

THỐNG KÊ SUY DIỄN

Chương 2

KIỂM ĐỊNH GIẢ THUYẾT THAM SỐ

1. Những vấn đề chung
2. Kiểm định một tham số tổng thể
3. Kiểm định hai tham số tổng thể
4. Kiểm định số nhiều tham số tổng thể

1. Những vấn đề chung.

1.1. Mệnh đề phủ định và logic phản nghiệm

- a) Mệnh đề phủ định:
Phủ định của mệnh đề B là mệnh đề có chân trị ngược với mệnh đề B, kí hiệu $\bar{B}$
- b) Ứng dụng trong chứng minh, lập luận:
- Không có trực tiếp B nhưng có “không B”
→ Phương pháp chứng minh phản chứng

1. Những vấn đề chung.

1.1. Mệnh đề phủ định và logic phản nghiệm

- b) Ứng dụng trong chứng minh, lập luận:
Nhắc lại Phương pháp chứng minh phản chứng $A \rightarrow B$
B1: Giả sử: B sai $\Leftrightarrow$ **không** (phủ định) B đúng.
B2: Kết hợp với những giả thiết trong A, suy luận dẫn đến sự mâu thuẫn với quy luật hiển nhiên hoặc với một trong những giả thiết cho trước.
B3: Kết luận B phải đúng.
Ví dụ kinh điển: Chứng minh căn 2 là số vô tỷ.

1. Những vấn đề chung.

1.1. Mệnh đề phủ định và logic phản nghiệm

- b) Ứng dụng trong chứng minh, lập luận:
- B có số lượng lớn phần tử → không B ít hơn.
 - Ví dụ: Trường đại học X có 30,000 sinh viên. Một tuyên bố “Tất cả sinh viên trường đại học X đều giao tiếp tốt bằng tiếng Anh với người nước ngoài” → “có sinh viên trường đại học X không giao tiếp tốt bằng tiếng Anh với người nước ngoài”

1. Những vấn đề chung.

1.2. Các khái niệm

- a) Giả thiết – giả thuyết: giống hay khác?
- Giả thiết là những điều cho trước được đảm bảo là đúng, có thể dùng làm căn cứ, cơ sở suy luận ra những kết quả khác.
 - **Giả thuyết** là phát biểu khẳng định hoặc phủ định nhưng chưa biết được tính đúng sai.
- giả thuyết ≠ giả thiết

1. Những vấn đề chung.

1.2. Các khái niệm

- b) Thông tin:
- Thông tin chuẩn (chung/tổng thể): là thông tin **mặc định**, chắc chắn, xảy ra theo quy luật tự nhiên, thường kì, thường có trước, tuyên bố trước của vấn đề.
 - Thông tin thu thập: là thông tin có từ mẫu dữ liệu khảo sát, thu thập, ..v.v..
- Khi thông tin mẫu thu thập >!<< thông tin chuẩn → xuất hiện dấu hiệu nghi vấn, thắc mắc → giả thuyết lựa chọn.

1. Những vấn đề chung.

1.2. Các khái niệm

- c) **Giả thuyết lựa chọn** (alternative hypothesis): là giả thuyết nghi ngờ (suy đoán) của vấn đề, phù hợp và được lập nên từ thông tin mẫu → thể hiện bằng câu hỏi có/không?
- d) **Giả thuyết không** (null hypothesis): là giả thuyết mà phát biểu luôn có từ **không**, phản nghiệm (phủ định) giả thuyết lựa chọn. Kí hiệu **H₀**.
- e) **Kiểm định**: là kiểm tra thẩm định

1. Những vấn đề chung.

1.3. Kiểm định giả thuyết thống kê

- a) **Khái niệm**: là dùng thống kê toán kiểm tra và thẩm định tiêu chuẩn quyết định việc **bác bỏ** hoặc **chấp nhận** giả thuyết không **H₀**.
- b) **Cơ sở logic**: logic phản nghiệm
- c) **Phân loại**:
 - Theo tham số: Không và có (trung bình, tỷ lệ, phương sai / độ lệch chuẩn).
 - Theo số lượng tổng thể: một, hai và nhiều (>2)
 - Theo hình thức: hai phía (bằng – khác), một phía (hơn – kém).

1. Những vấn đề chung.

1.4. Sai lầm

- Loại I: **bác bỏ** giả thuyết mà bản chất nó **Đúng**.
- Loại II: **chấp nhận** giả thuyết mà bản chất nó **Sai**.
- Khả năng vướng hai sai lầm này có mối quan hệ tỉ lệ nghịch.

1.5. Mức ý nghĩa

Là khả năng (xác suất) chấp nhận sai lầm, kí hiệu **α** (bù 1 với độ tin cậy), được chọn trước khi kiểm định.

1. Những vấn đề chung.

1.6. Quy trình kiểm định

Bước 1: Thiết lập **giả thuyết không H_0** .
Nguyên tắc: Phủ định giả thuyết lựa chọn hoặc phù hợp với thông tin chuẩn (chung, tổng thể)
Bước 2: Từ dữ liệu mẫu, xác định **cỡ mẫu** và tính các tham số **đặc trưng mẫu** cần thiết có liên quan.
Bước 3: Từ mức ý nghĩa α , dùng các bảng tra phân vị xác suất $\rightarrow$ giá trị phân phối tới hạn $T(\alpha)$:
Hai phía: $T_{\alpha/2}$; **Một phía:** T_{α}

1. Những vấn đề chung.

1.6. Quy trình thực hiện

Bước 4: Tính giá trị kiểm định **T**.
Bước 5: **Quyết định**
So sánh giá trị kiểm định **T** với giá trị tới hạn **$T(\alpha)$**
 $\rightarrow$ **Bác bỏ** hoặc **Chấp nhận** giả thuyết không H_0

1. Những vấn đề chung.

1.7. Yêu cầu kiểm định tham số:

- Thỏa mãn ít nhất một trong hai điều sau:
- Biến (vấn đề) kiểm định có phân bố thường (chuẩn)
 - Cỡ mẫu lớn (>30) để có thể xấp xỉ theo phân bố thường (chuẩn)

2. Kiểm định một tham số tổng thể.

- 2.1. Kiểm định trung bình
- 2.2. Kiểm định tỉ lệ
- 2.3. Kiểm định phương sai / độ lệch chuẩn

2.1. Kiểm định trung bình

Phương pháp:

Bước 1: Thiết lập giả thuyết không H_0 :

- Hai phía “ $\mu = \mu_0$ ”
- Một phía “ $\mu \leq \mu_0$ ” hoặc “ $\mu \geq \mu_0$ ”

Bước 2: Tính các tham số mẫu: trung bình mẫu $\bar{x}$, độ lệch chuẩn mẫu hiệu chỉnh s.

Bước 3: Tra giá trị phân phối tới hạn $T(\alpha)$ (dựa vào mức ý nghĩa α)

2.1. Kiểm định trung bình

▪ **Trường hợp 1.** Cỡ mẫu $n > 30$ hoặc $n \leq 30$ nhưng có phương sai σ^2 / độ lệch chuẩn σ tổng thể → Bảng giá trị phân vị hàm Lapalce:

Hai phía	Một phía
$0,5 - \frac{\alpha}{2} = \varphi(Z_{\alpha/2}) \rightarrow Z_{\alpha/2}$	$0,5 - \alpha = \varphi(Z_{\alpha}) \rightarrow Z_{\alpha}$

▪ **Trường hợp 2.** Cỡ mẫu $n \leq 30$ và **không** có phương sai σ^2 / độ lệch chuẩn σ tổng thể → Bảng phân vị Students:

Hai phía: $t_{\alpha/2}^{n-1}$, một phía: t_{α}^{n-1}

2.1. Kiểm định trung bình

Bước 4: Tính giá trị kiểm định:

$$Z \text{ or } t = \frac{\bar{x} - \mu_0}{\frac{\sigma \text{ or } s}{\sqrt{n}}} = \frac{\bar{x} - \mu_0}{(\sigma \text{ or } s) \sqrt{n}}$$

Bước 5: Quyết định:

- $|Z \text{ or } t| > Z(\alpha) \text{ or } t(\alpha) \rightarrow$ bác bỏ H_0
- $|Z \text{ or } t| \leq Z(\alpha) \text{ or } t(\alpha) \rightarrow$ chấp nhận H_0

2.1. Kiểm định trung bình

VD 1. Sở Điện lực A báo cáo rằng: trung bình mỗi hộ hàng tháng phải trả 250 ngàn đồng tiền điện, với độ lệch chuẩn là 20 ngàn. Người ta khảo sát ngẫu nhiên 500 hộ thì tính được trung bình hàng tháng mỗi hộ trả 252 ngàn đồng tiền điện. Trong kiểm định giả thuyết không H_0 : “trung bình mỗi hộ phải trả hàng tháng là 250 ngàn đồng tiền điện” với mức ý nghĩa $\alpha = 1\%$, hãy cho biết giá trị kiểm định và kết luận ?

2.1. Kiểm định trung bình

VD 2. Một nghiên cứu cho rằng số giờ tự học trung bình hàng ngày của sinh viên hiện nay không thấp hơn so với mức 1 giờ/ngày cách đây 10 năm. Một mẫu khảo sát ngẫu nhiên 120 sinh viên và tính được trung bình là 0,82 giờ/ngày với $s = 0,7531$ giờ/ngày. Với mức ý nghĩa 3%, hãy cho biết tính chính xác của kết luận trên?

2.1. Kiểm định trung bình

VD3. Trong một nhà máy gạo, trọng lượng đóng bao theo quy định của 1 bao gạo là 50 kg và độ lệch chuẩn là 0,3 kg. Cân thử 296 bao gạo của nhà máy này thì thấy trọng lượng trung bình là 49,97 kg. Kiểm định giả thuyết không H_0 : “trọng lượng trung bình mỗi bao gạo của nhà máy này là 50 kg” có giá trị kiểm định Z và kết luận là:

A. $z = -1,7205$; chấp nhận H_0 với mức ý nghĩa 6%.
B. $z = -1,7205$; bác bỏ H_0 , trọng lượng thực tế của bao gạo nhỏ hơn 50 kg với mức ý nghĩa 6%.
C. $z = 1,9732$; chấp nhận H_0 với mức ý nghĩa 4%.
D. $z = 1,9732$; bác bỏ H_0 , trọng lượng thực tế của bao gạo nhỏ hơn 50 kg với mức ý nghĩa 4%.

2.1. Kiểm định trung bình

VD 4. Một công ty cho biết mức lương trung bình của 1 kỹ sư ở công ty không thấp hơn 5,7 triệu đồng/tháng với độ lệch chuẩn 0,5 triệu đồng/tháng. Kỹ sư A đã thăm dò 18 kỹ sư ở công ty này thì thấy lương trung bình là 5,45 triệu đồng/tháng. Kỹ sư A quyết định rằng: nếu mức lương trung bình thực sự bằng hay cao hơn mức lương công ty đưa ra thì nộp đơn xin làm. Với mức ý nghĩa 5%, cho biết kết luận của kỹ sư A ?

2.1. Kiểm định trung bình

VD 5. Người ta kiểm tra ngẫu nhiên 38 cửa hàng của công ty A và có bảng doanh thu trong 1 tháng là

X (triệu đồng/tháng)	200	220	240	260
Số cửa hàng	8	16	12	2

Kiểm định giả thuyết không H_0 : “doanh thu trung bình hàng tháng của mỗi cửa hàng công ty là 230 triệu đồng”, mức ý nghĩa tối đa để giả thuyết không H_0 được chấp nhận là:

A. 3,4%; B. 4,2%; C. 5,6%; D. 7,8%.

2.1. Kiểm định trung bình

VD 6. Điểm trung bình môn Toán của sinh viên năm trước là 5,72. Năm nay, theo dõi 100 SV được số liệu

Điểm	3	4	5	6	7	8	9
Số sinh viên	3	5	27	43	12	6	4

Kiểm định giả thuyết không H_0 : “**điểm trung bình môn Toán của sinh viên năm nay bằng năm trước**”, mức ý nghĩa tối đa đề H_0 được chấp nhận là:

A. 13,94%; B. 13,62%; C. 11,74%; D. 11,86%.

2.1. Kiểm định trung bình

VD 7. Thời gian X (phút) giữa hai chuyến xe bus trong thành phố là biến ngẫu nhiên có phân phối chuẩn. Công ty xe bus nói rằng: trung bình cứ 5 phút lại có 1 chuyến xe bus. Người ta chọn ngẫu nhiên 8 thời điểm và ghi lại thời gian (phút) giữa hai chuyến xe bus là:

5,3; 4,5; 4,8; 5,1; 4,3; 4,8; 4,9; 4,7.

Với mức ý nghĩa 7%, hãy kiểm định lời nói trên ?

2.1. Kiểm định trung bình

VD 8. Chiều cao cây giống X trong một vườn ươm là biến ngẫu nhiên có phân phối chuẩn. Người ta đo ngẫu nhiên 25 cây giống này và có bảng số liệu:

X (m)	0,8	0,9	1,0	1,1	1,2	1,3
Số cây	1	2	9	7	4	2

Theo quy định của vườn ươm, khi nào chiều cao trung bình của cây hơn 1m thì đem ra trồng. Với mức ý nghĩa 5%, hãy kiểm định xem có thể mang cây ra trồng được không?

2.2. Kiểm định tỉ lệ

Phương pháp:

Bước 1: Thiết lập giả thuyết không H_0 :

- Hai phía “ $p = p_0$ ”
- Một phía “ $p \leq p_0$ ” hoặc “ $p \geq p_0$ ”

Bước 2: Tính các tham số mẫu: tỉ lệ mẫu $f = \frac{m}{n}$
 m : số cá thể có tính chất đang quan tâm của mẫu

2.2. Kiểm định tỉ lệ

Bước 3: Tra giá trị phân phối tới hạn $T(\alpha)$ (dựa vào mức ý nghĩa α) → Bảng giá trị phân phối hàm Laplace:

Hai phía	Một phía
$0,5 - \frac{\alpha}{2} = \varphi(Z_{\alpha/2}) \rightarrow Z_{\alpha/2}$	$0,5 - \alpha = \varphi(Z_{\alpha}) \rightarrow Z_{\alpha}$

Bước 4: Tính giá trị kiểm định:

$$Z = \frac{f - p_0}{\sqrt{\frac{p_0(1 - p_0)}{n}}} = \frac{f - p_0}{\sqrt{p_0(1 - p_0)}} \sqrt{n}$$

2.2. Kiểm định tỉ lệ

Bước 5: Quyết định:

- $|Z| > Z(\alpha) \rightarrow$ bác bỏ giả thuyết không H_0
- $|Z| \leq Z(\alpha) \rightarrow$ chấp nhận giả thuyết không H_0

2.2. Kiểm định tỉ lệ

VD 9. Một báo cáo cho biết có 58% người tiêu dùng Việt Nam quan tâm đến hàng Việt. Khảo sát ngẫu nhiên 1.000 người dân Việt Nam thấy có 612 người được hỏi là có quan tâm đến hàng Việt. Với mức ý nghĩa 5%, hãy kiểm định lại báo cáo trên ?

2.2. Kiểm định tỉ lệ

VD 10. Khảo sát ngẫu nhiên 400 sinh viên về mức độ nghiêm túc trong giờ học thì thấy 13 sinh viên thừa nhận có ngủ trong giờ học. Trong kiểm định giả thuyết không H_0 : “**có 2% sinh viên ngủ trong giờ học**”, mức ý nghĩa tối đa là bao nhiêu để H_0 được chấp nhận?

2.2. Kiểm định tỉ lệ

VD 11. Để kiểm tra 1 loại súng thể thao, người ta cho bắn 1.000 viên đạn vào 1 tấm bia thì có 650 viên trúng mục tiêu. Dựa vào mẫu trên và mức ý nghĩa 3%, có thể kết luận tỉ lệ bắn trúng của loại súng thể thao này nhỏ hơn 70% hay không?

2.2. Kiểm định tỉ lệ

VD 12. Công ty A tuyên bố rằng có 40% người tiêu dùng ưa thích sản phẩm của mình. Một cuộc điều tra 400 người tiêu dùng thấy có 179 người ưa thích sản phẩm của công ty A. Trong kiểm định giả thuyết không H_0 : “**có 40% người tiêu dùng thích sản phẩm của công ty A**”, mức ý nghĩa tối đa để giả thuyết không H_0 được chấp nhận là:
A. 7,86%; B. 6,48%; C. 5,24%; D. 4,32%.

2.3. Kiểm định phương sai / độ lệch chuẩn

Phương pháp:

Bước 1: Thiết lập giả thuyết không H_0 :

	Phương sai	Độ lệch chuẩn
Hai phía	$\sigma^2 = \sigma_0^2$	$\sigma = \sigma_0$
Một phía	$\sigma^2 \leq \sigma_0^2$ hoặc $\sigma^2 \geq \sigma_0^2$	$\sigma \leq \sigma_0$ hoặc $\sigma \geq \sigma_0$

Bước 2: Tính các tham số mẫu: phương sai S^2 hoặc độ lệch chuẩn mẫu S hiệu chỉnh

2.3. Kiểm định phương sai / độ lệch chuẩn

Bước 3: Tra giá trị phân phối tới hạn $T(\alpha)$ (dựa vào mức ý nghĩa α) → Bảng giá trị phân phối Chi bình phương:

- Hai phía: $\chi^2_{\alpha/2;n-1}$ và $\chi^2_{1-\alpha/2;n-1}$
- Một phía: $\chi^2_{\alpha;n-1}$ và $\chi^2_{1-\alpha;n-1}$

Bước 4: Tính giá trị kiểm định:

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

2.3. Kiểm định phương sai / độ lệch chuẩn

Bước 5: Quyết định:

	Hai phía	Một phía
Bác bỏ H_0	$\chi^2 < \chi^2_{1-\alpha/2;n-1}$ or $\chi^2 > \chi^2_{\alpha/2;n-1}$	$\chi^2 < \chi^2_{1-\alpha;n-1}$ or $\chi^2 > \chi^2_{\alpha;n-1}$
Chấp nhận H_0	$\chi^2_{1-\alpha/2;n-1} \leq \chi^2 \leq \chi^2_{\alpha/2;n-1}$	$\chi^2_{1-\alpha;n-1} \leq \chi^2 \leq \chi^2_{\alpha;n-1}$

2.3. Kiểm định phương sai / độ lệch chuẩn

Ví dụ. Một máy tiện quy định độ lệch chuẩn đường kính trục máy không quá 6. Người ta tiến hành 25 quan sát về đường kính trục máy và tính được phương sai là 36,266. Với mức ý nghĩa 10%, ta có kết luận thể nào về quá trình sản xuất.

3. Kiểm định hai tham số tổng thể

- 3.1. Kiểm định hai phương sai / độ lệch chuẩn
- 3.2. Kiểm định hai trung bình với hai mẫu phối hợp cặp
- 3.3. Kiểm định hai trung bình với hai mẫu độc lập
- 3.4. Kiểm định hai tỉ lệ

3.1. Kiểm định hai phương sai / ĐLC

Phương pháp:

Bước 1: Thiết lập giả thuyết không H_0 :

	Phương sai	Độ lệch chuẩn
Hai phía	$\sigma_A^2 = \sigma_B^2$	$\sigma_A = \sigma_B$
Một phía	$\sigma_A^2 \leq \sigma_B^2$ hoặc $\sigma_A^2 \geq \sigma_B^2$	$\sigma_A \leq \sigma_B$ hoặc $\sigma_A \geq \sigma_B$

Bước 2: Tính các tham số mẫu: phương sai S^2 hoặc độ lệch chuẩn mẫu S hiệu chỉnh hai mẫu

3.1. Kiểm định hai phương sai / ĐLC

Bước 3: Tính giá trị kiểm định:

$$F = \frac{S_A^2}{S_B^2} \text{ or } \frac{S_B^2}{S_A^2} > 1$$

Bước 4: Tra giá trị phân phối tới hạn $T(\alpha)$ (dựa vào mức ý nghĩa α) $\rightarrow$ Bảng giá trị phân vị Fisher:

Hai phía: $F_{\alpha/2}^{n_{tu}-1; n_{mẫu}-1}$, **một phía** $F_{\alpha}^{n_{tu}-1; n_{mẫu}-1}$

Bước 5: Quyết định:

- $F > F(\alpha) \rightarrow$ bác bỏ giả thuyết không H_0
- $F \leq F(\alpha) \rightarrow$ chấp nhận giả thuyết không H_0

3.1. Kiểm định hai phương sai

Ví dụ. Để kiểm tra độ chính xác của hai chiếc máy tiện, người ta chọn ngẫu nhiên từ máy X ra 16 sản phẩm và máy Y ra 13 sản phẩm. Phương sai về đường kính hai mẫu sản phẩm lần lượt là 17 và 26. Với mức ý nghĩa 5%, có thể kết luận hai máy có độ lệch chuẩn khác nhau không?

3.2. Kiểm định hai trung bình mẫu phối cặp

- a) **Điều kiện:**
Hai tổng thể cùng tồn tại cùng một môi trường hoặc được tạo ra từ một tổng thể ban đầu.
- b) **Dấu hiệu:**
Cỡ hai mẫu luôn bằng nhau và dữ liệu được liệt kê phối hợp sẵn thành từng cặp cá thể.
- c) **Phương pháp:**
Bước 1: Thiết lập giả thuyết không H_0 :

3.2. Kiểm định hai trung bình mẫu phối cặp

- Hai phía “ $\mu_A - \mu_B = \mu_0$ ”
- Một phía “ $\mu_A - \mu_B \leq \mu_0$ ” hoặc “ $\mu_A - \mu_B \geq \mu_0$ ”
- Nếu $\mu_0 = 0$ thì:
- Hai phía “ $\mu_A = \mu_B$ ”
- Một phía “ $\mu_A \leq \mu_B$ ” hoặc “ $\mu_A \geq \mu_B$ ”
- Bước 2: Tính các tham số mẫu: tính chênh lệch từng cặp cá thể $d_i = a_i - b_i \rightarrow$ trung bình mẫu $\bar{d}$ và độ lệch chuẩn mẫu hiệu chỉnh S_d của những chênh lệch này.

3.2. Kiểm định hai trung bình mẫu phối cặp

- Bước 3: Tra giá trị phân phối tới hạn $T(\alpha)$ (dựa vào mức ý nghĩa α)
- **Trường hợp 1.** Cỡ mẫu $n \leq 30 \rightarrow$ Bảng giá trị PP Students:

Hai phía: $t_{\alpha/2}^{n-1}$, một phía: t_{α}^{n-1}

▪ **Trường hợp 2.** Cỡ mẫu $n > 30 \rightarrow$ Bảng giá trị hàm Laplace

Hai phía	Một phía
$0,5 - \frac{\alpha}{2} = \varphi(Z_{\alpha/2}) \rightarrow Z_{\alpha/2}$	$0,5 - \alpha = \varphi(Z_{\alpha}) \rightarrow Z_{\alpha}$

3.2. Kiểm định hai trung bình mẫu phối cặp

Bước 4: Tính giá trị kiểm định:

$$t \text{ or } Z = \frac{\bar{d} - \mu_0}{\frac{S_d}{\sqrt{n}}} = \frac{\bar{d} - \mu_0}{S_d} \sqrt{n}$$

Bước 5: Quyết định:

- $|Z \text{ or } t| > Z(\alpha) \text{ or } t(\alpha) \rightarrow$ bác bỏ H_0
- $|Z \text{ or } t| \leq Z(\alpha) \text{ or } t(\alpha) \rightarrow$ chấp nhận H_0

3.2. Kiểm định hai trung bình mẫu phối cặp

VD. Một công ty thực hiện so sánh hai hình thức bán hàng một loại sản phẩm. Số liệu thu được từ 10 cửa hàng về số lượng bán hàng sản phẩm cho trong bảng. Hãy kiểm định sự khác biệt về trung bình của hai hình thức trên ở độ tin cậy 95%.

A	50	48	45	60	70	62	55	62	58	53
B	52	46	50	65	78	61	58	70	67	65

3.3. Kiểm định hai trung bình mẫu độc lập

Phương pháp:

Bước 1: Thiết lập giả thuyết không H_0 :

- Hai phía “ $\mu_A - \mu_B = \mu_0$ ”
- Một phía “ $\mu_A - \mu_B \leq \mu_0$ ” hoặc “ $\mu_A - \mu_B \geq \mu_0$ ”

Nếu $\mu_0 = 0$ thì:

- Hai phía “ $\mu_A = \mu_B$ ”
- Một phía “ $\mu_A \leq \mu_B$ ” hoặc “ $\mu_A \geq \mu_B$ ”

Bước 2: Tính các tham số mẫu: hai trung bình mẫu $\bar{x}_A, \bar{x}_B$ và hai độ lệch chuẩn mẫu hiệu chỉnh S_A, S_B .

3.3. Kiểm định hai trung bình mẫu độc lập

Bước 3: Tra giá trị phân phối tới hạn $T(\alpha)$ (dựa vào mức ý nghĩa α)

- Trường hợp 1.** Cỡ hai mẫu $n_A + n_B - 2 > 30$ hoặc $n_A + n_B - 2 \leq 30$ nhưng có phương sai / độ lệch chuẩn tổng thể $\rightarrow$ Phân vị thường (chuẩn) hàm Laplace:

Hai phía	Một phía
$0,5 - \frac{\alpha}{2} = \varphi(Z_{\alpha/2}) \rightarrow Z_{\alpha/2}$	$0,5 - \alpha = \varphi(Z_{\alpha}) \rightarrow Z_{\alpha}$

3.3. Kiểm định hai trung bình mẫu độc lập

Bước 3: Tra giá trị phân phối tới hạn $T(\alpha)$ (dựa vào mức ý nghĩa α)

- Trường hợp 2.** Cỡ hai mẫu $n_A + n_B - 2 \leq 30$ và không có phương sai / độ lệch chuẩn tổng thể $\rightarrow$ Bảng giá trị PP Students:
 - i) Phương sai hai tổng thể khác nhau:

Bậc tự do: $n = \frac{\frac{\left(\frac{S_A^2}{n_A} + \frac{S_B^2}{n_B}\right)^2}{\frac{S_A^2}{n_A} + \frac{S_B^2}{n_B}}}{\frac{\left(\frac{S_A^2}{n_A}\right)^2}{n_A - 1} + \frac{\left(\frac{S_B^2}{n_B}\right)^2}{n_B - 1}}$, Hai phía: $t_{\alpha/2}^n$, một phía: t_{α}^n

3.3. Kiểm định hai trung bình mẫu độc lập

Bước 3: Tra giá trị phân phối tới hạn $T(\alpha)$ (dựa vào mức ý nghĩa α)

- Trường hợp 2.** Cỡ hai mẫu $n_A + n_B - 2 \leq 30$ và không có phương sai / độ lệch chuẩn tổng thể $\rightarrow$ Bảng giá trị PP Students:
 - ii) Phương sai hai tổng thể bằng nhau:

Hai phía: $t_{\alpha/2}^{n_A + n_B - 2}$, một phía: $t_{\alpha}^{n_A + n_B - 2}$

3.3. Kiểm định hai trung bình mẫu độc lập

Bước 4: Tính giá trị kiểm định:

- Trường hợp 1.** Cỡ hai mẫu $n_A + n_B - 2 > 30$ hoặc $n_A + n_B - 2 \leq 30$ nhưng có phương sai / độ lệch chuẩn tổng thể:

$$Z = \frac{\bar{x}_A - \bar{x}_B - \mu_0}{\sqrt{\frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B}}}$$

Nếu cỡ hai mẫu $n_A + n_B - 2 > 30$ có thể sử dụng phương sai mẫu thay cho phương sai tổng thể.

3.3. Kiểm định hai trung bình mẫu độc lập

Bước 4: Tính giá trị kiểm định:

- Trường hợp 2.** Cỡ hai mẫu $n_A + n_B - 2 \leq 30$ và không có phương sai / độ lệch chuẩn tổng thể:

i) Phương sai hai tổng thể khác nhau:

$$t = \frac{\bar{x}_A - \bar{x}_B - \mu_0}{\sqrt{\frac{s_A^2}{n_A} + \frac{s_B^2}{n_B}}}$$

3.3. Kiểm định hai trung bình mẫu độc lập

Bước 4: Tính giá trị kiểm định:

- Trường hợp 2.** Cỡ hai mẫu $n_A + n_B - 2 \leq 30$ và không có phương sai / độ lệch chuẩn tổng thể:

ii) Phương sai hai tổng thể bằng nhau:

- Tính phương sai mẫu chung:

$$s^2 = \frac{(n_A - 1)s_A^2 + (n_B - 1)s_B^2}{n_A + n_B - 2}$$

- Giá trị kiểm định:

$$t = \frac{\bar{x}_A - \bar{x}_B - \mu_0}{\sqrt{\frac{s^2}{n_A} + \frac{s^2}{n_B}}}$$

3.3. Kiểm định hai trung bình mẫu độc lập

Bước 5: Quyết định:

- $|Z \text{ or } t| > Z(\alpha) \text{ or } t(\alpha) \rightarrow$ bác bỏ H_0
- $|Z \text{ or } t| \leq Z(\alpha) \text{ or } t(\alpha) \rightarrow$ chấp nhận H_0

3.3. Kiểm định hai trung bình mẫu độc lập

VD 1. Người ta tiến hành bón hai loại phân X, Y cho cây cà chua. Với 60 cây được bón phân X thì thu được trung bình 32,2 quả và độ lệch chuẩn của mẫu đã hiệu chỉnh là 8,5 quả; 72 cây được bón phân Y thu được trung bình 28,4 quả và độ lệch chuẩn của mẫu đã hiệu chỉnh là 9,3 quả. Với mức ý nghĩa 5%, hãy cho biết kết luận về sự khác biệt hai loại phân bón trên ?

3.3. Kiểm định hai trung bình mẫu độc lập

VD 2. Để so sánh mức lương trung bình của nhân viên nữ X (USD/giờ) và nam Y (USD/giờ) ở một công ty đa quốc gia, người ta tiến hành khảo sát ngẫu nhiên 100 nữ và 75 nam thì có kết quả:

$\bar{x} = 7,23, s_x = 1,64$ và $\bar{y} = 8,06, s_y = 1,85$.

Với mức ý nghĩa 3%, kiểm định giả thuyết không H_0 : “**mức lương trung bình nhân viên nam không cao hơn mức lương trung bình nhân viên nữ**” có giá trị điểm định và kết luận là:

- A. $z = 4,0957$; mức lương TB của nữ cao hơn nam.
- B. $z = 4,0957$; mức lương TB của nữ thấp hơn nam.
- C. $z = 3,0819$; mức lương TB của nữ cao hơn nam.
- D. $z = 3,0819$; mức lương TB của nữ thấp hơn nam.

3.3. Kiểm định hai trung bình mẫu độc lập

VD 3. Tuổi thọ (năm) của pin là biến ngẫu nhiên có phân phối chuẩn. Một công ty sản xuất thử nghiệm 10 chiếc pin loại X và 12 chiếc pin loại Y thì có kết quả:

$\bar{x} = 4,8, s_x = 1,1$ và $\bar{y} = 4,3, s_y = 0,7$.

Với mức ý nghĩa 10%, có thể kết luận “tuổi thọ trung bình của loại pin X cao hơn loại pin Y” được không?

3.3. Kiểm định hai trung bình mẫu độc lập

VD 4. Tuổi thọ (tháng) của thiết bị là biến ngẫu nhiên có phân phối chuẩn. Người ta kiểm tra ngẫu nhiên tuổi thọ của 16 thiết bị loại A , có kết quả:

114; 78; 96; 137; 78; 103; 126; 86;
99; 114; 72; 104; 73; 86; 117; 88.

Kiểm tra tuổi thọ 16 thiết bị loại B thấy có trung bình là 84 tháng và độ lệch chuẩn đã hiệu chỉnh là 19 tháng. Kiểm định giả thuyết không H_0 : “**tuổi thọ trung bình của thiết bị loại A và B là như nhau với mức ý nghĩa 5%**” có giá trị kiểm định và kết luận là:

A. $t = 2,0817$; tuổi thọ TB của 2 loại thiết bị khác nhau.
B. $t = 2,0817$; tuổi thọ TB của loại thiết bị A lớn hơn.
C. $t = 2,4616$; tuổi thọ TB của 2 loại thiết bị như nhau.
D. $t = 2,4616$; tuổi thọ TB của loại thiết bị A nhỏ hơn.

3.4. Kiểm định hai tỉ lệ

Phương pháp:

Bước 1: Thiết lập giả thuyết không H_0 :

- Hai phía “ $p_A - p_B = p_0$ ”
- Một phía “ $p_A - p_B \leq p_0$ ” hoặc “ $p_A - p_B \geq p_0$ ”

Nếu $p_0 = 0$ thì:

- Hai phía “ $p_A = p_B$ ”
- Một phía “ $p_A \leq p_B$ ” hoặc “ $p_A \geq p_B$ ”

3.4. Kiểm định hai tỉ lệ

Bước 2: Tính các tham số mẫu:

- Tỉ lệ hai mẫu: $f_A = \frac{m_A}{n_A}, f_B = \frac{m_B}{n_B}$
- Tỉ lệ mẫu chung: $f_0 = \frac{m_A + m_B}{n_A + n_B}$.

Bước 3: Tra giá trị phân phối tới hạn $T(\alpha)$ (dựa vào mức ý nghĩa α) $\rightarrow$ Bảng giá trị phân phối hàm Laplace:

Hai phía	Một phía
$0,5 - \frac{\alpha}{2} = \varphi(Z_{\alpha/2}) \rightarrow Z_{\alpha/2}$	$0,5 - \alpha = \varphi(Z_{\alpha}) \rightarrow Z_{\alpha}$

3.4. Kiểm định hai tỉ lệ

Bước 4: Tính giá trị kiểm định:

$$Z = \frac{f_A - f_B - p_0}{\sqrt{f_0(1 - f_0) \left(\frac{1}{n_A} + \frac{1}{n_B} \right)}}$$

Bước 5: Quyết định:

- $|Z| > Z(\alpha) \rightarrow$ bác bỏ giả thuyết không H_0
- $|Z| \leq Z(\alpha) \rightarrow$ chấp nhận giả thuyết không H_0

3.4. Kiểm định hai tỉ lệ

VD 5. Từ hai tổng thể X và Y người ta tiến hành kiểm tra 2 mẫu có kích thước $n_x = 1000, n_y = 1200$ về một tính chất A thì được $f_x = 0,27$ và $f_y = 0,3$. Với mức ý nghĩa 9%, hãy so sánh tỉ lệ p_x, p_y của hai tổng thể ?

3.4. Kiểm định hai tỉ lệ

VD 6. Kiểm tra 120 sản phẩm ở kho I thấy có 6 phế phẩm; 200 sản phẩm ở kho II thấy có 24 phế phẩm. Hỏi chất lượng hàng ở hai kho có khác nhau không với:
1) mức ý nghĩa 5%; 2) mức ý nghĩa 3%.

3.4. Kiểm định hai tỉ lệ

VD 7. Một công ty điện tử nghiên cứu thị trường về sở thích xem tivi của cư dân trong thành phố. Hỏi 400 người ở quận *X* thì có 270 người xem tivi ít nhất 1 giờ trong 1 ngày; 600 người ở quận *Y* có 450 người xem tivi ít nhất 1 giờ trong 1 ngày. Kiểm định giả thuyết không *H*₀: “**tỉ lệ cư dân xem tivi ít nhất 1 giờ trong 1 ngày ở quận X và Y như nhau**”, mức ý nghĩa tối đa đề *H* được chấp nhận là:
A. 0,96%; B. 2,84%; C. 4,06%; D. 6,14%.

3.4. Kiểm định hai tỉ lệ

VD 8. Trước bầu cử, người ta thăm dò 1000 cử tri thì thấy có 400 người nói rằng sẽ bỏ phiếu cho ông *A*. Một tuần sau (vẫn chưa bầu cử), người ta tổ chức 1 cuộc thăm dò khác và thấy có 680 trong số 1500 cử tri được hỏi sẽ bỏ phiếu cho ông *A*. Với mức ý nghĩa 5%, tỉ lệ đợt thăm dò sau có hơn đợt trước hay không?

BÀI TẬP TỔNG HỢP

Bài 1. Kiểm tra lượng kẹo X được bán ra hàng ngày tại một cửa hàng, có kết quả

X (kg)	25	30	35	40	45	50	55
Số ngày	9	23	27	30	25	20	5

- 1) Dựa vào mẫu trên, nếu muốn ước lượng lượng kẹo trung bình được bán ra hàng ngày ở cửa hàng này có độ chính xác là 1,231 kg thì đảm bảo độ tin cậy là bao nhiêu?
- 2) Hãy ước lượng tỉ lệ những ngày cửa hàng bán được nhiều hơn 40 kg với độ tin cậy 90%?
- 3) Bằng cách hạ giá bán, cửa hàng đã bán được lượng kẹo trung bình hàng ngày là 40,5 kg. Với mức ý nghĩa 5% hãy cho kết luận thực tế về việc hạ giá này?

BÀI TẬP TỔNG HỢP

Bài 2. Lượng chất đạm của gà ta thả vườn đạt chuẩn là cao hơn hay bằng 252g mỗi con. Khảo sát lượng chất đạm X (g) của loại gà này tại một nông trại, có kết quả

X (g)	248	250	252	254	256	258
Số con	5	17	31	45	19	3

- 1) Ước lượng lượng chất đạm trung bình có trong mỗi con gà của nông trại trên với độ tin cậy 95% ?
- 2) Để có thể nói tỉ lệ gà đạt chuẩn về chất đạm tại nông trại trên là 87% thì mức ý nghĩa tối đa là bao nhiêu ?

BÀI TẬP TỔNG HỢP

Bài 2. Lượng chất đạm của gà ta thả vườn đạt chuẩn là cao hơn hay bằng 252g mỗi con. Khảo sát lượng chất đạm X (g) của loại gà này tại một nông trại, có kết quả

X (g)	248	250	252	254	256	258
Số con	5	17	31	45	19	3

- 3) Người ta thử nuôi nhốt loại gà này cho đến lúc gà có cùng trọng lượng với cách nuôi như trên thì thấy lượng đạm trung bình là 249g. Với mức ý nghĩa 5% hãy cho kết luận thực tế của cách nuôi nhốt ?

4. Kiểm định số nhiều tham số tổng thể

4.1. Phân tích phương sai một nhân tố

4.2. Kiểm định bằng nhau số nhiều tỉ lệ

4.1. Phân tích phương sai một nhân tố.

- a) Phân tích phương sai (ANOVA):
- Là phương pháp kiểm định sự bằng nhau **số nhiều (>2) trung bình tổng thể** do một hoặc vài nhân tố định tính (độc lập) gây tác động, ảnh hưởng tạo nên.
 - Nói cách khác: là phương pháp kiểm định có hay không sự gây tác động, ảnh hưởng của một vài nhân tố định tính độc lập đến một nhân tố (biến) định lượng phụ thuộc.
 - Điều kiện: Phương sai trong các tổng thể bằng nhau (mặc định luôn thỏa mãn).

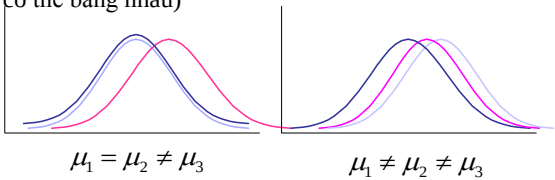
4.1. Phân tích phương sai một nhân tố.

- a) Phân tích phương sai (ANOVA):
- Ví dụ: Khối ngành học (kĩ thuật, kinh tế, xã hội) có ảnh hưởng đến việc học Anh văn của sinh viên Trường đại học X hay không?
 - Một nhân tố định tính gây ảnh hưởng: **Khối ngành học**. Nhân tố này chia tổng thể sinh viên Trường đại học X thành ba tổng thể: SV khối kĩ thuật, SV khối kinh tế, SV khối xã hội
 - Kiểm định sự bằng nhau điểm trung bình học Anh văn của ba tổng thể trên → Phân tích phương sai một nhân tố.

4.1. Phân tích phương sai một nhân tố.

b) Các giả thuyết:

- Giả thuyết lựa chọn: Nhân tố tác động có ảnh hưởng đến biến phụ thuộc $\Leftrightarrow$ Có sự khác nhau k trung bình tổng thể (**một số** trung bình tổng thể khác biệt với những trung bình tổng thể còn lại, các tổng thể còn lại có thể bằng nhau)

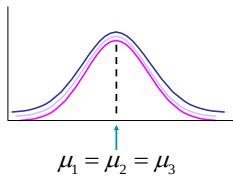


4.1. Phân tích phương sai một nhân tố.

b) Các giả thuyết:

- Giả thuyết không H_0 : Nhân tố tác động **không** ảnh hưởng đến biến phụ thuộc $\Leftrightarrow$ **Không** có sự khác nhau giữa k trung bình tổng thể

$\Leftrightarrow \mu_1 = \mu_2 = \dots = \mu_k = \mu_0$



4.1. Phân tích phương sai một nhân tố.

c) Phương pháp:

Giả sử có k tổng thể, gọi các kí hiệu sau:

- $n_1, n_2, \dots, n_k$ là k cỡ mẫu, n là cỡ bộ mẫu (k mẫu):
- ➔ $n = n_1 + n_2 + \dots + n_k$
- x_{ij} là giá trị thứ j của mẫu i, $1 \leq i \leq k, 1 \leq j \leq n_i$

4.1. Phân tích phương sai một nhân tố.

c) Phương pháp:

❑ Bước 1: Tính các giá trị trung bình

➤ Từng mẫu:

$$\bar{x}_i = \frac{\sum_{j=1}^{n_i} x_j}{n_i}, i = 1, 2, \dots, k$$

➤ Bộ mẫu:

$$\bar{x} = \frac{\sum_{i=1}^k \sum_{j=1}^{n_i} x_{ij}}{n} = \frac{\sum_{i=1}^k n_i \bar{x}_i}{n}$$

4.1. Phân tích phương sai một nhân tố.

c) Phương pháp:

❑ Bước 2: Tính các tổng bình phương độ lệch

➤ Tổng BPĐL nội bộ từng mẫu:

$$\begin{aligned} SS_i &= \sum_{j=1}^{n_i} x_{ij} - \bar{x}_i^2 = \sum_{j=1}^{n_i} x_{ij}^2 - n_i \bar{x}_i^2 \\ &= n_i \hat{S}_i^2 = n_i - 1 S_i^2 \end{aligned}$$

4.1. Phân tích phương sai một nhân tố.

c) Phương pháp:

❑ Bước 2: Tính các tổng bình phương độ lệch

➤ Tổng của các tổng BPĐL nội bộ:

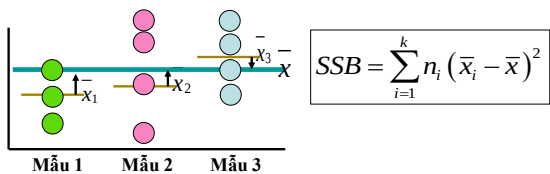
$$SSW = \sum_{i=1}^k SS_i = SS_1 + SS_2 + \dots + SS_k$$

Mẫu 1 Mẫu 2 Mẫu 3

4.1. Phân tích phương sai một nhân tố.

- c) Phương pháp:
- ❑

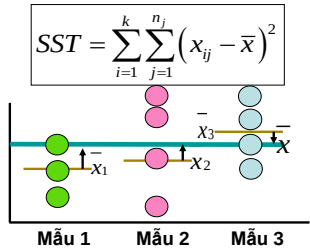
Bước 2: Tính các tổng bình phương độ lệch
 - Tổng BPĐL giữa trung bình từng mẫu với trung bình bộ mẫu:



4.1. Phân tích phương sai một nhân tố.

- c) Phương pháp:
- ❑

Bước 2: Tính các tổng bình phương độ lệch
 - Tổng BPĐL bộ mẫu:

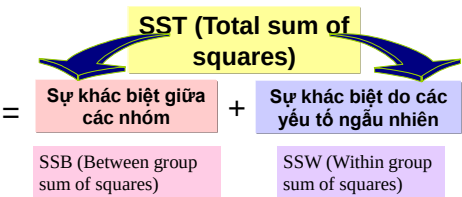


4.1. Phân tích phương sai một nhân tố.

- c) Phương pháp:
- ❑

Bước 2: Tính các tổng bình phương độ lệch

Ta có: $SST = SSW + SSB$



4.1. Phân tích phương sai một nhân tố.

- c) Phương pháp:
- ☐ Bước 3: Tính các ước lượng phương sai

$$MSW = \frac{SSW}{n - k}; MSB = \frac{SSB}{k - 1}$$
- ☐ Bước 4: Tính giá trị kiểm định

$$F = \frac{MSB}{MSW}$$
- ☐ Bước 5: Giá trị phân phối tới hạn: $F_{\alpha; k-1; n-k}$

→ Bác bỏ giả thuyết không H_0 khi $F > F_{\alpha; k-1; n-k}$

4.1. Phân tích phương sai một nhân tố.

c) Phương pháp:

Bảng ANOVA				
Biến thiên	Tổng bình phương độ lệch	Bậc tự do	Ước lượng phương sai	Giá trị kiểm định
Giữa các nhóm	SSB	$k - 1$	MSB	$F = MSB/MSW$
Nội bộ nhóm	SSW	$n - k$	MSW	
Tổng cộng	SST	$n - 1$		

4.1. Phân tích phương sai một nhân tố.

- d) Kiểm định từng cặp (sâu ANOVA, Turkey):
- Nếu giả thuyết không H_0 bị bác bỏ → có ít nhất một trung bình tổng thể không bằng những tổng thể khác
→ so sánh từng cặp trung bình tổng thể với nhau giúp tìm ra những trung bình tổng thể khác biệt đó.

Với k tổng thể thì số cặp kiểm định là:

$$\frac{k(k - 1)}{2}$$

4.1. Phân tích phương sai một nhân tố.

- d) Kiểm định từng cặp (sâu ANOVA, Turkey):
- Giả thuyết không cặp (f;g): $\mu_f = \mu_g$
 - Giá trị kiểm định:

$$D_{f;g} = \left| \bar{x}_f - \bar{x}_g \right|$$

4.1. Phân tích phương sai một nhân tố.

- d) Kiểm định từng cặp (sâu ANOVA, Turkey):
- Giá trị tới hạn:

$$T = q_{\alpha;k;n-k} \cdot \sqrt{\frac{MSW}{\min\{n_i\}}}$$

$q_{(\alpha;k;n-k)}$: bảng tra phân phối Turkey

- Bác bỏ giả thuyết không Ho khi:

$$D_{f;g} \geq T$$

4.1. Phân tích phương sai một nhân tố.

- e) Ví dụ:

Hiệp hội bảo vệ người tiêu dùng thực hiện một nghiên cứu để so sánh độ bền một loại vỏ xe của 3 nhãn hiệu. Chọn ngẫu nhiên một số sản phẩm của mỗi nhãn hiệu, đo độ bền (ngàn km). Độ bền 3 sản phẩm này có bằng nhau không với độ tin cậy 95%?	A	B	C
	25.40	23.40	20.00
	26.31	21.80	22.20
	24.10	23.50	19.75
	23.74	22.75	20.60
	25.10	21.60	20.40

4.2. Kiểm định bằng nhau số nhiều tỉ lệ

Phương pháp:

Bước 1: Thiết lập giả thuyết không H_0 :

$$p_1 = p_2 = \dots = p_k = 1/k.$$

Bước 2: Tính tần số tổng thể (lý thuyết):

$$E_i = n \cdot p_i = n/k = E; i = 1; 2; \dots; k$$

Bước 3: Tra giá trị tới hạn (từ mức ý nghĩa) $\rightarrow$ phân vị Chi bình phương $\chi^2_{(\alpha; k-1)}$

4.2. Kiểm định bằng nhau số nhiều tỉ lệ

Bước 4: Giá trị kiểm định:
$$\chi^2 = \frac{\sum_{i=1}^k O_i - E^2}{E}$$

 O_i : tần số mẫu i

Bước 5: Quyết định:

- $\chi^2 > \chi^2_{\alpha; k-1}$: bác bỏ giả thuyết không H_0
- $\chi^2 \leq \chi^2_{\alpha; k-1}$: chấp nhận giả thuyết không H_0

4.2. Kiểm định bằng nhau số nhiều tỉ lệ

Ví dụ. Một công ty muốn khảo sát xem có sự khác nhau của khách hàng với 4 màu sắc khác nhau sản phẩm xe gắn máy mới có khác nhau không? Bộ phận nghiên cứu chọn ngẫu nhiên 120 khách hàng tiềm năng thu thập ý kiến sở thích về 4 màu sắc xe cho kết quả như sau:

Màu xe	Trắng (1)	Đen (2)	Vàng (3)	Xanh (4)	Cộng
Số khách chọn	18	60	12	30	120

Với mức ý nghĩa 5%, hãy cho kết luận về việc này.
