HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG

o0o

Chương 5: Ước lượng tham số và kiểm định giả thiết thống kê

XÁC SUẤT THỐNG KÊ

Bộ môn Toán, Khoa Cơ bản 1

Hà Nội - 2023

Chương 5: Ước lượng tham số và kiểm định giả thiết thống kê

Chương 5: Ước lượng tham số và kiểm định giả thiết thống kê

Bài toán ước lượng tham số

Cho biến ngẫu nhiên X có quy luật phân bố xác suất đã biết nhưng chưa biết tham số θ nào đó của nó. Phải ước lượng (xác định một cách gần đúng) giá trị θ.

Khái niệm ước lượng điểm

- Giả sử cần ước lượng tham số θ của Bnn X. Từ tổng thể lập mẫu ngẫu nhiên kích thước n: $(X_1, X_2, \ldots, X_n)$. Nếu lấy thống kê $T = T(X_1, X_2, \ldots, X_n)$ để thay cho tham số θ thì T được gọi là một ước lượng điểm của θ.
- Với mẫu cụ thể $(x_1, x_2, \ldots, x_n)$, $\hat{\theta} = T(x_1, x_2, \ldots, x_n)$ cũng được gọi là một ước lượng điểm của θ.

Ước lượng không chệch

Thống kê $T = T(X_1, X_2, \ldots, X_n)$ được gọi là một ước lượng không chệch của tham số θ nếu $E(T) = \theta$. Ngược lại, nếu $E(T) \neq \theta$ thì T được gọi là một ước lượng chệch của θ.

Ví dụ 1. Giả sử biến ngẫu nhiên gốc X có

$$E(X) = \mu, D(X) = \sigma^2$$

$(X_1, X_2, \ldots, X_n)$ là một mẫu ngẫu nhiên. Khi đó

a) $\overline{X} = \dfrac{1}{n}\sum_{i=1}^{n} X_i$ là một ước lượng không chệch của μ.

b) $S^2 = \dfrac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$ là một ước lượng không chệch của σ^2.

Ước lượng vững

Thống kê $T = T(X_1, X_2, \dots, X_n)$ được gọi là một ước lượng vững của tham số θ nếu: với mọi $\varepsilon > 0$ cho trước, ta có

$$\lim_{n\to\infty} P(|T - \theta| < \varepsilon) = 1.$$

Ước lượng hiệu quả

Thống kê $T = T(X_1, X_2, \dots, X_n)$ được gọi là một ước lượng hiệu quả của tham số θ nếu nó là ước lượng không chệch có phương sai nhỏ nhất.

Chú ý

- Ước lượng điểm tốt nhất của kỳ vọng là trung bình mẫu.
- Ước lượng điểm tốt nhất của phương sai là phương sai mẫu.
- Ước lượng điểm tốt nhất của tỷ lệ cá thể có dấu hiệu A trong tổng thể là tần suất mẫu f, trong đó

$$f = \frac{\text{số phần tử có dấu hiệu A trong mẫu}}{\text{kích thước mẫu}}$$

Chương 5: Ước lượng tham số và kiểm định giả thiết thống kê

5.2.1 Các định nghĩa

Các định nghĩa

- Giả sử $L = L(X_1, X_2, \dots, X_n)$, $U = U(X_1, X_2, \dots, X_n)$ là hai thống kê có từ mẫu ngẫu nhiên $(X_1, X_2, \dots, X_n)$, θ là một tham số có mặt trong phân bố xác suất của tổng thể và $\alpha \in (0, 1)$.
- Khoảng $[L; U]$ được gọi là khoảng tin cậy của θ với độ tin cậy $1 - \alpha$ nếu

$$P(L \leq \theta \leq U) = 1 - \alpha$$

- $I = U - L$ được gọi là độ rộng của khoảng tin cậy.
- Với mẫu cụ thể $(x_1, x_2, \dots, x_n)$, ta tính được $L = \ell$, $U = u$. Khi đó khoảng tin cậy của θ với độ tin cậy $1 - \alpha$ là $[\ell; u]$

5.2.2 Khoảng tin cậy của kỳ vọng của biến ngẫu nhiên phân bố chuẩn

Bài toán

Giả sử biến ngẫu nhiên gốc X có phân bố chuẩn $N(\mu, \sigma^2)$ nhưng chưa biết μ. Với độ tin cậy $1 - \alpha$ cho trước, hãy tìm khoảng tin cậy của μ

Từ tổng thể ta lập mẫu ngẫu nhiên kích thước n:

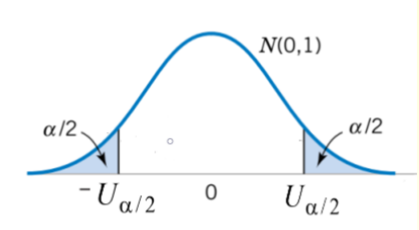
$$(X_1, X_2, \ldots, X_n)$$

Xét bài toán trong 3 trường hợp:

- Trường hợp 1: σ^2 đã biết
- Trường hợp 2: σ^2 chưa biết, kích thước mẫu lớn $(n \geq 30)$
- Trường hợp 3: σ^2 chưa biết, kích thước mẫu nhỏ $(n < 30)$

Trường hợp 1: σ^2 đã biết

- Xét thống kê $T = \dfrac{\overline{X} - \mu}{\sigma}\sqrt{n}$
- $T \sim N(0,1)$.



$$P\left(-U_{\alpha/2} \le \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \le U_{\alpha/2}\right) = 1 - \alpha$$

$$\Leftrightarrow P\left(\overline{X} - U_{\alpha/2}\frac{\sigma}{\sqrt{n}} \le \mu \le \overline{X} + U_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$$

Trường hợp 1: Phương sai σ^2 đã biết

Khoảng tin cậy của μ với độ tin cậy $1-\alpha$ là

$$\left[\overline{x} - U_{\alpha/2}\frac{\sigma}{\sqrt{n}}; \overline{x} + U_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right]$$

Ví dụ 2. Doanh số của một cửa hàng là biến ngẫu nhiên phân bố chuẩn với độ lệch chuẩn là 2 triệu/ tháng. Điều tra ngẫu nhiên doanh số của 600 cửa hàng có quy mô tương tự nhau tìm được doanh số trung bình là 18,5 triệu. Với độ tin cậy 95%, hãy ước lượng doanh số trung bình của các cửa hàng thuộc quy mô đó.

Định nghĩa

$\varepsilon = U_{\alpha/2}\dfrac{\sigma}{\sqrt{n}}$ được gọi là độ chính xác của ước lượng.

Chọn kích thước mẫu

Nếu muốn ước lượng có độ chính xác không vượt quá ε_0 với độ tin cậy $1-\alpha$ cho trước, cần chọn kích thước mẫu n là số tự nhiên nhỏ nhất thỏa mãn:

$$n \geq \frac{U^2_{\alpha/2}\sigma^2}{\varepsilon_0^2}$$

Ví dụ 3. Trong Ví dụ 2, nếu muốn độ chính xác của ước lượng không vượt quá 0,2 triệu thì cần điều tra doanh số của ít nhất bao nhiêu cửa hàng?

Trường hợp 2: σ^2 chưa biết, $n \geq 30$

- Xét thống kê $T = \dfrac{\overline{X} - \mu}{S}\sqrt{n}$
- T có phân bố xấp xỉ $N(0,1)$.

Trường hợp 2: Phương sai σ^2 chưa biết, $n \geq 30$

Khoảng tin cậy của μ với độ tin cậy $1-\alpha$ là

$$\left[\overline{x} - U_{\alpha/2}\frac{s}{\sqrt{n}}; \overline{x} + U_{\alpha/2}\frac{s}{\sqrt{n}}\right]$$

Chú ý: Trong trường hợp kích thước mẫu lớn, ta không cần giả thiết X có phân bố chuẩn.

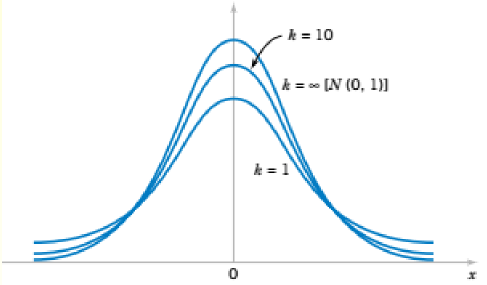
Ví dụ 4. Theo dõi ngẫu nhiên thời gian giao hàng (đơn vị: giờ) tới 60 địa chỉ trong nội thành của một dịch vụ chuyển phát nhanh, thu được kết quả sau:

Thời gian giao hàng	6-7	7-8	8-9	9-10	10-11	11-12	12-13
Số địa chỉ	2	4	10	16	13	10	5

Với độ tin cậy 0,95, hãy tìm khoảng tin cậy của thời gian giao hàng trung bình trong nội thành của dịch vụ chuyển phát nhanh nói trên.

Phân bố Student

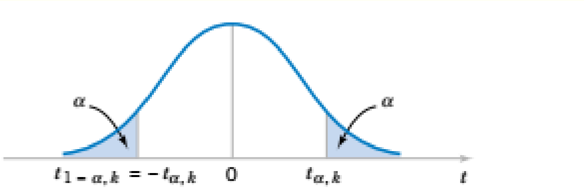
Cho $(X_1, X_2, \ldots, X_n)$ là một mẫu ngẫu nhiên từ một tổng thể có biến ngẫu nhiên gốc X. Nếu $X \sim N(\mu, \sigma^2)$ thì $\dfrac{\overline{X} - \mu}{S}\sqrt{n}$ có phân bố Student với $n - 1$ bậc tự do.



Hình 2.1: Đồ thị hàm mật độ của phân bố Student với các bậc tự do $k = 1, k = 10, k = \infty$

Chú ý:

- Khi số bậc tự do k tăng lên, phân bố Student sẽ hội tụ rất nhanh về phân bố chuẩn tắc. Do đó với k khá lớn ($k > 30$) có thể dùng phân bố chuẩn tắc thay cho phân bố Student.
- Nếu biến ngẫu nhiên T có phân bố Student với k bậc tự do thì giá trị tới hạn mức α của T, ký hiệu $t_{\alpha;k}$ hoặc $t_\alpha(k)$, là giá trị thỏa mãn $P(T > t_{\alpha;k}) = \alpha$.
- Giá trị của $t_{\alpha;k}$ được cho ở bảng Phụ lục III, nó là số nằm ở hàng k, cột α.



Trường hợp 3: σ^2 chưa biết, $n < 30$

- Xét thống kê $T = \dfrac{\overline{X} - \mu}{S}\sqrt{n}$
- T có phân bố Student với $n - 1$ bậc tự do.

Trường hợp 3: Phương sai σ^2 chưa biết, $n < 30$

Khoảng tin cậy của μ với độ tin cậy $1 - \alpha$ là

$$\left[\overline{x} - t_{\alpha/2}(n-1)\frac{s}{\sqrt{n}}; \overline{x} + t_{\alpha/2}(n-1)\frac{s}{\sqrt{n}}\right]$$

Ví dụ 5. Để ước lượng tuổi thọ trung bình của một loại sản phẩm, người ta chọn ngẫu nhiên 25 sản phẩm và thu được kết quả sau:

Tuổi thọ (giờ)	190	195	198	200	204	205
Số sản phẩm	4	4	2	8	6	1

Giả sử tuổi thọ sản phẩm là một biến ngẫu nhiên phân bố chuẩn, hãy ước lượng tuổi thọ trung bình của sản phẩm trên với độ tin cậy 95%.

Bài toán

Giả sử tỷ lệ cá thể có dấu hiệu A trong một tổng thể là p. Với độ tin cậy $1-\alpha$, hãy tìm khoảng tin cậy của p.

Chọn ngẫu nhiên n cá thể từ tổng thể. Gọi f là tỷ lệ cá thể có dấu hiệu A trong mẫu quan sát.

$$f = \frac{\text{số cá thể có dấu hiệu A trong mẫu}}{n}$$

Khoảng tin cậy của p với độ tin cậy $1-\alpha$ là:

$$\left[f - U_{\frac{\alpha}{2}}\sqrt{\frac{f(1-f)}{n}}; f + U_{\frac{\alpha}{2}}\sqrt{\frac{f(1-f)}{n}}\right]$$

với điều kiện n đủ lớn thoả mãn $nf \geq 10$ và $n(1-f) \geq 10$.

Ví dụ 6. Người ta lấy ngẫu nhiên từ một lô hàng ra 200 sản phẩm thì thấy có 182 sản phẩm đạt yêu cầu chất lượng. Với độ tin cậy 95%, hãy ước lượng tỷ lệ sản phẩm đạt yêu cầu chất lượng của lô hàng.

Bài toán ước lượng số cá thể có dấu hiệu A trong tổng thể

Giả sử một tổng thể có kích thước N (đã biết), trong đó có M (không biết) cá thể có dấu hiệu A. Với độ tin cậy $1-\alpha$, hãy tìm khoảng tin cậy của M.

Chọn ngẫu nhiên n cá thể từ tổng thể. Gọi f là tỷ lệ cá thể có dấu hiệu A trong mẫu quan sát.

Tỷ lệ cá thể có dấu hiệu A trong tổng thể là $p = \dfrac{M}{N}$, do đó

Khoảng tin cậy của M với độ tin cậy $1-\alpha$ là:

$$[Np_1; Np_2],$$

trong đó $[p_1; p_2]$ là khoảng tin cậy của p với độ tin cậy $1-\alpha$.

Ví dụ 7. Một vùng có 2000 hộ gia đình. Để điều tra nhu cầu tiêu dùng một loại hàng hóa tại vùng đó, người ta nghiên cứu ngẫu nhiên 100 gia đình và thấy có 60 gia đình có nhu cầu về loại hàng hóa trên. Với độ tin cậy 0,95 hãy ước lượng số gia đình trong vùng có nhu cầu về loại hàng hóa đó.

Bài toán ước lượng kích thước tổng thể

Giả sử một tổng thể có kích thước N (không biết). Với độ tin cậy $1-\alpha$, hãy tìm khoảng tin cậy của N.

Chọn ngẫu nhiên M cá thể từ tổng thể, đánh dấu các cá thể này, sau đó trả lại tổng thể. Sau một thời gian chọn ngẫu nhiên n cá thể. Gọi f là tỷ lệ cá thể bị đánh dấu trong mẫu quan sát.

Tỷ lệ cá thể bị đánh dấu trong tổng thể là $p = \dfrac{M}{N}$, do đó

Khoảng tin cậy của N với độ tin cậy $1-\alpha$ là:

$$\left[\frac{M}{p_2}; \frac{M}{p_1}\right],$$

trong đó $[p_1; p_2]$ là khoảng tin cậy của p với độ tin cậy $1-\alpha$.

Ví dụ 8. Tại một vùng rừng nguyên sinh, người ta đeo vòng cho 1000 con chim. Sau một thời gian bắt lại 200 con thì thấy có 40 con có đeo vòng. Ước lượng số chim trong vùng rừng đó với độ tin cậy 99%.

Chương 5: Ước lượng tham số và kiểm định giả thiết thống kê

5.3.1 Giả thiết thống kê

- Giả thiết thống kê là giả thiết về dạng phân bố xác suất của Bnn, về các tham số đặc trưng của Bnn hoặc về tính độc lập của các Bnn.
- Giả thiết thống kê đưa ra được ký hiệu là H_0 và được gọi là giả thiết không.
- Khi đưa ra một giả thiết thống kê, người ta còn nghiên cứu một mệnh đề mâu thuẫn với nó, gọi là đối thiết, ký hiệu H_1. Đối thiết H_1 sẽ được chấp nhận khi H_0 bị bác bỏ.

Ví dụ 9. Giả sử thời gian chờ trung bình để đặt hàng ở một nhà hàng thức ăn nhanh là 4,5 phút. Người quản lý nhà hàng muốn xác định xem thời gian chờ trung bình có thay đổi trong tháng vừa qua không. Ta đưa ra cặp giả thiết, đối thiết:

$$\begin{aligned} H_0: &\quad \mu = 4,5 \\ H_1: &\quad \mu \neq 4,5 \longleftarrow \text{đối thiết hai phía} \end{aligned}$$

Nếu nhà quản lý muốn xác định xem thời gian chờ trung bình trong tháng vừa qua có lớn hơn 4,5 phút không, ta có cặp giả thiết, đối thiết

$$\begin{aligned} H_0 &: \quad \mu = 4,5 \\ H_1 &: \quad \mu > 4,5 \longleftarrow \text{đối thiết một phía} \end{aligned}$$

Sau một thời gian cải tiến chất lượng dịch vụ, nhà quản lý muốn xác định xem thời gian chờ trung bình để đặt hàng có giảm không. Khi đó, cặp giả thiết, đối thiết được đưa ra là

$$\begin{aligned} H_0 &: \quad \mu = 4,5 \\ H_1 &: \quad \mu < 4,5 \longleftarrow \text{đối thiết một phía} \end{aligned}$$

- Kiểm định một giả thiết thống kê là dựa trên cơ sở mẫu đã có, ta xây dựng một quy tắc chấp nhận hoặc bác bỏ H_0.
- Quy tắc kiểm định dựa trên nguyên lý xác suất nhỏ: "Nếu một biến cố có xác suất rất nhỏ thì trong một phép thử biến cố đó coi như không xảy ra".

5.3.2 Tiêu chuẩn kiểm định giả thiết thống kê

- Từ biến ngẫu nhiên gốc X của tổng thể lập mẫu ngẫu nhiên $(X_1, X_2, \ldots, X_n)$.
- Chọn thống kê $T = T(X_1, X_2, \ldots, X_n)$ sao cho nếu H_0 đúng thì quy luật phân bố xác suất của T hoàn toàn xác định. Thống kê T được gọi là tiêu chuẩn kiểm định.
- Với mẫu cụ thể $(x_1, x_2, \ldots, x_n)$, $T_{qs} = T(x_1, x_2, \ldots, x_n)$ được gọi là giá trị quan sát của tiêu chuẩn kiểm định.

5.3.3 Miền bác bỏ giả thiết

- Sau khi đã chọn tiêu chuẩn kiểm định T, với α bé cho trước (thường α được lấy bằng 0,05 hoặc 0,01) ta có thể tìm được miền W_α sao cho với điều kiện giả thiết H_0 đúng, xác suất để T nhận giá trị trong miền W_α bằng α:

$$P(T \in W_\alpha | H_0) = \alpha.$$

- Giá trị α được gọi là mức ý nghĩa của tiêu chuẩn kiểm định và miền W_α được gọi là miền bác bỏ giả thiết H_0 với mức ý nghĩa α.

5.3.4 Quy tắc kiểm định giả thiết thống kê

- Nếu $T_{qs} \in W_\alpha$ thì H_0 sai, do đó ta bác bỏ H_0, chấp nhận H_1.
- Nếu $T_{qs} \notin W_\alpha$ thì không kết luận được là H_0 đúng mà chỉ có nghĩa là qua mẫu cụ thể này chưa khẳng định được là H_0 sai. Do đó ta chỉ có thể nói rằng chưa có cơ sở để bác bỏ H_0 (trên thực tế là vẫn chấp nhận H_0).

5.3.5 Các loại sai lầm mắc phải khi tiến hành kiểm định

- Sai lầm loại I: Bác bỏ giả thiết H_0 trong khi H_0 đúng.
  Xác suất của sai lầm loại I là: $P(T \in W_\alpha | H_0) = \alpha$.
- Sai lầm loại II: Chấp nhận giả thiết H_0 trong khi H_0 sai.
  Xác suất của sai lầm loại II là: $P(T \notin W_\alpha | H_1) = \beta$.

5.3.6 Các bước kiểm định giả thiết thống kê

- Phát biểu giả thiết H_0 và đối thiết H_1.
- Chọn tiêu chuẩn kiểm định T và xác định quy luật phân bố xác suất của T với điều kiện giả thiết H_0 đúng.
- Từ mẫu cụ thể tính giá trị quan sát của tiêu chuẩn kiểm định T_{qs}.
- Với mức ý nghĩa α, xác định miền bác bỏ W_α tốt nhất tùy thuộc vào đối thiết H_1.
- So sánh giá trị quan sát của tiêu chuẩn kiểm định T_{qs} với miền bác bỏ W_α và kết luận.

Chương 5: Ước lượng tham số và kiểm định giả thiết thống kê

Bài toán

Giả sử biến ngẫu nhiên gốc X trong tổng thể có phân bố chuẩn $N(\mu, \sigma^2)$. Với mức ý nghĩa α, hãy kiểm định cặp giả thiết, đối thiết:

$$\begin{aligned} H_0 : &\quad \mu = \mu_0 \\ H_1 : &\quad \mu \neq \mu_0 \text{ (hoặc } H_1 : \mu > \mu_0, \text{ hoặc } H_1 : \mu < \mu_0) \end{aligned}$$

Từ tổng thể ta lập mẫu ngẫu nhiên kích thước n:

$$(X_1, X_2, \ldots, X_n)$$

Trường hợp 1: phương sai σ^2 đã biết

- Nếu H_0 đúng thì $\overline{X} \sim N(\mu_0, \sigma^2/n)$, ta có thể xây dựng miền bác bỏ dựa vào giá trị tính được của $\overline{X}$.
- Tuy nhiên, để thuận tiện hơn ta sẽ chuẩn tắc hóa $\overline{X}$ và sử dụng tiêu chuẩn kiểm định là

$$T = \frac{\overline{X} - \mu_0}{\sigma}\sqrt{n}$$

- $T \sim N(0, 1)$.

Với mẫu cụ thể, ta thực hiện bài toán kiểm định $H_0 : \mu = \mu_0$ như sau:

1) Tính $T_{qs} = \dfrac{\overline{x} - \mu_0}{\sigma}\sqrt{n}$
2) Kết luận

Đối thiết	Bác bỏ H_0 nếu	Chấp nhận H_0 nếu
$H_1 : \mu \neq \mu_0$	$\lvert T_{qs}\rvert > U_{\alpha/2}$	$\lvert T_{qs}\rvert \leq U_{\alpha/2}$
$H_1 : \mu > \mu_0$	$T_{qs} > U_\alpha$	$T_{qs} \leq U_\alpha$
$H_1 : \mu < \mu_0$	$T_{qs} < -U_\alpha$	$T_{qs} \geq -U_\alpha$

Ví dụ 10. Theo tiêu chuẩn thì trọng lượng các gói mì chính được đóng trên một máy tự động là 450 g. Kiểm tra ngẫu nhiên 81 gói ta thấy trọng lượng trung bình là 445 g. Với mức ý nghĩa 0,05 có thể cho rằng trọng lượng các gói mì chính không đạt tiêu chuẩn hay không, biết rằng trọng lượng gói mì chính là biến ngẫu nhiên có phân bố chuẩn với độ lệch chuẩn là 36g?

Trường hợp 2: phương sai σ^2 chưa biết, $n \geq 30$

- Tiêu chuẩn kiểm định: $T = \dfrac{\overline{X} - \mu_0}{S}\sqrt{n}$
- Nếu H_0 đúng thì T có phân bố xấp xỉ $N(0, 1)$.

Chú ý: Trong trường hợp kích thước mẫu lớn ($n \geq 30$), ta không cần giả thiết X có phân bố chuẩn.

Với mẫu cụ thể, ta thực hiện bài toán kiểm định $H_0 : \mu = \mu_0$ như sau:

1) Tính $T_{qs} = \dfrac{\overline{x} - \mu_0}{s}\sqrt{n}$
2) Kết luận: Giống trường hợp 1.

Ví dụ 11. Lượng nước sạch của một gia đình 4 người ở Hà Nội sử dụng trong 6 tháng năm ngoái là 17 m^3. Theo dõi lượng nước sạch sử dụng trong 6 tháng năm nay của 60 gia đình 4 người thu được số liệu sau:

Lượng nước sạch (m^3)	15-16	16-17	17-18	18-19	19-20
Số gia đình tương ứng	7	15	21	12	5

a) Hãy ước lượng bằng khoảng tin cậy lượng nước sạch trung bình của các hộ sử dụng trong 6 tháng năm nay với độ tin cậy 95%.
b) Có ý kiến cho rằng lượng nước tiêu thụ năm nay tăng lên. Sử dụng bảng số liệu trên, hãy kiểm định ý kiến đó với mức ý nghĩa 2,5%.

Trường hợp 3: phương sai σ^2 chưa biết, $n < 30$

- Tiêu chuẩn kiểm định: $T = \dfrac{\overline{X} - \mu_0}{S}\sqrt{n}$
- Nếu H_0 đúng thì T có phân bố Student với $n - 1$ bậc tự do.

Với mẫu cụ thể, ta thực hiện bài toán kiểm định $H_0 : \mu = \mu_0$ như sau:

1) Tính $T_{qs} = \dfrac{\overline{x} - \mu_0}{s}\sqrt{n}$

2) Kết luận

Đối thiết	Bác bỏ H_0 nếu	Chấp nhận H_0 nếu
$H_1 : \mu \neq \mu_0$	$\lvert T_{qs}\rvert > t_{\alpha/2}(n-1)$	$\lvert T_{qs}\rvert \leq t_{\alpha/2}(n-1)$
$H_1 : \mu > \mu_0$	$T_{qs} > t_{\alpha}(n-1)$	$T_{qs} \leq t_{\alpha}(n-1)$
$H_1 : \mu < \mu_0$	$T_{qs} < -t_{\alpha}(n-1)$	$T_{qs} \geq -t_{\alpha}(n-1)$

Ví dụ 12. Trọng lượng đóng bao của các bao gạo trong kho là Bnn có phân bố chuẩn với trọng lượng trung bình theo quy định là 50 kg. Nghi ngờ gạo bị đóng thiếu, người ta đem cân ngẫu nhiên 25 bao và thu được các số liệu sau:

Trọng lượng bao (kg)	48,25	48,75	49,25	49,75	50,25
Số bao tương ứng	2	5	10	6	2

Với mức ý nghĩa 0,01, hãy kết luận về điều nghi ngờ nói trên.

Chương 5: Ước lượng tham số và kiểm định giả thiết thống kê

Bài toán

Giả sử tỷ lệ cá thể có dấu hiệu A trong một tổng thể là p. Với mức ý nghĩa α, hãy kiểm định cặp giả thiết, đối thiết:

$$\begin{aligned} H_0 &: \quad p = p_0 \\ H_1 &: \quad p \neq p_0 \ (\text{ hoặc } H_1 : p > p_0, \ \text{hoặc } H_1 : p < p_0) \end{aligned}$$

Chọn ngẫu nhiên n cá thể từ tổng thể. Gọi f là tỷ lệ cá thể có dấu hiệu A trong mẫu quan sát.

- Tiêu chuẩn kiểm định: $T = \dfrac{f - p_0}{\sqrt{p_0(1-p_0)}}\sqrt{n}$
- Nếu H_0 đúng thì T có phân bố xấp xỉ $N(0,1)$.

Với mẫu cụ thể, ta thực hiện bài toán kiểm định $H_0 : p = p_0$ như sau:

1) Tính $T_{qs} = \dfrac{f - p_0}{\sqrt{p_0(1-p_0)}}\sqrt{n}$
2) Kết luận

Đối thiết	Bác bỏ H_0 nếu	Chấp nhận H_0 nếu
$H_1 : p \neq p_0$	$\lvert T_{qs}\rvert > U_{\alpha/2}$	$\lvert T_{qs}\rvert \leq U_{\alpha/2}$
$H_1 : p > p_0$	$T_{qs} > U_\alpha$	$T_{qs} \leq U_\alpha$
$H_1 : p < p_0$	$T_{qs} < -U_\alpha$	$T_{qs} \geq -U_\alpha$

Ví dụ 13. Tỷ lệ bệnh nhân được chữa khỏi bệnh khi điều trị bằng thuốc A là 80%. Người ta đưa vào một loại thuốc mới để chữa bệnh cho 1000 người thì có 830 người khỏi bệnh. Với mức ý nghĩa 0,01 có thể cho rằng thuốc mới hiệu quả hơn không?