PS #4 - Question 3

```
set.seed(1234)
N = 1000 # sample size
```

(a) data generating:

```
x1 <- rnorm(N)
x2 <- rnorm(N, 0,5)
epsilon <- rnorm(N)

y = 1 + x1 + x2 + epsilon
```

OLS regression:

```
fit <- lm(y ~ x1 + x2)
fit$coefficients
```

```
## (Intercept)          x1          x2
##   1.0299573   1.0177020   0.9926205
```

(b)

residuals and regressors are orthogonal.

```
sum(fit$residuals)
```

```
## [1] -2.843906e-14
```

```
x1 %*% fit$residuals
```

```
##                [,1]
## [1,] -1.815215e-14
```

```
x2 %*% fit$residuals
```

```
##                [,1]
## [1,] -6.661338e-15
```

(c) true residuals are not orthogonal with residuals.

```
sum(epsilon)
```

```
## [1] 28.95114
```

```
x1 %*% epsilon
```

```
##          [,1]
## [1,] 14.77759
```

```
x2 %*% epsilon
```

```
##          [,1]
## [1,] -170.435
```

(d) auxiliary regression:

```
fit_aux = lm(x1 ~ x2)
fit_aux$coefficients
```

```
## (Intercept)          x2
## -0.02743232  0.01151237
```

```r
fit2 = lm(y ~ fit_aux$residuals + 0)
fit2$coefficients
```

```
## fit_aux$residuals
##          1.017702
```

T-stat

```r
sigma_hat = sum(sqrt(fit$residuals ** 2)) / (N-3)



# Auxiliary Regression of regressing X1 on constant and X2

x0 = matrix(rep(1, N), nrow = N)

X = matrix(cbind(x0, x1, x2), ncol=3)
beta = fit$coefficients

gamma_hat = solve(t(X[,c(1,3)]) %*% X[,c(1,3)]) %*% (t(X[,c(1,3)]) %*% y)
v1_hat = X[,2] - X[,c(1,3)] %*% gamma_hat


denom = sqrt(sigma_hat **2 / (t(v1_hat) %*% v1_hat))
numer = beta[2] - 0

numer / denom
```

```
##          [,1]
## [1,] 201.1393
```

```r
sqrt(diag(sigma_hat ** 2 / (t(X) %*% X)))
```

```
## [1] 0.025751958 0.025824429 0.005251165
```

Look at ith effect

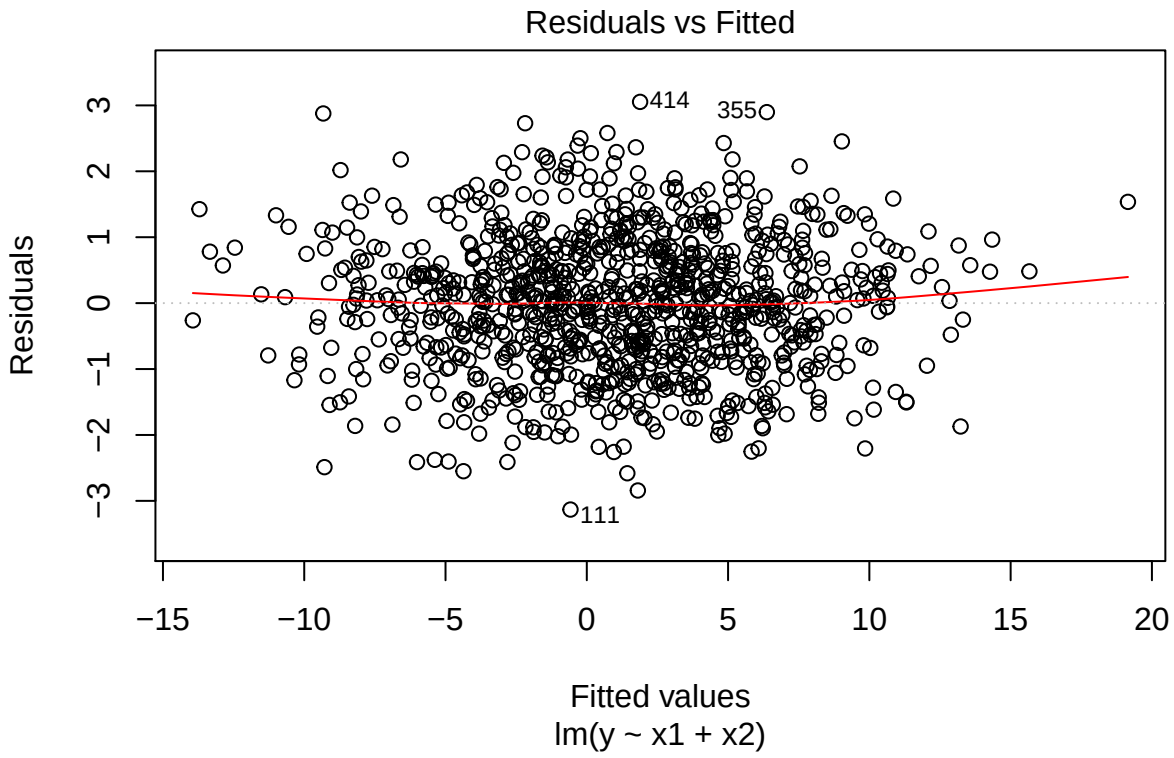
```r
i = 2
y_i = matrix(y, nrow=N)
y_i = y_i[-i,]
X_i = X
X_i <- X_i[-i,]

x1_i = x1[-i]
x2_i = x2[-i]

model_no_i = lm(y_i ~ x1_i + x2_i)
model_no_i$coefficients
```

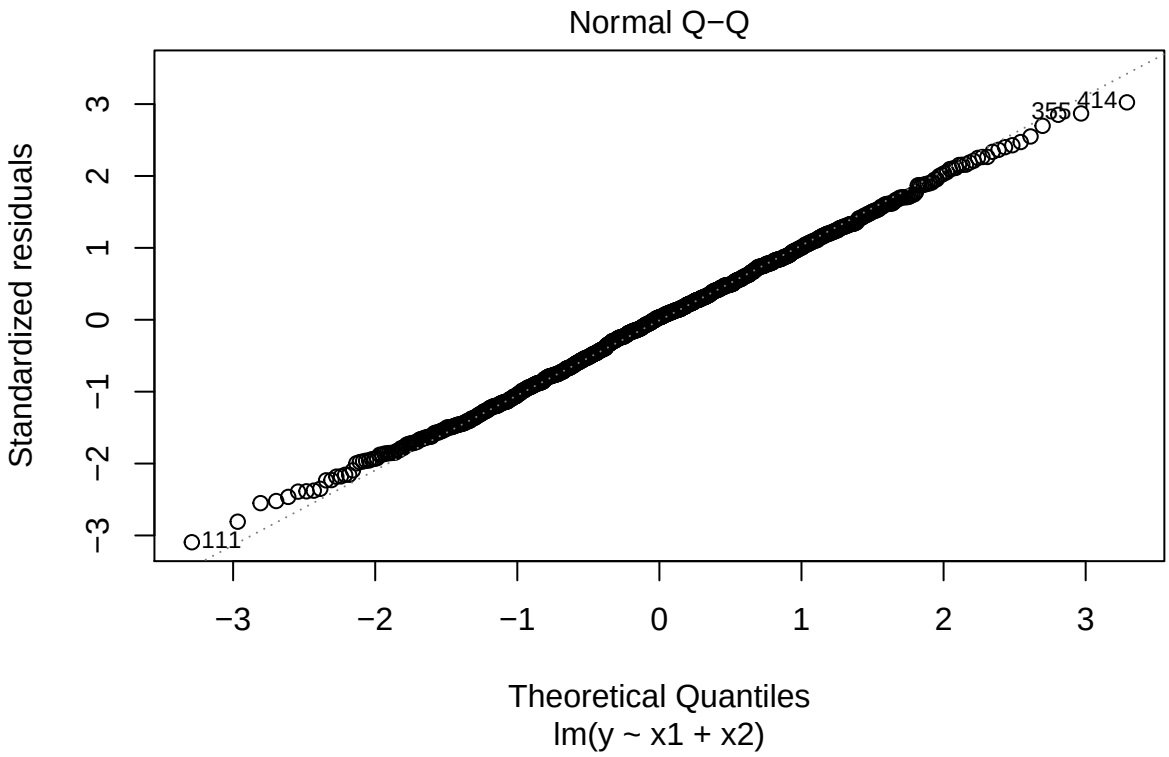
```
## (Intercept)        x1_i        x2_i
##   1.0300813   1.0177379   0.9926275
```

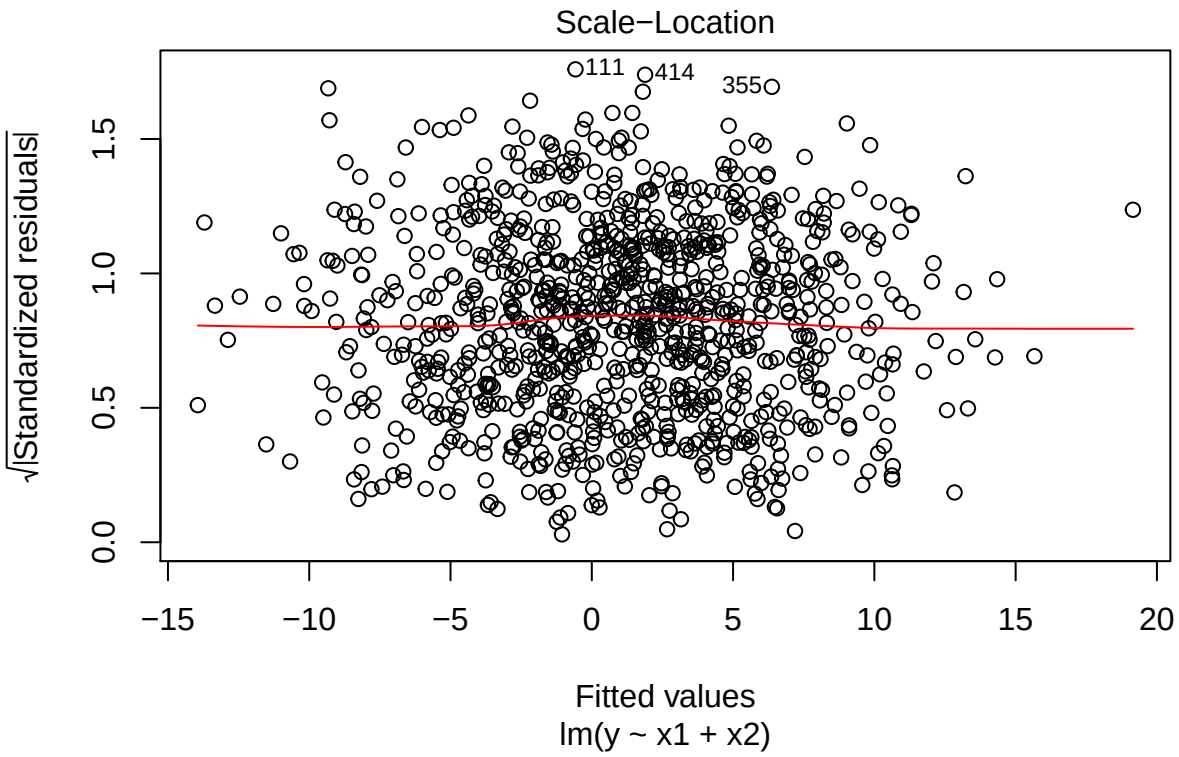
```r
model_no_i$coefficients - beta
```

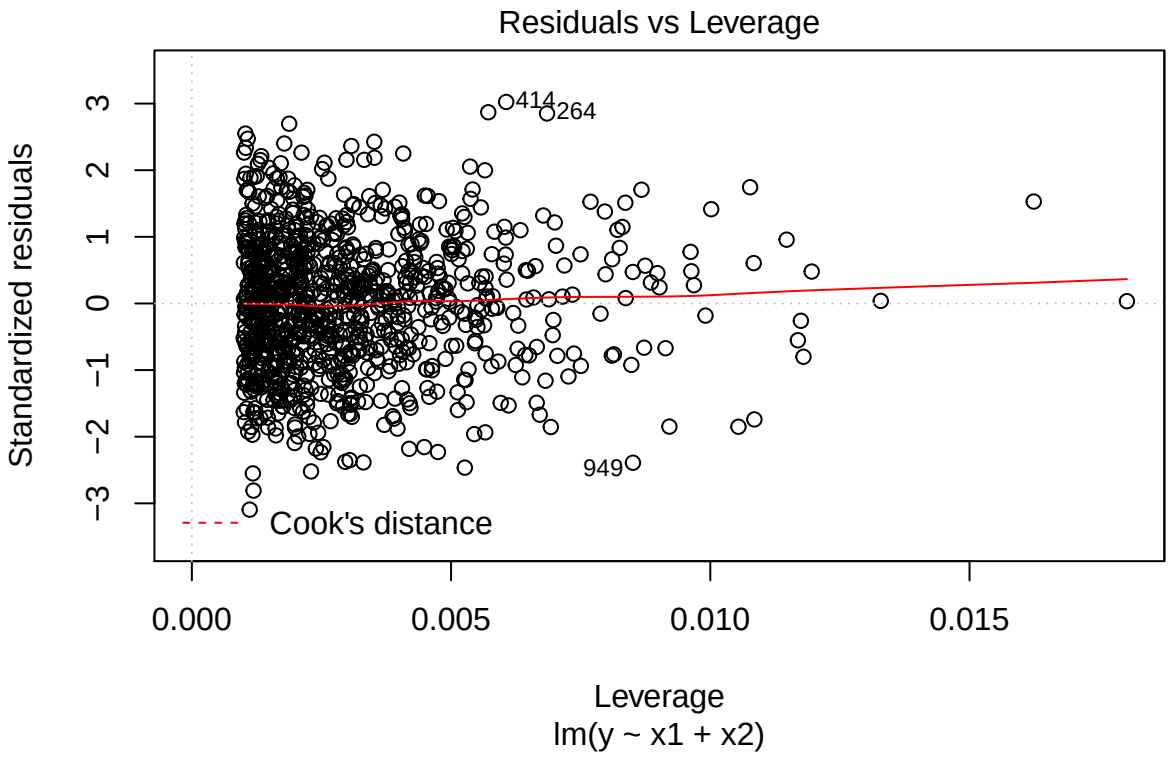
```
##  (Intercept)         x1_i         x2_i
## 1.239695e-04 3.585381e-05 6.958033e-06
```

```r
# Plots below will show most influential data points
plot(fit)
```




Normal Q−Q
Standardized residuals
Theoretical Quantiles
lm(y ~ x1 + x2)
111
355
414
−3
−2
−1
0
1
2
3


Scale−Location
111
414
355
√|Standardized residuals|
0.0
0.5
1.0
1.5
−15
−10
−5
0
5
10
15
20
Fitted values
lm(y ~ x1 + x2)


Residuals vs Leverage
Standardized residuals
3
2
1
0
−1
−2
−3
414
264
949
Cook's distance
0.000
0.005
0.010
0.015
Leverage
lm(y ~ x1 + x2)