HW6

Adhrit Srivastav, eid: Ams22362

10/22/2021

5.6 Not a condition?

B

5.7 ANOVA plots

D

5.11 North Carolina births

a) Random Assignment is impossible because we cannot randomly assign infant birth weight to racial/ethnic group of the mothers. The child and mothers are innately connected. Inference about cause is not possible because random assignment/randomization is not possible which is essential for the data for inference.

b) The samples must be randomized. The requirements of this kind of inference were satisfied.

5.14 Car Ages

a) ANOVA is used to compare multiple groups, and since there are four groups we can use ANOVA here.
b) The response variables are the ages of the cars, which is quantitative and thus ANOVA is usable here.
c) One way ANOVA can be used when sample sizes are different, thus ANOVA can be used when the four groups have different sample sizes.
d) The data was collected from 200 random people with parking permits. Thus we can generalize the results for the population of all people with a parking permit with a certain confidence level. As long as the sample is randomized, we can generalized it.

5.24 Turtles and habitat

a) H_0: $B_N = B_O = B_T$; H_1: at least one $B_i \neq 0$.
b) We need the measurements of the sizes of the turtles along with labels for which state the turtle hatchling is from.
c) We would need to make sure the data is randomized. It would help to ensure that the hatchlings are not related, because that could influence their growth rates to be similar. This would make the data less random and could potentially fail a condition for ANOVA.

5.27 Mouse serotonin: exploratory

```
mouse = read.csv("C:/Users/adhri/OneDrive/Documents/R/App_Reg_and_Time_Series/chpt5/datasets/MouseBrain
attach(mouse)

boxplot(Contacts~Genotype,data=mtcars, main="Contacts across Genotypes", xlab="Genotypes", ylab="Contact
```

Contacts across Genotypes
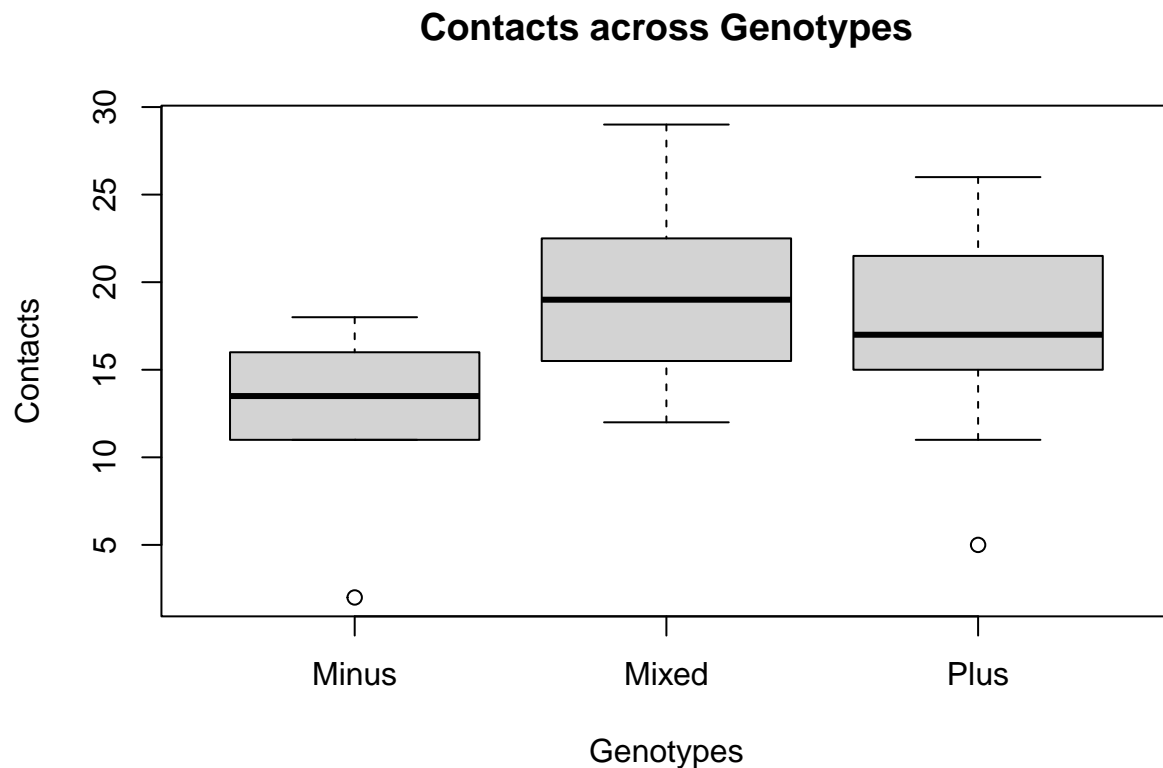
a) They reveal there are more median contacts in the mixed genotype. There appears to be more contacts in the mixed genotype in general. There are also appear to be fewer contacts in the minus group on average.

b) The conditions are met. The data is normal, each sample is independent from each other, and the variance condition is met.

5.29 Child poverty

One group has very small variance and the dotplot appears to have some clustering which means it might not pass the equal variance condition.

5.34 Aphid honeydew

a) The DF for aphid race - host plant is 5 (51 - 46 = 5). The SS is 24.9 (64.77 - 39.87 = 24.9). F-value is 5.7467 (4.9807/.8667 = 5.7467).

2

b) 6 different aphid/plant combinations were considered in this analysis. We know this becuase the number of groups is df + 1 and we can calculate group df by subtracting error df from total df.

c) The p-value is zero which indicates we can reject the null hypothesis and conclude that the amounts of honeydew produced by aphids differ for different combinations of type of aphid and type of host plant and that at least one $a_i \neq 0$.
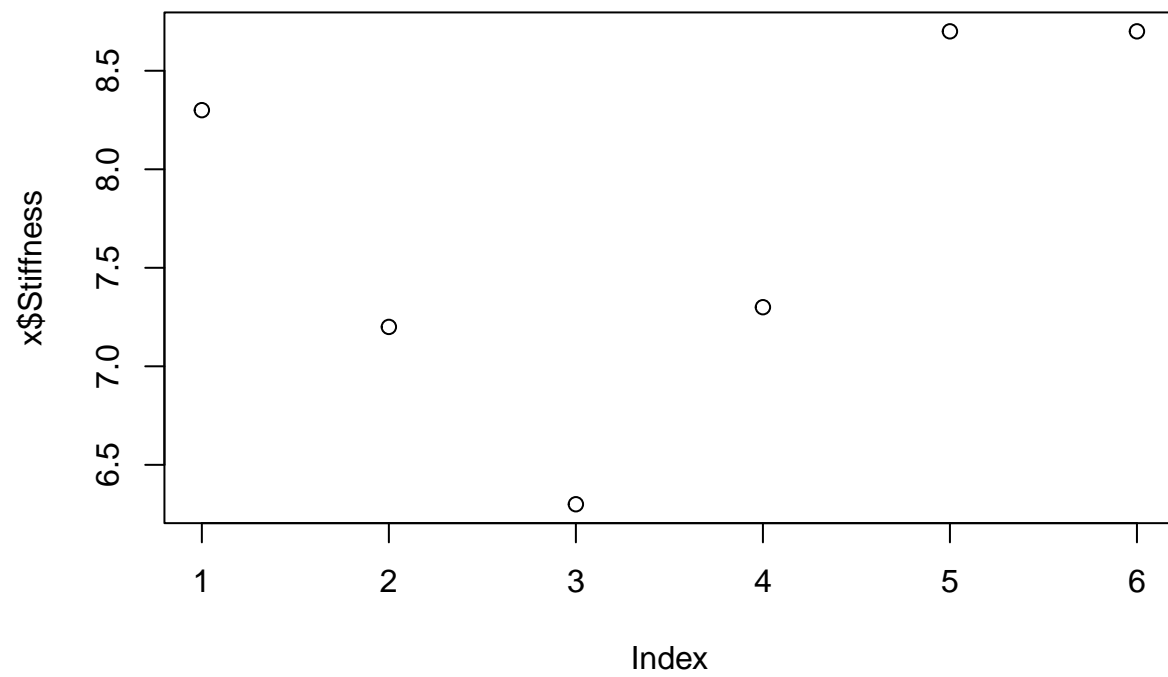
5.38 Meniscus: stiffness

```
meniscus = read.csv("C:/Users/adhri/OneDrive/Documents/R/App_Reg_and_Time_Series/chpt5/datasets/Meniscu
attach(meniscus)

x = meniscus[meniscus$Method == "1",]
y = meniscus[meniscus$Method == '2',]
z = meniscus[meniscus$Method == '3',]
boxplot(x$Stiffness, y$Stiffness, z$Stiffness)
```
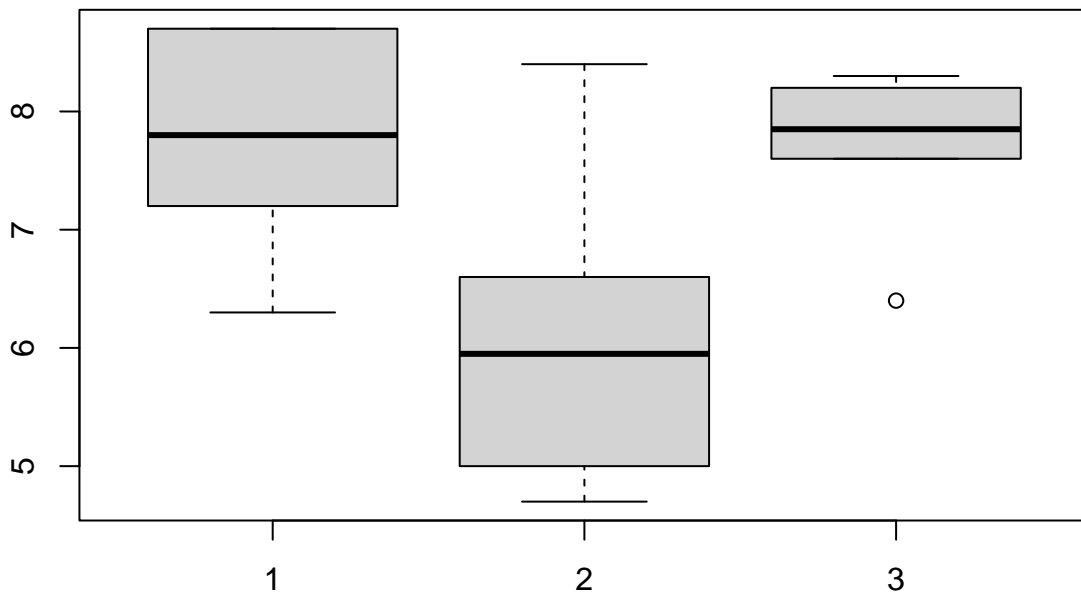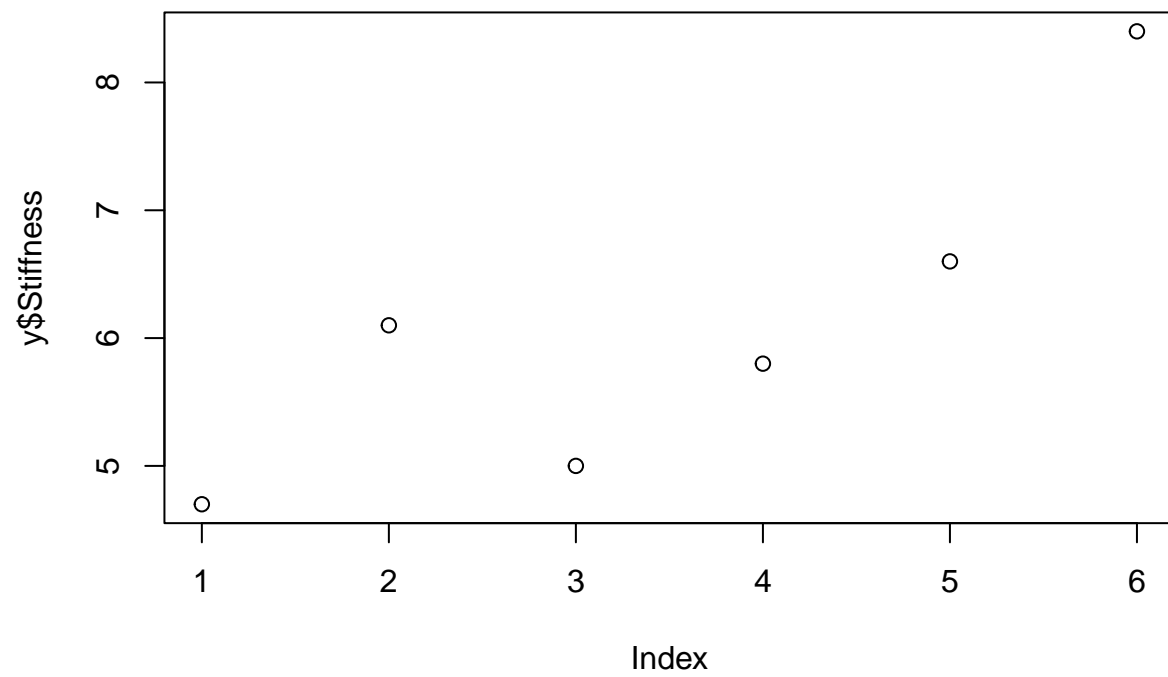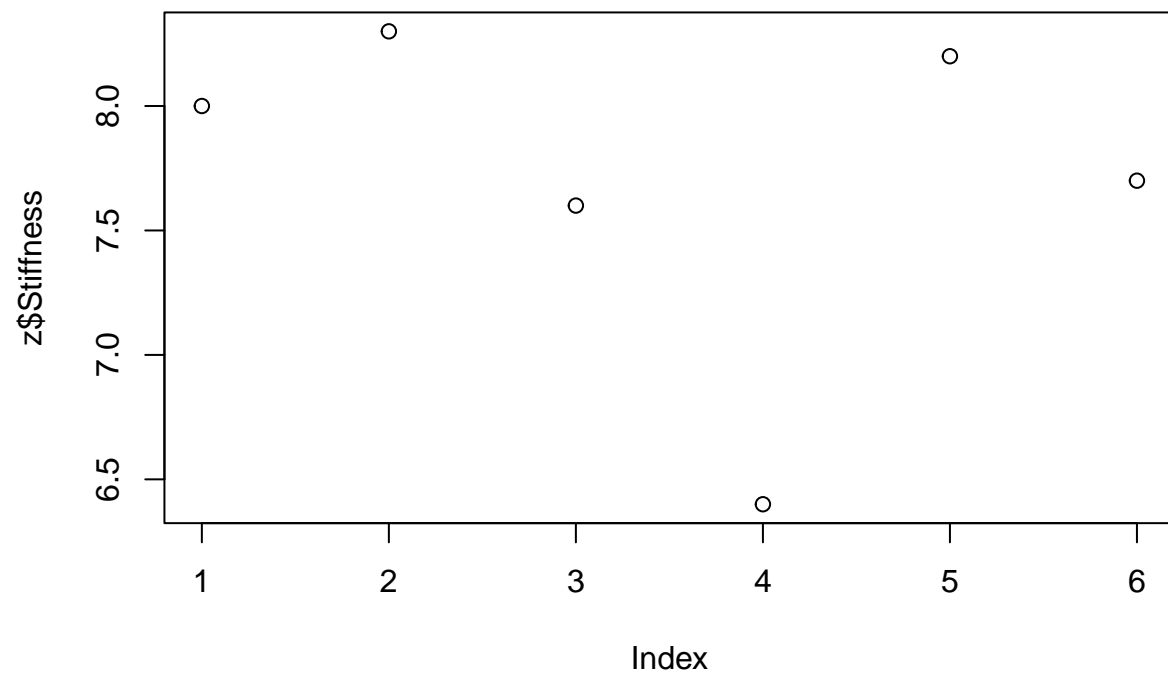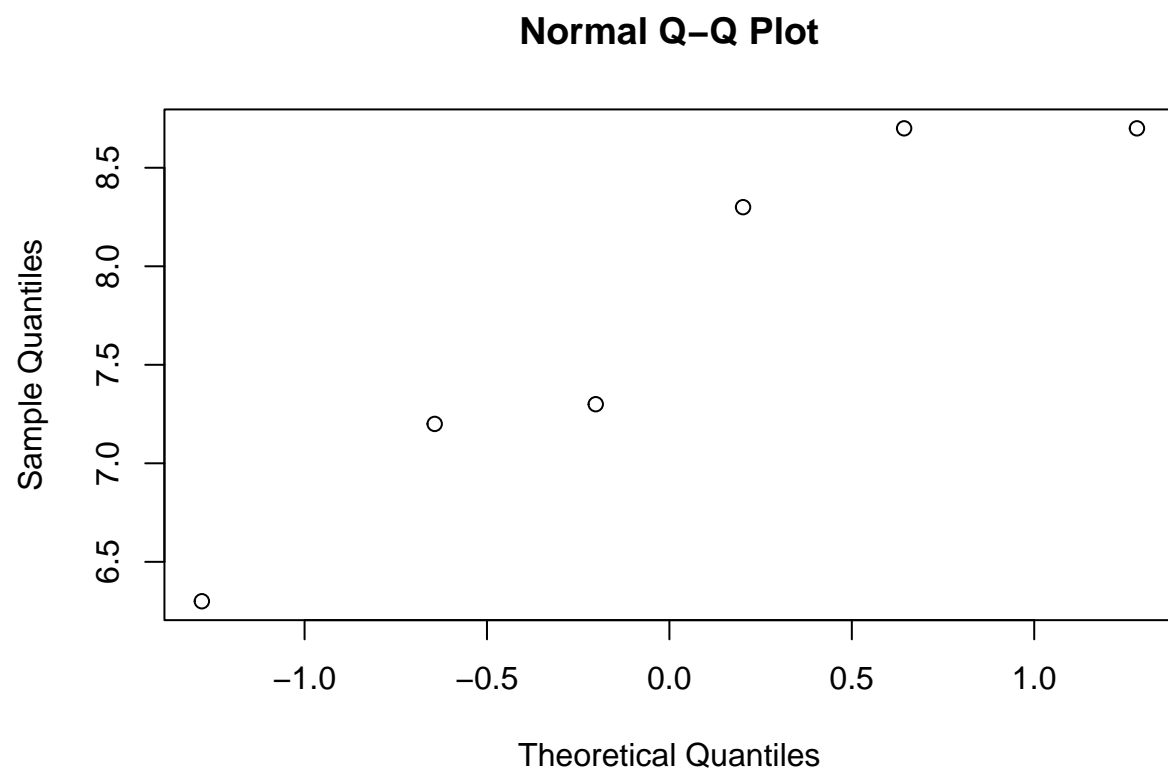


```
plot(x$Stiffness)
```

```
plot(y$Stiffness)
```
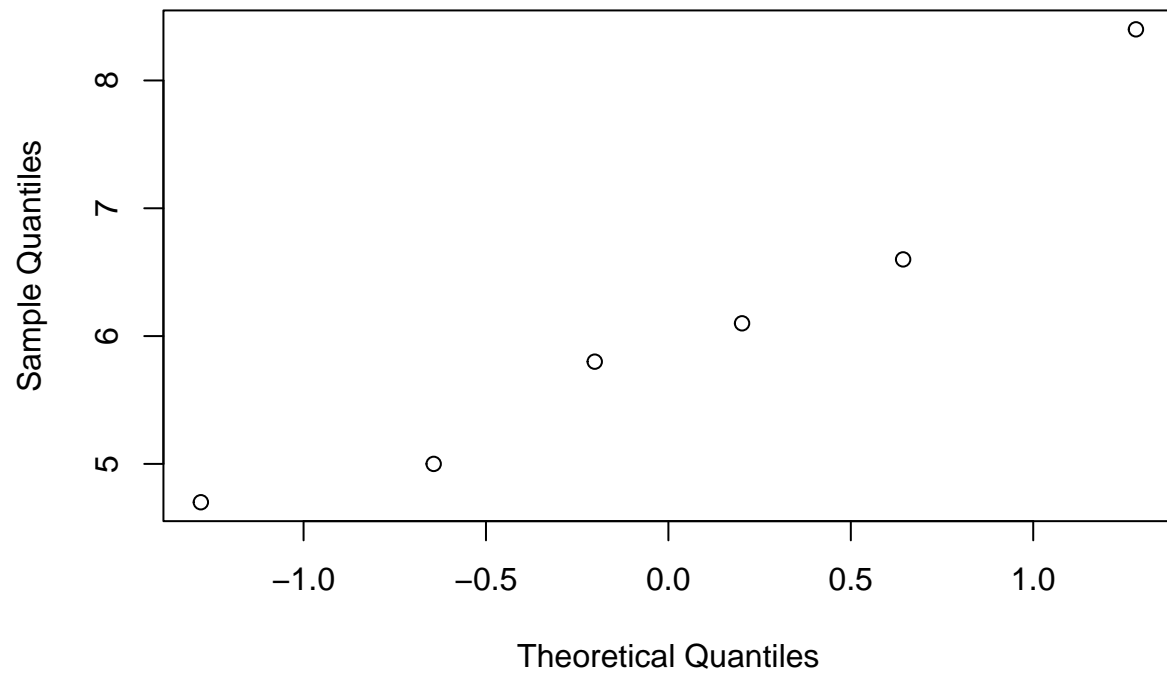
```
plot(z$Stiffness)
```

```
qqnorm(x$Stiffness)
```

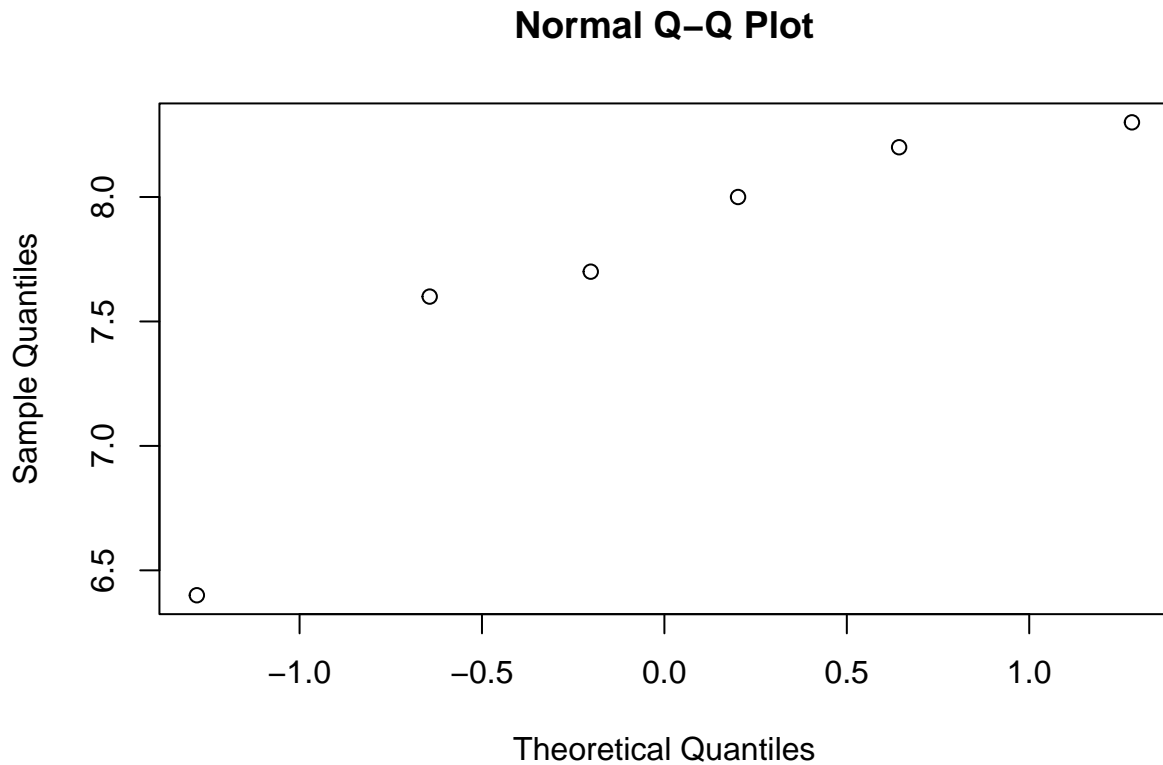# Normal Q-Q Plot


```
qqnorm(y$Stiffness)
```

Normal Q–Q Plot


```
qqnorm(z$Stiffness)
```

Normal Q–Q Plot


```
stiff = aov(Stiffness ~ factor(Method))
summary(stiff)
```

```
##                 Df Sum Sq Mean Sq F value Pr(>F)
## factor(Method)  2   10.57   5.285   4.981 0.0219 *
## Residuals       15  15.91   1.061
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

a) H_0: $\alpha_1 = \alpha_2 = \alpha_3 = 0$; H_1: at least one $\alpha_i \neq 0$.

b) The conditions are met. The data is normal, each sample is independent from each other, and it passes the equal variance condition. There is one outlier in the Method 3 group.

c) The p-value is .0219 which means we reject the null hypothesis. Thus the data provides strong evidence that the mean values of stiffness differs based on the type of meniscus repair.
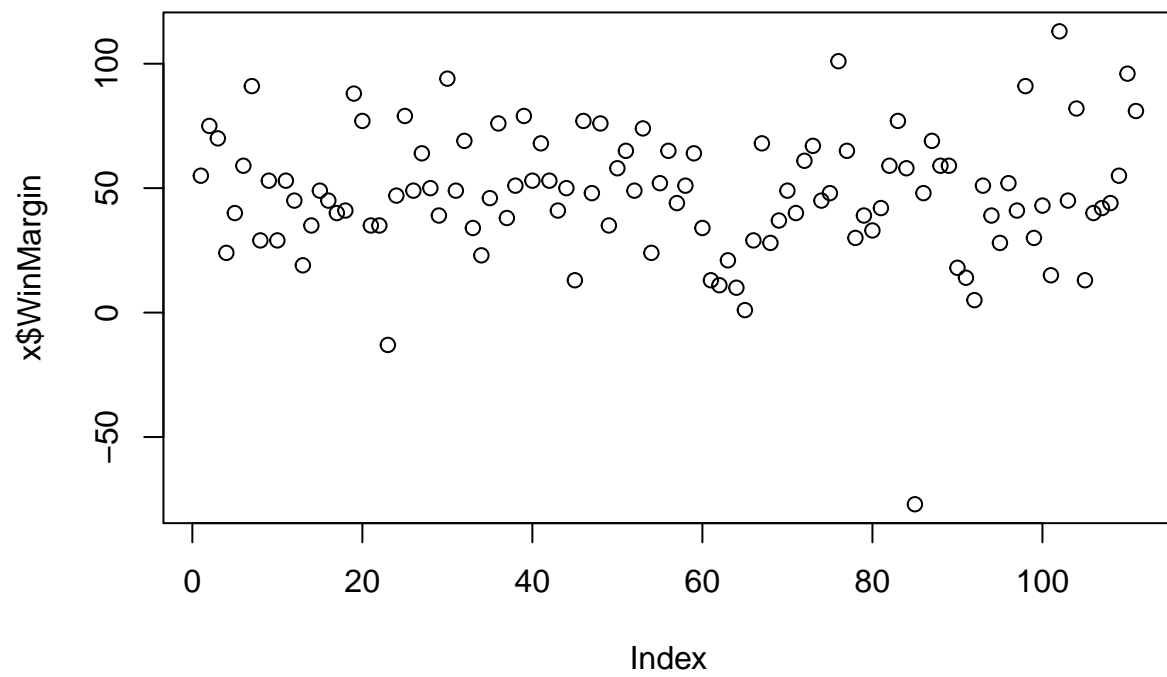
5.52 Words with Friends

```
words = read.csv("C:/Users/adhri/OneDrive/Documents/R/App_Reg_and_Time_Series/chpt5/datasets/WordsWithF:
attach(words)

plot(WinMargin ~ factor(BlanksNumber))
```
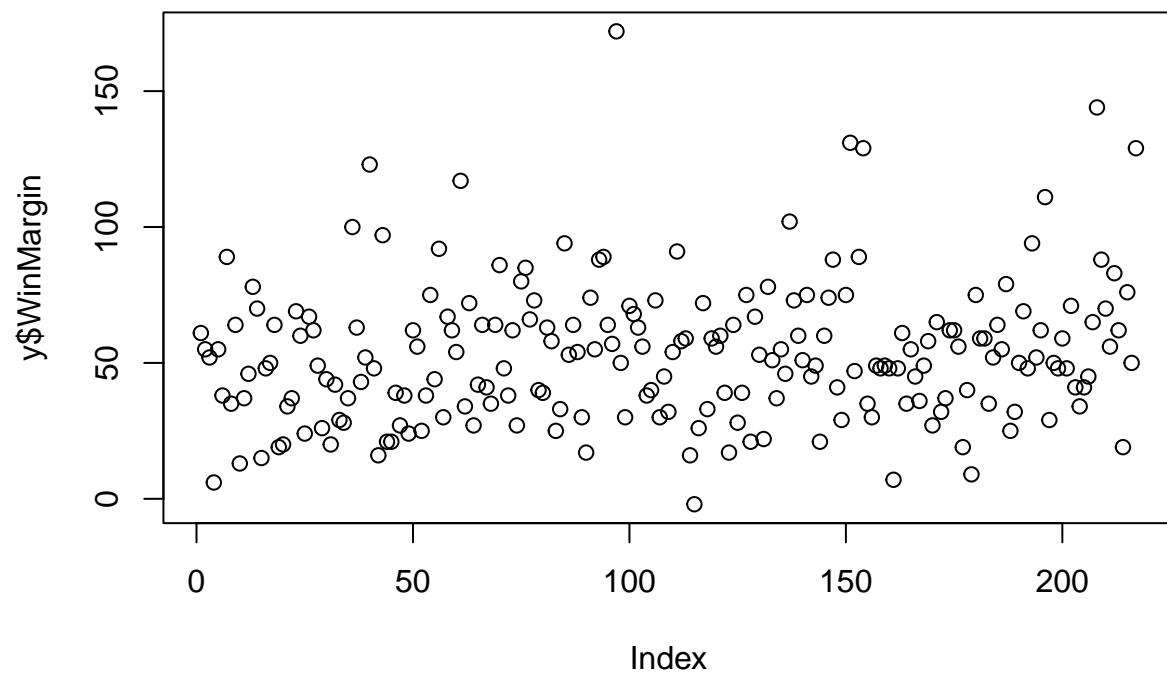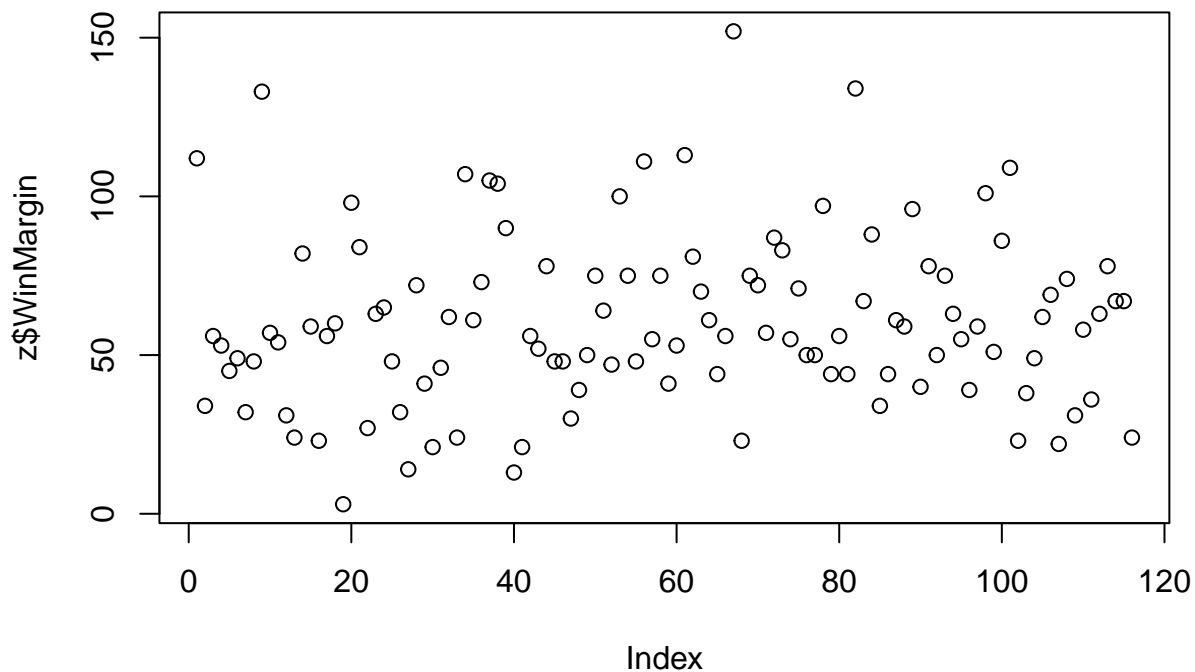
9
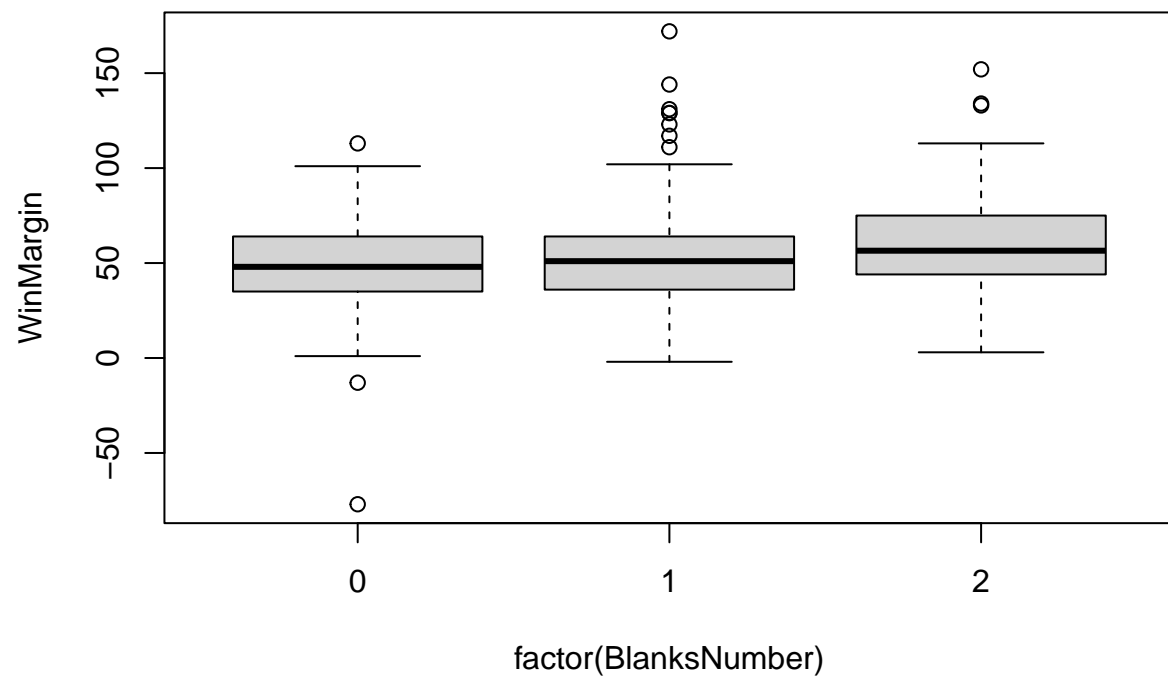
```
x = words[words$BlanksNumber == "0",]
y = words[words$BlanksNumber == "1",]
z = words[words$BlanksNumber == "2",]
plot(x$WinMargin)
```

```
plot(y$WinMargin)
```

```
plot(z$WinMargin)
```

```
blank = aov(WinMargin ~ factor(BlanksNumber))
summary(blank)
```

```
##                      Df Sum Sq Mean Sq F value  Pr(>F)
## factor(BlanksNumber)   2   9514    4757   6.988 0.00103 **
## Residuals            441 300202     681
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

a) The conditions are met. The data is normal, each sample is independent from each other, and it passes the equal variance condition.

b) The p-value is .00103 which is less than the alpha which means we can reject the null hypothesis and conclude that there is a noticeable difference between the final scores of the games depending on how many blank tiles the player receives.