

STAT 511 Solution to HW4 Spring 2008

1. (a) Since $x_1 + x_2 + x_3 = 1$, we know that

$$\begin{aligned} y_i &= \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \epsilon_i \\ &= \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 (1 - x_{1i} - x_{2i}) + \epsilon_i \\ &= \beta_3 + (\beta_1 - \beta_3)x_{1i} + (\beta_2 - \beta_3)x_{2i} + \epsilon_i \end{aligned}$$

Taking $\alpha_0 = \beta_3, \alpha_1 = \beta_1 - \beta_3, \alpha_2 = \beta_2 - \beta_3$, and vice versa. Thus (1A) and (1B) equal.

$$\begin{aligned} y_i &= \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_4 x_{1i} x_{2i} + \beta_5 x_{1i} x_{3i} + \beta_6 x_{2i} x_{3i} + \epsilon_i \\ &= \alpha_0 + \alpha_1 x_{1i} + \alpha_2 x_{2i} + \beta_4 x_{1i} x_{2i} + \beta_5 x_{1i} x_{3i} + \beta_6 x_{2i} x_{3i} + \epsilon_i \\ &= \alpha_0 + (\alpha_1 + \beta_5)x_{1i} + (\alpha_2 + \beta_6)x_{2i} - \beta_5 x_{1i}^2 - \beta_6 x_{2i}^2 + (\beta_4 - \beta_5 - \beta_6)x_{1i} x_{2i} + \epsilon_i \\ &= \alpha'_0 + \alpha'_1 x_{1i} + \alpha'_2 x_{2i} + \alpha'_3 x_{1i}^2 + \alpha'_4 x_{2i}^2 + \alpha'_5 x_{1i} x_{2i} + \epsilon_i, \end{aligned}$$

where α'_i are the corresponding coefficients. Therefore, (2A) and (2B) is equal.

- (b) Fitting the model (1B) using

```
lm(y~x1+x2,data=species)
```

Then we get the estimate of α as $\alpha = (0.68482, 0.32964, -0.02286)$. According to (a), we know the relationship between α and β . That is

$$\begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} \alpha_0 \\ \alpha_1 \\ \alpha_2 \end{pmatrix}$$

So an estimate of β is $\beta = (1.01446, 0.66196, 0.68482)$. The estimated covariance matrix for β is,

$$\begin{pmatrix} 0.0043770513 & 0.0009364724 & -0.005899193 \\ 0.0009364724 & 0.0023417792 & -0.004493887 \\ -0.0058991934 & -0.0044938865 & 0.013987728 \end{pmatrix}$$

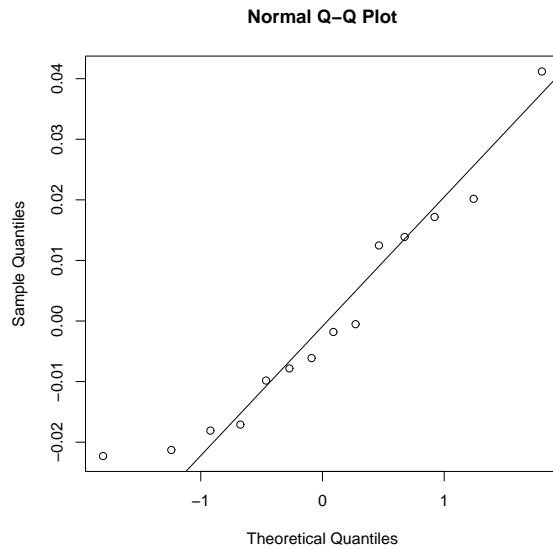
We also can fit model (1A) using

```
lm(y~x1+x2+x3-1,data=species)
```

- (c) We fit model (2B) and plot a normal qq plot using the following code and get a normal qq plot.

```
lm2<-lm(y~x1+x2+I(x1^2)+I(x2^2)+I(x1*x2),data=species)
qqnorm(lm2$residuals)
qqline(lm2$residuals)
```

- (d) Model comparison is used to test $H_0 : \alpha_3 = \alpha_4 = \alpha_5$. We get the following table:



```

Model 1: y ~ x1 + x2
Model 2: y ~ x1 + x2 + x11 + x22 + x12
Res. Df      RSS Df Sum of Sq      F Pr(>F)
1      11 0.0136460
2       8 0.0045154   3  0.0091306  5.3923 0.02528 *

```

Thus the quadratic curvature in response appear to be statistically detectable.

(e) The quadratic regression function is

$$y = -2.407 + 9.388x_1 + 8.283x_2 - 7.098x_1^2 - 5.738x_2^2 - 11.23x_1x_2$$

So the maximum is achieved at $(x_1, x_2) = (0.3999576, 0.3303831)$. The confidence interval can be found through following code:

```

eq<-matrix(c(2*7.098,11.23,11.23,2*5.738),2,2)
x0<-solve(eq)%*c(9.388,8.283)
C<-matrix(c(1,x0[1],x0[2],x0[1]^2,x0[2]^2,x0[1]*x0[2]),nrow=1,byrow=T)
beta<-lm(y~x1+x2+x11+x22+x12,data=species)$coefficients
library(MASS) X<-as.matrix(cbind(rep(1,14),species[,-(3:4)]))
C%%beta+qt(0.95,8)*sqrt((C%%solve(t(X)%*%X)%*%t(C))*0.000564)
C%%beta-qt(0.95,8)*sqrt((C%%solve(t(X)%*%X)%*%t(C))*0.000564)
C%%beta+qt(0.95,8)*sqrt((C%%solve(t(X)%*%X)%*%t(C)+1)*0.000564)
C%%beta-qt(0.95,8)*sqrt((C%%solve(t(X)%*%X)%*%t(C)+1)*0.000564)

```

Thus, the confidence interval for mean response estimate is (0.815062, 0.8632381) and the confidence interval for the prediction of the mean response is (0.788846, 0.8894541).

- Since we know that $C(x_1, x_2) = C(x_1, x_3)$, we can use x_1, x_3 as X^* to do the cell means model. In the model (2B) the model contains quadratic form, but in cell means model, the quadratic term are in the same space of $C(x_1, x_3)$, Thus we just need to consider the two way ANOVA with interaction for x_1, x_3 and treat them as factors each with 3 levels. Using the following R code we get the F value and p value.

```

lm.1<-lm(y~x1+x2+I(x1^2)+I(x2^2)+I(x1*x2),data=species)
lm.2<-lm(y~factor(x1)*factor(x3),data=species) anova(lm.1,lm.2)

```

The result is

	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
1	8	0.0045154				
2	5	0.0019715	3	0.0025439	2.1506	0.2124

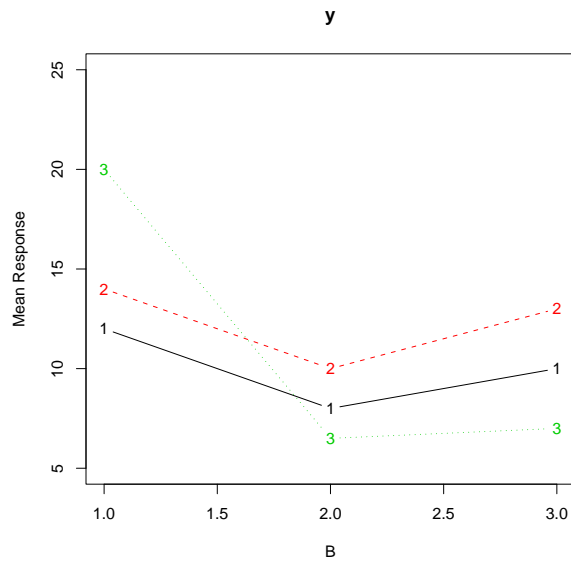
There is not enough evidence of "lack-of-fit", p-value=0.2124.

3. (a) Omitted

(b) The result is

	1	2	3
1	12	14	20.0
2	8	10	6.5
3	10	13	7.0

(c) The plot is as following:



(d) The result is as follows.

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
A	2	116.700	58.350	70.02	0.003037 **
B	2	4.050	2.025	2.43	0.235802
A:B	4	59.000	14.750	17.70	0.019990 *
Residuals	3	2.500	0.833		

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
B	2	26.250	13.125	15.75	0.025642 *
A	2	94.500	47.250	56.70	0.004138 **
B:A	4	59.000	14.750	17.70	0.019990 *
Residuals	3	2.500	0.833		

(e) We first write down the X matrix for sum restriction and SAS baseline restriction. Then use the following code to calculate the Type I, II and III sum-of-squares (for sum restriction). For the SAS baseline restriction, the code is similar, we do not display here. And we get the same type I and II sum of squares as sum restriction.

```

library(MASS) Px<-function(X) {X<-as.matrix(X)
  Px<-X%*%ginv(t(X)%*%X)%*%t(X)
  return(Px)}
X<-read.table("./hw_4_3_e_sum.R",header=F)
X<-as.matrix(X) y<-as.matrix(c(12,13,14,15,20,8,10,6,7,10,13,7))
#### Type I
P1<-Px(c(rep(1,12)))
Palpha<-Px(X[,1:3])
Pabeta<-Px(X[,1:5])
PX<-Px(X)
SSIalpha<-t(y)%*(Palpha-P1)%*y
SSIBeta<-t(y)%*(Pabeta-Palpha)%*y
SSIinter<-t(y)%*(PX-Pabeta)%*y
#### Type II
Pubeta<-Px(X[,c(1,4:5)])
Palpha<-Px(X[,c(1:5)])
Pualpha<-Px(X[,c(1:3)])
Pbeta<-Px(X[,1:5])
PX<-Px(X)
SSIIalpha<-t(y)%*(Palpha-Pubeta)%*y
SSIIbeta<-t(y)%*(Pbeta-Pualpha)%*y
SSIIinter<-t(y)%*(PX-Pabeta)%*y
#### Type III
Palpha<-Px(X[, -c(2:3)])
Pbeta<-Px(X[, -c(4:5)])
Palphabeta<-Px(X[, -c(6:9)])
PX<-Px(X)
SSIIIalpha<-t(y)%*(PX-Palpha)%*y
SSIIIbeta<-t(y)%*(PX-Pbeta)%*y
SSIIIinter<-t(y)%*(PX-Palphabeta)%*y

```

Then we get the Type I, II and III in following table:

	Type I SS	Type II SS	Type III SS
A	116.700	94.50	103.148
B	4.050	4.05	9.237
A:B	59.000	59.00	59.000

- (f) From part (e), we delete the missing data in y and corresponding X .

```

SSinter<-t(ymiss)%*(PX-Palphabeta)%*ymiss
P1<-diag(rep(1,11))
SSerror<-t(ymiss)%*(P1-PX)%*ymiss
F<-SSinter/SSerror
pf(F,3,3,lower.tail=F)

```

We get the F value as 0.7828571 with p-value 0.5773449, which means there is no evidence to shown the existence of interaction in the no missing data cell.

- (g) In the sum restriction, the OLS estimate of $\mu + \alpha_1 + \beta_3$ is the OLS estimate of $\mu + \alpha_1 - \beta_1 - \beta_2$.

```

Xmiss<-read.table("../hw_4_3_f_sum.txt",header=F)
Xmiss<-as.matrix(Xmiss[,1:5])
ymiss<-as.matrix(c(12,13,14,15,8,10,6,7,10,13,7)) C<-c(1,1,0,-1,-1)
Cbeta<-t(C)%*ginv(t(Xmiss)%*Xmiss)%*t(Xmiss)%*ymiss

```

Then the estimated mean response for the missing cell is 9.628571.