




8 — Regression

8.1 Linear Regression

Definition 8.1 — Regression Equation. The linear regression equation $\hat{y} = ax + b$ describes the linear equation that represents the "line of best fit". This line attempts to pass through as many of the points as possible. a is the slope of our linear regression equation and represents the rate of change in y versus x . b represents the y -intercept.

 The regression equation may also be written as $\hat{y} = bx + a$

The line of best first helps describe the dataset. It can also be used to make approximate predictions of how the data will behave.

Corollary 8.1 We can find the linear regression equation with the two following pieces of information:

$$\text{slope} = r \frac{s_y}{s_x}$$

The regression equation passes through the point (\bar{x}, \bar{y})

■ **Example 8.1** ■

8.2 Practice Problems

Problem 8.1 Marcus wants to investigate the relationship between hours of computer usage per day and number of minutes of migraines endured per day. After collecting data, He finds a correlation coefficient of 0.86, with $s_y = 375.55$ and $s_x = 1814.72$. The mean hours of computer usage from his data set was calculated to be 4.5 hours and the average number of minutes of migraine was calculated to be 25 minutes. Find the regression line that best fits his data.

Problem 8.2 Using the line that you calculated above, given 2 hours of computer usage, how many minutes of migraine would Marcus predict to follow?

Problem 8.3 Marcus coincidentally has a point in his data set that he collected for exactly 2 hours of computer usage. Given that the residual between his observed value for 2 hours of computer usage and the expected value (as calculated in the previous question) equals 1.89, how many minutes of migraine did Marcus observe for that point in his data set?