Homework 6

Name: Zixin Ouyang

1.

a.

```
> BFHS<-read.table("/Users/Constance/Desktop/BFHS.dat",header=TRUE,skip=6)
> BFHS
           Town Intervention ExternalComparison
1    Portsmouth        5.574              5.742
2    Darlington        5.812              5.831
3    Gloucester        5.651              5.714
4      Carlisle        5.438              6.067
5        Burton        5.595              5.702
6       Lincoln        5.530              5.717
7   Dunfermline        5.786              5.621
8      Bridgend        5.553              5.415
9          Bury        5.533              5.624
10 Huddersfield        5.594              5.752
11      Ipswich        5.391              5.585
12      Newport        5.629              5.619
13        Poole        5.370              5.515
> 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
```

b.

```
> t.test(BFHS$Intervention-BFHS$ExternalComparison,var.equal=TRUE)

	One Sample t-test

data:  BFHS$Intervention - BFHS$ExternalComparison
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
```

The p-value of t-test is 0.06067, which is larger than 0.05. So we conclude that there is no significant mean difference in the cholesterol levels of the intervention and comparison groups.

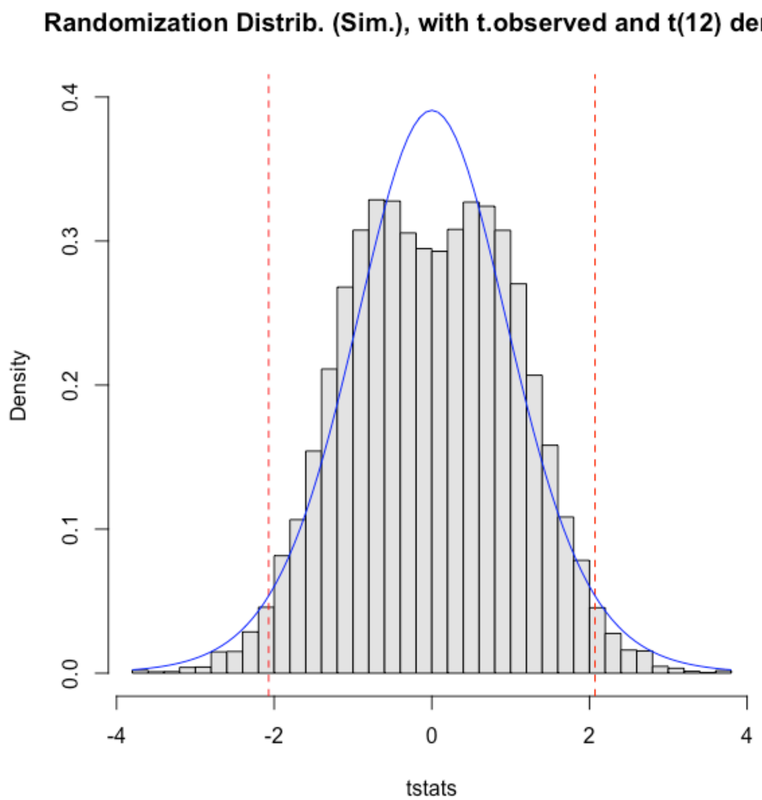
c.

```
> tstats<-replicate(100000,t.test((BFHS$Intervention-BFHS$ExternalComparison)*sample(c(-1,1),
13,replace=TRUE))$statistic)
> t.observed<-t.test(BFHS$Intervention-BFHS$ExternalComparison)$statistic
> approx.pval<-mean(abs(tstats)>=abs(t.observed))
> approx.pval
[1] 0.03922
```

The p-value is 0.03922, which is less than 0.05. So we conclude that there is significant mean difference in the cholesterol levels of the intervention and comparison groups. The conclusion change.

d.

```
> hist(tstats, breaks=50, freq=FALSE, col="grey90", ylim=c(0,0.4),main="Randomization Distrib
. (Sim.), with t.observed and t(12) density")
> abline(v=c(t.observed,-t.observed), lty=2, col="red")
> curve(dt(x, df=12), add=TRUE, col="blue")
```



2.

a.

```
> Barley<-read.csv("/Users/Constance/Desktop/Barley1928.csv",header=TRUE)
> table(Barley$Block)

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

There are six blocks: A, B, C, D, E, F, and there are 30 experimental units in total.

b.

```
> class(Barley$Block)
[1] "factor"
> class(Barley$Treatment)
[1] "integer"
```

The treatment is an integer variable, so we use factor(Treatment) when fitting the linear model.

```
> Barleymod<-lm(Yield~Block+factor(Treatment), data=Barley)
> summary(Barleymod)

Call:
lm(formula = Yield ~ Block + factor(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 ***
factor(Treatment)2    61.67      14.80   4.167 0.000477 ***
factor(Treatment)3    92.67      14.80   6.261 4.10e-06 ***
factor(Treatment)4   115.83      14.80   7.826 1.63e-07 ***
factor(Treatment)5   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
```

c.

```
> anova(Barleymod)
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 ***
factor(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
```

The p-value of treatment is 2.205e-07, which is less than 0.05, so treatment effects are significant.

d.

```
> TukeyHSD(aov(Yield~Block+factor(Treatment), data=Barley))$"factor(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
```

e. According to the Tukey intervals above, pairs (2,1), (3,1), (4,1), (5,1), (4, 2) and (5, 2) have significantly different means.

3.

a.

```
> Spelling<-read.csv("/Users/Constance/Desktop/Spelling1941.csv",header=TRUE)
> Spelling
   List Group Testing Number
1     A     1      MC     81
2     A     2      SD     41
3     A     3      WS     44
4     A     4      SW     53
5     B     1      SW     38
6     B     2      MC     97
7     B     3      SD     42
8     B     4      WS     49
9     C     1      WS     31
10    C     2      SW     43
11    C     3      MC     67
12    C     4      SD     36
13    D     1      SD     57
14    D     2      WS     33
15    D     3      SW     43
16    D     4      MC     81
```

The two blocking factors: List and Group.

b.

```
> matrix(Spelling$Testing,4,4)
     [,1] [,2] [,3] [,4]
[1,] "MC" "SW" "WS" "SD"
[2,] "SD" "MC" "SW" "WS"
[3,] "WS" "SD" "MC" "SW"
[4,] "SW" "WS" "SD" "MC"
```

List corresponds to row, and Group corresponds to column.

c.

```
> class(Spelling$List)
[1] "factor"
> class(Spelling$Group)
[1] "integer"
> class(Spelling$Testing)
[1] "factor"
```

The Group is an integer variable, so we use factor(Group) when fitting the linear model.

```
> Spellingmod<-lm(Number~List+factor(Group)+Testing, data=Spelling)
> summary(Spellingmod)

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
```

d.

```
> anova(Spellingmod)
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
```

The p-value of testing is 0.003254, which is less than 0.05, so treatment effects are statistically significant.

e.

```
> TukeyHSD(aov(Number~List+factor(Group)+Testing, data=Spelling))$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
```

f. According to the Tukey intervals above, pairs (SD, MC), (SW, MC), and (WS, MC) have significantly different means.

4.

(a) P (I chosen) = P (II chosen) = $\frac{1}{\frac{16!}{(4!)^4}} = \frac{1}{63063000}$

(b) P (I chosen) = $\frac{1}{(4!)^4} = \frac{1}{331776}$, P (II chosen) =0

(c) P (I chosen) = P(II chosen) = $\frac{1}{(4!)^4} = \frac{1}{331776}$

(d) Assignment I is a Latin square design, but Assignment II is not.

P (I chosen) = $\frac{1}{576}$, P (II chosen) =0