STA207 homework4

Juanjuan Hu, Zhen Zhang

February 7, 2016

25.7

(a)

```
library(xlsx)
```

```
## Loading required package: rJava
## Loading required package: xlsxjars
```

```
setwd("/Users/PullingCarrot/Desktop/201601-03/sta207/homework/hw4")
sodium = read.xlsx2("DataFor+PR25.7.xlsx",1)
sodium$Sodium = as.numeric(as.character(sodium$Sodium))
sodium$Brand = as.factor(sodium$Brand)
result = aov(Sodium ~ Brand, data=sodium)
summary(result)
```

```
##             Df Sum Sq Mean Sq F value Pr(>F)
## Brand        5  854.5  170.91   238.7 <2e-16 ***
## Residuals   42   30.1    0.72
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

$H_0 : {\sigma_\mu}^2 = 0$

$H_1 : {\sigma_\mu}^2 > 0$

$F^* = \frac{MSTR}{MSE} = 170.91/0.72 = 237.375$

$F(0.99, 5, 42) = 3.49$

$p - val = 0$

The decision rule is: if F^* is greater than 3.49, then reject H_0, otherwise, accept H_0. Here, $237.375 > 3.49$, so we reject H_0, concluding that the mean sodium content is not the same in all brands sold in the metropolitan area. The p-value is almost zero, which is less than 0.01, leading to the same conclusion.

(b)

```
sum(sodium$Sodium)/nrow(sodium)
```

```
## [1] 17.62917
```

$\hat{\mu}_{.} = \bar{Y}_{..} = 17.63$

$s^2 = MSTR/r * n = 170.91/48 = 3.560625$

$s(\bar{Y}_{..}) = \sqrt{3.560625} = 1.887$

$t(0.995, 5) = 4.032$

The confidence interval is $\bar{Y}_{..} \pm t(1-\alpha, r-1) * s(\bar{Y}_{..})$, which is $17.63 \pm 4.032 * 1.887$. i.e. [10.021, 25.237].

25.8

(a)

According to (25.16), compute L and U.

$F(\alpha/2; r-1, r(n-1)) = F(0.005, 5, 42) = 0.0799$

$F(1-\alpha/2; r-1, r(n-1)) = F(0.995, 5, 42) = 3.9528$

$L = \frac{1}{n}(\frac{MSTR}{MSE} * \frac{1}{F(1-\alpha/2; r-1, r(n-1))} - 1) = \frac{1}{8}(\frac{170.91}{0.72} * \frac{1}{F(0.995,5,42)} - 1) = 7.4292$

$U = \frac{1}{n}(\frac{MSTR}{MSE} * \frac{1}{F(\alpha/2; r-1, r(n-1))} - 1) = \frac{1}{8}(\frac{170.91}{0.72} * \frac{1}{F(0.005,5,42)} - 1) = 375.2083$

$L^* = \frac{L}{1+L} = 7.4292/(1+7.4292) = 0.8814$

$U^* = \frac{U}{1+U} = 375.2083/(1+375.2083) = 0.9973$

Therefore, the 99% confidence interval for $\frac{(\sigma_\mu)^2}{(\sigma_\mu)^2+(\sigma)^2}$ is [0.8814, 0.9973]. It means that with confidence coefficient 0.99, we conclude that the variability of the mean sodium for the different brands accounts for somewhere between 88 and 99.8 percent of the total variability of the sodium levels.

(b)

According to (25.6), an unbiased estimator of σ^2 is MSE = 0.72. According to (25.23), an unbiased point estimator of ${\sigma_\mu}^2$ is ${s_\mu}^2 = \frac{MSTR-MSE}{n} = (170.91-0.72)/8 = 21.2738$.

(c)

$\chi^2(1-\alpha/2; 42) = \chi^2(0.995; 42) = 69.336$

$\chi^2(\alpha/2; 42) = \chi^2(0.005; 42) = 22.1385$

According to (25.21), the CI is: $\frac{r(n-1)MSE}{\chi^2(1-\alpha/2;42)} \leq \sigma^2 \leq \frac{r(n-1)MSE}{\chi^2(\alpha/2;42)}$. Here the 99% confidence interval is: $\frac{6(8-1)0.72}{\chi^2(0.995;42)} \leq \sigma^2 \leq \frac{6(8-1)0.72}{\chi^2(0.005;42)}$, i.e. [0.4361, 1.3659].

(d)

$H_0 : {\sigma_\mu}^2 \leq 2\sigma^2$

$H_1 : {\sigma_\mu}^2 > 2\sigma^2$

$F^* = \frac{MSTR/2n+1}{MSE} = (170.91/17)/0.72 = 13.96$

$F(0.99, 5, 42) = 3.49$

The decision rule is: if F^* is greater than 3.49, then reject H_0, otherwise, accept H_0. Here, $13.96 > 3.49$, so we reject H_0, concluding that the variance of sodium content between brands is more than twice as great as that within brands.

(e)

$c_1 = 1/n = 1/8 = 0.125$, $c_2 = -c_1 = -0.125$

$\hat{L} = c_1 MSTR + c_2 MSE = 0.125 * 170.91 - 0.125 * 0.72 = 21.2737$

$df_1 = r - 1 = 5$, $df_2 = r(n-1) = 6 * 7 = 42$

$F_1 = F(1-\alpha/2, df_1, \inf) = F(0.995, 5, \inf) = 3.35$

$F_2 = F(1-\alpha/2, df_2, \inf) = F(0.995, 42, \inf) = 1.66$

$F_3 = F(1-\alpha/2, \inf, df_1) = F(0.995, \inf, 5) = 12.1$

$F_4 = F(1-\alpha/2, \inf, df_2) = F(0.995, \inf, 42) = 1.91$

$F_5 = F(1-\alpha/2, df_1, df_2) = F(0.995, 5, 42) = 3.95$

$F_1 = F(1-\alpha/2, df_2, df_1) = F(0.995, 42, 5) = 12.51$

$G_1 = 1 - \frac{1}{F_1} = 1 - 1/3.35 = 0.7015$

$G_2 = 1 - \frac{1}{F_2} = 1 - 1/1.66 = 0.3976$

$G_3 = \frac{(F_5-1)^2 - (G_1 F_5)^2 - (F_4-1)^2}{F_5} = 0.0497$

$G_4 = F_6((\frac{F_6-1}{F_6})^2 - (\frac{F_3-1}{F_6})^2 - {G_2}^2) = -1.2371$

$H_L = 14.99$

$H_U = 237.127$

According to (25.34), the 99% confidence interval is $\hat{L} - H_L \leq L \leq \hat{L} + H_U$, which is [21.2737-14.99, 21.2737+237.127], i.e. [6.284, 258.401]. It means that with confidence coefficient 0.99, we conclude that the variability of the mean sodium between brands is somewhere between 6.28 and 258.4.

25.16

(a)

```r
setwd("/Users/PullingCarrot/Desktop/201601-03/sta207/homework/hw4")
disk = read.xlsx2("DataForPR25.16.xlsx",1)
sapply(disk, class)
```

```
##    Minutes Technician       Make
##   "factor"   "factor"   "factor"
```

```r
disk$Minutes = as.numeric(as.character(disk$Minutes))
model = aov(Minutes ~ Make + Error(Make*Technician), data = disk)
summary(model)
```

```
##
## Error: Make
##      Df Sum Sq Mean Sq
## Make  2  28.31   14.16
##
## Error: Technician
```

```
##           Df Sum Sq Mean Sq F value Pr(>F)
## Residuals  2  24.58   12.29
##
## Error: Make:Technician
##           Df Sum Sq Mean Sq F value Pr(>F)
## Residuals  4   1215   303.8
##
## Error: Within
##           Df Sum Sq Mean Sq F value Pr(>F)
## Residuals 36   1872   52.01
```

$H_0 : {\sigma_{\alpha\beta}}^2 = 0$

$H_1 : {\sigma_{\alpha\beta}}^2 > 0$

$F^* = \frac{MSAB}{MSE} = 303.8/52.01 = 5.8412$

$F(0.99, 4, 36) = 3.89$

$p - val = 0.0009944268$

The decision rule is: if F^* is greater than 3.89, then reject H_0, otherwise, accept H_0. Here, $5.8412 > 3.49$, so we reject H_0, concluding that the two factors interact. The p-value is nearly 0.001, which is less than 0.01, leading to the same conclusion.

(b)

$MSAB = 303.8$, $MSE = 52.01$

The point estimator of ${\sigma_{\alpha\beta}}^2$ is $\frac{MSAB-MSE}{n} = \frac{303.8-52.01}{5} = 50.358$, which is no larger than σ^2.

(c)

$H_0 : {\sigma_\alpha}^2 = 0$

$H_1 : {\sigma_\alpha}^2 > 0$

$F^* = \frac{MSA}{MSE} = 12.29/52.01 = 0.2363$

$F(0.99, 2, 36) = 5.25$

$p - val = 0.7907587$

The decision rule is: if F^* is greater than 5.25, then reject H_0, otherwise, conclude H_0. Here, $0.2363 < 5.25$, so we cannot reject H_0, concluding that the factor A main effects are not present. The p-value is 0.79, which is above 0.01, leading to the same conclusion. It is meaningful to test this becuase if we can reject H_0, saying no A effects, then we can simplify our model.

(d)

H_0: all $\beta_j = 0$

H_1: not all β_j equals 0

$F^* = \frac{MSB}{MSAB} = 14.16/303.8 = 0.0466$

$F(0.99, 2, 4) = 18$

$p - val = 0.9549795$

The decision rule is: if F^* is greater than 18, then reject H_0, otherwise, conclude H_0. Here, $0.0466 < 18$, so we cannot reject H_0, concluding that the factor B main effects are not present. The p-value is 0.955, which is above 0.01, leading to the same conclusion.It is meaningful to test this becuase if we can reject H_0, saying no B effects, then we can simplify our model.

(e)

```r
mean(subset(disk, Make==1)$Minutes)
```

```
## [1] 56.13333
```

```r
mean(subset(disk, Make==2)$Minutes)
```

```
## [1] 56.6
```

```r
mean(subset(disk, Make==3)$Minutes)
```

```
## [1] 54.73333
```

$\bar{Y}_{.1.} = 56.13$, $\bar{Y}_{.2.} = 56.6$,$\bar{Y}_{.3.} = 54.73$

$\hat{D}_1 = \bar{Y}_{.1.} - \bar{Y}_{.2.} = -0.47$, $\hat{D}_2 = \bar{Y}_{.1.} - \bar{Y}_{.3.} = 1.4$,$\hat{D}_3 = \bar{Y}_{.2.} - \bar{Y}_{.3.} = 1.87$

$s(\hat{D}_i) = \sqrt{\frac{MSAB}{bn} * 2} = \sqrt{\frac{303.8}{5*3} * 2} = 6.3645$

$q(0.95, 3, 4) = 5.04$

$T = q(0.95, 3, 4) * (1/\sqrt{2}) = 3.5638$

Therefore, using the Turkey procedure and a 0.95 family confidence coefficient, the confidence intervals are:

$\hat{D}_1 - s(\hat{D}_1) * T \leq D_1 \leq \hat{D}_1 + s(\hat{D}_1) * T$, i.e. $-23.15 \leq D_1 \leq 22.22$

$\hat{D}_2 - s(\hat{D}_2) * T \leq D_2 \leq \hat{D}_2 + s(\hat{D}_2) * T$, i.e. $-21.28 \leq D_2 \leq 24.08$

$\hat{D}_3 - s(\hat{D}_3) * T \leq D_3 \leq \hat{D}_3 + s(\hat{D}_3) * T$, i.e. $-20.81 \leq D_3 \leq 24.55$

We can see that the three CIs all includes zero, meaning that there are no significant difference between the means for three disk drive maker.

(f)

$\hat{\mu}_{.1} = \bar{Y}_{.1.} = 56.13$

$c_1 = (3-1)/(5*3*3) = 2/45$, $c_2 = 1/(5*3*3) = 1/45$

$s(\hat{\mu}_{.1}) = \sqrt{c_1 * MSAB + c_2 * MSA} = 3.712$

$df = s(\hat{\mu}_{.1})^4/((c_1 * MSAB)^2/4 + (c_2 * MSA^2/2) = 4.16$

$t(0.995, 4) = 4.6$

Therefore, using Satterthwaite procedure, the condidence interval is:

$\hat{\mu}_{.1} - t(0.995, 4) * s(\hat{\mu}_{.1}) \leq \mu_{.1} \leq \hat{\mu}_{.1} + t(0.995, 4) * s(\hat{\mu}_{.1})$, i.e. $39.05 \leq \mu_{.1} \leq 73.21$

It means that with confidence coefficient 0.99, we conclude that the the mean minutes for make 1 is somewhere between 39.05 and 73.21.

(g)

$MS1 = MSA$, $MS2 = MSE$

$c1 = 1/15$,$c2 = -1/15$

$\hat{L} = (MS1 - MS2)/(5 * 3) = -2.648$

$df1 = 2$, $df2 = 36$

$F1 = F(1 - 0.01/2, df1, \text{inf}) = 5.2983$

$F2 = F(1 - 0.01/2, df2, \text{inf}) = 1.7106$

$F3 = F(1 - 0.01/2, \text{inf}, df1) = 199.4996$

$F4 = F(1 - 0.01/2, \text{inf}, df2) = 2.0127$

$F5 = F(1 - 0.01/2, df1, df2) = 6.1606$

$F6 = F(1 - 0.01/2, df2, df1) = 199.4718$

$G1 = 1 - 1/F1 = 0.8113$

$G2 = 1 - 1/F2 = 0.4154$

$G3 = ((F5 - 1)^2 - (G1 * F5)^2 - (F4 - 1)^2)/F5 = 0.1022$

$G4 = F6 * ((1 - 1/F6)^2 - (F3/F6 - 1/F6)^2 - G2^2) = -35.3895$

$H_L = \sqrt{(G1 * c1 * MS1)^2 + ((F4 - 1) * c2 * MS2)^2 - G3 * c1 * c2 * MS1 * MS2} = 3.605$

$H_U = \sqrt{((F3 - 1) * c1 * MS1)^2 + (G2 * c2 * MS2)^2 - G4 * c1 * c2 * MS1 * MS2} = 162.73$

Therefore, using MLS procedure, the confidence interval is:

$\hat{L} - H_L \leq L \leq \hat{L} + H_U$

-2.648-3.605 $\leq$ L $\leq$ -2.648+162.73, i.e. -6.253 $\leq$ L $\leq$ 160.082

A negativa variance is not meaningful, so the CI can be written as [0, 160.082].

It means that with confidence coefficient 0.99, we conclude that the variability of mean minutes for different techinician is somewhere between 0 and 160.082.