HW#10 Solution (week12 HW)

Due 11/21/2019

HW 24.6 Case Hardening data: CH24PR06.txt

```
HW6 = read.table(url(
  "https://raw.githubusercontent.com/npmldabook/Stat3119/master/Week-12/CH24PR06.txt"))
names(HW6) = c("response", "factorA", "factorB", "factorC", "Units")
HW6$factorA = as.factor(HW6$factorA);
HW6$factorB = as.factor(HW6$factorB)
HW6$factorC = as.factor(HW6$factorC)
str(HW6)
```

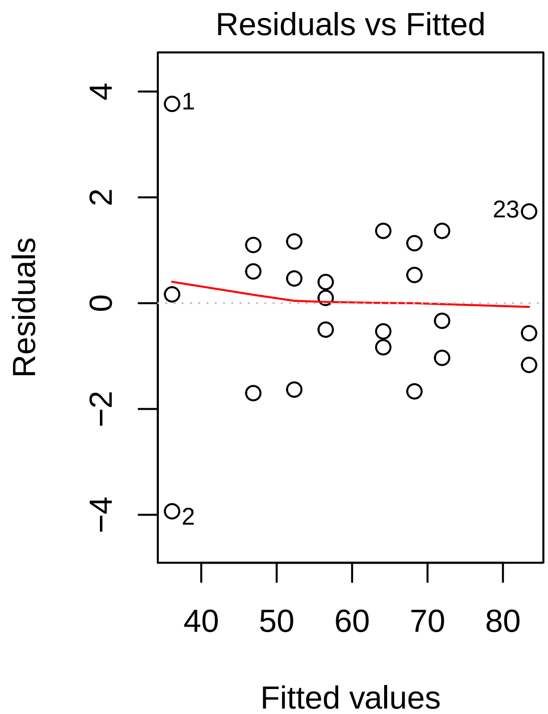
```
## 'data.frame':    24 obs. of  5 variables:
##  $ response: num  39.9 32.2 36.3 56 56.9 56.6 53.5 50.7 52.8 70.9 ...
##  $ factorA : Factor w/ 2 levels "1","2": 1 1 1 1 1 1 1 1 1 1 ...
##  $ factorB : Factor w/ 2 levels "1","2": 1 1 1 1 1 1 2 2 2 2 ...
##  $ factorC : Factor w/ 2 levels "1","2": 1 1 1 2 2 2 1 1 1 2 ...
##  $ Units   : int  1 2 3 1 2 3 1 2 3 1 ...
```

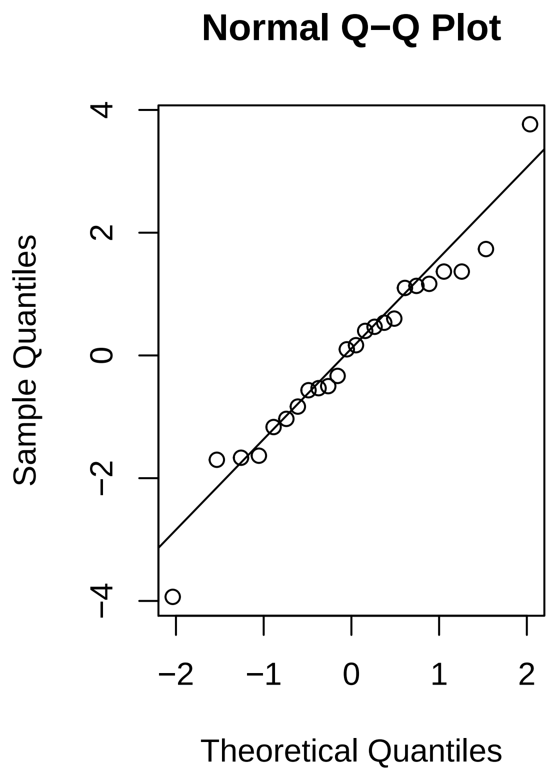
a-b) residuals plots

```
fit <- aov(response ~ factorA * factorB* factorC, data= HW6 )
fit$residuals
```

```
##          1          2          3          4          5          6
##  3.7666667 -3.9333333  0.1666667 -0.5000000  0.4000000  0.1000000
##          7          8          9         10         11         12
##  1.1666667 -1.6333333  0.4666667 -1.0333333  1.3666667 -0.3333333
##         13         14         15         16         17         18
## -1.7000000  1.1000000  0.6000000  1.1333333 -1.6666667  0.5333333
##         19         20         21         22         23         24
## -0.8333333  1.3666667 -0.5333333 -0.5666667  1.7333333 -1.1666667
```

```
par(mfrow=c(1,2))
plot(fit, 1)
#get correlation
QQstat<- qqnorm(fit$residuals) # same as plot(fit, 2)
qqline(fit$residuals)
```






```r
paste("The correlation coefficient is", round(cor(QQstat$x, QQstat$y),3))
```

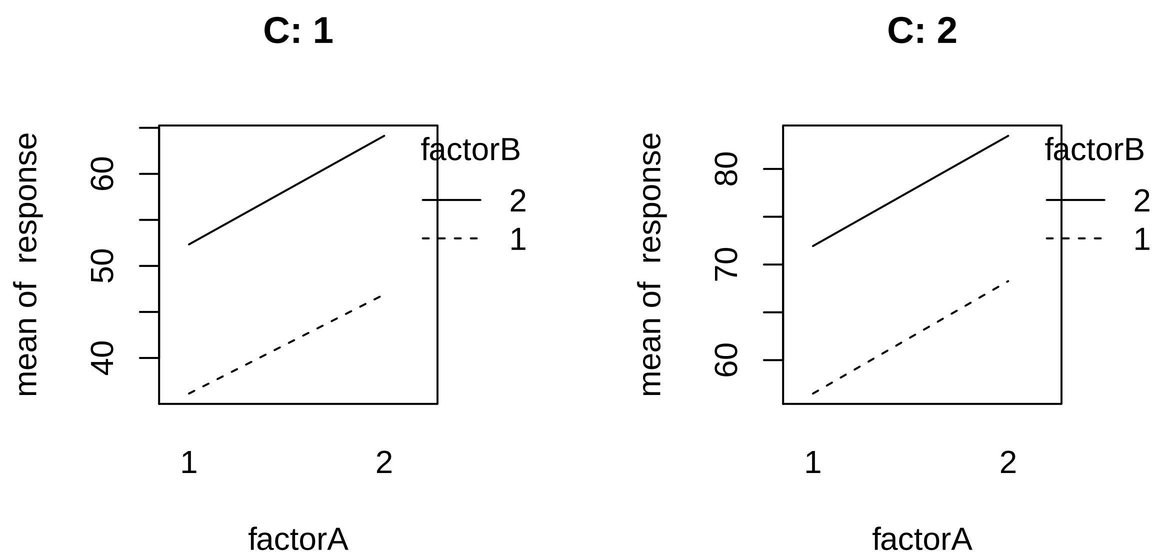
```
## [1] "The correlation coefficient is 0.976"
```

Results: The ANOVA assumptions appear to be reasonable.

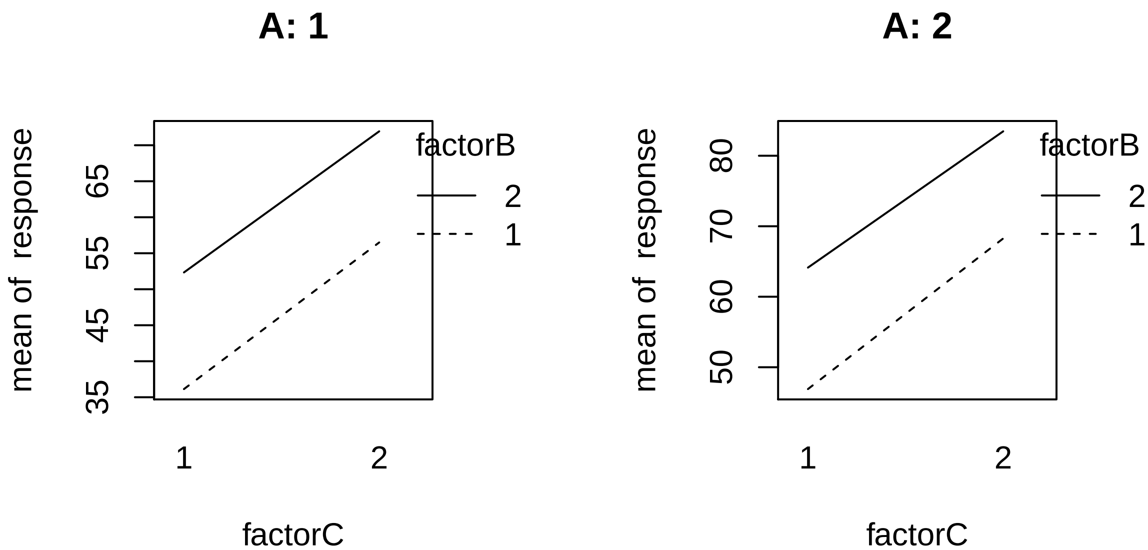
HW 24.7 Case Hardening data

a). Interacton plots

```r
par(mfrow= c(1,2),  pty="s",mar=c(4,4,5,5))

 with( HW6[HW6$factorC=='1',],
  interaction.plot(x.factor = factorA, trace.factor = factorB , response = response,
                   main= "C: 1" ) )

 with( HW6[HW6$factorC=='2',],
   interaction.plot(x.factor = factorA, trace.factor = factorB , response = response,
                   main= "C: 2" ))
```




```
par(mfrow= c(1,2),  pty="s",mar=c(4,4,5,5))

 with( HW6[HW6$factorA=='1',],
  interaction.plot(x.factor = factorC, trace.factor = factorB , response = response,
                   main= "A: 1" ) )

 with( HW6[HW6$factorA=='2',],
   interaction.plot(x.factor = factorC, trace.factor = factorB , response = response,
                   main= "A: 2" ))
```



b-e). ANOVA analysis

```
summary(fit)
```

```
##                         Df Sum Sq Mean Sq F value   Pr(>F)    
## factorA                  1  788.9   788.9 234.881 5.53e-11 ***
## factorB                  1 1539.2  1539.2 458.266 3.35e-13 ***
## factorC                  1 2440.2  2440.2 726.510 9.22e-15 ***
## factorA:factorB          1    0.2     0.2   0.071    0.793    
## factorA:factorC          1    0.2     0.2   0.060    0.810    
## factorB:factorC          1    2.9     2.9   0.875    0.363    
## factorA:factorB:factorC  1    0.6     0.6   0.179    0.678    
## Residuals               16   53.7     3.4                     
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Results: From the ANOVA table, P-value column, we can find the interactions (three-way, two way) were not significant. Then main effects were all significant. The test results agree with the findings from the interaction plots.

Detailed solutions for b-e:

c. H_0: all $(\alpha\beta\gamma)_{ijk}$ equal zero, H_a: not all $(\alpha\beta\gamma)_{ijk}$ equal zero. $F^* = .60167/3.35875 = .18$, $F(.975; 1, 16) = 6.12$. If $F^* \leq 6.12$ conclude H_0, otherwise H_a. Conclude H_0. P-value $= .68$

d. H_0: all $(\alpha\beta)_{ij}$ equal zero, H_a: not all $(\alpha\beta)_{ij}$ equal zero. $F^* = .24000/3.35875 = .07$, $F(.975; 1, 16) = 6.12$. If $F^* \leq 6.12$ conclude H_0, otherwise H_a. Conclude H_0. P-value $= .79$

H_0: all $(\alpha\gamma)_{ik}$ equal zero, H_a: not all $(\alpha\gamma)_{ik}$ equal zero. $F^* = .20167/3.35875 = .06$, $F(.975; 1, 16) = 6.12$. If $F^* \leq 6.12$ conclude H_0, otherwise H_a. Conclude H_0. P-value $= .81$

H_0: all $(\beta\gamma)_{jk}$ equal zero, H_a: not all $(\beta\gamma)_{jk}$ equal zero. $F^* = 2.94000/3.35875 = .875$, $F(.975; 1, 16) = 6.12$. If $F^* \leq 6.12$ conclude H_0, otherwise H_a. Conclude H_0. P-value $= .36$

e. H_0: all α_i equal zero $(i = 1, 2)$, H_a: not all α_i equal zero. $F^* = 788.90667/3.35875 = 234.88$, $F(.975; 1, 16) = 6.12$. If $F^* \leq 6.12$ conclude H_0, otherwise H_a. Conclude H_a. P-value $= 0+$

H_0: all β_j equal zero $(j = 1, 2)$, H_a: not all β_j equal zero. $F^* = 1,539.20167/3.35875 = 458.27$, $F(.975; 1, 16) = 6.12$. If $F^* \leq 6.12$ conclude H_0, otherwise H_a. Conclude H_a. P-value $= 0+$

H_0: all γ_k equal zero $(k = 1, 2)$, H_a: not all γ_k equal zero. $F^* = 2,440.1667/3.35875 = 726.51$, $F(.975; 1, 16) = 6.12$. If $F^* \leq 6.12$ conclude H_0, otherwise H_a. Conclude H_a. P-value $= 0+$

f). Kimball inequality

```r
1- (1-0.025)^7
```

```
## [1] 0.1624084
```

Results: $\alpha \leq 0.1624$.

HW 24.8 Inference.

Solution: apply standard formula in this chapter

a. $\hat{D}_1 = 65.69167 - 54.22500 = 11.46667,\ \hat{D}_2 = 67.96667 - 51.95000 = 16.01667$

$\hat{D}_3 = 70.04167 - 49.87500 = 20.16667,\ MSE = 3.35875,$

$s\{\hat{D}_i\} = .7482\ (i = 1, 2, 3),\ B = t(.99167; 16) = 2.673$

$$\begin{array}{ll} 11.46667 \pm 2.673(.7482) & 9.467 \leq D_1 \leq 13.467 \\ 16.01667 \pm 2.673(.7482) & 14.017 \leq D_2 \leq 18.017 \\ 20.16667 \pm 2.673(.7482) & 18.167 \leq D_3 \leq 22.167 \end{array}$$

b. $\bar{Y}_{222.} = 83.46667,\ s\{\bar{Y}_{222.}\} = 1.0581,\ t(.975;\ 16) = 2.120,$

$83.46667 \pm 2.120(1.0581),\ 81.2235 \leq \mu_{222} \leq 85.7098$

R implementation (a)

```
# paired difference : use emmeans to get the estimate and SE, or use formula

library(emmeans)
Est.meanA <- emmeans(fit, ~ factorA)
```

```
## NOTE: Results may be misleading due to involvement in interactions
```

```
Est.meanB <- emmeans(fit, ~ factorB)
```

```
## NOTE: Results may be misleading due to involvement in interactions
```

```
Est.meanC <- emmeans(fit, ~ factorC)
```

```
## NOTE: Results may be misleading due to involvement in interactions
```

```
L = list( L= c( -1, 1))    #L= mu2-mu1

# Bonferroni methods, alpha=0.05/3
confint(contrast(Est.meanA, L, adjust='none'), level=(1-0.05/3))
```

```
##  contrast estimate    SE df lower.CL upper.CL
##  L            11.5 0.748 16     9.47     13.5
##
## Results are averaged over the levels of: factorB, factorC
## Confidence level used: 0.983333333333333
```

```
confint(contrast(Est.meanB, L, adjust='none'), level=(1-0.05/3))
```

```
##  contrast estimate    SE df lower.CL upper.CL
##  L              16 0.748 16       14       18
##
## Results are averaged over the levels of: factorA, factorC
## Confidence level used: 0.983333333333333
```

```
confint(contrast(Est.meanC, L, adjust='none'), level=(1-0.05/3))
```

```
##  contrast estimate    SE df lower.CL upper.CL
##  L            20.2 0.748 16     18.2     22.2
##
## Results are averaged over the levels of: factorA, factorB
## Confidence level used: 0.983333333333333
```

R implementation (b)

```
predict(fit, newdata = data.frame(factorA ='2', factorB ='2',factorC ='2' ),
        interval = "confidence", level = .95)
```

```
##        fit      lwr      upr
## 1 83.46667 81.22359 85.70975
```

HW 24.15

```
HW6B = HW6[-(7:8),]
dim(HW6B)
```

```
## [1] 22  5
```

Solutions for a) to d):

a. $Y_{ijkm} = \mu_{...} + \alpha_1 X_{ijkm1} + \beta_1 X_{ijkm2} + \gamma_1 X_{ijkm3} + (\alpha\beta)_{11} X_{ijkm1} X_{ijkm2} + (\alpha\gamma)_{11} X_{ijkm1} X_{ijkm3}$
$+ (\beta\gamma)_{11} X_{ijkm2} X_{ijkm3} + (\alpha\beta\gamma)_{111} X_{ijkm1} X_{ijkm2} X_{ijkm3} + \epsilon_{ijkm}$

$$X_{ijk1} = \begin{matrix} 1 \text{ if case from level 1 for factor } A \\ -1 \text{ if case from level 2 for factor } A \end{matrix}$$

$$X_{ijk2} = \begin{matrix} 1 \text{ if case from level 1 for factor } B \\ -1 \text{ if case from level 2 for factor } B \end{matrix}$$

$$X_{ijk3} = \begin{matrix} 1 \text{ if case from level 1 for factor } C \\ -1 \text{ if case from level 2 for factor } C \end{matrix}$$

b. $Y_{ijkm} = \mu_{...} + \beta_1 X_{ijkm2} + \gamma_1 X_{ijkm3} + (\alpha\beta)_{11} X_{ijkm1} X_{ijkm2} + (\alpha\gamma)_{11} X_{ijkm1} X_{ijkm3}$
$+ (\beta\gamma)_{11} X_{ijkm2} X_{ijkm3} + (\alpha\beta\gamma)_{111} X_{ijkm1} X_{ijkm2} X_{ijkm3} + \epsilon_{ijkm}$

c. <u>Full model</u>:

$\hat{Y} = 60.01667 - 5.67500 X_1 - 8.06667 X_2 - 10.02500 X_3 + .04167 X_1 X_2$
$+ .15000 X_1 X_3 - .40833 X_2 X_3 + .10000 X_1 X_2 X_3,$

$SSE(F) = 49.4933$

<u>Reduced model</u>:

$\hat{Y} = 61.15167 - 9.20167 X_2 - 8.89000 X_3 - 1.09333 X_1 X_2 + 1.28500 X_1 X_3$
$-1.54333 X_2 X_3 - 1.03500 X_1 X_2 X_3,$

$SSE(R) = 667.8413$

H_0: $\alpha_1 = 0$, H_a: $\alpha_1 \neq 0$.

$F^* = (618.348/1) \div (49.4933/14) = 174.91$, $F(.975; 1, 14) = 6.298$.

If $F^* \leq 6.298$ conclude H_0, otherwise H_a. Conclude H_a. P-value = 0+

d. $\hat{D} = \hat{\mu}_{2..} - \hat{\mu}_{1..} = \hat{\alpha}_2 - \hat{\alpha}_1 = -2\hat{\alpha}_1 = 11.35000$, $s^2\{\hat{\alpha}_1\} = .18413$, $s\{\hat{D}\} = .8582$, $t(.975; 14) = 2.145$,

$11.35000 \pm 2.145(.8582)$, $9.509 \leq D \leq 13.191$

R immplementation a)-d)

```
Ind.A =  (HW6B$factorA=="1")*1+ (HW6B$factorA=="2")*(-1)
Ind.B =  (HW6B$factorB=="1")*1+ (HW6B$factorB=="2")*(-1)
Ind.C =  (HW6B$factorC=="1")*1+ (HW6B$factorC=="2")*(-1)


Model.full = lm(response~ Ind.A*Ind.B*Ind.C, data= HW6B )
summary(Model.full)
```

```
##
```

```
## Call:
## lm(formula = response ~ Ind.A * Ind.B * Ind.C, data = HW6B)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.9333 -0.7667  0.0500  0.9750  3.7667
##
## Coefficients:
##                      Estimate Std. Error t value Pr(>|t|)
## (Intercept)          60.01667    0.42910 139.866  < 2e-16 ***
## Ind.A                -5.67500    0.42910 -13.225 2.66e-09 ***
## Ind.B                -8.06667    0.42910 -18.799 2.49e-11 ***
## Ind.C               -10.02500    0.42910 -23.363 1.30e-12 ***
## Ind.A:Ind.B           0.04167    0.42910   0.097    0.924
## Ind.A:Ind.C           0.15000    0.42910   0.350    0.732
## Ind.B:Ind.C          -0.40833    0.42910  -0.952    0.357
## Ind.A:Ind.B:Ind.C     0.10000    0.42910   0.233    0.819
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.88 on 14 degrees of freedom
## Multiple R-squared:  0.9894, Adjusted R-squared:  0.9842
## F-statistic: 187.4 on 7 and 14 DF,  p-value: 9.674e-13
```

```r
Model.reduced = lm(response~  Ind.B*Ind.C+ Ind.A:Ind.B + Ind.A:Ind.C+ Ind.A:Ind.B:Ind.C, data= HW6B )
summary(Model.reduced)
```

```
##
## Call:
## lm(formula = response ~ Ind.B * Ind.C + Ind.A:Ind.B + Ind.A:Ind.C +
##     Ind.A:Ind.B:Ind.C, data = HW6B)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -13.620  -4.423   2.857   4.807   6.273
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)          61.152      1.492  40.986  < 2e-16 ***
## Ind.B                -9.202      1.492  -6.167 1.80e-05 ***
## Ind.C                -8.890      1.492  -5.958 2.62e-05 ***
## Ind.B:Ind.C          -1.543      1.492  -1.034    0.317
## Ind.B:Ind.A          -1.093      1.492  -0.733    0.475
## Ind.C:Ind.A           1.285      1.492   0.861    0.403
## Ind.B:Ind.C:Ind.A    -1.035      1.492  -0.694    0.498
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6.673 on 15 degrees of freedom
## Multiple R-squared:  0.8575, Adjusted R-squared:  0.8005
## F-statistic: 15.05 on 6 and 15 DF,  p-value: 1.389e-05
```

```r
anova(Model.reduced, Model.full )
```

```
## Analysis of Variance Table
##
## Model 1: response ~ Ind.B * Ind.C + Ind.A:Ind.B + Ind.A:Ind.C + Ind.A:Ind.B:Ind.C
## Model 2: response ~ Ind.A * Ind.B * Ind.C
##   Res.Df    RSS Df Sum of Sq      F   Pr(>F)
## 1     15 667.84
## 2     14  49.49  1    618.35 174.91 2.66e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

d) inference for factor A

Approach 1:

1) use formula (24.48a,b,c, page 1021)

```r
# For A1: the treatment means are 36.13, 56.5, 52.8, 71.93
# For A2: the treatment means 46.90000 68.26667 64.13333 83.46667

 (EstA1 = mean(Model.full$fitted.values[c(1,4,7,8)]))
```

```
## [1] 54.34167
```

```r
 (EstA2 = mean(Model.full$fitted.values[c(11,14,17,20)]))
```

```
## [1] 65.69167
```

```r
# the paired difference
(Diff=EstA2 -  EstA1)
```

```
## [1] 11.35
```

2) get SE estimate

```r
# The square root of MSE=sigma= 1.88
sigma= 1.88
(SE =  sigma/(2*2)*sqrt((1/3+1)+(1/3+1/3)+ (1/3+1/3) +(1/3+1/3)))
```

```
## [1] 0.8580987
```

3) get CI

```r
df= nrow(HW6B)- 8 #14=n_T-8

(t.value = qt(.975, df))
```

```
## [1] 2.144787
```

```
LCI= Diff- t.value*SE
UCI= Diff+ t.value*SE

paste(" Then 95% CI is (", round(LCI,4), "-", round(UCI,4),").")
```

```
## [1] " Then 95% CI is ( 9.5096 - 13.1904 )."
```

Approach 2: use the fact that

$D = \mu_{2..} - \mu_{1..} = \alpha_2 - \alpha_1 = (-\alpha_1) - \alpha_1 = -2\alpha_1$.

- From the regression output for the full model, $\hat{\alpha}_1 = -5.67500$, then the $\hat{D} = -2 * (-5.675) = 11.35$.
- From the regression output for the full model, $\sigma(\hat{\alpha_1}) = 0.4291$, then the $\sigma(\hat{D}) = 2{*}\sigma(\hat{\alpha_1}) = 2 * (0.4291) = 0.8582$.

Approach 3: use emmeans

```
LM.full2 = lm( response~ factorA* factorB* factorC, data=HW6B )
fit.emm <- emmeans( LM.full2, ~ factorA)
```

```
## NOTE: Results may be misleading due to involvement in interactions
```

```
L = list( L= c( -1, 1))   #L= mu2-mu1

# Bonferroni methods, alpha=0.05/3
confint(contrast(fit.emm, L, adjust='none'))
```

```
##  contrast estimate    SE df lower.CL upper.CL
##  L            11.3 0.858 14     9.51     13.2
##
## Results are averaged over the levels of: factorB, factorC
## Confidence level used: 0.95
```