HW 3 Problem 1

$\beta_{12} = \beta_{33}$ $\beta_{22}\beta_{32} = 0$

1) The null hypothesis is there is no mediation effect of x on y through m.

2) From code:

```
setwd("/Users/alwinlin/Desktop/DSC 382 Regression")
data <- read.csv(file = 'HW3_data.csv', header = TRUE)
eq1 <- lm(y ~ x, data = data) #First equation
eq2 <- lm(m ~ x, data = data) #Second equation regress m on y
eq3 <- lm(y ~ m + x, data = data) #Third equation regress the last equation with m & x
beta11 <- coef(eq1)[1]
beta12 <- coef(eq1)[2]
beta21 <- coef(eq2)[1]
beta22 <- coef(eq2)[2]
beta31 <- coef(eq3)[1]
beta32 <- coef(eq3)[2]
beta33 <- coef(eq3)[3]
```

$\beta_{11} = -1.61$ $\beta_{12} = 2.64$ $\beta_{21} = -0.772$

$\beta_{22} = 2.06$ $\beta_{31} = -0.666$

$\beta_{32} = 1.22$ $\beta_{33} = 0.133$

3) knowing $\sigma^2 = 1$ & $var = \sigma^2 C_j$ (slide 24)

$C_j = (X'X)^{-1}$ and var is the second diag

```
Matrix2 <- model.matrix(eq2)
vareq2 <- solve(t(Matrix2) %*% Matrix2)
var22 <- diag(vareq2)[2]   # Variance of beta22
Matrix3 <- model.matrix(eq3)
vareq3 <- solve(t(Matrix3) %*% Matrix3)
var32 <- diag(vareq3)[2]   # Variance of beta32
```

$var(\beta_{22}) = 0.00113$

$var(\beta_{32}) = 0.0084$

4) $$Z = \frac{\hat{\beta}_{12} - \hat{\beta}_{33}}{\sqrt{\beta_{22}^2 Var(\beta_{32}) + \beta_{32}^2 Var(\beta_{22})}} =$$

```
> Z_sobel
        x
12.99453
```

$\boxed{Z = 12.995}$

```
Z_sobel <- (beta12 - beta33) / sqrt((beta22^2 * var32) + (beta32^2 * var22))
```

5) $p = 2(1 - \Phi(|z|)) \approx \boxed{0}$

We reject the null hypothesis at the 5% significance level of no mediation effect

```
p_val <- 2 * (1 - pnorm(abs(Z_sobel)))
```

p_val	Named num 0

Problem 2

1) residual sum of squares $= \sum_{i=1}^{n}(y_i - x_i\beta)^2$, $\frac{d}{d\beta}(RSS) = -2\sum_{i=1}^{n} x_i(y_i - x_i\beta) = 0$

$$\to \sum_{i=1}^{n} x_i y_i - \beta\sum_{i=1}^{n} x_i^2 = 0 \to \hat{\beta} = \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2}$$

Since $\sum_{i=1}^{n} x_i^2 = n$

$$\boxed{\hat{\beta} = \frac{\sum_{i=1}^{n} x_i y_i}{n} = n^{-1}\sum_{i=1}^{n} x_i y_i}$$

2) RSS from above $= \sum_{i=1}^{n}(y_i - x_i\beta)^2$ expand $\to \sum_{i=1}^{n} y_i^2 - 2\beta\sum_{i=1}^{n} x_i y_i + \hat{\beta}^2\sum_{i=1}^{n} x_i^2$

We know $\sum_{i=1}^{n} x_i^2 = n$, so $\sum_{i=1}^{n} x_i y_i = n\beta$. Substitute into ↑ and get:

$$RSS = \sum_{i=1}^{n} y_i^2 - 2\hat{\beta}(n\hat{\beta}) + \hat{\beta}^2(n) = \boxed{\sum_{i=1}^{n} y_i^2 - n\hat{\beta}^2}$$

3) $F = \dfrac{\hat{e}_{red}'\hat{e}_{red} - \hat{e}'\hat{e}}{\hat{e}'\hat{e}/(n-p)}$ where $\hat{e}_{red}'\hat{e}_{red} = \sum_{i=1}^{n} y_i^2$, $\hat{e}'\hat{e} = \sum_{i=1}^{n} y_i^2 - n\hat{\beta}^2$

$p=1$

numerator $= \hat{e}_{red}'\hat{e}_{red} - \hat{e}'\hat{e} = \sum_{i=1}^{n} y_i^2 - \left(\sum_{i=1}^{n} y_i^2 - n\hat{\beta}^2\right) = n\hat{\beta}^2$

denominator $= \hat{e}'\hat{e}/n-p = \dfrac{\sum_{i=1}^{n} y_i^2 - n\hat{\beta}^2}{n-1}$

$$F = \frac{n\hat{\beta}^2}{\frac{\sum_{i=1}^{n} y_i^2 - n\hat{\beta}^2}{n-1}} = \boxed{\frac{(n-1)n\hat{\beta}^2}{\sum_{i=1}^{n} y_i^2 - n\hat{\beta}^2} \quad F \sim F_{1,n-1} \text{ under } H_0}$$

4) $\hat{\beta} = n^{-1}\sum_{i=1}^{n} x_i y_i \leftarrow y_i = \sigma\varepsilon_i$ so $\hat{\beta} = \frac{\sigma}{n}\sum_{i=1}^{n} x_i\varepsilon_i$

Mean: $E[\hat{\beta}] = \frac{\sigma}{n}\sum_{i=1}^{n} x_i E[\varepsilon_i] = 0$

Var: $var(\hat{\beta}) = \frac{\sigma^2}{n^2}\sum_{i=1}^{n} x_i^2 \leftarrow \sum_{i=1}^{n} x_i^2 = n \to var(\hat{\beta}) = \frac{\sigma^2}{n}$

$$\boxed{\hat{\beta} \mid H_0 \sim N\left(0, \frac{\sigma^2}{n}\right)}$$

5) From slide 40, $T = \dfrac{\hat{\beta}}{\hat{\sigma}\sqrt{c_1}} = \dfrac{\sqrt{n}\,\hat{\beta}}{\hat{\sigma}}$ $\quad T^2 = \dfrac{n\hat{\beta}^2}{\hat{\sigma}^2} = \dfrac{n\hat{\beta}^2}{\underbrace{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}_{\sum_{i=1}^{n} y_i^2 - n\hat{\beta}^2}/(n-1)} = \boxed{\dfrac{n\hat{\beta}^2(n-1)}{\sum_{i=1}^{n} y_i^2 - n\hat{\beta}^2}}$

boxed is same eq as F from Q3