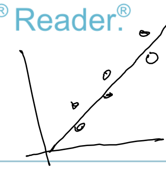
Section 3.1

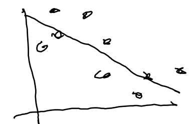
PART III

Linear Correlation



positive

$r \approx 1$



negative

$r \approx -1$

→ Meaning behind r.

→ Fact: In the numerator of the computational formula for r is the expression

$$\sum xy - \frac{1}{n}(\sum x)(\sum y).$$

This is actually algebraically equivalent to:

$$\sum xy - \frac{1}{n}(\sum x)(\sum y) = \sum(\underbrace{x-\bar{x}}_{\text{deviation of } x})(\underbrace{y-\bar{y}}_{\text{deviation of } y})$$

note also that the square roots in the denominator are closely related to the standard deviations, s_x & s_y, of x & y respectively.

$$s_x = \sqrt{\frac{\sum x^2 - \frac{1}{n}(\sum x)^2}{n-1}} \quad \text{and} \quad s_y = \sqrt{\frac{\sum y^2 - \frac{1}{n}(\sum y)^2}{n-1}}.$$

So r can be rewritten as follows:

$$r = \frac{\sum xy - \frac{1}{n}(\sum x)\cdot(\sum y)}{\sqrt{\sum x^2 - \frac{1}{n}(\sum x)^2}\cdot\sqrt{\sum y^2 - \frac{1}{n}(\sum y)^2}} = \frac{\sum(x-\bar{x})(y-\bar{y})}{(n-1)\underbrace{\sqrt{\frac{\sum x^2 - \frac{1}{n}(\sum x)^2}{n-1}}}_{s_x}\underbrace{\sqrt{\frac{\sum y^2 - \frac{1}{n}(\sum y)^2}{n-1}}}_{s_y}}$$

$$= \frac{\sum(x-\bar{x})(y-\bar{y})}{(n-1)s_x\cdot s_y} = \frac{\sum\overbrace{\left(\frac{x-\bar{x}}{s_x}\right)}^{z_x}\overbrace{\left(\frac{y-\bar{y}}{s_y}\right)}^{z_y}}{n-1}$$

$$= \frac{\sum z_x\cdot z_y}{n-1} = \boxed{\text{average of } z_x \cdot z_y \text{ values.}}$$

$$r = \frac{\sum z_x\cdot z_y}{n-1}$$