# CS624 HW1

## Ryan Richardson

### 29/09/2020

## Problem 1

### Part A

```
firstTest = data.frame("Method" = c("1","1","1","1","1","1"), "Score" = c(79,66,57,91,42,59))
secondTest = data.frame("Method" = c("2", "2", "2", "2", "2", "2"), "Score" = c(71,43,58,78,20,56))
firstTest$Method = as.factor(firstTest$Method)
secondTest$Method = as.factor(secondTest$Method)

shapiro.test(firstTest$Score)
```

**Normality**

```
##
##  Shapiro-Wilk normality test
##
## data:  firstTest$Score
## W = 0.97831, p-value = 0.9429
```
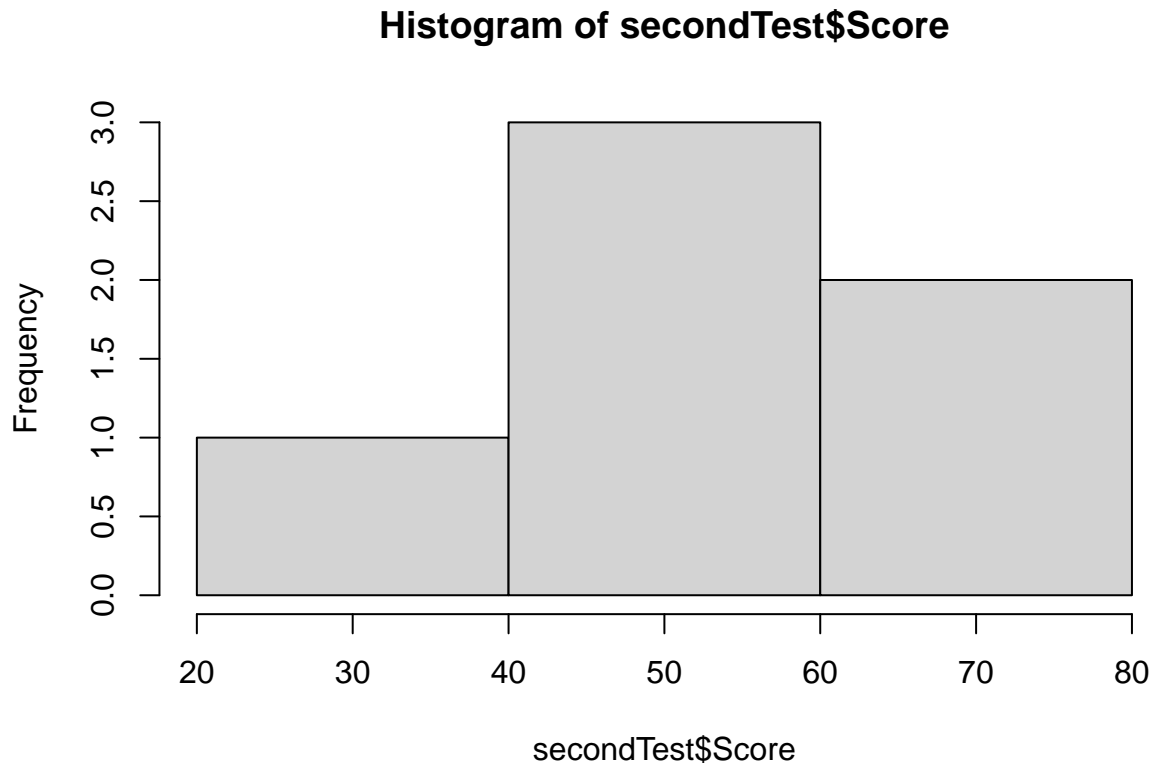
```
shapiro.test(secondTest$Score)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  secondTest$Score
## W = 0.95066, p-value = 0.7456
```

```
hist(firstTest$Score)
```

**Histogram of firstTest$Score**

```
hist(secondTest$Score)
```

## Histogram of secondTest$Score



Shapiro-Wilk Test for Normality has a p-value > .05 for both sets of scores, we can assume they are normally distributed. Plotting a histogram of both sets of scores also suggests normality, though there may ben an outlier on in the first set of points.

**Null Hypothesis**   The null hypothess is that the mean scores of the first 6 students are not different from the means of the second 6 students.

$$H_0 : \mu_1 = \mu_2$$

$$H_a : \mu_1 \neq \mu_2$$

**Two Sample T-Test & Findings**   Findings: Fail to reject $H_0 : \mu_1 = \mu_2$ with a p-value of .3298 at $\alpha = 0.01$

Confidence Interval (Difference of Means) : $(-23.93371, 46.60038)$

By failing to reject, we assume that there is no difference in the mean scores between the two teaching methods on different populations of students. Even so, the confidence interval suggests that the true value of the difference between teaching methods is somewhere between -23.93 and 46.60.

```
t.test(firstTest$Score, secondTest$Score , paired = FALSE, conf.level = 0.99, mu = 0)
```

```
##
##  Welch Two Sample t-test
##
## data:  firstTest$Score and secondTest$Score
## t = 1.026, df = 9.6833, p-value = 0.3298
## alternative hypothesis: true difference in means is not equal to 0
## 99 percent confidence interval:
```

```
##  -23.93371  46.60038
## sample estimates:
## mean of x mean of y
##  65.66667  54.33333
```

**Part B**

**Normality** We start by taking the differences between the scores from Method 1 and the scores form Method 2 for the same students. Running the differences through Shapiro-Wilkes test we get a p-value of .55, which allows us to assume normality for the paired t-test. However, the histogram of the differences looks like it is uniformly distrbuted, except for a single outlier.

```
differences = firstTest$Score - secondTest$Score
shapiro.test(differences)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  differences
## W = 0.92658, p-value = 0.5539
```

```
hist(differences)
```

## Histogram of differences
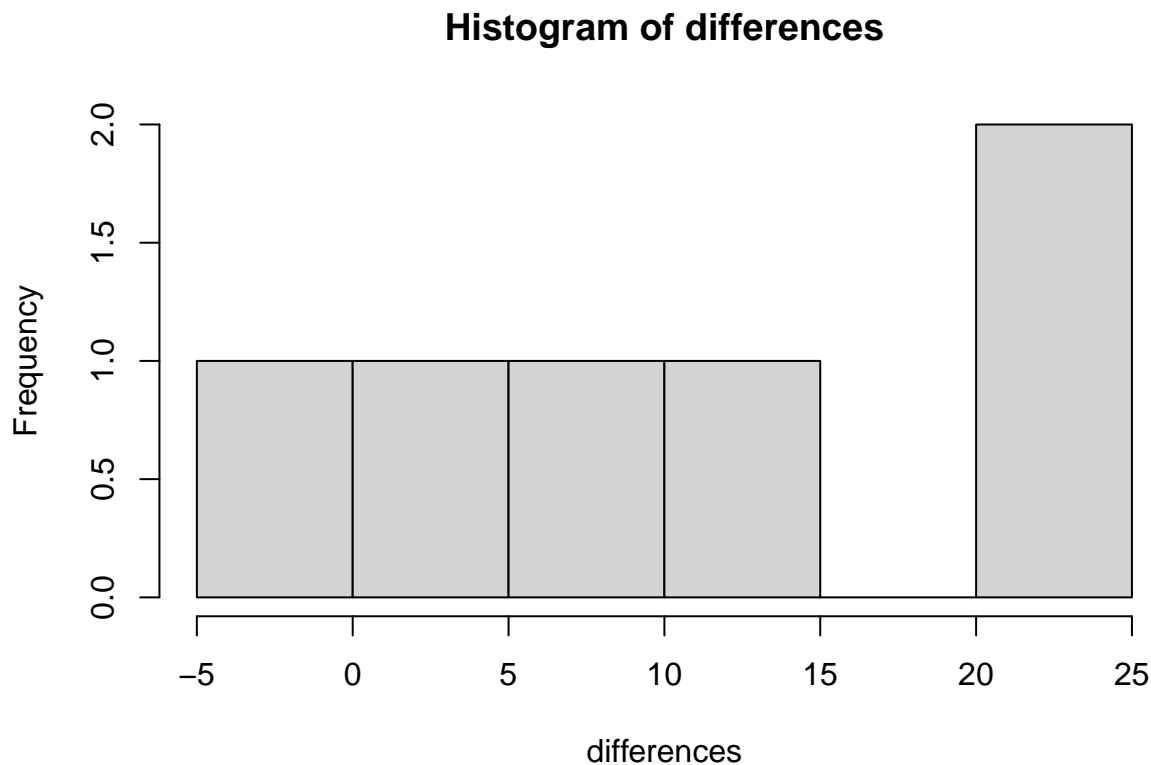


#### Null Hypothesis The null hypothesis is that the mean scores of the first test that the 6 students take is different from the mean scores of the second test that the same 6 students take at a later date.

$$H_0 : \mu_{s_1} = \mu_{s_2}$$
$$H_a : \mu_{s_1} \neq \mu_{s_2}$$

**Two Sample T-Test & Findings** Findings: Fail to reject $H_0 : \mu_1 = \mu_2$ with a p-value of 0.03721 at $\alpha = .01$

Confidence Interval (Difference of Means) : $(-4.884554, 27.551231)$

By failing to reject, we assume that there is no difference in the mean scores between the two teaching methods amongst the same students. Even so, the confidence interval suggests that the true value of the difference between teaching methods is somewhere between -4.89 and 27.55. If we were to use $\alpha = 0.95$, our findings would reject the null and would show a difference between teaching methods.

```
t.test(firstTest$Score, secondTest$Score , paired = TRUE, conf.level = 0.99, mu = 0)
```

```
##
##  Paired t-test
##
## data:  firstTest$Score and secondTest$Score
## t = 2.8177, df = 5, p-value = 0.03721
## alternative hypothesis: true difference in means is not equal to 0
## 99 percent confidence interval:
##  -4.884594 27.551261
## sample estimates:
## mean of the differences
##                11.33333
```

**Part C**

With a p-value of 0.098 we reject the null hypothesis of $\mu_1 = 80$, at $\alpha = .10$ significance.

Confidence Interval for the true mean is $(47.49, 83.84)$ at 90% confidence/

```
t.test(firstTest$Score, mu = 80)
```

```
##
##  One Sample t-test
##
## data:  firstTest$Score
## t = -2.0275, df = 5, p-value = 0.09843
## alternative hypothesis: true mean is not equal to 80
## 95 percent confidence interval:
##  47.49395 83.83939
## sample estimates:
## mean of x
##  65.66667
```

**Part D**

Both result say that there is no significant difference in the teaching methods when $\alpha = 0.01$. However, the results in part b do show a signficant difference at $\alpha = 0.05$ suggesting that the averages of the two samples are different. If we assume the calculted means are correct, then Method 1 is the better teaching method due to the higher mean.

There is very little data about how the teaching mehtods differ, the test content, how students perform in a related subject area, study time, student demographics, etc. In order to provide a better model, I think it would be useful to pull the student's class schedules, demographic information, and at least their grades from other classes in order to get a more accurate idea of what is actually influencing the student's grade.

I would also like to see multiple groups of students, one that was tested multiple times using Method 1 only and one that was retesetd multiple times using Method 2 only, and then to compare them. We can use that to get an estimate of how much students improve, or regess, at a base level for each method over time.

## Qusetion 2

### Part A

All of the p-values are below the taget level, $\alpha = 0.05$. as a result, we reject the null hypothesis that weight is normally distributed for any of the target groups.

```r
library(foreign)
wcgs <- read.dta('data/wcgs.dta')
wcgs$typchd69 = as.factor(wcgs$typchd69)
for(i in levels(wcgs$typchd69)){
    obs = wcgs[which(wcgs$typchd69 == i),]
    print(i)
    print(shapiro.test(obs$weight))
    hist(obs$weight)
}
```
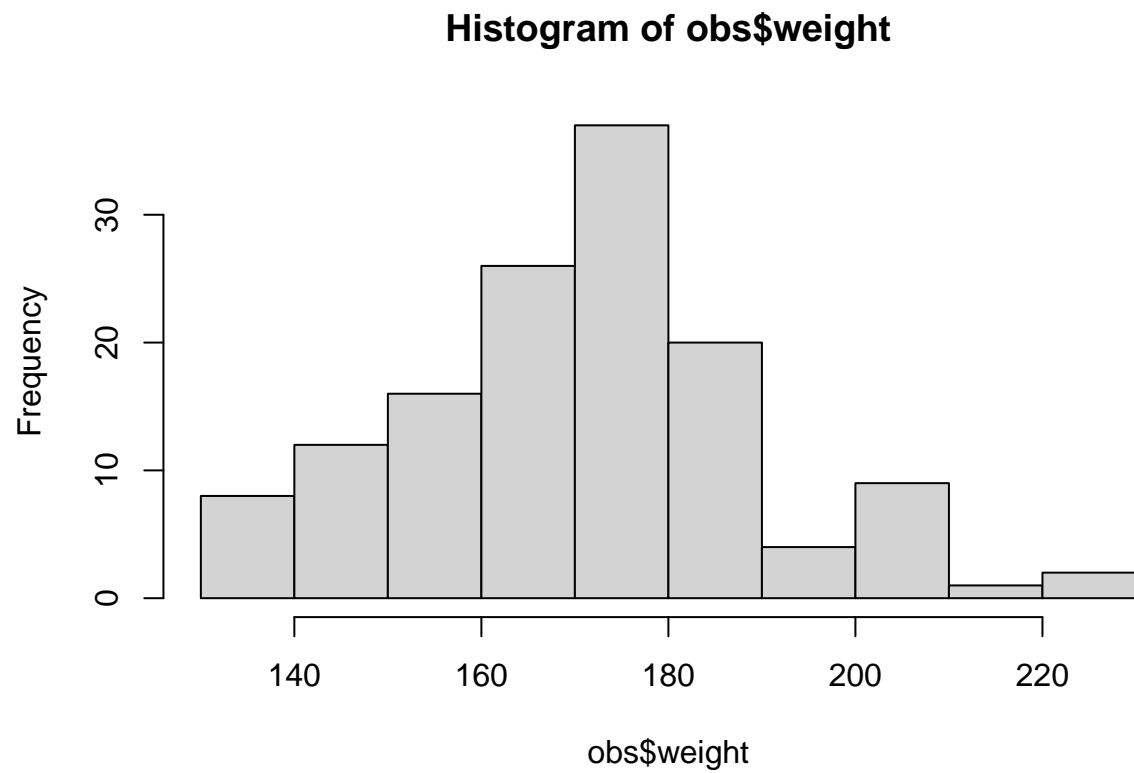
```
## [1] "0"
##
##  Shapiro-Wilk normality test
##
## data:  obs$weight
## W = 0.98321, p-value < 2.2e-16
```
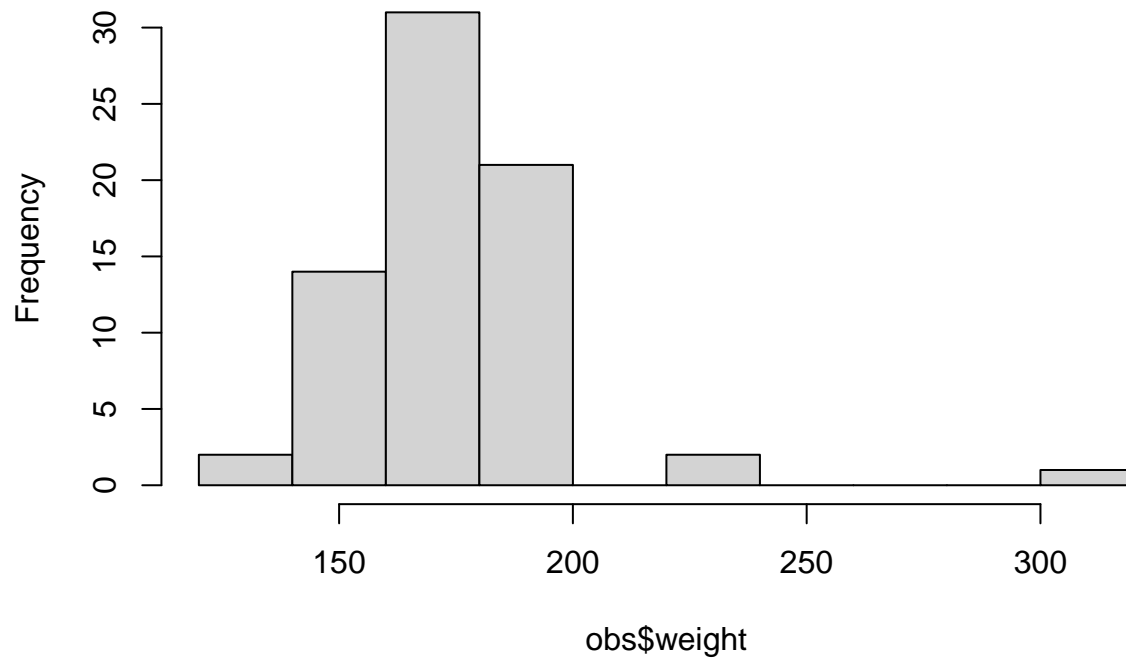
## Histogram of obs$weight
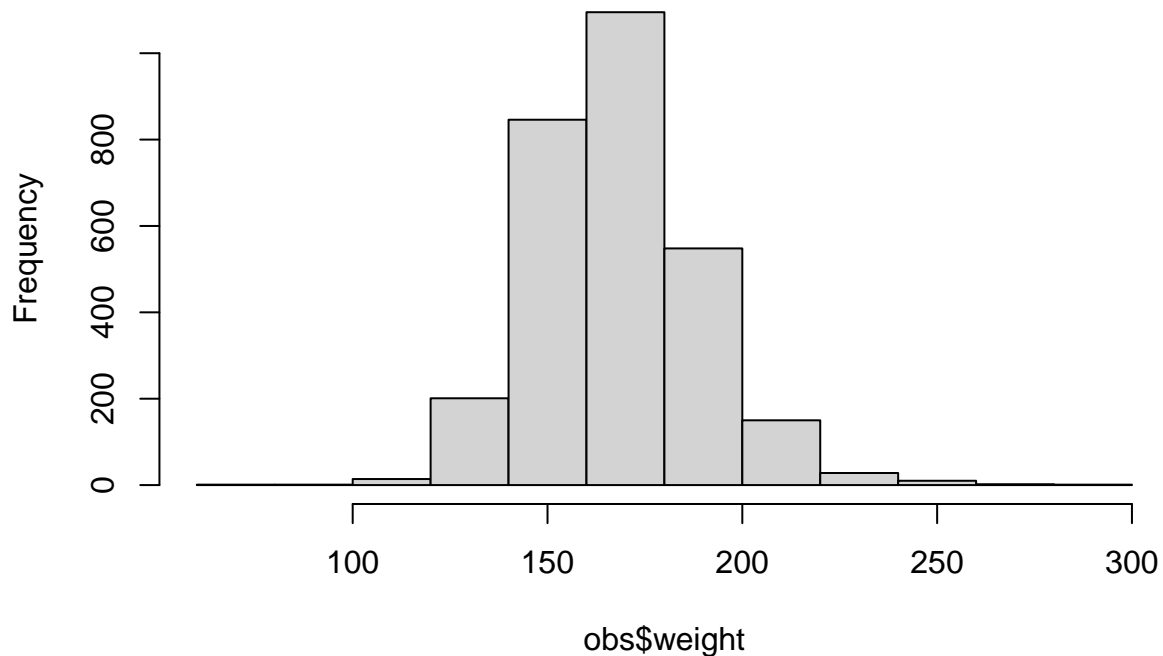


```
## [1] "1"
##
##  Shapiro-Wilk normality test
##
```

```
## data:  obs$weight
## W = 0.97991, p-value = 0.04348
```
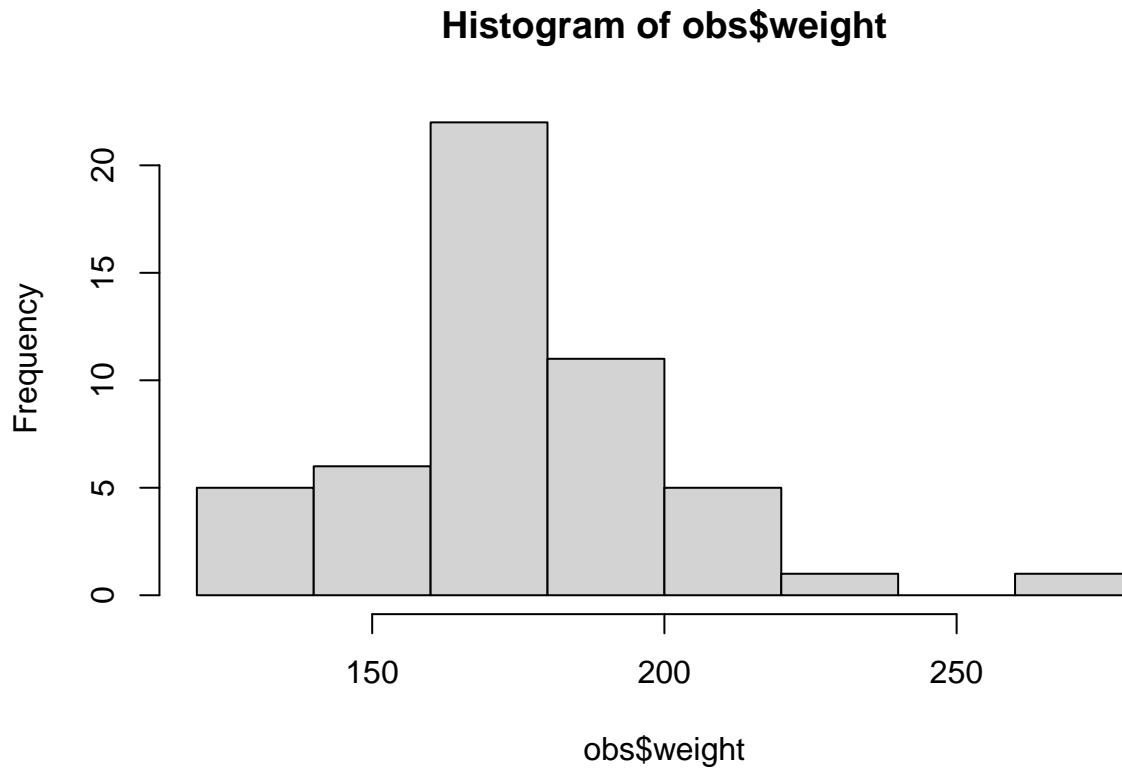
## Histogram of obs$weight



```
## [1] "2"
##
##  Shapiro-Wilk normality test
##
## data:  obs$weight
## W = 0.76603, p-value = 2.768e-09
```

# Histogram of obs$weight



```
## [1] "3"
##
##   Shapiro-Wilk normality test
##
## data:  obs$weight
## W = 0.94983, p-value = 0.031
```

# Histogram of obs$weight



## Part B

The Bartlett test returns a p-value of .03791, greater than the target level $\alpha = 0.05$, meaning we fail to reject the null and can assume the variances between the groups are homogenous.

```
bartlett.test(wcgs$weight ~ wcgs$typchd69)
```

```
##
##  Bartlett test of homogeneity of variances
##
## data:  wcgs$weight by wcgs$typchd69
## Bartlett's K-squared = 8.4303, df = 3, p-value = 0.03791
```

## Part C

**ANOVA**   With ANOVA we get a p-value of 0.0018 less than the target level $\alpha = 0.05$, which means we reject the null that there is a difference between the means of the 4 groups. This is supported by a boxplot across the 4 groups.

```
library(ggplot2)
weightAnova = aov(weight ~ typchd69, data=wcgs)
summary.aov(weightAnova)
```

```
##                Df  Sum Sq Mean Sq F value  Pr(>F)
## typchd69        3    6633  2210.9   4.987 0.00188 **
## Residuals    3150 1396551   443.3
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
ggplot(wcgs, aes(x=typchd69, y=weight)) + geom_boxplot()
```
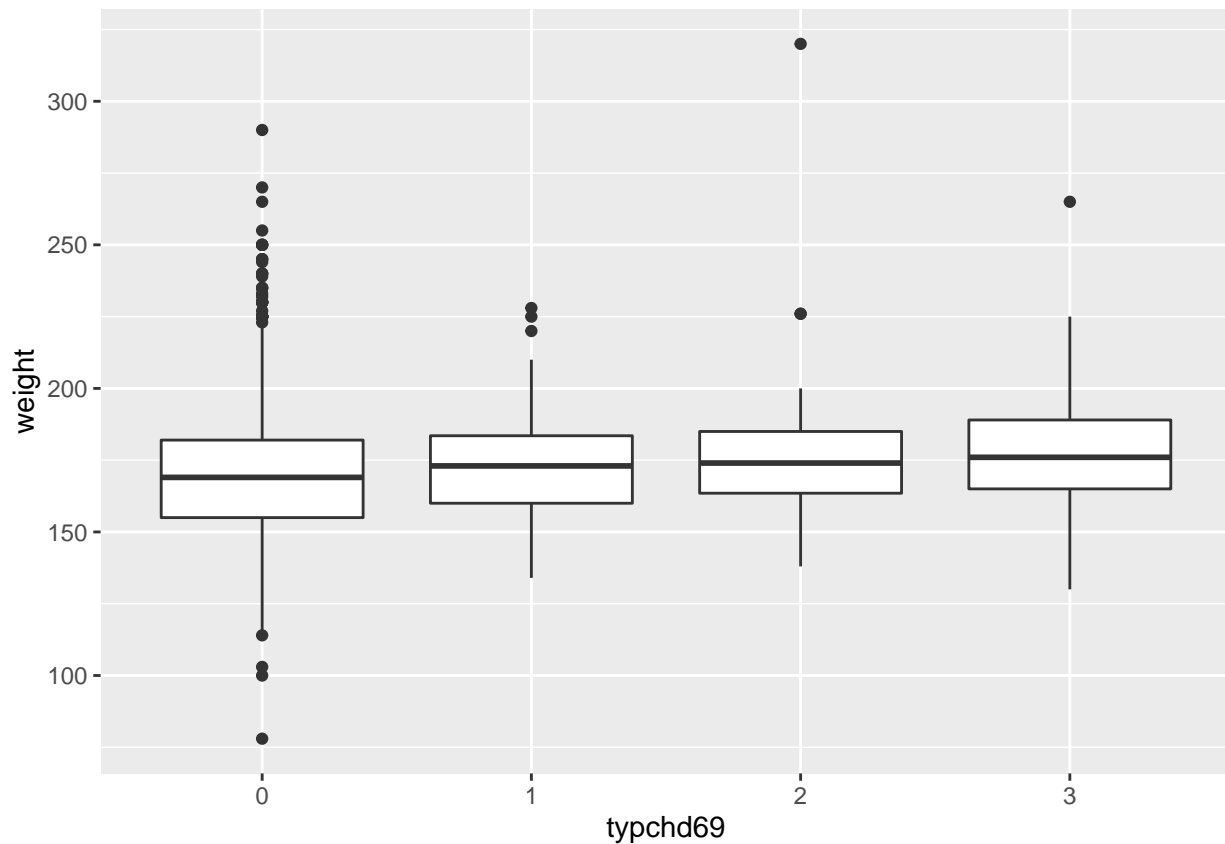


**Bonferroni**    The Bonferroni Adjusted Pairwise T-Test shows that there is only a significant difference at $\alpha = 0.05$ between groups 0 and 2 with a p value of .043

```
pairwise.t.test(wcgs$weight, wcgs$typchd69, p.adjust = "bonferroni")
```

```
##
##  Pairwise comparisons using t tests with pooled SD
##
## data:  wcgs$weight and wcgs$typchd69
##
##    0     1     2
## 1 0.574 -     -
## 2 0.043 1.000 -
## 3 0.103 1.000 1.000
##
## P value adjustment method: bonferroni
```

**Tukey's Honest Differences**    Tukey's HSD also contains only one p-value below our target level $\alpha = 0.05$, suggesting there is only a single signficant differences between the means of group 0 and group 2 across the four groups.

```
TukeyHSD(weightAnova)
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
```

```
##
## Fit: aov(formula = weight ~ typchd69, data = wcgs)
##
## $typchd69
##            diff        lwr        upr      p adj
## 1-0 3.0907682 -1.6746150  7.856151 0.3413649
## 2-0 6.7984365  0.2970962 13.299777 0.0363432
## 3-0 7.0933826 -0.5516399 14.738405 0.0801653
## 2-1 3.7076682 -4.2266921 11.642029 0.6261564
## 3-1 4.0026144 -4.8930598 12.898289 0.6543535
## 3-2 0.2949461 -9.6394167 10.229309 0.9998407
```

**Part D**

NOT DONE BECAUSE ANOVA DID NOT FAIL TO REJECT THE NULL

## Question 3

**Part A**

$$H_0 : \mu_1 = \mu_2 = \mu3 = \mu4$$

$$H_a : \mu_1 \neq \mu_2 \neq \mu3 \neq \mu4$$

```
diets <- read.table('data/diets.txt', header=TRUE)
diets$Diet = as.factor(diets$Diet)
```

**Part B**

Shapiro-Wilks returns a p-value of 0.3985, greater than our target $\alpha = 0.05$. We fail to reject the null and can assume that our weightloss data is not singificantly different from a normal distribution.

By extending the Shapiro test to look at normality across individual diet, we get even higher p-values across each group suggesting that each group is normally distributed as well.

Bartlett test for homogenous variance bewteen the different diets returns a p-value of 0.5482, greater than our target $\alpha = 0.05$. Which means that we fail to reject the null and cannot assume the variances are homogenous.

Despite this, we should still be able to conduct an ANOVA as the samples should be independently distributed across each separate diet.

```
library(RVAideMemoire)
```

```
## *** Package RVAideMemoire v 0.9-78 ***
```

```
shapiro <- shapiro.test(diets$WeightLoss)
shapiro
```

```
##
##  Shapiro-Wilk normality test
##
## data:  diets$WeightLoss
## W = 0.98818, p-value = 0.3985
```

```
shapByFactor <- byf.shapiro(WeightLoss~Diet, data = diets)
shapByFactor
```

```
##
##  Shapiro-Wilk normality tests
```

```
##
## data:  WeightLoss by Diet
##
##         W p-value
## 1 0.9759  0.8715
## 2 0.9608  0.4084
## 3 0.9681  0.4885
## 4 0.9779  0.5828
```

```
bart <- bartlett.test(WeightLoss ~ Diet, data = diets)
bart
```

```
##
##  Bartlett test of homogeneity of variances
##
## data:  WeightLoss by Diet
## Bartlett's K-squared = 2.1185, df = 3, p-value = 0.5482
```

**Part C**

Running ANOVA using tapply and mean() we get a p-value of 1.255527e-12, much less than our target $\alpha = 0.05$ which means that we reject the null that the mean between diets is the same. Essentially saying that there is a difference between the mean weight across the diets.

```
totalObservations = length(diets$WeightLoss)
totalLevels = length(levels(diets$Diet))

groupMeans <- tapply(diets$WeightLoss, diets$Diet, mean)
groupMeans
```

```
##          1         2         3         4
## 11.325000  9.492308 18.660000  0.100000
```

```
groupCounts <- tapply(diets$WeightLoss, diets$Diet, length)
groupCounts
```

```
##  1  2  3  4
## 20 26 30 42
```

```
overallMeans <- mean(groupMeans)
overallMeans
```

```
## [1] 9.894327
```

```
sst =  sum(groupCounts * ((groupMeans - overallMeans)^2))
sst
```

```
## [1] 6379.261
```

```
mst = sst/(totalLevels - 1)
mst
```

```
## [1] 2126.42
```

```
groupVariances =  tapply(diets$WeightLoss, diets$Diet, var)
groupVariances
```

```
##          1         2         3         4
##   53.64618   85.20714 100.35559   85.69415
```

```
sse = sum(groupVariances * (groupCounts - 1))
sse
```

## [1] 9573.228

```
mse = sse / (totalObservations - totalLevels)
mse
```

## [1] 83.97568

```
fStat = mst/mse
fStat
```

## [1] 25.32186

```
pf(fStat, 3, 114, lower.tail = FALSE)
```

## [1] 1.255527e-12

**Part D**

Running a Bonferroni Adjusted Pairwise comparison shows that there is a significant differences between every diet level except between Diets 1 and 2 at $\alpha = 0.05$

```
pairwise.t.test(diets$WeightLoss, diets$Diet, p.adjust = "bonferroni")
```

```
##
##  Pairwise comparisons using t tests with pooled SD
##
## data:  diets$WeightLoss and diets$Diet
##
##    1       2       3
## 2 1.00000 -       -
## 3 0.03897 0.00178 -
## 4 9.5e-05 0.00045 5.8e-13
##
## P value adjustment method: bonferroni
```

**Part E & F**

Runing the standard Anova, the Sum Squared Error value was a bit less than the Sum Squared Error value I calculated. This may be due to some additional measures taken when SSE is calculated, such as the removal of outliers. The result was a difference of 0.6 between the my FSTAT (25.32) and the FSTAT of the AOV Package (24.75). All in all, they were extremely close.

Running TukeyHSD we see significant differences between every diet level except between Diets 1 and 2 at $\alpha = 0.05$

```
dietAnova = aov(WeightLoss ~ Diet, data = diets)
summary(dietAnova )
```

```
##             Df Sum Sq Mean Sq F value   Pr(>F)
## Diet         3   6236    2078   24.75 2.09e-12 ***
## Residuals  114   9573      84
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
TukeyHSD(dietAnova)
```

```
##   Tukey multiple comparisons of means
```

```
##      95% family-wise confidence level
##
## Fit: aov(formula = WeightLoss ~ Diet, data = diets)
##
## $Diet
##             diff         lwr        upr      p adj
## 2-1  -1.832692  -8.9391049   5.273720 0.9073128
## 3-1   7.335000   0.4376455  14.232355 0.0324639
## 4-1 -11.225000 -17.7162655  -4.733734 0.0000929
## 3-2   9.167692   2.7656302  15.569754 0.0016652
## 4-2  -9.392308 -15.3546418  -3.429974 0.0004349
## 4-3 -18.560000 -24.2715522 -12.848448 0.0000000
```

**Part G**

Without the Bonferooni adjustment, we still see a signifiant difference between teh means of each diet except for diets 1 and 2 at $\alpha = 0.05$, but are drastically different compaared to teh adjustment, especially between groups 1 and 4, groups 2 and 4, groups 3 and 1, groups 2 and 1, and groups 2 and 3.

```r
pairwise.t.test(diets$WeightLoss, diets$Diet)
```

```
##
##  Pairwise comparisons using t tests with pooled SD
##
## data:  diets$WeightLoss and diets$Diet
##
##   1       2       3
## 2 0.50268 -       -
## 3 0.01299 0.00089 -
## 4 8.0e-05 0.00030 5.8e-13
##
## P value adjustment method: holm
```

**Part 4**

We have 16 total runs with 8 runs of A and 8 runs of B. When calculating the Wald Wolfowitz P Value we get p = 0.0001554 which is significantly less than our target $\alpha = 0.05$ as a result, we fail to reject the null that P(T = 16) and can assume the sequence is random.

```r
library(stringr)
stringSeq = 'aabbaaaabbbaababbaaabbbbbababbbabbb'
inputSeq = as.list(strsplit(stringSeq, "")[[1]])
convertedRuns = str_count(inputSeq, 'a') # convert to 1s and 0s to work with RLE
runLengths = rle(convertedRuns)
totalRuns = length(rle(convertedRuns)$lengths)
aChoose = choose(7,(totalRuns/2) - 1)
bChoose = choose(7,(totalRuns/2) - 1)
abChoose = choose((8+8), 8)
pv = (2 * aChoose * bChoose)/abChoose
pv
```

```
## [1] 0.0001554002
```