

- Covariance is a measure of the joint variability of two random variables.
- It is used to indicate if the variables tend to show similar/opposite behavior
- Covariance is also used when calculating variance of sum of R.V.s

Recall:

$$\begin{aligned} & \mathbf{Var}[X+Y] \\ = & \mathbf{E}\{[(X+Y)-(\mu_X+\mu_Y)]^2\} \\ = & \mathbf{E}[(X+Y)^2] - 2\mathbf{E}[(X+Y)(\mu_X+\mu_Y)] + (\mu_X+\mu_Y)^2 \\ = & \{\mathbf{E}[X^2]-\mu_X^2\} + \{\mathbf{E}[Y^2]-\mu_Y^2\} + \{2\mathbf{E}[XY]-2\mu_X\mu_Y\} \\ = & \mathbf{Var}[X] + \mathbf{Var}[Y] + 2(\mathbf{E}[XY]-\mu_X\mu_Y) \end{aligned}$$

Define:

$$\mathbf{Cov}[X, Y] = \mathbf{E}[XY] - \mu_X \mu_Y$$

$$Var[X + Y] = Var[X] + Var[Y] + 2Cov[X, Y]$$

Recall:

$$\begin{aligned} & \mathbf{Var}[X+Y] \\ = & \mathbf{E}\{[(X+Y)-(\mu_X+\mu_Y)]^2\} \\ = & \mathbf{E}[(X+Y)^2] - 2\mathbf{E}[(X+Y)(\mu_X+\mu_Y)] + (\mu_X+\mu_Y)^2 \\ = & \{\mathbf{E}[X^2]-\mu_X^2\} + \{\mathbf{E}[Y^2]-\mu_Y^2\} + \{2\mathbf{E}[XY]-2\mu_X\mu_Y\} \\ = & \mathbf{Var}[X] + \mathbf{Var}[Y] + 2(\mathbf{E}[XY]-\mu_X\mu_Y) \end{aligned}$$

Define:

$$\mathbf{Cov}[X,Y] = \mathbf{E}[XY] - \mu_X \mu_Y$$

$$Var[X + Y] = Var[X] + Var[Y] + 2Cov[X, Y]$$

Recall:

$$\begin{aligned} & \operatorname{Var}[X + Y] \\ = & \operatorname{E}\{[(X + Y) - (\mu_X + \mu_Y)]^2\} \\ = & \operatorname{E}[(X + Y)^2] - 2\operatorname{E}[(X + Y)(\mu_X + \mu_Y)] + (\mu_X + \mu_Y)^2 \\ = & \left\{\operatorname{E}[X^2] - \mu_X^2\right\} + \left\{\operatorname{E}[Y^2] - \mu_Y^2\right\} + \left\{2\operatorname{E}[XY] - 2\mu_X\mu_Y\right\} \\ = & \operatorname{Var}[X] + \operatorname{Var}[Y] + 2(\operatorname{E}[XY] - \mu_X\mu_Y) \end{aligned}$$

Define:

$$\mathbf{Cov}[X,Y] = \mathbf{E}[XY] - \mu_X \mu_Y$$

$$Var[X + Y] = Var[X] + Var[Y] + 2Cov[X, Y]$$

Recall:

$$\begin{aligned} & \mathbf{Var}[X+Y] \\ = & \mathbf{E}\{[(X+Y)-(\mu_X+\mu_Y)]^2\} \\ = & \mathbf{E}[(X+Y)^2] - 2\mathbf{E}[(X+Y)(\mu_X+\mu_Y)] + (\mu_X+\mu_Y)^2 \\ = & \{\mathbf{E}[X^2]-\mu_X^2\} + \{\mathbf{E}[Y^2]-\mu_Y^2\} + \{2\mathbf{E}[XY]-2\mu_X\mu_Y\} \\ = & \mathbf{Var}[X] + \mathbf{Var}[Y] + 2(\mathbf{E}[XY]-\mu_X\mu_Y) \end{aligned}$$

Define:

$$\mathbf{Cov}[X,Y] = \mathbf{E}[XY] - \mu_X \mu_Y$$

$$Var[X + Y] = Var[X] + Var[Y] + 2Cov[X, Y]$$

Recall:

$$\begin{aligned} & \mathbf{Var}[X+Y] \\ = & \mathbf{E}\{[(X+Y)-(\mu_X+\mu_Y)]^2\} \\ = & \mathbf{E}[(X+Y)^2] - 2\mathbf{E}[(X+Y)(\mu_X+\mu_Y)] + (\mu_X+\mu_Y)^2 \\ = & \{\mathbf{E}[X^2]-\mu_X^2\} + \{\mathbf{E}[Y^2]-\mu_Y^2\} + \{2\mathbf{E}[XY]-2\mu_X\mu_Y\} \\ = & \mathbf{Var}[X] + \mathbf{Var}[Y] + 2(\mathbf{E}[XY]-\mu_X\mu_Y) \end{aligned}$$

Define:

$$\mathbf{Cov}[X,Y] = \mathbf{E}[XY] - \mu_X \mu_Y$$

$$Var[X + Y] = Var[X] + Var[Y] + 2Cov[X, Y]$$

Recall:

$$\begin{aligned} & \operatorname{Var}[X+Y] \\ = & \operatorname{E}\{[(X+Y) - (\mu_X + \mu_Y)]^2\} \\ = & \operatorname{E}[(X+Y)^2] - 2\operatorname{E}[(X+Y)(\mu_X + \mu_Y)] + (\mu_X + \mu_Y)^2 \\ = & \left\{\operatorname{E}[X^2] - \mu_X^2\right\} + \left\{\operatorname{E}[Y^2] - \mu_Y^2\right\} + \left\{2\operatorname{E}[XY] - 2\mu_X\mu_Y\right\} \\ = & \operatorname{Var}[X] + \operatorname{Var}[Y] + 2(\operatorname{E}[XY] - \mu_X\mu_Y) \end{aligned}$$

Define:

$$\mathbf{Cov}[X,Y] = \mathbf{E}[XY] - \mu_X \mu_Y$$

$$Var[X + Y] = Var[X] + Var[Y] + 2Cov[X, Y]$$



Recall:

$$\begin{aligned} & \mathbf{Var}[X+Y] \\ = & \mathbf{E}\{[(X+Y)-(\mu_X+\mu_Y)]^2\} \\ = & \mathbf{E}[(X+Y)^2] - 2\mathbf{E}[(X+Y)(\mu_X+\mu_Y)] + (\mu_X+\mu_Y)^2 \\ = & \{\mathbf{E}[X^2]-\mu_X^2\} + \{\mathbf{E}[Y^2]-\mu_Y^2\} + \{2\mathbf{E}[XY]-2\mu_X\mu_Y\} \\ = & \mathbf{Var}[X] + \mathbf{Var}[Y] + 2(\mathbf{E}[XY]-\mu_X\mu_Y) \end{aligned}$$

Define:

$$\mathbf{Cov}[X,Y] = \mathbf{E}[XY] - \mu_X \mu_Y$$

$$Var[X + Y] = Var[X] + Var[Y] + 2Cov[X, Y]$$

Another definition:

$$Cov[X, Y] = E[(X - \mu_X)(Y - \mu_Y)]$$

This is also called (1,1)th central moment

$$\begin{aligned} & \mathbf{Cov}[X,Y] \\ = & \mathbf{E}[(X-\mu_X)(Y-\mu_Y)] \\ = & \mathbf{E}[XY-X\mu_Y-Y\mu_X+\mu_X\mu_Y] \\ = & \mathbf{E}[XY]-\mu_X\mu_Y \end{aligned}$$

Another definition:

$$Cov[X, Y] = E[(X - \mu_X)(Y - \mu_Y)]$$

This is also called (1,1)th central moment

$$\begin{aligned} & \textbf{Cov}[X, Y] \\ = & \textbf{E}[(X - \mu_X)(Y - \mu_Y)] \\ = & \textbf{E}[XY - X\mu_Y - Y\mu_X + \mu_X\mu_Y] \\ = & \textbf{E}[XY] - \mu_X\mu_Y \end{aligned}$$

Another definition:

$$Cov[X, Y] = E[(X - \mu_X)(Y - \mu_Y)]$$

This is also called (1,1)th central moment

$$\begin{aligned} & \mathbf{Cov}[X, Y] \\ = & \mathbf{E}[(X - \mu_X)(Y - \mu_Y)] \\ = & \mathbf{E}[XY - X\mu_Y - Y\mu_X + \mu_X\mu_Y] \\ = & \mathbf{E}[XY] - \mu_X\mu_Y \end{aligned}$$

Another definition:

$$Cov[X, Y] = E[(X - \mu_X)(Y - \mu_Y)]$$

This is also called (1,1)th central moment

$$\begin{aligned} & \mathbf{Cov}[X, Y] \\ = & \mathbf{E}[(X - \mu_X)(Y - \mu_Y)] \\ = & \mathbf{E}[XY - X\mu_Y - Y\mu_X + \mu_X\mu_Y] \\ = & \mathbf{E}[XY] - \mu_X\mu_Y \end{aligned}$$

Another definition:

$$Cov[X, Y] = E[(X - \mu_X)(Y - \mu_Y)]$$

This is also called (1,1)th central moment

$$\begin{aligned} & \mathbf{Cov}[X,Y] \\ = & \mathbf{E}[(X-\mu_X)(Y-\mu_Y)] \\ = & \mathbf{E}[XY-X\mu_Y-Y\mu_X+\mu_X\mu_Y] \\ = & \mathbf{E}[XY]-\mu_X\mu_Y \end{aligned}$$

Another definition:

$$Cov[X, Y] = E[(X - \mu_X)(Y - \mu_Y)]$$

This is also called (1,1)th central moment

$$\begin{aligned} & \mathbf{Cov}[X, Y] \\ = & \mathbf{E}[(X - \mu_X)(Y - \mu_Y)] \\ = & \mathbf{E}[XY - X\mu_Y - Y\mu_X + \mu_X\mu_Y] \\ = & \mathbf{E}[XY] - \mu_X\mu_Y \end{aligned}$$

Another definition:

$$Cov[X, Y] = E[(X - \mu_X)(Y - \mu_Y)]$$

This is also called (1,1)th central moment

$$\begin{aligned} & \mathbf{Cov}[X, Y] \\ = & \mathbf{E}[(X - \mu_X)(Y - \mu_Y)] \\ = & \mathbf{E}[XY - X\mu_Y - Y\mu_X + \mu_X\mu_Y] \\ = & \mathbf{E}[XY] - \mu_X\mu_Y \end{aligned}$$

- ▶ Cov[X, a] = 0
- $ightharpoonup \operatorname{Cov}[X,X] = \operatorname{Var}[X]$
- $ightharpoonup \mathbf{Cov}[X,Y] = \mathbf{Cov}[Y,X]$
- ▶ if Cov[X, Y] = 0, X and Y are orthogonal
- ▶ When $\mu_X = 0$ or $\mu_X = 0$, $\mathbf{Cov}[X, Y] = \mathbf{E}[XY]$
- When X, Y are s.i., then $\mathbf{E}[XY] = \mathbf{E}[X]\mathbf{E}[X] \Rightarrow \mathbf{Cov}[X, Y] = 0$

- ▶ Cov[X, a] = 0
- ightharpoonup Cov[X,X] = Var[X]
- $ightharpoonup \operatorname{Cov}[X,Y] = \operatorname{Cov}[Y,X]$
- ▶ if Cov[X, Y] = 0, X and Y are orthogonal
- ▶ When $\mu_X = 0$ or $\mu_X = 0$, $\mathbf{Cov}[X, Y] = \mathbf{E}[XY]$
- When X, Y are s.i., then $\mathbf{E}[XY] = \mathbf{E}[X]\mathbf{E}[X] \Rightarrow \mathbf{Cov}[X, Y] = 0$

- ▶ Cov[X, a] = 0
- ightharpoonup Cov[X,X] = Var[X]
- ▶ if Cov[X, Y] = 0, X and Y are orthogonal
- ▶ When $\mu_X = 0$ or $\mu_X = 0$, $\mathbf{Cov}[X, Y] = \mathbf{E}[XY]$
- When X, Y are s.i., then $\mathbf{E}[XY] = \mathbf{E}[X]\mathbf{E}[X] \Rightarrow \mathbf{Cov}[X, Y] = 0$

- ▶ Cov[X, a] = 0
- ightharpoonup Cov[X,X] = Var[X]
- $\blacktriangleright \ \mathbf{Cov}[X,Y] = \mathbf{Cov}[Y,X]$
- ▶ if Cov[X, Y] = 0, X and Y are orthogonal
- ▶ When $\mu_X = 0$ or $\mu_X = 0$, $\mathbf{Cov}[X, Y] = \mathbf{E}[XY]$
- When X, Y are s.i., then $\mathbf{E}[XY] = \mathbf{E}[X]\mathbf{E}[X] \Rightarrow \mathbf{Cov}[X, Y] = 0$

- ▶ Cov[X, a] = 0
- ightharpoonup Cov[X,X] = Var[X]
- $\blacktriangleright \ \mathbf{Cov}[X,Y] = \mathbf{Cov}[Y,X]$
- ▶ if Cov[X, Y] = 0, X and Y are orthogonal
- ▶ When $\mu_X = 0$ or $\mu_X = 0$, $\mathbf{Cov}[X, Y] = \mathbf{E}[XY]$
- When X, Y are s.i., then $\mathbf{E}[XY] = \mathbf{E}[X]\mathbf{E}[X] \Rightarrow \mathbf{Cov}[X, Y] = 0$

- ▶ Cov[X, a] = 0
- ightharpoonup Cov[X,X] = Var[X]
- $\blacktriangleright \ \mathbf{Cov}[X,Y] = \mathbf{Cov}[Y,X]$
- if Cov[X, Y] = 0, X and Y are orthogonal
- ▶ When $\mu_X = 0$ or $\mu_X = 0$, $\mathbf{Cov}[X, Y] = \mathbf{E}[XY]$
- When X, Y are s.i., then $\mathbf{E}[XY] = \mathbf{E}[X]\mathbf{E}[X] \Rightarrow \mathbf{Cov}[X, Y] = 0$

Covariance Matirx

For random vector



We can calculate the covariance matrix by intuitively generalize covariance into multiple dimensions

$$\boldsymbol{\Sigma} = \begin{bmatrix} \mathbf{Cov}[X_1, X_1] & \mathbf{Cov}[X_1, X_2] & \dots & \mathbf{Cov}[X_1, X_n] \\ \mathbf{Cov}[X_2, X_1] & \mathbf{Cov}[X_2, X_2] & \dots & \mathbf{Cov}[X_2, X_n] \\ \vdots & \ddots & & \vdots \\ \mathbf{Cov}[X_n, X_1] & \mathbf{Cov}[X_n, X_2] & \dots & \mathbf{Cov}[X_n, X_n] \end{bmatrix}$$

Covariance Matirx

For random vector

$$\begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{bmatrix}$$

We can calculate the covariance matrix by intuitively generalize covariance into multiple dimensions

$$\Sigma = \begin{bmatrix} \mathbf{Cov}[X_1, X_1] & \mathbf{Cov}[X_1, X_2] & \dots & \mathbf{Cov}[X_1, X_n] \\ \mathbf{Cov}[X_2, X_1] & \mathbf{Cov}[X_2, X_2] & \dots & \mathbf{Cov}[X_2, X_n] \\ \vdots & \ddots & & \vdots \\ \mathbf{Cov}[X_n, X_1] & \mathbf{Cov}[X_n, X_2] & \dots & \mathbf{Cov}[X_n, X_n] \end{bmatrix}$$

Covariance Matirx

For random vector

$$\begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{bmatrix}$$

We can calculate the covariance matrix by intuitively generalize covariance into multiple dimensions

$$\boldsymbol{\Sigma} = \begin{bmatrix} \mathbf{Cov}[X_1, X_1] & \mathbf{Cov}[X_1, X_2] & \dots & \mathbf{Cov}[X_1, X_n] \\ \mathbf{Cov}[X_2, X_1] & \mathbf{Cov}[X_2, X_2] & \dots & \mathbf{Cov}[X_2, X_n] \\ \vdots & \ddots & & \vdots \\ \mathbf{Cov}[X_n, X_1] & \mathbf{Cov}[X_n, X_2] & \dots & \mathbf{Cov}[X_n, X_n] \end{bmatrix}$$

- Covariance indicates how correlated are two R.V.
- ► But not comparable
- ▶ Cov[X, Y] = 1, Cov[Z, W] = 10, Z, W more correlated?

Define Pearson correlation coefficient

$$\rho_{XY} = \frac{\mathbf{Cov}[X,Y]}{\sigma_X \sigma_Y}$$

- ρ_{XY} is a measure of the dependence between X and Y
- ▶ $-1 \le \rho_{XY} \le 1$ (Without proof)
- $ho_{XY}=\pm 1$ if X and Y are linearly related
- ▶ If X and Y are s.i., they are uncorrelated, $\mathbf{Cov}[X, Y] = \rho_{XY} = 0$
- ▶ The converse is NOT TRUE

Define Pearson correlation coefficient

$$\rho_{XY} = \frac{\mathbf{Cov}[X, Y]}{\sigma_X \sigma_Y}$$

- ρ_{XY} is a measure of the dependence between X and Y
- ▶ $-1 \le \rho_{XY} \le 1$ (Without proof)
- $ho_{XY}=\pm 1$ if X and Y are linearly related
- If X and Y are s.i., they are uncorrelated, $\mathbf{Cov}[X, Y] = \rho_{XY} = 0$
- ▶ The converse is NOT TRUE

Define Pearson correlation coefficient

$$\rho_{XY} = \frac{\mathbf{Cov}[X, Y]}{\sigma_X \sigma_Y}$$

- $\triangleright \rho_{XY}$ is a measure of the dependence between X and Y
- ▶ $-1 \le \rho_{XY} \le 1$ (Without proof)
- $ho_{XY}=\pm 1$ if X and Y are linearly related
- If X and Y are s.i., they are uncorrelated $\mathbf{Cov}[X, Y] = \rho_{XY} = 0$
- ▶ The converse is NOT TRUE

Define Pearson correlation coefficient

$$\rho_{XY} = \frac{\mathbf{Cov}[X, Y]}{\sigma_X \sigma_Y}$$

- ρ_{XY} is a measure of the dependence between X and Y
- ▶ $-1 \le \rho_{XY} \le 1$ (Without proof)
- $ho_{XY} = \pm 1$ if X and Y are linearly related
- If X and Y are s.i., they are uncorrelated, $\mathbf{Cov}[X, Y] = \rho_{XY} = 0$
- ▶ The converse is NOT TRUE

Define Pearson correlation coefficient

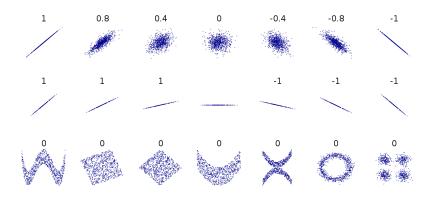
$$\rho_{XY} = \frac{\mathbf{Cov}[X, Y]}{\sigma_X \sigma_Y}$$

- ρ_{XY} is a measure of the dependence between X and Y
- ▶ $-1 \le \rho_{XY} \le 1$ (Without proof)
- $ho_{XY}=\pm 1$ if X and Y are linearly related
- If X and Y are s.i., they are uncorrelated, $\mathbf{Cov}[X, Y] = \rho_{XY} = 0$
- ► The converse is NOT TRUE

Define Pearson correlation coefficient

$$\rho_{XY} = \frac{\mathbf{Cov}[X, Y]}{\sigma_X \sigma_Y}$$

- ρ_{XY} is a measure of the dependence between X and Y
- ► $-1 \le \rho_{XY} \le 1$ (Without proof)
- $ho_{XY}=\pm 1$ if X and Y are linearly related
- If X and Y are s.i., they are uncorrelated, $\mathbf{Cov}[X, Y] = \rho_{XY} = 0$
- The converse is NOT TRUE



Define Spearman's rank correlation coefficient

$$\mathit{r_S} = \rho_{\mathit{rg}_X\mathit{rg}_Y} = \frac{\mathsf{Cov}[\mathit{rg}_X,\mathit{rg}_Y]}{\sigma_{\mathit{rg}_X}\sigma_{\mathit{rg}_Y}}$$

as the Pearson correlation coefficient between the ranked variables.

This can be calculated by scipy.stats.spearmanr

- Pearson's correlation works well if the relationship between variables is linear and if the variables are roughly normal
- Spearman's rank correlation mitigates the effect of outliers and skewed distributions

Define Spearman's rank correlation coefficient

$$r_S = \rho_{rg_X rg_Y} = \frac{\mathbf{Cov}[rg_X, rg_Y]}{\sigma_{rg_X} \sigma_{rg_Y}}$$

as the Pearson correlation coefficient between the ranked variables. This can be calculated by **scipy.stats.spearmanr**

- Pearson's correlation works well if the relationship between variables is linear and if the variables are roughly normal
- Spearman's rank correlation mitigates the effect of outliers and skewed distributions

Define Spearman's rank correlation coefficient

$$r_{S} = \rho_{rg_{X}rg_{Y}} = \frac{\mathbf{Cov}[rg_{X}, rg_{Y}]}{\sigma_{rg_{X}}\sigma_{rg_{Y}}}$$

as the Pearson correlation coefficient between the ranked variables. This can be calculated by **scipy.stats.spearmanr**

- Pearson's correlation works well if the relationship between variables is linear and if the variables are roughly normal
- Spearman's rank correlation mitigates the effect of outliers and skewed distributions

- Correlation coefficients measure the strength and sign of a relationship
- Not the slope
- A common way to show the slop is linear least squares fit
- We assume each y_i is roughly equal to $\alpha + \beta x_i$

- Correlation coefficients measure the strength and sign of a relationship
- Not the slope
- A common way to show the slop is linear least squares fit
- We assume each y_i is roughly equal to $\alpha + \beta x_i$

- Correlation coefficients measure the strength and sign of a relationship
- Not the slope
- A common way to show the slop is linear least squares fit
- We assume each y_i is roughly equal to $\alpha + \beta x_i$

- Correlation coefficients measure the strength and sign of a relationship
- Not the slope
- A common way to show the slop is linear least squares fit
- We assume each y_i is roughly equal to $\alpha + \beta x_i$

▶ Unless the correlation is perfect, we have deviation, or residual

$$\epsilon_i = \alpha + \beta x_i - y_i \neq 0$$

- Non-zero ϵ_i might be due to random factors like measurement error
- Might also be due to non-random factors that are unknown like model imperfection

Unless the correlation is perfect, we have deviation, or residual

$$\epsilon_i = \alpha + \beta x_i - y_i \neq 0$$

- Non-zero ϵ_i might be due to random factors like measurement error
- Might also be due to non-random factors that are unknown like model imperfection

► We want to minimize the squared residual:

$$\min_{\alpha,\beta} \sum \epsilon_i^2$$

- $ightharpoonup \hat{lpha}, \hat{eta}$ that minimize squared residual can be calculated easily
- Compute sample means $\bar{x}, \bar{y}, Var[X], Cov[X, Y]$

$$\begin{cases} \hat{\beta} = & \frac{\mathsf{Cov}[X, Y]}{\mathsf{Var}[X]} \\ \hat{\alpha} = & \bar{y} - \hat{\beta}\bar{x} \end{cases}$$

▶ We want to minimize the squared residual:

$$\min_{\alpha,\beta} \sum \epsilon_i^2$$

- ho \hat{lpha},\hat{eta} that minimize squared residual can be calculated easily
- ► Compute sample means $\bar{x}, \bar{y}, Var[X], Cov[X, Y]$

$$\begin{cases} \hat{\beta} = & \frac{\mathsf{Cov}[X, Y]}{\mathsf{Var}[X]} \\ \hat{\alpha} = & \bar{y} - \hat{\beta}\bar{x} \end{cases}$$

▶ We want to minimize the squared residual:

$$\min_{\alpha,\beta} \sum \epsilon_i^2$$

- $ightharpoonup \hat{lpha}, \hat{eta}$ that minimize squared residual can be calculated easily
- ► Compute sample means $\bar{x}, \bar{y}, Var[X], Cov[X, Y]$

$$\begin{cases} \hat{\beta} = & \frac{\operatorname{Cov}[X,Y]}{\operatorname{Var}[X]} \\ \hat{\alpha} = & \bar{y} - \hat{\beta}\bar{x} \end{cases}$$

We want to minimize the squared residual:

$$\min_{\alpha,\beta} \sum \epsilon_i^2$$

- ho \hat{lpha},\hat{eta} that minimize squared residual can be calculated easily
- ► Compute sample means $\bar{x}, \bar{y}, Var[X], Cov[X, Y]$

$$\begin{cases} \hat{\beta} = & \frac{\operatorname{Cov}[X,Y]}{\operatorname{Var}[X]} \\ \hat{\alpha} = & \bar{y} - \hat{\beta}\bar{x} \end{cases}$$

- Having a linear model fit to the data, we may want to know how good is it
- One common measure is predictive power of a model
- Also commonly known as "R-squared"

$$R^2 = 1 - \frac{\mathbf{Var}[\epsilon]}{\mathbf{Var}[Y]}$$

- Having a linear model fit to the data, we may want to know how good is it
- One common measure is predictive power of a model
- Also commonly known as "R-squared"

$$R^2 = 1 - \frac{\mathbf{Var}[\epsilon]}{\mathbf{Var}[Y]}$$

- Having a linear model fit to the data, we may want to know how good is it
- One common measure is predictive power of a model
- Also commonly known as "R-squared"

$$R^2 = 1 - \frac{\mathsf{Var}[\epsilon]}{\mathsf{Var}[Y]}$$

$$R^2 = 1 - \frac{\mathbf{Var}[\epsilon]}{\mathbf{Var}[Y]}$$

- ▶ $Var[Y] = \frac{1}{n} \sum (\bar{y} y_i)^2$ is the MSE when estimating Y using \bar{y}
- ▶ $Var[\epsilon] = \frac{1}{n} \sum (\hat{\alpha} + \hat{\beta}x_i y_i)^2$ is the MSE when estimating Y using the linear model
- Var[ε]/Var[Y] shows the performance difference caused by introducing linear model
- $R^2 = 0.60$, we can say the model explains 60% of the variability
- ▶ Or more precisely, reduces the MSE of prediction by 60%
- $ightharpoonup R^2 =
 ho^2$ (Can you show it in Jupyter notebook?)

$$R^2 = 1 - \frac{\mathbf{Var}[\epsilon]}{\mathbf{Var}[Y]}$$

- ▶ $Var[Y] = \frac{1}{n} \sum (\bar{y} y_i)^2$ is the MSE when estimating Y using \bar{y}
- ▶ $Var[\epsilon] = \frac{1}{n} \sum (\hat{\alpha} + \hat{\beta}x_i y_i)^2$ is the MSE when estimating Y using the linear mode
- Var[ε]/Var[Y] shows the performance difference caused by introducing linear model
- $R^2 = 0.60$, we can say the model explains 60% of the variability
- ▶ Or more precisely, reduces the MSE of prediction by 60%
- $ightharpoonup R^2 =
 ho^2$ (Can you show it in Jupyter notebook?)



$$R^2 = 1 - \frac{\mathbf{Var}[\epsilon]}{\mathbf{Var}[Y]}$$

- ► Var[Y] = $\frac{1}{n} \sum (\bar{y} y_i)^2$ is the MSE when estimating Y using \bar{y}
- ▶ $\mathbf{Var}[\epsilon] = \frac{1}{n} \sum (\hat{\alpha} + \hat{\beta}x_i y_i)^2$ is the MSE when estimating Y using the linear model
- Var[ε]/Var[Y] shows the performance difference caused by introducing linear model
- $ightharpoonup R^2 = 0.60$, we can say the model explains 60% of the variability
- ▶ Or more precisely, reduces the MSE of prediction by 60%
- $ightharpoonup R^2 =
 ho^2$ (Can you show it in Jupyter notebook?)

$$R^2 = 1 - \frac{\mathbf{Var}[\epsilon]}{\mathbf{Var}[Y]}$$

- ► Var[Y] = $\frac{1}{n}\sum (\bar{y} y_i)^2$ is the MSE when estimating Y using \bar{y}
- ▶ $Var[\epsilon] = \frac{1}{n} \sum (\hat{\alpha} + \hat{\beta}x_i y_i)^2$ is the MSE when estimating Y using the linear model
- Var[ϵ] / Var[Y] shows the performance difference caused by introducing linear model
- $R^2 = 0.60$, we can say the model explains 60% of the variability
- ▶ Or more precisely, reduces the MSE of prediction by 60%
- $ightharpoonup R^2 =
 ho^2$ (Can you show it in Jupyter notebook?)

$$R^2 = 1 - \frac{\mathsf{Var}[\epsilon]}{\mathsf{Var}[Y]}$$

- ► Var[Y] = $\frac{1}{n}\sum (\bar{y} y_i)^2$ is the MSE when estimating Y using \bar{y}
- ▶ $\mathbf{Var}[\epsilon] = \frac{1}{n} \sum (\hat{\alpha} + \hat{\beta}x_i y_i)^2$ is the MSE when estimating Y using the linear model
- Var[∈]/Var[Y] shows the performance difference caused by introducing linear model
- $R^2 = 0.60$, we can say the model explains 60% of the variability
- ▶ Or more precisely, reduces the MSE of prediction by 60%
- $ightharpoonup R^2 =
 ho^2$ (Can you show it in Jupyter notebook?)

$$R^2 = 1 - \frac{\mathsf{Var}[\epsilon]}{\mathsf{Var}[Y]}$$

- ► Var[Y] = $\frac{1}{n}\sum (\bar{y} y_i)^2$ is the MSE when estimating Y using \bar{y}
- ▶ $\mathbf{Var}[\epsilon] = \frac{1}{n} \sum (\hat{\alpha} + \hat{\beta}x_i y_i)^2$ is the MSE when estimating Y using the linear model
- Var[ϵ] / Var[Y] shows the performance difference caused by introducing linear model
- $R^2 = 0.60$, we can say the model explains 60% of the variability
- Or more precisely, reduces the MSE of prediction by 60%
- $ightharpoonup R^2 =
 ho^2$ (Can you show it in Jupyter notebook?)

$$R^2 = 1 - \frac{\mathsf{Var}[\epsilon]}{\mathsf{Var}[Y]}$$

- ► Var[Y] = $\frac{1}{n}\sum (\bar{y} y_i)^2$ is the MSE when estimating Y using \bar{y}
- ▶ $\mathbf{Var}[\epsilon] = \frac{1}{n} \sum (\hat{\alpha} + \hat{\beta}x_i y_i)^2$ is the MSE when estimating Y using the linear model
- Var[∈]/_{Var[Y]} shows the performance difference caused by introducing linear model
- $R^2 = 0.60$, we can say the model explains 60% of the variability
- Or more precisely, reduces the MSE of prediction by 60%
- $ightharpoonup R^2 =
 ho^2$ (Can you show it in Jupyter notebook?)

US spending on science, space, and technology

Suicides by hanging, strangulation and suffocation



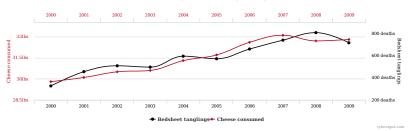
Number of people who drowned by falling into a pool correlates with

Films Nicolas Cage appeared in



Per capita cheese consumption

Number of people who died by becoming tangled in their bedsheets



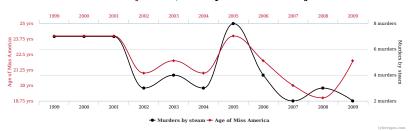


Per capita consumption of margarine



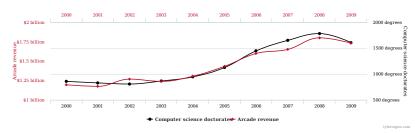
Age of Miss America

Murders by steam, hot vapours and hot objects



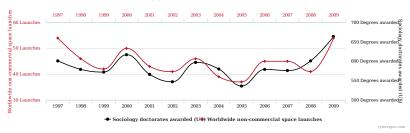
Total revenue generated by arcades correlates with

Computer science doctorates awarded in the US



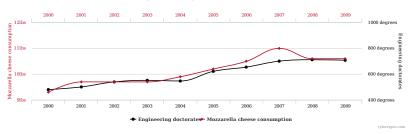
Worldwide non-commercial space launches

Sociology doctorates awarded (US)



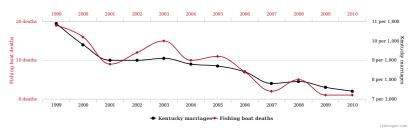
Per capita consumption of mozzarella cheese

Civil engineering doctorates awarded



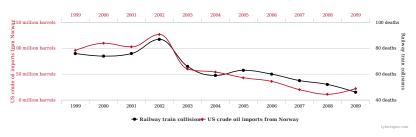
People who drowned after falling out of a fishing boat correlates with

Marriage rate in Kentucky



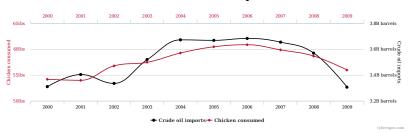
US crude oil imports from Norway

Drivers killed in collision with railway train



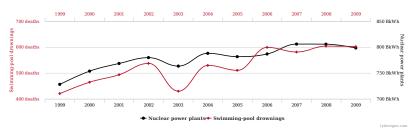
Per capita consumption of chicken

Total US crude oil imports



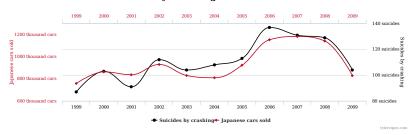
Number people who drowned while in a swimming-pool

Power generated by US nuclear power plants



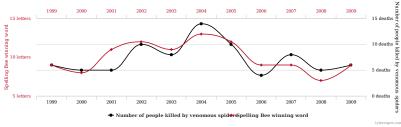
Japanese passenger cars sold in the US correlates with

Suicides by crashing of motor vehicle



Letters in Winning Word of Scripps National Spelling Beecorrelates with

Number of people killed by venomous spiders



Math doctorates awarded

Uranium stored at US nuclear power plants

