Stat525 HW6

Luke Geel

3/13/2021

1.42 a. likelihood <- function(bet){ lk <- (1/sqrt(32*pi))^6*exp((-1/32)*sum((y-bet*x)^2))}

b. b1=17: 9.451330e-30 b1=18 : 2.649043e-07 b1=19: 3.047285e-37 Largest likelihood value at b1 = 18

c. Maximum likelihood estimate is 17.93. This is consistent with part B as 17.93 is very close to 18 which was the largest likelihood function.

d. plot(beta_1, likelihood_values, type = "l") The point at which the likelihood function is maximized corresponds to the maximum likelihood estiamte from part c.

```
library(tidyverse)
```

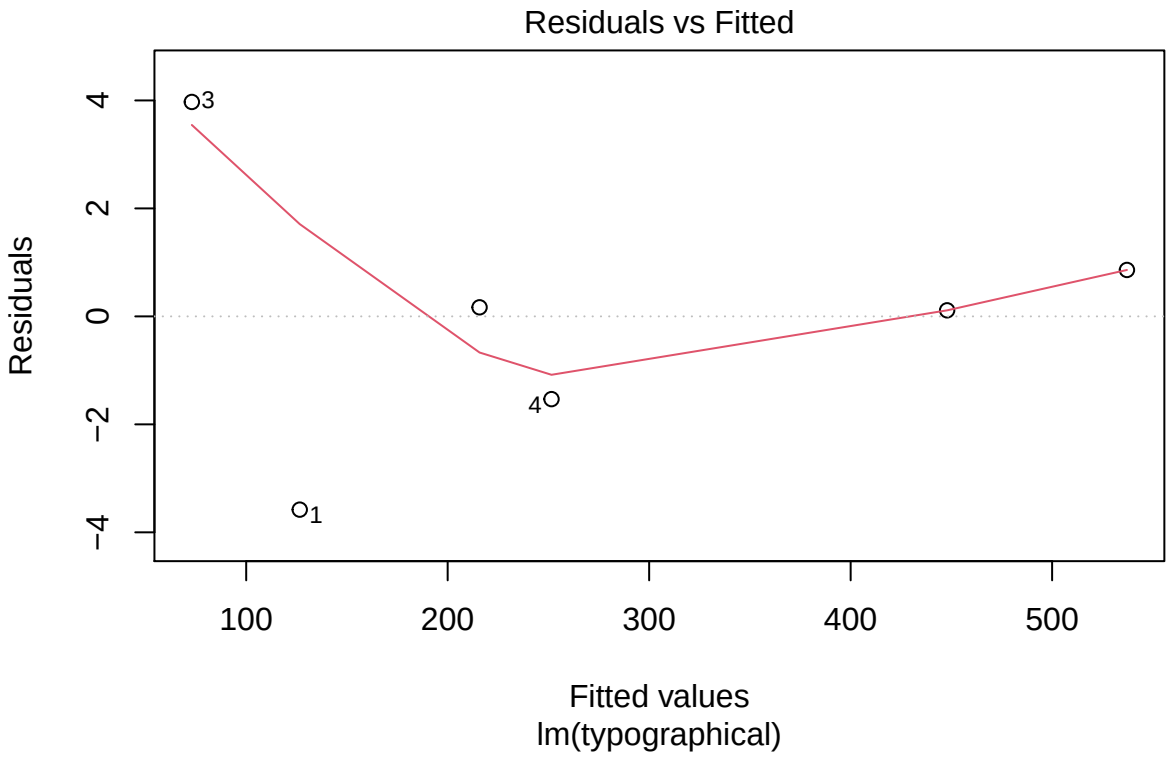
```
## -- Attaching packages ---------------------------------------------------------------------------
```

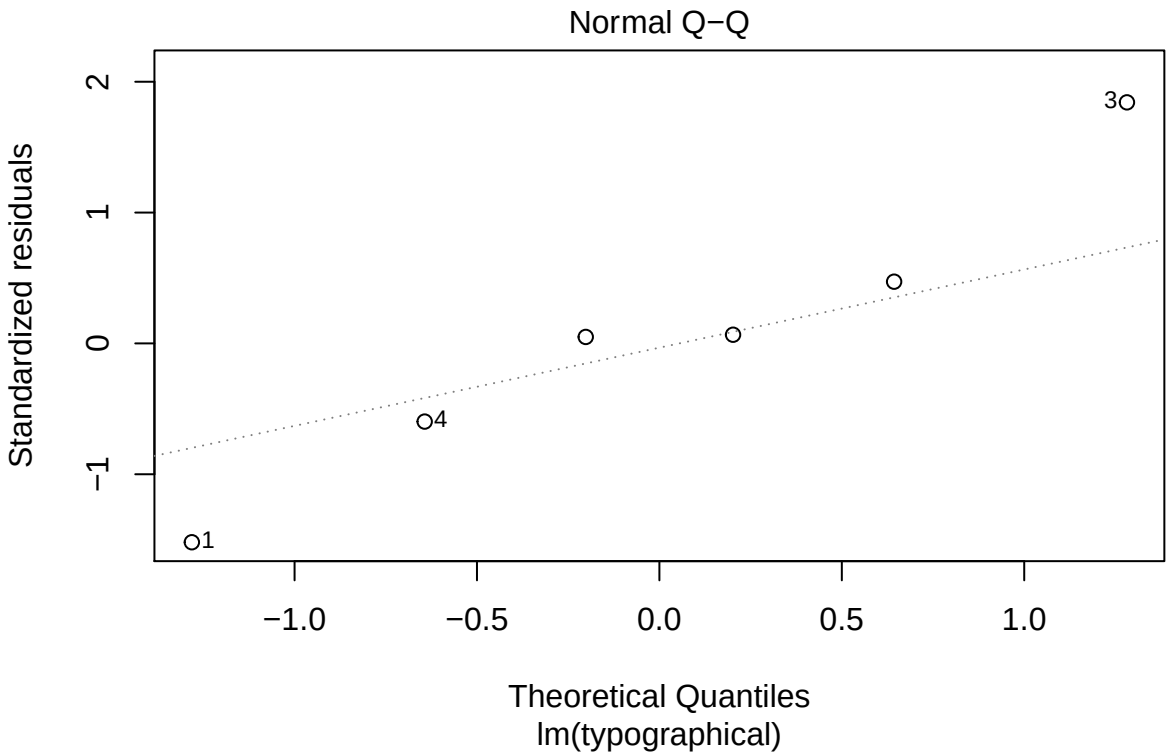
```
## v ggplot2 3.3.2     v purrr   0.3.4
## v tibble  3.0.3     v dplyr   1.0.2
## v tidyr   1.1.2     v stringr 1.4.0
## v readr   1.3.1     v forcats 0.5.0
```

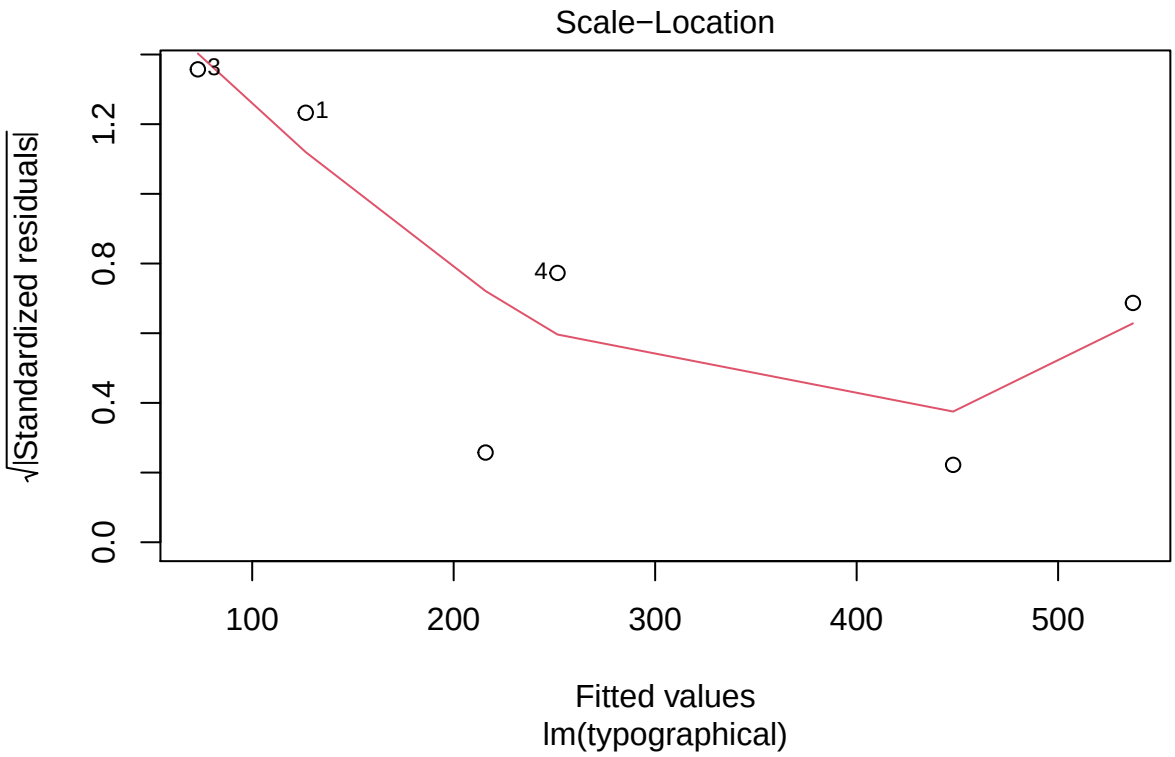
```
## -- Conflicts ------------------------------------------------------------------------------------
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

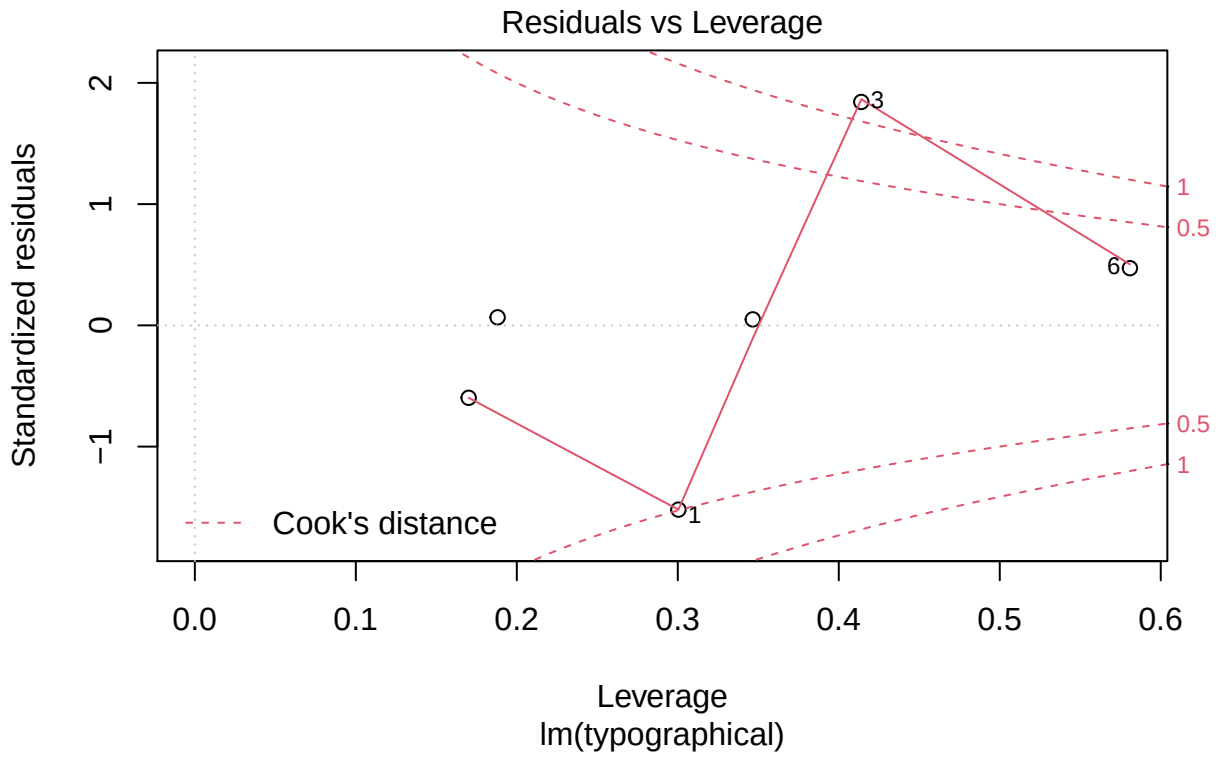
```
loadRData <- function(fileName){
  load(fileName)
  get(ls()[ls() != "fileName"])
}
typographical <- loadRData("/Users/lukegeel/Downloads/typographical_spring2021.RData")

view(typographical)
model <- lm(typographical)
plot(model)
```


Residuals vs Fitted
Residuals
4
2
0
−2
−4
3
1
4
100
200
300
400
500
Fitted values
lm(typographical)


Normal Q−Q
Standardized residuals
2
1
0
−1
3
4
1
−1.0
−0.5
0.0
0.5
1.0
Theoretical Quantiles
lm(typographical)


Scale−Location
√|Standardized residuals|
0.0
0.4
0.8
1.2
3
1
4
100
200
300
400
500
Fitted values
lm(typographical)



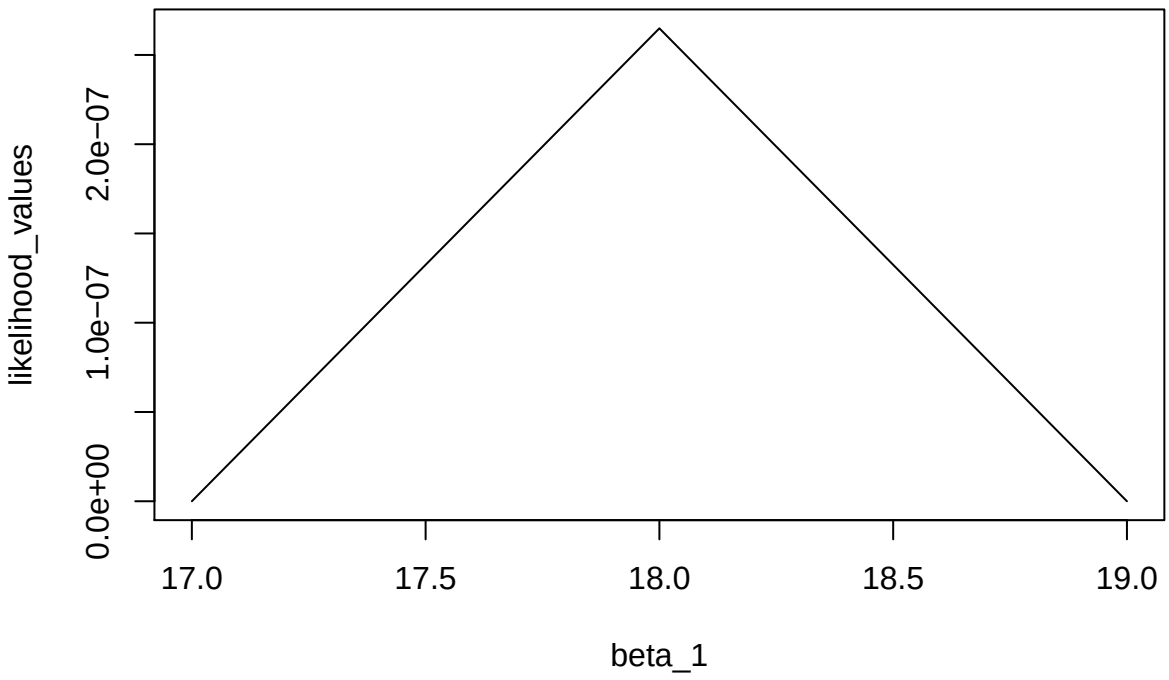

```
x <- cbind(7,12,4,14,25,30)
y <- cbind(128,213,75,250,446,540)
beta_1 = c(17,18,19)
likelihood <- function(bet){
lk <- (1/sqrt(32*pi))^6*exp((-1/32)*sum((y-bet*x)^2))}

likelihood_values <- numeric(0);
for(i in 1:length(beta_1)){
likelihood_values[i] <- likelihood(beta_1[i])}

print(likelihood_values)
```

```
## [1] 9.451330e-30 2.649043e-07 3.047285e-37
```

```
plot(beta_1, likelihood_values, type = "l")
```



2.62 The highest R^2 value is 0.89 and that represents personal income.

```
library(tidyverse)
loadRData <- function(fileName){
  load(fileName)
  get(ls()[ls() != "fileName"])
}
cdi <- loadRData("/Users/lukegeel/Downloads/cdi_spring2021.RData")
data2 <- cdi[c("X4", "X5","X6","X7","X8","X9","X10","X11","X12","X13")]
Y <- data2$X8
X1 <- data2$X4
X2 <- data2$X5
X3 <- data2$X6
X4 <- data2$X7

X6 <- data2$X9
X7 <- data2$X10
X8 <- data2$X11
X9 <- data2$X12
X10 <- data2$X13

linmod1 <- lm(Y~X1)
b01 <- linmod1$coef[1]
b11 <- linmod1$coef[2]

linmod2 <- lm(Y~X2)
```

```r
b02 <- linmod2$coef[1]
b12 <- linmod2$coef[2]

linmod3 <- lm(Y~X3)
b03 <- linmod3$coef[1]
b13 <- linmod3$coef[2]

linmod4 <- lm(Y~X4)
b04 <- linmod4$coef[1]
b14 <- linmod4$coef[2]

linmod5 <- lm(Y~X6)
b05 <- linmod5$coef[1]
b15 <- linmod5$coef[2]

linmod6 <- lm(Y~X7)
b06 <- linmod6$coef[1]
b16 <- linmod6$coef[2]

linmod7 <- lm(Y~X8)
b07 <- linmod7$coef[1]
b17 <- linmod7$coef[2]

linmod8 <- lm(Y~X9)
b08 <- linmod8$coef[1]
b18 <- linmod8$coef[2]

linmod9 <- lm(Y~X10)
b09 <- linmod9$coef[1]
b19 <- linmod9$coef[2]

summary(linmod1)$r.squared
```

```
## [1] 0.005889308
```

```r
summary(linmod2)$r.squared
```

```
## [1] 0.8793997
```

```r
summary(linmod3)$r.squared
```

```
## [1] 0.01518831
```

```r
summary(linmod4)$r.squared
```

```
## [1] 1.563569e-05
```

```r
summary(linmod5)$r.squared
```

```
## [1] 0.8826229
```

```r
summary(linmod6)$r.squared
```

```
## [1] 0.680936
```

```r
summary(linmod7)$r.squared
```

```
## [1] 8.625589e-05
```

```r
summary(linmod8)$r.squared
```

```
## [1] 0.06141623
```

```r
summary(linmod9)$r.squared
```

```
## [1] 0.002434574
```

2.66 a. e=rnorm(5,0,5)

b. for(i in 1:200){ x=c(4,8,12,16,20) e=rnorm(5,0,5) y=20+4*x+e LM=lm(y~x) C=LM$"coefficients" beta1[i]=C[2] }

c. mean = -0.8548 sd = 4.8478 These results are consistent with theoretical expectations. You would expect the mean to be 0 and standard deviation to be 5 but since the numbers are chosen at random the will be slightly off. -0.85 is very close to 0 and 4.85 is very close to 5 so these numbers aren't surprising to me.

```r
x=c(4,8,12,16,20)
xbar=mean(x)
e=rnorm(5,0,5)
y=20+4*x+e
LM <- lm(y~x)
C=LM$"coefficients"
SSRes <- sum((LM$residuals)^2)
MSRes <- SSRes/(3)
MSRes
```

```
## [1] 42.29563
```

```r
tva=qt(0.975,3)
C[1]+C[2]*10
```

```
## (Intercept)
##    65.02326
```

```r
Beta1=4
beta1=c()
for(i in 1:200){
  x=c(4,8,12,16,20)
  e=rnorm(5,0,5)
```

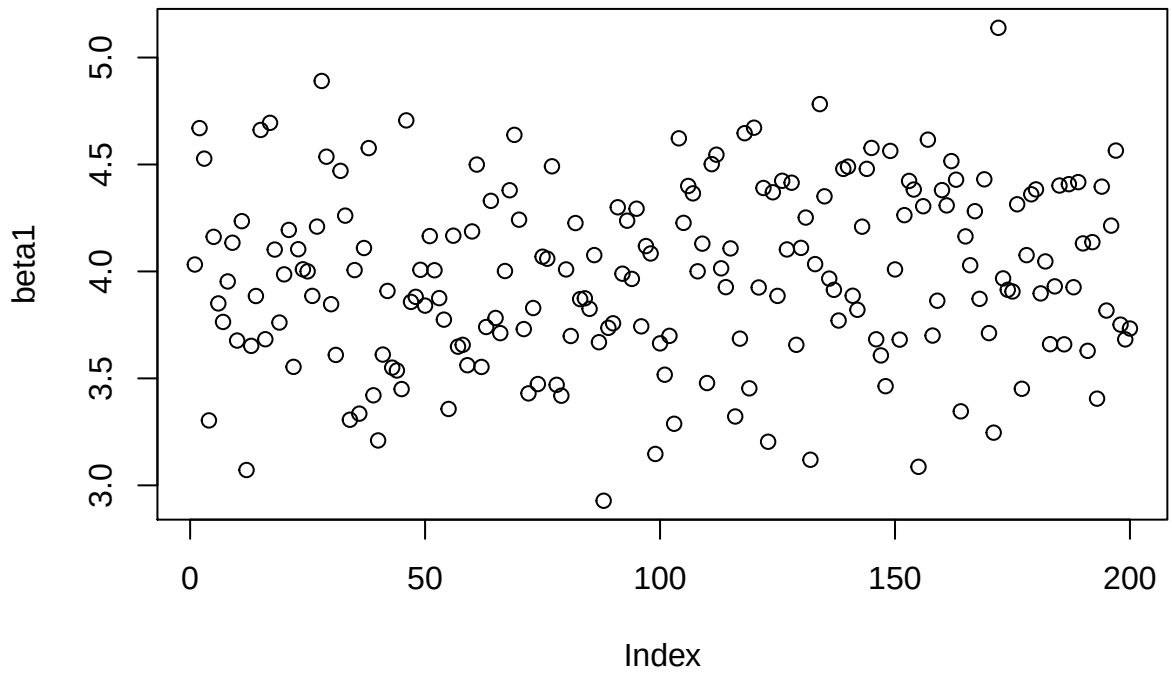
```
  y=20+4*x+e
  LM=lm(y~x)
  C=LM$"coefficients"
  beta1[i]=C[2]
}
mean(beta1)
```

```
## [1] 3.98527
```

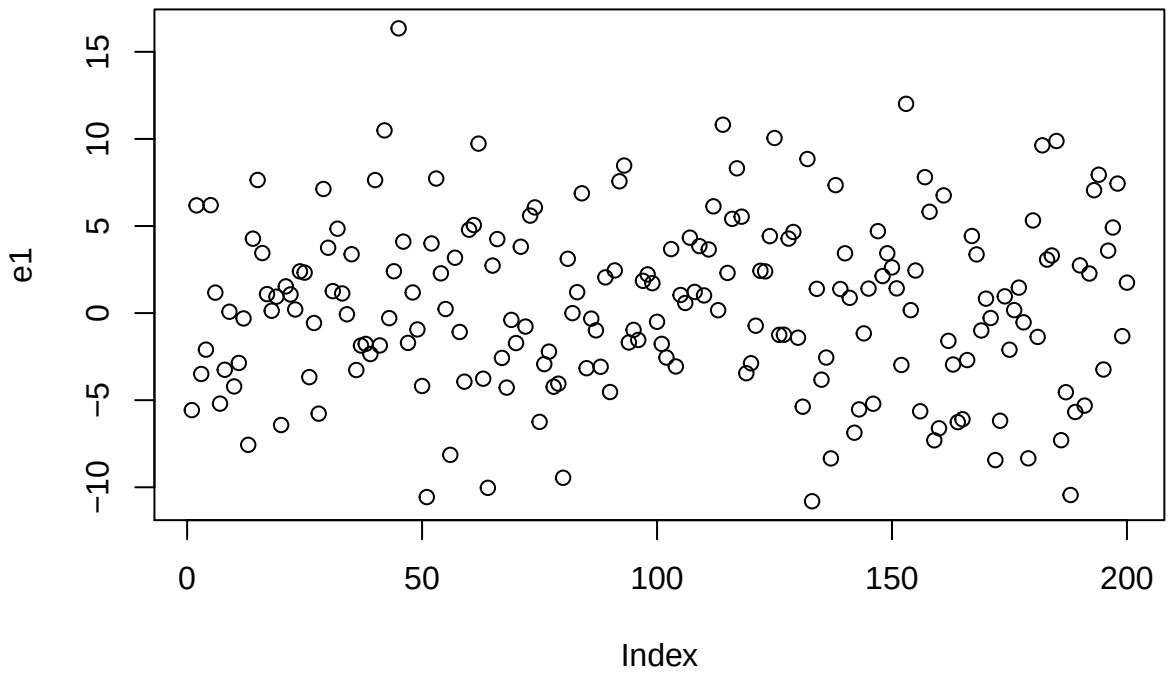
```
var(beta1)
```

```
## [1] 0.1658803
```

```
plot(beta1)
```




```
e1=rnorm(200,0,5)
plot(e1)
```




```
mean(e1)
```

```
## [1] 0.505179
```

```
sd(e1)
```

```
## [1] 4.851299
```

```
var(e1)
```

```
## [1] 23.5351
```