

## → Sample Covariance

Definition: Measure of how two variables are linearly related.

- Used to describe the degree to which 2 variables are related.
- Denoted by  $\text{cov}(x,y)$  or  $S_{xy}$

$$\text{cov}(x, y) = s_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{(n - 1)}$$

- »  $\text{Cov}(x,y)$  = sample covariance
- »  $n$  = number of observations
- »  $\bar{X}$  = Sample mean of X
- »  $\bar{Y}$  = Sample mean of Y

- If  $x$  and  $y$  are **positively related**, then small observations of  $x$  will tend to be associated with small observations of  $y$ .
- If  $x$  and  $y$  are **negatively related**, then small observations of  $x$  will tend to be associated with large observations of  $y$ .
  
- If the covariance is a **large positive number**, the two variables move in the same direction. It is positively linearly related.
- If the covariance is a **large negative number**, the two variables move in opposite directions. It is negatively linearly related.
- If the covariance is **close to 0**, then the two variables are unrelated.

**NOTE\*:** Population covariance is the same formula **EXCEPT** we divide by the whole population ( $N$ ) rather than  $(n-1)$

## → Coefficient of Correlation

Definition: A measurement of the strength of the relationship between two variables where one or both are ordinal.

- Measures the degree to which 2 variables are linearly related.
- Defined as sample covariance divided by the standard deviations of the variables.
- Denoted by the letter r

### Why we use coefficient of correlation:

Consider 2 sets of variables (x,y) and (p,q) and each variable is measured according to different units. It is hard to compare the degree to which it is related.

Coefficient of correlation allows us to compare the two sets of variables in the same unit by removing the scale.

$$r = \frac{S_{xy}}{S_x S_y}$$

» r = coefficient of correlation

»  $S_{xy}$  = Sample covariance

»  $S_x$  = Standard deviation of X

»  $S_y$  = Standard deviation of Y

WHERE:  $-1 \leq r \leq 1$ .

- A coefficient close to 1 means the closer the relationship is to being described by a straight line sloping upwards from left to right.
- A coefficient close to +1 indicates 2 variables are very strongly positively related.
- A coefficient close to -1 indicates 2 variables are strongly negatively related.
- A correlation close to 0 indicates that no straight line relationship exists. It may mean no pattern.
- A coefficient equal to positive or negative 1 suggests there exists a perfect linear relationship between x and y.

### Example:

A coefficient of correlation equal to -0.4 suggests that the relationship is negatively related. Also, since it is closer to 0, there is a weak linear relationship.