Stat 207 HW2

Cheng Luo 912466499
Fan Wu 912538518

January 21, 2015

1 25.7

(a)

```r
sod = read.table("CH25PR07.txt")
names(sod) = c("Y", "A", "B")
fit = lm(Y ~ factor(A), data = sod)
anova(fit)
```

```
## Analysis of Variance Table
##
## Response: Y
##           Df Sum Sq Mean Sq F value    Pr(>F)
## factor(A)  5 854.53 170.906  238.71 < 2.2e-16 ***
## Residuals 42  30.07   0.716
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Therefore, test whether or not the mean sodium content is the same in all brands sold in the metropolitan area:

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

$$\text{VS. } H_1{:}\sigma_\mu^2 \neq 0$$

$$F^* = \frac{MSA}{MSE} = 238.71$$

we can reject H_0 if $F^* > F(1-0.01;5,42) = 3.488235$,otherwise reject$H_1$

so that reject H_0 because $F^* > F(1-0.05;2,48) = 3.488235$,

therefore,the mean sodium content is not the same in all brands sold in the metropolitan area

(b)

```r
y_bar = 17.62917
s = sqrt(170.906/48)
t = qt(0.995,5)
c(y_bar-s*t, y_bar+s*t)
```

```
## [1] 10.02076 25.23758
```

$$\bar{Y}.. = 17.62917$$

$$s(\bar{Y}..) = \sqrt{\frac{MSTR}{nr}} = \sqrt{\frac{170.906}{48}} = 1.88694$$

$$t(1-\alpha/2, r-1) = t(1-0.01/2, 6-1) = 4.032143$$

Therefore, the confident interval is $10.021 \leq \mu. \leq 25.237$.

2 25.8

(a)
```
f_low = qf(0.005, 5, 42)
f_high = qf(0.995, 5, 42)
f_star = 238.71
n = 8
u = 1/n*(f_star/f_low - 1)
l = 1/n*(f_star/f_high - 1)
L_star = l/(l+1)
U_star = u/(u+1)
c(L_star, U_star)
```

```
## [1] 0.8812875 0.9973277
```

So a 99% confidence interval for $\sigma_\mu^2/(\sigma_\mu^2+\sigma^2)$ is (0.8812875, 0.9973277)

(b) Since

$$E(MSE) = \sigma^2$$
$$E(MSTR) = n\sigma_\mu^2 + \sigma^2$$

MSE = 0.716 estimates σ^2, and $s_\mu^2 = \frac{MSTR-MSE}{n} = 21.27375$ estimates σ_μ^2

(c)
```
sse = 0.716
x1 = qchisq(0.005, 42)
x2 = qchisq(0.995, 42)
c(30.07/x2, 30.07/x1)
```

```
## [1] 0.4336853 1.3582695
```

Since

$SSE/\sigma^2 \sim \chi^2_{(r(n-1))}$, r=6,n=8

then $\frac{SSE}{\chi^2_{(0.995,42)}} \le \sigma^2 \le \frac{SSE}{\chi^2_{(0.005,42)}}$,which means the 95% confidence interval of σ^2 is (0.4336853,1.3582695)

(d) test:

$$H_0{:}\sigma_\mu^2 \le 2\sigma^2$$
$$\text{VS. } H_1{:}\sigma_\mu^2 > 2\sigma^2$$
$$F^* = \frac{MSTR/(2n+1)}{MSE} = 14.04091$$

we can reject H_0 if $F^* > F(1-0.01;5,42) = 3.488235$,otherwise reject$H_1$

so that reject H_0 because $F^* > F(1-0.05;2,48) = 3.488235$,

therefore,the variance of sodium content between brands is more than twice as great as that within brands.

(e)

```
c1=0.125
c2=-.125
ms1=170.906
ms2=0.716
df1=5
df2=42
F1=qf(.995,5,Inf)
F2=qf(.995,42,Inf)
F3=qf(.995,Inf,5)
F4=qf(.995,Inf,42)
F5=qf(.995,5,42)
F6=qf(.995,42,5)
G1=1-1/F1
G2=1-1/F2
G3=((F5-1)^2-(G1*F5)^2-(F4-1)^2)/F5
G4=F6*( ((F6-1)/F6)^2 - ((F3-1)/F6)^2 - G2^2 )
Hl = sqrt( (G1*c1*ms1)^2 + ((F4-1)*c2*ms2)^2 - G3*c1*c2*ms1*ms2)
Hl

## [1] 14.98986

Hu = sqrt( (G2*c2*ms2)^2 + ((F3-1)*c1*ms1)^2 - G4*c1*c2*ms1*ms2)
Hu

## [1] 238.0569

sigma_mu = 21.27375
c(sigma_mu-Hl, sigma_mu+Hu)

## [1]   6.283885 259.330628
```

E(MSTR)= $n\sigma_\mu^2 + \sigma^2$ E(MSE)=σ^2

Base on $L = \sigma_\mu^2 = c_1E(MSTR) + c_2E(MSE)$

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

and $MSTR = 170.906, MSE = 0.716, df1 = 5, df2 = 42$

According to R code, $H_l = 14.98986$ $H_u = 238.0569$ $\sigma_\mu^2 = 21.27375$

so that $6.283885 \le \sigma_\mu^2 \le 259.330628$

Confidence interval is very large, because the small sample sizes and the difficulty in estimating variance component precisely.

3 25.11

Let $\bar{(\alpha\beta)}^*_{.j}$ denote the mean of the unrestricted interaction terms $(\alpha\beta)^*_{.1}$,$(\alpha\beta)^*_{.2}\cdots(\alpha\beta)^*_{.n}$

so that $(\alpha\beta)_{ij} = (\alpha\beta)^*_{ij} - \bar{(\alpha\beta)}^*_{.j}$

Therefore, $\sum_i(\alpha\beta)_{ij} = \sum_i((\alpha\beta)^*_{ij} - \bar{(\alpha\beta)}^*_{.j}) = \sum_i(\alpha\beta)^*_{ij} - \sum_i(\alpha\beta)^*_{ij} = 0$

but $\sum_j(\alpha\beta)_{ij} = \sum_j((\alpha\beta)^*_{ij} - \bar{(\alpha\beta)}^*_{.j}) = \sum_j(\alpha\beta)^*_{ij} - \sum_j\bar{(\alpha\beta)}^*_{.j}$, usually it doesn't equal zero.

4 25.12

We should choose Two factors model (A fixed, B random, ANOVA III, mixed model)

$$Y_{ijk} = \mu_{..} + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ijk}$$

We choose this model, because there are only 3 possible price, so it's fixed, but we choose 3 colors randomly from many colors to represent the range of different color, so we use this model.

5 25.16

(a)
```
dat = read.table('CH19PR16.txt')
names(dat) = c('y', 'A', 'B', 'k')
dat$A = as.factor(dat$A)
dat$B = as.factor(dat$B)
a = length(unique(dat$'A'))
b = length(unique(dat$'B'))
n = length(unique(dat$'k'))
model = aov(y ~ B + Error(A*B), data = dat)
model.aov = summary(model)
qf(1-0.01,4,36)

## [1] 3.890308

pf(303.8/52.01,4,36,lower.tail = FALSE)

## [1] 0.0009944442
```

Test:

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

$$\text{VS. } H_1{:}\sigma^2_{\alpha\beta} > 0$$

$$F^* = \frac{MSAB}{MSE} = 303.8/52.01 = 5.841184$$

we can reject H_0 if $F^* > F(1-0.01;4,36) = 3.890308$,otherwise reject$H_1$

so that reject H_0 because $F^* > 3.890308$,

therefore,there are two factors interact, and P-value of it is 0.0009944442

(b) Since

$$\text{E(MSAB)-E(MSE)}=n\sigma^2_{\alpha\beta}$$

then

$$s^2_{\alpha\beta} = (MSAB - MSE)/n = (303.8 - 52.01)/5 = 50.358$$

therefore,$s^2_{\alpha\beta} = 50.358$ is estimate of $\sigma^2_{\alpha\beta}$, $MSE = 52.01$ is estimate of σ^2, so that $\sigma^2_{\alpha\beta}$ appears to be small relative to σ^2.

(c) test:

$$H_0:\sigma^2_\alpha = 0$$
$$\text{VS. } H_1:\sigma^2_\alpha > 0$$
$$F^* = \frac{MSA}{MSE} = 12.29/52.01 = 0.2363007$$

we can reject H_0 if $F^* > F(1-0.01;2,36) = 5.247894$,otherwise reject$H_1$

so that reject H_1 because $F^* < 5.247894$,

therefore,no factor A main effects are present, but the interaction effects are present.

(d) test:

$$H_0\text{:all } \beta_j \text{ equal zero(j=1,2,3)}$$
$$\text{VS. } H_1\text{:not all } \beta_j \text{ equal zero(j=1,2,3)}$$
$$F^* = \frac{MSB}{MSAB} = 14.16/303.8 = 0.04660961$$

we can reject H_0 if $F^* > F(1-0.01;2,4) = 18$,otherwise rejectH_1

so that reject H_1 because $F^* < 18$,

therefore,no factor B main effects are present, but the interaction effects are present.

(e)
```
means_j = with(dat, by(y, B, mean))
means_j
```

```
## B: 1
## [1] 56.13333
## ------------------------------------------------------------
## B: 2
## [1] 56.6
## ------------------------------------------------------------
## B: 3
## [1] 54.73333

  D1 = means_j[1]-means_j[2]
  D2 = means_j[1]-means_j[3]
  D3 = means_j[2]-means_j[3]
  MSAB = 303.8
  s = sqrt(MSAB/(n*a)*(1+1))
  alpha = .05
  q. = 1/sqrt(2)*qtukey(1-alpha, b, (a-1)*(b-1))
  c(D1-s*q., D1+s*q.)

##         1         1
## -23.14962  22.21629

  c(D2-s*q., D2+s*q.)

##         1         1
## -21.28295  24.08295

  c(D3-s*q., D3+s*q.)

##         2         2
## -20.81629  24.54962
```

$$\bar{Y}_{.1.} = 56.13333, \bar{Y}_{.2.} = 56.6, \bar{Y}_{.3.} = 54.73333$$

$$\hat{D}_1 = \bar{Y}_{.1.} - \bar{Y}_{.2.} = -0.4666667, \hat{D}_2 = \bar{Y}_{.1.} - \bar{Y}_{.3.} = 1.4, \hat{D}_3 = \bar{Y}_{.2.} - \bar{Y}_{.3.} = 1.866667$$

$$S = \sqrt{\frac{MSAB}{na}\sum c_i} = 6.364485, T = \frac{1}{\sqrt{2}}\text{qtukey}(1 - alpha, b, (a-1)*(b-1)) = 3.563989$$

$$\text{base on}\hat{D}_i \pm S * T$$

$$-23.14962 \leq D_1 \leq 22.21629$$

$$-21.28295 \leq D_2 \leq 24.08295$$

$$-20.81629 \leq D_3 \leq 24.54962$$

It means D1,D2,D3 can equal to zero, there's no significant factor B effect.

(f)

```
mu_j1 = means_j[1]
mu_j1

##        1
## 56.13333

MSA = 12.29
MSAB = 303.8
c1 = (a-1)/(n*a*b)
c2 = 1/(n*a*b)
s = sqrt(c1*MSAB+c2*MSA)
s

## [1] 3.711514

df = s^4/( (c1*MSAB)^2/((a-1)*(b-1))+ (c2*MSA)^2/((a-1)))
t = qt(1-0.01/2,(df))
t

## [1] 4.485356

c(mu_j1-s*t, mu_j1+s*t)

##        1        1
## 39.48587 72.78079
```

$$\begin{aligned}
\hat{\mu}_{\cdot 1} =& 56.13333, MSA = 12.29, MSAB = 303.8 \\
c_1 =& \frac{a-1}{nab} = 0.04444444, c_2 = \frac{1}{nab} = 0.02222222 \\
s =& \sqrt{c_1 * MSAB + c_2 * MSA} = 3.711514 \\
df =& \frac{s^4}{\frac{(\frac{a-1}{nab}MSAB)^2}{(a-1)(b-1)} + \frac{(\frac{1}{nab}MSA)^2}{(a-1)}} = 4.160049 \\
t =& t(1-\alpha/2; df) = 4.485356
\end{aligned}$$

confidence limits $\hat{\mu}_{\cdot i} \pm s * t$

$$39.48587 \leq \mu_{\cdot 1} \leq 72.78079$$

We are 99% confident that $\mu_{\cdot 1}$ is between (39.48587, 72.78079)

(g)

```
c1=1/15
c2=-1/15
ms1=12.29
ms2=52.01
df1=2
df2=36
F1=qf(.995,2,Inf)
F2=qf(.995,36,Inf)
F3=qf(.995,Inf,2)
F4=qf(.995,Inf,36)
F5=qf(.995,2,36)
F6=qf(.995,36,2)
G1=1-1/F1
G2=1-1/F2
G3=((F5-1)^2-(G1*F5)^2-(F4-1)^2)/F5
G4=F6*( ((F6-1)/F6)^2 - ((F3-1)/F6)^2 - G2^2 )
Hl = sqrt( (G1*c1*ms1)^2 + ((F4-1)*c2*ms2)^2 - G3*c1*c2*ms1*ms2)
Hl
```

```
## [1] 3.613885
```

```
Hu = sqrt( (G2*c2*ms2)^2 + ((F3-1)*c1*ms1)^2 - G4*c1*c2*ms1*ms2)
Hu
```

```
## [1] 162.3423
```

```
sigma_mu =  (ms1-ms2)/(n*b)
c(max(0,0-Hl), 0+Hu)
```

```
## [1]   0.0000 162.3423
```

E(MSA)= $nb\sigma_\mu^2 + \sigma^2$ E(MSE)=σ^2
Base on $L = \sigma_\mu^2 = c_1 E(MSA) + c_2 E(MSE)$
then $c_1 = 1/(nb) = 0.06666667$, $c_2 = -1/(nb) = -0.06666667$
and $MSA = 12.29, MSE = 52.01, df1 = 2, df2 = 36$

According to MLS procedure, $H_l = 3.613885$ $H_u = 162.3423$ $\sigma_\mu^2 = -2.648$

since $\sigma_\mu^2 = -2.648 < 0$, so that $\sigma_\mu^2 = 0$

so that 99% confident interval is $max(0, 0 - H_l) \leq \sigma_\mu^2 \leq 0 + H_u$,which means $0 \leq \sigma_\mu^2 \leq 162.3423$

Confidence interval is very large, because the small sample sizes and the difficulty in estimating variance component precisely.

6 25.30

$$
\begin{aligned}
L \leq \frac{\sigma_\mu^2}{\sigma^2} \leq U \\
\frac{1}{L} \geq \frac{\sigma^2}{\sigma_\mu^2} \geq \frac{1}{U} \\
\frac{1+L}{L} \geq \frac{\sigma^2+\sigma_\mu^2}{\sigma_\mu^2} \geq \frac{1+U}{U} \\
\frac{L}{1+L} \leq \frac{\sigma_\mu^2}{\sigma^2+\sigma_\mu^2} \leq \frac{U}{1+U}
\end{aligned}
$$

7 25.32

$$
\begin{aligned}
Y_{ij} &= \mu_{\cdot\cdot} + \rho_i + \tau_j + \epsilon_{ij} \\
\text{then } \sigma^2\{Y_{ij}\} &= \sigma^2\{\mu_{\cdot\cdot} + \rho_i + \tau_j + \epsilon_{ij}\} = \sigma_\tau^2 + \sigma^2 \\
\sigma^2\{\bar{Y}_{\cdot j}\} &= \sigma^2\{\mu_{\cdot\cdot} + \frac{\sum \rho_i}{n_b} + \tau_j + \bar{\epsilon}_{\cdot j}\} = \sigma_\tau^2 + \sigma^2/n_b
\end{aligned}
$$

8 25.34

$$
\begin{aligned}
\sigma^2\{Y_{ij}, Y_{ij'}\} &= E\{(Y_{ij} - E(Y_{ij}))(Y_{ij'} - E(Y_{ij'}))\} \\
&= E\{[\mu_{\cdot\cdot} + \rho_i + \tau_j + \epsilon_{ij} - (\mu_{\cdot\cdot} + \tau_j)][\mu_{\cdot\cdot} + \rho_i + \tau_{j'} + \epsilon_{ij'} - (\mu_{\cdot\cdot} + \tau_{j'})]\} \\
&= E\{(\rho_i + \epsilon_{ij})(\rho_i + \epsilon_{ij'})\} \\
&= E(\rho_i^2) + E(\rho_i \epsilon_{ij'}) + E(\rho_i \epsilon_{ij}) + E(\epsilon_{ij}\epsilon_{ij'}) \\
&= (E(\rho_i))^2 + \sigma_\rho^2 \\
&= \sigma_\rho^2
\end{aligned}
$$