

Regression Basics (chap. 1)

Main formulas obtained on Thursday:

$$\text{Sample regression slope} \quad b_1 = \frac{S_{XY}}{S_{XX}} = \frac{s_{xy}}{s_x^2}$$

$$\text{Sample regression intercept} \quad b_0 = \bar{Y} - b_1 \cdot \bar{X}$$

$$\begin{aligned} \text{where} \quad S_{XX} &= \sum_{i=1}^n (X_i - \bar{X})^2 \\ S_{XY} &= \sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y}) \\ s_x^2 &= \frac{S_{XX}}{n-1} \text{ is the sample variance of } X \\ s_{xy} &= \frac{S_{XY}}{n-1} \text{ is the sample covariance of } X \text{ and } Y \end{aligned}$$

$$\text{Predicted values} \quad \hat{Y}_i = b_0 + b_1 \cdot X_i$$

$$\text{Residuals} \quad e_i = Y_i - \hat{Y}_i$$

$$\text{Error sum of squares} \quad \sum_{i=1}^n e_i^2$$

$$\text{Sample variance} \quad s^2 = \frac{\sum e_i^2}{n-2}$$

$$\text{Sample standard deviation} \quad s = \sqrt{s^2}$$

All this material is in Chap. 1 of our textbook.

##1-3. These exercises require the very basic understanding of linear regression, its meaning, and assumptions.

4-5. Instead of a sample covariance s_{xy} , these problems give us a correlation coefficient

$$r = \frac{s_{xy}}{s_x s_y}.$$

We can certainly use it to obtain the sample covariance $s_{xy} = r s_x s_y$ and then compute the slope and intercept from formulas above. But also, there is a shortcut, once we compare equations for the correlation coefficient r and the slope b_1 ,

$$r = \frac{s_{xy}}{s_x s_y} \quad \implies \quad b_1 = \frac{s_{xy}}{s_x^2} = r \frac{s_y}{s_x}.$$

This is how the slope b_1 depends on the correlation coefficient r .

6. To show unbiasedness, as always, we need to find $\mathbf{E}(b_0)$ and demonstrate that it equals β_0 .

7. All the needed R commands are in R lab #3 on Blackboard. SAS commands are also on Blackboard, in the folder "SAS help". We did a very similar example with Copiers data in class.