# MATH 668 Homework 5 Solutions

1. (a) First, we set up the design matrix and response vector.

```
setwd("C:/Users/ryan/Desktop/S18/668/data")
GaltonFamilies=read.table("HeightData.txt",header=TRUE)
y=GaltonFamilies$childHeight
g=as.numeric(GaltonFamilies$gender=="female")
m=GaltonFamilies$mother
f=GaltonFamilies$father
X=cbind(g,g*m,g*f,1-g,(1-g)*m,(1-g)*f);dimnames(X)=NULL
```

The maximum likeihood estimator $\hat{\boldsymbol{\beta}} = (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\boldsymbol{y} = \begin{pmatrix} 18.8335828 \\ 0.3034821 \\ 0.3725423 \\ 19.3128130 \\ 0.3287734 \\ 0.4175562 \end{pmatrix}$ can be obtained with the following command.

```
beta.hat=solve(t(X)%*%X)%*%t(X)%*%y
beta.hat
```

```
##            [,1]
## [1,] 18.8335828
## [2,]  0.3034821
## [3,]  0.3725423
## [4,] 19.3128130
## [5,]  0.3287734
## [6,]  0.4175562
```

Then the constrained MLE $\hat{\boldsymbol{\beta}}_c = \begin{pmatrix} 16.5212399 \\ 0.3176101 \\ 0.3928433 \\ 21.7362293 \\ 0.3176101 \\ 0.3928433 \end{pmatrix}$ can be obtained as follows (two equivalent methods).

```
C=rbind(c(0,1,0,0,-1,0),c(0,0,1,0,0,-1))
beta.hat.c=beta.hat-solve(t(X)%*%X)%*%t(C)%*%solve(C%*%solve(t(X)%*%X)%*%t(C))%*%C%*%beta.hat
beta.hat.c
```

```
##            [,1]
## [1,] 16.5212399
## [2,]  0.3176101
## [3,]  0.3928433
## [4,] 21.7362293
## [5,]  0.3176101
## [6,]  0.3928433
```

```
X.c=cbind(g,m,f,1-g);dimnames(X.c)=NULL
solve(t(X.c)%*%X.c)%*%t(X.c)%*%y
```

```
##            [,1]
## [1,] 16.5212399
## [2,]  0.3176101
## [3,]  0.3928433
```

```
## [4,] 21.7362293
```

(b) Now, we can compute the $F = \dfrac{\|\mathbf{X}(\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_c)\|^2/2}{\|\boldsymbol{y} - \mathbf{X}\boldsymbol{\beta}\|^2/(n-6)} = 0.4073867$ with the following code.

```
n=length(y)
(n-6)/2*sum((X%*%(beta.hat-beta.hat.c))^2)/sum((y-X%*%beta.hat)^2)
```

```
## [1] 0.4073867
```

(c) We first compute $\|\boldsymbol{y} - \mathbf{X}\boldsymbol{\beta}\|^2 = 4354.052$, $(\mathbf{C}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{C}^\top)^{-1} = \begin{pmatrix} 1218.74004 & 80.29407 \\ 80.29407 & 1411.68049 \end{pmatrix} = \begin{pmatrix} g_{11} & g_{12} \\ g_{12} & g_{22} \end{pmatrix}$, $\mathbf{C}\hat{\boldsymbol{\beta}} = \begin{pmatrix} \hat{\beta}_2 - \hat{\beta}_5 \\ \hat{\beta}_3 - \hat{\beta}_6 \end{pmatrix} = \begin{pmatrix} -0.02529128 \\ -0.04501389 \end{pmatrix} = \begin{pmatrix} k_1 \\ k_2 \end{pmatrix}$, $\dfrac{n-6}{2} = 464$, and $F_{.05,2,n-6} = 3.005424$ using the following code.

```
sum((y-X%*%beta.hat)^2)
```

```
## [1] 4354.052
```

```
g=solve(C%*%solve(t(X)%*%X)%*%t(C));g
```

```
##            [,1]       [,2]
## [1,] 1218.74004   80.29407
## [2,]   80.29407 1411.68049
```

```
k=C%*%beta.hat;k
```

```
##             [,1]
## [1,] -0.02529128
## [2,] -0.04501389
```

```
(n-6)/2
```

```
## [1] 464
```

```
qf(1-.05,2,n-6)
```

```
## [1] 3.005424
```

Then we have

$$P\left(\frac{(n-6)}{2}\frac{\begin{pmatrix}(\hat{\beta}_2 - \hat{\beta}_5) - (\beta_2 - \beta_5) \\ (\hat{\beta}_3 - \hat{\beta}_6) - (\beta_3 - \beta_6)\end{pmatrix}^\top \begin{pmatrix} 1218.74004 & 80.29407 \\ 80.29407 & 1411.68049 \end{pmatrix} \begin{pmatrix}(\hat{\beta}_2 - \hat{\beta}_5) - (\beta_2 - \beta_5) \\ (\hat{\beta}_3 - \hat{\beta}_6) - (\beta_3 - \beta_6)\end{pmatrix}}{\|\boldsymbol{y} - \mathbf{X}\hat{\boldsymbol{\beta}}\|^2} \leq 3.005424\right) = .95$$

so that a 95% confidence ellipse for $\begin{pmatrix}\beta_2 - \beta_5 \\ \beta_3 - \beta_6\end{pmatrix}$ is

$$464\frac{\begin{pmatrix}-0.02529128 - (\beta_2 - \beta_5) \\ -0.04501389 - (\beta_3 - \beta_6)\end{pmatrix}^\top \begin{pmatrix} 1218.74004 & 80.29407 \\ 80.29407 & 1411.68049 \end{pmatrix} \begin{pmatrix}-0.02529128 - (\beta_2 - \beta_5) \\ -0.04501389 - (\beta_3 - \beta_6)\end{pmatrix}}{4354.052} \leq 3.005424 \Rightarrow$$

$$\begin{pmatrix}k_1 - (\beta_2 - \beta_5) \\ k_2 - (\beta_3 - \beta_6)\end{pmatrix}^\top \begin{pmatrix} g_{11} & g_{12} \\ g_{12} & g_{22} \end{pmatrix} \begin{pmatrix}k_1 - (\beta_2 - \beta_5) \\ k_2 - (\beta_3 - \beta_6)\end{pmatrix} \leq f$$

where $f = 28.20209$.

```
f=2*sum((y-X%*%beta.hat)^2)*qf(1-.05,2,n-6)/(n-6);f
```

```
## [1] 28.20209
```

Then we see that the left side of the inequality can be expressed as
$g_{11}(\beta_2 - \beta_5 - k_1)^2 + 2g_{12}(\beta_2 - \beta_5 - k_1)(\beta_3 - \beta_6 - k_2) + g_{22}(\beta_3 - \beta_6 - k_2)^2$

$= g_{11}(\beta_2-\beta_5)^2 + g_{22}(\beta_3-\beta_6)^2 + 2g_{12}(\beta_2-\beta_5)(\beta_3-\beta_6) - 2(k_1g_{11}+k_2g_{12})(\beta_2-\beta_5) - 2(k_1g_{12}+k_2g_{22})(\beta_3-\beta_6) + (k_1^2g_{11} + 2k_1k_2g_{12} + k_2^2g_{22})$
so that
$A(\beta_2-\beta_5)^2 + B(\beta_3-\beta_6)^2 + C(\beta_2-\beta_5)(\beta_3-\beta_6) + D(\beta_2-\beta_5) + E(\beta_3-\beta_6) \le 1$ is the 95% confidence ellipse
where $A = \dfrac{g_{11}}{f - k_1^2g_{11} - 2k_1k_2g_{12} - k_2^2g_{22}} = 49.9908$,
$B = \dfrac{g_{22}}{f - k_1^2g_{11} - 2k_1k_2g_{12} - k_2^2g_{22}} = 57.90491$,
$C = \dfrac{2g_{12}}{f - k_1^2g_{11} - 2k_1k_2g_{12} - k_2^2g_{22}} = 6.587073$,
$D = \dfrac{-2(k_1g_{11}+k_2g_{12})}{f - k_1^2g_{11} - 2k_1k_2g_{12} - k_2^2g_{22}} = 2.825173$, and
$E = \dfrac{-2(k_1g_{12}+k_2g_{22})}{f - k_1^2g_{11} - 2k_1k_2g_{12} - k_2^2g_{22}} = 5.379647$ as computed below.

```
f2=f-g[1,1]*k[1]^2-2*g[1,2]*k[1]*k[2]-g[2,2]*k[2]^2
A=g[1,1]/f2;A
```

```
## [1] 49.9908
```

```
B=g[2,2]/f2;B
```

```
## [1] 57.90491
```

```
C=2*g[1,2]/f2;C
```

```
## [1] 6.587073
```

```
D=-2*(k[1]*g[1,1]+k[2]*g[1,2])/f2;D
```

```
## [1] 2.825173
```

```
E=-2*(k[1]*g[1,2]+k[2]*g[2,2])/f2;E
```

```
## [1] 5.379647
```

2. $\hat{\boldsymbol\beta} = (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\boldsymbol{y}$

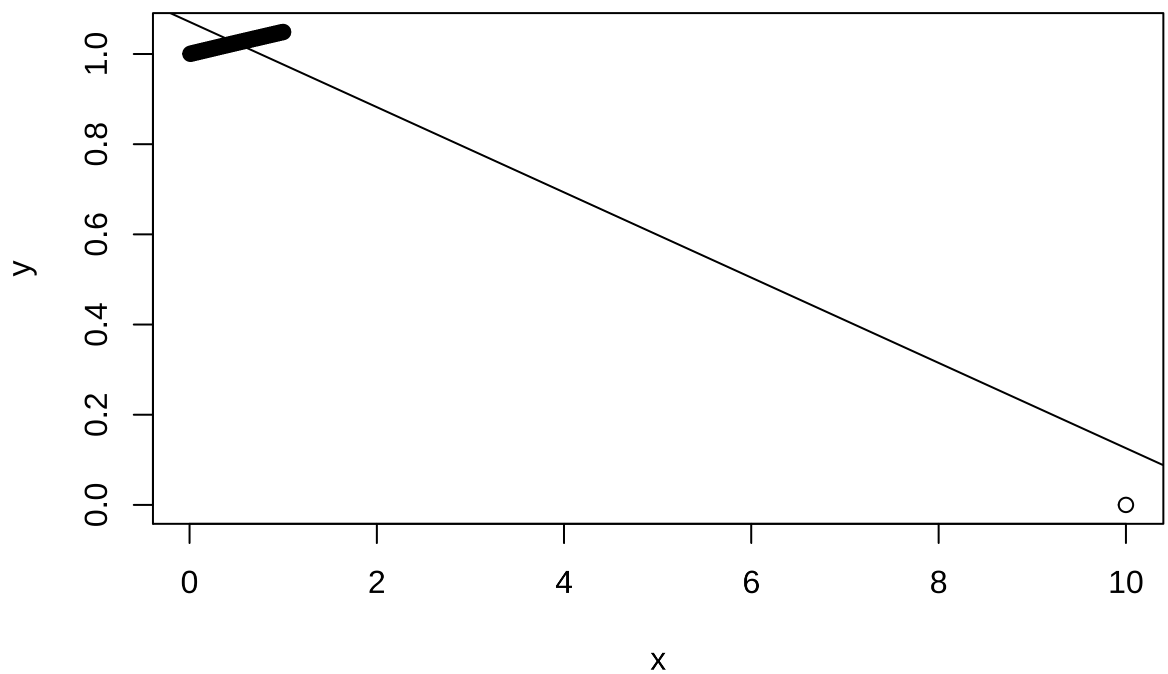
$$
\begin{aligned}
&= \left(\left(\mathbf{X}_{(n)}^\top, \mathcal{X}_n\right)\begin{pmatrix}\mathbf{X}_{(n)}\\ \mathcal{X}_n^\top\end{pmatrix}\right)^{-1}\left(\mathbf{X}_{(n)}^\top, \mathcal{X}_n\right)\begin{pmatrix}\boldsymbol{y}_{(n)}\\ y_n\end{pmatrix}\\
&= \left(\mathbf{X}_{(n)}^\top\mathbf{X}_{(n)} + \mathcal{X}_n\mathcal{X}_n^\top\right)^{-1}(\mathbf{X}_{(n)}^\top\boldsymbol{y}_{(n)} + \mathcal{X}_n y_n)\\
&= \left\{\left(\mathbf{X}_{(n)}^\top\mathbf{X}_{(n)}\right)^{-1} - \left(\mathbf{X}_{(n)}^\top\mathbf{X}_{(n)}\right)^{-1}\mathcal{X}_n\left(1+\mathcal{X}_n^\top\left(\mathbf{X}_{(n)}^\top\mathbf{X}_{(n)}\right)^{-1}\mathcal{X}_n\right)^{-1}\mathcal{X}_n^\top\left(\mathbf{X}_{(n)}^\top\mathbf{X}_{(n)}\right)^{-1}\right\}(\mathbf{X}_{(n)}^\top\boldsymbol{y}_{(n)} + \mathcal{X}_n y_n)\\
&= \left(\mathbf{M} - \frac{\mathbf{M}\mathcal{X}_n\mathcal{X}_n^\top\mathbf{M}}{1+\mathcal{X}_n^\top\mathbf{M}\mathcal{X}_n}\right)(\mathbf{X}_{(n)}^\top\boldsymbol{y}_{(n)} + \mathcal{X}_n y_n)\\
&= \mathbf{M}\mathbf{X}_{(n)}^\top\boldsymbol{y}_{(n)} + \mathbf{M}\mathcal{X}_n y_n - \frac{\mathbf{M}\mathcal{X}_n\mathcal{X}_n^\top\mathbf{M}}{1+\mathcal{X}_n^\top\mathbf{M}\mathcal{X}_n}\mathbf{X}_{(n)}^\top\boldsymbol{y}_{(n)} - \frac{\mathbf{M}\mathcal{X}_n\mathcal{X}_n^\top\mathbf{M}}{1+\mathcal{X}_n^\top\mathbf{M}\mathcal{X}_n}\mathcal{X}_n y_n\\
&= \mathbf{M}\mathbf{X}_{(n)}^\top\boldsymbol{y}_{(n)} + \mathbf{M}\mathcal{X}_n y_n - \frac{\mathbf{M}\mathcal{X}_n\mathcal{X}_n^\top\mathbf{M}}{1+\mathcal{X}_n^\top\mathbf{M}\mathcal{X}_n}\mathbf{X}_{(n)}^\top\boldsymbol{y}_{(n)} - \frac{\mathbf{M}\mathcal{X}_n\mathcal{X}_n^\top\mathbf{M}}{1+\mathcal{X}_n^\top\mathbf{M}\mathcal{X}_n}\mathcal{X}_n y_n\\
&= \hat{\boldsymbol\beta}_{(n)} + \frac{\mathbf{M}\mathcal{X}_n}{1+\mathcal{X}_n^\top\mathbf{M}\mathcal{X}_n}\left(y_n(1+\mathcal{X}_n^\top\mathbf{M}\mathcal{X}_n) - \mathcal{X}_n^\top\mathbf{M}\mathbf{X}_{(n)}^\top\boldsymbol{y}_{(n)} - \mathcal{X}_n^\top\mathbf{M}\mathcal{X}_n y_n\right)\\
&= \hat{\boldsymbol\beta}_{(n)} + \frac{\mathbf{M}\mathcal{X}_n}{1+\mathcal{X}_n^\top\mathbf{M}\mathcal{X}_n}\left(y_n - \mathcal{X}_n^\top\mathbf{M}\mathbf{X}_{(n)}^\top\boldsymbol{y}_{(n)}\right)\\
&= \hat{\boldsymbol\beta}_{(n)} + \frac{\mathbf{M}\mathcal{X}_n}{1+\mathcal{X}_n^\top\mathbf{M}\mathcal{X}_n}\left(y_n - \mathcal{X}_n^\top\hat{\boldsymbol\beta}_{(n)}\right)
\end{aligned}
$$

3. First, we create the data in R using the following commands.

```
set.seed(123)
x0=1:100/100
x=c(x0,10)
y0=sqrt(1+.1*x0+rnorm(100,sd=.0001))
y=c(y0,0)
```

(a) The following commands create a scatterplot with the regression line superimposed.

```
plot(x,y)
lm.model=lm(y~x)
abline(lm.model$coef)
```




(b) We see that $\hat{\varepsilon}_{101} = -0.1258052$, $r_{101} = -9.949851$, and $t_{101} = -4602.297$ as computed below.

```
epsilon.hat=residuals(lm.model); epsilon.hat[101]
```

```
##        101
## -0.1258052
```

```
r=rstandard(lm.model); r[101]
```

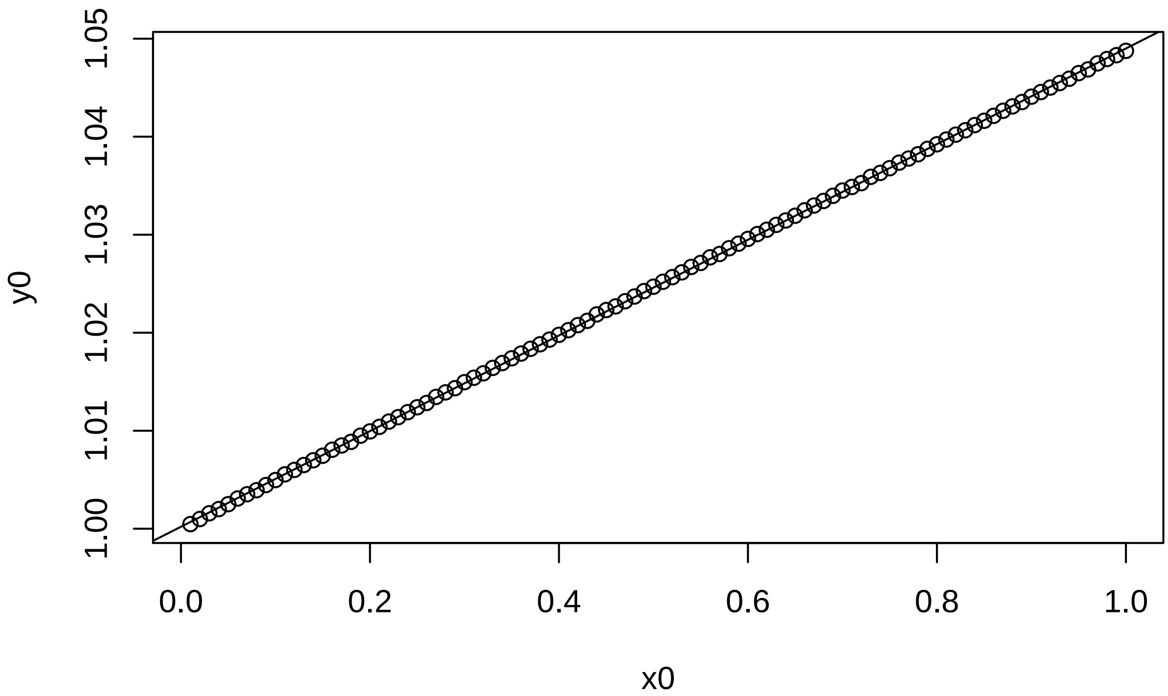
```
##       101
## -9.949851
```

```
t=rstudent(lm.model); t[101]
```

```
##       101
## -4602.297
```

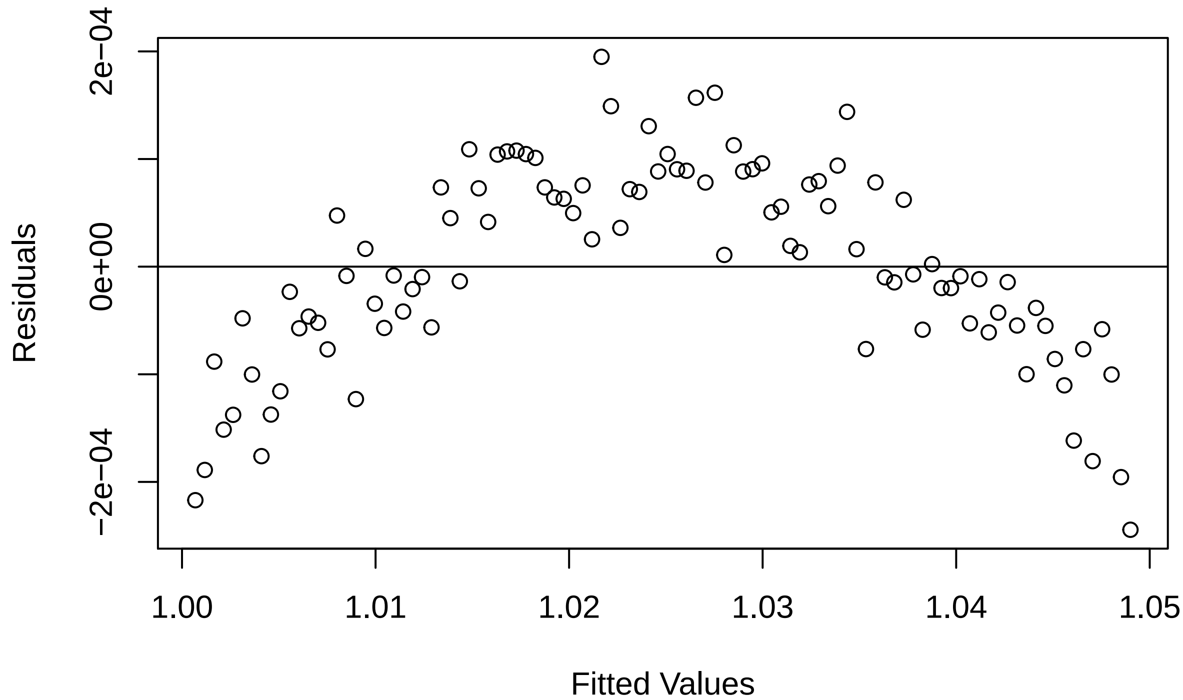
(c) The following commands create a scatterplot with the regression line superimposed for the data set with the last observation removed.

```
plot(x0,y0)
lm.model0=lm(y0~x0)
abline(lm.model0$coef)
```




(d) Here is a residual plot for the fitted values with the last observation removed.

```
plot(lm.model0$fit,lm.model0$resid,xlab="Fitted Values",ylab="Residuals")
abline(h=0)
```



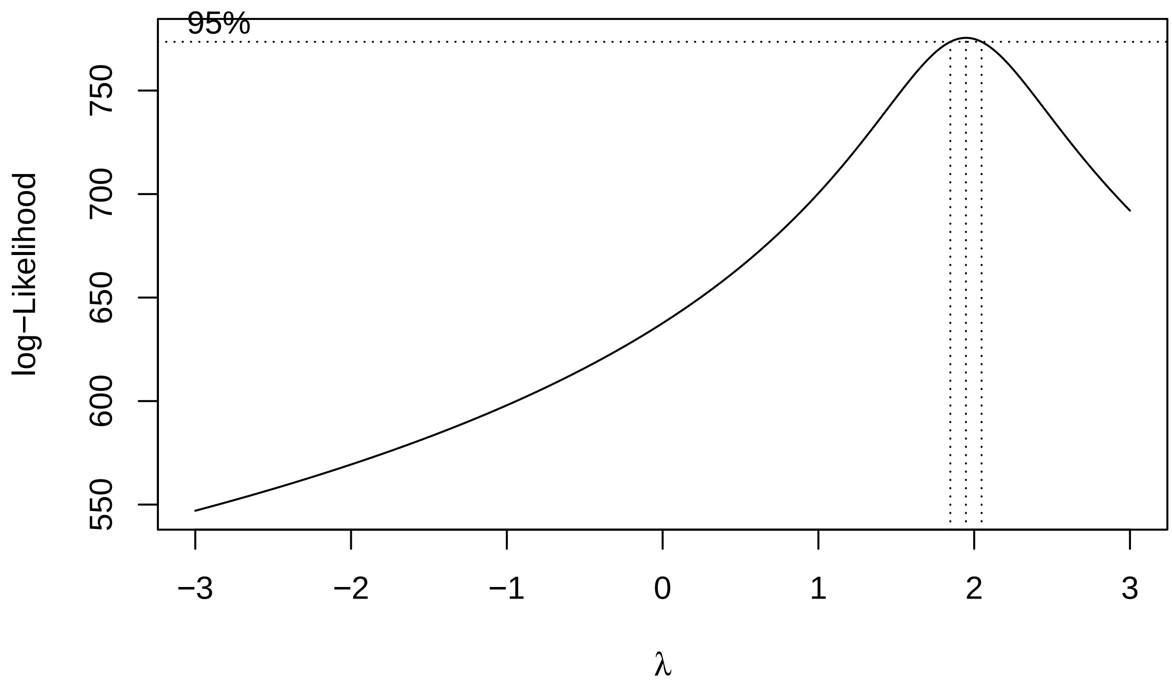

Clearly, there is a pattern so the linearity assumption is not random.

(e) We can obtain $\lambda = 1.947$ in the Box-Cox transformation as follows.

```
require(MASS)
```

```
## Loading required package: MASS
```

```
bc=boxcox(y0~x0,lambda=seq(-3,3,by=.001))
```

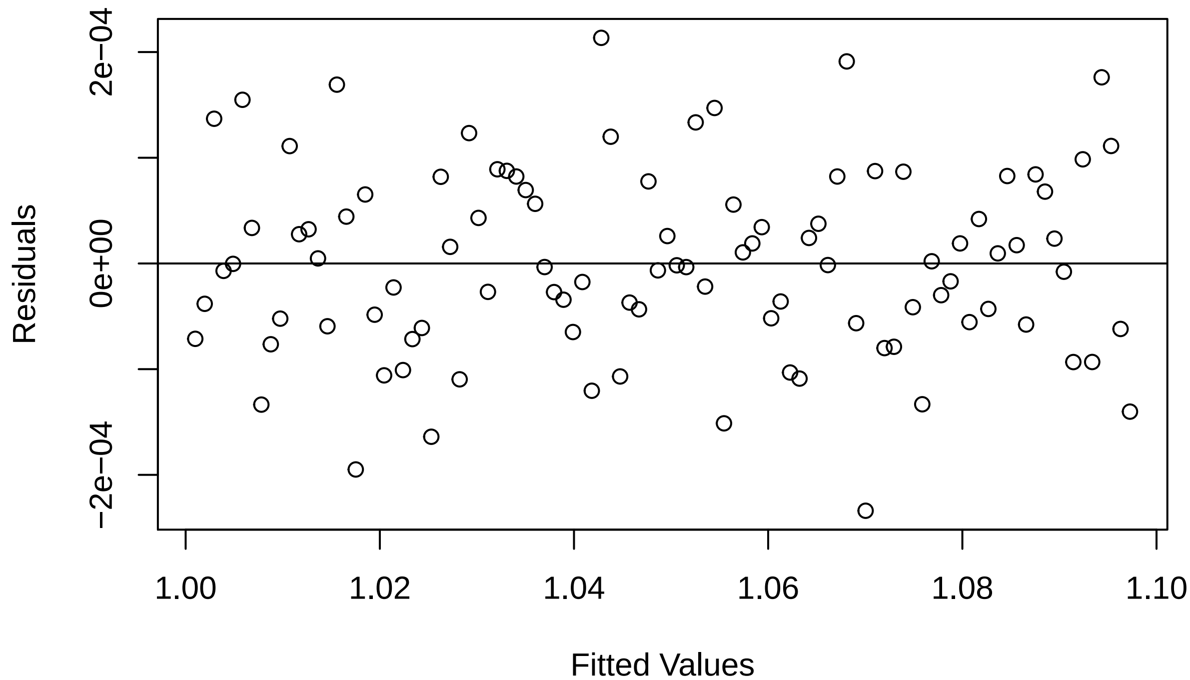



```
bc$x[which.max(bc$y)]
```

```
## [1] 1.947
```

Then it is interesting to look at the residual plot for the transformed data to see that the new model is more reasonable.

```
u=y0^1.947
lm.model.bc=lm(u~x0)
plot(lm.model.bc$fit,lm.model.bc$resid,xlab="Fitted Values",ylab="Residuals")
abline(h=0)
```


Residuals
2e−04
0e+00
−2e−04
1.00
1.02
1.04
1.06
1.08
1.10
Fitted Values