Hw3

Regression and Predictive

Rafael Espinosa

PROBLEM 1

(1) The null hypothesis of the Sobel test is that there is no mediation effect by $\beta_{22}=0$ or $\beta_{32}=0$ which both imply that $\beta_{12}=\beta_{33}$. So formally we use the null hypothesis $\boxed{H_0: \beta_{12}=\beta_{33}}$

(2)
With the first model, $\hat{y}$ dependent of X, we obtain using R

$$\boxed{\hat{y} = \hat{\beta}_{11} + \hat{\beta}_{12} X}$$

$$\hat{\beta} = (\underbrace{2.977290}_{\hat{\beta}_{11}}, \underbrace{1.784123}_{\hat{\beta}_{12}})$$

With the second model, m dependent of X, we obtain using R

$$\boxed{m = \beta_{21} + \beta_{22} X}$$

$$\hat{\beta} = (\underbrace{0.9740537}_{\hat{\beta}_{21}}, \underbrace{0.91042}_{\hat{\beta}_{22}})$$

with the third model, $\hat{y}$ dependent of m and X, we obtain using R

$$\boxed{\hat{y} = \hat{\beta}_{31} + \hat{\beta}_{32}\, m + \hat{\beta}_{33}\, X}$$

$$\hat{\beta} = (\underbrace{1.178648}_{\hat{\beta}_{31}}, \underbrace{1.846554}_{\hat{\beta}_{32}}, \underbrace{0.1029832}_{\hat{\beta}_{33}})$$

(3) The variances are using R

$$\boxed{\begin{aligned} Var(\hat{\beta}_{22}) &= 0.0054901181 \\ Var(\hat{\beta}_{32}) &= 0.004840097 \end{aligned}}$$

where we used that each β follows a normal distribution with

$$\hat{\beta} = (\beta, C_j\, \sigma^2)$$

where C_j is the jth component of the $(XX')^{-1}$ matrix

(4) $\hat{\beta}_{12} = 1.784123$

$\hat{\beta}_{33} = 0.1029832$

$\hat{\beta}_{22} = 0.9104200$

$\hat{\beta}_{32} = 1.8465540$

$Var(\hat{\beta}_{22}) = 0.0054901181$

$Var(\hat{\beta}_{32}) = 0.004840097$

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

$$Z = \frac{1.784123 - 0.1029832}{\sqrt{(0.9104200)^2 0.004840097 + (1.8465540)^2 (0.0054901181)}}$$

$$\boxed{Z = 11.15031}$$

(5) The p-value according to R is 0. This is expected due to the big z value obtained.

Given that the p-value is zero, we empathically reject the null hypothesis (No mediation effect).

Problem 2

(1) We want to minimize the sum of squares.

So, we calculate

$$\frac{\partial}{\partial \hat{\beta}}\left[\sum_{i=1}^{N}\left(y_i - x_i\hat{\beta}\right)^2\right] = 0$$

$$\frac{\partial}{\partial \hat{\beta}}\left[\sum_{i=1}^{N}\left(y_i - x_i\hat{\beta}\right)^2\right] = -2\sum_{i=1}^{N}\left(y_i - x_i\hat{\beta}\right)x_i = 0$$

$$\Rightarrow \sum_{i=1}^{N} y_i x_i - \sum_{i=1}^{N} x_i^2 \hat{\beta} = 0$$

$$\Rightarrow$$

$$\sum_{i=1}^{N} y_i x_i - \hat{\beta}\underbrace{\sum_{i=1}^{N} x_i^2}_{1} = 0$$

$$\Rightarrow$$

$$\boxed{\hat{\beta} = \sum_{i=1}^{N} y_i x_i = z}$$

(2) $$\sum_{i=1}^{N}\left(y_i - \hat{y}_i\right)^2 = \sum_{i=1}^{N}\left(y_i - x_i\hat{\beta}\right)^2$$

$$= \sum_{i=1}^{N}\left(y_i^2 - 2y_i x_i\hat{\beta} + x_i^2\hat{\beta}^2\right)$$

$$= \sum_{i=1}^{N} y_i^2 - 2\hat{\beta}\sum_{i=1}^{N} y_i x_i + \hat{\beta}^2\underbrace{\sum_{i=1}^{N} x_i^2}_{1}$$

$$= \sum_{i=1}^{N} y_i^2 - 2\hat{\beta}\sum_{i=1}^{N} y_i x_i + \hat{\beta}^2$$

Using the result from part (1), i.e. that $\hat{\beta} = \sum_{i=1}^{N} y_i x_i = z$, we obtain

$$\boxed{\sum_{i=1}^{N}\left(y_i - \hat{y}_i\right)^2 = \sum_{i=1}^{N} y_i^2 - 2z^2 + z^2 = \sum_{i=1}^{N} y_i^2 - z^2}$$

(3) The F-statistic for testing $H_0: \beta = 0$ is calculated by

$$F = \frac{\hat{e}_{red}'\hat{e}_{red} - \hat{e}'\hat{e}}{\hat{e}'\hat{e}/(n-p)}$$

where for this case $\hat{e}_{red}'\hat{e}_{red} = \sum_{i=1}^{N} y_i^2$

and $\hat{e}'\hat{e} = \sum_{i=1}^{N}(y_i - \hat{y}_i)^2 = \sum_{i=1}^{N} y_i^2 - z^2$ and $p=1$,

So,

$$F = \frac{\sum_{i=1}^{N} y_i^2 - \left(\sum_{i=1}^{N} y_i^2 - z^2\right)}{\left(\sum_{i=1}^{N} y_i^2 - z^2\right)/(n-1)}$$

$$= \frac{z^2}{\left(\sum_{i=1}^{N} y_i^2 - z^2\right)/(n-1)} = \frac{z^2(n-1)}{\sum_{i=1}^{N} y_i^2 - z^2}$$

(4) Given that σ is unknown, the distribution of $z = \hat{\beta} = \sum_{i=1}^{N} x_i y_i$ will be a t distribution with $n-1$ degrees of freedom.

(5) We have that (according to our result in (3)),

$$F = \frac{z^2(n-1)}{\sum_{i=1}^{N} y_i^2 - z^2} = \frac{z^2}{\left(\frac{1}{n-1}\right)\left(\sum_{i=1}^{N} y_i^2 - z^2\right)}$$

$$= \frac{z^2}{\left(\frac{1}{n-1}\right) \sum_{i=1}^{N}(y_i - \hat{y}_i)^2} = \frac{z^2}{\hat{\sigma}^2}$$

where we used what we found in (2)

$$\sum_{i=1}^{N} y_i^2 - z^2 = \sum_{i=1} (y_i - \hat{y}_i)^2$$

On the other hand, we can write a student t random variable with $n-1$ degrees of freedom as

$$T = z' \Big/ \sqrt{X_{n-1}^2 / (n-1)}$$

where Z' is a standard normal and independent of χ - squared χ^2_{n-p} variable

We can use

$$Z' = \frac{Z/\sigma}{\sqrt{\sum x_i^2}} \quad \text{and} \quad \frac{\chi_{n-1}}{n-1} = \frac{\hat{\sigma}^2}{\sigma^2}$$

$$\text{where } \hat{\sigma} = \frac{1}{n-1} \sum_{i=1}^{N} (y_i - \hat{y}_i)^2, \quad \sum_{i=1}^{N} x_i^2 = 1$$

so

$$T = \frac{(Z/\sigma)}{\sqrt{\frac{\hat{\sigma}}{\sigma^2}}} = \frac{Z}{\hat{\sigma}} = \frac{Z}{\hat{\sigma}}$$

Hence $F = T^2$, so the F test and T test are the same.