STAT3008 Exercise 8 Solutions

(2011-2012 2nd Semester)

Q1. (5.1)

5.1.1

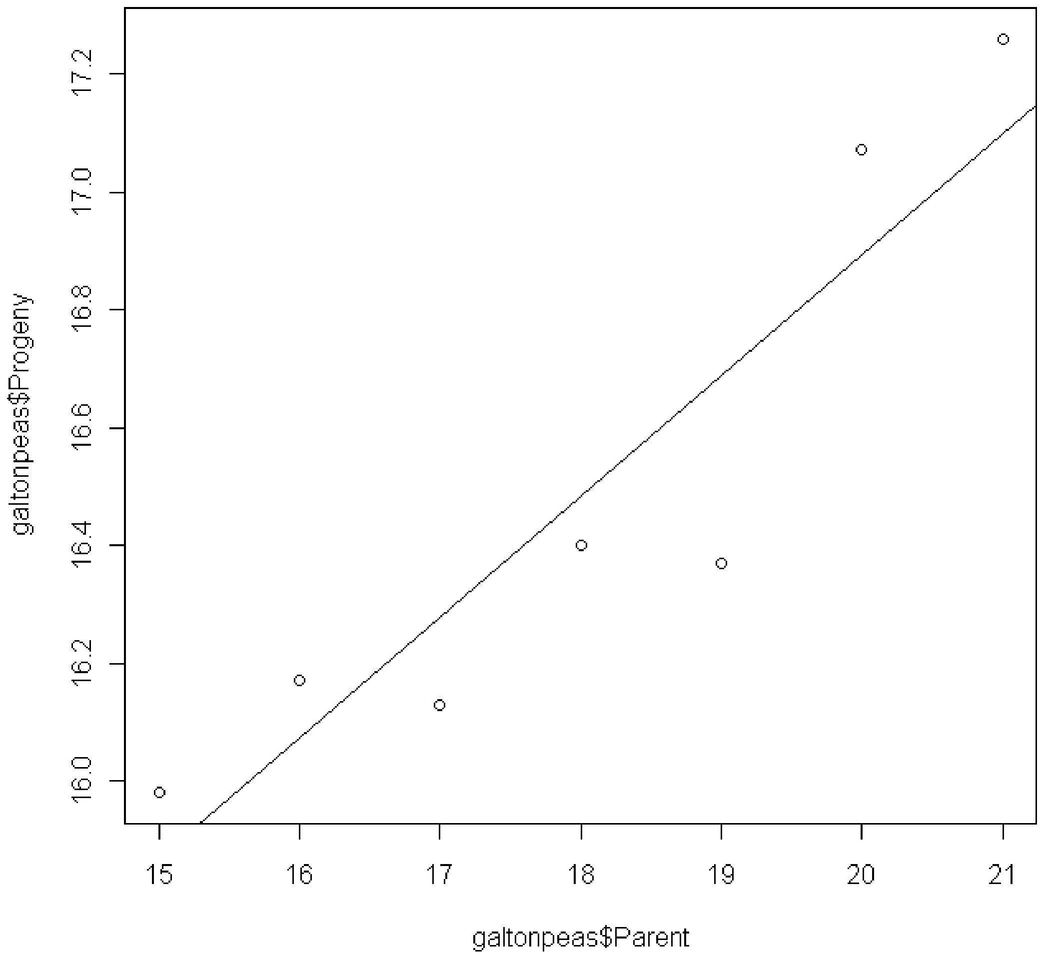
R Codes:

```
library(alr3)
plot(galtonpeas$Parent,galtonpeas$Progeny)
```

5.1.2

R Codes:

```
m1=lm(Progeny ~ Parent, weights = 1/SD^2,data=galtonpeas)
abline(coef(m1)[1],coef(m1)[2])
```



5.1.3

R Codes:

summary(m1)

$\widehat{\beta_1} = 0.2048, se(\widehat{\beta_1}) = 0.0382$

$obs.T.S. = \left| \frac{\widehat{\beta_1} - 1}{se(\widehat{\beta_1})} \right| = 20.723$

$t_{(0.05,5)} = 2.015 < obs.T.S.$

The null hypothesis is rejected.

(i)

$$\frac{(n-2)\widehat{\sigma}^2}{\sigma^2} \sim \chi^2_{n-2}$$

$obs.T.S. = \frac{5(0.11)^2}{0.3} = 0.2017$

$\chi^2_{(0.05,5)} = 11.0705 < obs.T.S.$

Thus, there is lack of fit.

(ii)

Since sample size is unknown, the lack of fit test cannot be performed.

Q2. (6.1)

6.1.1

R Codes:

library(alr3)

m1 <- lm(Y ~ X1+X2+I(X1^2)+I(X2^2)+X1:X2, data=cakes)

summary(m1)

Call:

lm(formula = Y ~ X1 + X2 + I(X1^2) + I(X2^2) + X1:X2, data = cakes)

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	-2.204e+03	2.416e+02	-9.125	1.67e-05	***
X1	2.592e+01	4.659e+00	5.563	0.000533	***
X2	9.918e+00	1.167e+00	8.502	2.81e05	***
I(X1^2)	-1.569e-01	3.945e-02	-3.977	0.004079	**
I(X2^2)	-1.195e-02	1.578e-03	-7.574	6.46e-05	***
X1:X2	-4.163e-02	1.072e-02	-3.883	0.004654	**

Therefore, the significance levels are all less than 0.05.

6.1.2

R Codes:

```
x1.max="(b2*b5-2*b1*b4)/(4*b3*b4-b5^2)"
x2.max="(b1*b5-2*b2*b3)/(4*b3*b4-b5^2)"
delta.method(m1,x1.max)
delta.method(m1,x2.max)
```

Therefore, $\widetilde{X_1} = 35.828, \widetilde{X_2} = 352.59$.

OR

$$E(Y \mid X) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1^{\;2} + \beta_4 X_2^{\;2} + \beta_5 X_1 X_2$$

$$\frac{dE(Y \mid X)}{dX_1} = \beta_1 + 2\beta_3 X_1 + \beta_5 X_2 = 0,$$

$$\frac{dE(Y \mid X)}{dX_2} = \beta_2 + 2\beta_4 X_2 + \beta_5 X_1 = 0$$

$$\widetilde{X_1} = \frac{\beta_2\beta_5 - 2\beta_1\beta_4}{4\beta_3\beta_4 - \beta_5^2}, \widetilde{X_2} = \frac{\beta_1\beta_5 - 2\beta_2\beta_2}{4\beta_3\beta_4 - \beta_5^2}$$

$$\widetilde{X_1} = 35.828, \widetilde{X_2} = 352.59$$

Q3. (6.4)

R Codes:

```
library(alr3)
twins$C=factor(twins$C)
m1=lm(IQf~IQb,data=twins)
m2=update(m1,~.+C)
m3=update(m1,~C:IQb)
m4=update(m1,~C*IQb)
anova(m1,m2,m4)
anova(m1,m3,m4)
```

Since all the p-values are larger than 0.05, the simplest model without any classification is preferred.

Q4. (6.14)

6.14.1

For a male, $\hat{\beta}_0 + \hat{\beta}_1 \times 0 = \hat{\eta}_0 + \hat{\eta}_1 \times 2 => \hat{\beta}_0 = \hat{\eta}_0 + 2\hat{\eta}_1$

For a female, $\hat{\beta}_0 + \hat{\beta}_1 \times 1 = \hat{\eta}_0 + \hat{\eta}_1 \times 1 => \hat{\beta}_0 + \hat{\beta}_1 = \hat{\eta}_0 + \hat{\eta}_1$

$\hat{\eta}_1 = -\hat{\beta}_1 = 571, \hat{\eta}_0 = \hat{\beta}_0 + 2\hat{\beta}_1 = 17681$

6.14.2

For a male, $\hat{\beta}_0 + \hat{\beta}_1 \times 0 = \hat{\eta}_0 + \hat{\eta}_1 \times (-1) => \hat{\beta}_0 = \hat{\eta}_0 - \hat{\eta}_1$

For a female, $\hat{\beta}_0 + \hat{\beta}_1 \times 1 = \hat{\eta}_0 + \hat{\eta}_1 \times 1 => \hat{\beta}_0 + \hat{\beta}_1 = \hat{\eta}_0 + \hat{\eta}_1$

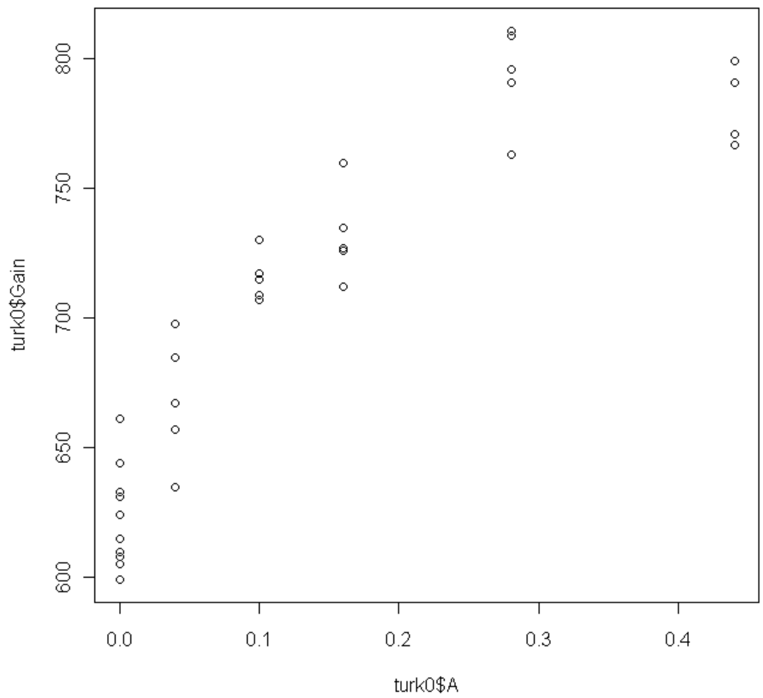
$\hat{\eta}_1 = \frac{\hat{\beta}_1}{2} = -285.5, \hat{\eta}_0 = \hat{\beta}_0 + \frac{\hat{\beta}_1}{2} = 17937.5$

Q5. (6.15)

6.15.1

R Codes:

```
library(alr3)
plot(turk0$A,turk0$Gain)
```



When A increases, Gain increases. However, Gain eventually drops when A exceeds 0.4. Thus, a simple linear regression model is not plausible.

6.15.2

R Codes:

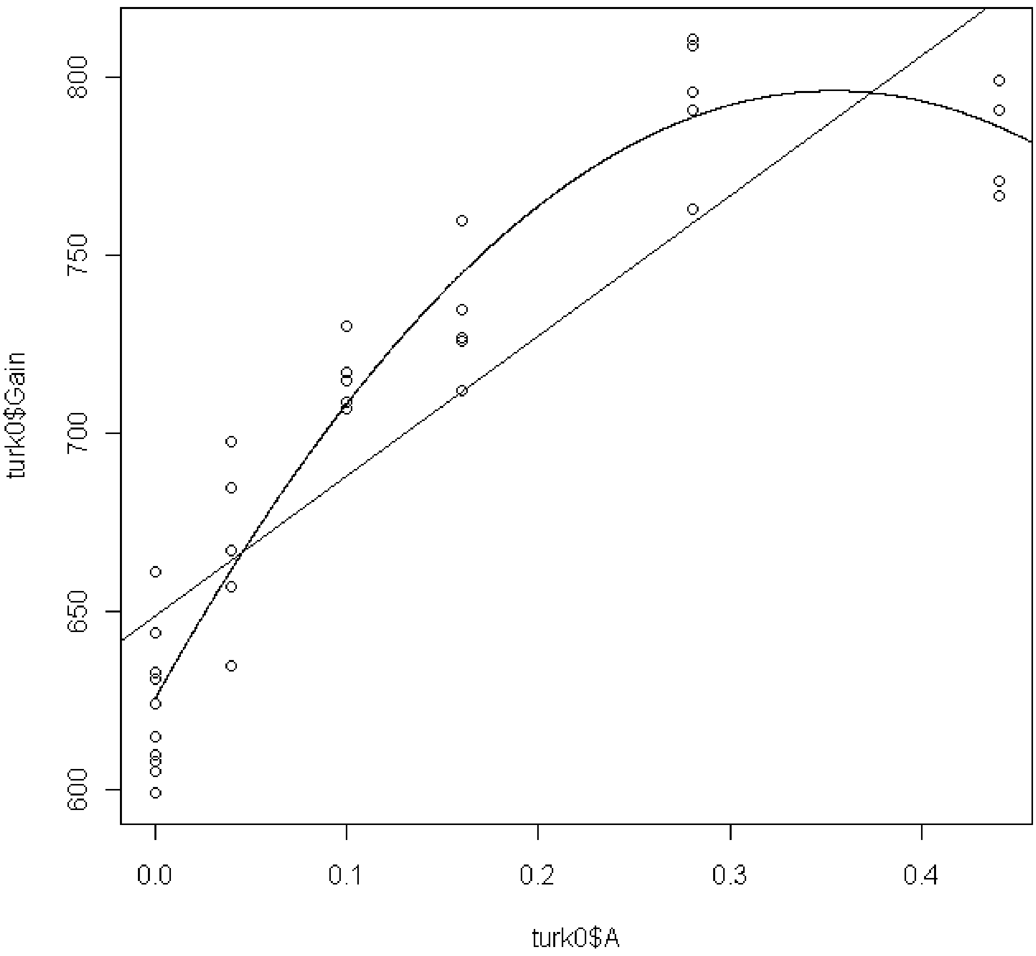
```
m11=lm(Gain~A,data=turk0)
m12=lm(Gain~factor(A),data=turk0)
m21=lm(Gain~A+I(A^2),data=turk0)
m22= lm(Gain~factor(A)+I(A^2),data=turk0)
anova(m11,m12)
anova(m21,m22)

Analysis of Variance Table

Model 1: Gain ~ A
Model 2: Gain ~ factor(A)
  Res.Df   RSS Df Sum of Sq      F    Pr(>F)
1     33 35176
2     29  9824  4     25353 18.711 1.062e-07 ***

Analysis of Variance Table

Model 1: Gain ~ A + I(A^2)
Model 2: Gain ~ factor(A) + I(A^2)
  Res.Df     RSS Df Sum of Sq     F Pr(>F)
1     32 11339.9
2     29  9823.6  3    1516.2 1.492 0.2374
```

From the p –values of the above tables, there is lack of fit for model 1. But there is no lack of fit for model 2.

6.15.3

R Codes:

```
x=sort(runif(10000,0,0.6))
y=coef(m21)[1]+x*coef(m21)[2]+x^2*coef(m21)[3]
plot(turk0$A,turk0$Gain)
abline(m11)
lines(x,y,type="l")
```



800
750
700
650
600
turk0$Gain
0.0
0.1
0.2
0.3
0.4
turk0$A