

$r=0$



$r=0.5$



$r=0.7$



$r=0.9$



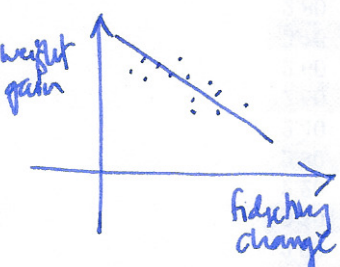
$r=0.99$

note correlation is same if you swap x, y coordinates
correlation $\nrightarrow$ causation.

§2.4 Least squares regression (best fit straight line / linear regression)

example (2.19)

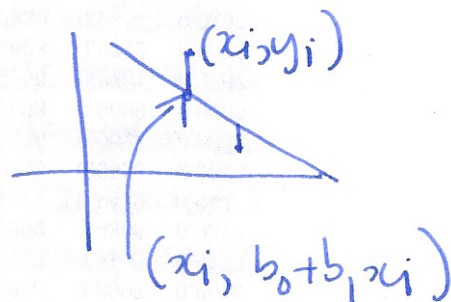
change calories used in fidgeting	-94	-57	-29	135		690
weight gain	4.2	3.0	3.7	2.7		1.1



Q: how do we find a best fit straight line?

A: here's one way. data points (x_i, y_i) .

consider same line $y = b_0 + b_1 x$.



choose the line which
minimize the squares of the vertical errors.
this is called the least-squares regression line
(not symmetric in (x, y) !)

data (x_1, y_1) (x_2, y_2) ... (x_n, y_n)

$$\sum (\text{error})^2 = \sum_{i=1}^n (y_i - b_0 - b_1 x_i)^2$$

Fact the values of b_0, b_1 which minimize this is

(in example 2.19 $b_1 = -0.0034$ $b_0 = 2.388$).

$$b_1 = r \frac{s_y}{s_x}$$

$$b_0 = \bar{y} - b_1 \bar{x}$$

where $r =$ correlation coefficient!

warning: extrapolation not always useful beyond size of data.

