Chương 5

Kiểm định giả tuyết thống kê

5.1 Khái niệm

Định nghĩa 5.1. .

Giả thuyết thống kê là một phát biểu về những tham số của một hay nhiều tổng thể(μ, p, σ^2), hoặc về tính độc lập của các đặc điểm trong tổng thể.

Kiểm định giả thuyết thống kê là việc dùng các phương pháp thống kê, dựa trên dữ liệu thu được từ mẫu điều tra cho kết luận về việc chấp nhận hay bác bỏ một giả thuyết thống kê nào đó.

Gọi

- H: giả thuyết (giả thuyết không),

- $\overline{H}$: giả thuyết đối (đối thuyết)

Giả thuyết H là một phát biểu sẽ được kiểm tra thông qua một phép kiểm định có ý nghĩa, nghĩa là thông qua đó ta có bằng chứng đủ mạnh để bác bỏ H.

Ví dụ 5.1. *Một nhà sản xuất cho rằng trọng lượng trung bình của 1 gói mì là 75g.*

Để kiểm tra điều này đúng hay sai, ta thực hiện các bước sau

- Gọi μ trọng lượng trunh bình của gói mì.

- Đặt giả thuyết

$$\begin{cases} H : \mu = 75 \\ \overline{H} : \mu \neq 75 \end{cases}$$

- Lấy mẫu một số gói mì.

- Kiểm định, tính toán.

- Kết luận.

Các quyết định dựa vào mẫu quan sát nên có thể dẫn đến sai lầm. Có hai loại sai lầm

Định nghĩa 5.2. .

i. Sai lầm loại I: là sai lầm nếu **bác bỏ** giả thuyết H khi nó **đúng**.

Xác suất của sai lầm loại I:

$$P\{\text{Sai lầm loại I}\} = P\{\textbf{bác bỏ } H|H \text{ đúng}\} = \alpha.$$

Trong đó: α: *mức ý nghĩa*.

ii. Sai lầm loại II: là sai lầm nếu **chấp nhận** giả thuyết H khi nó **sai**.

Xác suất của sai lầm loại II:

$$P\{\text{Sai lầm loại II}\} = P\left\{\textbf{chấp nhận } H|H\text{sai}\right\} = \beta$$

Ta có:

	Bác bỏ	Chấp nhận
H đúng	Sai lầm loại I $P\{\textbf{bác bỏ } H\|H \text{ đúng}\} = \alpha$	$P\{\text{chấp nhận } H\|H \text{ đúng}\} = 1-\alpha$
H sai	$P\{\text{bác bỏ } H\|H \text{ sai}\} = 1-\beta$	Sai lầm loại II $P\left\{\textbf{chấp nhận } H\|H\text{sai}\right\} = \beta$

Sai lầm loại I và loại II luôn tồn tại trong bài toán kiểm định, ta không thể đồng thời làm giảm 2 sai lầm này. Khi làm giảm sai lầm loại I ta sẽ làm tăng sai lầm loại II và ngược lại.

Ta sẽ ấn định $P\,(\text{sai lầm loại I}) = \alpha$, khi đó xác suất sai lầm loại II sẽ nhỏ nhất.

Ví dụ 5.2. *Một nhà kinh doanh sau khi áp dụng các biên pháp khuyến mãi muốn biết lợi nhuận có tăng hay không.*

Giải. Giả thuyết

H: lợi nhuận tăng. (LN tăng); $\overline{H}$: lợi nhuận không tăng.

Xét

Thực tế	Bác bỏ	Chấp nhận
lợi nhuận tăng	Sai lầm loại I $P\{\textbf{bác bỏ } H\| \text{ LN tăng}\} = \alpha$	kết luận đúng $P\{\text{chấp nhận } H\| \text{ LN tăng}\} = 1-\alpha$
lợi nhuận giảm	kết luận đúng $P\{\text{bác bỏ } H\| \text{ LN giảm}\} = 1-\beta$	Sai lầm loại II $P\left\{\textbf{chấp nhận } H\|\text{LN giảm}\right\} = \beta$

Ta thấy, *"lợi nhuận không tăng mà bảo tăng"* là sai lầm nghiêm trọng hơn *"lợi nhuận có tăng mà bảo không"*. Do đó, nên đặt giả thuyết: H: lợi nhuận không tăng.

Chú ý: nên đặt sai lầm nghiêm trọng hơn là sai lầm loại I.

5.2 Kiểm định giả thuyết về tham số

Gọi

- θ là tham số cần kiểm định $(\theta = \mu, p, \sigma^2)$.
- θ_0 là giá trị cho trước.
- Ta có

Kiểm định 2 phía	**Kiểm định một phía**	
	phía trái	phía phải
$\begin{cases} H : \theta = \theta_0 \\ \overline{H} : \theta \neq \theta_0 \end{cases}$	$\begin{cases} H : \theta = \theta_0 \\ \overline{H} : \theta < \theta_0 \end{cases}$	$\begin{cases} H : \theta = \theta_0 \\ \overline{H} : \theta > \theta_0 \end{cases}$

$\triangleright$ **Các bước kiểm định**

Bước 0: Xác định tham số mình quan tâm trong một bài toán cụ thể.

Bước 1: Phát biểu giả thuyết không H về tham số quan tâm.

Bước 2: Chỉ định đối thuyết $\overline{H}$ sẽ thay thế H

Bước 3: Xác định mức ý nghĩa.

Bước 4: Thu thập dữ liệu mẫu và tính giá trị thống kê kiểm định $(\overline{x}, s, f)$.

Bước 5: Kiểm định giả thuyết thống kê.

Bước 6: Kết luận bác bỏ hay chấp nhận H.

Các phần sau chỉ trình bày bước 5, và 6.

5.3 So sánh trung bình với một số

Giả thuyết: cho

- μ: trung bình của tổng thể.
- μ_0: giá trị cho trước
- α: mức ý nghĩa.

Cách làm:

Bước 5:

	Kiểm định 2 phía $\begin{cases} H: \mu = \mu_0 \\ \overline{H}: \mu \neq \mu_0 \end{cases}$	Kiểm định trái $\begin{cases} H: \mu = \mu_0 \\ \overline{H}: \mu < \mu_0 \end{cases}$	kiểm định phải $\begin{cases} H: \mu = \mu_0 \\ \overline{H}: \mu > \mu_0 \end{cases}$
$\begin{cases} \left[\begin{array}{l} \sigma \text{ biết} \\ \sigma \text{ chưa biết} \end{array}\right. \\ n \geq 30. \end{cases}$	$\varphi(C) = \dfrac{\gamma}{2} \Rightarrow C$	$\varphi(C) = 0.5 - \alpha \Rightarrow C$	$\varphi(C) = 0.5 - \alpha \Rightarrow C$
$\begin{cases} \sigma \text{ chưa biết} \\ n < 30. \end{cases}$	$C = t\left(n-1, \dfrac{\alpha}{2}\right)$	$C = t(n-1, \alpha)$	$C = t(n-1, \alpha)$

Xác định giá trị thống kê

$$t = \frac{(\overline{x} - \mu_0)\sqrt{n}}{\square}$$

Trong đó

- $\square = \sigma$ nếu σ biết. - $\square = s$ nếu σ chưa biết.

Bước 6: Kết luận

	Kiểm định 2 phía $\begin{cases} H: \mu = \mu_0 \\ \overline{H}: \mu \neq \mu_0 \end{cases}$	Kiểm định trái $\begin{cases} H: \mu = \mu_0 \\ \overline{H}: \mu < \mu_0 \end{cases}$	kiểm định phải $\begin{cases} H: \mu = \mu_0 \\ \overline{H}: \mu > \mu_0 \end{cases}$
Chấp nhận H nếu	$\lvert t \rvert \leq C$	$-t \leq C$	$t \leq C$
Chấp nhận $\overline{H}$ nếu	$\lvert t \rvert > C$	$-t > C$	$t > C$

Ví dụ 5.3. *Điểm trung bình môn XSTK năm học trước là 5,75. năm nay theo dõi 100 sinh viên được số liệu*

Điểm	*3*	*4*	*5*	*6*	*7*	*8*	*9*
Số SV	*3*	*5*	*27*	*43*	*12*	*6*	*4*

Với mứ ý nghĩa 1%, phải chăng điểm trung bình môn XSTK năm nay cao hơn năm trước?

Giải.

- Gọi μ là điểm trung bình môn toán của năm nay.

- Giả thuyết

$$\begin{cases} H: \mu = 5,72 \\ \overline{H}: \mu > 5,72 \end{cases}$$

- Ta có $\alpha = 1\% = 0,01$, $\quad n = 100$, $\quad \overline{x} = 5,9$ và $s = 1,21$.

$\Rightarrow \sigma$ chưa biết, $n \geq 30$

- $\varphi(C) = 0.5 - \alpha = 0,5 - 0,01 = 0,49$

$\Rightarrow C = 2,33$

- $t = \dfrac{(5,9 - 5,72)\sqrt{100}}{1,21} \approx 1,4876.$

- Vì $t = 1,4876 < C = 2,33$, nên ta chấp nhận H.

- Vậy điểm trung bình môn XSTK năm nay không cao hơn năm trước với mức ý nghĩa 1%.

Ví dụ 5.4. *Trong năm 2001, Bộ Lao động Mỹ báo cáo rằng thu nhập trung bình giờ của công nhân sản xuất ở Mỹ là 14,32 USD/giờ. Một mẫu khảo sát 75 công nhân năm 2003 cho thấy trung bình mẫu là 14,68 USD/giờ. Giả sử độ lệch chuẩn tổng thể là $\sigma = 1,45$ USD, có thể kết luận rằng thu nhập trung bình đã tăng lên so với năm 2001 được không? Biết $\alpha = 0,05$.*

Giải

Ví dụ 5.5. *Một nhà nghiên cứu nhân chủng học muốn tìm hiểu xem chiều cao trung bình của thanh niên có thay đổi không so với mức 1,70m cách đây 10 năm. Một mẫu gồm 121 thanh niên được chọn và từ mẫu đó tính được $\overline{x} = 1,72m$ với độ lệch chuẩn $s = 0,02m$. Với mức ý nghĩa $\alpha = 0,05$ ta có thể kết luận gì về sự thay đổi chiều cao trung bình của thanh niên.*

Giải

5.4 So sánh tỉ lệ với một số

Giả thuyết: cho

- p: tỉ lệ của tổng thể.
- p_0: giá trị cho trước
- f: tỉ lệ mẫu.
- α: mức ý nghĩa.

Cách làm:

Bước 5:

Kiểm định 2 phía $\begin{cases} H: p = p_0 \\ \overline{H}: p \neq p_0 \end{cases}$	Kiểm định trái $\begin{cases} H: p = p_0 \\ \overline{H}: p < p_0 \end{cases}$	kiểm định phải $\begin{cases} H: p = p_0 \\ \overline{H}: p > p_0 \end{cases}$
$\varphi(C) = \frac{\gamma}{2} \Rightarrow C$	$\varphi(C) = 0.5 - \alpha \Rightarrow C$	$\varphi(C) = 0.5 - \alpha \Rightarrow C$

Xác định giá trị thống kê

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

Bước 6: Kết luận

	Kiểm định 2 phía $\begin{cases} H: p = p_0 \\ \overline{H}: p \neq p_0 \end{cases}$	Kiểm định trái $\begin{cases} H: p = p_0 \\ \overline{H}: p < p_0 \end{cases}$	kiểm định phải $\begin{cases} H: p = p_0 \\ \overline{H}: p > p_0 \end{cases}$
Chấp nhận H nếu	$\lvert t\rvert \leq C$	$-t \leq C$	$t \leq C$
Chấp nhận $\overline{H}$ nếu	$\lvert t\rvert > C$	$-t > C$	$t > C$

Ví dụ 5.6. *Theo thống kê của bộ y tế, 12% dân cư tỉnh A mắc bệnh đau mắt hột. Khi kiểm tra 200 người tại tỉnh A đã phát hiện 21 người đau mắt hột. Số liệu này có khẳng định kết luận mà bộ y tế đưa ra không, với mức ý nghĩa 1%.*

Giải

- Gọi p là tỉ lệ người dân tỉnh A mắc bệnh đau mắt hột.
- Giả thuyết

$$\begin{cases} H: p = 0,12 \\ \overline{H}: p \neq 0,12 \end{cases}$$

- Ta có $\alpha = 1\% = 0,01 \Rightarrow \gamma = 0,99, \qquad n = 200, \qquad f = \frac{21}{200} = 0,105.$

- $\varphi(C) = \dfrac{\gamma}{2} = \dfrac{0,99}{2} = 0,495$

$\Rightarrow C = 2,58$

- $t = (0,105 - 0,12)\sqrt{\dfrac{200}{0,12(1-0,12)}} \approx -0,6528.$

- Vì $|t| = 0,6528 < C = 2,58$, nên ta chấp nhận H.

Vậy, báo cáo của bộ y tế được chấp nhận là đúng với mức ý nghĩa 1%.

Ví dụ 5.7. *Tỉ lệ học sinh tốt nghiệp THPT năm ngoái của tỉnh A là* 88%*. Trong kì thi năm nay, trong 100 em được chọn ngẫu nhiên có 82 em thi đỗ. Với mức ý nghĩa* 5%*, có thể kết luận rằng tỉ lệ học sinh thi đỗ năm nay thấp hơn năm ngoái hay không?*

Giải

5.5 So sánh hai trung bình

Giả thuyết:

- Tổng thể I: $\begin{cases} \mu_1, \sigma_1, \\ \overline{x_1}, s_1, \text{cỡ mẫu } m. \end{cases}$
- Tổng thể II: $\begin{cases} \mu_2, \sigma_2, \\ \overline{x_2}, s_2, \text{cỡ mẫu } n. \end{cases}$

Cách làm:

Bước 5:

	Kiểm định 2 phía $\begin{cases} H: \mu_1 = \mu_2 \\ \overline{H}: \mu_1 \neq \mu_2 \end{cases}$	Kiểm định trái $\begin{cases} H: \mu_1 = \mu_2 \\ \overline{H}: \mu_1 < \mu_2 \end{cases}$	kiểm định phải $\begin{cases} H: \mu_1 = \mu_2 \\ \overline{H}: \mu_1 > \mu_2 \end{cases}$
$\left[\begin{array}{l} \sigma_1, \sigma_2 \text{ biết} \\ \sigma_1, \sigma_2 \text{ chưa biết} \end{array}\right.$ và $m, n \geq 30$	$\varphi(C) = \dfrac{\gamma}{2}$ $\Rightarrow C$	$\varphi(C) = 0.5 - \alpha$ $\Rightarrow C$	$\varphi(C) = 0.5 - \alpha$ $\Rightarrow C$
σ_1, σ_2 chưa biết và $m, n \leq 30$	C = $t\left(m+n-2, \dfrac{\alpha}{2}\right)$	C = $t(m+n-2, \alpha)$	C = $t(m+n-2, \alpha)$

Xác định giá trị thống kê

$$t = \frac{\overline{x_1} - \overline{x_2}}{\sqrt{\dfrac{\square_1}{m} + \dfrac{\square_2}{n}}}$$

Trong đó

- $\square_1 = \sigma_1^2, \square_2 = \sigma_2^2$, nếu σ_1, σ_2 biết.
- $\square_1 = s_1^2, \square_2 = s_2^2$, nếu σ_1, σ_2 chưa biết và $m, n \geq 30$.
- $\square_1 = \square_2 = s^2 = \dfrac{(m-1)s_1^2 + (n-1)s_2^2}{m+n-2}$, nếu σ_1, σ_2 chưa biết và $m, n < 30$.

Bước 6: Kết luận

	Kiểm định 2 phía $\begin{cases} H: \mu_1 = \mu_2 \\ \overline{H}: \mu_1 \neq \mu_2 \end{cases}$	Kiểm định trái $\begin{cases} H: \mu_1 = \mu_2 \\ \overline{H}: \mu_1 < \mu_2 \end{cases}$	kiểm định phải $\begin{cases} H: \mu_1 = \mu_2 \\ \overline{H}: \mu_1 > \mu_2 \end{cases}$
Chấp nhận H nếu	$\lvert t\rvert \leq C$	$-t \leq C$	$t \leq C$
Chấp nhận $\overline{H}$ nếu	$\lvert t\rvert > C$	$-t > C$	$t > C$

Ví dụ 5.8. *Gọi X, Y là khối lượng trẻ sơ sinh trai và gái. Quan sát ngẫu nhiên 22 bé trai, ta được* $\overline{x} = 3200g; s_x = 400g$ *và quan sát 20 bé gái, ta được* $\overline{y} = 3000g, s_y = 380g$.

Với mức ý nghĩa 5%, *hãy cho biết khối lượng của trẻ sơ sinh trai và gái có giống nhau không?*

Giải. Ta có:

- Trẻ sơ sinh nam: $\overline{x} = 3200g; s_x = 400g, m = 22$

- Trẻ sơ sinh gái: $\overline{y} = 3000g, s_y = 380g, n = 20$

$\Rightarrow \sigma_x, \sigma_y$ chưa biết và $m, n < 30$.

Gọi

- μ_x khối lượng trung bình của trẻ sơ sinh trai.

- μ_y khối lượng trung bình của trẻ sơ sinh gái.

Giả thuyết

$$\begin{cases} H : \mu_1 = \mu_2 \\ \overline{H} : \mu_1 \neq \mu_2 \end{cases}$$

Ta có

- $C = t\left(m+n-2, \dfrac{\alpha}{2}\right) = t\left(22+20-2, \dfrac{0,05}{2}\right) = t\,(40; 0,025) = 2,021.$

- $s^2 = \dfrac{(m-1)s_1^2 + (n-1)s_2^2}{m+n-2} = \dfrac{21.400^2 + 19.380^2}{40} = 152590.$

- $t = \dfrac{3200 - 3000}{\sqrt{\dfrac{152590}{22} + \dfrac{152590}{20}}} \approx 1,6572.$

$\Rightarrow |t| = 1,6572 \leq C = 2,021. \Rightarrow$ chấp nhận H.

Vậy, khối lượng trẻ sơ sinh trai và gái là như nhau với mức ý nghĩa 5%.

Ví dụ 5.9. *Điểm trung bình học tập của 50 học sinh lớp 5 trường TH Quận 5 là* $6,72$, *phương sai mẫu là* $(0,72)^2$. *Điểm trung bình học tập của 80 học sinh lớp 5 trường TH quận 1 là* $6,46$, *phương sai mẫu là* $(0,91)^2$.

1. Với mức ý nghĩa 5%, *hãy cho biết điểm trung bình học tập của học sinh trường TH quận 5 và trường TH quận 1 có khác nhau không.*

2. Với mức ý nghĩa 5%, *hãy cho biết điểm trung bình học tập của học sinh trường TH quận 5 có lớn hơn trường TH quận 1 không.*

Giải

5.6 So sánh hai tỉ lệ

Giả thuyết:

- Tổng thể I: $\begin{cases} p_1 \text{ tỉ lệ tổng thể,} \\ \text{tỉ lệ mẫu} f_1, \text{cỡ mẫu } m. \end{cases}$

- Tổng thể II: $\begin{cases} p_2 \text{tỉ lệ tổng thể,} \\ \text{tỉ lệ mẫu} f_2, \text{cỡ mẫu } n. \end{cases}$

Cách làm:

Bước 5:

Kiểm định 2 phía $\begin{cases} H : p_1 = p_2 \\ \overline{H} : p_1 \neq p_2 \end{cases}$	Kiểm định trái $\begin{cases} H : p_1 = p_2 \\ \overline{H} : p_1 < p_2 \end{cases}$	kiểm định phải $\begin{cases} H : p_1 = p_2 \\ \overline{H} : p_1 > p_2 \end{cases}$
$\varphi(C) = \dfrac{\gamma}{2} \Rightarrow C$	$\varphi(C) = 0.5 - \alpha \Rightarrow C$	$\varphi(C) = 0.5 - \alpha \Rightarrow C$

Xác định giá trị thống kê

$$f = \frac{m.f_1 + n.f_2}{m+n}.$$

$$\Rightarrow t = \frac{f_1 - f_2}{\sqrt{f.(1-f).\left(\dfrac{1}{m} + \dfrac{1}{n}\right)}}.$$

Bước 6: Kết luận

	Kiểm định 2 phía $\begin{cases} H : p_1 = p_2 \\ \overline{H} : p_1 \neq p_2 \end{cases}$	Kiểm định trái $\begin{cases} H : p_1 = p_2 \\ \overline{H} : p_1 < p_2 \end{cases}$	kiểm định phải $\begin{cases} H : p_1 = p_2 \\ \overline{H} : p_1 > p_2 \end{cases}$
Chấp nhận H nếu	$\lvert t \rvert \leq C$	$-t \leq C$	$t \leq C$
Chấp nhận $\overline{H}$ nếu	$\lvert t \rvert > C$	$-t > C$	$t > C$

Ví dụ 5.10. *Kiểm tra 100 sản phẩm ở kho I, thấy có sáu phế phẩm. Kiểm tra 200 sản phẩm ở kho II, thấy có 24 phế phẩm. Với mức ý nghĩa* 5%*, chất lượng hàng ở hai kho có khác nhau hay không?*

Giải. Ta có

- Kho I: m = 100; $f_1 = \dfrac{6}{100}$

- Kho II: n = 200; $f_2 = \dfrac{24}{200}$

- $\alpha = 0,05 \Rightarrow \gamma = 0,95$

Gọi

- p_1 tỉ lệ phế phẩm ở kho I.

- p_2 tỉ lệ phế phẩm ỏ kho II.

Giả thuyết

$$\begin{cases} H : p_1 = p_2 \\ \overline{H} : p_1 \neq p_2 \end{cases}$$

Ta có

- $\varphi(C) = \dfrac{0,95}{2} = 0,475 \Rightarrow C = 1,96.$

- $f = \dfrac{100.0,06 + 200.0,12}{100 + 200} = 0,1$

- $t = \dfrac{0,06 - 0,12}{\sqrt{0,1.(1-0,1).\left(\dfrac{1}{100} + \dfrac{1}{200}\right)}} \approx -1,633$

$\Rightarrow |t| = 1,633 < C = 1,96 \Rightarrow$ chấp nhận H.

Vậy, Chất lượng hàng hóa tại hai kho là như nhau với mức ý nghĩa 5%.

Ví dụ 5.11. *Trong năm 2018, tại xí nghiệp A có 200 công nhân thì có 30 công nhân xin nghỉ việc, tại xí nghiệp B có 350 công nhân thì có 65 công nhân xin nghỉ việc. Với múc ý nghĩa 5% có thể cho rằng tỉ lệ công nhân thôi việc tại xí nghiệp A thấp hơn xí nghiệp B hay không?*

Giải