EE2211 Pre-Tutorial 3

Dr Feng LIN

feng_lin@nus.edu.sg

Agenda

- Recap
- Self-learning
- Tutorial 3

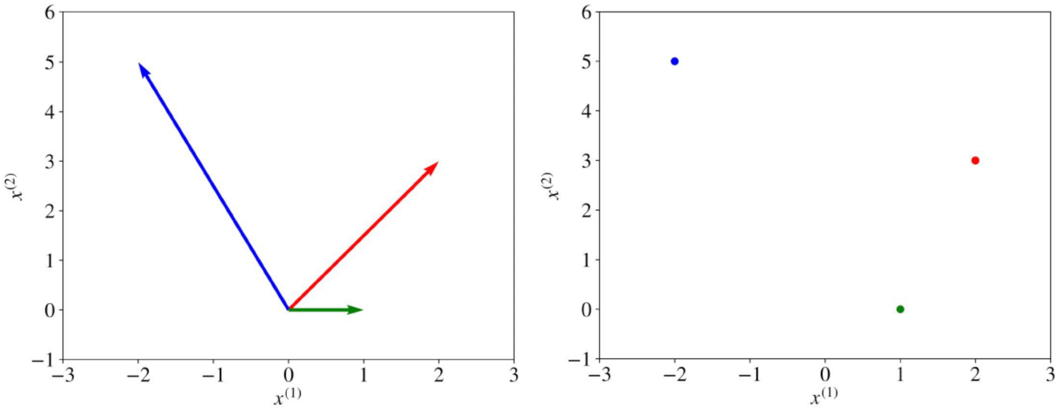
Recap

- Linear Algebra
    - Vectors
    - Matrices
    - Linear Equations
- Causality and Simpson’s paradox
- Probability
    - Axioms of Probability
    - Random Variable
    - Basic rules
    - Bayes' rule

Vectors

- A **vector** is an ordered list of scalar values
  - Denoted by a bold character, e.g. $\boldsymbol{x}$ or $\boldsymbol{a}$
- **Vectors** can be visualized as, in a multi-dimensional space,
  - arrows that point to some directions, or e.g. Velocity vector
  - points e.g. color points

Illustrations of three two-dimensional vectors, $\mathbf{a} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$, $\mathbf{b} = \begin{bmatrix} -2 \\ 5 \end{bmatrix}$, and $\mathbf{c} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$



Matrices

- A matrix is a rectangular array of numbers arranged in rows and columns
  - ➢ Denoted with bold capital letters, such as $\boldsymbol{X}$ or $\boldsymbol{W}$
  - ➢ An example of a matrix with two rows and three columns:

$$\mathbf{X} = \begin{bmatrix} 2 & 4 & -3 \\ 21 & -6 & -1 \end{bmatrix}$$

e.g. image frame

- Note:
  - ➢ For elements in matrix X, we shall use the indexing $x_{1,1}$ where the first and second indices indicate the row and the column position.
  - ➢ Usually, for input data, rows represent samples and columns represent features

Linear Equation

Linear dependence and independence

- A collection of d-vector $\mathbf{x}_1, \cdots, \mathbf{x}_m$ (with $m \geq 1$) is called linearly dependent if

$$\beta_1 \mathbf{x}_1 + \cdots + \beta_m \mathbf{x}_m = 0$$

  Holds for some $\beta_1, \cdots, \beta_m$ that are not all zeros.

- A collection of d-vector $\mathbf{x}_1, \cdots, \mathbf{x}_m$ (with $m \geq 1$) is called linearly independent if

$$\beta_1 \mathbf{x}_1 + \cdots + \beta_m \mathbf{x}_m = 0$$

  only hold for $\beta_1 = \cdots = \beta_m = 0$.

Note: If all rows or columns of a square matrix **X** are linearly **independent**, then **X** is **invertible**.

Causality

- Causality, or causation is:
  - The influence by which one event or process (i.e., cause) contributes to another (i.e. effect),
  - The cause is partly responsible for the effect, and the effect is partly dependent on the cause

- Causality relates to an extremely very wide domain of subjects: philosophy, science, management, humanity.

- Causality research is extremely complex
  - Researcher can never be completely certain that there are **no other factors** influencing the causal relationship,
  - In most cases, we can only say "probably" causal.

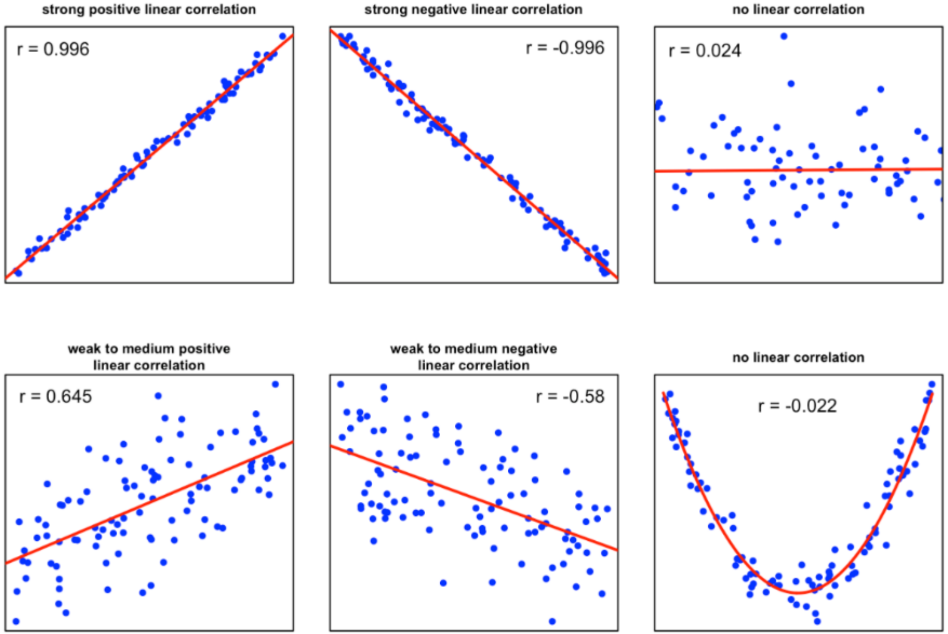
Correlation (vs Causality)

- Linear correlation coefficient, r, which is also known as the Pearson Coefficient.

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}} = \frac{s_{xy}}{s_x s_y},$$

Strong linear relationship	$r > 0.9$
Medium linear relationship	$0.7 < r \le 0.9$
Weak linear relationship	$0.5 < r \le 0.7$
No or doubtful linear relationship	$0 < r \le 0.5$

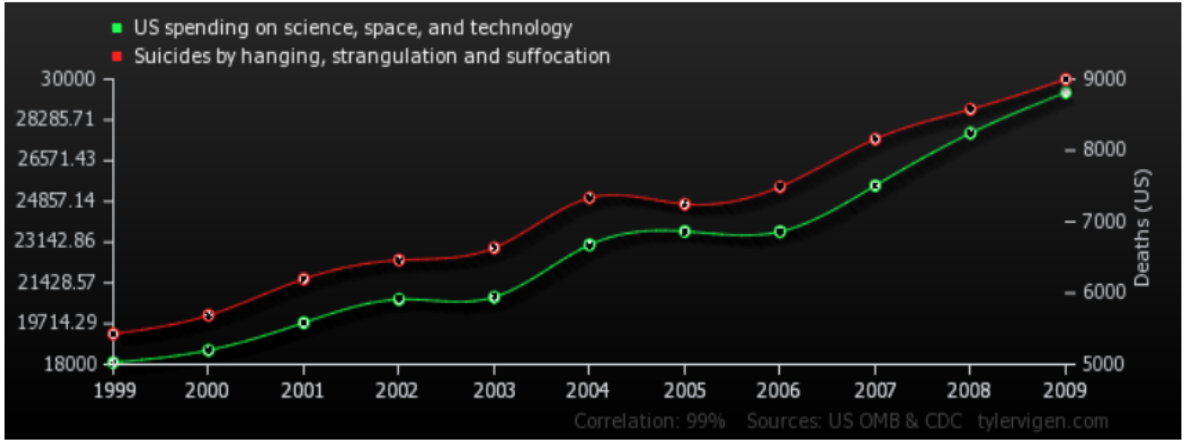
The same holds for negative values.



https://www.geo.fu-berlin.de/en/v/soga-py/Basics-of-statistics/Descriptive-Statistics/Measures-of-Relation-Between-Variables/Correlation/index.html

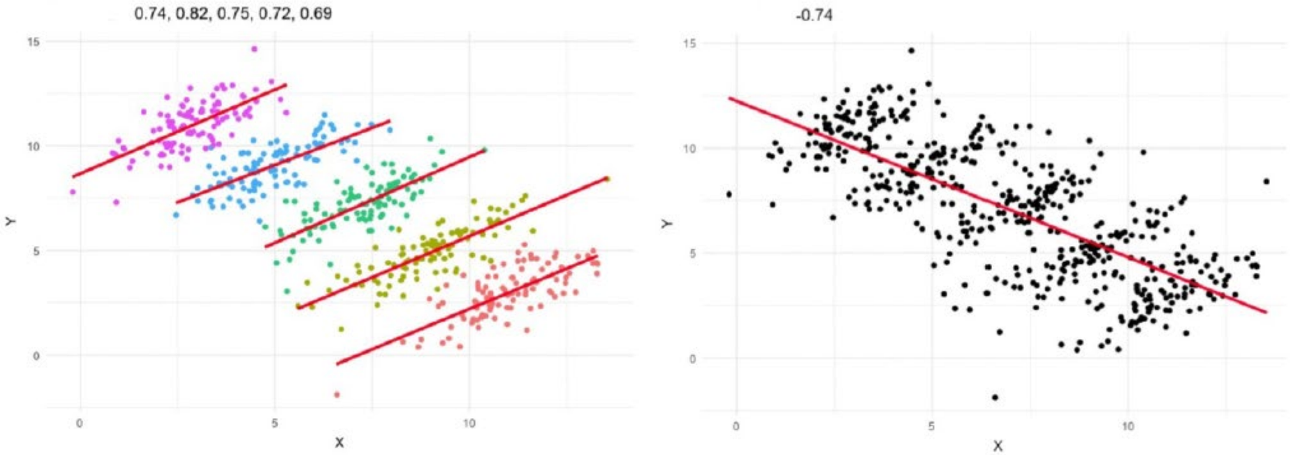
Correlation does not imply causation!

- Some great examples of correlations that can be calculated but are clearly not causally related appear at http://tylervigen.com/



Simpson’s paradox

- **Simpson's paradox** is a phenomenon in probability and statistics, in which a trend appears in several different groups of data but disappears or reverses when these groups are combined.



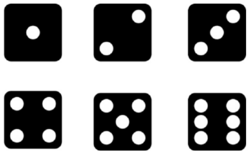
The same set of samples!

Probability

- We describe a *random experiment* by describing its procedure and observations of its *outcomes*.

Outcome

- Outcomes are mutual exclusive in the sense that only one outcome occurs in a specific trial of the random experiment.
  - This also means an outcome is not decomposable.
  - All unique outcomes form a *sample space.*



Sample space

- A subset of sample space S, denoted as A, is an *event* in a random experiment $A \subset S$, that is meaningful to an application.
  - Example of an event: faces with numbers no greater than 3

$$P(A) = \frac{\text{\# of favorable outcomes}}{\text{\# of total outcomes}}$$

Event

Axioms of Probability

- Assume events $A \subseteq S$ and $B \subseteq S$.

1. [**Nonnegativity**] $Pr(A) \geq 0$ for every event A.
2. [**Normalization**] $Pr(S) = 1$.
3. [**Additivity**] If $A \cap B = \emptyset$, then $Pr(A \cup B) = Pr(A) + Pr(B)$.

   otherwise $Pr(A \cup B) = Pr(A) + Pr(B) - Pr(A \cap B)$

   It can be extended to more events:

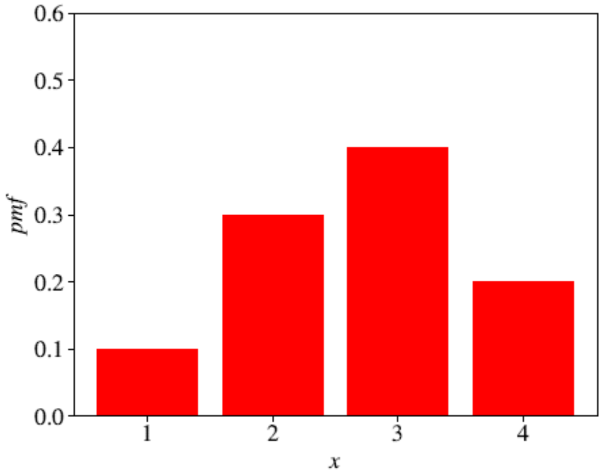
   If $A_1, A_2, \ldots$ are disjoint, then $Pr(\cup_{k=1}^{\infty} A_k) = \Sigma_{k=1}^{\infty} Pr(A_k)$.

Random Variable

- A random variable, usually written as an italic capital letter, like X, is a variable whose possible values are numerical outcomes of a random event.
- There are two types of random variables: discrete and continuous.

Discrete Random Variable

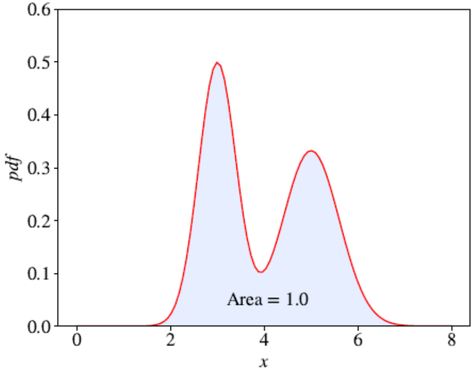
- A discrete random variable (DRV) takes on only a countable number of distinct values such as red, orange, blue or 1, 2, 3.
- The probability distribution of a discrete random variable is described by a list of probabilities associated with each of its possible values.
- This list of probabilities is called a probability mass function (pmf).
  - Like a histogram, except that here the probabilities sum to 1



A probability mass function

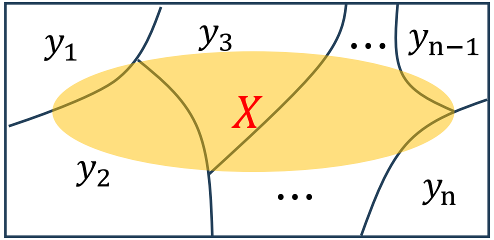
Continuous Random Variable

- A continuous random variable (CRV) takes an infinite number of possible values in some interval.
  - ➢ Examples include height, weight, and time.
  - ➢ The number of values of a continuous random variable X is infinite, the probability $\Pr(X = c)$ for any c is 0.
  - ➢ Therefore, instead of the list of probabilities, the probability distribution of a CRV (a continuous probability distribution) is described by a **probability density function** (pdf).
  - ➢ The pdf is a function whose range is nonnegative and the area under the curve is equal to 1.



A probability density function

Two Basic Rules



- Sum Rule

$$\Pr(X = x) = \sum_{Y} \Pr(X = x, Y = y_i)$$

- Product Rule

$$\Pr(X = x, Y = y) = \Pr(Y = y | X = x)\, P(X = x)$$

Bayes' Rule

- The conditional probability $\Pr(Y = y|X = x)$ is the probability of the random variable Y to have a specific value y, given that another random variable X has a specific value of x.

- The **Bayes' Rule** (also known as the **Bayes' Theorem**):

likelihood prior

$$\Pr(Y = y|X = x) = \frac{\Pr(X = x|Y = y)\Pr(Y = y)}{\Pr(X = x)}$$

posterior evidence

Summary

- (Very Gentle) Introduction to Linear Algebra
- Causality and Simpson’s paradox
- Random Variable, Bayes’ Rule

THANK YOU