Linear Regression Example

Trending vehicle miles-per-gallon versus vehicle weight

Data provided:

$mpg := (35\ \ 30\ \ 28\ \ 22\ \ 18\ \ 17\ \ 15)^T$ $\qquad y := mpg$

$weight := (2000\ \ 2200\ \ 2600\ \ 2900\ \ 3400\ \ 3800\ \ 4100)^T$ $\qquad x := weight$

$n := length(mpg)$ $\qquad n = 7$ $\qquad i := 0 .. n - 1$

Form of regression equation:

$$y = a_0 + a_1 \cdot x$$

Using regression equations:

$x\_bar := \frac{\sum_i x_i}{n}$ $\qquad x\_bar = 3 \times 10^3$ $\qquad y\_bar := \frac{\sum_i y_i}{n}$ $\qquad y\_bar = 23.571$

$$a_1 := \frac{n \cdot \sum_i x_i \cdot y_i - \sum_i x_i \cdot \sum_i y_i}{n \cdot \sum_i (x_i)^2 - \left(\sum_i x_i\right)^2} \qquad a_1 = -9.188 \times 10^{-3}$$

$a_0 := y\_bar - a_1 \cdot x\_bar$ $\qquad a_0 = 51.137$

Using MathCAD's built-in functions:

$a_1 := slope(x, y)$ $\qquad a_1 = -9.188 \times 10^{-3}$

$a_0 := intercept(x, y)$ $\qquad a_0 = 51.137$

Computing the standard error and the correlation coefficient:

$$S_t := \sum_i \left(y_i - y\_bar\right)^2 \qquad S_t = 341.714$$

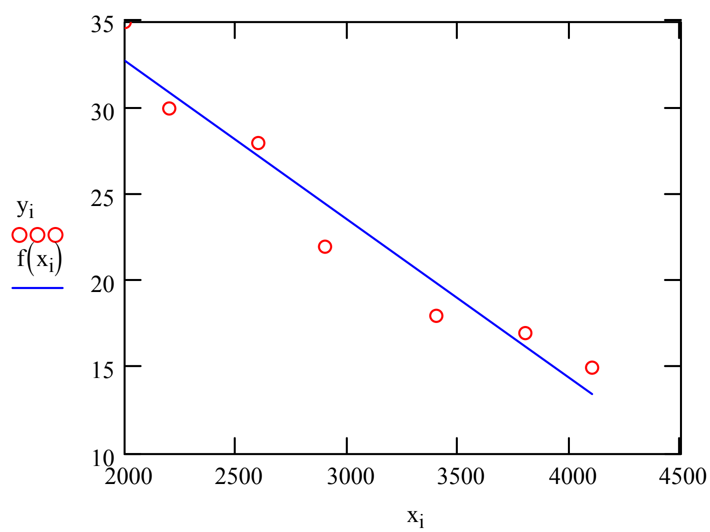
$$S_r := \sum_i \left[y_i - \left(a_0 + a_1 \cdot x_i\right)\right]^2 \qquad S_r = 19.199$$

$$s_{y\_x} := \sqrt{\frac{S_r}{n-2}} \qquad s_{y\_x} = 1.96$$

$$r := \sqrt{\frac{S_t - S_r}{S_t}} \qquad r = 0.972$$

Plotting the original data and the regression line:

$$f(x) := a_0 + a_1 \cdot x$$



Now we can estimate a vehicles mpg, given its weight. For example:

$$weight := 3100$$

$$mpg := f(weight) \qquad mpg = 22.653$$