Homework 3 Solutions

Problem 1

Given the provided dataset, consider the model regressing y on x and m. Conduct a Sobel test to see whether there is a mediation effect. You need to consider the three relevant models,

$$y_i = \beta_{11} + \beta_{12}x_i + \epsilon_i,$$

$$m_i = \beta_{21} + \beta_{22}x_i + \epsilon_i,$$

and

$$y_i = \beta_{31} + \beta_{32}m_i + \beta_{33}x_i + \epsilon_i.$$

Suppose the variance of ϵ_i is known with $\sigma^2 = 1$ for all the three models. Use the level of significance $\alpha = 0.05$

(1) (4 pts) What is the null hypothesis of the Sobel test?

Solution:

$$\text{The null hypothesis is } H_0: \ \beta_{12} = \beta_{33}$$

(2) (8 pts) Regress the three models and find the estimators.

Solution:

$$\hat{\beta}_1 = (2.98,\ 1.78) \qquad \hat{\beta}_2 = (0.97,\ 0.91) \qquad \hat{\beta}_3 = (1.18,\ 1.85,\ 0.10)$$

(3) (10 pts) Find the variance of $\hat{\beta}_{22}$ and $\hat{\beta}_{32}$.

Solution:

$$Var(\hat{\beta}_{22}) = 0.0055 \qquad Var(\hat{\beta}_{32}) = 0.0048$$

(4) (10 pts) Find the z test statistic.

Solution:

$$z = \frac{\hat{\beta}_{12} - \hat{\beta}_{33}}{\sqrt{\hat{\beta}_{22}^2 Var(\hat{\beta}_{32}) + \hat{\beta}_{32}^2 Var(\hat{\beta}_{22})}} = 11.15$$

(5) (8 pts) Find the p-value and the conclusion of the test.

Solution:

$$p = 2(1 - \text{pnorm}(z, 0, 1)) = 0 < \alpha = 0.05$$

So, we reject the null hypothesis of no mediation effect.

Problem 2

Consider the usual linear model,

$$y_i = x_i\beta + \sigma\epsilon_i,$$

where σ is unknown and $\sum_{i=1}^n x_i^2 = 1$.

(1) (12 pts) Show that the least square estimator $\hat{\beta} = \sum_{i=1}^{n} x_i y_i$, and write this as z.

Solution:

$$I(\beta) = \sum_{i=1}^{n} (y_i - x_i\beta)^2 = \sum_{i=1}^{n} (y_i^2 + x_i^2\beta^2 - 2x_i y_i \beta)$$

$$\frac{dI(\beta)}{d\beta} = \sum_{i=1}^{n} (2x_i^2\beta - 2x_i y_i) = 2\beta \sum_{i=1}^{n} x_i^2 - 2\sum_{i=1}^{n} x_i y_i = 2\beta - 2\sum_{i=1}^{n} x_i y_i = 0$$

$$\text{So} \quad \hat{\beta} = \sum_{i=1}^{n} x_i y_i$$

(2) (12 pts) Show that the residual sum of squares is given by

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

Solution:

$$\begin{aligned} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 &= \sum_{i=1}^{n} (y_i - x_i\hat{\beta})^2 = \sum_{i=1}^{n} (y_i - x_i z)^2 \\ &= \sum_{i=1}^{n} y_i^2 - 2z \sum_{i=1}^{n} x_i y_i + z^2 \sum_{i=1}^{n} x_i^2 \\ &= \sum_{i=1}^{n} y_i^2 - 2z^2 + z^2 = \sum_{i=1}^{n} y_i^2 - z^2 \end{aligned}$$

(3) (12 pts) Hence, show that the F statistic for testing $H_0 : \beta = 0$ is given by

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

Solution:

Here the reduced model is

$$y_i = \sigma\epsilon_i \qquad \hat{y}_{red} = 0 \qquad \hat{e}'_{red}\hat{e}_{red} = \sum_{i=1}^{n} y_i^2$$

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

$$F = \frac{\hat{e}'_{red}\hat{e}_{red} - \hat{e}'\hat{e}}{\hat{e}'\hat{e}/(n-p)} = \frac{\sum_{i=1}^{n} y_i^2 - (\sum_{i=1}^{n} y_i^2 - z^2)}{(\sum_{i=1}^{n} y_i^2 - z^2)/(n-1)} = \frac{z^2(n-1)}{\sum_{i=1}^{n} y_i^2 - z^2}$$

(4) (12 pts) If the null hypothesis is true, what is the distribution of z?

Solution:

If the null hypothesis $H_0 : \beta = 0$ is true, than

$$y_i = \sigma\epsilon_i \qquad z = \sum_{i=1}^{n} x_i y_i = \sigma \sum_{i=1}^{n} x_i \epsilon_i \qquad \epsilon_i \overset{iid}{\sim} N(0,1)$$

$$E(z) = \sigma \sum_{i=1}^{n} x_i E(\epsilon_i) = 0 \qquad Var(z) = \sigma^2 \sum_{i=1}^{n} x_i^2 Var(\epsilon_i) = \sigma^2$$

$$\text{so} \quad z \sim N(0, \sigma^2)$$

(5) (12 pts) Hence, show that the Student–t test statistic for testing the null hypothesis is

$$T = \frac{z}{\hat{\sigma}}, \qquad \text{with} \quad \hat{\sigma}^2 = \frac{1}{n-1} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

and confirm that the F test and the T test are the same.

Solution:

$$\frac{z}{\sigma} \sim N(0,1) \qquad (n-1)\frac{\hat{\sigma}^2}{\sigma^2} \sim \chi^2_{n-1}$$

$$\text{So} \quad T = \frac{z}{\hat{\sigma}} = \frac{z/\sigma}{\sqrt{(n-1)\frac{\hat{\sigma}^2}{\sigma^2}/(n-1)}} \sim t_{n-1}$$

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

$$\text{So} \quad T^2 = \frac{z^2}{\hat{\sigma}^2} = \frac{z^2(n-1)}{\sum_{i=1}^{n} y_i^2 - z^2} = F$$