Linear Regression

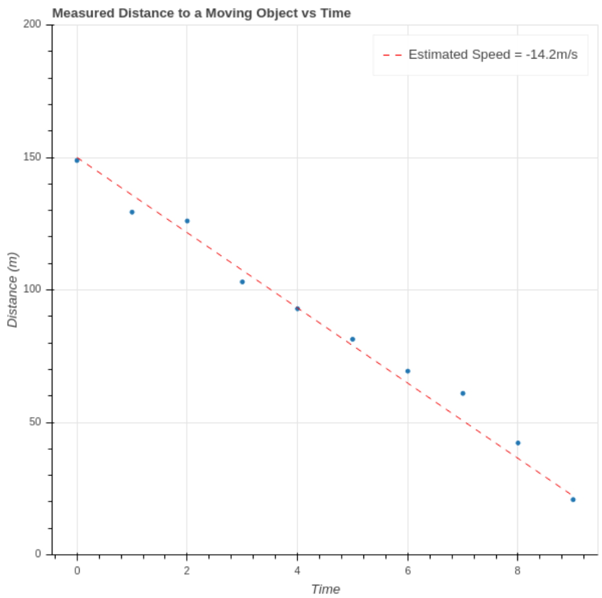
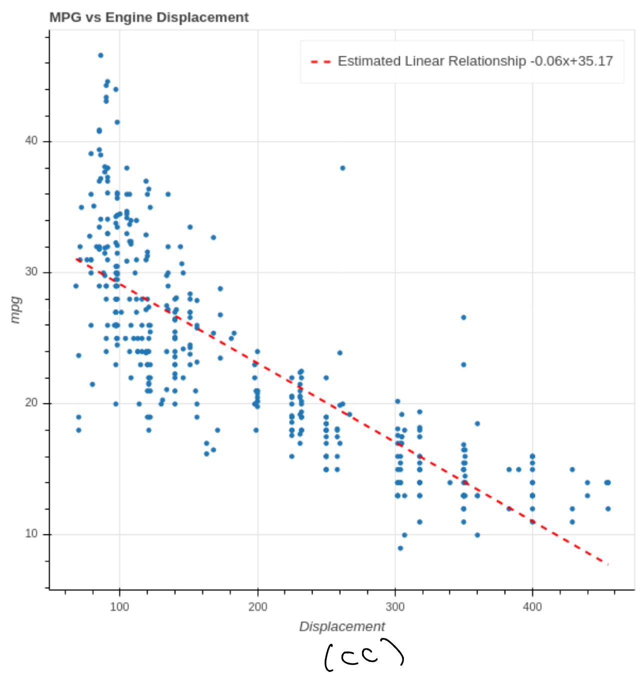


Figure 1: Physics Experiment


MPG vs Engine Displacement
Estimated Linear Relationship -0.06x+35.17
mpg
Displacement
(cc)
10
20
30
40
100
200
300
400

Linear Regression

Least Squares (Ordinary Least Squares or OLS)

Given data points $\{(x_i, y_i)\}_{i=1}^{N}$

Find slope m and intercept b so that

$$y_i \approx m x_i + b \quad \text{for } i = 1, \ldots, N$$

Mean squared error

$$MSE(m, b) = \frac{1}{N} \sum_{i=1}^{N} (y_i - m x_i - b)^2$$

"Best" choice of m and b are the values that minimize the MSE.

Use Calc to find them. Let $E = MSE(m, b)$

Solve $\frac{\partial E}{\partial m} = 0$ $\quad \frac{\partial E}{\partial b} = 0$ to find the minimum.

$$MSE = \frac{1}{N} \sum_{i=1}^{N} (y_i - m x_i - b)^2$$

a quadratic f^n of m and b

$$MSE = Am^2 + Bmb + Cb^2 + Dm + Eb + F$$

$$\frac{\partial E}{\partial m}$$

$$E = \frac{1}{N}\sum (y_i - m x_i - b)^2$$

$$\frac{\partial E}{\partial m} = \frac{1}{N}\sum_{i=1}^{N} 2(y_i - m x_i - b)[-x_i]$$

$$= -2 \cdot \frac{1}{N}\sum_{i=1}^{N} (y_i x_i - m x_i^2 - b x_i)$$

$$= -2 \cdot \left[\frac{1}{N}\sum y_i x_i - m\left(\frac{1}{N}\sum x_i^2\right) - b\left(\frac{1}{N}\sum x_i\right)\right]$$

$$\frac{1}{N}\sum x_i y_i = S_{xy} \qquad \frac{1}{N}\sum x_i^2 = S_{xx} \qquad \frac{1}{N}\sum x_i = \overline{x}$$

$$\frac{\partial E}{\partial m} = -2\left[S_{xy} - m S_{xx} - b\overline{x}\right] = 0$$

$$\frac{\partial E}{\partial b}$$

$$E = \frac{1}{N}\sum_{1}^{N} (y_i - m x_i - b)^2$$

$$\frac{\partial E}{\partial b} = \frac{1}{N}\sum^{N} 2(y_i - m x_i - b)[-1]$$

$$= \frac{-2}{N}\left[\sum y_i - m\sum x_i - b\sum_{i=1}^{N} 1\right]$$

$$= -2\left[\frac{1}{N}\sum y_i - m\left(\frac{1}{N}\sum x_i\right) - b\right]$$

$$\overline{y} = \frac{1}{N}\sum y_i$$

$$\frac{\partial E}{\partial b} = -2\left[\overline{y} - m\overline{x} - b\right] = 0$$

$$\bar{y} = m\bar{x} + b$$

$$S_{xy} = m S_{xx} + b\bar{x}$$

$$\bar{x}\bar{y} = m\bar{x}^2 + b\bar{x}$$

$$S_{xy} - \bar{x}\bar{y} = m(S_{xx} - \bar{x}^2)$$

$$m = \frac{S_{xy} - \bar{x}\bar{y}}{S_{xx} - \bar{x}^2}$$

$$\bar{y} = m\bar{x} + b$$

$$S_{xy} = m S_{xx} + b\bar{x}$$

$$\bar{x} S_{xy} = m\bar{x} S_{xx} + b\bar{x}^2$$

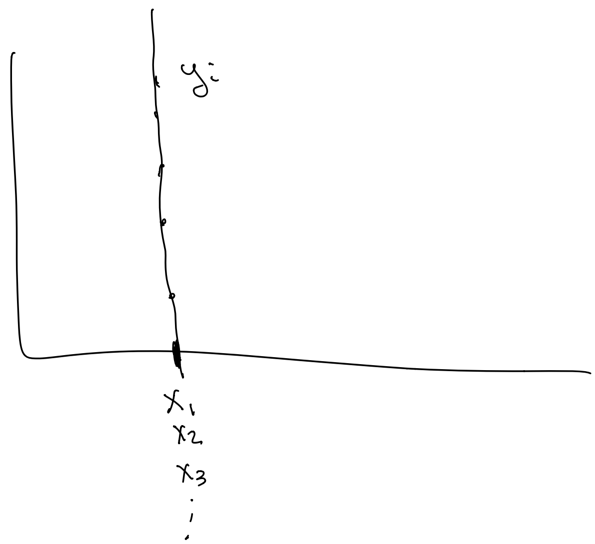
$$S_{xx}\bar{y} = m\bar{x} S_{xx} + b S_{xx}$$

$$\bar{x} S_{xy} - \bar{y} S_{xx} = b(\bar{x}^2 - S_{xx})$$

$$b = \frac{\bar{y} S_{xx} - \bar{x} S_{xy}}{S_{xx} - \bar{x}^2}$$

$S_{xx} \neq \bar{x}^2$!!

$$\frac{1}{N}\sum x_i^2 \neq \left[\frac{1}{N}\sum x_i\right]^2$$



y_i
x_1
x_2
x_3

$$N \sum x_i^2 = \left(\sum x_i\right)^2$$ if and only if all x_i are equal this time.

$X = (x_1, \ldots, x_N)$

$E = (1, \ldots, 1)$

$\|X\|^2 = \sum x_i^2 \qquad \|E\|^2 = N$

$\left(\sum x_i\right)^2 = |X \cdot E|^2$

Cauchy-Swartz:

$|a \cdot b| \le \|a\| \, \|b\|$

with $=$ only if $a = \lambda b$ for some λ.

$|X \cdot E|^2 \le \|X\|^2 \|E\|^2$

$=$ only if $X = \lambda E$

$(x_1, \ldots, x_N) = \lambda(1, 1, \ldots, 1)$

$\boxed{x_i = \lambda}$