

Simple Linear Regression Overview (Answer Key)

Econ 4050 (Soliman)

2026-01-20

1. How would you define the residual in a simple linear regression?

The residual is the difference between the actual observed value of the outcome variable and the value predicted by the regression model. For observation (i), the residual is:

$$\hat{u}_i = y_i - \hat{y}_i$$

It represents how far the model's prediction is from the observed data point.

2. What does ordinary least squares (OLS) estimation do?

a) Definition in your own words

Ordinary least squares chooses the regression line that fits the data as closely as possible by making the squared prediction errors as small as possible, on average.

b) Formal definition

OLS chooses the coefficients that minimize the sum of squared residuals:

$$\min_{\beta_0, \beta_1} \sum_{i=1}^n (y_i - (\beta_0 + \beta_1 x_i))^2$$

3. Interpret each coefficient in the regression

The estimated regression is:

$$\widehat{\text{overdose deaths}} = 24.39 + 0.0505 \times \text{per capita pills}$$

- **Intercept (24.39):** When per capita pill dispensing is equal to zero, the model predicts approximately 24.4 overdose deaths in a county-year. This is a mechanical prediction from the linear model and may not correspond to a realistic or observed scenario.
- **Slope coefficient on per capita pills (0.0505):** A one-unit increase in per capita pill dispensing is associated with an increase of about 0.05 overdose deaths in a county in a given year, on average, according to the linear regression model.

4. What is the implicit assumption we discussed that one must be careful to consider when using OLS?

The key assumption discussed is **linearity**. OLS assumes that the relationship between the explanatory variable (per capita pill dispensing) and the outcome variable (overdose deaths) can be well approximated by a straight line. If the true relationship is nonlinear, the linear regression model may provide a poor or misleading summary of the relationship.