Question 5 Solution:

Part a: Computing y,X,β and ϵ

$$\boldsymbol{y} = \begin{bmatrix} 27.3 \\ 42.7. \\ 38.7 \\ 4.5 \\ 23.0 \\ 166.3 \\ 109.7 \\ 80.1 \\ 150.7 \\ 20.3 \\ 189.7 \\ 131.3 \\ 404.2 \\ 149 \end{bmatrix}. \quad \boldsymbol{X} = \begin{bmatrix} 1 & 13.1 \\ 1 & 15.3 \\ 1 & 25.8 \\ 1 & 1.8 \\ 1 & 4.9 \\ 1 & 55.4 \\ 1 & 39.3 \\ 1 & 26.7 \\ 1 & 47.5 \\ 1 & 6.6 \\ 1 & 94.7 \\ 1 & 61.1 \\ 1 & 135.6 \\ 1 & 47.6 \end{bmatrix} \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} \quad \boldsymbol{\epsilon} = \begin{bmatrix} \epsilon_0 \\ \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \epsilon_4 \\ \epsilon_5 \\ \epsilon_6 \\ \epsilon_7 \\ \epsilon_8 \\ \epsilon_9 \\ \epsilon_{10} \\ \epsilon_{11} \\ \epsilon_{12} \\ \epsilon_{13} \end{bmatrix}$$

$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$ *becomes,*

$$\begin{bmatrix} 27.3 \\ 42.7. \\ 38.7 \\ 4.5 \\ 23.0 \\ 166.3 \\ 109.7 \\ 80.1 \\ 150.7 \\ 20.3 \\ 189.7 \\ 131.3 \\ 404.2 \\ 149 \end{bmatrix}. = \begin{bmatrix} 1 & 13.1 \\ 1 & 15.3 \\ 1 & 25.8 \\ 1 & 1.8 \\ 1 & 4.9 \\ 1 & 55.4 \\ 1 & 39.3 \\ 1 & 26.7 \\ 1 & 47.5 \\ 1 & 6.6 \\ 1 & 94.7 \\ 1 & 61.1 \\ 1 & 135.6 \\ 1 & 47.6 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} + \begin{bmatrix} \epsilon_0 \\ \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \epsilon_4 \\ \epsilon_5 \\ \epsilon_6 \\ \epsilon_7 \\ \epsilon_8 \\ \epsilon_9 \\ \epsilon_{10} \\ \epsilon_{11} \\ \epsilon_{12} \\ \epsilon_{13} \end{bmatrix}$$

Part b: Solving the least squares estimator

$\boldsymbol{b} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{y}$

```
```{r}
b = solve(t(X)%*%X, t(X)%*%y)
b
```
```

```
          [,1]
[1,] -1.233836
[2,]  2.701553
```

Part c:

$s^2 = \frac{SS_{Res}}{n-p}$

```
```{r}
e = y - X%*%b
e #Residual errors


```
```

```
             [,1]
 [1,]   -6.8565106
 [2,]    2.3000724
 [3,]  -29.7662361
 [4,]    0.8710405
 [5,]   10.9962256
 [6,]   17.8677893
 [7,]    4.7627957
 [8,]    9.2023660
 [9,]   23.6100596
[10,]    3.7035852
[11,]  -64.9032511
[12,]  -32.5310639
[13,]   39.1032233
[14,]   21.6399042
```

```{r}
n = 14 #sample size
p = 2 #number of parameters
SSRes = sum(e^2)
ssquared = SSRes/(n-p)
ssquared


```

```
[1] 777.1528
```

Part d:

$\boldsymbol{t^T b = [1, 28] b = b_0 + 28 b_1}$

````
```{r}
c(1,28)%*%b
```
````

```
      [,1]
[1,] 74.40965
```

Part e:

$z_i = \frac{e_i}{\sqrt{(s^2(1-H_{ii}))}}$

```
```{r}
a = solve(t(X)%*%X)
a
```
```

```
              [,1]          [,2]
[1,]  0.163081936 -2.230009e-03
[2,] -0.002230009  5.425812e-05
```

```
```{r}
H = X%*%a%*%t(X)
H


```
```

```{r}
z = e/sqrt(ssquared * (1 - diag(H)))
z[13]
```

```
[1] 2.104999
```

Part f:

$D_i = \frac{z_i^2 H_{ii}}{(k+1)(1-H_{ii})}$

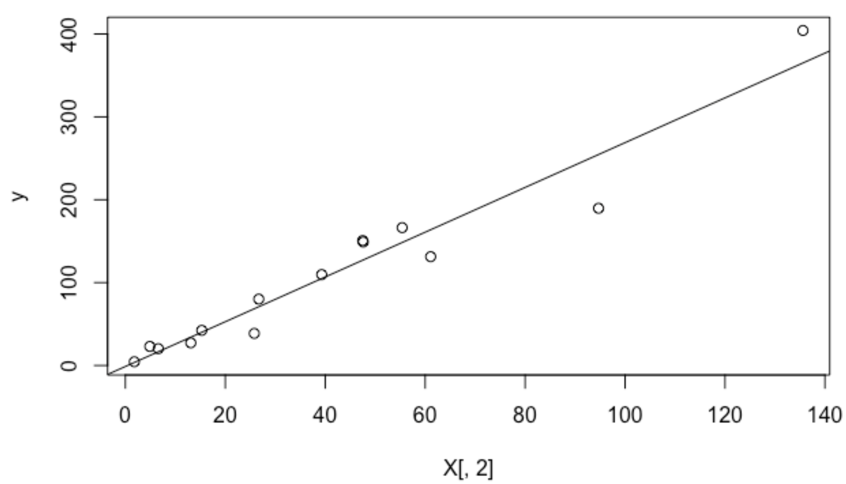
```{r}
k = 1
D = z^2 * (diag(H)/(1-diag(H))) * 1/(k+1)
D[13]
```

```
[1] 2.774008
```

Part g:

```{r}
plot(X[,2],y)
abline(b[1],b[2])

```



Full explaination: The Cook's distance certainly indicates it should be of some concern; however looking at the plot, it seems that the fit is actually okay. There is considerable evidence for heteroskedasticity — the variance increases with x (the design variable). Sea scallops has (by far) the largest x and so may be prone to a larger variance than the remaining points. The high Cook's distance therefore comes primarily from a very high leverage, rather than a bad fit to the model.

END OF ASSIGNMENT!!