R Assignment 6

Jacey Davies

12/2/2020

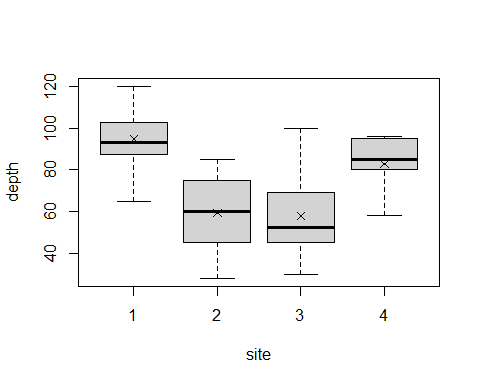
## Exercise 19.1

Consider the following data: (see problem). These figures provide the depths (in centimeters) at which important archaeological finds were made at four sites in New Mexico. Store these data in your R workspace, with one vector containing depth and the other vector containing hte site of each observation.

depth = c(93,120,65,105,115,82,99,87,100,90,78,95,93,88,110,85,45,80,28,75,70,65,55,50,40,100,75,65,40,73,65,50,30,45,50,45,55,96,58,95,90,65,80,85,95,82)  
site = rep(factor(c(1,2,3,4)),times=c(15,10,12,9))

1. Produce side-by-side boxplots of the depths split by group, and use additional points to mark the locations of the sample means.

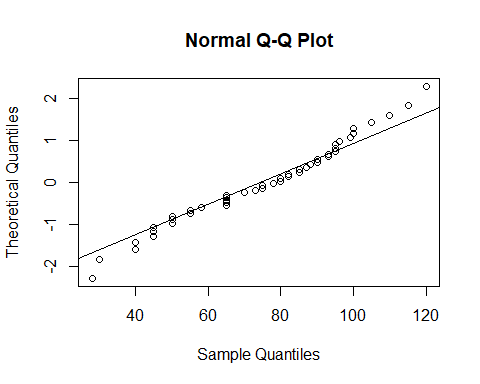
means = c(mean(depth[1:15]),mean(depth[16:25]),mean(depth[26:37]),mean(depth[38:46]))  
boxplot(depth~site)  
points(1:4,means,pch=4)



1. Assuming independence, execute diagnostic checks for normality and equality of variances.

Checking for normality:

qqnorm(depth,datax=TRUE)  
qqline(depth,datax=TRUE)



The data appears approximately normal.

Checking for equality of variances:

library(car)

## Warning: package 'car' was built under R version 4.0.3

## Loading required package: carData

## Warning: package 'carData' was built under R version 4.0.3

leveneTest(depth~site)

## Levene's Test for Homogeneity of Variance (center = median)  
## Df F value Pr(>F)  
## group 3 0.7978 0.5021  
## 42

P-value = 0.5021

The p-value is large, meaning that the variances are likely equal.

1. Perform and conclude a one-way ANOVA test for evidence of a difference between the means.

H0 : µ1 = µ2 = µ3 = µ4

Ha : At least two of the means are not equal

finds.anova = aov(depth~site)  
summary(finds.anova)

## Df Sum Sq Mean Sq F value Pr(>F)   
## site 3 12397 4132 15.14 7.99e-07 \*\*\*  
## Residuals 42 11465 273   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

P-value = 7.99 x 10-7

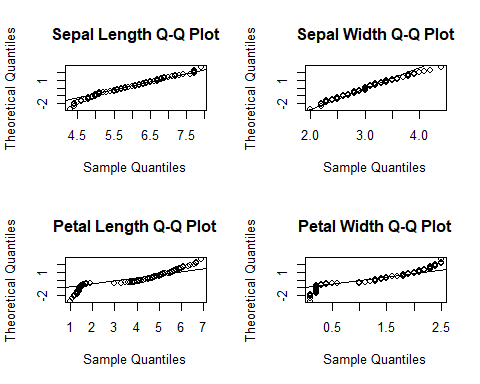
The p-value is very small, meaning that there is significant evidence for a difference between the means.

In Section 14.4, you looked at the data set providing measurements on petal and sepal sizes for three species of iris flowers. This is available in R as iris.

1. Based on diagnostic checks for normality and equality of variances, decide which of the four outcome measurements (sepal length/width and petal length/width) would be suitable for ANOVA (using the species as the group variable).

Checking for normality:

library(datasets)  
data("iris")  
  
par(mfrow = c(2,2))  
  
qqnorm(iris$Sepal.Length,datax=TRUE, main = "Sepal Length Q-Q Plot")  
qqline(iris$Sepal.Length,datax=TRUE)  
  
qqnorm(iris$Sepal.Width,datax=TRUE, main = "Sepal Width Q-Q Plot")  
qqline(iris$Sepal.Width,datax=TRUE)  
  
qqnorm(iris$Petal.Length,datax=TRUE, main = "Petal Length Q-Q Plot")  
qqline(iris$Petal.Length,datax=TRUE)  
  
qqnorm(iris$Petal.Width,datax=TRUE, main = "Petal Width Q-Q Plot")  
qqline(iris$Petal.Width,datax=TRUE)



The sepal lengths and sepal widths appear to be approximately normal, while the petal lengths and petal widths do not.

Checking for equality of variances:

leveneTest(Sepal.Length~Species,data=iris)

## Levene's Test for Homogeneity of Variance (center = median)  
## Df F value Pr(>F)   
## group 2 6.3527 0.002259 \*\*  
## 147   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

leveneTest(Sepal.Width~Species,data=iris)

## Levene's Test for Homogeneity of Variance (center = median)  
## Df F value Pr(>F)  
## group 2 0.5902 0.5555  
## 147

leveneTest(Sepal.Width~Species,data=iris)

## Levene's Test for Homogeneity of Variance (center = median)  
## Df F value Pr(>F)  
## group 2 0.5902 0.5555  
## 147

leveneTest(Petal.Length~Species,data=iris)

## Levene's Test for Homogeneity of Variance (center = median)  
## Df F value Pr(>F)   
## group 2 19.48 3.129e-08 \*\*\*  
## 147   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

The sepal lengths, petal lengths, and petal widths all have small p-values, meaning that the variances are not equal. However, the sepal widths have a large p-value, so its variances are equal.

1. Carry out one-way ANOVA for any suitable measurement variables.

Since the sepal widths were the only variable to meet the normality and equality of variances requirements, they will undergo the one-way ANOVA.

H0 : µsetosa = µversicolor = µvirginia

Ha : At least two of the means are not equal

sw.anova = aov(Sepal.Width~Species,data=iris)  
summary(sw.anova)

## Df Sum Sq Mean Sq F value Pr(>F)   
## Species 2 11.35 5.672 49.16 <2e-16 \*\*\*  
## Residuals 147 16.96 0.115   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

P-value < 2 x 10-16

Since the p-value is very small, there is a significant difference in the means of the sepal widths of the flower species.