SP25 7140: Homework 1

Danbo CHEN

March 21, 2025

1 Question 1

Proof:

We are to show the following identity is true:

$$\frac{n\sum x_i y_i}{n\sum x_i^2 - \left(\sum x_i\right)^2} = \frac{n\bar{x}\bar{y} + \sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2}$$

$$n\bar{x}\bar{y} = n \cdot \frac{\sum x_i}{n} \cdot \frac{\sum y_i}{n} = \frac{\sum x_i \sum y_i}{n}$$

Now expand the full numerator:

$$n\bar{x}\bar{y} + \sum(x_i - \bar{x})(y_i - \bar{y}) = \frac{\sum x_i \sum y_i}{n} + \sum\left[x_i y_i - x_i\bar{y} - \bar{x}y_i + \bar{x}\bar{y}\right]$$

Break this sum:

$$= \frac{\sum x_i \sum y_i}{n} + \sum x_i y_i - \bar{y}\sum x_i - \bar{x}\sum y_i + n\bar{x}\bar{y}$$

Now simplify each term:

$$\bar{y}\sum x_i = \frac{\sum y_i}{n} \cdot \sum x_i = \frac{\sum x_i \sum y_i}{n}$$

$$\bar{x}\sum y_i = \frac{\sum x_i \sum y_i}{n}, \quad n\bar{x}\bar{y} = \frac{\sum x_i \sum y_i}{n}$$

Substitute back:

$$= \frac{\sum x_i \sum y_i}{n} + \sum x_i y_i - \frac{\sum x_i \sum y_i}{n} - \frac{\sum x_i \sum y_i}{n} + \frac{\sum x_i \sum y_i}{n}$$

Simplify:

$$= \sum x_i y_i$$

$$\sum(x_i - \bar{x})^2 = \sum x_i^2 - 2\bar{x}\sum x_i + n\bar{x}^2$$

Substitute $\bar{x} = \frac{\sum x_i}{n}$:

$$n\bar{x}^2 = \frac{(\sum x_i)^2}{n}$$

Now:

$$\sum(x_i - \bar{x})^2 = \sum x_i^2 - 2 \cdot \frac{\sum x_i}{n} \cdot \sum x_i + \frac{(\sum x_i)^2}{n}$$

$$= \sum x_i^2 - \frac{2(\sum x_i)^2}{n} + \frac{(\sum x_i)^2}{n} = \sum x_i^2 - \frac{(\sum x_i)^2}{n}$$

Final RHS Expression

$$\frac{\sum x_i y_i}{\sum x_i^2 - \frac{(\sum x_i)^2}{n}} = \frac{n \sum x_i y_i}{n \sum x_i^2 - (\sum x_i)^2}$$

Which matches the left-hand side.

2 Question 2

2.1

Based on Model 2.2 in the handout, suppose you want to examine how individuals' experience affects their wage. In addition to just experience itself, you suspect experience may play a different role depending on gender. You further assume that experience does **not** make a difference across occupations.

We propose the following model:

$$Y = \beta_0 + \beta_1 \text{Exper} + \beta_2 \text{Female} + \beta_3 (\text{Female} \times \text{Exper}) + \beta_4 D_{\text{Doc}} + \beta_5 D_{\text{Prof}} + \varepsilon$$

where:

- Female = 1 if individual is female, 0 otherwise
- $D_{\text{Doc}} = 1$ if occupation is doctor, 0 otherwise
- $D_{\text{Prof}} = 1$ if occupation is professor, 0 otherwise (Lawyer is omitted)
- Exper = years of experience

To calculate the wage of a **female doctor** with sample average experience $\overline{\text{Exper}}$, substitute into the model:

$$Y_{\text{Female Doc}} = \beta_0 + \beta_1 \overline{\text{Exper}} + \beta_2(1) + \beta_3(1 \times \overline{\text{Exper}}) + \beta_4(1)$$

Simplifying:

$$Y_{\text{Female Doc}} = (\beta_0 + \beta_2 + \beta_4) + (\beta_1 + \beta_3) \cdot \overline{\text{Exper}}$$

2.2

Estimate the following model:

$$Y = \beta_0 + \beta_{MD} D_{MD} + \beta_{FP} D_{FP} + \beta_{MP} D_{MP} + \beta_{FL} D_{FL} + \beta_{ML} D_{ML} + \varepsilon$$

where $D_{MD}, D_{FP}, D_{MP}, D_{FL}, D_{ML}$ are dummy variables for Male Doctor, Female Professor, Male Professor, Female Lawyer, and Male Lawyer respectively. The omitted category is Male Doctor.

To calculate the wage rate of a **female professor**, we substitute $D_{FP} = 1$, all others $= 0$:

$$Y_{\text{FP}} = \beta_0 + \beta_{FP}$$

Compare this to the wage rate of female professor in model M32 in the handout. Discuss differences in estimates and compare R^2 and adjusted R^2 values between these two models.

2.3

Suppose you suspect the impact of experience on wage follows a quadratic pattern. Based on Model M5, add the squared experience term and estimate the model:

$$Y = \gamma_0 + \gamma_{DD} D_{DD} + \gamma_{DP} D_{DP} + \gamma_E \text{Exper} + \gamma_{ED} (D_{DD} \cdot \text{Exper}) + \gamma_{EP} (D_{DP} \cdot \text{Exper}) + \gamma_{E2} \text{Exper}^2 + \varepsilon$$

To calculate the wage of a **doctor** with average experience $\overline{\text{Exper}}$, substitute $D_{DD} = 1$, $D_{DP} = 0$:

$$Y_{\text{Doc}} = \gamma_0 + \gamma_{DD} + \gamma_E \overline{\text{Exper}} + \gamma_{ED} \cdot \overline{\text{Exper}} + \gamma_{E2} \cdot \overline{\text{Exper}}^2$$

Simplify:

$$Y_{\text{Doc}} = (\gamma_0 + \gamma_{DD}) + (\gamma_E + \gamma_{ED}) \cdot \overline{\text{Exper}} + \gamma_{E2} \cdot \overline{\text{Exper}}^2$$

3 Question 3

3.1

Table 1: Log-Transformed Model Results

	Dependent variable:
	log(Value)
Building	-0.287^{***}
	(0.039)
Genertn	-0.038^{***}
	(0.006)
log(Rain)	0.754^{***}
	(0.029)
Orchard	1.102^{***}
	(0.062)
Range	1.164^{***}
	(0.067)
Crop	0.995^{***}
	(0.068)
Constant	-1.866^{***}
	(0.120)
Observations	651
R^2	0.691
Adjusted R^2	0.688
Residual Std. Error	0.466 (df = 644)
F Statistic	239.993^{***} (df = 6; 644)
Note:	*p<0.1; **p<0.05; ***p<0.01

3.2

Table 2: Model Comparison

	Dependent variable:		
	log(Value)	Value	log(Value)
	(1)	(2)	(3)
Building	-0.286^{***}	-0.659^{***}	-0.287^{***}
	(0.039)	(0.074)	(0.039)
Genertn	-0.031^{***}	-0.077^{***}	-0.038^{***}
	(0.007)	(0.012)	(0.006)
Rain	0.023^{***}		
	(0.001)		
log(Rain)		1.902^{***}	0.754^{***}
		(0.055)	(0.029)
Orchard	1.095^{***}	1.584^{***}	1.102^{***}
	(0.064)	(0.120)	(0.062)
Range	1.202^{***}	1.533^{***}	1.164^{***}
	(0.068)	(0.128)	(0.067)
Crop	1.012^{***}	1.010^{***}	0.995^{***}
	(0.069)	(0.131)	(0.068)
Constant	-0.184^{**}	-2.745^{***}	-1.866^{***}
	(0.090)	(0.231)	(0.120)
Observations	651	651	651
R^2	0.679	0.735	0.691
Adjusted R^2	0.676	0.732	0.688
Residual Std. Error (df = 644)	0.475	0.899	0.466
F Statistic (df = 6; 644)	226.761^{***}	297.239^{***}	239.993^{***}
Note:			*p<0.1; **p<0.05; ***p<0.01