ANOVA etc.

Matthew Allen

07/26/2018

Create a Word docx from this R Markdown file for the following exercise. Submit the R markdown file and resulting Word docx file via D2L Dropbox.

## Exercise 1

In the Lesson 3 presentation you saw how to use the Wilcoxon Rank Sum test to compare the difference in median repair times for Macs and PCs. You’ll find the repair dataset in the DS705data package. In this problem we’ll test the hypothesis that the population mean repair times are different for Macs and PCs at the 5% significance level using three different approaches.

### Part 1a

Even though repair times for both computer types are skewed right go ahead and use t.test to test if the population mean times are significantly different. Include your R code below and write a conclusion to the test for practice.

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

# perform t-test to check whether mean repair time for mac and pc are significantly different.  
mac.repair.time <- repair$time[which(repair$type=="Mac")]  
#hist(mac.repair.time)  
pc.repair.time <- repair$time[which(repair$type=="PC")]  
#hist(pc.repair.time)  
t.test(mac.repair.time, pc.repair.time, alternative = 'two.sided')$p.value

## [1] 0.0550238

At a 5% level of significance, we fail to reject the hypotheis that the mean repair times for Macs and PCs are different.

### Part 1b

Now use the boot package to construct a 95% BCa confidence interval for the difference in population mean repair times. Use at least 5000 resamples. Use that confidence interval to write a hypothesis test conclusion to this hypothesis test. (Review: you made similar bootstrap CI’s in Lesson 3.)

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

# Construct a 95% BCa confidence interval for the difference in population mean repair times.  
library(boot)  
bootMeanDiff <- function(d,i){  
 means <- tapply(d[i,1],d[,2],mean)  
 means[1]-means[2]  
 }  
boot.object <- boot(repair, bootMeanDiff, R = 5000,   
 strata = repair$type)  
boot.ci(boot.object,conf=.95,type='bca')$bca[4:5]

## [1] -3.5015842 -0.2226549

The 95% interval constructed by bootstrapping is [-3.4218950, -0.1100385]. The interval does not include zero. Thus at a 5% level of significance, we have evidence that the difference in mean repair times is different from 0.

### Part 1c

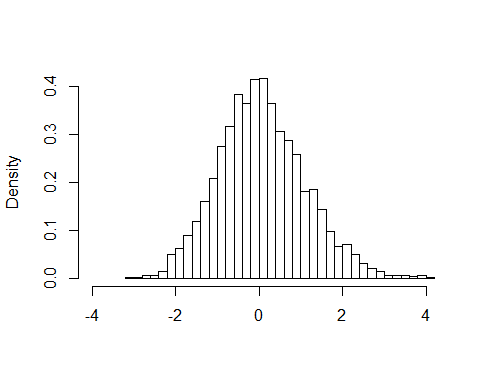
Follow along with with Two Means example in the Bootstrap Hypothesis Testing presentation to bootstrap the two means t test to see if there is a significant difference in population mean repair times. Include a histrogram of the boostrapped t-distribution and write a conclusion to the hypothesis test. (NOTE: in the P value computation slide the last part got cut off, the full code is P <- 2\*min( sum( bootdist < toriginal), sum( bootdist > toriginal ) )/5000.)

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

# two means bootstrap  
xnull <- c(mac.repair.time-mean(mac.repair.time), pc.repair.time-mean(pc.repair.time))  
g <- repair$type  
rs <- rbind(replicate( 5000, sample( xnull[1:62], replace = T) ),  
 replicate( 5000, sample( xnull[63:122], replace = T) ) )  
bootdist <- apply(rs, 2, function(c) t.test(c~g)$statistic )  
  
# compute observed test stat  
toriginal <- t.test( repair$time~g)$statistic  
# for asymmetric distributions the two-tailed P-value is   
# ambiguous. A common solution is to find smaller of the   
# left and right tails and then double it.   
P <- 2\*min( sum( bootdist < toriginal), sum( bootdist > toriginal ) )/5000  
P

## [1] 0.0376

hist( bootdist, probability = TRUE, breaks = 40, main = "",   
 xlab="",xlim=c(-4,4), ylim = c(0,.4))



#curve( dt( x, df = 19), add = TRUE)  
#abline( v = toriginal, col = 'blue')

The p-value is 0.0304. Thus at a 5% level of significance, we have evidence that the difference in mean repair times is different from 0.

### Part 1d

The bootstrap and theoretical t-distributions give different results here. Which do you trust? Why?

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

Both bootstrap methods lead to the same conclusion. The distributions of repair time for macs and pcs are not normal, which is a condition for the t-test. Since the condition of normality is not present, I trust the bootstrap method more.

## Exercise 2

This exercise is based on the data and experimental design from exercises 8.42 & 8.43 in the Ott textbook.

A small corporation makes insulation shields for electrical wires using three different types of machines. The corporation wants to evaluate the variation in the inside diameter dimension of the shields produced by the machines. A quality engineer at the corporation randomly selects shields produced by each of the machines and records the inside diameters of each shield (in millimeters). The goal is to determine whether the location parameters (i.e. mean or median) of the three machines differ. The data set shields is in the DS705data package. The R code to load it is already below.

### Part 2a

Construct side-by-side boxplots for the inside diameters of the insulation shields for the three machines.

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

shields\_box\_plot <- ggplot(aes(x = Machine, y = Diameter), data = shields) + geom\_boxplot()  
shields\_box\_plot

## 

### Part 2b

Comment on what you see in the boxplots. How do the medians compare visually? Do the samples look like they have roughly the same variability? Is there severe skewness or outliers in any of the samples? Be specific.

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

Machines A and B have rougly the same medians, but Machine A definitely has a different median. The machines do not look like they have the same variability for diameter. Machine C appears to be right skewed.

### Part 2c

Which data conditions for ANOVA appear not to be met, if any? Provide reasoning for your answer.

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

The response variable Diameter is not normally distributed for each population. Machine B and C are definitely skewed.

The response variable Diameter does not have have equal variance for each population. The variance of Machine C is greater than Machine A and B.

It is unclear whether samples have been selected randomly, and independently.

### Part 2d

Conduct an analysis of variance test (the standard one that assumes normality and equal variance). (i) State the null and alternative hypotheses, (ii) use R to compute the test statistic and p-value, and (iii) write a conclusion in context at .

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

# Compute F statistic and P-Value.  
anova( lm( Diameter ~ Machine, shields ) )

## Analysis of Variance Table  
##   
## Response: Diameter  
## Df Sum Sq Mean Sq F value Pr(>F)   
## Machine 2 4141 2070.52 2.727 0.09393 .  
## Residuals 17 12907 759.26   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

1. At a 5% level of significace, we fail to reject that all Machines produce the same inside diameter dimension for shields.

### Part 2e

Conduct an analysis of variance test with the Welch correction. (i) State the null and alternative hypotheses, (ii) use R to compute the test statistic and p-value, and (iii) write a conclusion in context at .

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

# conduct anova with welch  
oneway.test(Diameter~Machine,data=shields,var.equal=F)

##   
## One-way analysis of means (not assuming equal variances)  
##   
## data: Diameter and Machine  
## F = 3.974, num df = 2.0000, denom df = 8.4087, p-value = 0.06096

## At a 5% level of significace, we fail to reject that all Machines produce the same inside diameter dimension for shields.

### Part 2f

Which data conditions for Welch ANOVA are not met, if any? Provide reasoning for your answer.

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

The response variable Diameter is not normally distributed for each population. Machine B and C are definitely skewed.

It is unclear whether samples have been selected randomly, and independently.

### Part 2g

Conduct a Kruskal-Wallis test. (i) State the null and alternative hypotheses, (ii) use R to compute the test statistic and p-value, and (iii) write a conclusion in context using .

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

1. the population distributions are the same the population distributions are not the same

# Conduct a Kruskal-Wallis test.  
kruskal.test( Diameter ~ Machine, data = shields )

##   
## Kruskal-Wallis rank sum test  
##   
## data: Diameter by Machine  
## Kruskal-Wallis chi-squared = 9.8914, df = 2, p-value = 0.007114

1. At a 5% level of significance, there is evidence that the distribution of diameters are different between the machines, or that the distribution of diameters are shifted from each other. That is they have different medians.

### Part 2h

Which data conditions for the Kruskal-Wallis test are not met, if any? Provide reasoning for your answer.

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

The different machines’ diameter distributions may not have equal variances.

### Part 2i

Conduct a bootstrapped ANOVA test using pooled residuals and unequal variances as in the notes. (i) State the null and alternative hypotheses, (ii) use R to compute the test statistic and p-value, and (iii) write a conclusion in context . Do not use a helper function, instead mimic the code in the notes using a for loop to construct the boostrapped sampling distribution.

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

1. the population distributions are the same the population distributions are not the same

resA <- shields$Diameter[shields$Machine=='A'] - mean(shields$Diameter[shields$Machine=='A'])  
resB <- shields$Diameter[shields$Machine=='B'] - mean(shields$Diameter[shields$Machine=='B'])  
resC <- shields$Diameter[shields$Machine=='C'] - mean(shields$Diameter[shields$Machine=='C'])  
pop.null <- data.frame(res=c(resA,resB,resC),shields$Machine)  
with(pop.null, tapply( res, shields$Machine, mean) )

## A B C   
## 5.329071e-16 -3.552714e-16 2.131975e-15

B <- 10000; Fstar1 <- numeric(B)  
for (i in 1:B){  
 pop.null <- data.frame(   
 res = sample( c(resA, resB, resC), replace = T), shields$Machine )  
 Fstar1[i] <- oneway.test( res~shields$Machine, data=pop.null,   
 var.equal=FALSE)$statistic  
}  
F.obs <- oneway.test( Diameter ~ Machine, data = shields)$statistic  
  
Fstar1[is.na(Fstar1)] <- 100\*F.obs  
p.approx1 <- sum( Fstar1 > F.obs )/B; p.approx1

## [1] 0.0573

1. At a 5% level of significance, we fail to reject that the distributions are the same.

### Part 2j

Repeat the bootstrapped ANOVA test using unpooled residuals and unequal variances as in the notes. (i) State the null and alternative hypotheses, (ii) use R to compute the test statistic and p-value, and (iii) write a conclusion in context . Go ahead and use the helper function or t1waybt do do this problem.

### -|-|-|-|-|-|-|-|-|-|-|- Answer 2j -|-|-|-|-|-|-|-|-|-|-|-

1. the population distributions are the same the population distributions are not the same

B <- 10000; Fstar2 <- numeric(B)  
for (i in 1:B){  
 pop.null <- data.frame(   
 res = c( sample( resA, replace = T ),   
 sample( resB, replace = T ),   
 sample( resC, replace = T ) ), shields$Machine )  
 Fstar2[i] <- oneway.test( res~shields$Machine, data=pop.null,   
 var.equal=FALSE)$statistic  
}  
Fstar2[is.na(Fstar2)] <- 100\*F.obs  
p.approx2 <- sum( Fstar2 > F.obs )/B; p.approx2

## [1] 0.3396

#install.packages("WRS2")  
require('WRS2') # install this package if needed

## Loading required package: WRS2

## Warning: package 'WRS2' was built under R version 3.5.1

# use 10% trimmed means   
t1waybt(Diameter~Machine,data=shields,tr=0.1,nboot=10000)

## Warning in t1waybt(Diameter ~ Machine, data = shields, tr = 0.1, nboot  
## = 10000): Some bootstrap estimates of the test statistic could not be  
## computed.

## Call:  
## t1waybt(formula = Diameter ~ Machine, data = shields, tr = 0.1,   
## nboot = 10000)  
##   
## Effective number of bootstrap samples was 9969.  
##   
## Test statistic: 2.2086   
## p-value: 0.30876   
## Variance explained 0.449   
## Effect size 0.67

1. At a 5% level of significance, we fail to reject that the distributions are the same.

### Part 2k

Which seems better and why, the bootstrap procedure with the pooled or unpooled residuals?

### -|-|-|-|-|-|-|-|-|-|-|- Answer 2k -|-|-|-|-|-|-|-|-|-|-|-

The bootstrap procedure with pooled residuals seems better. It is visually apparent from the boxplot that the means are different. The bootstrap procedure with pooled residuals agrees with this.

### Part 2l

Do any of the four statistical inference procedures used here provide a clear answer to the question of whether or not the three machines produce the same average inside diameter for the insulation shields?

### -|-|-|-|-|-|-|-|-|-|-|- Answer 2l -|-|-|-|-|-|-|-|-|-|-|-

There is no clear answer.

### Part 2m

If you were responsible for conducting the statistical analysis here, what would you report to the engineer?

### -|-|-|-|-|-|-|-|-|-|-|- Answer 2m -|-|-|-|-|-|-|-|-|-|-|-

I would say the results are inconclusive.

### Part 2n

What impact do you think samples of sizes 5, 5, and 10 play here?

### -|-|-|-|-|-|-|-|-|-|-|- Answer 2n -|-|-|-|-|-|-|-|-|-|-|-

The sample sizes are somewhat small. An increase in the sample size especially for the bootstrap method, would increase the power of the tests.

### Part 2o

Often the Kruskall Wallis test is presented as a test of

the population distributions are all the same

the population distributions are not all the same,

but this is not what KW tests as this example shows. Take 3 random samples of size 100 from normal distributions having mean 0 and standard deviations 1, 10, and 50. If KW were testing the hypotheses above, then we should reject since these three distributions are clearly different. Run the KW test. You should get a large -value. Why did you get a large -value when the distributions are so different?

### -|-|-|-|-|-|-|-|-|-|-|- Answer 2o -|-|-|-|-|-|-|-|-|-|-|-

set.seed(321)  
x <- c( rnorm(100,0,1), rnorm(100,0,10), rnorm(100,0,50))  
groups <- factor( (rep( c('A','B','C'), each=100 ) ) )  
# Conduct a Kruskal-Wallis test.  
kruskal.test(x ~ groups)

##   
## Kruskal-Wallis rank sum test  
##   
## data: x by groups  
## Kruskal-Wallis chi-squared = 0.84306, df = 2, p-value = 0.656

The condition that the populations have equal variances for the Kruskall Wallis test is violated. The three normal distributions have standard deviations of 1, 10 and 50.