 95 percent confidence interval:  
## -0.3900065 0.2800410  
## sample estimates:  
## difference in location   
## -0.04997005

With 95% Confidence, the shift in the distributions of delay times for Omega and Skybird airlines is between -.33 and .24.

### 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 reason I used this is because when I performed the welch-shapiro test, I rejected the null hypothesis which means that the data is not normally distributed. To perform a T-test, the data must be normally distributed. The null hypothesis that the distributions were equal was not rejected. Zero is included in the confidence interval which means there is not sufficient evidence to claim that the population means are different.

### Part 2g

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

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

Replace this text with your answer.

### Part 2h

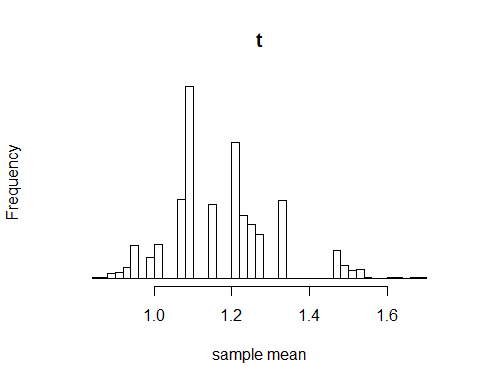
Construct a 95% bootstrap confidence interval to compare population medians for the delay times over the past 10 years for the two airlines. Use the Bca method and a bootstrap sample size of 2000. Compare it to the interval produced by the wilcox.test function.

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

library(boot)

## Warning: package 'boot' was built under R version 3.6.2

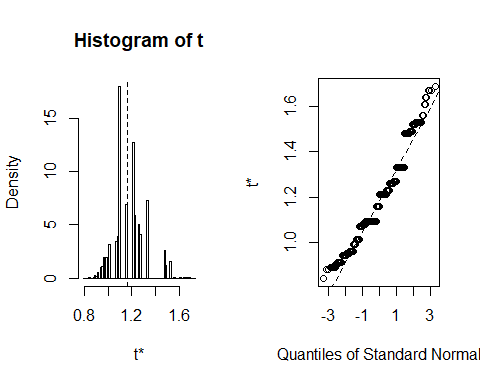
bootMedian <- function(x, i){  
 #x is a vector of data  
 # i is a vector of indices for the resampled elements of x  
 return(median(x[i])) # gets returned  
}  
  
boot.object <- boot(delays$delay,bootMedian,R=2000)  
  
hist(boot.object$t,breaks =40,main='t',xlab='sample mean',yaxt = 'n')



str(boot.object)

## List of 11  
## $ t0 : num 1.16  
## $ t : num [1:2000, 1] 1.23 1.01 1.21 1.09 1.16 0.94 1.16 1.33 1.09 1.09 ...  
## $ R : num 2000  
## $ data : num [1:95] 0.38 0.32 0.61 1.26 2.04 2 1.49 0.56 3.97 0.89 ...  
## $ seed : int [1:626] 10403 1 -2054213166 430028900 -1155737689 1107523837 -243659532 -1789187110 -780517595 1151665727 ...  
## $ statistic:function (x, i)   
## ..- attr(\*, "srcref")= 'srcref' int [1:8] 5 15 9 1 15 1 5 9  
## .. ..- attr(\*, "srcfile")=Classes 'srcfilecopy', 'srcfile' <environment: 0x000000001263d388>   
## $ sim : chr "ordinary"  
## $ call : language boot(data = delays$delay, statistic = bootMedian, R = 2000)  
## $ stype : chr "i"  
## $ strata : num [1:95] 1 1 1 1 1 1 1 1 1 1 ...  
## $ weights : num [1:95] 0.0105 0.0105 0.0105 0.0105 0.0105 ...  
## - attr(\*, "class")= chr "boot"  
## - attr(\*, "boot\_type")= chr "boot"

plot(boot.object)



BCI <- boot.ci(boot.object,conf = .95,type="bca")  
BCI

## BOOTSTRAP CONFIDENCE INTERVAL CALCULATIONS  
## Based on 2000 bootstrap replicates  
##   
## CALL :   
## boot.ci(boot.out = boot.object, conf = 0.95, type = "bca")  
##   
## Intervals :   
## Level BCa   
## 95% ( 0.91, 1.48 )   
## Calculations and Intervals on Original Scale

The interval retuirned by the bootstrap method is .91 and 1.48. This is different than the the produced from the wilcox.test method. With the wilcox.test method,zero was contained in the interval which means that there is not sufficient evidence to claim that the population means are different. We are dealinhg with medians in the bootstrap method but the fact that both are positive indicates that there is sufficient evidence to conclude that the the population medians are different.