Stat 207 HW1

Cheng Luo 912466499
Fan Wu 912538518

January 15, 2015

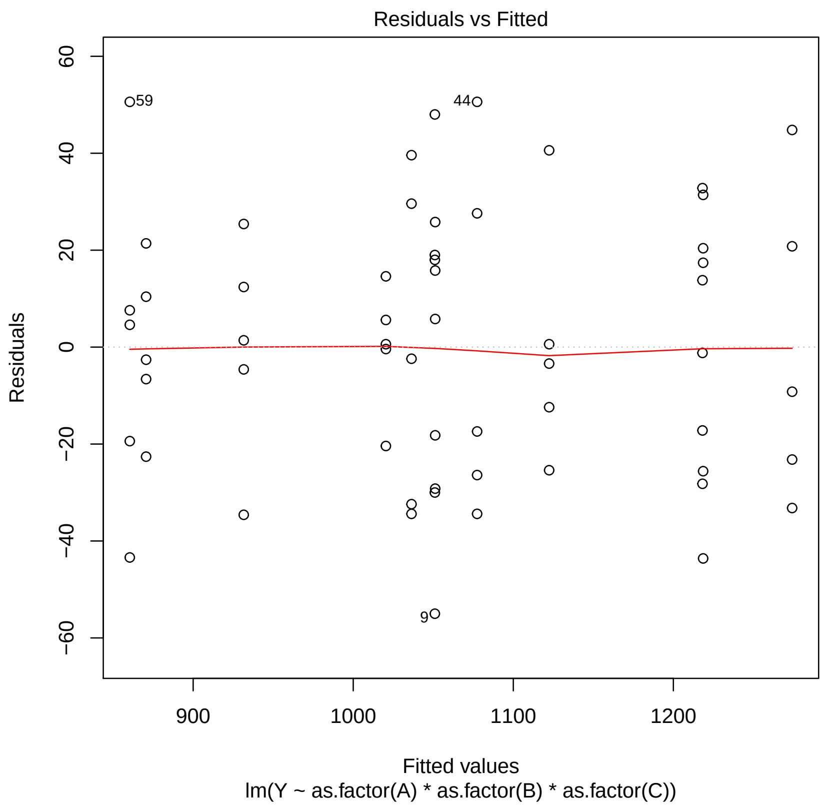
1 24.12

(a)

```
elec = read.table("CH24PR12.txt")
names(elec) = c("Y", "A", "B", "C", "T")
n = length(unique(elec$T))
a = length(unique(elec$A))
b = length(unique(elec$B))
c = length(unique(elec$C))
fit = lm(Y ~ as.factor(A) * as.factor(B) * as.factor(C) , elec)
res = resid(fit)
res
```

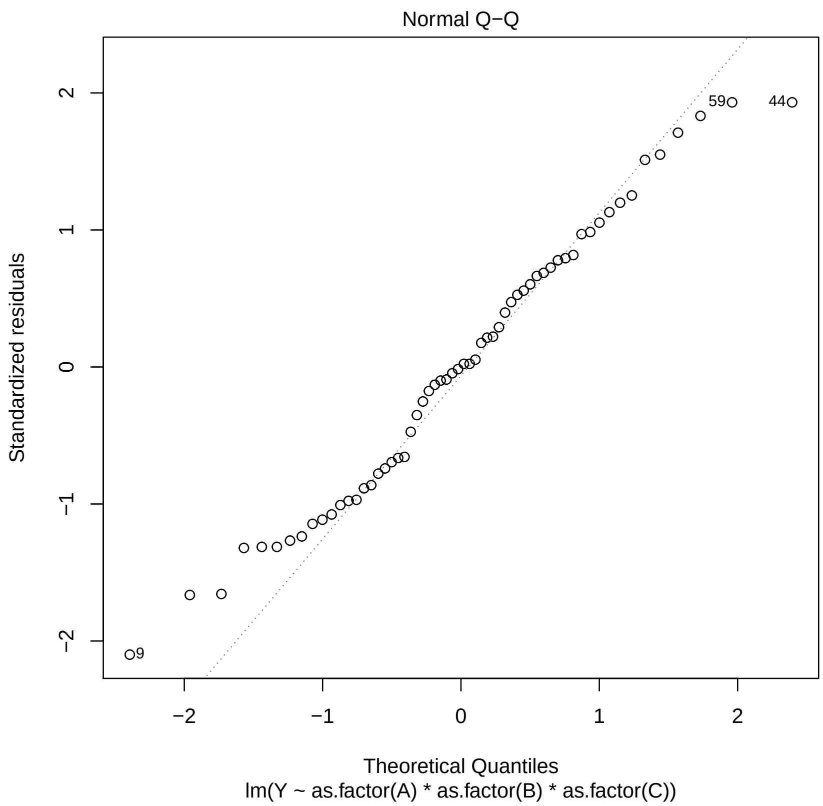
```
##     1     2     3     4     5     6     7     8     9    10    11    12
##  31.4 -43.6  17.4  20.4 -25.6 -30.0  48.0  18.0 -55.0  19.0  44.8 -23.2
##    13    14    15    16    17    18    19    20    21    22    23    24
## -33.2  20.8  -9.2  -3.4 -12.4   0.6 -25.4  40.6  -1.2 -28.2 -17.2  13.8
##    25    26    27    28    29    30    31    32    33    34    35    36
##  32.8 -18.2  15.8   5.8  25.8 -29.2  29.6  39.6 -32.4 -34.4  -2.4  -6.6
##    37    38    39    40    41    42    43    44    45    46    47    48
## -22.6  10.4  21.4  -2.6  27.6 -34.4 -26.4  50.6 -17.4  -4.6  12.4  25.4
##    49    50    51    52    53    54    55    56    57    58    59    60
## -34.6   1.4   0.6  -0.4  14.6 -20.4   5.6 -19.4   4.6 -43.4  50.6   7.6
```

```
plot(fit, which = 1)
```



The residuals versus fitted values plots shows no sign for unequal variance.

(b)
```
plot(fit, which = 2)
```




```
cor(sort(res), ppoints(res))

## [1] 0.9930971
```

The normal QQ plot and correlation indicates approximately normal distribution of residuals with slight light tail, so that the normality assumption seems to be reasonable here.

2 24.13

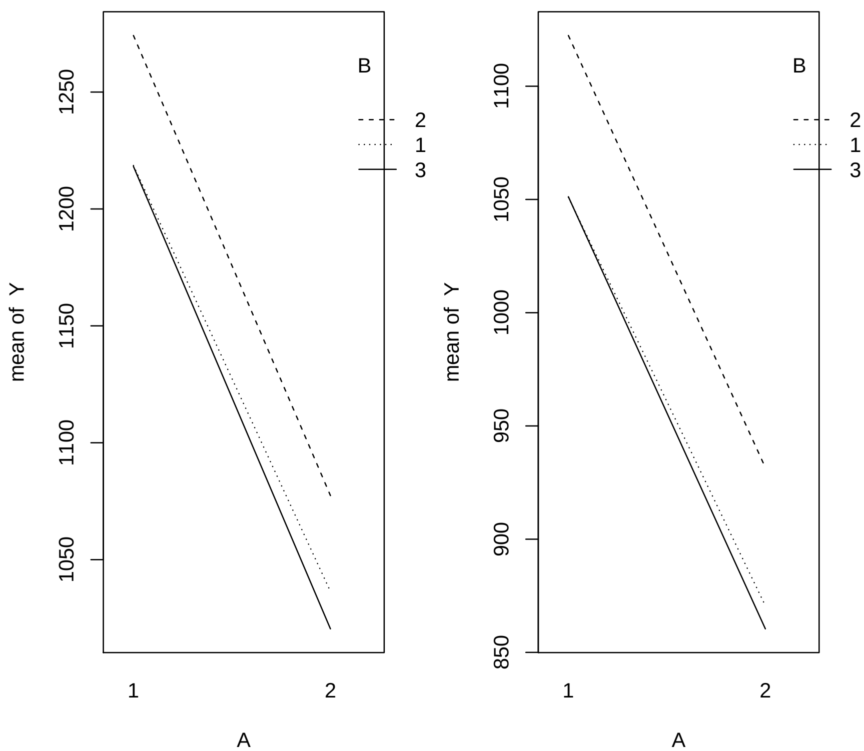
(a) $\bar{Y}_{ijk\cdot}$ are as follow:

```
means = sapply(with(elec, split(Y,list(A,B,C))), mean)
means

##  1.1.1  2.1.1  1.2.1  2.2.1  1.3.1  2.3.1  1.1.2  2.1.2  1.2.2  2.2.2
```

```
## 1218.6 1036.4 1274.2 1077.4 1218.2 1020.4 1051.0  870.6 1122.4  931.6
##  1.3.2  2.3.2
## 1051.2  860.4
```

```
elec.C = split(elec, elec$C)
par(mfrow = c(1,2))
with(elec.C$'1', interaction.plot(A, B, Y))
with(elec.C$'2', interaction.plot(A, B, Y))
```



From AB plots of the estimated treatment means, the AB curves seem to be parallel, which means there's no interaction between AB, moreover, main effect A and main effect B are present.

(b)
```
anova(fit)
```

```
## Analysis of Variance Table
```

```
##
## Response: Y
##                                          Df Sum Sq Mean Sq  F value
## as.factor(A)                              1 540361  540361 629.7603
## as.factor(B)                              2  49320   24660  28.7396
## as.factor(C)                              1 382402  382402 445.6679
## as.factor(A):as.factor(B)                 2    543     271   0.3161
## as.factor(A):as.factor(C)                 1     91      91   0.1064
## as.factor(B):as.factor(C)                 2    911     456   0.5310
## as.factor(A):as.factor(B):as.factor(C)   2     19      10   0.0111
## Residuals                                48  41186     858
##                                             Pr(>F)
## as.factor(A)                             < 2.2e-16 ***
## as.factor(B)                              6.22e-09 ***
## as.factor(C)                             < 2.2e-16 ***
## as.factor(A):as.factor(B)                   0.7305
## as.factor(A):as.factor(C)                   0.7457
## as.factor(B):as.factor(C)                   0.5914
## as.factor(A):as.factor(B):as.factor(C)     0.9890
## Residuals
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

(c) For three-factor interaction:

$$H_0\text{:all } (\alpha\beta\gamma)_{ijk}\text{=}0$$
$$\text{VS. } H_1\text{:not all } (\alpha\beta\gamma)_{ijk} = 0$$
$$F^* = \frac{MSABC}{MSE} = 0.0111$$
$$\text{we can reject } H_0 \text{ if } F^* > F(1-0.05; 2, 48) = 3.19\text{,otherwise reject} H_1$$
$$\text{reject } H_1 \text{ because } F^* < F(1-0.05; 2, 48) = 3.19,$$

therefore,we conclude H_0 at 0.05 level, and there's no three-factor interaction effect, and P-value = 0.9890

(d) For AB interaction:

$$H_0\text{:all } (\alpha\beta)_{ij}\text{=}0$$
$$\text{VS. } H_1\text{:not all } (\alpha\beta)_{ij} = 0$$
$$F^* = \frac{MSAB}{MSE} = 0.3161$$
$$\text{we can reject } H_0 \text{ if } F^* > F(1-0.05; 2, 48) = 3.19\text{,otherwise reject} H_1$$
$$\text{reject } H_1 \text{ because } F^* < F(1-0.05; 2, 48) = 3.19,$$

therefore,we conclude H_0 at 0.05 level, and there's no AB interaction effect, and P-value = 0.7305

For AC interaction:

$$H_0\text{:all } (\alpha\gamma)_{ik}\text{=0}$$
$$\text{VS. } H_1\text{:not all } (\alpha\gamma)_{ik} = 0$$
$$F^* = \frac{MSAC}{MSE} = 0.1064$$
we can reject H_0 if $F^* > F(1-0.05;1,48) = 4.04$,otherwise reject$H_1$
reject H_1 because $F^* < F(1-0.05;2,48) = 4.04$,
therefore,we conclude H_0 at 0.05 level, and there's no AC interaction effect, and P-value = 0.7457

For BC interaction:

$$H_0\text{:all } (\beta\gamma)_{jk}\text{=0}$$
$$\text{VS. } H_1\text{:not all } (\beta\gamma)_{jk} = 0$$
$$F^* = \frac{MSBC}{MSE} = 0.5310$$
we can reject H_0 if $F^* > F(1-0.05;2,48) = 3.19$,otherwise reject$H_1$
reject H_1 because $F^* < F(1-0.05;2,48) = 3.19$,
therefore,we conclude H_0 at 0.05 level, and there's no BC interaction effect, and P-value = 0.5914

(e) For A main effect:

$$H_0\text{:all } (\alpha)_i\text{=0}$$
$$\text{VS. } H_1\text{:not all } (\alpha)_i = 0$$
$$F^* = \frac{MSA}{MSE} = 629.7603$$
we can reject H_0 if $F^* > F(1-0.05;2,48) = 4.04$,otherwise reject$H_1$
reject H_0 because $F^* > F(1-0.05;2,48) = 4.04$,
therefore,we conclude H_1 at 0.05 level, and there's A main effect, and P-value = 2.2e-16

For B main effect:

$$H_0\text{:all } (\beta)_j\text{=0}$$
$$\text{VS. } H_1\text{:not all } (\beta)_j = 0$$
$$F^* = \frac{MSB}{MSE} = 28.7396$$
we can reject H_0 if $F^* > F(1-0.05;2,48) = 3.19$,otherwise reject$H_1$
reject H_0 because $F^* > F(1-0.05;2,48) = 3.19$,
therefore,we conclude H_1 at 0.05 level, and there's B main effect, and P-value = 6.22e-09

For C main effect:

$$H_0\text{:all } (\gamma)_k\text{=0}$$
$$\text{VS. } H_1\text{:not all } (\gamma)_k = 0$$
$$F^* = \frac{MSC}{MSE} = 445.6679$$
we can reject H_0 if $F^* > F(1-0.05; 2, 48) = 4.04$,otherwise reject$H_1$

reject H_0 because $F^* > F(1-0.05; 2, 48) = 3.19$,

therefore,we conclude H_1 at 0.05 level, and there's C main effect, and P-value = 2.2e-16

(f) A,B and C main effects are present, but there's no any interaction effect. As for upper bound for the family level of significance, using kimball inequality,$\alpha < 1-(1-\alpha_1)(1-\alpha_2)\cdots(1-\alpha_7) = 1-(1-0.05)^7 = 0.3016627$

(g) The results in part(f) confirm my graphic analysis in part(a)

3 24.14

(a)
```
means = with(elec, sapply(list(A, B, C), function(x) by(Y, x, mean)))
means

## [[1]]
## x: 1
## [1] 1155.933
## --------------------------------------------------------
## x: 2
## [1] 966.1333
##
## [[2]]
## x: 1
## [1] 1044.15
## --------------------------------------------------------
## x: 2
## [1] 1101.4
## --------------------------------------------------------
## x: 3
## [1] 1037.55
##
## [[3]]
## x: 1
## [1] 1140.867
## --------------------------------------------------------
## x: 2
## [1] 981.2
```

```r
  D1.hat = means[[1]][1] - means[[1]][2]
  D2.hat = means[[2]][1] - means[[2]][2]
  D3.hat = means[[2]][1] - means[[2]][3]
  D4.hat = means[[2]][2] - means[[2]][3]
  D5.hat = means[[3]][1] - means[[3]][2]
  mse = anova(fit)["Residuals", 3]
  S_D1 = sqrt(mse/(n*b*c)*2)
  S_D2 = sqrt(mse/(n*a*c)*2)
  S_D3 = sqrt(mse/(n*a*c)*2)
  S_D4 = sqrt(mse/(n*a*c)*2)
  S_D5 = sqrt(mse/(n*a*b)*2)
  B = qt(1-.1/(2*5), 48)
  c(D1.hat-B*S_D1, D1.hat+B*S_D1)

##        1        1
## 171.5984 208.0016

  c(D2.hat-B*S_D2, D2.hat+B*S_D2)

##         1         1
## -79.54229 -34.95771

  c(D3.hat-B*S_D3, D3.hat+B*S_D3)

##         1         1
## -15.69229  28.89229

  c(D4.hat-B*S_D4, D4.hat+B*S_D4)

##        2        2
## 41.55771 86.14229

  c(D5.hat-B*S_D5, D5.hat+B*S_D5)

##        1        1
## 141.4651 177.8682
```

$$171.5984 \leq D_1 \leq 208.0016$$
$$-79.54229 \leq D_2 \leq -34.95771$$
$$-15.69229 \leq D_3 \leq 28.89229$$
$$41.55771 \leq D_4 \leq 86.14229$$
$$141.4651 \leq D_5 \leq 177.8682$$

(b)
```
dat.231 = elec[with(elec, A == 2 & B == 3 & C == 1), ]
means = mean(dat.231$Y)
s = sqrt(mse/n)
t. = qt(1 - .05/2, (n - 1)*a*b*c)
c(means - t.*s, means + t.*s)

## [1]  994.0608 1046.7392
```

Therefore,μ_{231} with a 95 percent confidence interval is $994.0608 \leq \mu_{231} \leq 1046.7392$

4 24.16

(a)
```
  dat.reg = elec[-c(19, 40, 43), ]
  x1 = dat.reg$A
  x1 = replace(x1, which(dat.reg$A == 2), -1)
  x2 = dat.reg$B
  x2 = replace(x2, which(dat.reg$B == 3), -1)
  x2 = replace(x2, which(dat.reg$B == 2), 0)
  x3 = dat.reg$B
  x3 = replace(x3, which(dat.reg$B == 3), -1)
  x3 = replace(x3, which(dat.reg$B == 1), 0)
  x3 = replace(x3, which(dat.reg$B == 2), 1)
  x4 = dat.reg$C
  x4 = replace(x4, which(dat.reg$C == 2), -1)
  model.full = lm(dat.reg$Y ~ x1*(x2+x3)*x4)
  sse.full = sum(resid(model.full)^2)
```

For full model:

$$
\begin{aligned}
Y_{ijkm} = \mu_{\dots} &+ \alpha_1 X_{ijkm1} + \beta_1 X_{ijkm2} + \beta_2 X_{ijkm3} + \gamma X_{ijkm4} \\
&+ (\alpha\beta)_{11} X_{ijkm1} X_{ijkm2} + (\alpha\beta)_{12} X_{ijkm1} X_{ijkm3} + (\alpha\gamma)_{11} X_{ijkm1} X_{ijkm4} \\
&+ (\beta\gamma)_{11} X_{ijkm2} X_{ijkm4} + (\beta\gamma)_{21} X_{ijkm3} X_{ijkm4} + (\alpha\beta\gamma)_{111} X_{ijkm1} X_{ijkm2} X_{ijkm4} \\
&+ (\alpha\beta\gamma)_{121} X_{ijkm1} X_{ijkm3} X_{ijkm4} + \epsilon_{ijkm}
\end{aligned}
$$

$$
X_{ijkm1} = \begin{cases} 1 & \text{if factor A is level 1} \\ -1 & \text{if factor A is level 2} \end{cases} \tag{1}
$$

$$
X_{ijkm2} = \begin{cases} 1 & \text{if factor B is level 1} \\ -1 & \text{if factor B is level 3} \\ 0 & \textit{else} \end{cases} \tag{2}
$$

$$X_{ijkm3} = \begin{cases} 1 & \text{if factor B is level 2} \\ -1 & \text{if factor B is level 3} \\ 0 & else \end{cases} \tag{3}$$

$$X_{ijkm4} = \begin{cases} 1 & \text{if factor C is level 1} \\ -1 & \text{if factor C is level 2} \end{cases} \tag{4}$$

(b)
```
model.red = lm(dat.reg$Y ~ x1*(x2+x3)*x4 -x4)
sse.red = sum(resid(model.red)^2)
```

For reduced model:

$$\begin{aligned} Y_{ijkm} &= \mu_{...} + \alpha_1 X_{ijkm1} + \beta_1 X_{ijkm2} + \beta_2 X_{ijkm3} \\ &\quad + (\alpha\beta)_{11} X_{ijkm1} X_{ijkm2} + (\alpha\beta)_{12} X_{ijkm1} X_{ijkm3} + (\alpha\gamma)_{11} X_{ijkm1} X_{ijkm4} \\ &\quad + (\beta\gamma)_{11} X_{ijkm2} X_{ijkm4} + (\beta\gamma)_{21} X_{ijkm3} X_{ijkm4} + (\alpha\beta\gamma)_{111} X_{ijkm1} X_{ijkm2} X_{ijkm4} \\ &\quad + (\alpha\beta\gamma)_{121} X_{ijkm1} X_{ijkm3} X_{ijkm4} + \epsilon_{ijkm} \end{aligned}$$

(c)
```
summary(model.full)

##
## Call:
## lm(formula = dat.reg$Y ~ x1 * (x2 + x3) * x4)
##
## Residuals:
##    Min     1Q Median     3Q    Max
##  -55.0  -23.2    0.6   20.4   50.6
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1062.1667     3.9426 269.409  < 2e-16 ***
## x1            94.8250     3.9426  24.052  < 2e-16 ***
## x2           -17.8542     5.5757  -3.202   0.0025 **
## x3            42.4708     5.6571   7.508 1.81e-09 ***
## x4            79.8000     3.9426  20.241  < 2e-16 ***
## x1:x2         -4.3375     5.5757  -0.778   0.4407
## x1:x3          2.0125     5.6571   0.356   0.7237
## x1:x4          0.2083     3.9426   0.053   0.9581
## x2:x4          3.3875     5.5757   0.608   0.5465
## x3:x4         -5.3375     5.6571  -0.944   0.3505
## x1:x2:x4       0.4042     5.5757   0.072   0.9425
```

```
## x1:x3:x4      -1.9458     5.6571  -0.344   0.7325
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 29.63 on 45 degrees of freedom
## Multiple R-squared:  0.9595,Adjusted R-squared:  0.9496
## F-statistic: 96.96 on 11 and 45 DF,  p-value: < 2.2e-16
```

```
summary(model.red)
```

```
##
## Call:
## lm(formula = dat.reg$Y ~ x1 * (x2 + x3) * x4 - x4)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -130.11  -73.71   31.51   74.71  137.88
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1063.731     12.393  85.834  < 2e-16 ***
## x1            96.390     12.393   7.778  6.3e-10 ***
## x2           -14.725     17.523  -0.840    0.405
## x3            40.906     17.784   2.300    0.026 *
## x1:x2        -10.596     17.503  -0.605    0.548
## x1:x3          9.836     17.744   0.554    0.582
## x1:x4          1.773     12.393   0.143    0.887
## x2:x4          3.388     17.529   0.193    0.848
## x3:x4        -10.032     17.770  -0.565    0.575
## x1:x2:x4       3.534     17.523   0.202    0.841
## x1:x3:x4      -3.511     17.784  -0.197    0.844
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 93.15 on 46 degrees of freedom
## Multiple R-squared:  0.5909,Adjusted R-squared:  0.502
## F-statistic: 6.645 on 10 and 46 DF,  p-value: 2.836e-06
```

```
sse.full
```

```
## [1] 39499.9
```

```
sse.red
```

```
## [1] 399106.9
```

SSE(F)=39499.9, df(F)=45

SSE(R)=399106.9, df(R)=46

$$H_0\text{:all } (\gamma)_1 = 0$$
$$\text{VS. } H_1\text{:not all } (\gamma)_1 = 0$$
$$F^* = \frac{(SSE(R)-SSE(F))/(df(R)-df(F))}{SSE(F)/df(F)} = \frac{359607/1}{39499.9/45} = 409.6799$$

we can reject H_0 if $F^* > F(1-0.05; 1, 45) = 4.056612$,otherwise reject$H_1$

reject H_0 because $F^* > F(1-0.05; 1, 45) = 3.19$,

therefore,we conclude H_1 at 0.05 level, and there's C main effect, and P-value is 3.114909e-24, very close to zero.

```
  f.star = (sse.red - sse.full) / (sse.full/45)
  qf(1 - .05, 1, 45)

## [1] 4.056612

  pf(f.star, 1, 45, lower = FALSE)

## [1] 3.114909e-24
```

(d)

```
  D.hat = 159.6
  n.s = with(dat.reg, by(Y, list(A, B, C), length))
  s = sqrt(sse.full/(nrow(dat.reg) - a*b*c) / (a^2*b^2) * sum(1/n.s))
  s

## [1] 7.885161

  t. = qt(1 - .05/2, nrow(dat.reg) - a*b*c)
  c(D.hat - s*t., D.hat + s*t.)

## [1] 143.7185 175.4815
```

$\hat{D} = \hat{\mu}_{..1} - \hat{\mu}_{..2} = \hat{\gamma}_1 - \hat{\gamma}_2 = 2\hat{\gamma}_1 = 2 * 79.8 = 159.6$

MSE $= \frac{SSE(F)}{df(F)} = 877.7756$

$Var(\bar{Y}_{ijk\cdot}) = \frac{MSE}{n_{ijk}}$

$Var(\hat{\mu}_{..1}) = Var(\frac{\sum_{i=1}^{a}\sum_{j=1}^{b}\bar{Y}_{ij1\cdot}}{ab}) = \frac{MSE}{a^2b^2}\sum_{i=1}^{a}\sum_{j=1}^{b}\frac{1}{n_{ij1}}$

$S(\hat{D}) = \sqrt{Var(\hat{\mu}_{..1}) + Var(\hat{\mu}_{..2})} = \sqrt{\frac{MSE}{a^2b^2}\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{c}\frac{1}{n_{ijk}}} = 7.885161$

t = t(1- 0.05/2; 45) = 2.014103

Therefore,$\hat{D} \pm S(\hat{D}) * t$, which means $143.7185 \leq D \leq 175.4815$

5 24.18

```
  n = 5
  s = sqrt(29^2/(n*b*c)*2)
  B = qt(1 - .1/(2*5), n*a*b*c - a*b*c)
  B*s

## [1] 18.01992

  n = 7
  s = sqrt(29^2/(n*a*c)*2)
  B = qt(1 - .1/(2*5), n*a*b*c - a*b*c)
  B*s

## [1] 18.44065

  n = 5
  s = sqrt(29^2/(n*a*b)*2)
  B = qt(1 - .1/(2*5), n*a*b*c - a*b*c)
  B*s

## [1] 18.01992
```

We find that

- for L_1, if the precision of each of the estimates should not exceed, the smallest sample size is 5
- for L_2, L_3 and L_4, if the precision of each of the estimates should not exceed, the smallest sample size is 7
- for L_1, if the precision of each of the estimates should not exceed, the smallest sample size is 5.

Therefore, the required sample size should be n ≥ 7.

6 24.19

$$\begin{aligned}\sum_i (\alpha\beta\gamma)_{ijk} &= \sum_i (\mu_{ijk} - \mu_{ij\cdot} - \mu_{i\cdot k} - \mu_{\cdot jk} + \mu_{i\cdot\cdot} + \mu_{\cdot j\cdot} + \mu_{\cdot\cdot k} - \mu_{\cdot\cdot\cdot}) \\ &= a\mu_{\cdot jk} - a\mu_{\cdot j\cdot} - a\mu_{\cdot\cdot k} - a\mu_{\cdot jk} + a\mu_{\cdot\cdot\cdot} + a\mu_{\cdot j\cdot} + a\mu_{\cdot\cdot k} - a\mu_{\cdot\cdot\cdot} \\ &= 0\end{aligned}$$

7 24.20

The model without three-factor interaction is:

$$Y_{ijk} = \mu_{\cdot\cdot\cdot} + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} + \epsilon_{ijk}$$

	SS	d.f.	MS
A	SSA	a-1	MSA
B	SSB	b-1	MSB
C	SSC	c-1	MSC
AB	SSAB	(a-1)(b-1)	MSAB
AC	SSAC	(a-1)(c-1)	MSAC
BC	SSBC	(b-1)(c-1)	MSBC
Residual	SSE	(a-1)(b-1)(c-1)	MSE
Total	SSTO	abc-1	

8 24.21

$$\begin{aligned}
Var(\hat{L}) &= Var(\sum \sum c_{ij} \bar{Y}_{ij..}) \\
&= \sum \sum c_{ij}^2 Var(\bar{Y}_{ij..}) \\
&= \sum \sum c_{ij}^2 \frac{\sigma^2}{cn} \\
&= \frac{\sigma^2}{cn} \sum \sum c_{ij}^2
\end{aligned}$$