



ISO 9001: 2015



School of  
Engineering & Technology

# Chương 2. Kiểm định giả thuyết thống kê

## 2.1. Kiểm định giả thuyết thống kê

# Nội dung

- 2.1. Mô hình kiểm định ý nghĩa thống kê (test of significance)
- 2.2. Mô hình kiểm định giả thuyết (test of hypothesis)
- 2.3. Mô hình hỗn hợp
- 2.4. Trị số P và vấn đề của trị số P
- 2.5. Ra quyết định trong kiểm định

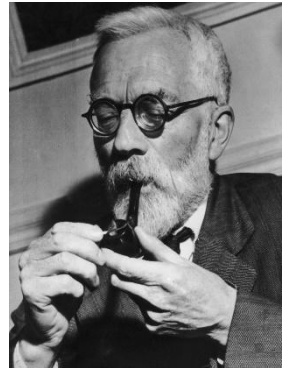
# Lý thuyết khoa học

- Theo Karl Popper, một giả thuyết được xem là mang **tính khoa học** nếu giả thuyết đó có khả năng **“phản nghiệm”** (falsified).
- Đặc điểm duy nhất để phân biệt một lý thuyết khoa học thực thụ với giả khoa học (pseudoscience) là thuyết khoa học luôn có đặc tính có thể “bị bác bỏ” hay “khả năng phản nghiệm” bằng những thực nghiệm đơn giản.
- Ví dụ: lý thuyết khoa học **“vi khuẩn *V. cholerae* gây bệnh dịch tả”** có thể bị bác bỏ nếu có một bệnh nhân dịch tả nhưng không nhiễm vi khuẩn *V. cholerae*.



## 2.1. Mô hình kiểm định ý nghĩa thống kê (test of significance)

- **Ronald A. Fisher (1890-1962)** đề xuất một phương pháp định lượng để phản nghiệm một giả thuyết khoa học gọi là **test of significance** (phương pháp kiểm định ý nghĩa thống kê).
- Fisher quan niệm rằng thống kê là một bộ phận quan trọng của phương pháp **suy luận quy nạp** (inductive inference), tức là phương pháp suy luận dựa vào quan sát **các mẫu (sample)** và **khái quát** cho **một tổng thể (population)**
- **Phương pháp kiểm định ý nghĩa thống kê gồm 3 bước:**
  - Bước 1, phát biểu một giả thuyết vô hiệu (null hypothesis)
    - Giả thuyết vô hiệu là giả thuyết ngược lại với giả thuyết mà nghiên cứu muốn kiểm định. Chẳng hạn nếu muốn kiểm định giả thuyết vaccine có tác dụng giảm tỷ lệ nhiễm COVID-19, thì giả thuyết vô hiệu là “**tỷ lệ nhiễm COVID-19 ở nhóm bệnh nhân sử dụng vaccine bằng với nhóm sử dụng giả dược**”. Gọi giả thuyết vô hiệu là  $H_0$ .
  - Bước 2, thu thập dữ liệu.
    - Trong ví dụ trên, số liệu thu thập là số lượng nhiễm COVID. Gọi dữ liệu là  $D$ .
  - Bước 3, ước tính xác suất quan sát dữ liệu  $D$  nếu giả thuyết  $H_0$  đúng.
    - Hay ước tính  $P(D|H_0)$ . Đây chính là tính trị số  $p$  (p-value)

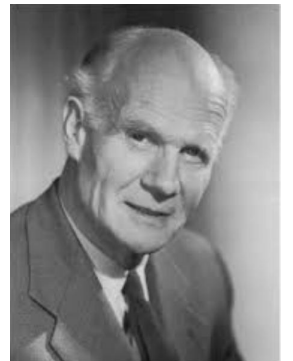


## 2.2. Mô hình kiểm định giả thuyết (test of hypothesis)

- Jerzy Neyman và Egon Pearson, cùng lúc với Fisher, phát triển một phương pháp rất khác gọi là **Test of Hypothesis (kiểm định giả thuyết)**.
- Hai ông bác bỏ quan niệm suy luận theo quy nạp mà cho rằng thống kê học là một phương pháp hay cơ chế để hướng dẫn chúng ta đi đến một quyết định đúng về lâu về dài. Hai ông cho rằng phương pháp của Fisher vô nghĩa.
- Phương pháp kiểm định giả thuyết gồm 4 bước:**
  - Bước 1: Phát biểu giả thuyết chính ( $H_1$ ) và giả thuyết vô hiệu ( $H_0$ )
  - Bước 2: Quyết định mức độ  $\alpha$  và  $\beta$  có thể chấp nhận được và ước tính cỡ mẫu cần giả thuyết.  $\alpha$  là xác suất bác bỏ  $H_1$  trong khi  $H_1$  đúng,  $\beta$  là xác suất bác bỏ  $H_0$  trong khi  $H_0$  đúng.
  - Bước 3: Thu thập dữ liệu liên quan đến giả thuyết
  - Bước 4: Nếu dữ liệu nằm trong khoảng bác bỏ giả thuyết  $H_0$  thì chấp nhận giả thuyết  $H_1$ ; nếu không thì chấp nhận giả thuyết  $H_0$ .
- Nguyên lý của mô hình của Neyman và Pearson là dựa vào dữ liệu để chọn một mô hình mà về lâu về dài chúng ta không quá sai. Vì thế ta thường chọn  $\alpha = 5\%$  và  $\beta = 10\%$  đến  $20\%$ .
- Fisher và Neyman-Pearson tranh luận dai dẳng...



Jerzy Neyman (1894-1981)







Egon Pearson (1895-1980)

## 2.3. Mô hình hỗn hợp

- Mô hình hỗn hợp kết hợp 2 mô hình của Fisher và Neyman-Pearson.
- Mô hình này sử dụng kết quả kiểm định của Fisher để đi đến quyết định chấp nhận hay bác bỏ giả thuyết  $H_0$  hay  $H_1$  theo mô hình của Neyman và Pearson.
- Mô hình nghiên cứu lâm sàng đối chứng ngẫu nhiên (Randomized Controlled clinical Trial - RCT), 5 bước:
  - Bước 1, định nghĩa một giả thuyết vô hiệu và một giả thuyết chính. VD: trong một nghiên cứu lâm sàng, gồm 2 nhóm bệnh nhân, nhóm 1 điều trị bằng thuốc A, nhóm 2 điều trị bằng giả dược (placebo). Giả thuyết vô hiệu  $H_0$  phát biểu rằng độ hiệu nghiệm của thuốc A tương đương với placebo.
  - Bước 2, xác định xác suất  $\alpha$  (sai số loại I) và  $\beta$  (sai số loại II), và ước tính cỡ mẫu dựa vào hai xác suất này.
  - Bước 3, thu thập dữ liệu liên quan đến giả thuyết, gọi dữ liệu là  $D$ .
  - Bước 4, sử dụng phương pháp kiểm định ký nghĩa thống kê của Fisher ước tính xác suất  $P(D|H_0)$ , gọi trị số này là  $P$ .
  - Bước 5, nếu  $P < 0.05$ , bác bỏ giả thuyết  $H_0$ . Chú ý, bác bỏ  $H_0$  không có nghĩa là chấp nhận  $H_1$ .

# Sai lầm loại I và Sai lầm loại II

- Sai lầm loại I (Type 1 error):
  - Bác bỏ  $H_0$  trong khi nó Đúng – False Positive (Dương tính giả)
- Sai lầm loại II (Type II error)
  - Chấp nhận  $H_0$  trong khi nó Sai – False Negative (Âm tính giả)

	Null Hypothesis is TRUE	Null Hypothesis is FALSE
Reject null hypothesis	 Type I Error (False positive)	 Correct Outcome! (True positive)
Fail to reject null hypothesis	 Correct Outcome! (True negative)	 Type II Error (False negative)

**Giả thuyết vô hiệu ( $H_0$ )**

**Bác bỏ**

**Chấp nhận**

**Đúng**

Sai lầm loại I  
Dương tính giả  
 $P = \alpha$

Đúng  
Âm tính thật  
 $P = 1 - \alpha$

**Sai**

Đúng  
Dương tính thật  
 $P = 1 - \beta$

Sai lầm loại II  
Âm tính giả  
 $P = \beta$



## 2.3. Mô hình hỗn hợp

- VD: Nghiên cứu hiệu quả của thuốc Ramipril trong phòng chống tử vong và đột quy. Nghiên cứu trên 2 nhóm bệnh nhân, nhóm 1 sử dụng Ramipril, nhóm 2 sử dụng placebo. Bắt đầu xác định  $\alpha = 0.05$  và  $\beta = 0.08$ , các nhà nghiên cứu đi xác định cỡ mẫu. Sau 3 năm thu thập số liệu, có kết quả

	Tỉ số nguy cơ với khoảng tin cậy 95%	Trị số P
Nhồi máu cơ tim, đột quy, tử vong	0.83 (0.75-0.91)	0.0002
Tỉ lệ mới mắc bệnh tiểu đường	0.69 (0.56-0.85)	0.0006

- Vì trị số P thấp hơn  $\alpha = 0.05$  nên nhà nghiên cứu cho rằng sự khác biệt về tỉ lệ tử vong và đột quy giữa 2 nhóm là có ý nghĩa thống kê (significant different).
- Tuy nhiên: **Trị số P không nói lên trực tiếp rằng thuốc Ramipril có tác dụng giảm tử vong và đột quy.** Trị số P ở đây nói rằng nếu thuốc Rampril không có tác dụng giảm tử vong và đột quy ( $H_0$  đúng), thì xác suất các nhà nghiên cứu quan sát được số liệu là 0.0002.



## 2.4. Vấn đề của trị số P

### Vấn đề logic

- Tiến trình nghiên cứu có thể khái quát như sau:
  - Đề ra một giả thuyết vô hiệu ( $H_0$ )
  - Từ giả thuyết vô hiệu, đề ra một giả thuyết chính ( $H_1$ )
  - Tiến hành thu thập dữ liệu  $D$
  - Phân tích dữ kiện, tính toán xác suất  $D$  xảy ra nếu  $H_0$  là thật,  $P(D|H_0)$ . Bước này chính là xác định trị số  $P$ .
- Trị số  $P$  là một xác suất có điều kiện:  $P(D|H_0)$ . Nó không cho ta biết sự khả dĩ của giả thuyết.
- Mà **trị số  $P$  là xác suất dữ liệu  $D$  xảy ra nếu giả thuyết vô hiệu  $H_0$  là đúng.**
- Trị số  $P$  không trực tiếp cho ta một ý niệm gì về giả thuyết  $H_1$ , nó chỉ gián tiếp cung cấp bằng chứng để ta chấp nhận hoặc bác bỏ giả thuyết vô hiệu  $H_0$ .
- Logic của trị số  $P$  là quá trình chứng minh đảo ngược (proof by contradiction):
  - Mệnh đề 1: Nếu giả thuyết vô hiệu đúng, thì sự kiện này không thể xảy ra.
  - Mệnh đề 2: Sự kiện xảy ra.
  - Mệnh đề 3 (kết luận): Giả thuyết vô hiệu không thể đúng.
- Do đó, trị số  $P$  gián tiếp phản ánh xác suất của mệnh đề 3. Nó ước tính mức độ khả dĩ của dữ liệu  $P(D|H_0)$ , chứ không nói cho ta biết mức độ khả dĩ của giả thuyết.
- Thực tế, ta cần biết với dữ liệu  $D$  đang có, thì xác suất  $H_1$  đúng là bao nhiêu, tức  $P(H_1|D)$ , chứ ta không cần biết  $P(D|H_0)$ .

## 2.4. Vấn đề của trị số P

### Vấn đề kiểm định nhiều giả thuyết

- Trong thực tế, ta có thể kiểm định nhiều giả thuyết cùng lúc.
  - VD: ta phân tích tính hiệu quả của thuốc Ramipril qua nhiều tiêu chí như nguy cơ tử vong, nhồi máu cơ tim, đột quỵ, tỉ lệ mắc bệnh tiểu đường... của từng giới tính, từng nhóm tuổi... Mỗi phân tích có thể xem như một kiểm định giả thuyết.
- Vấn đề: nếu ta kiểm định nhiều giả thuyết, và mỗi giả thuyết ta chấp nhận một sai sót 5%. Vậy trong n thử nghiệm, ta tuyên bố k thử nghiệm có ý nghĩa thống kê (tức  $P < 0.05$ ), thì xác suất có ít nhất một giả thuyết sai là bao nhiêu?
  - VD: Nếu ta thử nghiệm 03 giả thuyết, xác suất mà ta đúng cả ba là  $0.95 \times 0.95 \times 0.95 = 0.8574$ . Như vậy, xác suất có ít nhất một sai lầm trong 3 tuyên bố có ý nghĩa thống kê là  $1 - 0.8574 = 0.1426$ . Tức khoảng 14%.
- Vậy làm thế nào?
  - Trước khi kiểm định nhiều giả thuyết, ta không nên giữ mức  $\alpha = 0.05$ , mà nên chia nhỏ con số này ra số giả thuyết. Như vậy, với 3 giả thuyết, ta nên dùng  $\alpha = 0.016$ . Chỉ khi nào trị số P nhỏ hơn giá trị này, ta mới có thể khẳng định 3 giả thuyết trên có ý nghĩa thống kê.

# Kiểm định thống kê và phân phối mẫu

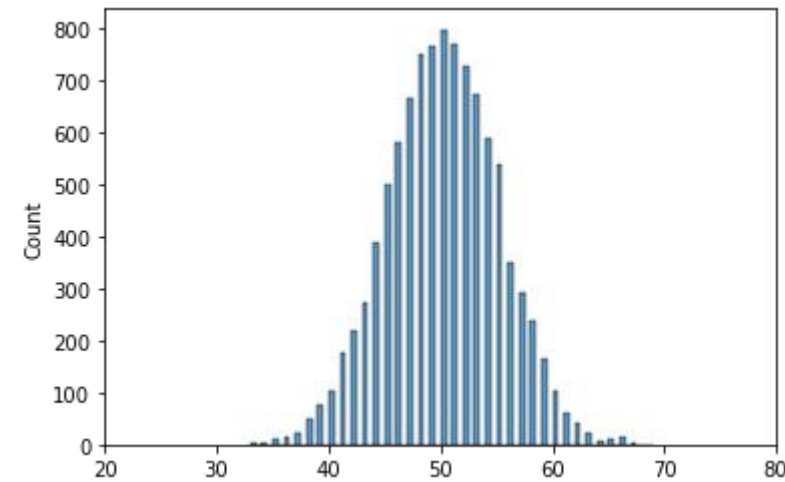
- Sau khi có dữ liệu, vấn đề là ta muốn nêu chính xác với giá trị nào của dữ liệu thì ta có thể bác bỏ giả thuyết  $H_0$  và giá trị nào thì chấp nhận  $H_0$ .
- Chúng ta có thể sử dụng phân phối để đánh giá.
- VD: Giả sử, có một bài toán mà tỷ lệ người có thể giải được là 50%. Vậy thử nghiệm với 100 người, có thể có 60 người giải được hay không?
- $X \sim \text{Binomial}(0, N)$

```
from myst_nb import glue
from numpy import random
import matplotlib.pyplot as plt
import seaborn as sns

# sample from a binomial distribution
data = random.binomial(n=100, p=.5,
size=10000)

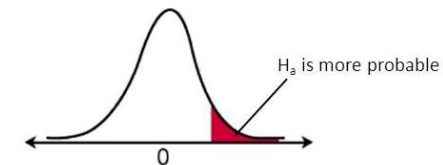
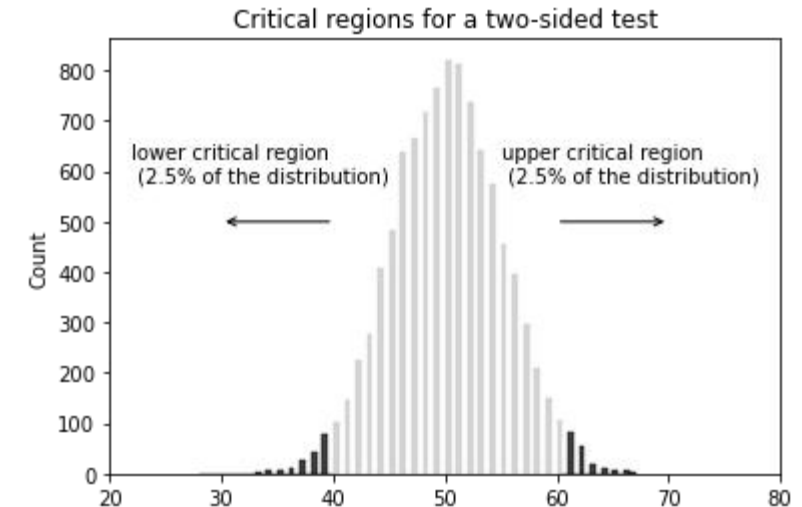
esp = sns.histplot(data,
bins=20, binwidth=0.5)
esp.set(xlim=(20, 80))

glue("estimation-fig", esp, display=False)
```



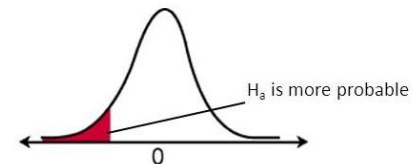
## 2.5. Ra quyết định trong kiểm định

- Miền bác bỏ (critical regions): Vùng tới hạn của một phép thử tương ứng với những giá trị mà tại đó ta bác bỏ giả thuyết vô hiệu.
  - X có thể rất lớn hoặc rất nhỏ để bác bỏ  $H_0$ .
  - Nếu  $\alpha = 0.05$ , vùng tới hạn phải phủ 5% của phân bố.
- Miền chấp nhận: là miền chứa các giá trị giúp  $H_0$  không bị bác bỏ.
- Giá trị tới hạn (critical value): Giá trị ngăn cách giữa miền bác bỏ và miền chấp nhận
- **Kiểm định một phía (one-sided):** Khi giả thuyết ngược lại  $H_1$  có tính chất một phía thì việc kiểm định được gọi là kiểm định một phía
  - $H_0: \theta \leq \theta_0$
  - $H_1: \theta > \theta_0$
- **Kiểm định hai phía (two-sided):** Khi giả thuyết ngược lại  $H_1$  có tính chất hai phía thì việc kiểm định được gọi là kiểm định hai phía
  - $H_0: \theta = \theta_0$
  - $H_1: \theta \neq \theta_0$



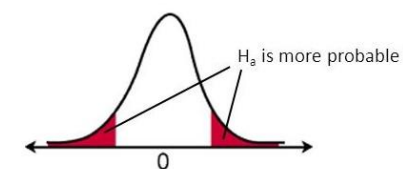
Right-tail test

$H_a: \mu > \text{value}$



Left-tail test

$H_a: \mu < \text{value}$



Two-tail test

$H_a: \mu \neq \text{value}$

# Các bước của việc kiểm định giả thuyết thống kê

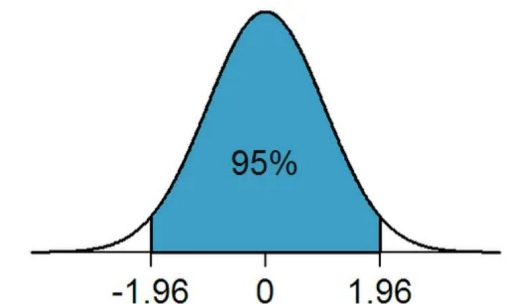
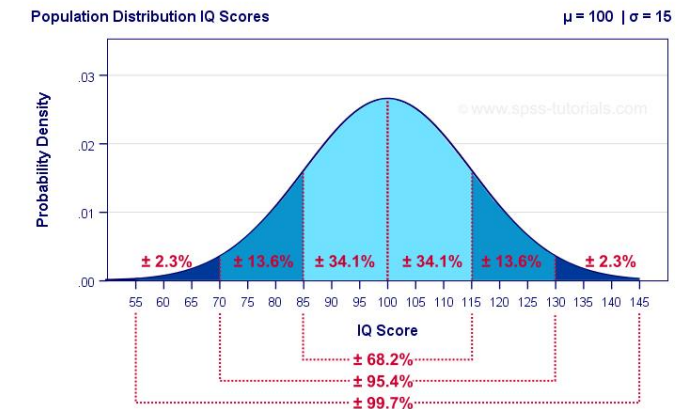
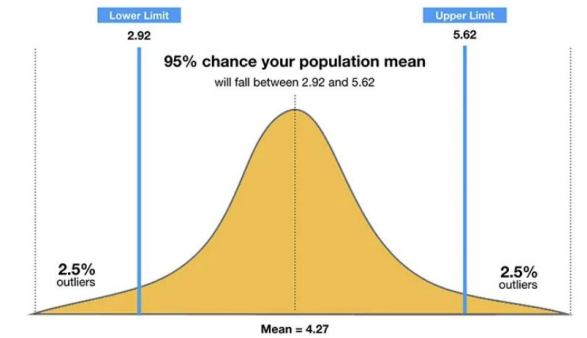
- Bước 1: Thành lập giả thuyết  $H_0$
- Bước 2: Thành lập giả thuyết  $H_1$
- Bước 3: Xác định mức ý nghĩa  $\alpha$
- Bước 4: Chọn các tham số thống kê thích hợp cho việc kiểm định và xác định các miền bác bỏ, miền chấp nhận và giá trị tới hạn.
- Bước 5: Tính toán các giá trị của tham số thống kê trong việc kiểm định.
- Bước 6: Ra quyết định: Nếu các giá trị tính toán rơi vào miền bác bỏ thì bác bỏ  $H_0$ . Ngược lại sẽ chấp nhận  $H_0$ .

# Kiểm định dựa vào khoảng tin cậy

- Khoảng tin cậy 95% (confidence interval) là khoảng giá trị mà ta tin 95% rằng nó chứa giá trị trung bình của biến.
- Công thức tính khoảng tin cậy

$$\bar{x} \pm z \frac{s}{\sqrt{n}}$$

- $\bar{x}$ : giá trị trung bình
- $z$ : giá trị z-score tại vị trí tới hạn (1.96, tính theo độ lệch chuẩn)
- $s$ : độ lệch chuẩn
- $n$ : kích thước mẫu



# Tóm tắt chương 2

- Mô hình kiểm định ý nghĩa thống kê (Fisher)
- Mô hình kiểm định giả thuyết thống kê (Neyman-Pearson)
- Mô hình RCT (nghiên cứu lâm sàng đối sánh ngẫu nhiên)
- Trị số P
- Sai lầm loại I vs. Sai lầm loại II
- Kiểm định một phía vs. hai phía
- Các bước của việc kiểm định giả thuyết thống kê