ISLP — Chapter 3: Conceptual Exercises (Linear Regression)

Conceptual 1 (Exercise 3.7 #6)

Show that the OLS regression line passes through $(\bar{x}, \bar{y})$.

Solution. The model is $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$. OLS minimizes $\sum(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2$. The first normal equation is

$$\sum(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0 \quad \Rightarrow \quad n\hat{\beta}_0 + \hat{\beta}_1 \sum x_i = \sum y_i.$$

Dividing by n,

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}.$$

Substitute $x = \bar{x}$:

$$\hat{y}(\bar{x}) = \hat{\beta}_0 + \hat{\beta}_1 \bar{x} = \bar{y} - \hat{\beta}_1 \bar{x} + \hat{\beta}_1 \bar{x} = \bar{y}.$$

Hence, the fitted line always goes through $(\bar{x}, \bar{y})$.

Interpretation. This shows that OLS balances the errors so that the regression passes through the centroid of the data cloud.

—

Conceptual 2 (Exercise 3.7 #7)

Prove that in simple regression $R^2 = r^2$.

Solution. Assume centered data ($\bar{x} = \bar{y} = 0$). Then

$$\hat{\beta}_1 = \frac{\sum x_i y_i}{\sum x_i^2} = \frac{S_{XY}}{S_{XX}}.$$

The explained sum of squares is

$$\text{SSR} = \sum \hat{y}_i^2 = \hat{\beta}_1^2 S_{XX} = \frac{S_{XY}^2}{S_{XX}}.$$

The total sum of squares is $\text{SST} = \sum y_i^2 = S_{YY}$. Thus

$$R^2 = \frac{\text{SSR}}{\text{SST}} = \frac{S_{XY}^2}{S_{XX} S_{YY}} = \left(\frac{S_{XY}}{\sqrt{S_{XX} S_{YY}}} \right)^2 = r^2.$$

Interpretation. In simple regression, the model's explanatory power is exactly the squared correlation.

—

Conceptual 3 (Exercise 3.7 #1)

Explain the difference between correlation and regression.

Solution. Correlation is a symmetric statistic: $r_{XY} = r_{YX}$. It measures the strength of linear association between X and Y but does not distinguish dependent from independent variable. Regression, on the other hand, is directional: it models Y as a function of X (or several X's). It produces coefficients that can be used for prediction and inference.

Interpretation. Correlation is about *association*; regression is about *prediction and explanation.*

—

Conceptual 4 (Exercise 3.7 #2)

What does R^2 represent?

Solution. R^2 is defined as

$$R^2 = 1 - \frac{\text{SSE}}{\text{SST}} = \frac{\text{SSR}}{\text{SST}}.$$

It measures the proportion of variability in Y explained by the linear model. It is bounded between 0 and 1.

Interpretation. R^2 close to 1 means the model explains most of the variation in Y. R^2 close to 0 means the model has almost no explanatory power.