Stat 207 HW3

Cheng Luo 912466499
Fan Wu 912538518

January 26, 2015

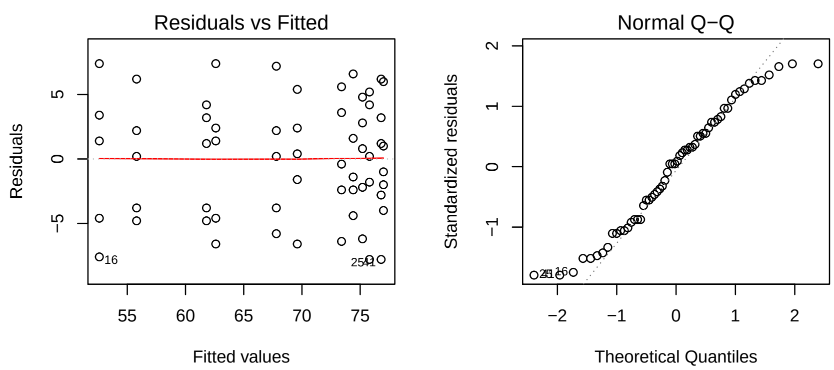
1 26.4

(a)

```
dat = read.table("CH26PR04.txt")
names(dat) = c("Y", "A", "B", "k")
dat$A = factor(dat$A)
dat$B = factor(dat$B)
a = length(unique(dat$A))
b = length(unique(dat$B))
n = length(unique(dat$k))
model = aov(Y ~ A + A/B, data = dat)
resid(model)
```

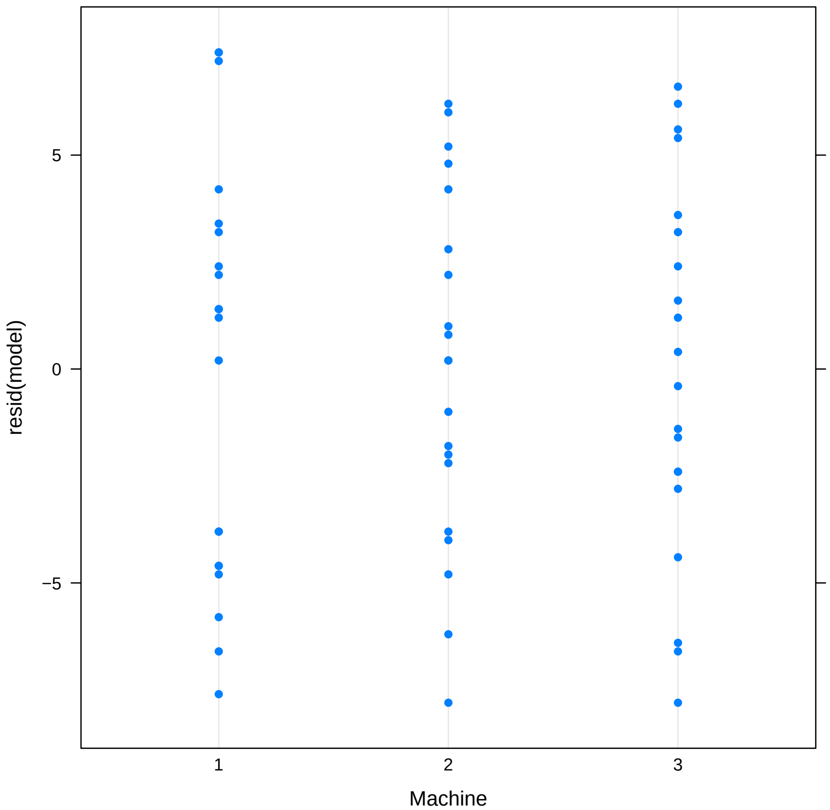
```
##    1    2    3    4    5    6    7    8    9   10   11   12   13   14   15
##  3.2 -3.8  1.2 -4.8  4.2  0.2 -5.8  7.2 -3.8  2.2 -6.6  2.4 -4.6  7.4  1.4
##   16   17   18   19   20   21   22   23   24   25   26   27   28   29   30
## -7.6  3.4  1.4 -4.6  7.4 -1.8  5.2  0.2  4.2 -7.8 -6.2  0.8  4.8  2.8 -2.2
##   31   32   33   34   35   36   37   38   39   40   41   42   43   44   45
## -3.8  0.2  6.2  2.2 -4.8 -4.0  1.0  6.0 -2.0 -1.0 -7.8  6.2 -2.8  1.2  3.2
##   46   47   48   49   50   51   52   53   54   55   56   57   58   59   60
## -6.6  0.4  2.4 -1.6  5.4  6.6 -2.4 -1.4  1.6 -4.4 -6.4  5.6 -0.4  3.6 -2.4
```

```
par( mfrow = c(2,2))
plot(model, which = 1)
plot(model, which = 2)
```



The residuals versus fitted values plots shows no sign for unequal variance.And the QQ-plot indicates approximately normal distribution with slightly light tail, so that normality assumption seems to be reasonable, we can use model(26.7) here.

(b)
```r
require("lattice")
## Loading required package:  lattice
dotplot(resid(model) ~ dat$A, xlab = "Machine" )
```

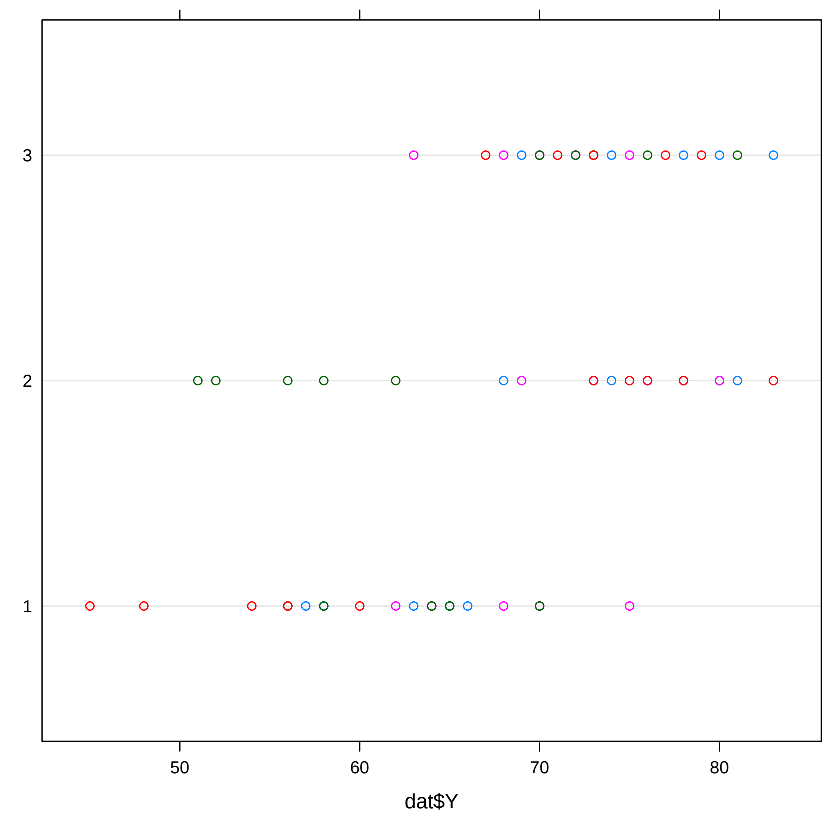


The plot shows no sign for unequal variance, so it support the assumption of constancy of the error variance.

2 26.5

(a) No.

(b)
```
dotplot(dat$A ~ dat$Y, groups = dat$B)
```



It seems operator effect are present.

(c)
```
summary(model)

##             Df Sum Sq Mean Sq F value   Pr(>F)
## A            2   1696   847.8   35.92 2.90e-10 ***
## A:B          9   2272   252.5   10.70 6.99e-09 ***
## Residuals   48   1133    23.6
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Test the mean outputs differ for three machine:

$$H_0\text{:all } \alpha_i \text{ equal zero(i=1,2,3)}$$
$$\text{VS. } H_1\text{:not all } \alpha_i \text{ equal zero}$$
$$F^* = \frac{MSA}{MSE} = 847.8/23.6 = 35.92$$

we can reject H_0 if $F^* > F(1-0.01;2,48) = 5.076664$,otherwise reject$H_1$

so that reject H_0 because $F^* > 5.076664$,

therefore, the mean outputs differ for three machine, and the P-value is 2.90e-10

(d) Test the mean outputs differ for the operator:

H_0:all $\beta_{j(i)}$ equal zero(i=1,2,3)

VS. H_1:not all $\beta_{j(i)}$ equal zero

$$F^* = \frac{MSB(A)}{MSE} = 252.5/23.6 = 10.7$$

we can reject H_0 if $F^* > F(1-0.01;9,48) = 2.801816$,otherwise reject$H_1$

so that reject H_0 because $F^* > 2.801816$,

therefore, our conclusion implies that operator within at least one machine differ in terms of mean shifts effects, and the P-value is 6.99e-09

(e)
```
model1 = aov(Y ~ A + A/B, data = dat)
```

H_0:all $\beta_{j(i)}$ equal zero(i=1,2,3)

VS. H_1:not all $\beta_{j(i)}$ equal zero

$$F^* = \frac{MSB(A)}{MSE} = 252.5/23.6 = 10.7$$

we can reject H_0 if $F^* > F(1-0.01;9,48) = 2.801816$,otherwise reject$H_1$

so that reject H_0 because $F^* > 2.801816$,

therefore, our conclusion implies that operator within at least one machine differ in terms of mean shifts effects, and the P-value is 6.99e-09

H_0:all $\beta_{j(i)}$ equal zero(i=1,2,3)

VS. H_1:not all $\beta_{j(i)}$ equal zero

$$F^* = \frac{MSB(A)}{MSE} = 252.5/23.6 = 10.7$$

we can reject H_0 if $F^* > F(1-0.01;9,48) = 2.801816$,otherwise reject$H_1$

so that reject H_0 because $F^* > 2.801816$,

therefore, our conclusion implies that operator within at least one machine differ in terms of mean shifts effects, and the P-value is 6.99e-09

H_0:all $\beta_{j(i)}$ equal zero(i=1,2,3)

VS. H_1:not all $\beta_{j(i)}$ equal zero

$$F^* = \frac{MSB(A)}{MSE} = 252.5/23.6 = 10.7$$

$$\text{we can reject } H_0 \text{ if } F^* > F(1-0.01; 9, 48) = 2.801816\text{,otherwise reject} H_1$$
$$\text{so that reject } H_0 \text{ because } F^* > 2.801816,$$

therefore, our conclusion implies that operator within at least one machine differ in terms of mean shifts effects, and the P-value is 6.99e-09

(f)

$$\begin{aligned}
\alpha &\le 1-(1-\alpha_1)...(1-\alpha_5) \\
&= 1-(1-0.05)^5 \\
&= 0.2262191
\end{aligned}$$

We conclude that three machine differ in mean output, 4 operators in machine 1 have different mean output effect, 4 operators in machine 2 have different mean output effect, but 4 operators in machine 3 do not have different mean output effect.

3 26.6

(a)

```
means = with(dat, by(Y, A, mean))
D1 = means[1] - means[2]
D2 = means[1] - means[3]
D3 = means[2] - means[3]
tukey = 1/sqrt(2)*qtukey(0.95, 3, 48)
tukey
```

```
## [1] 2.418488
```

```
mse = 23.6
s = sqrt(2*mse/(b*n))
s
```

```
## [1] 1.536229
```

```
c(D1-s*tukey, D1+s*tukey)
```

```
##          1          1
## -13.465351  -6.034649
```

```
c(D2-s*tukey, D2+s*tukey)
```

```
##          1          1
## -16.065351  -8.634649
```

```
c(D3-s*tukey, D3+s*tukey)
```

```
##         2         2
## -6.315351  1.115351
```

$$\bar{Y}_{1..} = 61.2, \bar{Y}_{2..} = 70.95, \bar{Y}_{3..} = 73.55$$

$$\hat{D}_1 = \bar{Y}_{1..} - \bar{Y}_{2..} = -9.75, \hat{D}_2 = \bar{Y}_{1..} - \bar{Y}_{3..} = -12.35, \hat{D}_3 = \bar{Y}_{2..} - \bar{Y}_{3..} = -2.6$$

$$S = \sqrt{\frac{MSE}{bn} * 2} = 1.536229, Tukey = \frac{1}{\sqrt{2}}\text{qtukey}(1 - alpha, a, ab(n-1)) = 2.418488$$

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

$$-13.465351 \le D_1 \le -6.034649$$

$$-16.065351 \le D_2 \le -8.634649$$

$$-6.315351 \le D_3 \le 1.115351$$

We conclude that with 95% family confidence that the mean output is highest in machine 3, and the difference between machine 2 and machine 3 is not statistically significant.

(b)

```
means = with(dat, by(Y, list(A,B) , mean))
D1 = means[1,1] - means[1,2]
D2 = means[1,1] - means[1,3]
D3 = means[1,1] - means[1,4]
D4 = means[1,2] - means[1,3]
D5 = means[1,2] - means[1,4]
D6 = means[1,3] - means[1,4]
B = qt(1-0.05/(2*6), (n-1)*a*b)
B

## [1] 2.752023

mse = 23.6
s = sqrt(2*mse/(n))
s

## [1] 3.072458

c(D1-s*B, D1+s*B)

## [1] -14.455477   2.455477

c(D2-s*B, D2+s*B)

## [1] -9.255477  7.655477

c(D3-s*B, D3+s*B)
```

```
## [1]  0.7445233 17.6554767

  c(D4-s*B, D4+s*B)

## [1] -3.255477 13.655477

  c(D5-s*B, D5+s*B)

## [1]  6.744523 23.655477

  c(D6-s*B, D6+s*B)

## [1]  1.544523 18.455477
```

$$\bar{Y}_{11\cdot} = 61.8, \bar{Y}_{12\cdot} = 67.8, \bar{Y}_{13\cdot} = 62.6, \bar{Y}_{14\cdot} = 52.6$$
$$\hat{D}_1 = \bar{Y}_{11\cdot} - \bar{Y}_{12\cdot} = -6, \hat{D}_2 = \bar{Y}_{11\cdot} - \bar{Y}_{13\cdot} = -0.8, \hat{D}_3 = \bar{Y}_{11\cdot} - \bar{Y}_{14\cdot} = -9.2$$
$$\hat{D}_4 = \bar{Y}_{12\cdot} - \bar{Y}_{13\cdot} = 5.2, \hat{D}_5 = \bar{Y}_{12\cdot} - \bar{Y}_{14\cdot} = 15.2, \hat{D}_6 = \bar{Y}_{13\cdot} - \bar{Y}_{14\cdot} = 10$$
$$S = \sqrt{\frac{MSE}{n} * 2} = 3.072458, B = t(1 - \alpha/(2*6), ab(n-1)) = 2.752023$$
$$\text{base on} \hat{D}_i \pm S * B$$
$$-14.455477 \leq D_1 \leq 2.455477$$
$$-9.255477 \leq D_2 \leq 7.655477$$
$$0.7445233 \leq D_3 \leq 17.6554767$$
$$-3.255477 \leq D_4 \leq 13.655477$$
$$6.744523 \leq D_5 \leq 23.655477$$
$$1.544523 \leq D_6 \leq 18.455477$$

We conclude that with 95% family confidence in machine 1 the differences between operator 1 and operator 2, operator 1 and operator 3, operator 2 and operator 3 are not statistically significant.

(c)
```
  L_hat = (means[1,1]+means[1,2]+means[1,3])/3 - means[1,4]
  s = sqrt(mse/(n)*((1/3)^2*3+1))
  s

## [1] 2.508652

  t = qt(1-0.01/2, a*b*(n-1))
  t
```

```
## [1] 2.682204

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

## [1]  4.737951 18.195382
```

$$
\begin{aligned}
\hat{L} &= \frac{\bar{Y}_{11.} + \bar{Y}_{12.} + \bar{Y}_{13.}}{3} + \bar{Y}_{14.} = 11.46667 \\
c_1 &= c_2 = c_3 = 1/3, c_4 = -1 \\
S &= \sqrt{\frac{MSE}{n} * \sum_i c_i^2} = 2.508652 \\
t &= t(1-\alpha/2, ab(n-1)) = 2.682204
\end{aligned}
$$

base on $\hat{L} \pm S * t$

$4.737951 \le D_1 \le 18.195382$

We are 99% confident that L is between 0.737951 and 18.195382.

4 26.7

(a) $\beta_{j(i)}$ are $i.i.d$ $N(0, \sigma_\beta^2$, and $\beta_{j(i)}$ and ϵ_{ijk} are independent.

(b)

```
  model_new = aov(Y ~ A+ Error(A/B), data = dat)
  summary(model_new)

##
## Error: A
##   Df Sum Sq Mean Sq
## A  2   1696   847.8
##
## Error: A:B
##           Df Sum Sq Mean Sq F value Pr(>F)
## Residuals  9   2272   252.5
##
## Error: Within
##           Df Sum Sq Mean Sq F value Pr(>F)
## Residuals 48   1133    23.6

  s_square = (252.5-23.6)/n
  s_square

## [1] 45.78
```

$$E(MSB(A)) = \sigma^2 + n\sigma_\beta^2$$
$$E(MSE) = \sigma^2$$
$$\hat{\sigma}_\beta^2 = s_\beta^2 = (MSB(A) - MSE)/n = 45.78$$

(c) Test:

$$H_0{:}\sigma_\beta^2 = 0$$
$$\text{VS. } H_1{:}\sigma_\beta^2 \neq 0$$
$$F^* = \frac{MSB(A)}{MSE} = 252.5/23.6 = 10.7$$

we can reject H_0 if $F^* > F(1 - 0.01; 9, 48) = 1.765318$,otherwise reject$H_1$

so that reject H_0 because $F^* > 1.765318$,

therefore, our conclusion implies that operator within at least one machine differ in terms of mean shifts effects, and the P-value is 6.976208e-09

```r
c1=1/5
c2=-1/5
ms1=252.5
ms2=23.6
df1=9
df2=48
F1=qf(.95,df1,Inf)
F2=qf(.95,df2,Inf)
F3=qf(.95,Inf,df1)
F4=qf(.95,Inf,df2)
F5=qf(.95,df1,df2)
F6=qf(.95,df2,df1)
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] 23.77552

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

## [1] 86.09976
```

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

## [1]  22.00448 131.87976
```

$$\text{E(MSB(A))}= nb\sigma_\beta^2 + \sigma^2 \qquad \text{E(MSE)}=\sigma^2$$
$$\text{Base on } L = \sigma_\mu^2 = c_1 E(MSB(A)) + c_2 E(MSE)$$
$$\text{then } c_1 = 1/(n) = 0.2,\ c_2 = -1/(n) = -0.2$$
$$\text{and } MSB(A) = 252.5, MSE = 23.6, df1 = 9, df2 = 48$$

According to MLS procedure, $H_l = 22.00448 \qquad H_u = 131.87976 \qquad \sigma_\beta^2 = 45.78$

so that 90% confident interval is $s_\beta^2 - H_l \leq \sigma_\mu^2 \leq s_\beta^2 + H_u$,which means $22.00448 \leq \sigma_\beta^2 \leq 131.87976$

(d) Test the mean outputs differ for three machine:

$$H_0\text{:all } \alpha_i \text{ equal zero(i=1,2,3)}$$
$$\text{VS. } H_1\text{:not all } \alpha_i \text{ equal zero}$$
$$F^* = \frac{MSA}{MSB(A)} = 847.8/252.5 = 3.357624$$

we can reject H_0 if $F^* > F(1-0.1; 2, 9) = 3.006452$,otherwise reject$H_1$

so that reject H_0 because $F^* > 3.006452$,

therefore, the mean outputs differ for three machine, and the P-value is 0.08140399

(e)

```r
means = with(dat, by(Y, A, mean))
D1 = means[1] - means[2]
D2 = means[1] - means[3]
D3 = means[2] - means[3]
tukey = 1/sqrt(2)*qtukey(0.9, 3, 9)
tukey

## [1] 2.344595

msb_a = 252.5
s = sqrt(2*msb_a/(b*n))
s

## [1] 5.024938

c(D1-s*tukey, D1+s*tukey)
```

```
##           1           1
## -21.531444   2.031444
```

```r
c(D2-s*tukey, D2+s*tukey)
```

```
##           1           1
## -24.131444  -0.568556
```

```r
c(D3-s*tukey, D3+s*tukey)
```

```
##           2           2
## -14.381444   9.181444
```

$$\bar{Y}_{1..} = 61.2, \bar{Y}_{2..} = 70.95, \bar{Y}_{3..} = 73.55$$

$$\hat{D}_1 = \bar{Y}_{1..} - \bar{Y}_{2..} = -9.75, \hat{D}_2 = \bar{Y}_{1..} - \bar{Y}_{3..} = -12.35, \hat{D}_3 = \bar{Y}_{2..} - \bar{Y}_{3..} = -2.6$$

$$S = \sqrt{\frac{MSB(A)}{bn} * 2} = 5.024938, Tukey = \frac{1}{\sqrt{2}}\text{qtukey}(1 - alpha, a, a(b-1) = 2.344595$$

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

$$-21.531444 \le D_1 \le 2.031444$$

$$-24.131444 \le D_2 \le -0.568556$$

$$-14.381444 \le D_3 \le 9.181444$$

We conclude that with 95% family confidence that the mean output is highest in machine 3, and the differences between machine 2 and machine 3, machine 1 and machine 2 are not statistically significant.

(f)

```r
dat_reg = dat
med1 = median(dat_reg$Y[dat_reg$A==1])
med2 = median(dat_reg$Y[dat_reg$A==2])
med3 = median(dat_reg$Y[dat_reg$A==3])
d = abs(dat_reg$Y - c(rep(med1,20),rep(med2,20),rep(med3,20)))
model_fi = aov(d ~ factor(dat_reg$A))
summary(model_fi)
```

```
##                   Df Sum Sq Mean Sq F value Pr(>F)
## factor(dat_reg$A)  2    109   54.52   1.923  0.155
## Residuals         57   1616   28.35
```

Test:

Set $d_{ij} = |Y_{ij} - \tilde{Y}_i|$

H_0:all $\sigma^2(\beta_{j(i)})$ are equal(i=1,2,3)

VS. H_1:not all $\sigma^2(\beta_{j(i)})$ are equal zero

$F^*_{BF} = \frac{MSTR(d)}{MSE(d)} = 109.85/34.02 = 3.228983$

we can reject H_0 if $F^* > F(1-0.01;2,9) = 8.021517$,otherwise reject$H_1$

so that reject H_1 because $F^* < 8.021517$,

therefore, all $\sigma^2(\beta_{j(i)})$ are equal(i=1,2,3)

5 26.19

6 26.20

7 26.24

$$SSB + SSAB = na\sum_j(\bar{Y}_{\cdot j \cdot} - \bar{Y}_{\cdots})^2 + n\sum_i\sum_j(\bar{Y}_{ij\cdot} - \bar{Y}_{i\cdot\cdot} - \bar{Y}_{\cdot j\cdot} + \bar{Y}_{\cdots})^2$$

8 26.25

(a) Since

$$\bar{Y}_{ijk} = \mu_{ij} + \alpha_i + \beta_{j(i)} + \epsilon_{ijk}$$

then:

$$\begin{aligned}
\sigma^2\{\bar{Y}_{i\cdot\cdot}\} &= \sigma^2\{\mu_{i\cdot\cdot} + \alpha_i + \bar{\beta}_{\cdot(i)} + \bar{\epsilon}_{i\cdot\cdot}\} \\
&= \sigma^2\{\bar{\beta}_{\cdot(i)} + \bar{\epsilon}_{i\cdot\cdot}\} \\
&= \frac{\sigma^2_\beta}{b} + \frac{\sigma^2}{bn} \qquad \text{,since } \beta \text{ and } \epsilon \text{ are independent}
\end{aligned}$$

$$\begin{aligned}
\sigma^2\{\bar{Y}_{\cdots}\} &= \sigma^2\{\mu_{\cdots} + \bar{\beta}_{\cdot(\cdot)} + \bar{\epsilon}_{\cdots}\} \qquad \text{,since } \sum_i \alpha = 0 \\
&= \sigma^2\{\bar{\beta}_{\cdot(\cdot)} + \bar{\epsilon}_{\cdots}\} \\
&= \frac{\sigma^2_\beta}{ab} + \frac{\sigma^2}{abn} \qquad \text{,since } \beta \text{ and } \epsilon \text{ are independent}
\end{aligned}$$

(b)

$$\begin{aligned}
E(MSB(A)) &= \sigma^2 + n\sigma^2_\beta \\
E(MSE) &= \sigma^2 \\
s^2_\beta &= (MSB(A) - MSE)/n \\
\hat{\sigma}^2_\beta &= max(0, s^2_\beta) = max(0, (MSB(A) - MSE)/n)
\end{aligned}$$

9 26.28