

Introduction to the Least-Squares Regression Line

The goal in regression is often to describe a relationship between a **response** variable Y and an **explanatory** variable X and/or to predict Y for a given value of X .

The response variable is often referred to as the **dependent** variable because we can imagine that the value of Y will *depend* on the corresponding value of X .

We call X the explanatory variable because we think of X as a variable that *explains* variation in Y values. The explanatory X is often called the **independent** variable.

For any scatterplot exhibiting the relationship between Y and X , there are infinitely many lines that could be drawn to summarize the relationship between Y and X . The least-squares regression line is the “best” line in the sense that it makes the sum of the squares of the vertical distances from the data points to the line as small as possible.

The equation of the least-squares regression line is

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X \quad \text{where} \quad \hat{\beta}_1 = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{(n-1)s_X^2} \quad \text{and} \quad \hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}.$$

Alternatively, the equation of the least-squares regression line is

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X \quad \text{where} \quad \hat{\beta}_1 = \frac{r \cdot S_Y}{S_X} \quad \text{and} \quad \hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}.$$

The latter equation shows that we expect an increase of one standard deviation on the X scale to be accompanied by an increase of r standard deviations on the Y scale.

1. Examine the scatter plot on the back of this sheet that illustrates the relationship between the latitude (X) of 56 U.S. cities and their average minimum January temperature (Y) in degrees Fahrenheit. Select the appropriate sample correlation coefficient from the following list: -0.848 -0.124 0 0.124 0.848.
2. The mean latitude of the 56 cities is 39.0 degrees north of the equator. The mean of the 56 average minimum January temperatures is 26.5 degrees Fahrenheit. The standard deviations of the latitudes and temperatures are 5.4 and 13.4, respectively. Find the equation of the least-squares regression line for predicting temperature from latitude. Plot this line on the scatter plot.
3. Predict the average minimum January temperature for Ames, Iowa, which is located 42.0 degrees north of the equator.
4. Predict the average minimum January temperature at the equator. Does this seem reasonable? Explain.