

KIỂM ĐỊNH GIẢ THUYẾT

Ngày 22 tháng 7 năm 2024

Nội dung

- 1 Các khái niệm và định nghĩa
- 2 Các tiêu chuẩn kiểm định giả thuyết thống kê

Giả thiết thống kê

Kiểm định giả thiết thống kê là một trong các bài toán cơ bản của thống kê toán. Giả thiết thống kê là các giả thiết nói về các tham số, luật phân phối hoặc tính độc lập của các biến ngẫu nhiên. Việc tìm ra kết luận bác bỏ hay chấp nhận một giả thiết thống kê được gọi là một trắc nghiệm hay kiểm định giả thiết thống kê.

Giả thiết thống kê

Để đặt một giả thiết thống kê ta lưu ý các vấn đề sau:

- Giả thiết đặt ra trái với điều cần chứng minh, cần thuyết phục. Như vậy, khi ta bác bỏ giả thiết tức là điều ta cần chứng minh là đúng.
- Giả thiết đặt ra sao cho khi chấp nhận hoặc bác bỏ nó sẽ có tác dụng trả lời câu hỏi bài toán đặt ra.
- Giả thiết đặt ra phải đảm bảo khi nó đúng thì sẽ xác định được luật phân phối xác suất của biến ngẫu nhiên dùng làm tiêu chuẩn kiểm định.
- Giả thiết đặt ra thường là so sánh cái chưa biết với cái đã biết, tức là các thông tin trong quá khứ.
- Giả thiết đặt ra thường mang nghĩa “không khác nhau”, hoặc “sự khác nhau không có ý nghĩa”, hoặc “bằng nhau” giữa các tham số kiểm định.

Ví dụ

Khi khảo sát khối lượng gà xuất chuồng ở một trại gà trong nhiều lần, người ta thấy khối lượng trung bình là 2.7 kg/con. Người ta thử nghiệm một loại thức ăn mới với hy vọng làm tăng khối lượng gà xuất chuồng. Để giải quyết vấn đề “liệu thức ăn mới có làm thay đổi khối lượng gà xuất chuồng hay không” người ta đặt giả thiết như sau:

Ví dụ

- Khối lượng gà xuất chuồng khi sử dụng thức ăn mới không thay đổi, nghĩa là vẫn ở mức 2.7 kg/con. Như vậy, khi bác bỏ giả thiết này có nghĩa là vấn đề cần chứng minh “thức ăn mới làm thay đổi khối lượng gà xuất chuồng” là đúng. Dù kết luận là chấp nhận hay bác bỏ giả thiết này thì ta cũng đều có kết luận về vấn đề đặt ra.
- Mặt khác, nếu coi khối lượng mỗi con gà khi xuất chuồng là một biến ngẫu nhiên, hiển nhiên khối lượng trung bình của mỗi con gà là kỳ vọng của biến ngẫu nhiên đó. Như vậy, khi giả thiết đặt ra đúng có nghĩa là xác định được kỳ vọng của biến ngẫu nhiên đó. Và do đó, ta có thể xác định được luật phân phối xác suất của một biến ngẫu nhiên liên quan đến khối lượng gà xuất chuồng.

Ví dụ

Nếu đặt θ là khối lượng trung bình gà xuất chuồng khi sử dụng thức ăn mới, θ_0 là khối lượng trung bình gà xuất chuồng khi chưa sử dụng thức ăn mới (ở đây $\theta_0 = 2.7$ kg/con). Giả thiết đặt ra như trên có thể viết thành: $H_0 : \theta = \theta_0$. Giả thiết này được gọi là giả thiết cần kiểm định hay giả thiết không. Chữ “không” ở đây có nghĩa là các tham số “không khác nhau”, hay sự khác nhau giữa các tham số là “không có nghĩa”. Để kiểm định ta đặt ra một giả thiết đối với giả thiết H_0 được gọi là giả thiết đối hay “đối thiết”.

Ví dụ

Với giả thiết H_0 vừa nêu ta có hai cách đặt đối thiết

- $H_1 : \theta \neq \theta_0$, kiểm định giả thiết với đối thiết dạng này ta gọi là kiểm định giả thiết hai phía hay trắc nghiệm hai đuôi.
- $H_1 : \theta > \theta_0$ hoặc $H_1 : \theta < \theta_0$, kiểm định giả thiết dạng này được gọi là kiểm định giả thiết một phía hay trắc nghiệm một đuôi. Tùy theo sự thay đổi của tham số cần kiểm định là tăng hay giảm mà ta quyết định kiểm định một phía phải $H_1 : \theta > \theta_0$ hay trái $H_1 : \theta < \theta_0$.

Mức ý nghĩa và miền bác bỏ

Giả sử với giả thiết H_0 và đối thiết H_1 được đặt ra. Từ mẫu ngẫu nhiên $X = (X_1, X_2, \dots, X_n)$, ta xét thống kê

$T = \varphi \circ (X_1, X_2, \dots, X_n, \theta_0)$ sao cho khi H_0 đúng, luật phân phối xác suất của T xác định và giá trị của T có thể tính được trên mẫu cụ thể. Thống kê T khi đó được gọi là tiêu chuẩn kiểm định.

Ký hiệu Ω là tập hợp tất cả các giá trị của mẫu ngẫu nhiên

$X = (X_1, X_2, \dots, X_n)$ và $t = \varphi(x_1, x_2, \dots, x_n)$ là giá trị của T tại mẫu cụ thể (x_1, x_2, \dots, x_n) .

Mức ý nghĩa và miền bác bỏ

Dựa vào T , với mức α cho trước ta tìm một miền W_α sao cho $P(T \in W_\alpha) = \alpha$. Miền W_α được gọi là miền bác bỏ giả thiết H_0 và α được gọi là mức ý nghĩa của phép kiểm định. Thường người ta lấy $\alpha \in (1%; 5\%)$. Khi thực hiện phép thử, ta lấy mẫu cụ thể (x_1, x_2, \dots, x_n) nếu $t = \varphi(x_1, x_2, \dots, x_n) \in W_\alpha$ thì H_0 bị bác bỏ, ngược lại ta chấp nhận H_0 .

Lưu ý:

- Khi chấp nhận H_0 không đồng nghĩa là H_0 đúng mà là với mẫu hiện có chưa đủ bằng chứng để bác bỏ H_0 .
- Trong thực hành ta nói là “có thể chấp nhận H_0 ” hoặc “chưa đủ cơ sở để bác bỏ H_0 ”.

Mức ý nghĩa và miền bác bỏ

- Đối với trắc nghiệm hai đuôi $H_1 : \theta \neq \theta_0$ miền W_α tương đương với $\{(x_1, x_2, \dots, x_n) : |t| = |\varphi(x_1, x_2, \dots, x_n)| > t_{\frac{\alpha}{2}}\}$ được lấy sao cho $P(|T| > t_{\frac{\alpha}{2}}) = \alpha$.
- Đối với trắc nghiệm một đuôi dạng $H_1 : \theta > \theta_0$ miền bác bỏ được lấy sao cho $\{(x_1, x_2, \dots, x_n) : |t| = |\varphi(x_1, x_2, \dots, x_n)| > t_\alpha\}$, với $P(|T| > t_\alpha) = \alpha$.
- Đối với trắc nghiệm một đuôi dạng $H_1 : \theta < \theta_0$ miền bác bỏ được lấy sao cho $\{(x_1, x_2, \dots, x_n) : |t| = |\varphi(x_1, x_2, \dots, x_n)| < t_\alpha\}$, với $P(|T| < t_\alpha) = \alpha$.

Các giá trị $t_{\frac{\alpha}{2}}$ và t_α được gọi là giá trị tới hạn ứng với mức ý nghĩa α .

Sai lầm loại 1 và sai lầm loại 2

- Sai lầm loại 1 là sai lầm khi ta bác bỏ giả thiết H_0 trong khi thực tế nó đúng. Xác suất mắc sai lầm loại 1 này bằng mức ý nghĩa α . Nghĩa là $P(T \in W_\alpha | H_0) = \alpha$. Nếu α càng bé thì khả năng phạm sai lầm càng ít.
- Sai lầm loại 2 là sai lầm khi ta chấp nhận H_0 trong khi thực tế nó sai. Xác suất sai lầm loại 2 là $\beta = P(T \notin W_\alpha | H_1)$ là xác suất T không thuộc miền bác bỏ với điều kiện H_0 sai. Như vậy, xác suất không mắc sai lầm loại 2 là $1 - \beta = P(T \in W_\alpha | H_1)$ được gọi là năng lực của phép kiểm định.

Sai lầm loại 1 và sai lầm loại 2

Sai lầm loại 1 và sai lầm loại 2 đều gây ra tác hại. Chẳng hạn, chấp nhận một lô hàng xấu hoặc từ chối một lô hàng tốt; đánh rớt một thí sinh giỏi hay cho đậu một thí sinh kém; ... đều là các sai lầm tai hại. Do đó, để hạn chế các sai lầm này người ta thường cố định mức xác suất sai lầm loại 1 ở mức ý nghĩa α cho trước. Sau đó chọn các tiêu chuẩn tốt nhất để sai lầm loại 2 nhỏ nhất nghĩa là năng lực kiểm định cao nhất.

Kiểm định giả thiết về trung bình tổng thể

Bài toán: Cho biến ngẫu nhiên X có kỳ vọng là μ chưa biết (còn gọi là trung bình tổng thể).

Với μ_0 cho trước và mẫu (X_1, X_2, \dots, X_n) đặc trưng X , kiểm định giả thiết:

$$H_0 : \mu = \mu_0$$

ở mức α cho trước.

Kiểm định giả thiết về trung bình tổng thể

| Giả thiết $H_0: \mu = \mu_0$ | Đối thiết dạng 1: $H_1: \mu \neq \mu_0$ | Đối thiết dạng 2: $H_1: \mu > \mu_0$ | Đối thiết dạng 3: $H_1: \mu < \mu_0$ |
|---|--|---|---|
| Trường hợp 1: $X \sim N(\mu, \sigma^2)$ biết σ . Nếu H_0 đúng thì $Z = \frac{\bar{X} - \mu_0}{\sigma} \sqrt{n} \sim N(0,1)$. | | | |
| Giá trị tới hạn mức α | $gtth = z_{\alpha/2}$ | $gtth = z_\alpha$ | $gtth = -z_\alpha$ |
| H_0 bị bác bỏ khi | $ z > gtth$ | $z > gtth$ | $z < gtth$ |
| Trường hợp 2: $X \sim N(\mu, \sigma^2)$ chưa biết σ . Nếu H_0 đúng thì | | | |
| | $T = \frac{\bar{X} - \mu_0}{S} \sqrt{n} \sim t(n-1)$ | | |
| Giá trị tới hạn mức α | $gtth = t_{\alpha/2}^{(n-1)}$ | $gtth = t_\alpha^{(n-1)}$ | $gtth = -t_\alpha^{(n-1)}$ |
| H_0 bị bác bỏ khi | $ t > gtth$ | $t > gtth$ | $t < gtth$ |
| Trường hợp 3: Chưa biết phân phối của X mẫu lớn ($n \geq 30$). Nếu H_0 đúng thì | | | |
| | $Z = \frac{\bar{X} - \mu_0}{S} \sqrt{n} \sim N(0,1)$ | | |
| Giá trị tới hạn mức α | $gtth = z_{\alpha/2}$ | $gtth = z_\alpha$ | $gtth = -z_\alpha$ |
| H_0 bị bác bỏ khi | $ z > gtth$ | $z > gtth$ | $z < gtth$ |

Ví dụ

Một phân xưởng của một nhà máy sản xuất một loại trục máy. Sau một thời gian hoạt động, chọn ngẫu nhiên 20 trục máy ta đo được đường kính trung bình là 251.25 và độ lệch chuẩn mẫu là 7.7111. (đơn vị là mm). Biết rằng đường kính của các trục máy của phân xưởng có phân phối chuẩn và đường kính quy định của một trục máy là 250 mm. Ở mức ý nghĩa $\alpha = 5\%$, hãy cho kết luận xem đường kính trục máy do phân xưởng đó sản xuất có đúng quy định không?

Ví dụ

Sản phẩm của một xí nghiệp đúc cho phép số khuyết tật trung bình cho một sản phẩm là 3. Sau một đợt cải tiến kỹ thuật, người ta lấy ngẫu nhiên 36 sản phẩm để kiểm tra thấy trung bình có 2,7222 khuyết tật với độ lệch chuẩn $s = 1,86$. Hãy cho kết luận về hiệu quả của đợt cải tiến kỹ thuật đối với số khuyết tật trung bình của một sản phẩm ở mức ý nghĩa 5%.

Ví dụ

Công ty ABC muốn sản xuất loại bóng đèn có tuổi thọ trung bình $\mu = 1600$ giờ. Nếu thời gian dùng ngắn hơn 1600 giờ thì công ty sẽ mất khách hàng; nếu thời gian dùng dài hơn thì chi phí sản xuất tăng lên. Để biết xem quy trình sản xuất có tốt không, công ty chọn một mẫu ngẫu nhiên gồm 64 bóng đèn đốt thử và thấy tuổi thọ trung bình của chúng là 1575 giờ với độ lệch chuẩn là 121 giờ. Hãy cho kết luận về quy trình sản xuất ở mức ý nghĩa 5%.

Kiểm định giả thiết về tỉ lệ tổng thể (mẫu lớn)

Bài toán:

Giả sử một tổng thể các đối tượng được khảo sát có tỉ lệ đôi
tượng có dấu hiệu A nào đó là p chưa biết. Có thể xem tổng thể
đó là biến ngẫu nhiên X có phân phối Bernoulli với tham số p . Do
đó, p cũng chính là kỳ vọng của biến ngẫu nhiên X . Với mẫu kích
thước n ta tính tỉ lệ mẫu, ký hiệu \bar{P} . Lưu ý mẫu phải lấy khá lớn
sao cho $n \geq 30$.

Xét bài toán kiểm định giả thiết:

$$H_0 : p = p_0$$

ở mức ý nghĩa α , chú ý $np_0 \geq 5$ và $n(1 - p_0) \geq 5$

Kiểm định giả thiết về tỉ lệ tổng thể (mẫu lớn)

| Giả thiết $H_0: p = p_0$ | Đối thiết dạng 1: $H_1: p \neq p_0$ | Đối thiết dạng 2 $H_1: p > p_0$ | Đối thiết dạng 3 $H_1: p < p_0$ |
|--|--|---|---|
| Nếu H_0 đúng thì $Z = \frac{\bar{P} - p_0}{\sqrt{p_0(1-p_0)}} \cdot \sqrt{n} \sim N(0,1).$ | | | |
| Giá trị tới hạn mức α | $gtth = z_{\alpha/2}$ | $gtth = z_\alpha$ | $gtth = -z_\alpha$ |
| H_0 bị bác bỏ khi | $ z > gtth$ | $z > gtth$ | $z < gtth$ |

Ví dụ

Tại một địa phương, bệnh B có tỉ lệ 34%. Sau một đợt điều trị, kiểm tra lại trên 100 người, thấy có 24 người bị bệnh B. Hỏi đợt điều trị có thực sự làm giảm tỉ lệ bệnh B ở địa phương trên không? (kết luận ở mức $\alpha = 0.05$)

Kiểm định giả thiết về phương sai

Giả sử tổng thể X có phân phối $N(\mu, \sigma^2)$ trong đó σ chưa biết.
Dựa vào mẫu cỡ n , chúng ta kiểm định giả thiết $H_0 : \sigma^2 = \sigma_0^2$ ở mức ý nghĩa α cho trước.

Kiểm định giả thiết về phương sai

Trường hợp μ đã biết

| Giả thiết $H_0: \sigma = \sigma_0$ | Đối thiết dạng 1: $H_1: \sigma \neq \sigma_0$ | Đối thiết dạng 2: $H_1: \sigma > \sigma_0$ | Đối thiết dạng 3: $H_1: \sigma < \sigma_0$ |
|---|---|--|--|
| Nếu H_0 đúng thì $Y = \frac{\sum_{i=1}^n (X_i - \mu)^2}{\sigma_0^2} \sim \chi^2(n)$. | | | |
| Giá trị tới hạn mức α | $gtth = \chi^2_{\alpha/2}(n)$ và $\chi^2_{1-\alpha/2}(n)$ | $gtth = \chi^2_{\alpha}(n)$ | $gtth = \chi^2_{1-\alpha}(n)$ |
| H_0 bị bác bỏ khi | $y > \chi^2_{\alpha/2}(n)$ hoặc $y < \chi^2_{1-\alpha/2}(n)$ | $y > gtth$ | $y < gtth$ |

Ví dụ

Bằng phương pháp cũ, người ta tìm được hàm lượng đậm trong một loại hạt đạt mức trung bình là 4,2% với độ lệch chuẩn 0,45%. Người ta làm với phương pháp mới lặp lại 5 lần với kết quả như sau: 2,3%; 2,4%; 4,0%; 5,5%; 5,8%. Hãy cho kết luận về hiệu quả của hai phương pháp trên ở mức ý nghĩa 1%.

Kiểm định giả thiết về phương sai

Trường hợp μ chưa biết

| Giả thiết $H_0: \sigma = \sigma_0$ | Đối thiết dạng 1: $H_1: \sigma \neq \sigma_0$ | Đối thiết dạng 2: $H_1: \sigma > \sigma_0$ | Đối thiết dạng 3: $H_1: \sigma < \sigma_0$ |
|---|--|--|--|
| Nếu H_0 đúng thì $Y = \frac{(n-1)S^2}{\sigma_0^2} \sim \chi^2(n-1)$ | | | |
| Giá trị tới hạn mức α | $gtth = \chi_{\alpha/2}^2(n-1)$ và $\chi_{1-\alpha/2}^2(n-1)$ | $gtth = \chi_{\alpha}^2(n-1)$ | $gtth = \chi_{1-\alpha}^2(n-1)$ |
| H_0 bị bác bỏ khi | $y < \chi_{\alpha/2}^2(n-1)$ hoặc $y > \chi_{1-\alpha/2}^2(n-1)$ | $y > gtth$ | $y < gtth$ |

Ví dụ

Chủ hãng sản xuất một loại thiết bị cho biết sai số đo của thiết bị này có độ lệch chuẩn là 5 mm. Kiểm tra một mẫu gồm 19 thiết bị loại này tính được phương sai mẫu là $s^2 = 33mm^2$. Với mức ý nghĩa 5%, hãy cho nhẫn xét về ý kiến của chủ hãng. Biết sai số đo của thiết bị có phân phối chuẩn.

So sánh hai trung bình với hai mẫu lớn độc lập

Giả sử ta muốn so sánh hai kỳ vọng của hai biến ngẫu nhiên X , Y , độc lập. Từ tổng thể ta lấy hai mẫu độc lập nhau: mẫu đặc trưng X là (X_1, X_2, \dots, X_n) và mẫu đặc trưng Y là (Y_1, Y_2, \dots, Y_n) . Dự vào hai mẫu trên ta kiểm định giả thiết $H_0 : \mu_X = \mu_Y$ ở mức ý nghĩa α .

So sánh hai trung bình với hai mẫu lớn độc lập

| Giả thiết $H_0: \mu_X = \mu_Y$ | Đối thiết dạng 1: $H_1: \mu_X \neq \mu_Y$ | Đối thiết dạng 2: $H_1: \mu_X > \mu_Y$ | Đối thiết dạng 3: $H_1: \mu_X < \mu_Y$ |
|--|--|--|--|
| Trường hợp 1: X, Y có phân phối chuẩn biết σ_X, σ_Y . Nếu H_0 đúng thì: | | | |
| | $Z = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}} \sim N(0,1)$ | | |
| Giá trị tối hạn mức α | $gtth = z_{\alpha/2}$ | $gtth = z_\alpha$ | $gtth = -z_\alpha$ |

So sánh hai trung bình với hai mẫu lớn độc lập

| | | | |
|--|---|-------------------------------|--------------------------------|
| H_0 bị bác bỏ khi | $ z > gtth$ | $z > gtth$ | $z < gtth$ |
| Trường hợp 2: X, Y có phân phối chuẩn cùng phương sai | | | |
| | $T = \frac{\bar{X} - \bar{Y}}{\sqrt{S^2 \left(\frac{1}{n} + \frac{1}{m} \right)}} \sim t(n+m-2)$ với $S^2 = \frac{(n-1)s_X^2 + (m-1)s_Y^2}{n+m-2}$ | | |
| Giá trị tới hạn mức α | $gtth = t_{\alpha/2}^{(n+m-2)}$ | $gtth = t_{\alpha}^{(n+m-2)}$ | $gtth = -t_{\alpha}^{(n+m-2)}$ |
| H_0 bị bác bỏ khi | $ t > gtth$ | $t > gtth$ | $t < gtth$ |

So sánh hai trung bình với hai mẫu lớn độc lập

Trường hợp 3: Chưa biết phân phối của X và Y mẫu lớn ($n \geq 30; m \geq 30$)

$$Z = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{s_X^2}{n} + \frac{s_Y^2}{m}}} \sim N(0,1)$$

| | | | |
|------------------------------|-----------------------|-------------------|--------------------|
| Giá trị tới hạn mức α | $gtth = z_{\alpha/2}$ | $gtth = z_\alpha$ | $gtth = -z_\alpha$ |
| H_0 bị bác bỏ khi | $ z > gtth$ | $z > gtth$ | $z < gtth$ |

Ví dụ

Người ta cho hai nhóm học sinh, theo thứ tự, đại diện cho hai trường A và B, làm một bài kiểm tra. Nhóm thứ nhất gồm 40 học sinh, có điểm trung bình 7,4; nhóm thứ hai gồm 50 học sinh, có điểm trung bình 7,8. Dựa vào mẫu trên, có thể kết luận rằng: Điểm trung bình của trường B tốt hơn điểm trung bình của trường A không? (kết luận ở mức ý nghĩa 4%). Biết rằng điểm số của mỗi học sinh của hai trường A và B có phân phối chuẩn với độ lệch chuẩn, theo thứ tự, là 0,8 và 0,7.

Ví dụ

Trong một công ty sản xuất bóng đèn, người ta muốn kiểm tra sự làm việc của hai phân xưởng A và B. Một mẫu gồm $n = 10$ bóng đèn của phân xưởng A cho tuổi thọ trung bình 4000 giờ với độ lệch chuẩn 200 giờ. Một mẫu gồm $m = 8$ bóng đèn của phân xưởng B cho tuổi thọ trung bình 4300 giờ với độ lệch chuẩn 250 giờ. Biết rằng tuổi thọ của mỗi bóng đèn của hai phân xưởng A và B, theo thứ tự, tuân theo luật phân phối chuẩn có cùng phương sai. Hãy cho kết luận về sự khác nhau giữa tuổi thọ trung bình của hai loại bóng đèn trên ở mức ý nghĩa 1%.

Ví dụ

Giám đốc một hãng sản xuất thép muốn xác định xem có sự khác nhau về năng suất làm việc mỗi công nhân giữa ca ngày và tối hay không. Điều tra một mẫu gồm 100 công nhân của ca ngày sản xuất được trung bình 74,3 (đơn vị: phần/1 giờ) với độ lệch chuẩn 16; một mẫu khác gồm 90 công nhân ca tối thấy năng suất là 69,7 (đơn vị: phần/1 giờ) với độ lệch chuẩn 18. Hãy xem có sự khác nhau về năng suất giữa hai ca không, với mức ý nghĩa 5%.