Chapter 4 Multiple Regression

Arthur Chen

February 18, 2023

Exercise Set A

Problem 7

Yule's regression model for pauperism:

$$\Delta Paup = a + b \times \Delta Out + c \times \Delta Old + d \times \Delta Pop + error$$

can be translated into matrix notation: $Y = X\beta + \epsilon$. We assume that Y_i are the observed values of $X\beta + \epsilon$, that the ϵ_i are iid with mean zero and variance σ^2, and that ϵ is independent of X. For the metropolitan unions and the period 1871-81:

Part a

What are X and Y?

Take all of the entries of Table 1.3 and subtract 100 from them. Y is the column vector of Paup, and X is the columns Out, Old, and Pop, with a column of all 1's added on. (I will assume that the 1's column is added to the left of the X's.)

Part b

What are the observed values of X_{41}? X_{42}? Y_4?

$X_{41} = 1$, since that's the column of all ones. X_{42} is the Out ratio for Chelsea, which is $21 - 100 = -79$. Y_4 is the Pauper for Chelsea, which is $64 - 100 = -36$.

Part c

Where do we look in $(X'X)^{-1}X'Y$ to find the estimated coefficient of ΔOut?

In the second entry, since the Out column is right after the Intercept column.

Problem Set B

Problem 14

We have from the equations in Chapter 2 that

$$\hat{a} = \bar{y} - slope\bar{x}$$
$$\hat{b} = corr(x, y)\frac{\mathrm{SD}(y)}{\mathrm{SD}(x)}$$

In this case, X is

$$X = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}$$

so we get that

$$X^T X = \begin{pmatrix} n & \sum X_i \\ \sum X_i & \sum X_i^2 \end{pmatrix}$$

We notice that

$$\det(X^T X) = n \sum X_i - (\sum X_i)^2 = n^2(\frac{1}{n} \sum X_i - \overline{X}^2) = n^2 \mathrm{Var}(X)$$

and so

$$(X^T X)^{-1} = \frac{1}{n^2 \mathrm{Var}(X)} \begin{pmatrix} \sum X_i^2 & -\sum X_i \\ -\sum X_i & n \end{pmatrix}$$

Multiplying out $X^T Y$ gives

$$X^T Y = \begin{pmatrix} \bar{Y} \\ \sum X_i Y_i \end{pmatrix}$$

and so

$$\hat{\beta} = \frac{1}{n^2 \mathrm{Var}(X)} \begin{pmatrix} \sum X_i^2 & -\sum X_i \\ -\sum X_i & n \end{pmatrix} \frac{1}{n^2 \mathrm{Var}(X)} \begin{pmatrix} \sum X_i^2 & -\sum X_i \\ -\sum X_i & n \end{pmatrix}$$

Calculating the coefficients,

$$\hat{\beta}_2 = \frac{\frac{1}{n} \sum X_i Y_i - \bar{x}\bar{y}}{\mathrm{Var}x} = \frac{\mathrm{Cov}(x, y)}{\mathrm{Var}(x)} = \frac{\mathrm{Corr}(x, y)\mathrm{SD}(y)}{\mathrm{SD}(x)}$$

which matches $\hat{b}$.