Stat 424: Homework 4

Evangeline Lim

11/1/2019

Question 1 a) and b)

```
## Warning in read.table("weight.txt", header = F): incomplete final line
## found by readTableHeader on 'weight.txt'
```

```
##
##  One Sample t-test
##
## data:  scale_I - scale_II
## t = -3.0813, df = 5, p-value = 0.02743
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  -3.9742161 -0.3591172
## sample estimates:
## mean of x
## -2.166667
```

```
## [1] 2.776445
```

```
## [1] TRUE
```

```
## [1] 0.03688492
```

$H_0 = t_I = t_{II}$ Given that the t_{paired} is greater than the t_{crit}, we reject the null hypothesis. Scale I does not give the same measurements as Scale II. The p-value is 0.03688492.

Question 2

```
## Analysis of Variance Table
##
## Response: r
##           Df  Sum Sq Mean Sq F value   Pr(>F)
## treatment  5 135.417 27.0833 18.2584 0.003154 **
## block      1  14.083 14.0833  9.4944 0.027429 *
## Residuals  5   7.417  1.4833
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
## [1] 9.4944
```

```
## [1] f= 9.4944         t-paired squared= 9.49440969
```

We see that from the previous question 1, the f statistic from the ANOVA is approximately the same as t-paired squared. Hence proven.

***2.** Prove the equivalence of the F-test statistic in the ANOVA for the paired comparison experiment and the square of the paired t statistic given in (3.1).

$$F = \frac{MSB}{MSR} = \frac{SSB/(k-1)}{SSR/(N-k)}$$

$$= \frac{SSB(k-1)/\hat{\sigma}^2(k-1)^2}{SSR(N-k)/\hat{\sigma}^2(N-k)^2}$$

For paired, $k=2$

$$= \frac{SSB/\hat{\sigma}^2}{SSR(N-2)/\hat{\sigma}^2(N-2)^2}$$

The numerator is a χ_1^2 distribution with $df = 1$ so, $SSB/\hat{\sigma}^2 \sim \chi_1^2$

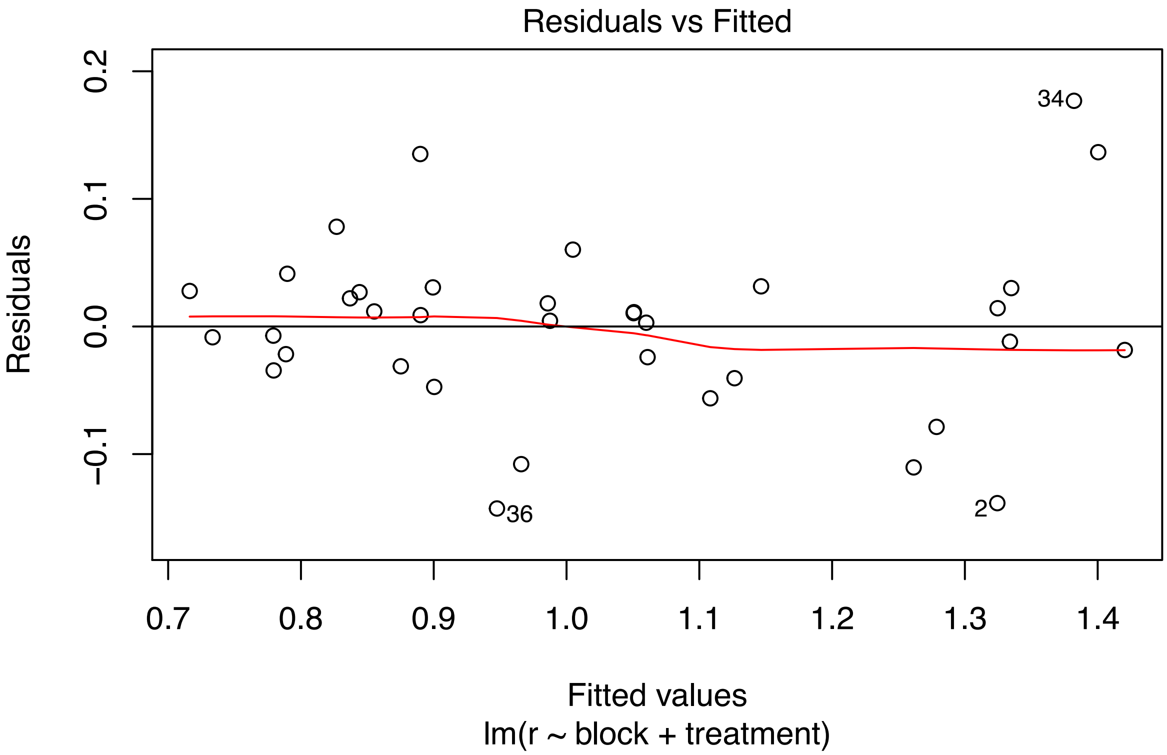
and $SSR(N-2)/\hat{\sigma}^2(N-2)^2 \sim \dfrac{\chi^2_{N-2}}{(N-2)^2}$

$\therefore$ We have the ratio of two χ^2 distributions

$$\frac{\chi_1^2}{\chi^2_{N-2}/(N-2)^2} \sim t^2_{N-2}$$

// proven

Question 4



The residuals plot shows that the spread of the plots above and below the (0,0) line are relatively equal but there is a large variance in the plots and a few outliers namely 2,34,36. Homoskedacity is also violated because the variance at the two ends are more than the variance at the middle.

```
## Analysis of Variance Table
##
## Response: r
##           Df  Sum Sq  Mean Sq F value    Pr(>F)
## block      8 0.49958 0.062448  7.2791 4.902e-07 ***
## treatment  9 1.82549 0.202832 23.6429 < 2.2e-16 ***
## Residuals 72 0.61769 0.008579
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
## [1] TRUE
```

$H_0 = \tau_1 = ... = \tau_k$ where k = 1,..10.

Since F test is smaller than the F critical, we reject the null hypothesis that the four methods produce the same sheer strength. So, we proceed to Tukey multiple comparisons method.

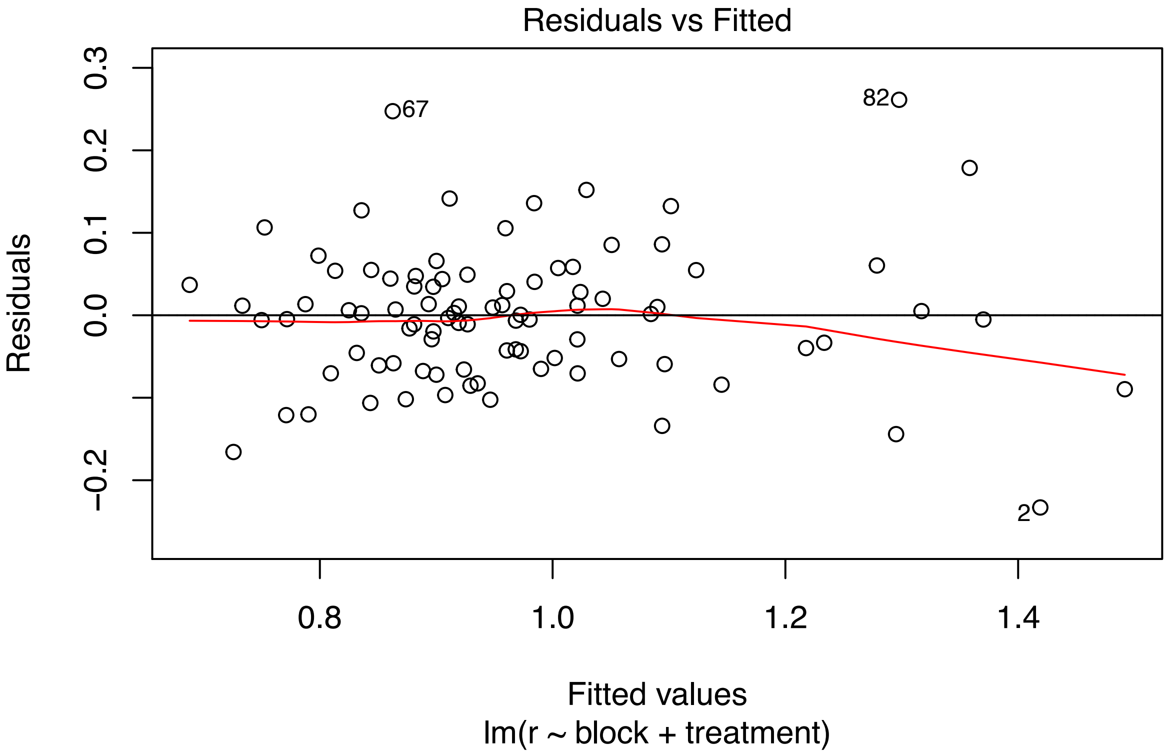
```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = r ~ block + treatment)
##
## $block
##        diff          lwr           upr     p adj
## 2-1 -0.1239 -0.256370081  0.0085700811 0.0849755
## 3-1 -0.1021 -0.234570081  0.0303700811 0.2665442
```

```
## 4-1 -0.1404 -0.272870081 -0.0079299189 0.0295456
## 5-1 -0.1857 -0.318170081 -0.0532299189 0.0008702
## 6-1  0.0726 -0.059870081  0.2050700811 0.7121220
## 7-1 -0.0489 -0.181370081  0.0835700811 0.9581095
## 8-1 -0.0607 -0.193170081  0.0717700811 0.8674905
## 9-1 -0.1213 -0.253770081  0.0111700811 0.0989720
## 3-2  0.0218 -0.110670081  0.1542700811 0.9998337
## 4-2 -0.0165 -0.148970081  0.1159700811 0.9999801
## 5-2 -0.0618 -0.194270081  0.0706700811 0.8556260
## 6-2  0.1965  0.064029919  0.3289700811 0.0003391
## 7-2  0.0750 -0.057470081  0.2074700811 0.6750968
## 8-2  0.0632 -0.069270081  0.1956700811 0.8397064
## 9-2  0.0026 -0.129870081  0.1350700811 1.0000000
## 4-3 -0.0383 -0.170770081  0.0941700811 0.9908034
## 5-3 -0.0836 -0.216070081  0.0488700811 0.5361693
## 6-3  0.1747  0.042229919  0.3071700811 0.0021943
## 7-3  0.0532 -0.079270081  0.1856700811 0.9327336
## 8-3  0.0414 -0.091070081  0.1738700811 0.9847874
## 9-3 -0.0192 -0.151670081  0.1132700811 0.9999364
## 5-4 -0.0453 -0.177770081  0.0871700811 0.9733822
## 6-4  0.2130  0.080529919  0.3454700811 0.0000759
## 7-4  0.0915 -0.040970081  0.2239700811 0.4112260
## 8-4  0.0797 -0.052770081  0.2121700811 0.5997617
## 9-4  0.0191 -0.113370081  0.1515700811 0.9999389
## 6-5  0.2583  0.125829919  0.3907700811 0.0000010
## 7-5  0.1368  0.004329919  0.2692700811 0.0376668
## 8-5  0.1250 -0.007470081  0.2574700811 0.0795704
## 9-5  0.0644 -0.068070081  0.1968700811 0.8253524
## 7-6 -0.1215 -0.253970081  0.0109700811 0.0978322
## 8-6 -0.1333 -0.265770081 -0.0008299189 0.0473943
## 9-6 -0.1939 -0.326370081 -0.0614299189 0.0004266
## 8-7 -0.0118 -0.144270081  0.1206700811 0.9999985
## 9-7 -0.0724 -0.204870081  0.0600700811 0.7151456
## 9-8 -0.0606 -0.193070081  0.0718700811 0.8685406
```

```
## [1] Pairs that are significant:
```

```
## [1] 4-1, 5-1, 6-2, 6-3, 6-4, 6-5, 7-5, 9-6, 8-6
```

```r
##Question 6 Residuals
plot(lm, which=1)
abline(0,0)
```



From the residuals plot there are a few outliers but the spread above and below the y=0 line seems even with quite a bit of variability.

Question 12

```
##    P.O C.W HT M.B
## 1   10  24 32   M
## 2   13  18 22   M
## 3   17  17 30   M
## 4   16  17 35   M
## 5   15  15 32   M
## 6   14  23 28   M
## 7   11  14 27   M
## 8   14  18 28   M
## 9   15  12 30   M
## 10  16  11 30   M
## 11  25  20 26   B
## 12  40  16 40   B
## 13  30  17 28   B
## 14  17  18 38   B
## 15  16  15 38   B
## 16  45  16 30   B
## 17  49  19 26   B
## 18  33  14 38   B
## 19  30  15 45   B
## 20  20  24 38   B
```

```
##   M.B HT
## 1   M 32
```

```
## 2    M 22
## 3    M 30
## 4    M 35
## 5    M 32
## 6    M 28
## 7    M 27
## 8    M 28
## 9    M 30
## 10   M 30

##     M.B C.W
## 1    M  24
## 2    M  18
## 3    M  17
## 4    M  17
## 5    M  15
## 6    M  23
## 7    M  14
## 8    M  18
## 9    M  12
## 10   M  11

##     M.B P.O
## 1    M  10
## 2    M  13
## 3    M  17
## 4    M  16
## 5    M  15
## 6    M  14
## 7    M  11
## 8    M  14
## 9    M  15
## 10   M  16

##     M.B HT
## 11   B 26
## 12   B 40
## 13   B 28
## 14   B 38
## 15   B 38
## 16   B 30
## 17   B 26
## 18   B 38
## 19   B 45
## 20   B 38

##     M.B C.W
## 11   B  20
## 12   B  16
## 13   B  17
## 14   B  18
## 15   B  15
## 16   B  16
## 17   B  19
## 18   B  14
## 19   B  15
```

```
## 20   B  24
```

```
##    M.B P.O
## 11   B  25
## 12   B  40
## 13   B  30
## 14   B  17
## 15   B  16
## 16   B  45
## 17   B  49
## 18   B  33
## 19   B  30
## 20   B  20
```

```
## [1] z_m_ht =         2.50688555650678
```

```
## [1] z_m_cw =         2.88355847666212
```

```
## [1] z_m_po =         1.60721321741199
```

```
## [1] z_b_ht =         3.78444212729165
```

```
## [1] z_b_cw =        2.1897895988487
```

```
## [1] z_b_po =         4.87816154302987
```

We see that B-PO combination has the largest variance followed by B-HT, M-CW, M-HT, B-CW, M-PO. This is in line with the boxplot's spread of each combination.

Question 13

We want to let $\alpha_1 = 0$ and $\beta_1 = 0$

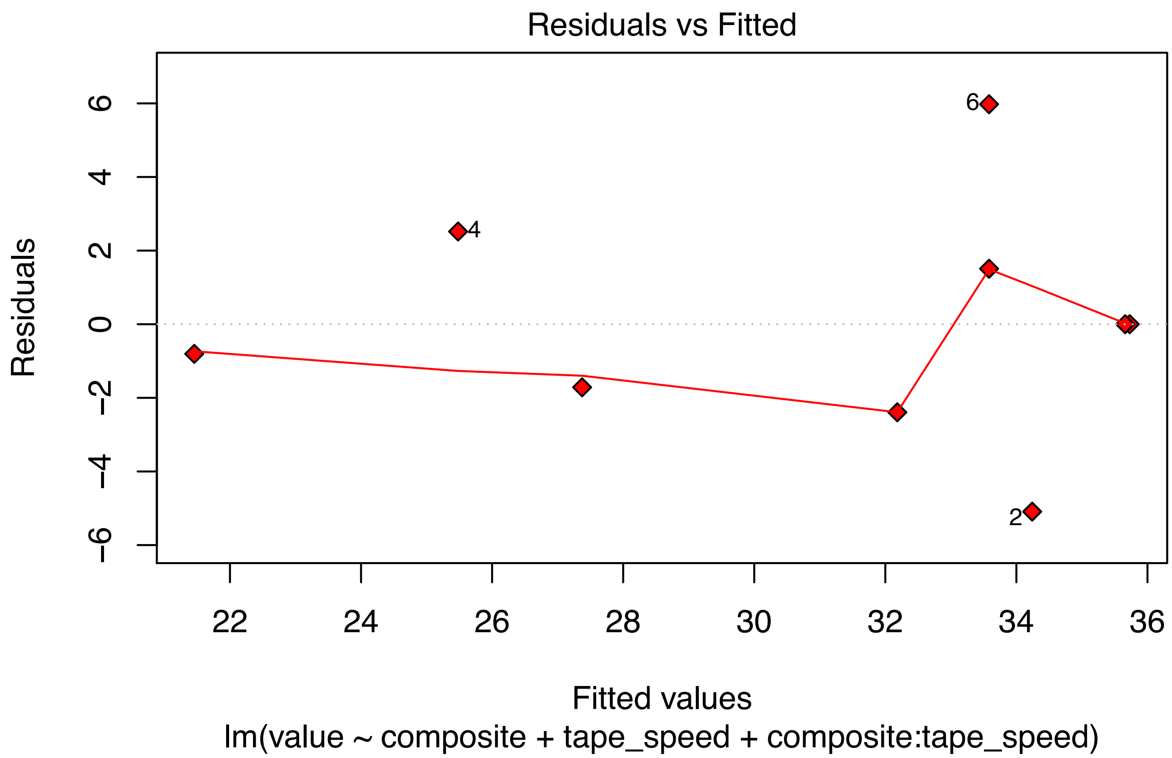
```
## Analysis of Variance Table
##
## Response: value
##                      Df  Sum Sq Mean Sq F value Pr(>F)
## composite             2 183.507  91.754  3.4594 0.1663
## tape_speed            1  11.449  11.449  0.4317 0.5581
## composite:tape_speed  2   9.080   4.540  0.1712 0.8504
## Residuals             3  79.569  26.523
```

```
## Warning in model.matrix.default(mt, mf, contrasts): variable 'f' is absent,
## its contrast will be ignored
```

```
##
## Call:
## lm(formula = cts$value ~ cts$composite + cts$tape_speed + cts$composite:cts$tape_speed,
##     contrasts = list(f = "contr.sum"))
##
## Residuals:
##          1          2          3          4          5          6
## -1.713e+00 -5.091e+00 -2.776e-16  2.519e+00  1.507e+00  5.977e+00
##          7          8          9
## -8.053e-01 -2.393e+00  3.886e-16
##
## Coefficients:
##                                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)                      29.2196     6.1304   4.766   0.0175 *
```

```
## cts$composite50W                  5.6636     8.3764   0.676   0.5474
## cts$composite60W                  6.5322     9.2637   0.705   0.5315
## cts$tape_speed                   -0.2876     0.3465  -0.830   0.4674
## cts$composite50W:cts$tape_speed   0.1876     0.4875   0.385   0.7261
## cts$composite60W:cts$tape_speed   0.2842     0.4953   0.574   0.6063
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.15 on 3 degrees of freedom
## Multiple R-squared:  0.7194, Adjusted R-squared:  0.2518
## F-statistic: 1.539 on 5 and 3 DF,  p-value: 0.3838
```

```
##   (Intercept) cts$composite50W cts$composite60W cts$tape_speed
## 1           1                0                0           6.42
## 2           1                1                0           6.42
## 3           1                0                1           6.42
## 4           1                0                0          13.00
## 5           1                1                0          13.00
## 6           1                1                0          13.00
## 7           1                0                0          27.00
## 8           1                1                0          27.00
## 9           1                0                1          27.00
##   cts$composite50W:cts$tape_speed cts$composite60W:cts$tape_speed
## 1                            0.00                            0.00
## 2                            6.42                            0.00
## 3                            0.00                            6.42
## 4                            0.00                            0.00
## 5                           13.00                            0.00
## 6                           13.00                            0.00
## 7                            0.00                            0.00
## 8                           27.00                            0.00
## 9                            0.00                           27.00
## attr(,"assign")
## [1] 0 1 1 2 3 3
## attr(,"contrasts")
## attr(,"contrasts")$`cts$composite`
## [1] "contr.treatment"
```



There are too little samples to properly interpret this residual plot. However, the varaince looks very large and it violates linearity seeing how the points are scattered all over the place. It is hard to determine for sure whether the point 6 and 2 are outliers or not since the sample is so small.

Appendix

```
##Question 1
data = read.table("weight.txt", header=F)
##a)
scale_I = data[1,]
scale_II = data[2,]
t.test(scale_I-scale_II)
t = -3.0813

(tcrit = qt(1-0.025,df = 4))

abs(t)>tcrit

##b)
(p = 2*pt(-abs(t),df=4))
```

```
##Question 2

b=2
k=6
f <- c(1,2,3,4,5,6)
r <- c(t(as.matrix(data)))
treatment = gl(k, 1, b*k, factor(f))
block = gl(b, k, b*k)
```

```r
lm <- lm(r ~ treatment + block)

anova(lm)

#From ANOVA
(F_test = 9.4944)

a <- c("f= 9.4944", "t-paired squared=", t^2 )
print(a, quote=F)
```

```r
##Question 4
data4 = read.table("girder.txt", header = T)

drops <- "Girder"
newData4 = data4[ , !(names(data4) %in% drops)]
b = nrow(data4)
r = c(t(as.matrix(newData4)))
f <-c(colnames(newData4))
k = length(f)

treatment = gl(k, 1, b*k, factor(f))
block = gl(b, k, b*k)

lm <- lm(r ~ block + treatment)
plot(lm, which=1)
abline(0,0)
```

```r
##Question 6

data6 = read.csv("fullgirder.csv", header = T)

drops <- "Girder"
newData6 = data6[ , !(names(data6) %in% drops)]
b = nrow(data6)
r = c(t(as.matrix(newData6)))
f <-c(colnames(newData6))
k = length(f)

treatment = gl(k, 1, b*k, factor(f))
block = gl(b, k, b*k)

lm <- lm(r ~ block + treatment)
anova(lm)


f_test = 23.6429

f_crit = qf(1-0.05,k-1,(b-1)*(k-1))

f_test > f_crit
```

```r
##Question 6 Tukey
av <- aov(r~ block + treatment)
```

```
mc <- TukeyHSD(x=av, 'block', conf.level=0.95)
mc

sig <- c("4-1,","5-1,", "6-2,", "6-3,", "6-4,","6-5,","7-5,","9-6,","8-6")

print("Pairs that are significant:", quote=F)
print(sig, quote=F)
```

```
##Question 6 Residuals
plot(lm, which=1)
abline(0,0)
```

```
##Question 12
data12 <- read.table("bolt.txt", header=T)
data12
m_ht<- data12[1:10,] %>% select(M.B, HT)
m_ht
m_cw <- data12[1:10,] %>% select(M.B, C.W)
m_cw
m_po <- data12[1:10,] %>% select(M.B, P.O)
m_po
b_ht<- data12[11:20,] %>% select(M.B, HT)
b_ht
b_cw <- data12[11:20,] %>% select(M.B, C.W)
b_cw
b_po <- data12[11:20,] %>% select(M.B, P.O)
b_po

z_m_ht = c("z_m_ht = ", log((sd(m_ht$HT))^2))
print(z_m_ht,quote=F)
z_m_cw = c("z_m_cw = ", log((sd(m_cw$C.W))^2))
print(z_m_cw,quote=F)
z_m_po = c("z_m_po =",log(((sd(m_po$P.O))^2)))
print(z_m_po,quote=F)
z_b_ht = c("z_b_ht =",log((sd(b_ht$HT))^2))
print(z_b_ht,quote=F)
z_b_cw = c("z_b_cw =",log((sd(b_cw$C.W))^2))
print(z_b_cw,quote=F)
z_b_po = c( "z_b_po =", log((sd(b_po$P.O))^2))
print(z_b_po,quote=F)
```

```
##Question 13
cts<-read.table("composite.txt",header=T)

#ANOVA
lm <-lm(value~composite + tape_speed + composite:tape_speed, data=cts)
anova(lm)

#Linear contrast/parameter estimation and model matrix
lm_s <- lm(cts$value ~ cts$composite+cts$tape_speed+cts$composite:cts$tape_speed, contrasts = list(f =
summary(lm_s)
m = model.matrix(lm_s)
m
```

```r
#Residual analysis
plot(lm, which =1 ,pch=23, bg="red", cex=1)
```