Section 10.2 - Regression

Given a collection of paired sample data, the **regression equation**

$$\hat{y} = ax + b$$

algebraically describes the relationship between the two variables x and y. The graph of the regression equation is called the **regression line**.

The regression line is the line that "best" fits the data in the least-squares sense.

The values of $\hat{y}$ predicted by the regression equation and corresponding to the values of x probably do not equal the actual y data values.

For a pair of sample x- and y-values, the **residual** is the difference between the observed sample value of y and the y-value predicted by the regression equation.

$$\text{residual} = \text{sample } y \text{ - predicted } y = y - \hat{y}$$

The regression line is the line that best fits the data in the sense that the sum of the squares of the residuals is a minimum, i.e. $\sum(y - \hat{y})^2$ is a minimum.