Chapra [1] Example 14.6

James E. Stine and Marcus Mellor
Electrical and Computer Engineering Department
Oklahoma State University
Stillwater, OK 74078, USA

Introduction

In regression analysis, it is important to distinguish between the observed values y_i, the fitted values $\hat{y}_i$, and the residuals $e_i = y_i - \hat{y}_i$. The total unexplained variation in the data is captured by the error sum of squares:

$$SS_E = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2.$$

Fitted Values

For simple linear regression, the fitted values are determined from the least–squares regression line:

$$\hat{y}_i = b_0 + b_1 \cdot x_i,$$

with slope and intercept given by

$$a_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}, \qquad a_0 = \bar{y} - a_1 \cdot \bar{x}.$$

Example with Salary Data

Given the dataset

$$(x, y) = \{(10, 25), (20, 70), (30, 380), (40, 550), (50, 610), (60, 1220), (70, 830), (80, 1450)\},$$

the sample means are

$$\bar{x} = 45, \qquad \bar{y} = 641.875.$$

The relevant sums are

$$\sum (x_i - \bar{x})^2 = 4{,}200, \qquad \sum (x_i - \bar{x})(y_i - \bar{y}) = 81{,}775.$$

Hence

$$a_1 = \frac{81{,}775}{4200} = 19.47, \qquad a_0 = 641.875 - (19.47)(45) = -234.29.$$

The fitted regression line is therefore

$$\hat{y} = -234.29 + 19.47 \cdot x.$$

Residuals

The residual for observation i is

$$e_i = y_i - \hat{y}_i.$$

Example calculations:

$$\hat{y}_1 = -234.29 + 19.47(10) = -39.58, \quad e_1 = 25 - (-39.58) = 64.58,$$

$$\hat{y}_2 = -234.29 + 19.47(20) = 155.12, \quad e_2 = 70 - 155.12 = -85.12.$$

Each squared residual e_i^2 contributes to the error sum of squares:

$$SS_E = \sum_{i=1}^{n} e_i^2.$$

Explained and Total Variation

To connect residuals to the decomposition of sums of squares, consider:

$$(y_i - \bar{y})^2 \quad \text{measures total deviation from the mean,}$$

$$(\hat{y}_i - \bar{y})^2 \quad \text{measures the portion explained by the regression.}$$

Example calculations:

$$(y_1 - \bar{y})^2 = (25 - 641.875)^2 = 380{,}534.8,$$

$$(\hat{y}_1 - \bar{y})^2 = (-39.58 - 641.875)^2 = 464{,}385.5.$$

Thus, each observation contributes to SS_T and SS_R, while e_i^2 contributes to SS_E. The relationship

$$SS_T = SS_R + SS_E$$

always holds when the regression model includes an intercept.

Conclusion

By computing fitted values $\hat{y}_i$, residuals e_i, and their squares e_i^2, we quantify the error sum of squares SS_E, which measures the unexplained variation in the regression model.

References

[1] S. C. Chapra and D. Clough, *Applied Numerical Methods with Python for Engineers and Scientists*, 1st ed. McGraw-Hill Education, 2023.