"You Do It" Solution

Lecture 10: Simple Linear Regression

Prepared as a classroom-ready detailed solution set for ECOM5005

Exercise 1: Prediction and Residuals from a Regression Equation

Question

A regression between foot length (response variable in cm) and height (explanatory variable in inches) for 33 students resulted in the following regression equation:

$$\widehat{y} = 10.9 + 0.23x.$$

(a) One student in the sample was 73 inches tall with a foot length of 29 cm. What is the predicted foot length for this student?
(b) One student in the sample was 73 inches tall with a foot length of 29 cm. What is the residual for this student?
(c) What is the estimated average foot length for students who are 70 inches tall?

Detailed Solution

Step 1: Identify the variables and the fitted equation.
The response variable is foot length in centimetres, Y, and the explanatory variable is height in inches, X. The fitted regression equation is

$$\widehat{y} = 10.9 + 0.23x,$$

where $\widehat{y}$ is the foot length predicted by the regression line.
Step 2: Part (a) - predict the foot length for a 73-inch student.
Substitute $x = 73$:

$$\widehat{y} = 10.9 + 0.23(73) = 10.9 + 16.79 = \boxed{27.69 \text{ cm}}.$$

Thus, a 73-inch student is predicted to have a foot length of approximately **27.69 cm**.
Step 3: Part (b) - calculate the residual.
A residual is the observed value minus the predicted value:

$$e = y - \widehat{y}.$$

Here $y = 29$ cm and $\widehat{y} = 27.69$ cm, so

$$e = 29 - 27.69 = \boxed{1.31 \text{ cm}}.$$

The residual is **positive**; therefore, this student's actual foot length is 1.31 cm larger than predicted. The model under-predicts this observation by 1.31 cm.
Step 4: Part (c) - estimate the average foot length for 70-inch students.
Substitute $x = 70$:

$$\widehat{y} = 10.9 + 0.23(70) = 10.9 + 16.10 = \boxed{27.00 \text{ cm}}.$$

Therefore, the estimated average foot length for students who are 70 inches tall is **27.00 cm**. The phrase *estimated average* refers to the mean response predicted by the regression line at $x = 70$; individual 70-inch students may lie above or below that fitted value.

Exercise 1: Interpretation and Final Answer

Reading the Regression Equation

The fitted equation

$$\widehat{y} = 10.9 + 0.23x$$

contains two estimated coefficients:

- The **slope** is $b_1 = 0.23$. For each additional one inch of height, the model predicts that average foot length increases by approximately **0.23 cm**.
- The **intercept** is $b_0 = 10.9$. Algebraically, it is the predicted foot length when height is zero inches. Such an interpretation is not practically useful here because zero inches is not a meaningful student height and lies far outside the relevant data range.

Observed Value, Predicted Value, and Residual

For the 73-inch student:

$$\text{Observed foot length} = 29.00 \text{ cm},$$

$$\text{Predicted foot length} = 27.69 \text{ cm},$$

so

$$\text{Residual} = 29.00 - 27.69 = 1.31 \text{ cm}.$$

The residual is the vertical difference between the observed point and the fitted regression line. A **positive** residual means that the observed response lies above the regression line; a **negative** residual would mean that it lies below the line.

Answer Summary

(a) Predicted foot length for a 73-inch student:

$$\boxed{27.69 \text{ cm}}.$$

(b) Residual for the student whose observed foot length is 29 cm:

$$\boxed{29 - 27.69 = 1.31 \text{ cm}}.$$

The positive residual means the actual foot length is 1.31 cm greater than predicted.
(c) Estimated average foot length for 70-inch students:

$$\boxed{27.00 \text{ cm}}.$$

Final Check

For prediction questions in simple linear regression, remember:

- substitute the specified x value into $\widehat{y} = b_0 + b_1 x$ to obtain the predicted response;
- calculate a residual as $e = y - \widehat{y}$, not $\widehat{y} - y$;
- a positive residual means the observed Y is above the fitted line;
- interpret the slope in the units of Y *per one unit of* X; here, cm of foot length per inch of height;
- distinguish an **individual observed response** from the regression line's **estimated average response** at a given value of X.