SAMSUNG

Samsung Innovation Campus

Artificial Intelligence Course

Together for Tomorrow!
Enabling People
Education for Future Generations

Chapter 4.

Probability and Statistics

AI Course

Chapter Description

Chapter objectives

- ✓ Be able to make probabilistic assessment.
- ✓ Be able to carry out descriptive statistics and exploratory data analysis.
- ✓ Be able to apply hypothesis testing.

Chapter contents

Unit 1.

Understanding of Probability

Random Experiment and Event

Random experiment

- Experiment where all possible outcomes can be observed.
- It can be repeated as many times as possible under the exact same conditions.
- The result or outcome is produced randomly.

Sample space

- Suppose that the possible outcomes of a random experiment are $e_1, e_2, e_3, \ldots, e_N$.
- You denote the sample space of the random experiment by S. This can be expressed as a set.

$$S = \{e_1, e_2, e_3, \ldots, e_N\}$$

Ex Rolling a dice, $S = \{1, 2, 3, 4, 5, 6\}$.

Ex Flipping a coin, $S = \{T, H\}$ where T = tail and H = head.

- An event is a subset of the sample space.
  - **Ex** The whole of $S \rightarrow$ "total event"
  - **Ex** Empty set $\phi \rightarrow$ "empty event"
  - **Ex** Subsets composed of individual elements of S: $\{e_1\}$, $\{e_2\}$, ... $\rightarrow$ "elementary events"
  - **Ex** Other subsets: $\{e_1, e_2\}$, $\{e_7, e_9, e_{13}\}$, ...

Events of rolling a dice

- Sample space: $S = \{1, 2, 3, 4, 5, 6\}$

Ex Event of outcome three: $E_1 = \{3\}$ → "Elementary event"

Ex Event of outcome equal or larger than three: $A = \{3, 4, 5, 6\}$

Ex Event of odd outcome: $O = \{1, 3, 5\}$

Ex Event of even outcome: $E = \{2, 4, 6\}$

Ex Event of even outcome that is equal or larger than three: $A \cap E = \{4, 6\}$

Ex Event of even outcome or outcome that is equal or larger than three: $A \cup E = \{2, 3, 4, 5, 6\}$

Events of rolling a dice (continued)

- Ex We notice that $E \cap O = \phi$ → "E and O are mutually exclusive events."
- Ex Event of even outcome that is **not** equal or larger than three: $E - A = \{2\}$
- Ex Event of outcome equal or larger than three that is **not** even: $A - E = \{3, 5\}$
- Ex Event of outcome that is **not** equal or larger than three: $A^c = S - A = \{1, 2\}$ → "Complementary event"

- Events of flipping two coins: one dime (10 cents) and one quarter (25 cents)
  - Sample space: $S = \{(H,H),(H,T),(T,H),(T,T)\}$

    Ex Event of both coins showing the tail side: $E_1 = \{(T,T)\}$

    Ex Event of both coins showing the same side: $E_2 = \{(H,H),(T,T)\}$

    Ex Event of at least one coin showing the tail side: $E_3 = \{(T,T),(T,H),(H,T)\}$

Definition of Probability

Mathematical definition of probability

- Suppose that N is the number of elementary events of the sample space.
- Suppose that N_A is the number of all possible outcomes corresponding to the event A.

$$P(A) = \frac{N_A}{N}$$

Probabilities of rolling two dices

- Sample space: $S = \{(1,1),(1,2),(1,3),\ldots,(6,6)\}$, N=36

Ex The probability of dices showing different numbers

$$P = \frac{36-6}{36} = \frac{5}{6}$$

Ex The probability of dices showing numbers that sum up to 7

The event we are interested in is $A = \{(1,6),(2,5),(3,4),(4,3),(5,2),(6,1)\}$. So, N_A=6.

$$P = \frac{6}{36} = \frac{1}{6}$$

Probabilities of rolling two dices

- Sample space: $S = \{(1,1),(1,2),(1,3),\ldots,(6,6)\}$, N=36

Ex The probability of dices showing numbers of which the product is even

There are 27 combinations where the product can be an even number.

$$\left.\begin{array}{l}(\text{Even}) \rightarrow (\text{Even}) = (\text{Even}) \Rightarrow 3 \times 3 = 9 \text{ cases} \\ (\text{Even}) \rightarrow (\text{Odd}) = (\text{Even}) \Rightarrow 3 \times 3 = 9 \text{ cases} \\ (\text{Odd}) \rightarrow (\text{Even}) = (\text{Even}) \Rightarrow 3 \times 3 = 9 \text{ cases}\end{array}\right\} \text{27 in total}$$

$$P = \frac{27}{36} = \frac{3}{4}$$

Statistical (empirical) definition of probability

- Total number of observations of a random experiment is N.
- The number of observations corresponding to the event A is N_A.
- Ideally, the probability is given by the following limit:

$$P(A) = \lim_{N \to \infty} \frac{N_A}{N}$$

- However, the following is more realistic:

$$P(A) = \frac{N_A}{N}$$

Statistical (empirical) definition of probability

Ex According to the data, out of 1,000 newborns, 985 live beyond the first year.

What is the survival probability in one year?

$$P = \frac{980}{1000} = 0.98$$

Ex A dice was rolled 10,000 times out of which 1,650 times the outcome 'one' was obtained.

What is the probability of the outcome 'one'?

$$P = \frac{1650}{10000} = 0.165 \approx \frac{1}{6}$$

Theorems of Probability

- For an arbitrary event A, sample space S and empty event ϕ:

$$0 \leq P(A) \leq 1$$

$$P(S) = 1$$

$$P(\phi) = 0$$

- Given the elementary events e_i and the corresponding probabilities p_i, the "normalization" holds:

$$p_1 + p_2 + \cdots + p_N = 1$$

Unit 1.

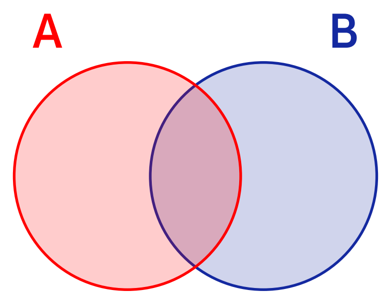
Understanding of Probability

Probability Rules

Sum of probabilities

- For arbitrary events A and B:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$ "**Sum rule**"



Sum of probabilities

Ex When we roll a dice, what is the probability that the outcome is odd **or** larger than three?

If the event $A =$ 'odd outcome' and $B =$ 'outcome larger than three,'

$$A = \{1, 3, 4\}$$

$$B = \{4, 5, 6\}$$

a) The union of A and B is the event we are interested in.

So,

$$A \cup B = \{1, 3, 4, 5, 6\} \rightarrow P(A \cup B) = \frac{5}{6}$$

Sum of probabilities

Ex When we roll a dice, what is the probability that the outcome is odd **or** larger than three?

If the event $A =$ 'odd outcome' and $B =$ 'outcome larger than three,'

$$A = \{1, 3, 4\}$$

$$B = \{4, 5, 6\}$$

b) We have $P(A) = \frac{3}{6}$, $P(B) = \frac{3}{6}$, $P(A \cap B) = P(\{4\}) = \frac{1}{6}$.

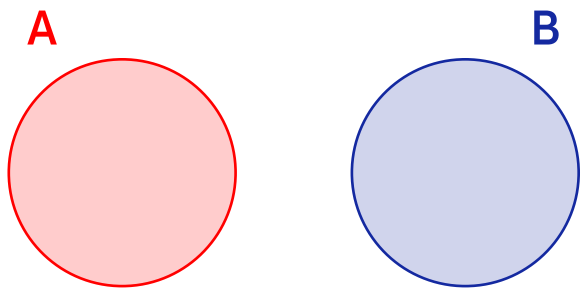
Using the sum rule,

$$P(A \cup B) = P(A) + P(B) - \mathrm{P}(A \cap B) = \frac{3}{6} + \frac{3}{6} - \frac{1}{6} = \frac{5}{6}$$

Sum of probabilities

- If the events A and C are **mutually exclusive** ("$A \cap B = \phi$"):

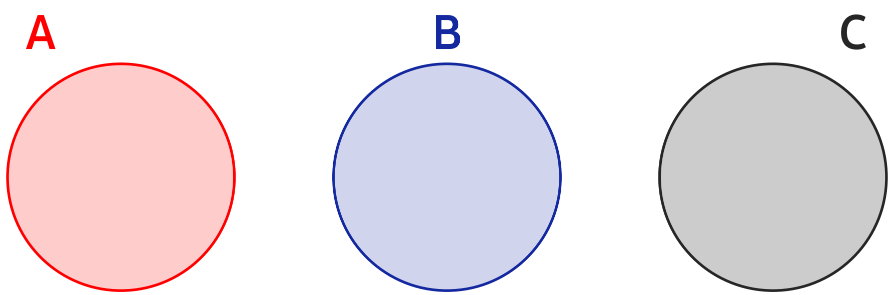
$$P(A \cup B) = P(A) + P(B)$$



Sum of probabilities

- If the events A, B, and C are **mutually exclusive** ("$A \cap B = \phi$," "$A \cap C = \phi$," "$B \cap C = \phi$"):

$$P(A \cup B \cup C) = P(A) + P(B) + P(C)$$



Sum of probabilities

Ex Out of 100 people, 30 are of blood type A, 28 of B, 17 of O, and 25 of AB.

If we randomly pick one person, what is the probability of that person's blood type being A or AB?

a) Blood types are mutually exclusive events.

So,

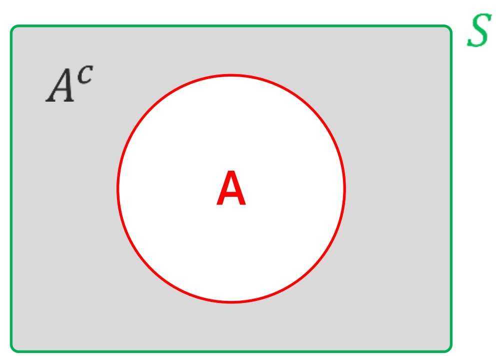
$$P(A \cup AB) = P(A) + P(AB) = \frac{30}{100} + \frac{25}{100} = \frac{11}{20}$$

Complementary events

- For an event A and its complementary event A^c, the following relation holds:

$$A \cup A^c = S \Rightarrow P(A) + P(A^c) = 1$$

$$\Rightarrow P(A) = 1 - P(A^c)$$



Complementary events

Ex When we flip five coins, what is the probability that at least one comes out head?

a) If the event A= 'at least one coin comes out head,' its complementary is

A^c= 'all the coins come out tail.'

Thus,

$$P(A^c) = \frac{1}{2^5} = \frac{1}{32}$$

b) Between the event and its complementary, we have $P(A)$+$P(A\hat{}c$)=1.

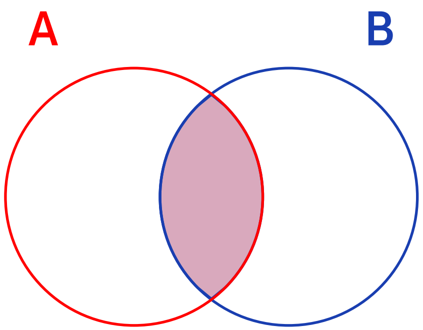
So,

$$P(A) = 1 - P(A^c) = 1 - \frac{1}{32} = \frac{31}{32}$$

Complementary events

- Suppose that you have the events A and B of which probabilities are non-zero.
- The probability of the B conditional on the A is denoted by $P(B|A)$.
- The following relation holds among the probabilities.

$$P(B|A) = \frac{P(A \cap B)}{P(A)}$$



- We can also rearrange the above relation as following.

$$\begin{aligned} P(A \cap B) &= P(A|B)P(B) \\ &= P(B|A)P(A) \end{aligned}$$

Conditional probability

Ex In a sack, there are two white balls and three red balls.

Suppose that two balls are taken out one after another.

Calculate the probability of both balls being white.

a) Suppose that the first ball is **not** put back into the sack.

Let us denote by A = event that the first ball is white, and

B = event that the second ball is white.

$$P(A \cap B) = P(B|A)P(A) = \frac{1}{4} \times \frac{2}{5} = \frac{1}{10}$$

Conditional probability

Ex In a sack, there are two white balls and three red balls.

Suppose that two balls are taken out one after another.

Calculate the probability of both balls being white.

b) Suppose that the first ball is put back into the sack before the second ball is withdrawn.

This is like a “reset.” The events A and B are independent from each other.

$$P(A \cap B) = P(B|A)P(A) = P(B)P(A) = \frac{2}{5} \times \frac{2}{5} = \frac{4}{25}$$

For the given events A and B, you can think of the following cases.

1) If the events A and B are dependent on each other, then:

$$P(A \cap B) = P(A|B)P(B)$$
$$= P(B|A)P(A)$$

and

$$P(A \cap B) \neq P(A)P(B)$$

2) If the events A and B are independent from each other, then:

$$P(A \cap B) = P(B)P(A)$$

3) If the events A and B are mutually exclusive, then:

$$P(A \cap B) = 0$$

Bayes' Theorem

- Let us remember the following relation:

$$\begin{aligned} P(A \cap B) &= P(A|B)P(B) \\ &= P(B|A)P(A) \end{aligned}$$

- Using the above relation, we can state the so-called Bayes' theorem.

$$P(A|B)P(B) = P(B|A)P(A)$$

- We may transform the Bayes' theorem in the following way:

$$P(A|B)P(B) = P(B|A)P(A)$$

$$\downarrow$$

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

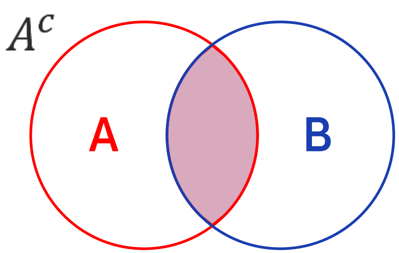
$$= \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|A^c)P(A^c)}$$

- We may transform the Bayes' theorem in the following way:

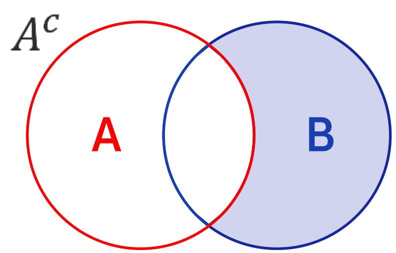
$$P(A|B) = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|A^c)P(A^c)}$$

- We may transform the Bayes' theorem in the following way:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|A^c)P(A^c)}$$



$P(B \cap A)$

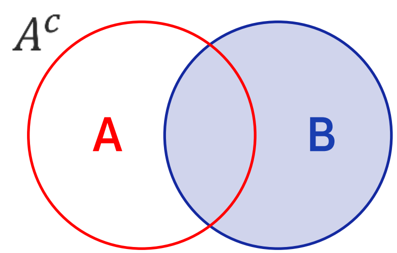


$P(B \cap A^c)$

- We may transform the Bayes' theorem in the following way:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|A^c)P(A^c)}$$

$$\|$$



$P(B)$

Bayes' theorem example

Ex There are 100 coins out of which one is abnormal with both sides as heads (H).

The remaining 99 coins are normal with head (H) on one side and tail (T) on the other.

A coin is randomly drawn and then thrown five times in a row.

In all five times, the coin lands showing the head side only.

What is the probability that this coin is that abnormal one?

a) Let us define as A the event that the coin is abnormal.

Then A^c is the event that the coin is normal.

Let us also define B the event that the coin lands five times showing the head side.

Bayes' theorem example

Ex There are 100 coins out of which one is abnormal with both sides as heads (H).

The remaining 99 coins are normal with head (H) on one side and tail (T) on the other.

A coin is randomly drawn and then thrown five times in a row.

In all five times, the coin lands showing the head side only.

What is the probability that this coin is that abnormal one?

b) The answer we seek is given by $P(A|B)$:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|A^c)P(A^c)} = \frac{1 \times \frac{1}{100}}{1 \times \frac{1}{100} + \left(\frac{1}{2}\right)^5 \times \frac{99}{100}} \cong \mathbf{0.244}$$

Bayes' theorem example

Ex A particular type of cancer affects 0.1% of the population.

A new diagnostic method has been devised.

→ This method can diagnose as positive (+) in 99% of the true cases (D).

→ This method can diagnose as negative (-) in 95% of the false cases (D^c).

This diagnostic method was administered to a patient giving the positive (+) result.

What is the probability that this patient actually has the cancer?

a) We have $P(D) = 0.001$, $P(D^c) = 0.999$, $P(+|D) = 0.99$ and $P(-|D^c) = 0.95$.

We can also derive that $P(+|D^c) = 1 - P(-|D^c) = 0.05$.

Bayes' theorem example

Ex A particular type of cancer affects 0.1% of the population.

A new diagnostic method has been devised.

→ This method can diagnose as positive (+) in 99% of the true cases (D).

→ This method can diagnose as negative (-) in 95% of the false cases (D^c).

This diagnostic method was administered to a patient giving the positive (+) result.

What is the probability that this patient actually has the cancer?

b) The answer we seek is given by $P(D|+)$:

$$P(D|+) = \frac{P(+|D)P(D)}{P(+|D)P(D) + P(+|D^c)P(D^c)} = \frac{0.99 \times 0.001}{0.99 \times 0.001 + 0.05 \times 0.999} \cong \mathbf{0.02}$$

Unit 1.

Understanding of Probability

Random Variable

- Random variable: a function that assigns a number to the outcome of a random experiment.

  Ex In a random experiment of coin flipping, assign 1 to the head (H) and assign 0 to the tail (T).

- There are two types of random variables:

  1) **Discrete:** for random experiments with a finite number of the possible outcomes.

  Ex Random variable that represents coin flipping experiment. Possible outcomes $= \{H, T\}$.

  Ex Random variable that represents dice rolling experiment. Possible outcomes $= \{1,2,3,4,5,6\}$.

  2) **Continuous:** for random experiments with an infinite number of the possible outcomes.

  Ex Random variable that represents the heights of people.

  Ex Random variable that represents the wages of people.

Discrete probability distribution function $P(x)$:

- Maps the values of a discrete random variable to the corresponding probabilities.
- You will denote in upper case a random variable and in lower case a particular value of it.

Ex Given a random variable X,
the probability of X taking on a value x is denoted by $P(X=x)$ or $P(x)$.

Continuous probability density function $f(x)$:

- When integrated, gives the interval probabilities. $\Leftarrow$ More about this in the next unit.

Properties of **discrete** probability distribution

1) $0 \leq P(x) \leq 1$

2) $\sum_{all\ x_i} P(x_i) = 1$

Population and Sample

Population

- The entirety of the data set is subject to the analysis.
- It can be either "real" or "idealized."
- The properties of a population are called parameters.

Ex Mean, standard deviation, variance, etc. of a population

Sample

- It is a subset of the population.
- The properties of a sample are called statistics.

Ex Mean, standard deviation, variance, etc. of a sample

- We could calculate the mean, standard deviation, variance, etc. using a probability distribution function.
  - These quantities are properties of an "idealized" group.
  - They can be interpreted as parameters of an idealized population.

Population Mean

Properties of the population mean

- It is also called the **expected value** of a random variable: $E[X]$.

1) For the discrete case with $P(x)$ = probability distribution function:

$$\mu = E[X] = \sum_{all\ x} x\, P(x)$$

2) For the continues case with $f(x)$ = probability density function:

$$\mu = E[X] = \int x\, f(x) dx$$

Properties of the population mean

1) $E[c] = c$

2) $E[c\,X] = c\,E[X]$

3) $E[X + c] = E[X] + c$

4) $E[X + Y] = E[X] + E[Y]$

Note: c stands for a constant.

Population Variance

- The population variance is denoted by σ^2 or $Var(X)$.

1) For the discrete case with $P(x)$ = probability distribution function:

$$\sigma^2 = Var(X) = \sum_{all\ x} (x - \mu)^2 \ P(x)$$

2) For the continues case with $f(x)$ = probability density function:

$$\sigma^2 = Var(X) = \int (x - \mu)^2 f(x) dx$$

Note: In both cases, $\sigma^2 = E[X^2] - (E[X])^2$.

Note: The population standard deviation: $\sigma = \sqrt{\sigma^2}$.

Properties of the population variance

1) $Var(c) \ = 0$

2) $Var(X + c) = Var(X)$

3) $Var(c\ X) = c^{2}\ Var(X)$

4) $Var(X + Y) = Var(X) + Var(Y) + 2\ Cov(X, Y)$

$$= Var(X) + Var(Y)$$ $\Leftarrow$ Only when X and Y are independent to each other!

Note: c stands for a constant.

Unit 1.

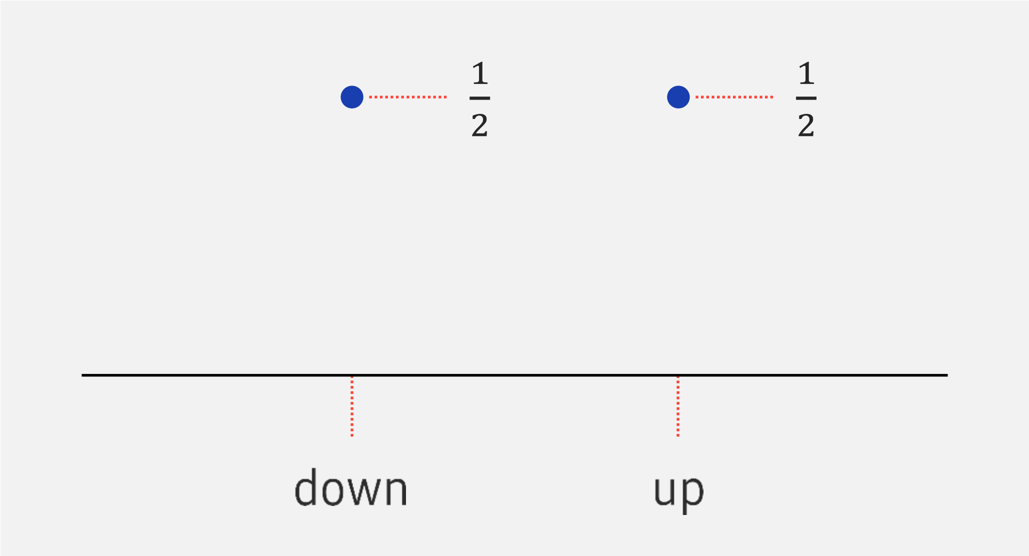
Understanding of Probability

Discrete Probability Distribution

Bernoulli random variable and probability distribution

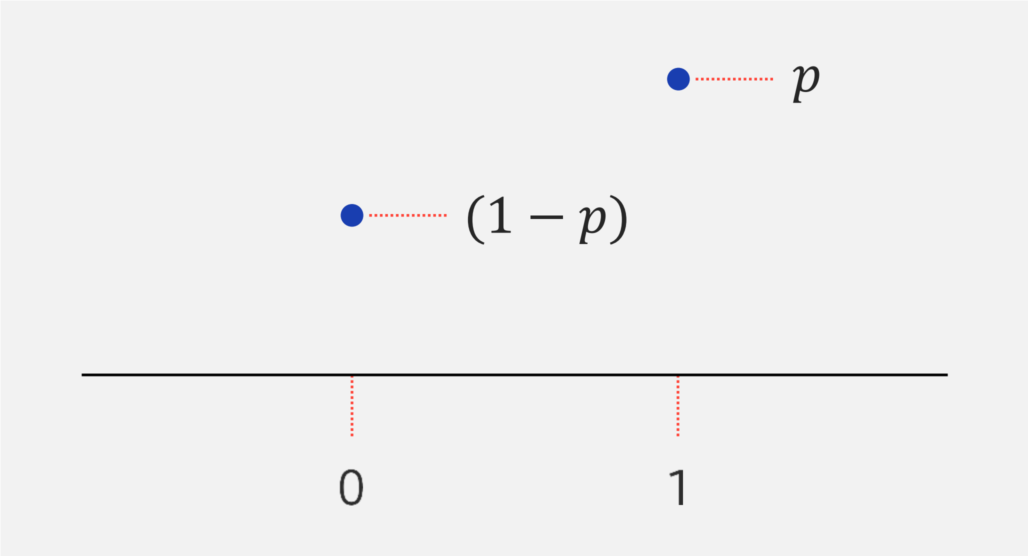
Bernoulli random variable and probability distribution

Ex Coin flipping



Bernoulli random variable and probability distribution

Ex General binary outcome situation



Bernoulli random variable and probability distribution

1) "X is a Bernoulli random variable." $\Leftrightarrow X \sim Ber(p)$

2) $P(x) = p^x(1-p)^{1-x}$

- Mean $= p$
- Variance $= p\,(1-p)$
- Standard deviation $= \sqrt{p\,(1-p)}$

Bernoulli random variable and probability distribution

Bernoulli random variable and probability distribution

1) "X is a Binomial random variable." $\Leftrightarrow X \sim Bin(n, p)$

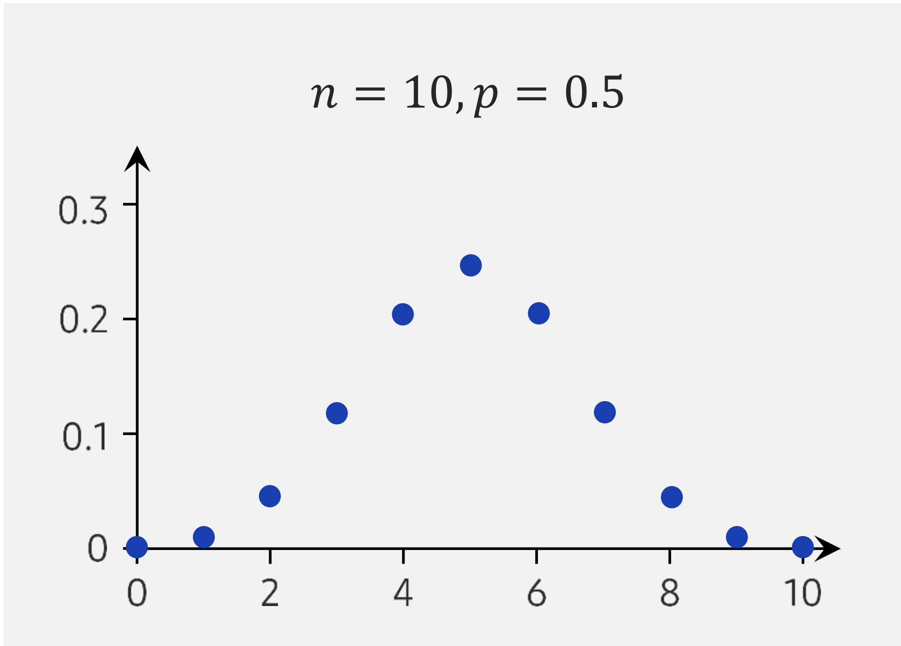
2) $X_{bin} = X_{Ber} + X_{Ber} + \cdots + X_{Ber}$

$\leftarrow \quad \mathbf{n} \quad \rightarrow$

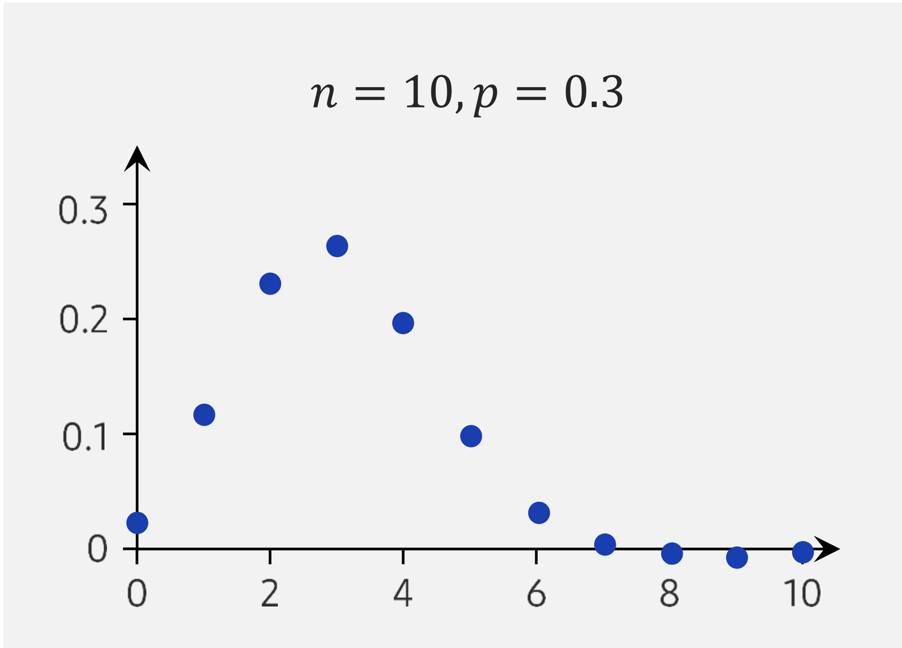
3) $P(x) = \binom{n}{x} p^x (1-p)^{n-x} \quad \Leftarrow \quad 0 \le x \le n$

- Mean $= n\,p$
- Variance $= n\,p\,(1-p)$
- Standard deviation $= \sqrt{n\,p\,(1-p)}$

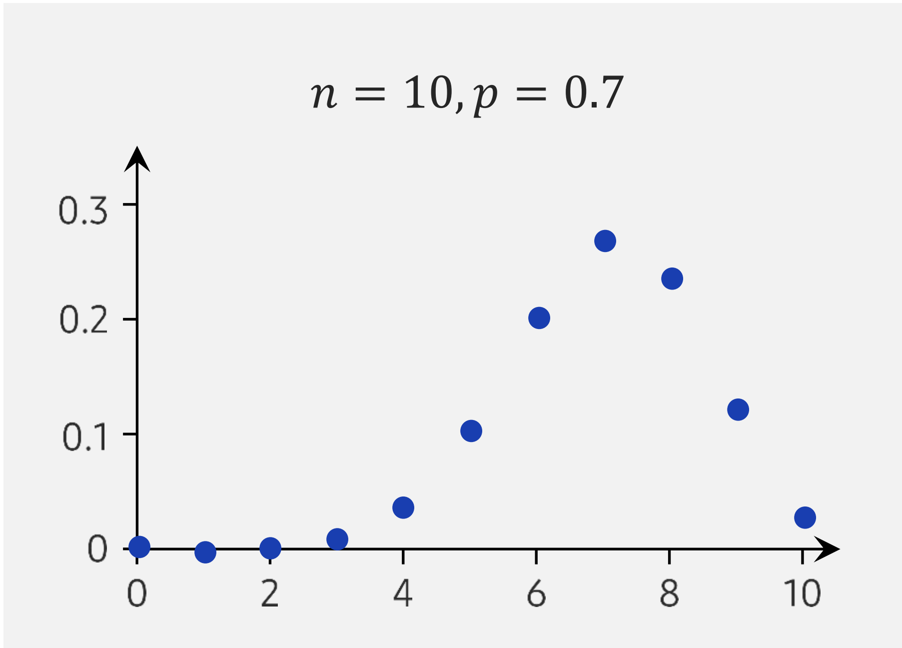
Bernoulli random variable and probability distribution



Bernoulli random variable and probability distribution



Bernoulli random variable and probability distribution



Bernoulli random variable and probability distribution

Ex It is known that the daily probability of raining is 20% for the following five days.

What is the probability of **zero** rainy day?

$$P(0) = \binom{5}{0} 0.2^0 (1-0.2)^{\ 5-0} = \frac{5!}{0!\ 5!} 0.8^5 = 0.8^5 = 0.328$$

Bernoulli random variable and probability distribution

Ex It is known that the daily probability of raining is 20% for the following five days.

What is the probability of **exactly two** rainy days?

$$P(2) = \binom{5}{2} 0.2^2(1-0.2)^{5-2} = \frac{5!}{2!\,3!} 0.2^2 \times 0.8^3 = 10 \times 0.2^2 \times 0.8^3 = 0.205$$

Bernoulli random variable and probability distribution

Ex It is known that the daily probability of raining is 20% for the following five days.

What is the probability of **two or fewer** rainy days?

$$P(X \le 2) = P(0) + P(1) + P(2)$$

$$= \binom{5}{0} 0.2^0 (1-0.2)^{5-0} + \binom{5}{1} 0.2^1 (1-0.2)^{5-1} + \binom{5}{2} 0.2^2 (1-0.2)^{5-2}$$

$$= 1 \times 1 \times 0.8^5 + 5 \times 0.2 \times 0.8^4 + 10 \times 0.2^2 \times 0.8^3$$

$$= 0.328 + 0.41 + 0.205$$

$$= 0.942$$

Poisson random variable and probability distribution

- It is named after the French scientist Simeon D. Poisson.
- It describes the frequency of events for a given interval of time or space.

Ex Number of emails received per hour

Ex Number of earthquakes per year

Ex Number of chocolate chips in a cookie

Poisson random variable and probability distribution

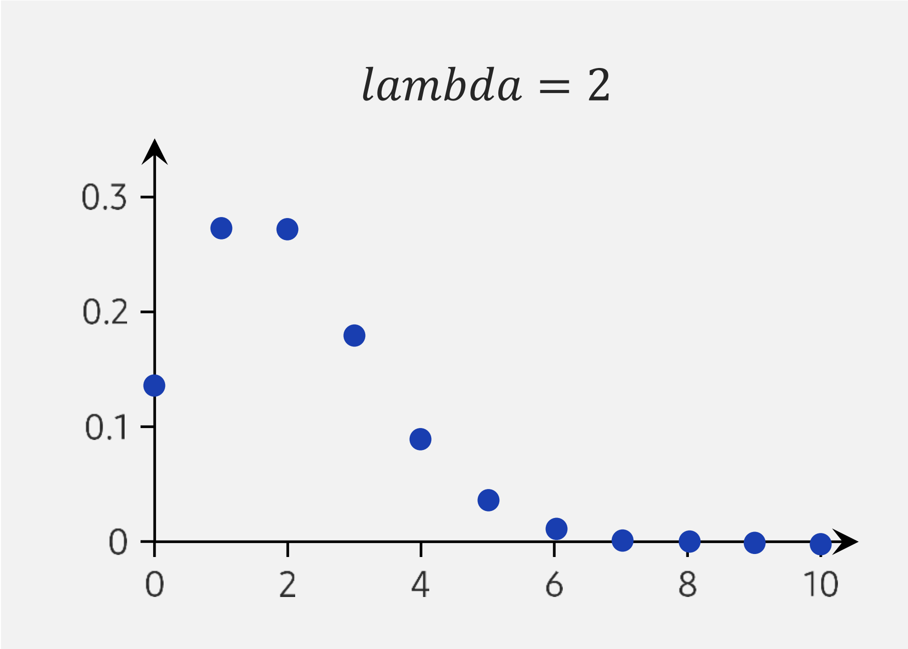
1) "X is a Poisson random variable." $\Leftrightarrow X \sim Pois(\lambda)$

2) $P(x) = \dfrac{\lambda^{x} e^{-\lambda}}{x!} \quad \Leftarrow \; 0 \leq x$

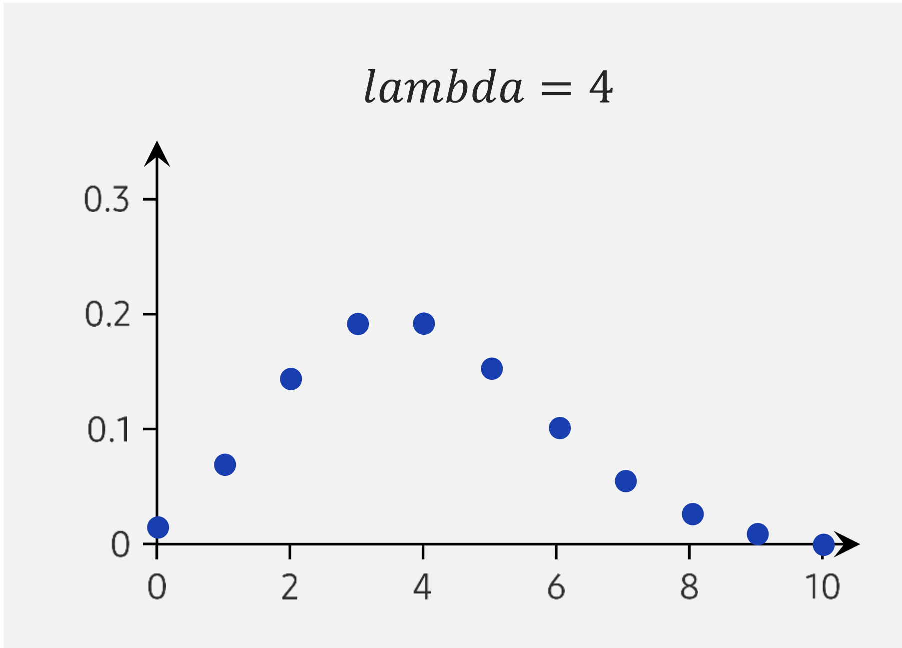
- Mean $= \lambda$
- Variance $= \lambda$
- Standard deviation $= \sqrt{\lambda}$

3) By increasing n while decreasing p, the Binomial distribution converges to the Poisson with $\lambda = n\,p$.

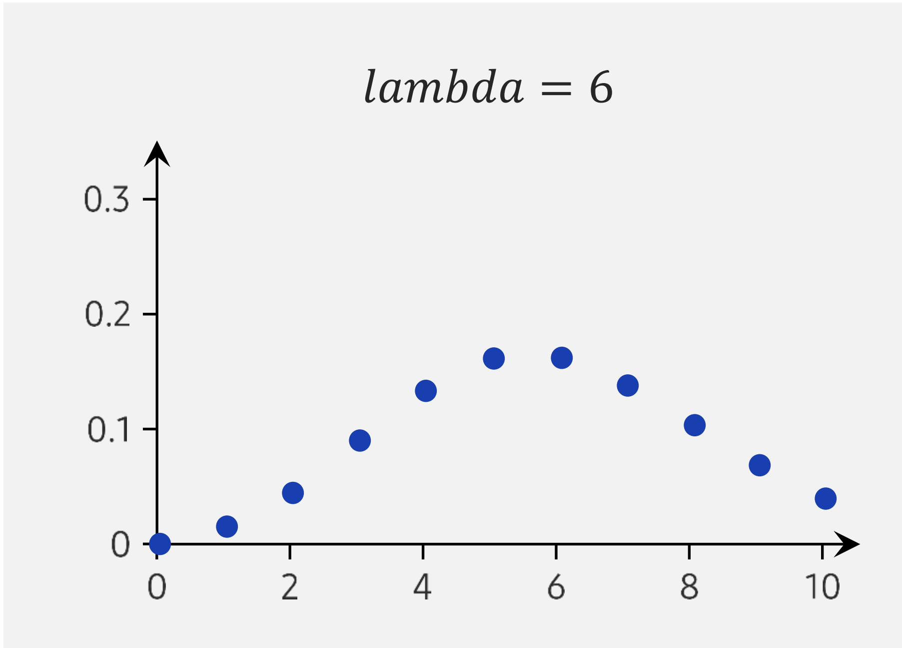
Poisson random variable and probability distribution



Poisson random variable and probability distribution



Poisson random variable and probability distribution



Poisson random variable and probability distribution

Ex During the last 100 days, 40 spam mails were received.

What is the probability that there will be zero spam mail tomorrow?

You can calculate that $\lambda = \frac{40}{100} = 0.4$.

Then,

$$P(0) = \frac{0.4^0 e^{-0.4}}{0!} = e^{-0.4} = 0.6703$$

Poisson random variable and probability distribution

Ex During the last 100 days, 40 spam mails were received.

What is the probability that there will be more than one spam mails tomorrow?

You can calculate that $\lambda = \frac{40}{100} = 0.4$.

Then, do $P(1) + P(2) + P(3) + \cdots$??. $\Rightarrow$ Hard!

Instead, you can calculate by doing $1 - P(0) \; = \; 1 - 0.6703 \; = \; 0.3297$.

Coding Exercise #0201

Follow practice steps on 'ex_0201.ipynb' file.

Unit 2.

Understanding of Statistics I

Continuous Probability Density

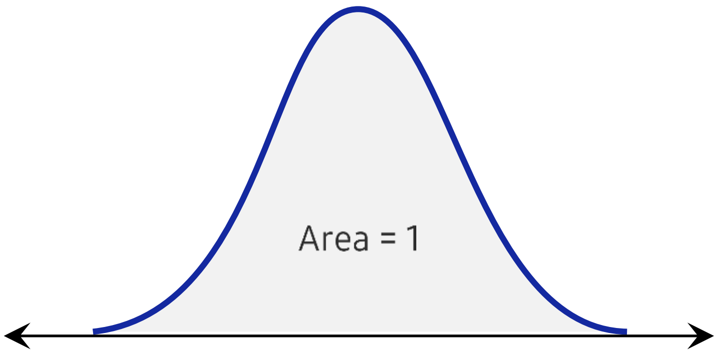
Continuous random variable and probability density

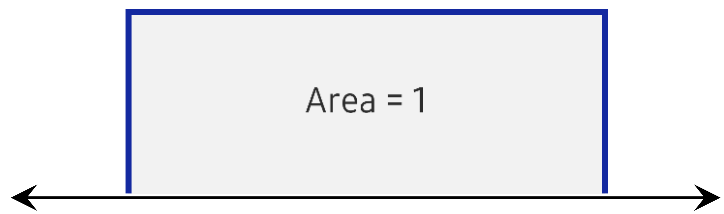
- Infinite number of possible values
- The probability at a specific value is zero: $P(X = \boldsymbol{x_0}) = 0$
- Non-zero probability only for intervals: $P(x_1 \leq X \leq x_2) = \int_{x_1}^{x_2} f(x)dx$
- Cumulative probability: $CDF(x) = P(-\infty < X \leq x)$
  $= \int_{-\infty}^{x} f(y)dy$
- We can calculate $P(x_1 \leq X \leq x_2) = CDF(x_2) - CDF(x_1)$.

Properties of **continuous** probability density

1) $0 \leq f(x)$

2) $\int f(x)dx = 1$





Uniform random variable and probability density

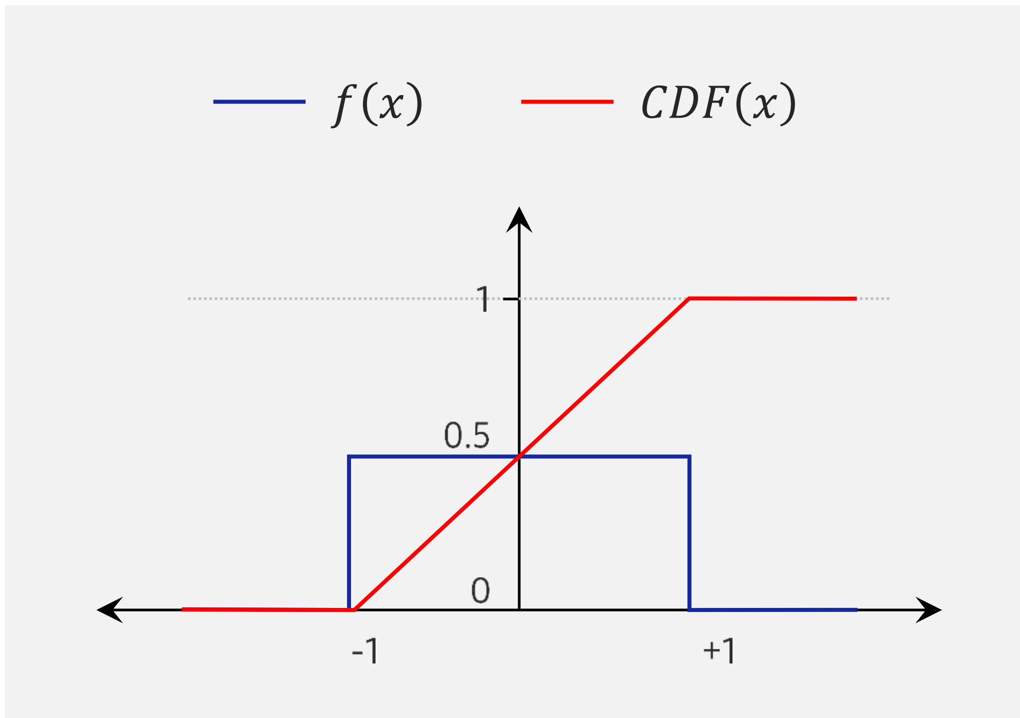
1) "X is a uniform random variable in the interval $[a, b]$." $\Leftrightarrow X \sim Unif(a, b)$

2) $f(x) = \frac{1}{(b-a)}$ $\Leftarrow$ defined in the interval $[a, b]$ and zero elsewhere

- Mean $= \frac{1}{2}(a + b)$
- Variance $= \frac{1}{12}(b - a)^2$
- Standard deviation $= \frac{1}{\sqrt{12}}(b - a)$

Uniform random variable and probability density

Ex $a = -1, b = +1$



Uniform random variable and probability density

Ex The lifespan of a light bulb is uniformly distributed between 10,000 and 20,000 hours.

What is the probability that the light bulb will last between 12,000 and 15,000 hours of usage?

$$P(12000 \leq X \leq 15000) = \int_{12000}^{15000} \frac{1}{20000 - 10000} dx = 0.3$$

Uniform random variable and probability density

Ex The lifespan of a light bulb is uniformly distributed between 10,000 and 20,000 hours.
What is the probability that the light bulb will last 15,000 hours or more?

$$P(15000 \leq X \leq 20000) = \int_{15000}^{20000} \frac{1}{20000 - 10000} dx = 0.5$$

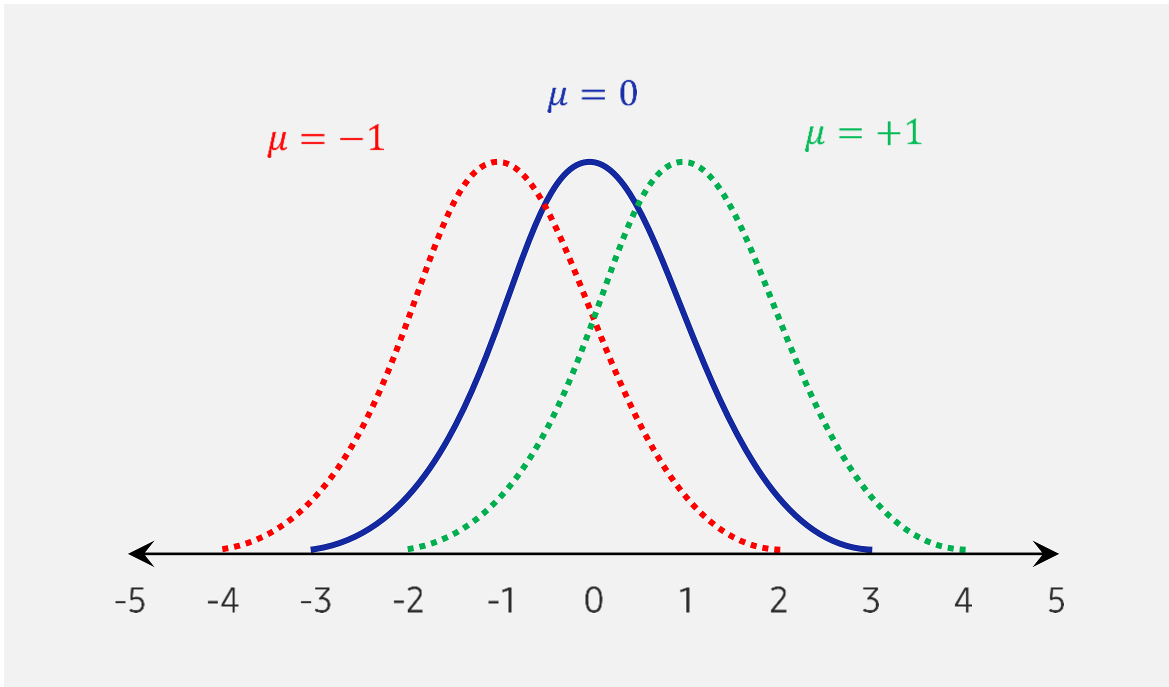
Normal random variable and probability density

1) "X is a normal random variable with mean μ and variance σ^2." $\Leftrightarrow X \sim N(\mu, \sigma^2)$

2) $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ $\Leftarrow$ defined in the interval $(-\infty, +\infty)$

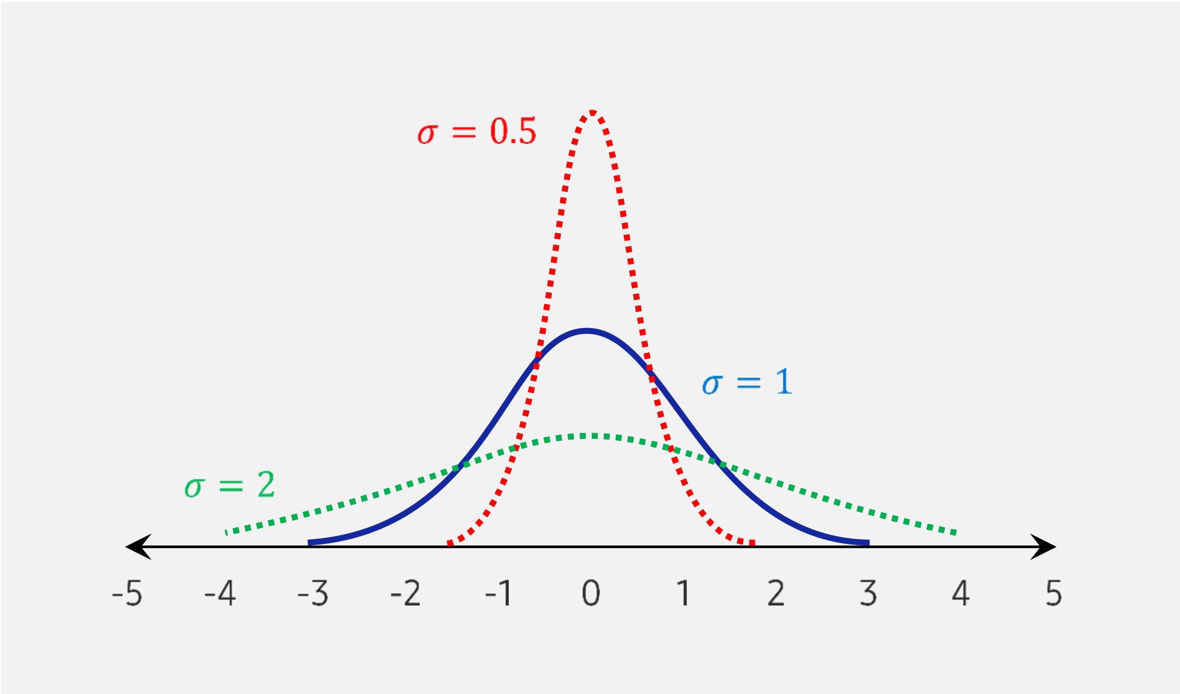
- Mean $= \mu$
- Variance $= \sigma^2$
- Standard deviation $= \sigma$

Normal random variable and probability density



- μ is the so-called "**location**" parameter.

Normal random variable and probability density



- σ is the so-called "**shape**" parameter.

Standard normal random variable and probability density

1) "X is a standard normal random variable." $\Leftrightarrow X \sim N(0,1)$

2) $f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ $\Leftarrow$ defined in the interval $(-\infty, +\infty)$

- Mean $= 0$
- Variance $= 1$
- Standard deviation $= 1$

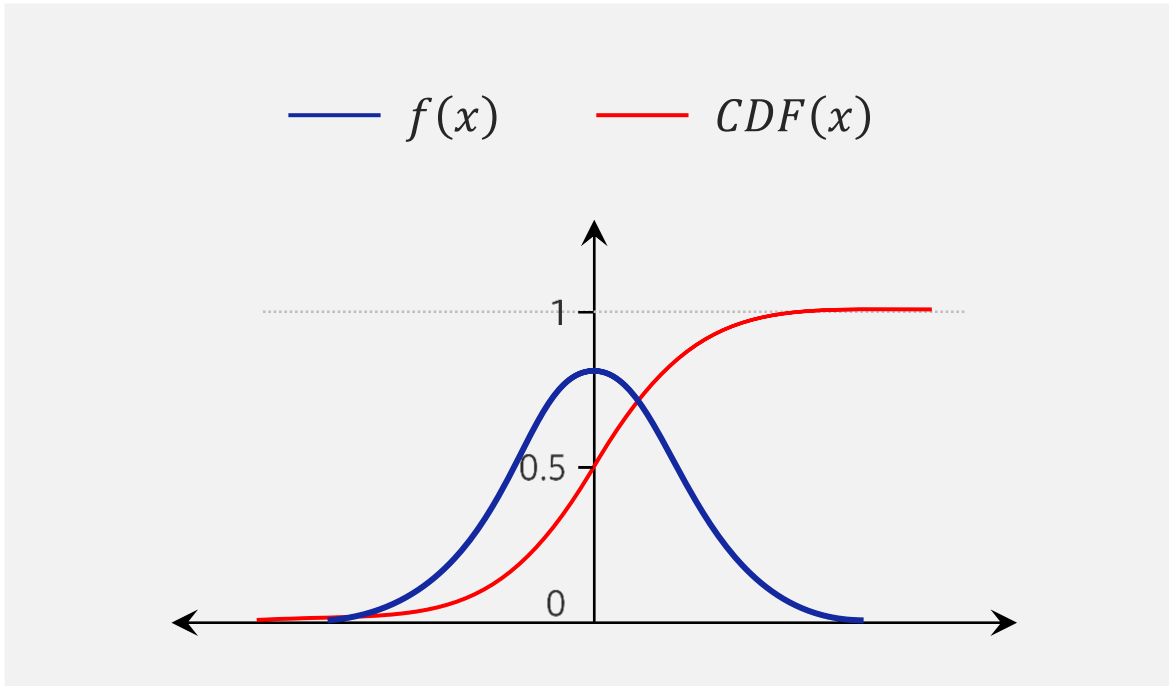
Normal random variable and probability density: "addition and subtraction"

- Suppose that $X \sim N(\mu_1, \sigma_1^2)$ and $Y \sim N(\mu_2, \sigma_2^2)$.
- And, also suppose that X and Y are independent to each other.

1) $X + Y \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$

2) $X - Y \sim N(\mu_1 - \mu_2, \sigma_1^2 + \sigma_2^2)$

Standard normal random variable and probability density



- Suppose that we have a random variable $X \sim N(\mu, \sigma^2)$, then it is possible to convert to another random variable Z such that $Z \sim N(0,1)$. This conversion process is called "standardization."

$$Z = \frac{X - \mu}{\sigma}$$

- Standardized values are called "z-scores":

$$z\text{−}score = \frac{x - \mu}{\sigma}$$

- It is possible to go in the opposite direction: from the standard normal to the normal.

$$X = \sigma\, Z + \mu$$

Standard normal random variable and probability density

Ex A barista takes in average 50 seconds to make a cup of coffee with the standard deviation of 20 seconds. What is the probability that the barista will take 48 ~ 54 seconds for your next cup of coffee? Use the table of the standard normal CDFs.

1) First of all, let us standardize:

$$x_1 = 48 \quad \Rightarrow z_1 = \frac{x_1 - \mu}{\sigma} = \frac{48-50}{20} = -\frac{2}{20} = -0.1$$

$$x_2 = 54 \quad \Rightarrow z_2 = \frac{x_2 - \mu}{\sigma} = \frac{54-50}{20} = \frac{4}{20} = 0.2$$

z	$\mathrm{CDF}(z)$
-0.2	0.4207
-0.1	0.4602
0	**0.5**
0.1	0.5398
0.2	0.5793

Standard normal random variable and probability density

Ex A barista takes in average 50 seconds to make a cup of coffee with the standard deviation of 20 seconds. What is the probability that the barista will take 48 ~ 54 seconds for your next cup of coffee? Use the table of the standard normal CDFs.

2) $P(z_1 \leq Z \leq z_2) = CDF(z_2) - CDF(z_1)$

$$= CDF(0.2) - CDF(-0.1)$$

$$= 0.5793 - 0.4602 = 0.1191$$

z	CDF(z)
-0.2	0.4207
-0.1	0.4602
0	**0.5**
0.1	0.5398
0.2	0.5793

Chi-square random variable and probability density

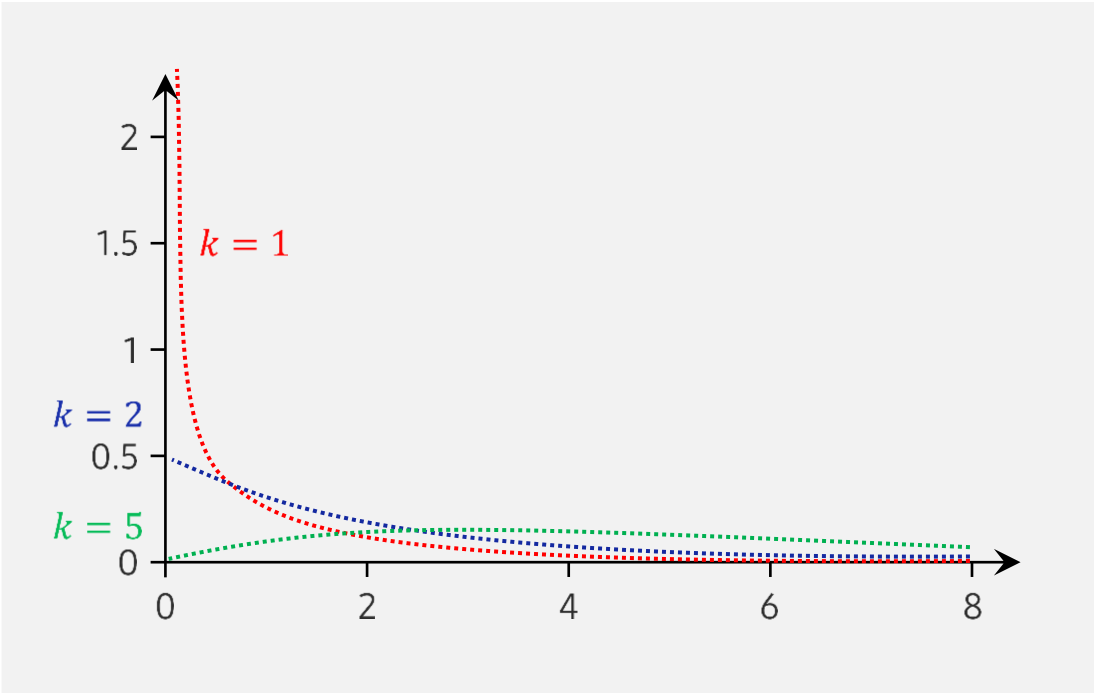
1) The chi-square random variable Q is the sum of squares of $\boldsymbol{k}$ independent standard normal random variables X_i: $Q = X_1^2 + X_2^2 + \cdots + X_k^2$.

2) "X is a chi-square random variable with the **degree of freedom $\mathbf{k}$**." $\Leftrightarrow Q \sim \chi^2(k)$

3) $f(x) = \frac{1}{2^{\frac{k}{2}}\Gamma(\frac{k}{2})} x^{\frac{k}{2}-1}$ $\Leftarrow$ defined in the interval $(0, +\infty)$

- Mean $= k$
- Variance $= 2k$
- Standard deviation $= \sqrt{2k}$

Chi-square random variable and probability density



Chi-square random variable and probability density: "addition"

Suppose that $Q_1 \sim \chi^2(\boldsymbol{k_1})$ and $Q_2 \sim \chi^2(\boldsymbol{k_2})$ which are independent to each other.

Then, $Q_1 + Q_2 \sim \chi^2(\boldsymbol{k_1} + \boldsymbol{k_2})$.

The proof is straightforward.

$$Q_1 + Q_2 = \underbrace{\{X^2 + X^2 + \cdots + X^2\}}_{\boldsymbol{k_1}} + \underbrace{\{X^2 + \cdots + X^2\}}_{\boldsymbol{k_2}}$$

Thus, by regrouping, $Q_1 + Q_2 \sim \chi^2(\boldsymbol{k_1} + \boldsymbol{k_2})$.

Student-t random variable and probability density

1) When $Q \sim \chi^2(\nu)$ and $Z \sim N(0,1)$, then the Student-t random variable is defined as following:

$$T = \frac{Z}{\sqrt{Q/\nu}}$$

2) "T is a Student-t random variable with the **degree of freedom** ν." $\Leftrightarrow T \sim t(\nu)$

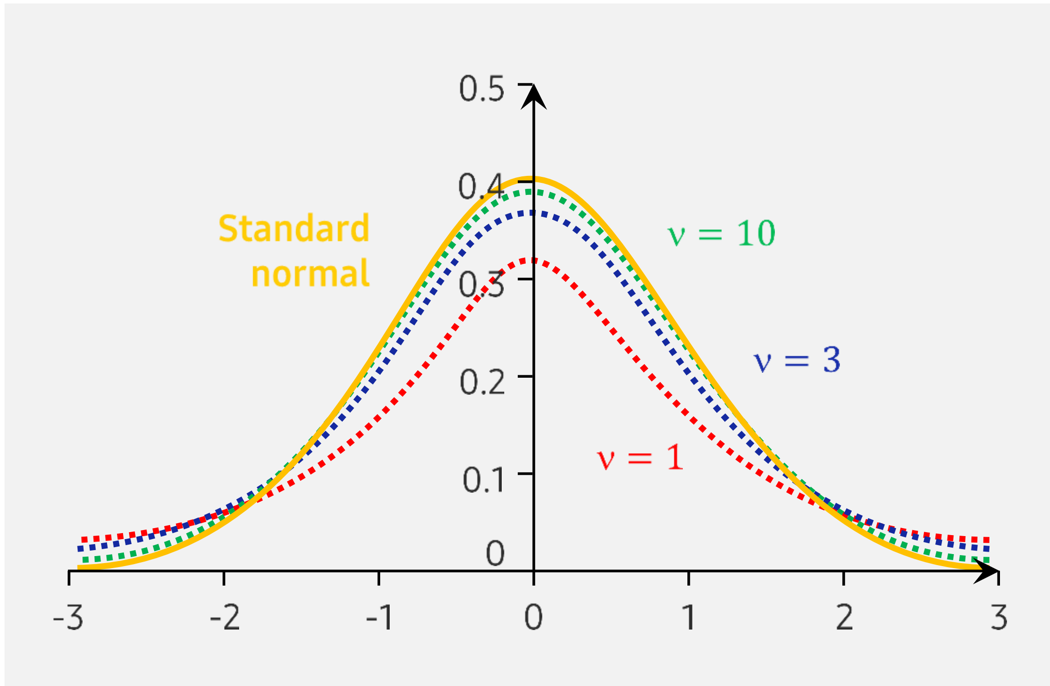
3) $f(x) = \dfrac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\,\Gamma\left(\frac{\nu}{2}\right)}\left(1 + \dfrac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}}$ $\Leftarrow$ defined in the interval $(-\infty, +\infty)$

- Mean $= 0$
- Variance $= \frac{\nu}{\nu-2}$ $\Leftarrow$ when $\nu > 2$
- Standard deviation $= \sqrt{\frac{\nu}{\nu-2}}$ $\Leftarrow$ when $\nu > 2$

Student-t random variable and probability density

- It is useful when doing the interval estimation with small sample size.
- It is also useful when we do the hypothesis testing of the means.
- The degree of freedom and the sample size are related: $\nu = n - 1$.
- As the degree of freedom (~sample size) increases, the student-t converges to the standard normal.

Student-t random variable and probability density



- As the degree of freedom ν increases, the student-t converges to the standard normal.

F random variable and probability density

1) When $Q_1 \sim \chi^2(d_1)$ and $Q_2 \sim \chi^2(d_2)$, then the F random variable is defined as following:

$$X = \frac{Q_1 / d_1}{Q_2 / d_2}$$

2) "X is a F random variable with the **degrees of freedom $\boldsymbol{d_1}$ and $\boldsymbol{d_2}$**." $\Leftrightarrow X \sim F(d_1, d_2)$

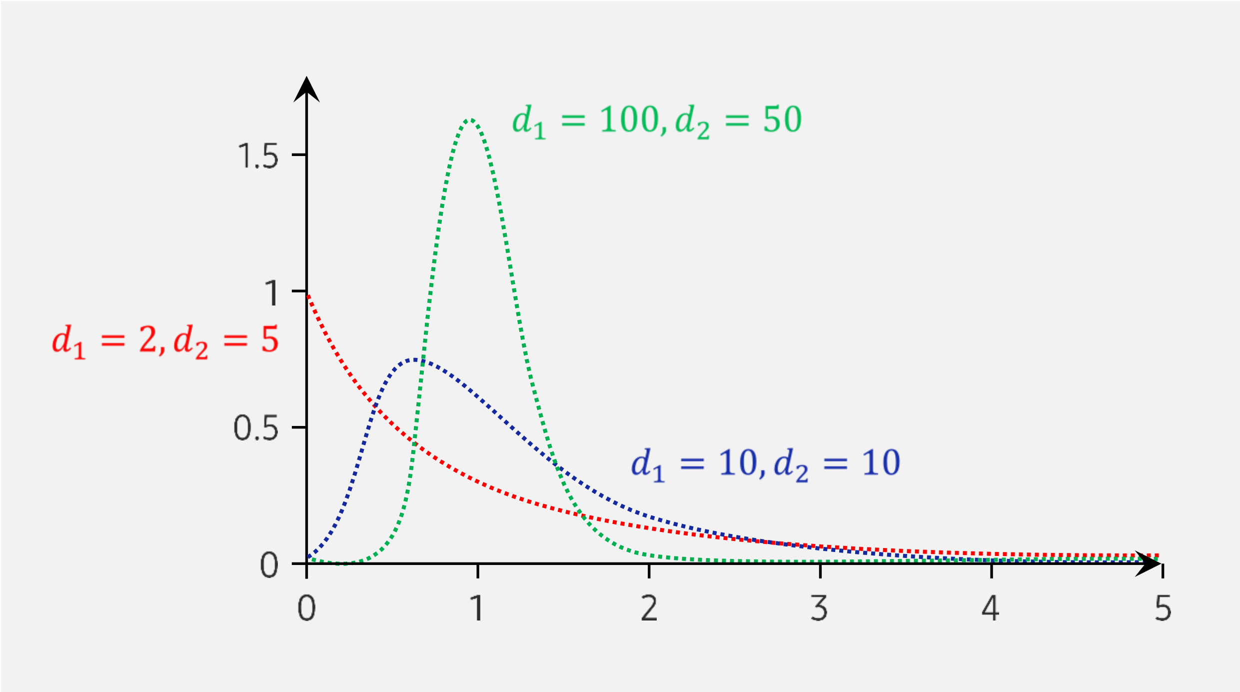
3) d_1 = degree of freedom for the numerator, d_2 = the degree of freedom for the denominator

- Mean $= \frac{d_2}{d_2-2}$ $\Leftarrow$ when $d_2 > 2$
- Variance $= \frac{2d_2^2(d_1+d_2-2)}{d_1(d_2-4)(d_2-2)^2}$ $\Leftarrow$ when $d_2 > 4$
- Standard deviation $= \sqrt{\frac{2d_2^2(d_1+d_2-2)}{d_1(d_2-4)(d_2-2)^2}}$ $\Leftarrow$ when $d_2 > 4$

F random variable and probability density

- It is useful when comparing variances of two samples.
- It is also used for ANOVA (Analysis of Variance) that compares the group means.
- As d_2 increases, the F random variable converges to ${}^{Q_1}/_{d_1}$.

F random variable and probability density



- There are two degrees of freedom: d_1 and d_2.

Coding Exercise #0202

Follow practice steps on 'ex_0202.ipynb' file.

Unit 2.

Understanding of Statistics I

2.1. Continuous Probability Density

2.2. Conjoint Probability

Conjoint Probability

- Bivariate conjoint probability: for two random variables X and Y.
- For the discrete case: $P(x, y) = P(X = x, Y = y)$
  - The probability when X = x AND Y = y
- For the continuous case: $f(x,y) = f(X{=}x, Y{=}y)$
  - The probability density when X = x AND Y = y

Covariance

- It is about two variables X and Y with an implicit conjoint probability.
- It can be calculated as expected value: $Cov(X,Y)=E[(X-\mu_x)(Y-\mu_y)]$.
- A rather "simpler" expression: $Cov(X,Y)=E[X\,Y]-E[X]\,E[Y]$
- When $X=Y$, we have $Cov(X,X)=Var(X)$.

Correlation

- Correlation and covariance are related:

$$Cor(X,Y) = \frac{Cov(X,Y)}{\sigma_X \sigma_Y}$$

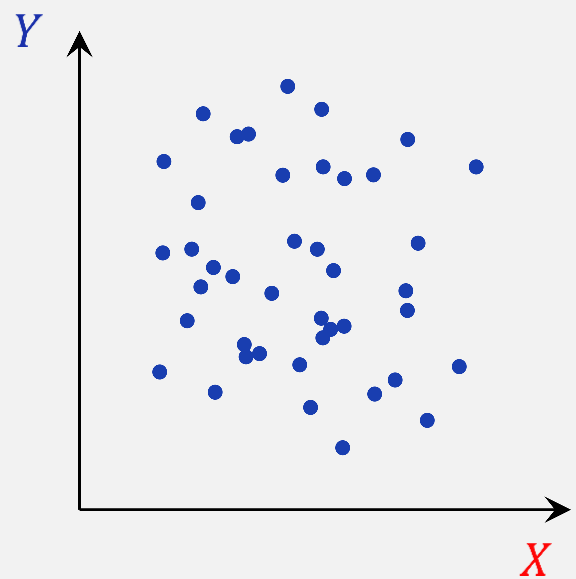
- Value in the interval [−1,1]
- Correlation is a measure of the linear relationship between X and Y.

a) $Cor(X,Y) > 0$: a positive linear relationship between X and Y

b) $Cor(X,Y) < 0$: a negative linear relationship between X and Y

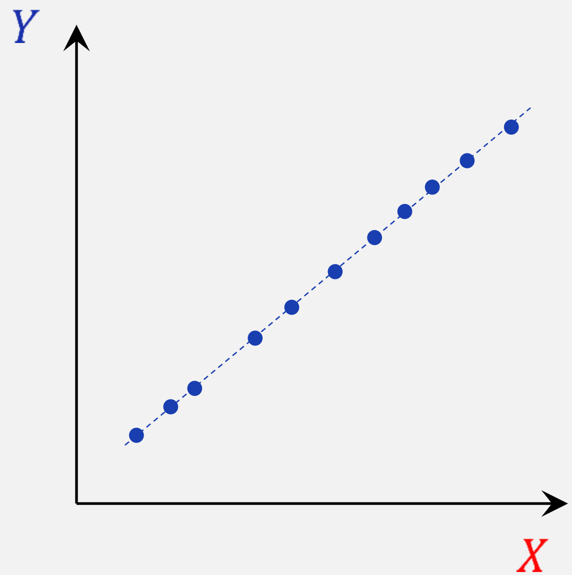
c) $Cor(X,Y) = 0$: no linear relationship between X and Y

Correlation

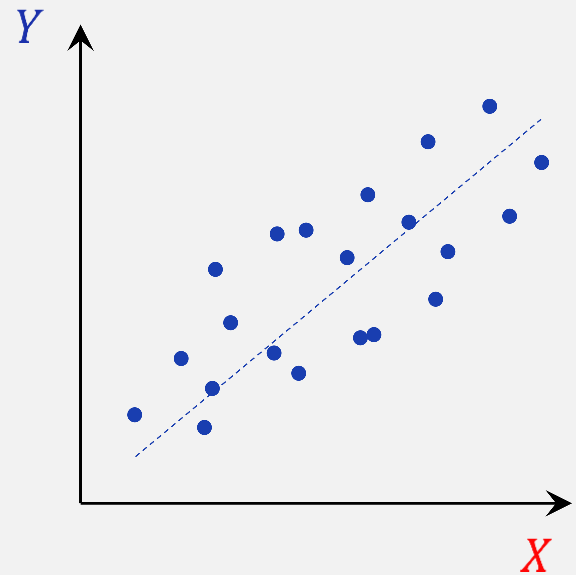


No linear relationship

Correlation

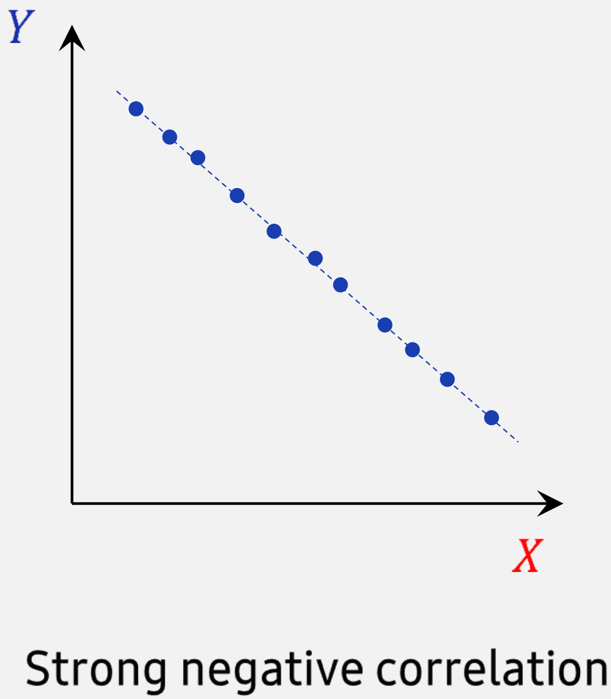


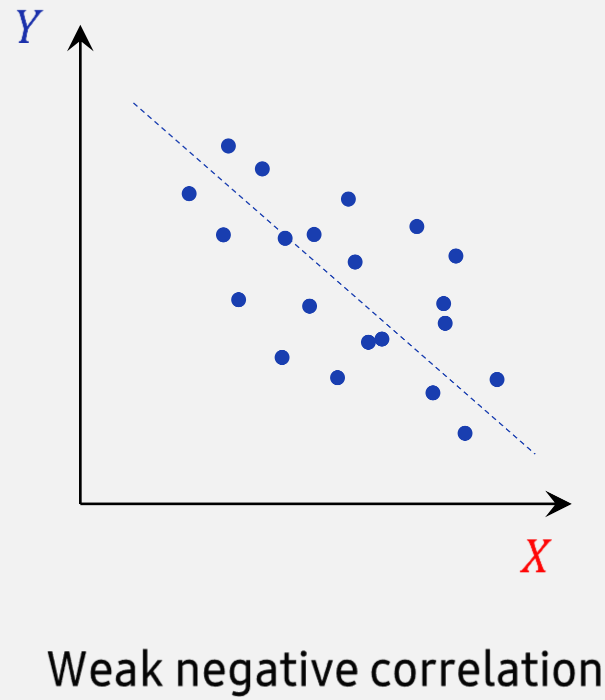
Strong positive correlation



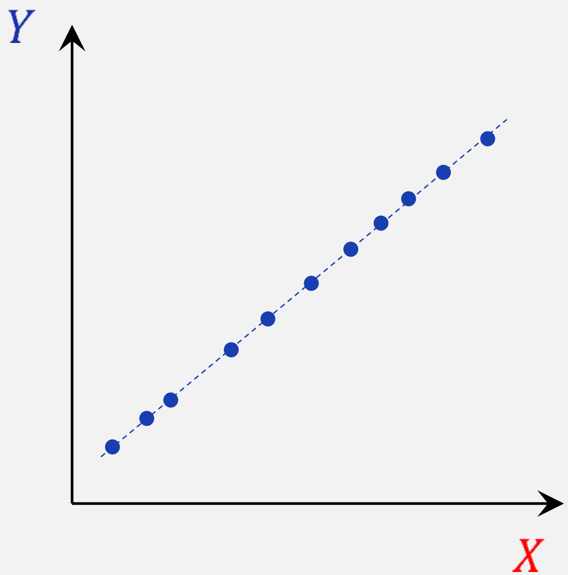
Weak positive correlation

Correlation

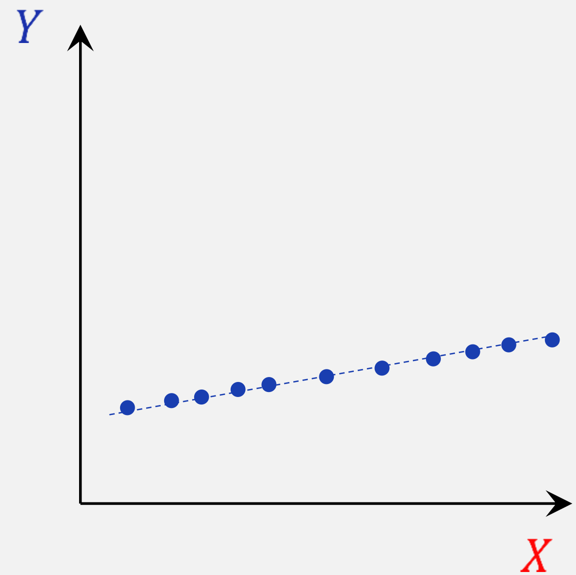




Correlation

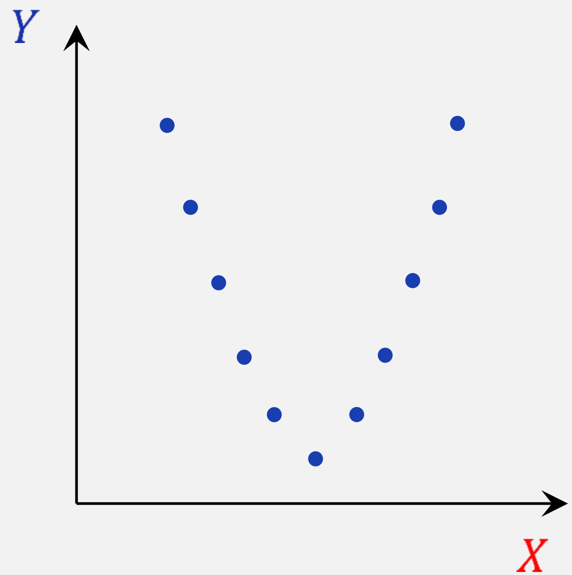


Strong positive correlation

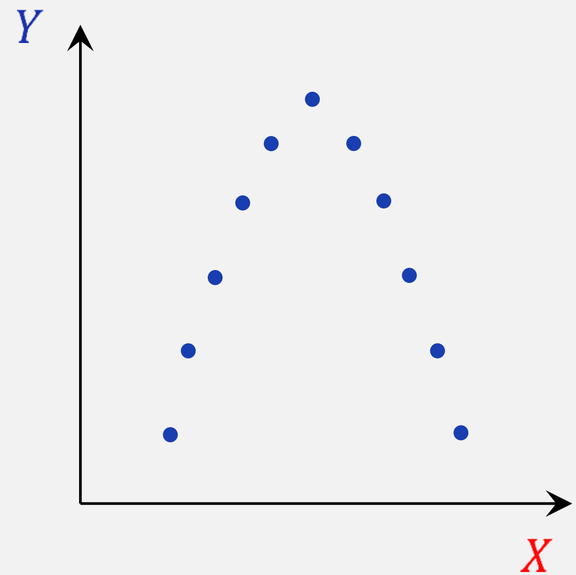


Strong positive correlation

Correlation



No linear relationship



No linear relationship

Dependence and correlation

- If X and Y are independent: $P(X,Y)=P(X)\ \ P(Y)$ or $f(X,Y)=f(X)\ \ f(Y)$.

  Then, $Cov(X,Y)=E[X\ Y]-E[X]E[Y]=E[X]E[Y]-E[X]E[Y]=0$.

  As $Cor(X,Y)=\frac{Cov(X,Y)}{\sigma_X\sigma_Y}$, we conclude that $Cor(X,Y)=0$.

  So, independence **always** means uncorrelatedness.

- However, X and Y being uncorrelated (*) cannot guarantee the independence all the time.

(*) that is $Cor(X,Y)=0$

Dependence and correlation

Ex Let's consider tossing two coins represented by the random variables X and Y.

The sample space is composed of HH, HT, TH, TT.

a) The individual probability distributions are:

$$\left.\begin{aligned} P(X{=}H) &= \frac{1}{2} \\ P(X{=}T) &= \frac{1}{2} \end{aligned}\right\} P(X)$$

$$\left.\begin{aligned} P(Y{=}H) &= \frac{1}{2} \\ P(Y{=}T) &= \frac{1}{2} \end{aligned}\right\} P(Y)$$

Dependence and correlation

Ex Let's consider tossing two coins represented by the random variables X and Y.

The sample space is composed of HH, HT, TH, TT.

b) In the conjoint probability, you can check that indeed X and Y are independent:

$$P(X{=}H, Y{=}H) = \frac{1}{4} = P(X{=}H) \times P(Y{=}H)$$

$$P(X{=}H, Y{=}T) = \frac{1}{4} = P(X{=}H) \times P(Y{=}T)$$

$$P(X{=}T, Y{=}H) = \frac{1}{4} = P(X{=}T) \times P(Y{=}H)$$

$$P(X{=}T, Y{=}T) = \frac{1}{4} = P(X{=}T) \times P(Y{=}T)$$

Dependence and correlation

Ex The random variable X has equal probability 1/3 for the each one of the outcomes -1, 0, 1.

And, there is another random variable Y defined as $Y=X^2$.

Then, the conjoint probabilities are as following:

$$P(X=-1, Y=1) = \frac{1}{3} \quad , \quad P(X=0, Y=0) = \frac{1}{3} \quad , \quad P(X=1, Y=1) = \frac{1}{3}$$

1) Let us calculate the $Cor(X,Y)$:

$$E[X]=-1\times P(X=-1)+0\times P(X=0)+1\times P(X=1)=-\frac{1}{3}+0+\frac{1}{3}=0$$

$$E[Y]=0\times P(Y=0)+1\times P(Y=1)=0+1\times\frac{2}{3}=\frac{2}{3}$$

$$E[X\ Y]=E[X\ X^2]=E[X^3]=-1\times P(X=-1)+0\times P(X=0)+1\times P(X=1)=0$$

So, $Cov(X,Y)=E[X\ Y]-E[X]\ E[Y]=0-0\times\frac{2}{3}=0$ **and** $Cor(X,Y)=0$, too.

Dependence and correlation

Ex The random variable X has equal probability 1/3 for the each one of the outcomes -1, 0, 1.

And, there is another random variable Y defined as $Y=X^2$.

Then, the conjoint probabilities are as following:

$$P(X=-1, Y=1) = \frac{1}{3}, P(X=0, Y=0) = \frac{1}{3}, P(X=1, Y=1) = \frac{1}{3}$$

2) Now that you know X and Y are uncorrelated, check for the independence.

$$P(X=-1, Y=1) = \frac{1}{3} \neq P(X=-1) \times P(Y=1) = \frac{1}{3} \times \frac{2}{3} = \frac{2}{9}$$

$$P(X=0, Y=0) = \frac{1}{3} \neq P(X=0) \times P(Y=0) = \frac{1}{3} \times \frac{1}{3} = \frac{1}{9}$$

$$P(X=1, Y=1) = \frac{1}{3} \neq P(X=1) \times P(Y=1) = \frac{1}{3} \times \frac{2}{3} = \frac{2}{9}$$

Summarizing $P(X,Y) \neq P(X)P(Y)$ so that X and Y are **not** independent.

Here, we found a case where X and Y are uncorrelated but not independent!

Unit 3.

Understanding of Statistics II

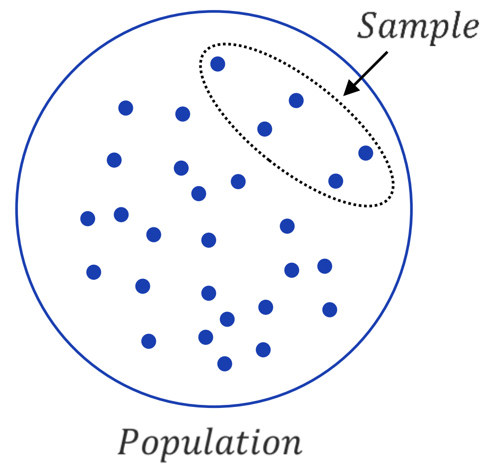
Population vs. Sample

Population

- Whole data set is subject to the analysis. It can be real or conceptual.

Sample

- It is a subset of the population.
- You analyze samples because the population is hard or impossible to reach.
- You would like to draw conclusions about the population using its samples.



Descriptive vs. Inferential

Descriptive statistics

- It summarizes the data without generalization as a primary goal.
- It extracts the properties of data as sample statistics.

Inferential statistics

- It is an analysis of the samples with the purpose of making generalized statements about the population.

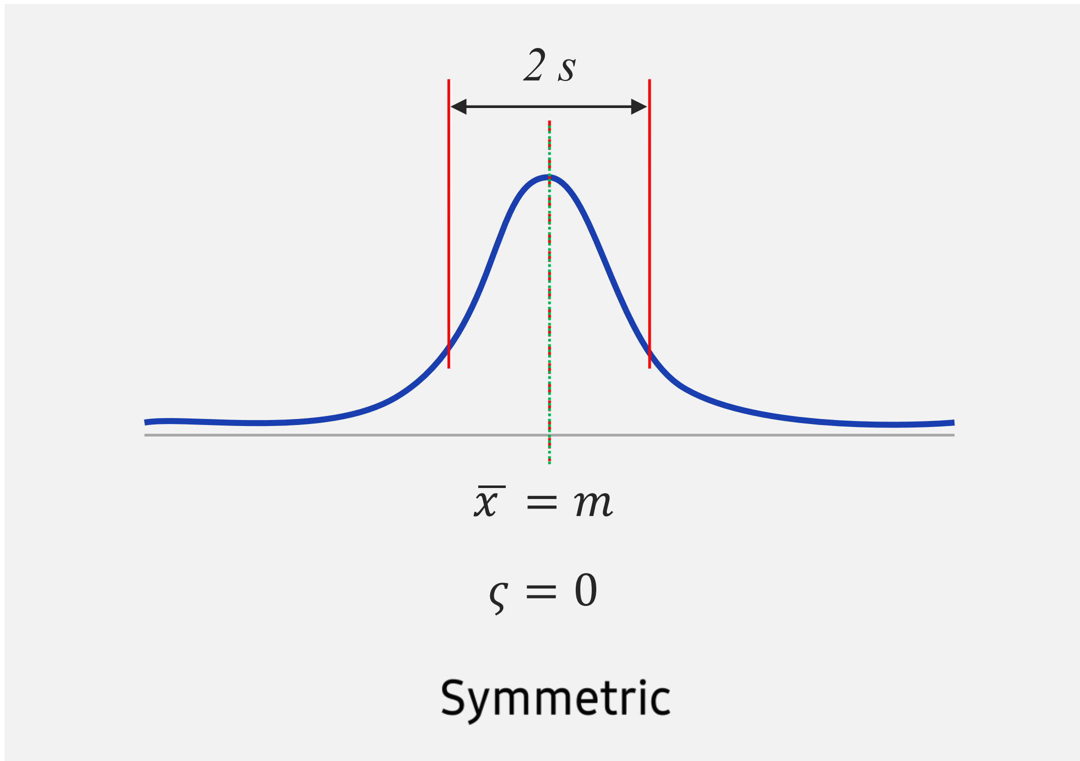
The main difference between descriptive and inferential statistics lies in the goal.

Descriptive Statistics

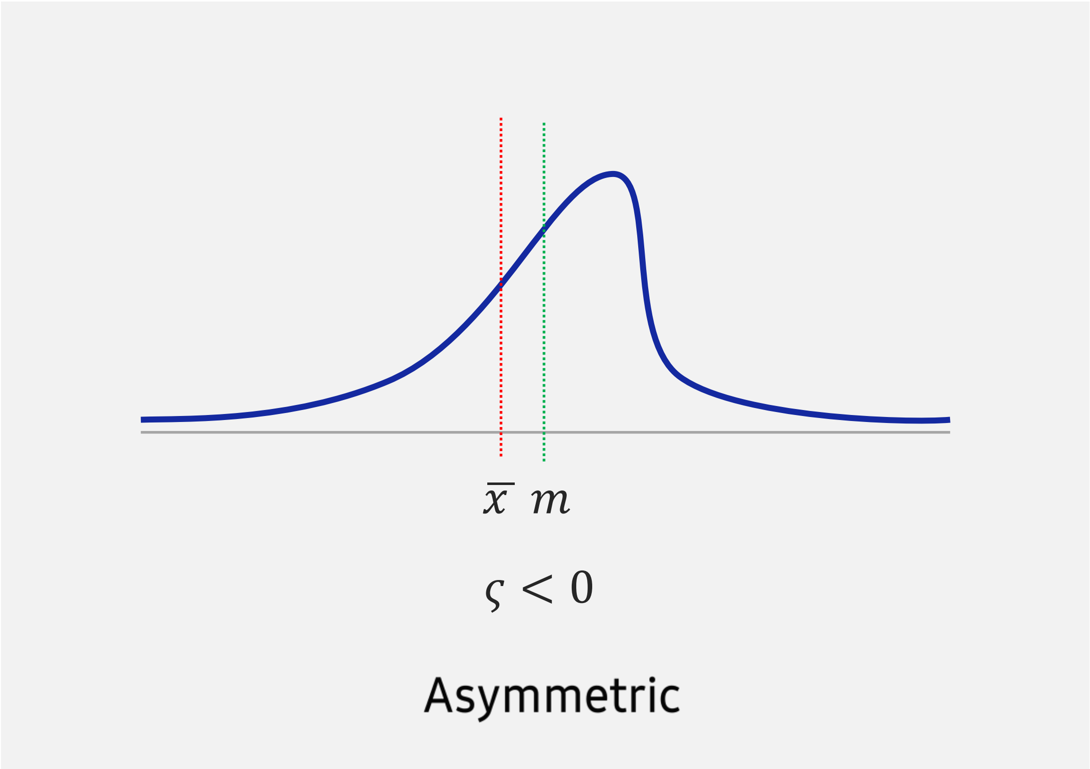
Sample statistics

- Mean value: $\overline{x}$
- Median: m
- Variance: s^2
- Standard deviation: $s = \sqrt{s^2}$
- Covariance: s_{XY}
- Correlation: r
- Skewness: ς
- Kurtosis: κ

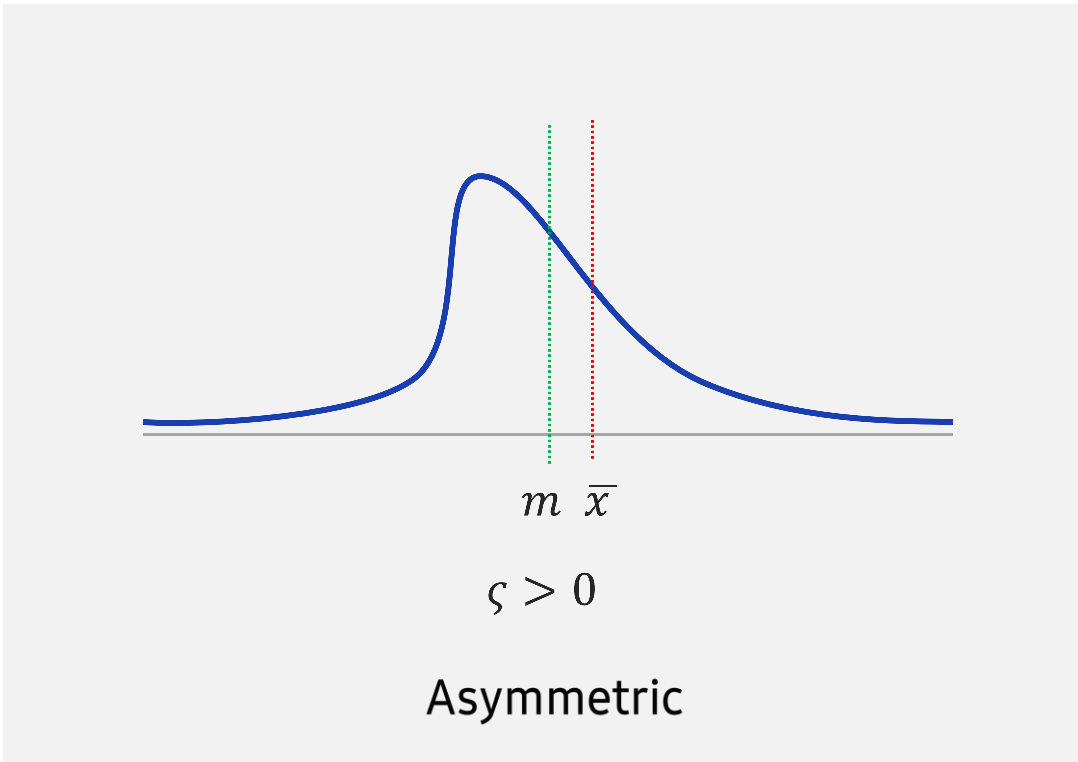
Sample statistics



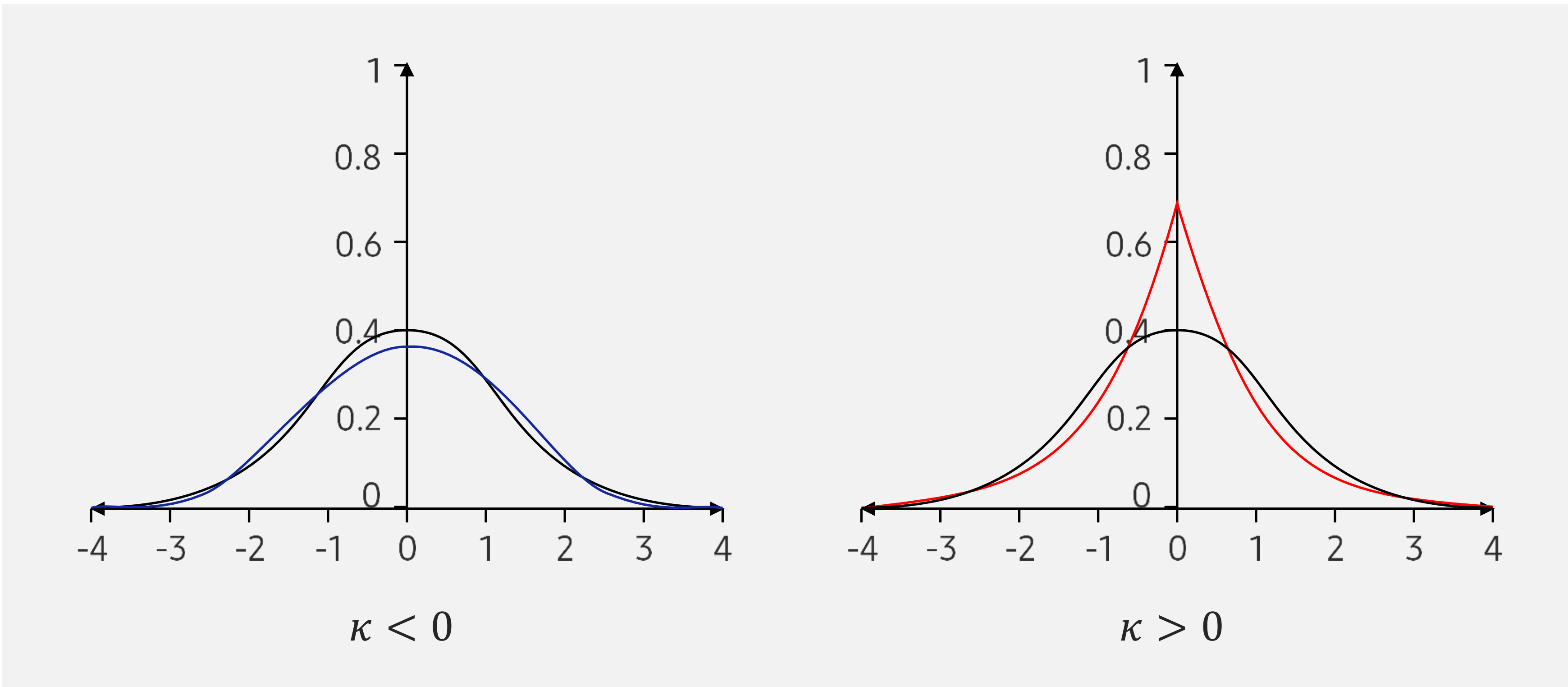
Sample statistics



Sample statistics



Sample statistics

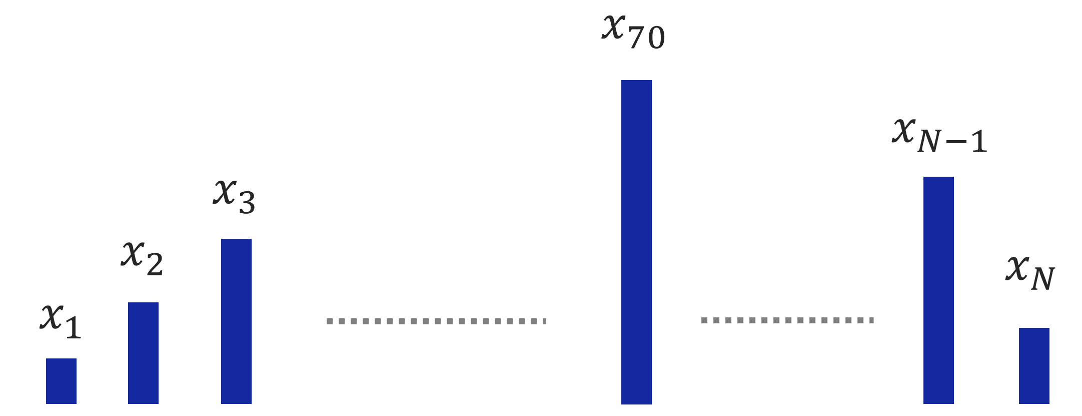


Sample statistics

- Sample variance: $s^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}$
- Sample covariance: $s_{XY} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{n-1}$
- Sample correlation: $r = \frac{s_{XY}}{s_X s_Y}$

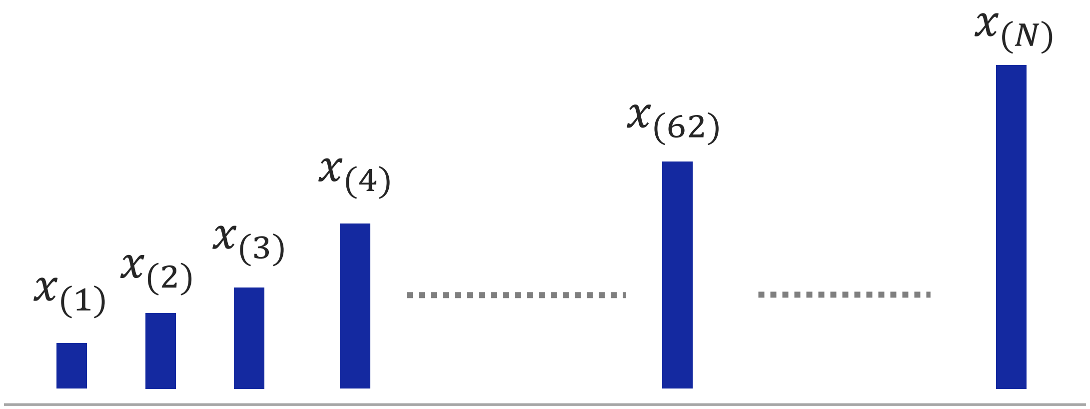
Sample statistics: quantile

- Let's suppose that you have a sample of values $x_1, x_2, x_3, \cdots, x_N$. You can represent them as bars.



Sample statistics: quantile

- Now, let us sort the values from the smallest to the largest and get $x_{(1)}, x_{(2)}, x_{(3)}, \cdots, x_{(N)}$.



- α *quantile*= value at the position $int(N\times\alpha)$

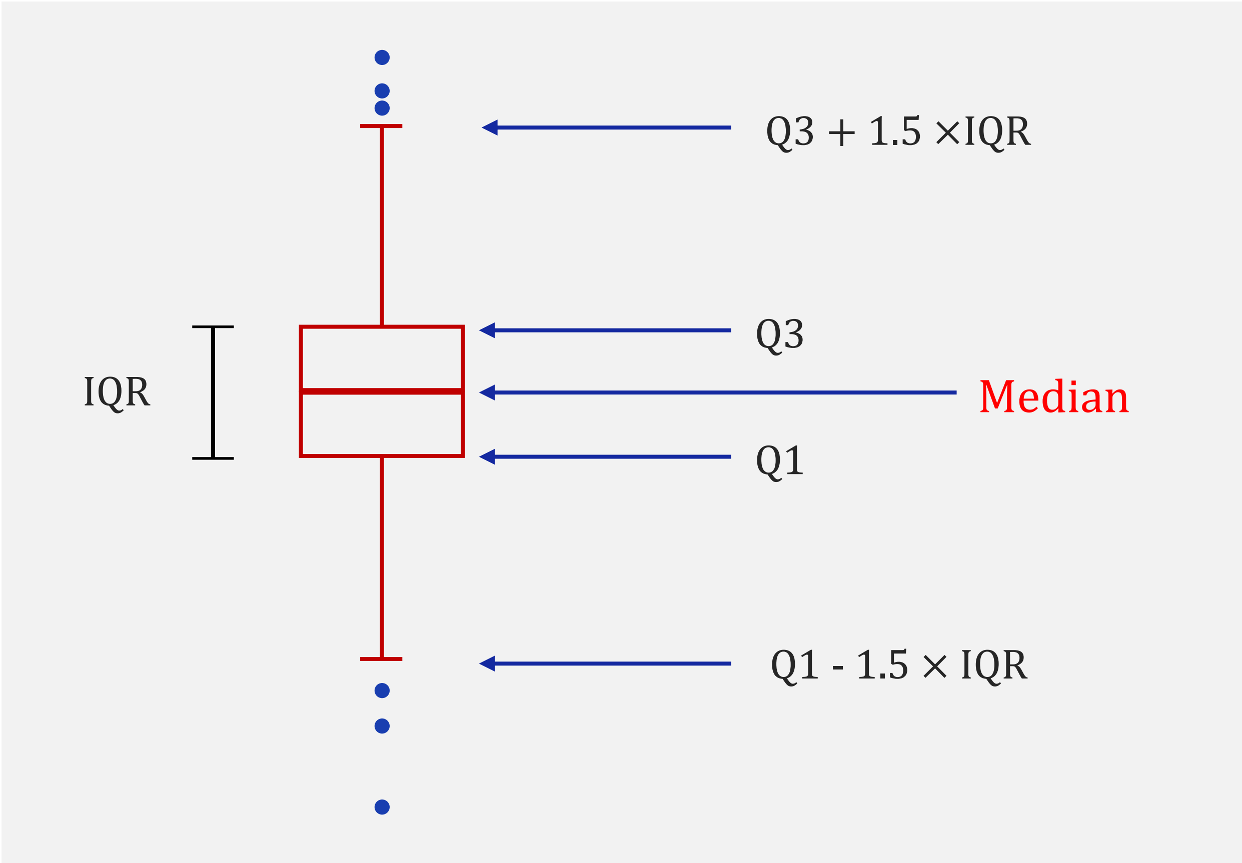
Sample statistics: quantile

- Quantile: α is probability 0 ~ 1.
- Percentile: α is given as percentage 0% ~ 100%.
  - Minimum = 0% percentile
  - Maximum = 100% percentile

$$\text{Range} = \text{Maximum} - \text{Minimum}$$

- Quartile: α is subdivided into four equal intervals.
  - 1st quartile (Q1): 25% percentile
  - 2nd quartile (Q2): 50% percentile = Median
  - 3rd quartile (Q3): 75% percentile

Boxplot



Coding Exercise #0203

Follow practice steps on 'ex_0203.ipynb' file.

Unit 3.

Understanding of Statistics II

Inferential Statistics

Population vs. Sample

- You analyze samples because the population is hard or impossible to reach.
- You would like to draw conclusions about the population using its samples.
- However, you should carefully analyze the uncertainties associated with the samples.

Population parameters vs. Sample statistics

- The population parameters are properties of a given population.
- There is one population, and the population parameters are calculated once for each.
- Sample statistics are properties of a given sample.
- In principle, you can draw several samples and calculate the same sample statistics many times.

Sampling method

- Simple random sampling: values drawn with equal probability with or without replacement
- Weighted random sampling: values drawn with varying probabilities
- Stratified sampling: reflects the proportions of the strata
- Systematic sampling: values are drawn with an implicit periodicity
- Cluster sampling: takes a representative cluster

Population parameters vs. Sample statistics

	Population	Sample
Size	N	n
Mean	$\mu = \frac{\sum_{i=1}^{N} x_i}{N}$	$\overline{x} = \frac{\sum_{i=1}^{n} x_i}{n}$
Variance	$\sigma^2 = \frac{\sum_{i=1}^{N}(x_i - \mu)^2}{N}$	$s^2 = \frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n-1}$
Standard Deviation	$\sigma = \sqrt{\sigma^2}$	$s = \sqrt{s^2}$

- Connect the sample statistics with the population parameters through statistical inference.

Central Limit Theorem

Coin flipping experiment

- Let us flip the coin twice and group the results as a sample (n=2). We assign T = 0 and H =1.
- As we draw several samples, we get the corresponding sample means: $\overline{x_1}$, $\overline{x_2}$, $\overline{x_3}$,....

Ex

i	Sample	$\overline{\mathbf{x}}_i$
1	1,1	1
2	0,1	0.5
3	1,0	0.5
4	0,0	0
⋮	⋮	⋮

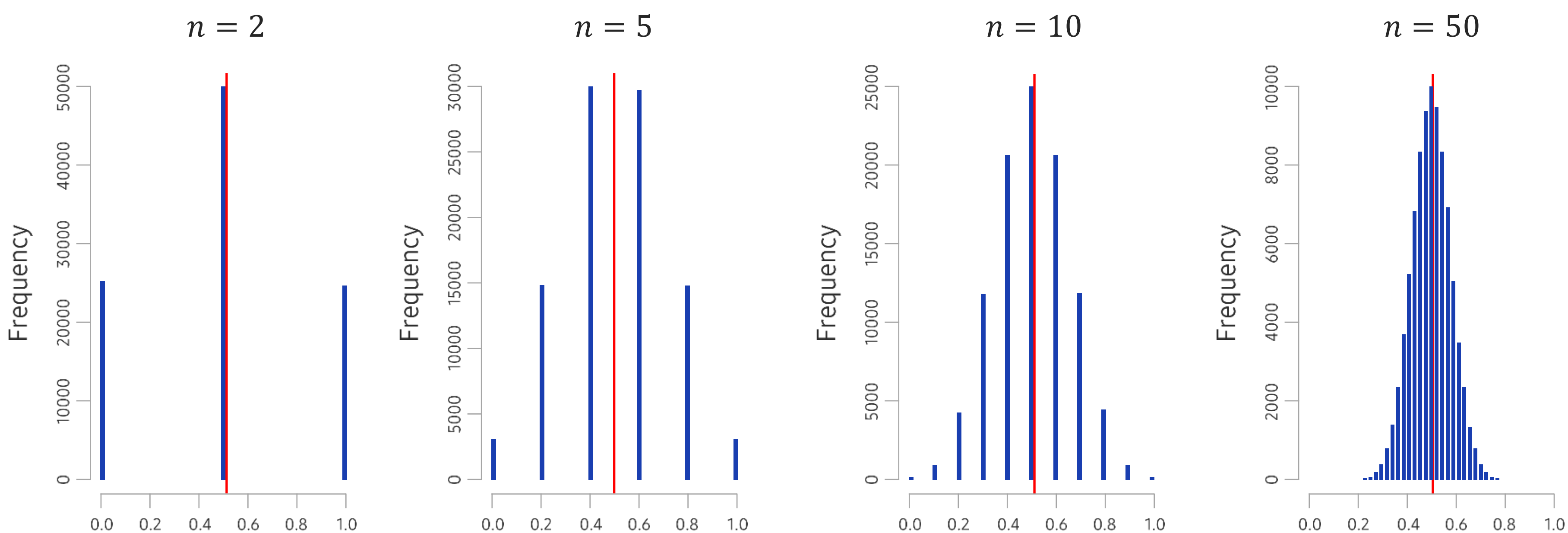
Coin flipping experiment

- Let us flip the coin three times and group the results as a sample (n=3).
- As we draw several samples, we get the corresponding sample means: $\overline{x_1}$, $\overline{x_2}$, $\overline{x_3}$,….

Ex

i	Sample	$\overline{x}_i$
1	1,0,1	2/3
2	0,1,0	1/3
3	1,0,0	1/3
4	0,0,0	0
⋮	⋮	⋮

Coin flipping experiment



- Histograms of the sample means of varying sample size
- As the sample size increases, the histograms converge to the normal distribution.

Coin flipping experiment

- The sample means $\overline{x}$ are also randomly distributed.
- You can denote the sample mean by $\overline{X}$ (upper case) and treat it as a random variable.
- Then, we obtain the following results:

a) Expected value of the $\overline{X}$ is the population mean : $E[\overline{X}]=\mu$.

b) Variance and standard deviation of the $\overline{X}$:

$$Var(\overline{X}) = \sigma_{\overline{X}}^2 = \frac{\sigma^2}{n} = \frac{p(1-p)}{n} = \frac{0.25}{n}$$

$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}} = \frac{\sqrt{p(1-p)}}{\sqrt{n}} = \frac{0.5}{\sqrt{n}}$$

"Standard error"

- σ^2= population variance, s^2= sample variance, $\sigma_{\overline{X}}^2$= variance of the sample mean.

Central limit theorem

- The distribution of the sample means converges to the normal distribution as n increases.

  1) This is true not just for the coin flipping (binomial).

  2) You can verify the same phenomena for other underlying distributions (discrete and continuous).

- For large enough n, analogous tendency can be verified for the sample statistics beside the mean.

Realistic considerations

- In reality, you will not have several samples but one.
- You apply the central limit theorem and assume that the sample distribution is normal.
- For the statistical inference, you will take advantage of the properties of the normal distribution.

Standardization

- When the sample size n is big enough $\overline{X}$ follows the normal distribution $N(\mu, \sigma$^2).
- You can standardize $\overline{X}$ and get

$$Z = \frac{\overline{X} - \mu}{\sigma / \sqrt{n}}$$

which follows the Standard Normal distribution N(0,1).

- When the sample size n is small or the σ is not known, you define

$$t = \frac{\overline{X} - \mu}{s / \sqrt{n}}$$

which follows the student-t probability density with degree of freedom $n-1$.

Sampling distribution of proportions

- Let us consider the case where the population follows Bernoulli distribution.
- Let us suppose that the success probability is p.
- We denote the sample proportion (of success) as $\widehat{P}$ and have

a) Expected value: $E\left[\hat{P}\right] = p$

b) Standard error: $\sigma_{\widehat{P}} = \frac{\sigma}{\sqrt{n}} = \sqrt{\frac{p(1-p)}{n}}$

- When $n\,p > 10$ and $n(1-p) > 10$, then we can apply the central limit theorem: $\frac{\hat{P}-p}{\sigma_{\widehat{P}}} \sim N(0,1)$.

Sampling distribution of the difference

- Let us consider two populations. Samples of size n_1 and n_2 are drawn from each population.
- Then for the difference of the sample means, you have

a) Expected value: $E[\overline{X_1} - \overline{X_2}] = E[\overline{X_1}] - E[\overline{X_2}] = \mu_1 - \mu_2$

b) Standard error: $\sigma_{\overline{X_1}-\overline{X_2}} = \sqrt{\sigma_{\overline{X_1}}^2 + \sigma_{\overline{X_2}}^2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$

- When n_1 and n_2 are large enough, you can apply the central limit theorem: $\frac{\overline{X_1}-\overline{X_2}-(\mu_1-\mu_2)}{\sigma_{\overline{X_1}-\overline{X_2}}} \sim N(0,1)$.

Standard errors

Statistic	Standard Error	Explanation
Mean	$\frac{\sigma}{\sqrt{n}}$	When $n \geq 30$, the sample mean approximately follows the normal distribution.
Proportion	$\sqrt{\frac{p(1-p)}{n}}$	When $n\,p > 10$ and $n(1-p) > 10$, the sample proportion approximately follows the normal distribution.
Median	$\sigma\sqrt{\frac{\pi}{2n}}$	When $n \geq 30$, the sample mean approximately follows the normal distribution.
Standard Deviation	a) $\frac{\sigma}{\sqrt{2n}}$ b) $\sqrt{\frac{\mu_4-\sigma^4}{4n\sigma^2}}$	a) When the population follows the normal distribution b) When the population does not follow the normal distribution
Variance	a) $\sigma^2\sqrt{\frac{2}{n}}$ b) $\sqrt{\frac{\mu_4-\sigma^2}{n}}$	a) When the population follows the normal distribution b) When the population does not follow the normal distribution The sample variance follows the chi-square distribution.
Correlation	$\sqrt{\frac{1-r^2}{n-2}}$	r is the sample correlation. Fisher's z-transformation is required.

Unit 3.

Understanding of Statistics II

Estimation Theory

Point estimation vs Interval estimation

Ex Say you want to survey the mean weight of adult males. The possible answers are:

a) 72 kg

b) 71 kg ~ 74 kg

c) 70 kg ~ 75 kg

- There are many ways of answering to the same question.
- Providing a single value like in a) is called “point estimation.” ⇐ using an estimator
- Providing an interval like in b) and c) is called “interval estimation.”

Estimator

What makes a "good" estimator?

1) Unbiasedness

2) Efficiency

3) Consistency

Unbiasedness

- Let us suppose that θ is the population parameter, and $\hat{\theta}$ is the corresponding estimator.
- A unbiased estimator satisfies the following condition: $E\left[\hat{\theta}\right] = \theta$

Ex $\frac{\sum_{i=1}^{n} x_i}{n}$ is an unbiased estimator of the population mean μ.

Ex $\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{\boldsymbol{n-1}}$ is an unbiased estimator of the population variance σ^2.

Efficiency

- Let us suppose that there are several unbiased estimators $\hat{\theta}_1, \hat{\theta}_2, \ldots$

$$E[\hat{\theta}_1] = \theta \ , \ E[\hat{\theta}_2] = \theta \ , \ E[\hat{\theta}_3] = \theta \ , \ldots$$

- The estimator with the least uncertainty (variance or standard deviation) is the efficient estimator:

Ex If $Var(\hat{\theta}_1)$ is at the minimum, then $\hat{\theta}_1$ is the efficient estimator.

Consistency

- A biased estimator for small n may become unbiased for large enough n because of the consistency.

Ex $\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{\boldsymbol{n}}$ is biased estimator of σ^2 for small n but becomes unbiased for large enough n.

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

- When the sample size n can be increased, the consistency becomes an important criteria.

Interval Estimation of the Mean

Confidence interval

- It is the interval that may contain the true value of a population parameter with a given confidence level.

  a) Significance level: α

  b) Confidence level: $(1-\alpha) \Leftarrow$ A probability

Confidence interval of the mean

- Let us suppose that the sample size is big enough that the sample mean is distributed normally.
- The 95% confidence interval can be obtained in the following way:

$$P(-1.96 < Z < 1.96) = 0.95$$

$$\Downarrow$$

$$P\left(-1.96 < \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} < 1.96\right) = 0.95$$

$$\Downarrow$$

$$P\left(-1.96\, {}^{\sigma}/_{\sqrt{n}} < \bar{X}-\mu < 1.96\, {}^{\sigma}/_{\sqrt{n}}\right) = 0.95$$

$$\Downarrow$$

Confidence interval of the mean

- Let us suppose that the sample size is big enough that the sample mean is distributed normally.
- The 95% confidence interval can be obtained in the following way:

$$P\left(-1.96\,{}^{\sigma}/_{\sqrt{n}} < \bar{X} - \mu < 1.96\,{}^{\sigma}/_{\sqrt{n}}\right) = 0.95$$

$$P\left(-\bar{X} - 1.96\,{}^{\sigma}/_{\sqrt{n}} < -\mu < -\bar{X} + 1.96\,{}^{\sigma}/_{\sqrt{n}}\right) = 0.95$$

$$P\left(\bar{X} - 1.96\,{}^{\sigma}/_{\sqrt{n}} \leq \mu \leq \bar{X} + 1.96\,{}^{\sigma}/_{\sqrt{n}}\right) = 0.95$$

Confidence interval of the mean

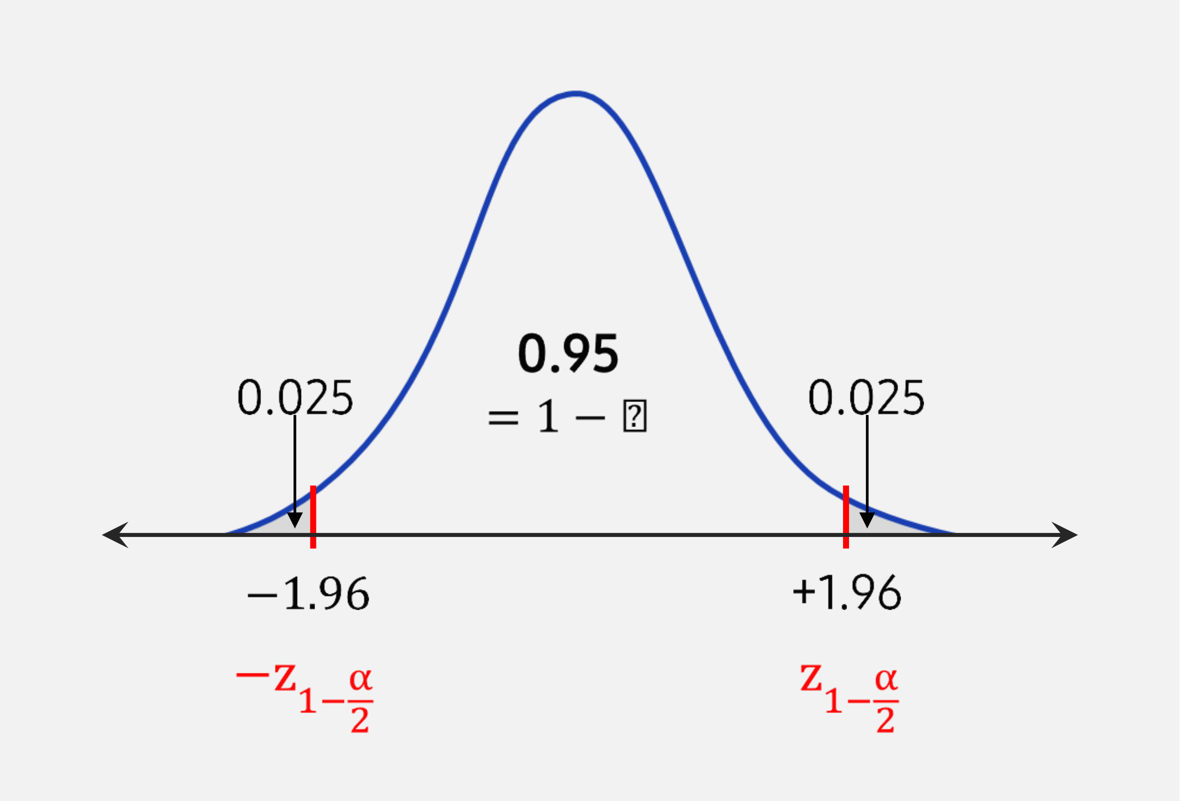
- The 95% confidence interval can be obtained in the following way:

Lower bound: $\bar{X} - 1.96\, {}^{\sigma}/_{\sqrt{n}}$ Upper bound: $\bar{X} + 1.96\, {}^{\sigma}/_{\sqrt{n}}$

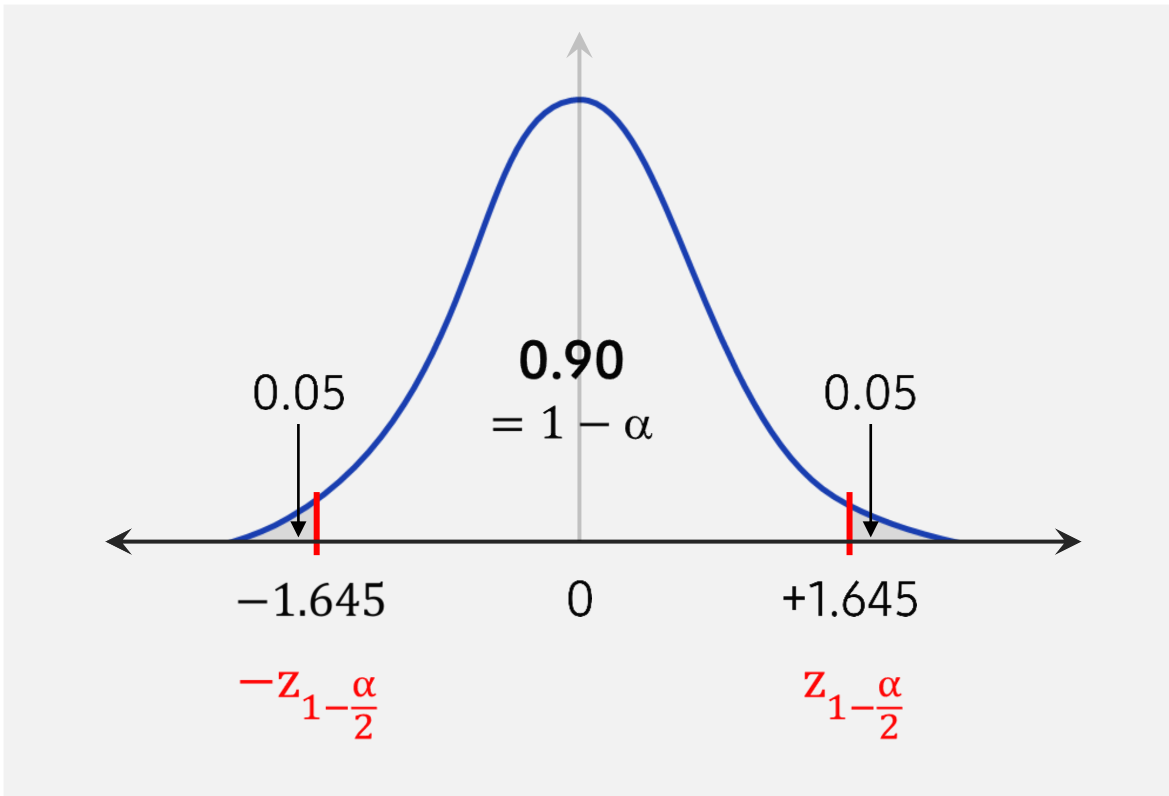
[$\leftarrow$ Confidence Interval $\rightarrow$]

- But, where did the number 1.96 come from?
- 1.96 is the 0.975 quantile of the normal distribution $\Leftrightarrow$ the position where the CDF = 0.975.

Confidence interval of the mean: 95%



Confidence interval of the mean: 90%



Confidence interval of the mean

- Of arbitrary confidence level = (1−α)

Lower bound: $\bar{X} - z_{1-\frac{\alpha}{2}}\,{}^{\sigma}/_{\sqrt{n}}$ Upper bound: $\bar{X} + z_{1-\frac{\alpha}{2}}\,{}^{\sigma}/_{\sqrt{n}}$

[← Confidence Interval →]

- When the sample size is small and the population variance is not known, use student-t.

Lower bound: $\bar{X} - t_{1-\frac{\alpha}{2}}\,{}^{s}/_{\sqrt{n}}$ Upper bound: $\bar{X} + t_{1-\frac{\alpha}{2}}\,{}^{s}/_{\sqrt{n}}$

[← Confidence Interval →]

Confidence interval of the mean

Question:

Isn't it always good to have as large a confidence level as possible?

Confidence interval of the mean

Answer:

All the other conditions being equal, larger confidence level means larger uncertainty.
The confidence interval broadens.

Confidence interval of the mean

[99% confidence interval]

[95% confidence interval]

[90% confidence interval]

How to narrow the confidence interval of the mean while keeping the confidence level high?

- Let us remember that the bounds of the confidence interval was given by:

$$\bar{X} \pm z_{1-\frac{\alpha}{2}} \times \frac{\sigma}{\sqrt{n}} \quad \text{or} \quad \bar{X} \pm t_{1-\frac{\alpha}{2}} \times \frac{s}{\sqrt{n}}$$

- Smaller σ or s can narrow the confidence interval $\Leftarrow$ **you cannot** control it!
- Larger sample size n can narrow the confidence interval $\Leftarrow$ in principle, **you can** control it!
- If W is the targeted half width of the confidence interval, apply the formula:

$$z_{1-\frac{\alpha}{2}} \times \frac{\sigma}{\sqrt{n}} = W \quad \Rightarrow \quad \boldsymbol{n} = \left[\frac{z_{1-\alpha/2} \times \sigma}{W}\right]^2$$

Coding Exercise #0204

Follow practice steps on 'ex_0204.ipynb' file.

Coding Exercise #0205

Follow practice steps on 'ex_0205.ipynb' file.

Interval Estimation of the Correlation

Confidence interval of the correlation

- There are different kinds of correlation: Pearson, Spearman, Kendall, etc.
- Spearman and Kendall are the so-called “rank correlations.”
- Let us focus on the “usual type,” that is, Pearson correlation.

Confidence interval of the correlation

- The distribution of the sample correlation r deviates significantly from the normal distribution.
- This is true even for large enough sample size, because the correlation is bound to be $[-1, +1]$.
- To overcome these "distortions," it is necessary to do Fisher's z-transformation.

a) Transformed correlation: $z = arctanh(r)$

b) Standard error after the transformation: $\sigma_Z = \frac{1}{\sqrt{n-3}}$

c) Inverse transformation: $r = tanh(z)$

Confidence interval of the correlation

- Of arbitrary confidence level = (1−α)

$$tanh\left(z - z_{1-\frac{\alpha}{2}} \times \sigma_z\right) \qquad tanh\left(z + z_{1-\frac{\alpha}{2}} \times \sigma_z\right)$$

[← Confidence Interval →]

Ex 95% confidence interval: $[tanh(z - 1.96\ \sigma_z), tanh(z + 1.96\ \sigma_z)]$

Coding Exercise #0206

Follow practice steps on 'ex_0206.ipynb' file.

Unit 4.

Statistical Hypothesis Testing

Hypothesis Testing

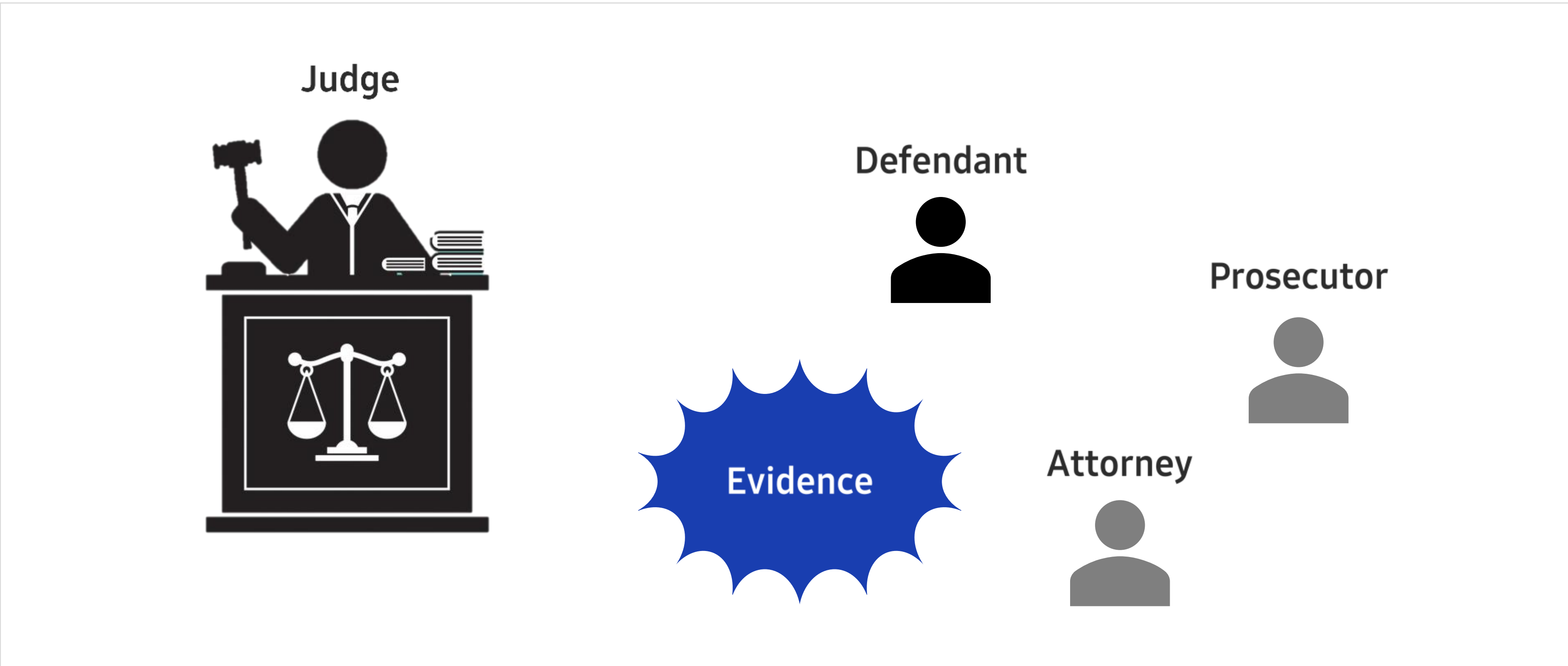
Principles of the hypothesis testing in statistics

- It tests a hypothesis about the population parameters using the sample statistics.

  Ex "The average sleep time of adults in USA is 7 hours." ⇐ True or False?

  Ex "The average wage in the zip code area 91885 is 85,000 USD." ⇐ True or False?

Principles of the hypothesis testing in statistics



Principles of the hypothesis testing in statistics

- The defendant is assumed innocent until proven otherwise.
- The defendant doesn’t need to prove his/her innocence.
- The main objective of the trial is to prove the guiltiness of the defendant.
- The judge delivers the verdict based on the presented evidence.

Principles of the hypothesis testing in statistics

- Null hypothesis H_0: what is assumed until proven otherwise
- Alternative hypothesis H_1: what needs to be proven
- Test statistic: a quantity used as the evidence; calculated from the sample
- p-value: probability of observing the current test statistic or more extreme one assuming that the null hypothesis is true

  a) If the p-value is small: weakens the assumption of the null hypothesis

  b) If the p-value is large: strengthens the assumption of the null hypothesis

Principles of the hypothesis testing in statistics

- Significance level α:

  a) It's the reference probability used when deciding whether the p-value is small enough or not.

  b) If p-value ≥ α, the H_0 is maintained.

  If p-value < α, the H_0 is rejected in favor of the H_1.

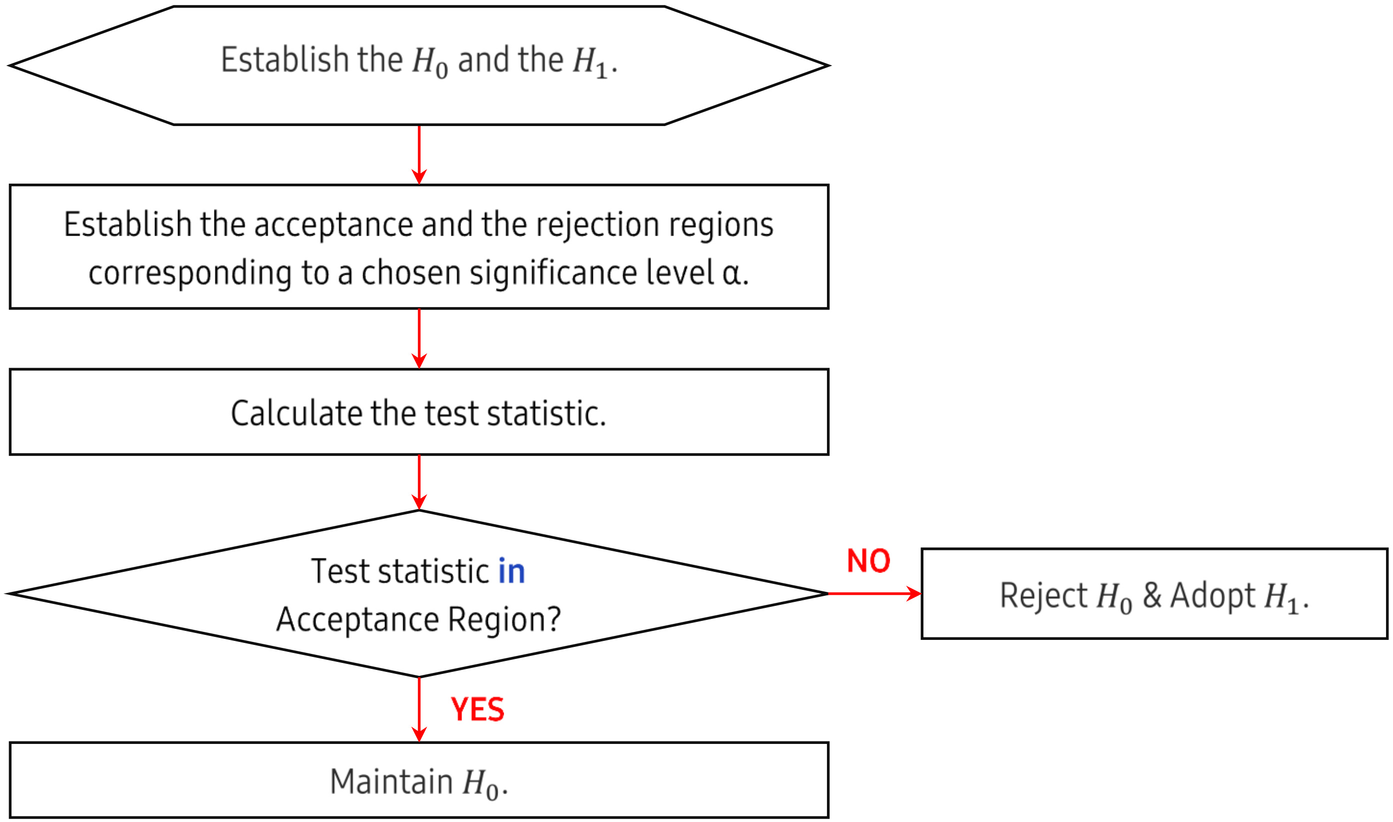
  c) This can also be interpreted as the maximum probability of rejecting the H_0 even when it's true.

  d) 1−α is the probability that the true H_0 is retained.

- Statistical power: probability of rejecting the false H_0 in favor of H_1

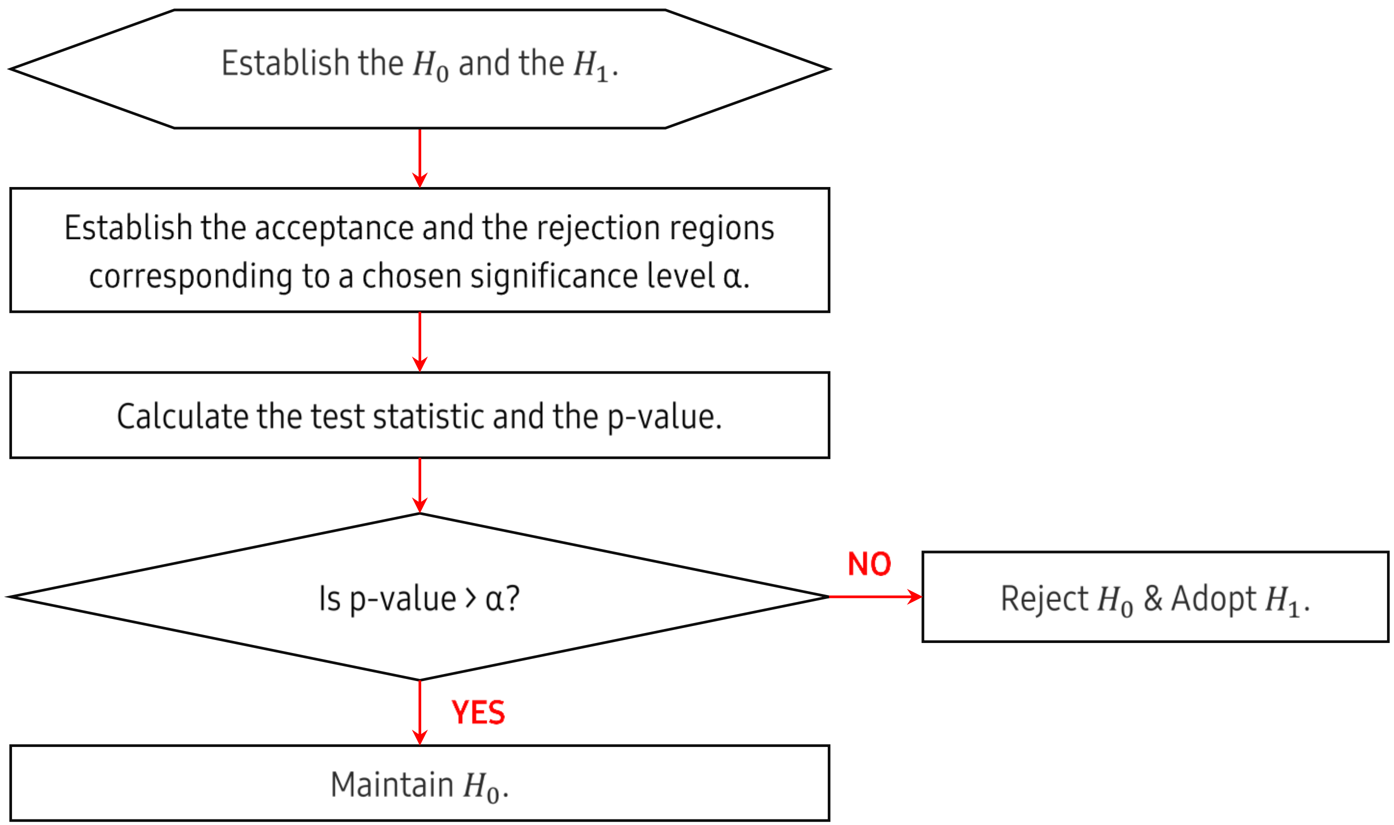
Principles of the hypothesis testing in statistics

- Procedure using the **test statistic**:



Principles of the hypothesis testing in statistics

- Procedure using the **p-value**:



Principles of the hypothesis testing in statistics

TEST RESULT \ ACTUALLY	H_0 IS TRUE	H_0 IS FALSE
H_0 RETAINED	Correct Decision Probability $= 1 - \alpha$	Type 2 error Probability$= \beta$
H_0 REJECTED & H_1 ADOPTED	Type 1 error Probability $= \alpha$	Correct Decision Probability $= 1 - \beta$

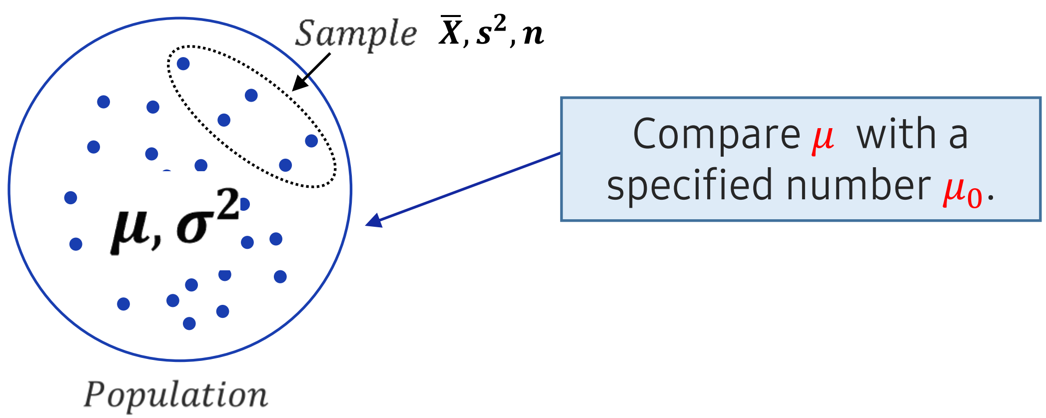
Unit 4.

Statistical Hypothesis Testing

Hypothesis Test of the Means

One sample t-test

- There is one population and one sample.



- Student-t distribution is used to interpret the test statistic calculated as following:

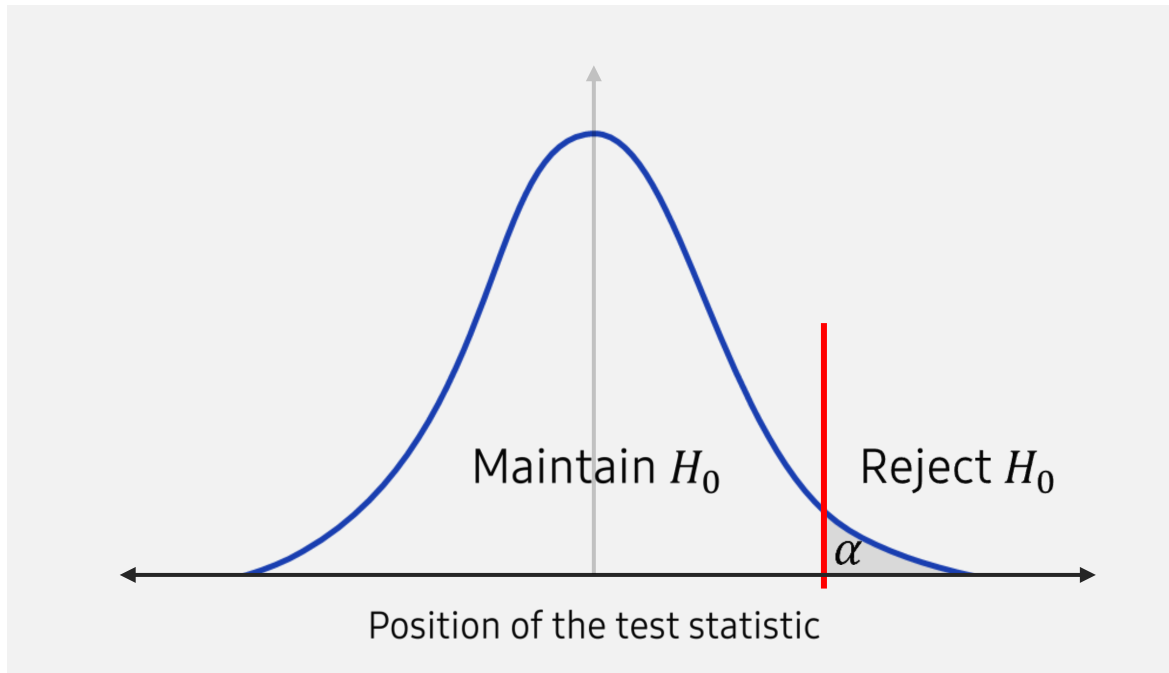
$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

One sample t-test

- Right tail test:

$H_0 : \mu \leq \mu_0$
$H_1 : \mu > \mu_0$ “larger”

- p-value = $P(X > test\ statistic)$

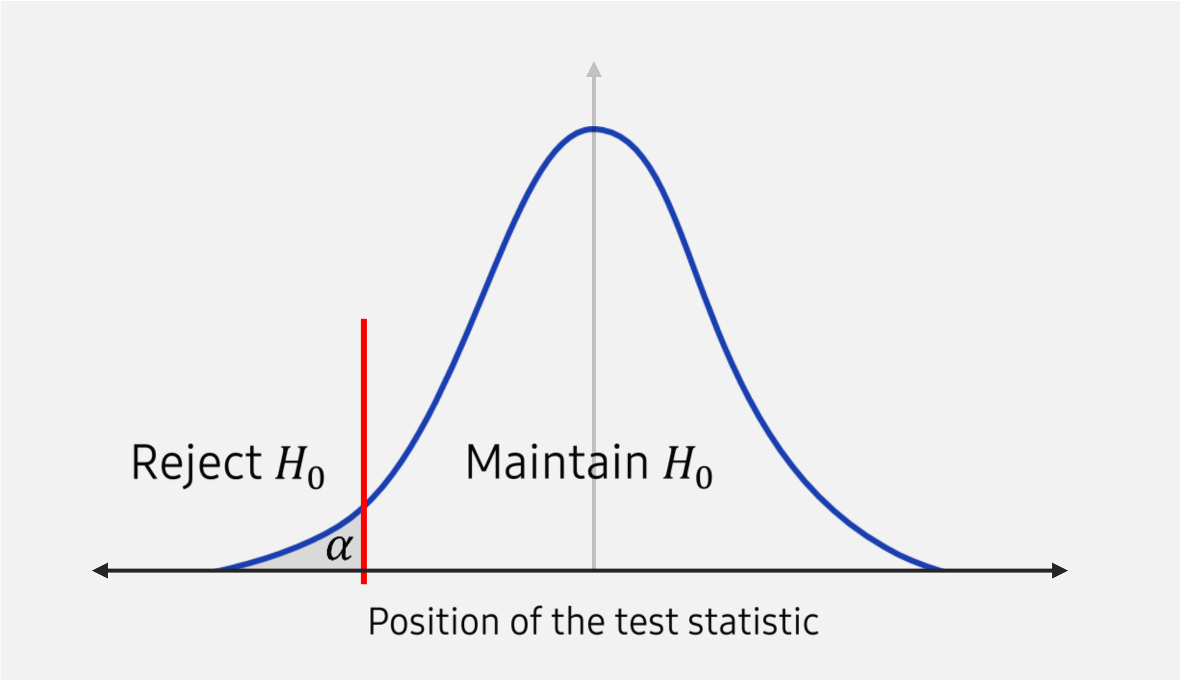


One sample t-test

- Left tail test:

$$H_0 \;:\; \mu \geq \mu_0$$
$$H_1 \;:\; \mu < \mu_0 \qquad \text{“smaller”}$$

- p-value = $P(X < test\ statistic)$

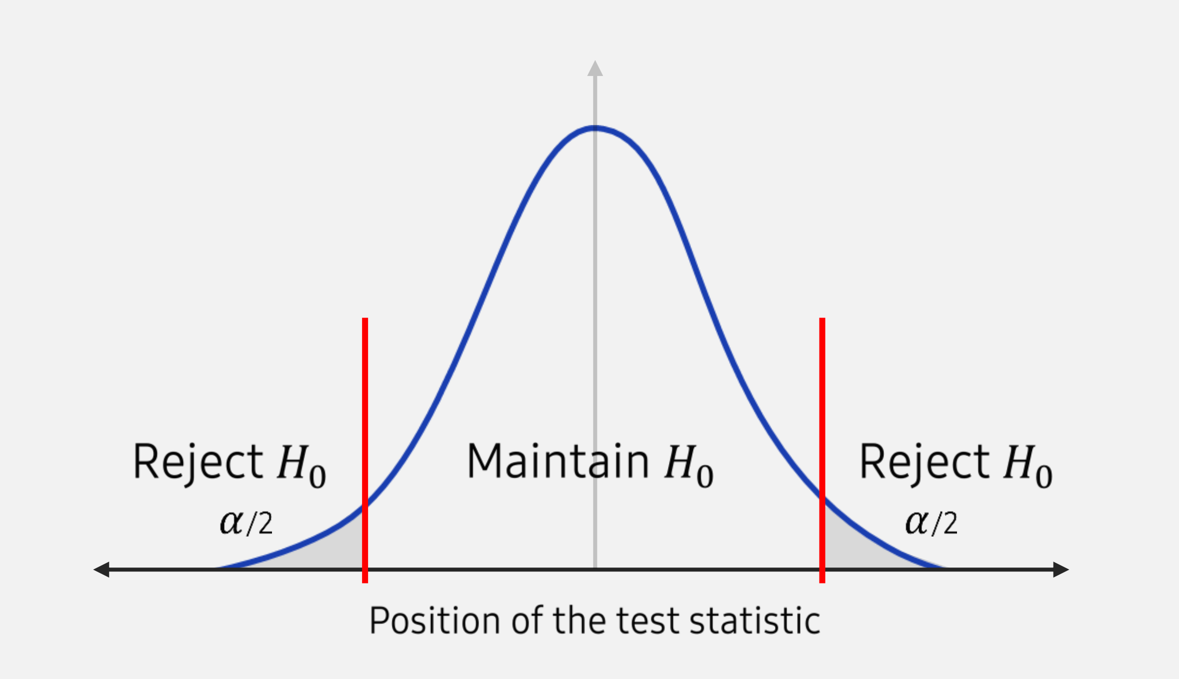


One sample t-test

- Two tail test:

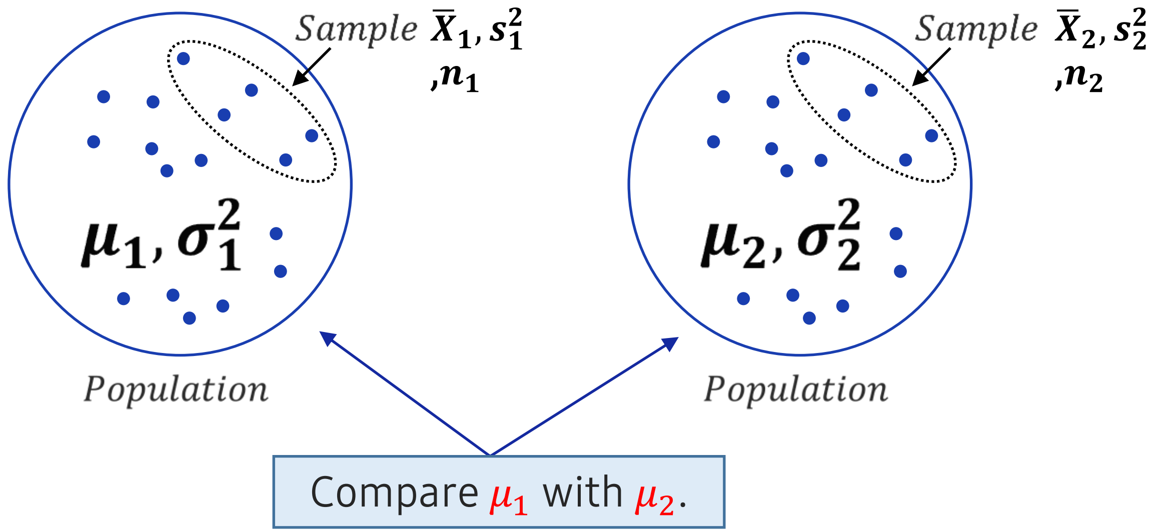
$$H_0 : \mu = \mu_0$$
$$H_1 : \mu \neq \mu_0 \quad \text{“different”}$$

- p-value $=2 \times P(X > |test\ statistic|)$



Independent two sample t-test

- There are two populations and two samples.



Independent two sample t-test

- Right tail test

$$H_0 : \mu_1 - \mu_2 \leq 0 \quad \Leftrightarrow \quad \mu_1 \leq \mu_2$$
$$H_1 : \mu_1 - \mu_2 > 0 \quad \Leftrightarrow \quad \mu_1 > \mu_2$$

- Left tail test

$$H_0 : \mu_1 - \mu_2 \geq 0 \quad \Leftrightarrow \quad \mu_1 \geq \mu_2$$
$$H_1 : \mu_1 - \mu_2 < 0 \quad \Leftrightarrow \quad \mu_1 < \mu_2$$

- Two tail test

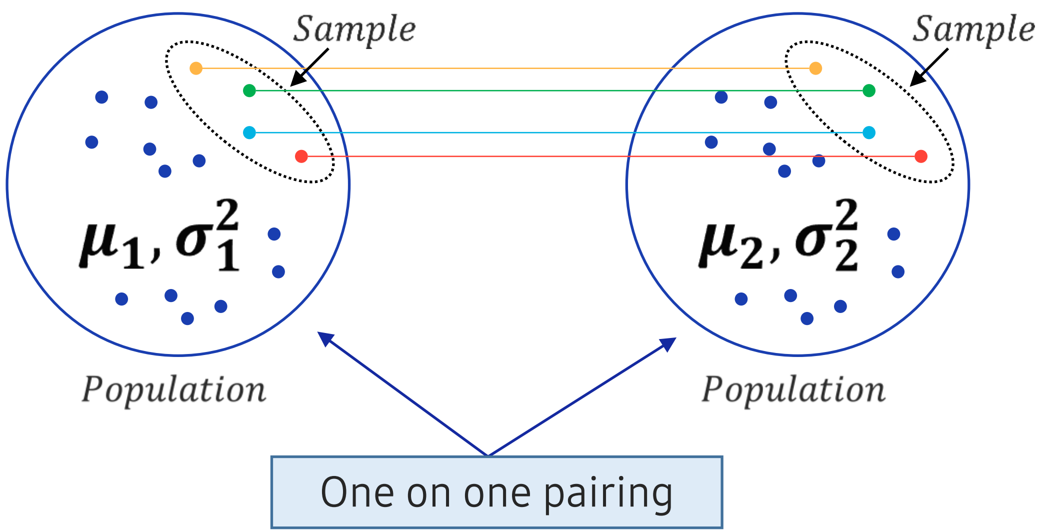
$$H_0 : \mu_1 - \mu_2 = 0 \quad \Leftrightarrow \quad \mu_1 = \mu_2$$
$$H_1 : \mu_1 - \mu_2 \neq 0 \quad \Leftrightarrow \quad \mu_1 \neq \mu_2$$

You should consider two cases:

1) Equal variances.
2) Unequal variances.

Paired two sample t-test

- There are two populations and two samples. There is “one on one” pairing.



Ex Change in the blood pressure of the same test subjects before and after taking a new drug

Analysis of Variance (ANOVA)

- So far with the t-test, we had one or two groups (samples).
- ANOVA can detect differences in the means of two or more groups.

Null hypothesis $\boldsymbol{H_0}$: $\mu_1 = \mu_2 = \mu_3 = \ldots$

Alternative hypothesis $\boldsymbol{H_1}$: There is at least one case where $\mu i \neq \mu j$

- Assumptions: 1) The distribution of the data is normal.
  2) The group variances are the same.
  3) The groups are independent from each other.
- F distribution is used to calculate the p-value.

Coding Exercise #0207

Follow practice steps on 'ex_0207.ipynb' file.

Hypothesis Test of the Frequencies

Chi-squared test for one way table

- One way table or "frequency table" summarizes a categorical variable.
- It compares the observed frequencies with a given model (expected frequencies).

  Null hypothesis H_0 : The observed frequency table and the expected model agree.

  Alternative hypothesis H_1 : The observed frequency table and the expected model are different.
- It is also called the "goodness of fit test."

Chi-squared test for one way table

- The test statistic is:

$$test\ statistic = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$$

- E_i are the expected frequencies, and O_i are the observed frequencies.
- k is the number of categories or types.
- The test statistic follows the chi-square distribution of degree of freedom = $k - 1$.

Chi-squared test for two way table

- A contingency table summarizes two categorical variables.

  Ex Confusion matrix (machine learning)

- It uses the frequencies to test the existence of relationship between two categorical variables.

  Null hypothesis H_0 : The categorical variables are independent.

  Alternative hypothesis H_1 : The categorical variables are not independent.

- It is also called the "independence test."

Chi-squared test for two way table

- The test statistic is:

$$test\ statistic = \sum_{i=1}^{r}\sum_{j=1}^{c}\frac{\left(O_{ij}-E_{ij}\right)^2}{E_{ij}}$$

- E_i are the expected frequencies, and O_i are the observed frequencies.
- r is the number of rows, and c is the number of columns in the two way table.
- The test statistic follows the chi-square distribution of degree of freedom = $(r - 1) \times (c - 1)$.

Coding Exercise #0208

Follow practice steps on 'ex_0208.ipynb' file.

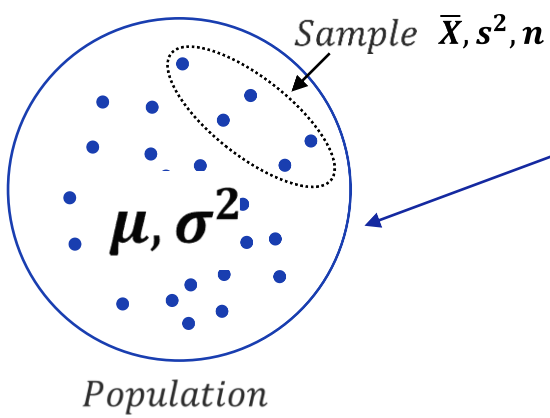
Coding Exercise #0209

Follow practice steps on 'ex_0209.ipynb’ file.

Hypothesis Test of the Variances

One sample t-test

- There is one population and one sample.



Compare σ^2 with a specified number $\sigma_0{}^2$.

Chi-squared test of variance

- There are left tail test, right tail test, and two tail test. ⇐ Just like in t-test
- The test statistic is calculated as: (n= sample size).

$$test\ statistic = \frac{(n-1)S^2}{\sigma_0^2}$$

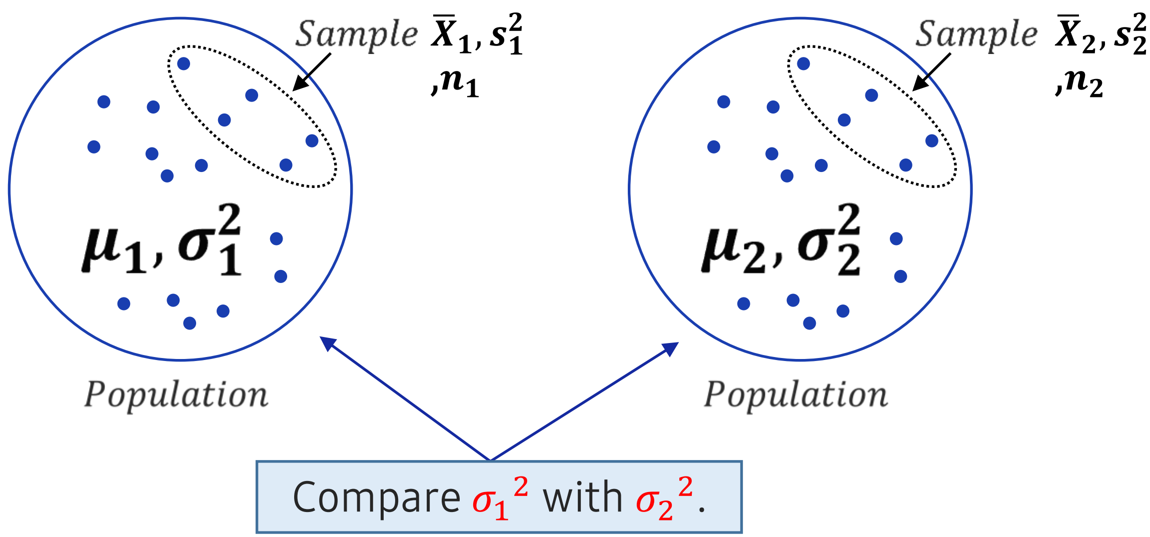
- The test statistic follows the chi-square distribution of degree of freedom = $n - 1$.

Coding Exercise #0210

Follow practice steps on 'ex_0210.ipynb' file.

F-test of variance ratio

- There are two populations and two samples.



F-test of variance ratio

- There are left tail test, right tail test, and two tail test.
- The test statistic is calculated as a ratio of the sample variances:

$$test\ statistic = \frac{S_1^2}{S_2^2}$$

- The test statistic follows the F distribution $F(n_1 - 1, n_2 - 1)$. Here, n_1 and n_2 are the sample sizes.

Hypothesis Test Summary

Here, this summarizes the hypothesis tests we have covered so far.

HYPOTHESIS TEST	PROBABILITY DENSITY DISTRIBUTION
One sample t-test, Independent two sample t-test, Paired sample t-test	Student-t
ANOVA	F
Chi-squared test of one way table, Chi-squared test of two way table	Chi-square
Chi-squared test of variance	Chi-square
F-test of variance ratio	F

Coding Exercise #0211

Follow practice steps on 'ex_0211.ipynb' file.

Coding Exercise #0212

Follow practice steps on 'ex_0212.ipynb' file.

End of Document

SAMSUNG

Together for Tomorrow!

Enabling People

Education for Future Generations