HW10

Problem 1.

a)

This part has been done on Rstudio.

```{r}
y <- c(1, 0, 0, 1, 2, 3, 3)
x1 <- c(-3, -2, -1, 0, 1, 2, 3)
x2 <- c(5, 0, -3, -4, -3, 0, 5)
x3 <- c(-1, 1, 1, 0, -1, -1, 1)
df <- data.frame(y, x1, x2, x3)
df
```

Description: df [7 x 4]

y <dbl>	x1 <dbl>	x2 <dbl>	x3 <dbl>
1	-3	5	-1
0	-2	0	1
0	-1	-3	1
1	0	-4	0
2	1	-3	-1
3	2	0	-1
3	3	5	1

7 rows

Results:

Alejandro C Parra Garcia

```{r}
m1 <- lm(y ~ x1 + x2 + x3, data = df)
summary(m1)
```

```
Call:
lm(formula = y ~ x1 + x2 + x3, data = df)

Residuals:
       1        2        3        4        5        6 
-0.02381  0.07143 -0.07143  0.04762 -0.07143  0.07143 
       7 
-0.02381 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  1.42857    0.03367   42.43 2.88e-05 ***
x1           0.50000    0.01684   29.70 8.38e-05 ***
x2           0.11905    0.00972   12.25 0.001172 ** 
x3          -0.50000    0.03637  -13.75 0.000833 ***
---
Signif. codes:  
0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.08909 on 3 degrees of freedom
Multiple R-squared:  0.9975,	Adjusted R-squared:  0.9951 
F-statistic:   407 on 3 and 3 DF,  p-value: 0.0002058
```

We get the following equation:

$$\hat{y} = 1.42857 + 0.5 * x_1 + 0.11905 * x_2 - 0.5 * x_3 + \varepsilon$$

And the following $\hat{\beta}$:

$$\hat{\beta} = \begin{bmatrix} 1.42857 \\ 0.5 \\ 0.11905 \\ -0.5 \end{bmatrix}$$

b)

$$\hat{y} = 1.42857 + 0.5 * 1 + 0.11905 * (-3) - 0.5 * (-1)$$

$$\hat{y} = 2.07142$$

```{r}
x_1=1
x_2=-3
x_3=-1
y_hat = 1.42857 + 0.50000 * x_1 + 0.11905 * x_2 -0.50000 * x_3
y_hat
```

```
[1] 2.07142
```

The output is 2.07142, while the original y for those x_1, x_2, x_3 values was 2.

This mismatch is due to the ε, it is a random error, as the equation is an approximation to the true data.

c)

$$PI_\alpha = a^T\hat{\beta} \pm t_{\frac{\alpha}{2},n-k-1} * \hat{\sigma}\sqrt{1 + a^T(X^TX)^{-1}a}$$

$$X = \begin{bmatrix} 1 \\ -3 \\ -1 \end{bmatrix}$$

$$a = \begin{bmatrix} 1 \\ 1 \\ -3 \\ -1 \end{bmatrix}$$

$$SSE = Y^TY - \left(X\hat{\beta}\right)^T Y = 0.0238$$

$$\hat{\sigma}^2 = \frac{SSE}{n-k-1} = \frac{0.0238}{7-3-1} = 0.007933333$$

$$\hat{\sigma} = \sqrt{\hat{\sigma}^2} = \sqrt{0.007933333} = 0.08906926$$

$$t_{\frac{\alpha}{2},n-k-1} = t_{\frac{0.05}{2},7-3-1} = t_{0.025,3} = 3.182$$

$$a^T\hat{\beta} \pm t_{\frac{\alpha}{2},n-k-1} * \hat{\sigma}\sqrt{1 + a^T(X^TX)^{-1}a} =$$

$$= \begin{bmatrix} 1 \\ 1 \\ -3 \\ -1 \end{bmatrix}^T \begin{bmatrix} 1.42857 \\ 0.5 \\ 0.11905 \\ -0.5 \end{bmatrix} \pm 3.182 * 0.08906926 \sqrt{1 + \begin{bmatrix} 1 \\ 1 \\ -3 \\ -1 \end{bmatrix}^T \left(\begin{bmatrix} 1 \\ -3 \\ -1 \end{bmatrix}^T \begin{bmatrix} 1 \\ -3 \\ -1 \end{bmatrix} \right)^{-1} \begin{bmatrix} 1 \\ 1 \\ -3 \\ -1 \end{bmatrix}} =$$

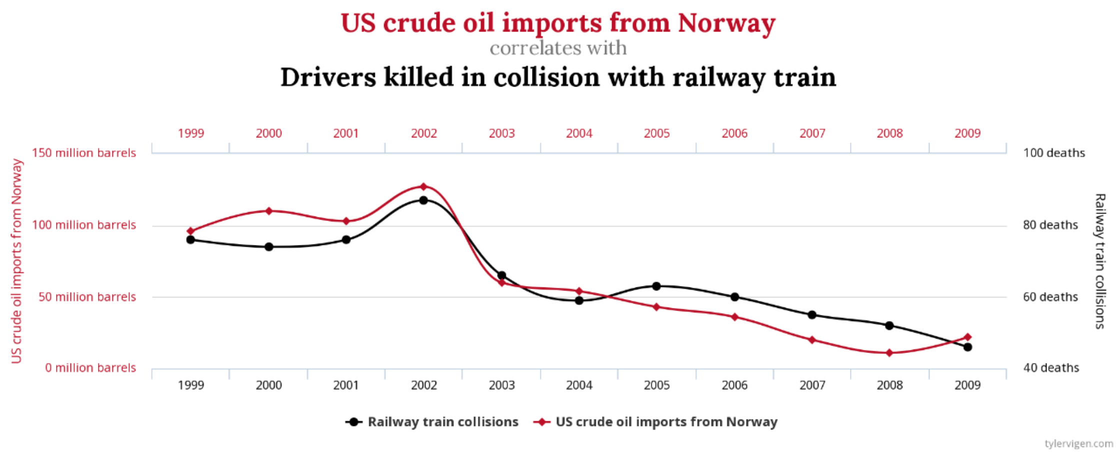
$$= 2.07142 \pm 0.341561 = (1.729859, 2.412981)$$

$$PI_{0.05} = (1.729859, 2.412981)$$

Problem 2.

1)

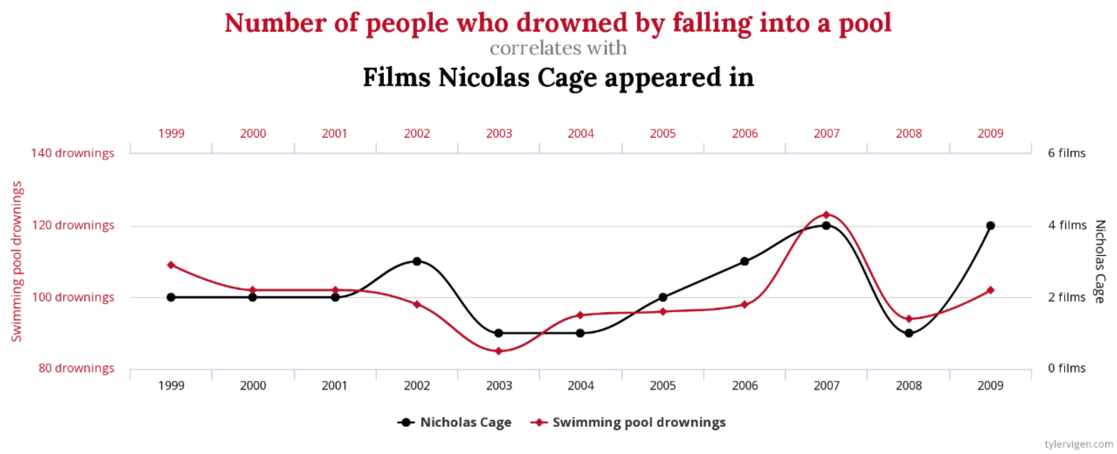
For example, US crude imports from Norway in number of barrels and the number of drivers killed in collision with trains:



r=0.954509

2)

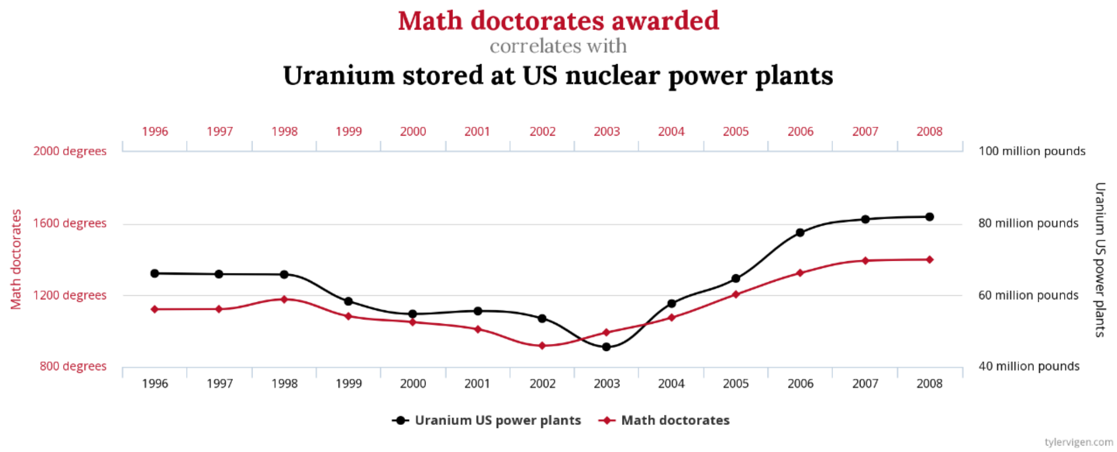
Another one can be the number of Films Nicolas Cage appeared in and the number of people who drowned by falling into a pool:



r=0.666004

3)

Math doctorates awarded and the amount of Uranium stored in the US nuclear power plants:



r=0.952257

Source: http://www.tylervigen.com/spurious-correlations