Seminar 1

Giulio Rossetti*

giuliorossetti94.github.io

January 16, 2025

* email: giulio.rossetti.1@wbs.ac.uk

Roadmap

Exercise 1

Exercise 2

Exercise 3

Exercise 1

Model

$$Wage = \beta_0 + \beta_1 Educ + u$$

- OLS estimates of the intercept and slope parameters:

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \quad \text{and} \quad \hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = \frac{cov(x,y)}{var(x)}$$

where

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i \quad , \quad \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \quad \text{and} \quad cov(x,y) = \sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})/n$$

Exercise 1

Cont'd

$$\hat{\beta}_1 = \frac{99.43}{95.67} = 1.04$$

$$\hat{\beta}_0 = 6.78 - 1.04 \times 13.17 = -6.90.$$

Exercise 1

Goodness-of-Fit

- The R^2 of the regression is a measure of Goodness-of-Fit and is defined as:

$$R^2 = \frac{SSE}{SST} = 1 - \frac{SSR}{SST} \in [0,1]$$
$$= \frac{var(X\beta)}{var(Y)} = 1 - \frac{var(u)}{var(Y)}.$$

- Where:
  - $SST = \sum_{i=1}^{n}(y_i - \bar{y})^2 = 206.48 =$ Total Sum of Squares
  - $SSR = \sum_{i=1}^{n} \hat{u}_i^2 = 103.13 =$ Residual Sum of Squares
- Therefore, the R^2 is equal to:

$$R^2 = \frac{SSE}{SST} = 1 - \frac{SSR}{SST} = 1 - \frac{103.13}{206.48} = 0.50$$

Exercise 1

Extended Model and Interpretation

- Suppose we estimate the model:

$$Wage = \beta_0 + \beta_1 Educ + \beta_2 Expert + u$$

- Motivation for including another factor:
  - Multiple Regression Model (MLR) allows investigating the marginal effect of multiple explanatory factors.
  - Holds fixed other factors otherwise hidden in u.
- $\hat{\beta}_j \equiv$ change in the dependent variable due to a one-unit increase in the j-th explanatory variable, *ceteris paribus*.

Roadmap

Exercise 1

Exercise 2

Exercise 3

Exercise 2

Model

$$Fert = \beta_0 + \beta_1 Educ + u$$

- u includes unobservable factors (e.g., income, intelligence).
- Potential correlation with Education.
- Omitted Variable Bias (OVB) may arise.
- unlikely to uncover the ceteris paribus effect

Roadmap

Exercise 1

Exercise 2

Exercise 3

Exercise 3

College GPA Prediction

Model:

$$colgpa = \beta_0 + \beta_1 hsperc + \beta_2 sat + u$$

Source	SS	df	MS		
Model	490.606706	2	245.303353	Number of obs	= 4137
Residual	1303.58897	4134	.315333567	F(2, 4134)	= 777.92
				Prob > F	= 0.0000
Total	1794.19567	4136	.433799728	Root MSE	= .56155

colgpa	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
hsperc	-.0135192	.0005495	-24.60	0.000	-.0145965	-.012442
sat	.0014762	.0000653	22.60	0.000	.0013482	.0016043
_cons	1.391757	.0715424	19.45	0.000	1.251495	1.532018

Exercise 3

College GPA Prediction

Model:

$$colgpa = \underbrace{\beta_0}_{1.392} + \underbrace{\beta_1}_{-0.013} \ hsperc + \underbrace{\beta_2}_{0.0015} \ sat + u$$

- Predicted GPA for $hsperc = 20$, $sat = 1050$: 2.707.

$$\begin{aligned}\hat{y} &= X\hat{\beta} \\ &= 1.392 - 0.012 \times 20 + 0.0015 \times 1050 = 2.707\end{aligned}$$

- Expected GPA for a student in the top 20% of the high school graduating class with a SAT score of 1050 is 2.707.

Exercise 3

Predicted GPA Difference (A vs. B)

- Goal:
  - Students A and B have the same `hsperc`.
  - A's SAT score is 200 points lower than B's.
- Equation:

$$\Delta \text{colgpa} = \hat{\beta}_1 \Delta \text{hsperc} + \hat{\beta}_2 \Delta \text{sat}$$

$$= -0.013 \times 0 + 0.0015 \times (-200) = -0.3$$

- Result: A's GPA is expected to be 0.3 lower than B's.

Exercise 3

SAT Difference for GPA Gap

- Goal: Find Δsat for a GPA difference of 0.5, keeping `hsperc` fixed.
- Equation:

$$\Delta\text{colgpa} = \hat{\beta}_1 \Delta\text{hsperc} + \hat{\beta}_2 \Delta\text{sat}$$

$$0.50 = 0.0015 \Delta\text{sat}$$

- Result:

$$\Delta\text{sat} = \frac{0.50}{0.0015} = 333.33$$

- Interpretation: expected GPA of one student is one-half point higher than the one of a peer if he/she gets a 333.33 higher SAT score.

Exercise 3

Goodness-of-Fit

- From the STATA output, we have:

$$R^2 = \frac{SSE}{SST} = 1 - \frac{SSR}{SST} = 1 - \frac{1303.59}{1794.20} = 0.27.$$

- large fraction of in-sample variation in y which is not explained by x (relevant variables omitted?).