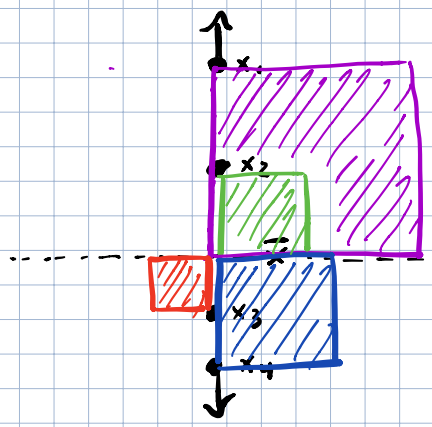
Variance

- Suppose I have a sample of $x_1, x_2, \ldots, x_n$ values.
- What is sample variance? $s^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}$



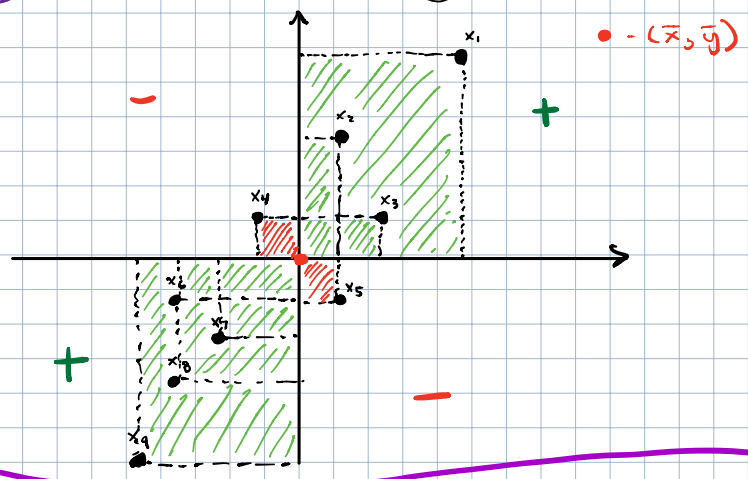
$$\sum_{i=1}^{n} (x_i - \bar{x})^2$$

Covariance - Cov(X, Y)

Tells us direction of relationship between X and Y.

$\begin{cases} X - x_1, x_2, x_3, \ldots, x_n & \text{can find } \bar{x} \\ Y - y_1, y_2, y_3, \ldots, y_n & \text{can find } \bar{y} \end{cases}$

formula: $\frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{(n-1)}$



Correlation: Tells us both direction and strength of a linear relationship between two variables.

ρ - population correlation coefficent

r - sample correlation coefficient

$$r = \frac{1}{n-1} \sum \left(\frac{x_i - \bar{x}}{s_x} \right) \left(\frac{y_i - \bar{y}}{s_y} \right)$$

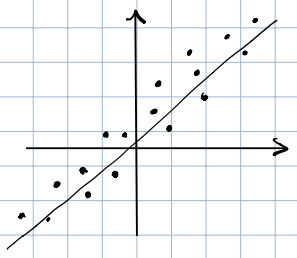
what AP gives you.

$$= \frac{Cov(X, Y)}{s_x \, s_y}$$

Correlation and Regression.

- Suppose I have a sample of n ordered pairs (x_i, y_i) for $i = 1, 2, 3, \ldots, n$.
- how might I visualize a relationship between the variables X and Y?

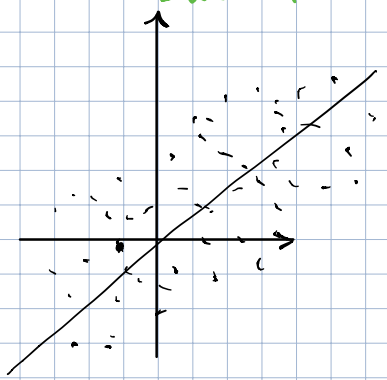
- Strong positive



Scatter-Plot - 1st step in any regression analysis.

- r, the correlation coefficient tells us the strength and direction of a linear relationship.

- weak positive



- no relationship.