Covariance and Correlation coefficient

PROBABILITY AND STATISTICS FOR ECE

Covariance

- 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

Covariance

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:

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

So:

$$\mathbf{Var}[X+Y]=\mathbf{Var}[X]+\mathbf{Var}[Y]+2\mathbf{Cov}[X,Y]$$

Covariance

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:

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

So:

$$\mathbf{Var}[X+Y]=\mathbf{Var}[X]+\mathbf{Var}[Y]+2\mathbf{Cov}[X,Y]$$

Covariance

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:

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

So:

$$\mathbf{Var}[X+Y]=\mathbf{Var}[X]+\mathbf{Var}[Y]+2\mathbf{Cov}[X,Y]$$

Covariance

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:

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

So:

$$\mathbf{Var}[X+Y]=\mathbf{Var}[X]+\mathbf{Var}[Y]+2\mathbf{Cov}[X,Y]$$

Covariance

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:

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

So:

$$\mathbf{Var}[X+Y]=\mathbf{Var}[X]+\mathbf{Var}[Y]+2\mathbf{Cov}[X,Y]$$

Covariance

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:

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

So:

$$\mathbf{Var}[X+Y]=\mathbf{Var}[X]+\mathbf{Var}[Y]+2\mathbf{Cov}[X,Y]$$

Covariance

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:

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

So:

$$\mathbf{Var}[X+Y]=\mathbf{Var}[X]+\mathbf{Var}[Y]+2\mathbf{Cov}[X,Y]$$

Covariance

Another definition:

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

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

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

This can be calculated by **numpy.cov**

Covariance

Another definition:

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

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

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

This can be calculated by **numpy.cov**

Covariance

Another definition:

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

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

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

This can be calculated by **numpy.cov**

Covariance

Another definition:

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

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

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

This can be calculated by **numpy.cov**

Covariance

Another definition:

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

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

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

This can be calculated by **numpy.cov**

Covariance

Another definition:

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

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

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

This can be calculated by **numpy.cov**

Covariance

Another definition:

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

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

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

This can be calculated by **numpy.cov**

Covariance Properties

- $\mathbf{Cov}[X, a] = 0$
- $\mathbf{Cov}[X, X] = \mathbf{Var}[X]$
- $\mathbf{Cov}[X, Y] = \mathbf{Cov}[Y, X]$
- if $\mathbf{Cov}[X, Y] = 0$, 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 Properties

- $\mathbf{Cov}[X, a] = 0$
- $\mathbf{Cov}[X, X] = \mathbf{Var}[X]$
- $\mathbf{Cov}[X, Y] = \mathbf{Cov}[Y, X]$
- if $\mathbf{Cov}[X, Y] = 0$, 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 Properties

- $\mathbf{Cov}[X, a] = 0$
- $\mathbf{Cov}[X, X] = \mathbf{Var}[X]$
- $\mathbf{Cov}[X, Y] = \mathbf{Cov}[Y, X]$
- if $\mathbf{Cov}[X, Y] = 0$, 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 Properties

- $\mathbf{Cov}[X, a] = 0$
- $\mathbf{Cov}[X, X] = \mathbf{Var}[X]$
- $\mathbf{Cov}[X, Y] = \mathbf{Cov}[Y, X]$
- if $\mathbf{Cov}[X, Y] = 0$, 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 Properties

- $\mathbf{Cov}[X, a] = 0$
- $\mathbf{Cov}[X, X] = \mathbf{Var}[X]$
- $\mathbf{Cov}[X, Y] = \mathbf{Cov}[Y, X]$
- if $\mathbf{Cov}[X, Y] = 0$, 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 Properties

- $\mathbf{Cov}[X, a] = 0$
- $\mathbf{Cov}[X, X] = \mathbf{Var}[X]$
- $\mathbf{Cov}[X, Y] = \mathbf{Cov}[Y, X]$
- if $\mathbf{Cov}[X, Y] = 0$, 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

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

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

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

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

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

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

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

Correlation Coefficient

- Covariance indicates how correlated are two R.V.
- But not comparable
- $\mathbf{Cov}[X, Y] = 1$, $\mathbf{Cov}[Z, W] = 10$, Z, W more correlated?

Correlation Coefficient

Define Pearson correlation coefficient

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

This can be calculated by **numpy.corrcoef**

- ρ_{XY} is a measure of the dependence between X and Y
- $-1 \leq \rho_{XY} \leq 1$ (Without proof)
- $\rho_{XY} = \pm 1$ 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

Correlation Coefficient

Define Pearson correlation coefficient

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

This can be calculated by **numpy.corrcoef**

- ρ_{XY} is a measure of the dependence between X and Y
- $-1 \leq \rho_{XY} \leq 1$ (Without proof)
- $\rho_{XY} = \pm 1$ 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

Correlation Coefficient

Define Pearson correlation coefficient

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

This can be calculated by **numpy.corrcoef**

- ρ_{XY} is a measure of the dependence between X and Y
- $-1 \leq \rho_{XY} \leq 1$ (Without proof)
- $\rho_{XY} = \pm 1$ 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

Correlation Coefficient

Define Pearson correlation coefficient

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

This can be calculated by **numpy.corrcoef**

- ρ_{XY} is a measure of the dependence between X and Y
- $-1 \leq \rho_{XY} \leq 1$ (Without proof)
- $\rho_{XY} = \pm 1$ 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

Correlation Coefficient

Define Pearson correlation coefficient

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

This can be calculated by **numpy.corrcoef**

- ρ_{XY} is a measure of the dependence between X and Y
- $-1 \leq \rho_{XY} \leq 1$ (Without proof)
- $\rho_{XY} = \pm 1$ 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

Correlation Coefficient

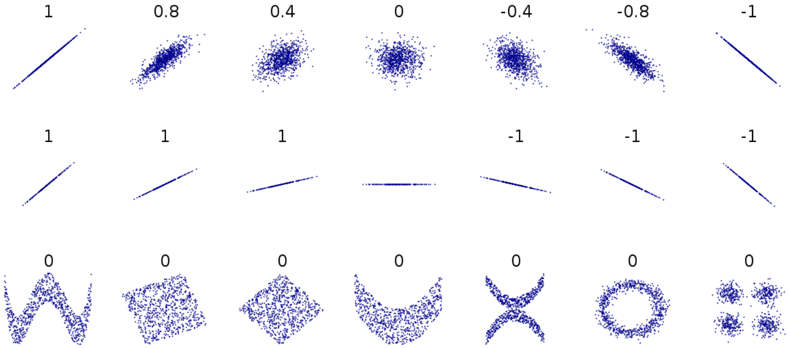
Define Pearson correlation coefficient

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

This can be calculated by **numpy.corrcoef**

- ρ_{XY} is a measure of the dependence between X and Y
- $-1 \leq \rho_{XY} \leq 1$ (Without proof)
- $\rho_{XY} = \pm 1$ 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

Correlation Coefficient



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 Coefficient

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 Coefficient

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

Least Square Fit

- 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 $\alpha + \beta x_i$

Least Square Fit

- 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 $\alpha + \beta x_i$

Least Square Fit

- 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 $\alpha + \beta x_i$

Least Square Fit

- 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 $\alpha + \beta x_i$

Least Square Fit

- 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

Least Square Fit

- 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

Least Square Fit

- We want to minimize the squared residual:

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

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

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

scipy.stats.linregress

Least Square Fit

- We want to minimize the squared residual:

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

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

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

scipy.stats.linregress

Least Square Fit

- We want to minimize the squared residual:

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

- $\hat{\alpha}, \hat{\beta}$ that minimize squared residual can be calculated easily
- Compute sample means $\bar{x}, \bar{y}, \textbf{Var}[X], \textbf{Cov}[X, Y]$
- $$\begin{cases} \hat{\beta} = & \frac{\textbf{Cov}[X,Y]}{\textbf{Var}[X]} \\ \hat{\alpha} = & \bar{y} - \hat{\beta}\bar{x} \end{cases}$$

scipy.stats.linregress

Least Square Fit

- We want to minimize the squared residual:

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

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

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

scipy.stats.linregress

Goodness of Fit

- 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"

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

Goodness of Fit

- 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"

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

Goodness of Fit

- 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"

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

Goodness of Fit

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

- $\mathbf{Var}[Y] = \frac{1}{n}\sum(\bar{y} - y_i)^2$
  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
- ${}^{\mathbf{Var}[\epsilon]}/_{\mathbf{Var}[Y]}$ 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%
- $R^2 = \rho^2$ (Can you show it in Jupyter notebook?)

Goodness of Fit

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

- $\mathbf{Var}[Y] = \frac{1}{n}\sum(\bar{y} - y_i)^2$
  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
- ${}^{\mathbf{Var}[\epsilon]}/_{\mathbf{Var}[Y]}$ 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%
- $R^2 = \rho^2$ (Can you show it in Jupyter notebook?)

Goodness of Fit

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

- $\mathbf{Var}[Y] = \frac{1}{n}\sum(\bar{y} - y_i)^2$
  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
- ${}^{\mathbf{Var}[\epsilon]}/_{\mathbf{Var}[Y]}$ 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%
- $R^2 = \rho^2$ (Can you show it in Jupyter notebook?)

Goodness of Fit

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

- $\textbf{Var}[Y] = \frac{1}{n}\sum(\bar{y} - y_i)^2$
  is the MSE when estimating Y using $\bar{y}$
- $\textbf{Var}[\epsilon] = \frac{1}{n}\sum(\hat{\alpha} + \hat{\beta}x_i - y_i)^2$
  is the MSE when estimating Y using the linear model
- ${}^{\textbf{Var}[\epsilon]}/_{\textbf{Var}[Y]}$ 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%
- $R^2 = \rho^2$ (Can you show it in Jupyter notebook?)

Goodness of Fit

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

- $\mathbf{Var}[Y] = \frac{1}{n}\sum(\bar{y} - y_i)^2$
  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
- ${}^{\mathbf{Var}[\epsilon]}/_{\mathbf{Var}[Y]}$ 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%
- $R^2 = \rho^2$ (Can you show it in Jupyter notebook?)

Goodness of Fit

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

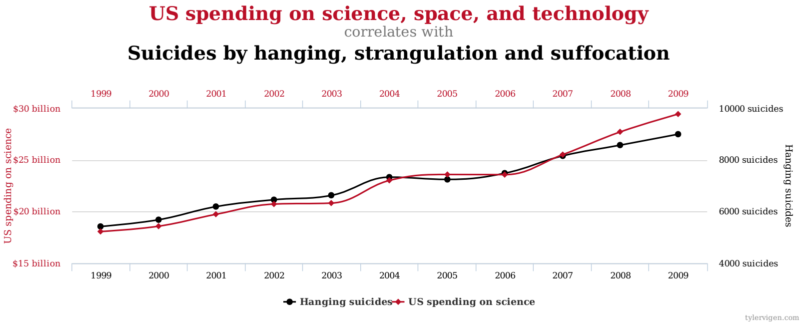
- $\textbf{Var}[Y] = \frac{1}{n}\sum(\bar{y} - y_i)^2$
  is the MSE when estimating Y using $\bar{y}$
- $\textbf{Var}[\epsilon] = \frac{1}{n}\sum(\hat{\alpha} + \hat{\beta}x_i - y_i)^2$
  is the MSE when estimating Y using the linear model
- ${}^{\textbf{Var}[\epsilon]}/_{\textbf{Var}[Y]}$ shows the 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%
- $R^2 = \rho^2$ (Can you show it in Jupyter notebook?)

Goodness of Fit

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

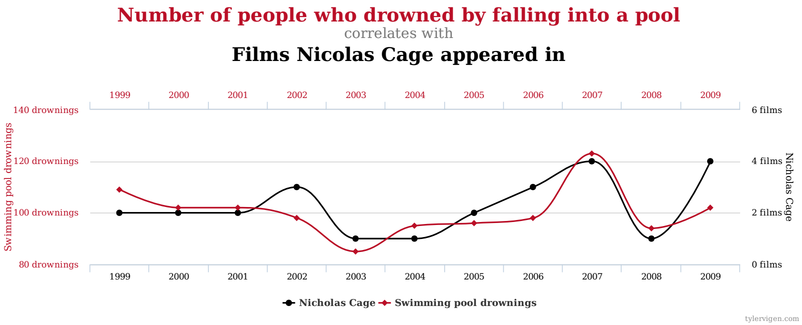
- $\textbf{Var}[Y] = \frac{1}{n}\sum(\bar{y} - y_i)^2$
  is the MSE when estimating Y using $\bar{y}$
- $\textbf{Var}[\epsilon] = \frac{1}{n}\sum(\hat{\alpha} + \hat{\beta}x_i - y_i)^2$
  is the MSE when estimating Y using the linear model
- ${}^{\textbf{Var}[\epsilon]}/_{\textbf{Var}[Y]}$ shows the 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%
- $R^2 = \rho^2$ (Can you show it in Jupyter notebook?)

Correlation and Causation

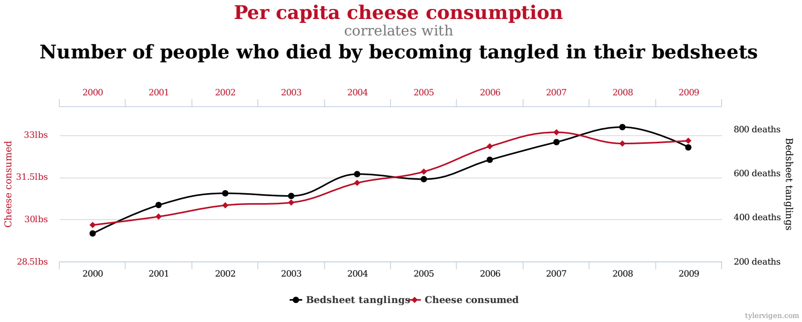


Charts from the book **Spurious Correlations**
`http://www.tylervigen.com/spurious-correlations`

Correlation and Causation

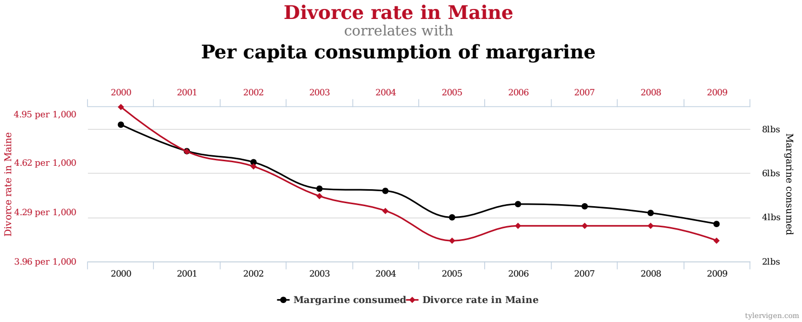


Charts from the book **Spurious Correlations**
`http://www.tylervigen.com/spurious-correlations`

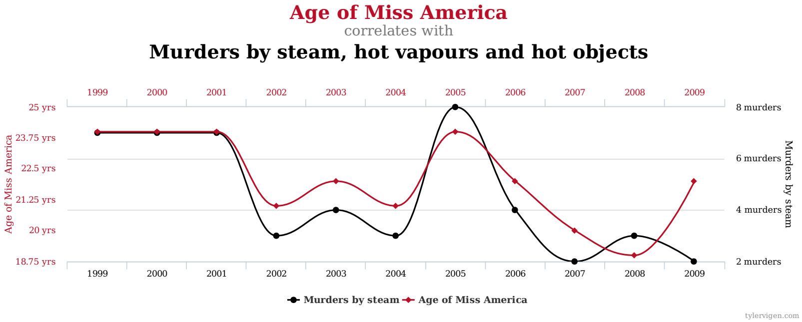


Charts from the book **Spurious Correlations**
`http://www.tylervigen.com/spurious-correlations`

Correlation and Causation

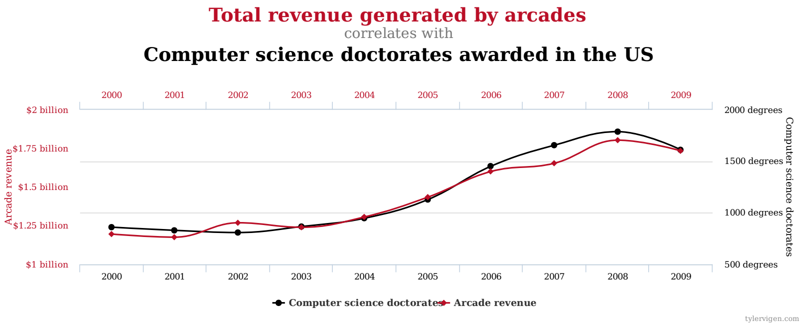


Charts from the book **Spurious Correlations**
`http://www.tylervigen.com/spurious-correlations`



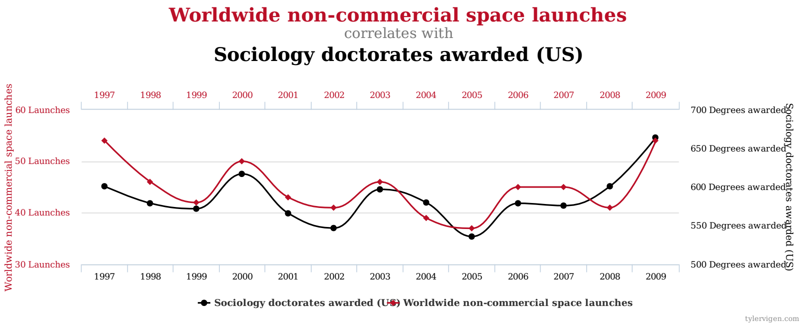
Charts from the book **Spurious Correlations**
`http://www.tylervigen.com/spurious-correlations`

Correlation and Causation



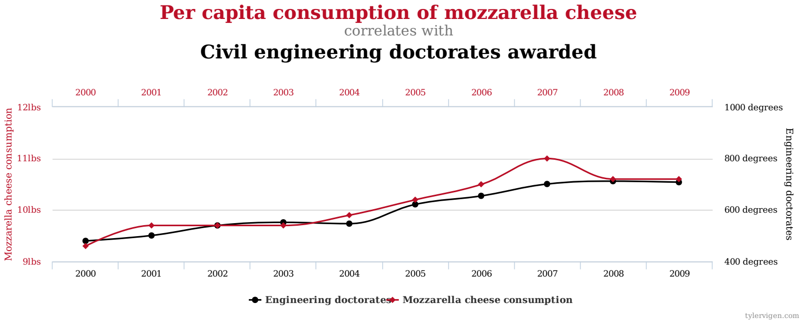
Charts from the book **Spurious Correlations**
`http://www.tylervigen.com/spurious-correlations`

Correlation and Causation



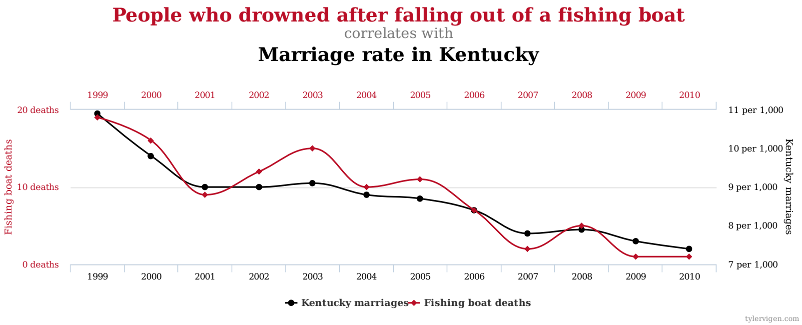
Charts from the book **Spurious Correlations**
`http://www.tylervigen.com/spurious-correlations`

Correlation and Causation



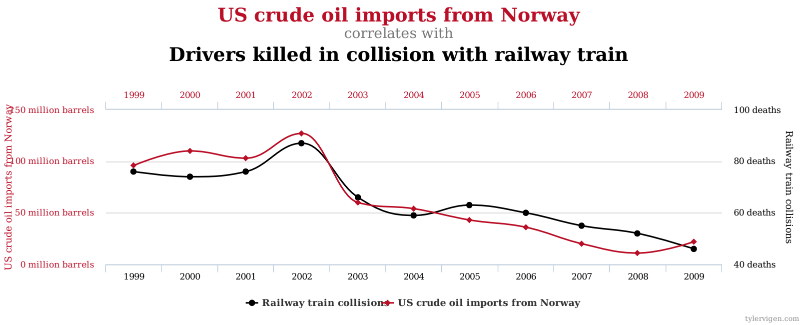
Charts from the book **Spurious Correlations**
`http://www.tylervigen.com/spurious-correlations`

Correlation and Causation



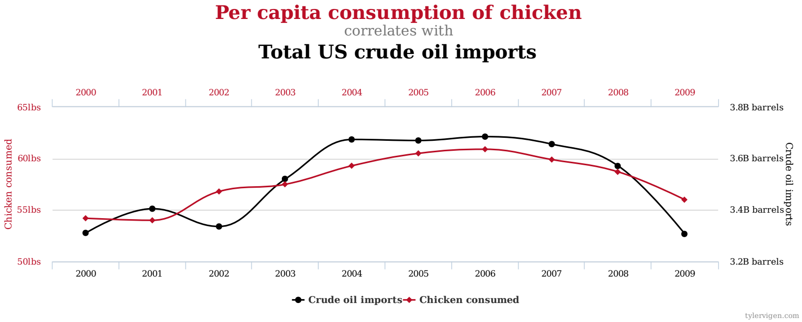
Charts from the book **Spurious Correlations**
`http://www.tylervigen.com/spurious-correlations`

Correlation and Causation



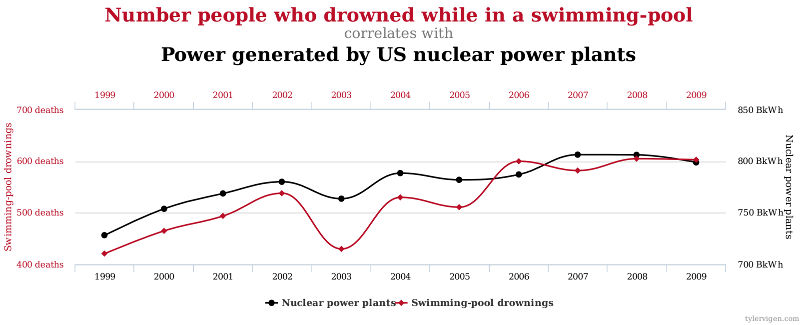
Charts from the book **Spurious Correlations**
`http://www.tylervigen.com/spurious-correlations`

Correlation and Causation



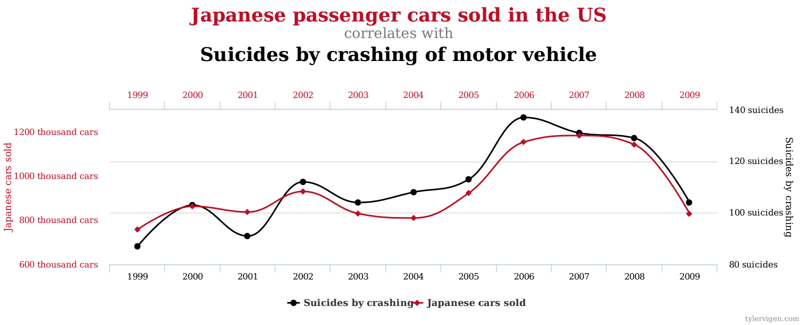
Charts from the book **Spurious Correlations**
`http://www.tylervigen.com/spurious-correlations`

Correlation and Causation



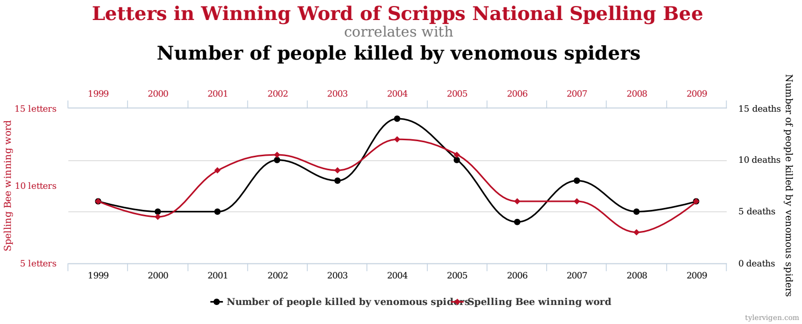
Charts from the book **Spurious Correlations**
`http://www.tylervigen.com/spurious-correlations`

Correlation and Causation



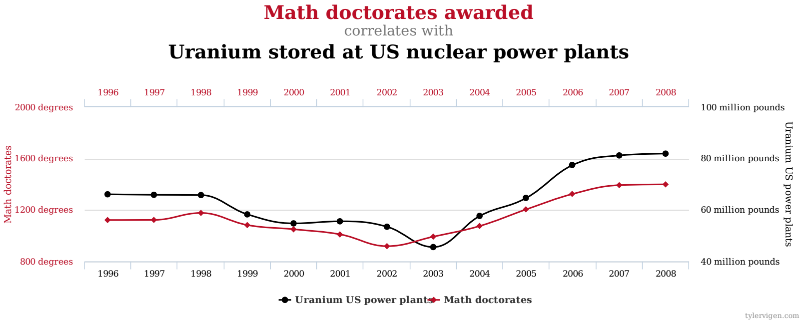
Charts from the book **Spurious Correlations**
`http://www.tylervigen.com/spurious-correlations`

Correlation and Causation



Charts from the book **Spurious Correlations**
`http://www.tylervigen.com/spurious-correlations`

Correlation and Causation



Charts from the book **Spurious Correlations**
`http://www.tylervigen.com/spurious-correlations`