HW6 STAT425

Aldo Sanjoto

December 1, 2017

1a)

```
bfhs = read.table("BFHS.dat", header = TRUE)
summary(bfhs)
```

```
##           Town    Intervention   ExternalComparison
##  Bridgend   :1   Min.   :5.370   Min.   :5.415
##  Burton     :1   1st Qu.:5.530   1st Qu.:5.619
##  Bury       :1   Median :5.574   Median :5.702
##  Carlisle   :1   Mean   :5.574   Mean   :5.685
##  Darlington :1   3rd Qu.:5.629   3rd Qu.:5.742
##  Dunfermline:1   Max.   :5.812   Max.   :6.067
##  (Other)    :7
```

```
dif = bfhs$Intervention - bfhs$ExternalComparison
t.test(dif)
```

```
##
##  One Sample t-test
##
## data:  dif
## t = -2.0702, df = 12, p-value = 0.06067
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  -0.228613888  0.005844657
## sample estimates:
##  mean of x
## -0.1113846
```

1b) No significance difference between the intervention and external comparison as the p-value is greater than 0.05

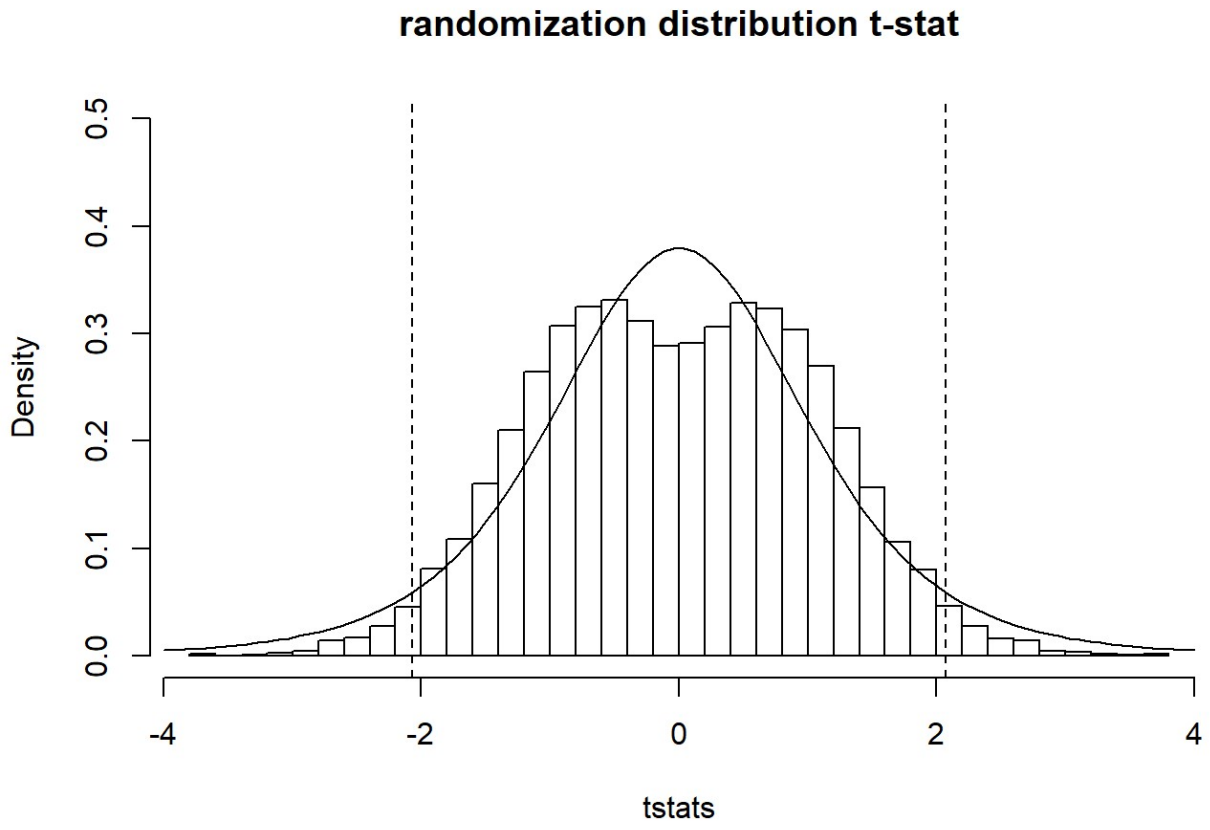
```
tstats = replicate(100000, t.test(dif * sample(c(-1,1),13,replace=TRUE))$statistic)
tobserved = t.test(dif)$statistic
approx_pval = mean(abs(tstats) >= abs(tobserved))
approx_pval
```

```
## [1] 0.03881
```

1c) There's a significant difference between the intervention and external comparison as the p-value is lesser than 0.05.

1d)

```
hist(tstats, main = "randomization distribution t-stat", freq = FALSE, breaks = 30, yl
im = c(0, 0.5))
lines(seq(-4,4,len=100), dt(seq(-4,4,len=100), df = 5))
abline(v = c(tobserved,-tobserved), lty=2)
```




```
barley = read.csv("Barley1928.csv")
table(barley$Block)
```

```
##
## A B C D E F
## 5 5 5 5 5 5
```

2a) There are 5 blocks with a total of 30 experimental units.

2b)

```
barley$Treatment = as.factor(barley$Treatment)
fit = lm(formula = Yield ~ Block + Treatment, data = barley)
summary(fit)
```

```
##
## Call:
## lm(formula = Yield ~ Block + Treatment, data = barley)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -42.57 -12.37  -4.10  15.58  36.63
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   222.77      14.80  15.051 2.26e-12 ***
## BlockB          0.40      16.21   0.025 0.980562
## BlockC         11.40      16.21   0.703 0.490078
## BlockD         79.20      16.21   4.885 8.96e-05 ***
## BlockE         57.80      16.21   3.565 0.001940 **
## BlockF         64.60      16.21   3.984 0.000730 ***
## Treatment2     61.67      14.80   4.167 0.000477 ***
## Treatment3     92.67      14.80   6.261 4.10e-06 ***
## Treatment4    115.83      14.80   7.826 1.63e-07 ***
## Treatment5    128.00      14.80   8.648 3.43e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 25.64 on 20 degrees of freedom
## Multiple R-squared:  0.8779, Adjusted R-squared:  0.823
## F-statistic: 15.98 on 9 and 20 DF,  p-value: 2.662e-07
```

```
anova(fit)
```

```
## Analysis of Variance Table
##
## Response: Yield
##           Df Sum Sq Mean Sq F value    Pr(>F)
## Block      5  31634  6326.8  9.6274 8.458e-05 ***
## Treatment  4  62903 15725.9 23.9298 2.205e-07 ***
## Residuals 20  13143   657.2
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

2c) Treatment effects are statistically significant because p-value < 0.05

2d)

```
TukeyHSD(aov(fit))$Treatment
```

```
##          diff        lwr       upr        p adj
## 2-1  61.66667  17.377951 105.95538 3.874425e-03
## 3-1  92.66667  48.377951 136.95538 3.673806e-05
## 4-1 115.83333  71.544617 160.12205 1.501867e-06
## 5-1 128.00000  83.711284 172.28872 3.175531e-07
## 3-2  31.00000 -13.288716  75.28872 2.607787e-01
## 4-2  54.16667   9.877951  98.45538 1.201739e-02
## 5-2  66.33333  22.044617 110.62205 1.900159e-03
## 4-3  23.16667 -21.122049  67.45538 5.349191e-01
## 5-3  35.33333  -8.955383  79.62205 1.597099e-01
## 5-4  12.16667 -32.122049  56.45538 9.206894e-01
```

2e) Based on the Tukey intervals, 2-1, 3-1, 4-1, 5-1, 4-2, 5-2 are the pairs of treatments that have significantly different means

```
spelling = read.csv("Spelling1941.csv")
```

3a) the 2 blocking factors are “List” and “Group”

```
matrix(spelling$Testing, nrow = 4, ncol = 4)
```

```
##      [,1] [,2] [,3] [,4]
## [1,] "MC" "SW" "WS" "SD"
## [2,] "SD" "MC" "SW" "WS"
## [3,] "WS" "SD" "MC" "SW"
## [4,] "SW" "WS" "SD" "MC"
```

3b) Column = List, Row = Group

3c)

```
fit_2 = lm(formula = Number ~ List + factor(Group) + Testing, data = spelling)
summary(fit_2)
```

```
##
## Call:
## lm(formula = Number ~ List + factor(Group) + Testing, data = spelling)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -10.000  -3.500  -1.125   4.188  12.250
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)      83.500      7.948  10.505 4.37e-05 ***
## ListB             1.750      7.109   0.246  0.81376
## ListC           -10.500      7.109  -1.477  0.19016
## ListD            -1.250      7.109  -0.176  0.86621
## factor(Group)2    1.750      7.109   0.246  0.81376
## factor(Group)3   -2.750      7.109  -0.387  0.71224
## factor(Group)4    3.000      7.109   0.422  0.68774
## TestingSD       -37.500      7.109  -5.275  0.00187 **
## TestingSW       -37.250      7.109  -5.240  0.00194 **
## TestingWS       -42.250      7.109  -5.943  0.00101 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 10.05 on 6 degrees of freedom
## Multiple R-squared:  0.893,  Adjusted R-squared:  0.7324
## F-statistic: 5.563 on 9 and 6 DF,  p-value: 0.02458
```

```
anova(fit_2)
```

```
## Analysis of Variance Table
##
## Response: Number
##               Df Sum Sq Mean Sq F value   Pr(>F)
## List           3  359.5  119.83  1.1855 0.391361
## factor(Group)  3   74.5   24.83  0.2457 0.861666
## Testing        3 4626.5 1542.17 15.2564 0.003254 **
## Residuals      6  606.5  101.08
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

3d) Treatment effects are statistically significant because p-value < 0.05

3e)

```
TukeyHSD(aov(fit_2))$Testing
```

```
##         diff       lwr       upr       p adj
## SD-MC -37.50 -62.11023 -12.88977 0.007497477
## SW-MC -37.25 -61.86023 -12.63977 0.007749362
## WS-MC -42.25 -66.86023 -17.63977 0.004107668
## SW-SD   0.25 -24.36023  24.86023 0.999982530
## WS-SD  -4.75 -29.36023  19.86023 0.905385377
## WS-SW  -5.00 -29.61023  19.61023 0.892265654
```

3f) Based on the Tukey intervals, SD-MC, SW-MC, WS-MC are the pairs of treatments that have significantly different means