Xác suất thống kê
KIỂM ĐỊNH GIẢ THUYẾT THỐNG KÊ

Nguyễn Thị Hồng Nhung
nthnhung@hcmus.edu.vn

Khoa Toán
Trường Đại học Khoa Học Tự Nhiên Hồ Chí Minh

Ngày 26 tháng 3 năm 2024

Nội dung

Định nghĩa

Ví dụ 1:

Giám đốc một nhà máy sản xuất bo mạch chủ máy vi tính tuyên bố rằng tuổi thọ trung bình của một bo mạch chủ do nhà máy sản xuất là 5 năm; đây là một giả thuyết về kỳ vọng của biến ngẫu nhiên $X :=$ tuổi thọ của một bo mạch chủ. Để đưa ra kết luận là chấp nhận hay bác bỏ giả thuyết trên, ta cần dựa vào mẫu điều tra và quy tắc kiểm định thống kê.

Định nghĩa

Định nghĩa

- Giả thuyết thống kê là những phát biểu về các tham số, quy luật phân phối, hoặc tính độc lập của các đại lượng ngẫu nhiên.

Định nghĩa

Định nghĩa

- **Giả thuyết thống kê** là những phát biểu về các tham số, quy luật phân phối, hoặc tính độc lập của các đại lượng ngẫu nhiên.
- **Kiểm định giả thuyết** là quá trình mà qua đó có thể quyết định bác bỏ giả thuyết hay không, dựa vào mẫu $(X_1, \ldots, X_n)$ lấy từ tổng thể.

Định nghĩa

Định nghĩa

Trong bài toán kiểm định giả thuyết,

- Giả thuyết không (null hypothesis) là giả thuyết cần được kiểm định. Ký hiệu: H_0.

Định nghĩa

Định nghĩa

Trong bài toán kiểm định giả thuyết,

- Giả thuyết không (null hypothesis) là giả thuyết cần được kiểm định. Ký hiệu: H_0.
- Đối thuyết (alternative hypothesis) là giả thuyết được xem xét để thay thế giả thuyết không. Ký hiệu: H_1.

Định nghĩa

Định nghĩa

Trong bài toán kiểm định giả thuyết,

- Giả thuyết không (null hypothesis) là giả thuyết cần được kiểm định. Ký hiệu: H_0.
- Đối thuyết (alternative hypothesis) là giả thuyết được xem xét để thay thế giả thuyết không. Ký hiệu: H_1.

Định nghĩa

Định nghĩa

Trong bài toán kiểm định giả thuyết,

- Giả thuyết không (null hypothesis) là giả thuyết cần được kiểm định. Ký hiệu: H_0.
- Đối thuyết (alternative hypothesis) là giả thuyết được xem xét để thay thế giả thuyết không. Ký hiệu: H_1.

Định nghĩa

Xét bài toán kiểm định tham số, giả sử ta quan trắc mẫu ngẫu nhiên $(X_1, \ldots, X_n)$ từ biến ngẫu nhiên X có hàm mật độ xác suất $f(x;\theta)$ phụ thuộc vào tham số θ. Gọi Θ là không gian tham số, và Θ_0 và Θ_0^c là hai tập con rời nhau của Θ sao cho $\Theta_0 \cup \Theta_0^c = \Theta$. Giả thuyết (giả thuyết không) và đối thuyết của bài toán có dạng như sau

$$\begin{cases} H_0 : \theta \in \Theta_0 \\ H_1 : \theta \in \Theta_0^c \end{cases}$$

Ví dụ-Giả thuyết không và đối thuyết

1. Gọi μ độ thay đổi trung bình trong huyết áp của một bệnh nhân sau khi dùng thuốc bác sĩ điều trị cần quan tâm đến giả thuyết sau

$$
\begin{array}{ll}
H_0 : \mu = 0 & \text{Không có ảnh hưởng thuốc lên huyết áp của bệnh nhân} \\
H_1 : \mu \neq 0 & \text{Có ảnh hưởng của thuốc lên huyết áp của bệnh nhân.}
\end{array}
$$

Ví dụ-Giả thuyết không và đối thuyết

1. Gọi μ độ thay đổi trung bình trong huyết áp của một bệnh nhân sau khi dùng thuốc bác sĩ điều trị cần quan tâm đến giả thuyết sau

$$\begin{array}{ll} H_0 : \mu = 0 & \text{Không có ảnh hưởng thuốc lên huyết áp của bệnh nhân} \\ H_1 : \mu \neq 0 & \text{Có ảnh hưởng của thuốc lên huyết áp của bệnh nhân.} \end{array}$$

2. Một khách hàng quan tâm đến tỷ lệ sản phẩm kém chất lượng trong một lô hàng mua của một nhà cung cấp. Giả sử tỷ lệ sản phẩm kém tối đa được phép là 5%. Khách hàng cần quan tâm đến giả thuyết sau

$$\begin{array}{ll} H_0 : p \leq 0.05 & \text{Tỷ lệ sản phẩm kém ở mức chấp nhận được} \\ H_1 : p > 0.05 & \text{Tỷ lệ sản phẩm kém cao hơn mức cho phép} \end{array}$$

Cách đặt giả thuyết

1. Giả thuyết được đặt ra với ý đồ bác bỏ nó, nghĩa là giả thuyết đặt ra ngược lại với điều ta muốn chứng minh, muốn thuyết phục.

Cách đặt giả thuyết

1. Giả thuyết được đặt ra với ý đồ bác bỏ nó, nghĩa là giả thuyết đặt ra ngược lại với điều ta muốn chứng minh, muốn thuyết phục.
2. Giả thuyết được đặt ra sao cho khi chấp nhận hay bác bỏ nó sẽ có tác dụng trả lời bài toán thực tế đặt ra.

Cách đặt giả thuyết

1. Giả thuyết được đặt ra với ý đồ bác bỏ nó, nghĩa là giả thuyết đặt ra ngược lại với điều ta muốn chứng minh, muốn thuyết phục.
2. Giả thuyết được đặt ra sao cho khi chấp nhận hay bác bỏ nó sẽ có tác dụng trả lời bài toán thực tế đặt ra.
3. Giả thuyết được đặt ra sao cho nếu nó đúng thì ta sẽ xác định được quy luật phân phối xác suất của đại lượng ngẫu nhiên được chọn làm tiêu chuẩn kiểm định.

Cách đặt giả thuyết

1. Giả thuyết được đặt ra với ý đồ bác bỏ nó, nghĩa là giả thuyết đặt ra ngược lại với điều ta muốn chứng minh, muốn thuyết phục.
2. Giả thuyết được đặt ra sao cho khi chấp nhận hay bác bỏ nó sẽ có tác dụng trả lời bài toán thực tế đặt ra.
3. Giả thuyết được đặt ra sao cho nếu nó đúng thì ta sẽ xác định được quy luật phân phối xác suất của đại lượng ngẫu nhiên được chọn làm tiêu chuẩn kiểm định.
4. Khi đặt giả thuyết, ta thường so sánh cái chưa biết với cái đã biết. Cái chưa biết là điều mà ta cần kiểm định, kiểm tra, làm rõ. "Cái đã biết" là những thông tin trong quá khứ, các định mức kinh tế, kỹ thuật.

Cách đặt giả thuyết

Tổng quát, một bài toán kiểm định giả thuyết cho tham số θ sẽ có một trong 3 dạng dưới đây (θ_0 là giá trị kiểm định đã biết):

Cách đặt giả thuyết

Tổng quát, một bài toán kiểm định giả thuyết cho tham số θ sẽ có một trong 3 dạng dưới đây (θ_0 là giá trị kiểm định đã biết): Hai phía:

$$\begin{cases} H_0 & : \theta = \theta_0 \\ H_1 & : \theta \neq \theta_0 \end{cases}$$

Cách đặt giả thuyết

Tổng quát, một bài toán kiểm định giả thuyết cho tham số θ sẽ có một trong 3 dạng dưới đây (θ_0 là giá trị kiểm định đã biết): Hai phía:

$$\begin{cases} H_0 & : \theta = \theta_0 \\ H_1 & : \theta \neq \theta_0 \end{cases}$$

Một phía bên trái:

$$\begin{cases} H_0 & : \theta \geq \theta_0 \\ H_1 & : \theta < \theta_0 \end{cases}$$

Cách đặt giả thuyết

Tổng quát, một bài toán kiểm định giả thuyết cho tham số θ sẽ có một trong 3 dạng dưới đây (θ_0 là giá trị kiểm định đã biết): Hai phía:

$$\begin{cases} H_0 & : \theta = \theta_0 \\ H_1 & : \theta \neq \theta_0 \end{cases}$$

Một phía bên trái:

$$\begin{cases} H_0 & : \theta \geq \theta_0 \\ H_1 & : \theta < \theta_0 \end{cases}$$

Một phía bên phải:

$$\begin{cases} H_0 & : \theta \leq \theta_0 \\ H_1 & : \theta > \theta_0 \end{cases}$$

Cách đặt giả thuyết

Tổng quát, một bài toán kiểm định giả thuyết cho tham số θ sẽ có một trong 3 dạng dưới đây (θ_0 là giá trị kiểm định đã biết): Hai phía:

$$\begin{cases} H_0 & : \theta = \theta_0 \\ H_1 & : \theta \neq \theta_0 \end{cases}$$

Một phía bên trái:

$$\begin{cases} H_0 & : \theta \geq \theta_0 \\ H_1 & : \theta < \theta_0 \end{cases}$$

Một phía bên phải:

$$\begin{cases} H_0 & : \theta \leq \theta_0 \\ H_1 & : \theta > \theta_0 \end{cases}$$

Miền bác bỏ-Tiêu chuẩn kiểm định

Định nghĩa

Xét bài toán kiểm định giả thuyết H_0 và đối thuyết H_1. Giả sử rằng H_0 đúng, từ mẫu ngẫu nhiên $X = (X_1, \ldots, X_n)$ chọn hàm $Z = h(X_1, \ldots, X_n; \theta_0)$ sao cho với số $\alpha > 0$ bé tùy ý ta có thể tìm dược tập hợp W_α thỏa điều kiện

$$P(Z \in W_\alpha) = \alpha$$

Miền bác bỏ-Tiêu chuẩn kiểm định

Định nghĩa

Xét bài toán kiểm định giả thuyết H_0 và đối thuyết H_1. Giả sử rằng H_0 đúng, từ mẫu ngẫu nhiên $X = (X_1, \ldots, X_n)$ chọn hàm $Z = h(X_1, \ldots, X_n; \theta_0)$ sao cho với số $\alpha > 0$ bé tùy ý ta có thể tìm dược tập hợp W_α thỏa điều kiện

$$P(Z \in W_\alpha) = \alpha$$

Tập hợp W_α gọi là miền bác bỏ giả thuyết H_0 và phần bù W_α^c gọi là miền chấp nhận giả thuyết H_0. Đại lượng ngẫu nhiên $Z = h(X_1, \ldots, X_n; \theta_0)$ gọi là tiêu chuẩn kiểm định giả thuyết H_0. Giá trị α gọi là mức ý nghĩa của bài toán kiểm định.

Miền bác bỏ-Tiêu chuẩn kiểm định

Định nghĩa

Xét bài toán kiểm định giả thuyết H_0 và đối thuyết H_1. Giả sử rằng H_0 đúng, từ mẫu ngẫu nhiên $X = (X_1, \dots, X_n)$ chọn hàm $Z = h(X_1, \dots, X_n; \theta_0)$ sao cho với số $\alpha > 0$ bé tùy ý ta có thể tìm được tập hợp W_α thỏa điều kiện

$$P(Z \in W_\alpha) = \alpha$$

Tập hợp W_α gọi là miền bác bỏ giả thuyết H_0 và phần bù W_α^c gọi là miền chấp nhận giả thuyết H_0. Đại lượng ngẫu nhiên $Z = h(X_1, \dots, X_n; \theta_0)$ gọi là tiêu chuẩn kiểm định giả thuyết H_0. Giá trị α gọi là mức ý nghĩa của bài toán kiểm định.

Miền bác bỏ-Tiêu chuẩn kiểm định

Thực hiện quan trắc dựa trên mẫu ngẫu nhiên $(X_1, \ldots, X_n)$ ta thu được mẫu thực nghiệm $(x_1, \ldots, x_n)$.

Miền bác bỏ-Tiêu chuẩn kiểm định

Thực hiện quan trắc dựa trên mẫu ngẫu nhiên $(X_1, \ldots, X_n)$ ta thu được mẫu thực nghiệm $(x_1, \ldots, x_n)$. Từ mẫu thực nghiệm này, ta tính được giá trị của Z là $z = h(x_1, \ldots, x_n; \theta_0)$.

Miền bác bỏ-Tiêu chuẩn kiểm định

Thực hiện quan trắc dựa trên mẫu ngẫu nhiên $(X_1, \ldots, X_n)$ ta thu được mẫu thực nghiệm $(x_1, \ldots, x_n)$. Từ mẫu thực nghiệm này, ta tính được giá trị của Z là $z = h(x_1, \ldots, x_n; \theta_0)$.

- Nếu $z \in W_\alpha$ thì ta bác bỏ giả thuyết H_0.

Miền bác bỏ-Tiêu chuẩn kiểm định

Thực hiện quan trắc dựa trên mẫu ngẫu nhiên $(X_1, \ldots, X_n)$ ta thu được mẫu thực nghiệm $(x_1, \ldots, x_n)$. Từ mẫu thực nghiệm này, ta tính được giá trị của Z là $z = h(x_1, \ldots, x_n; \theta_0)$.

- Nếu $z \in W_\alpha$ thì ta bác bỏ giả thuyết H_0.
- Nếu $z \in W_\alpha^c$ thì ta chưa đủ cơ sở để bác bỏ H_0.

Sai lầm loại I và loại II

Trong bài toán kiểm định giả thuyết thống kê, ta có thể mắc phải các sai lầm sau

Sai lầm loại I và loại II

Trong bài toán kiểm định giả thuyết thống kê, ta có thể mắc phải các sai lầm sau

a. Sai lầm loại I: là sai lầm mắc phải khi ta bác bỏ H_0 trong khi thực tế giả thuyết H_0 đúng. Xác suất mắc sai lầm loại I được gọi là mức ý nghĩa của kiểm định α.

$$\alpha = P(W_\alpha | H_0)$$

Sai lầm loại I và loại II

Trong bài toán kiểm định giả thuyết thống kê, ta có thể mắc phải các sai lầm sau

a. Sai lầm loại I: là sai lầm mắc phải khi ta bác bỏ H_0 trong khi thực tế giả thuyết H_0 đúng. Xác suất mắc sai lầm loại I được gọi là mức ý nghĩa của kiểm định α.

$$\alpha = P(W_\alpha | H_0)$$

b. Sai lầm loại II: là sai lầm mắc phải khi ta chấp nhận giả thuyết H_0 trong khi thực tế H_0 sai. Xác suất mắc sai lầm loại II được kí hiệu β.

$$\beta = P(W_\alpha^c | H_1)$$

Sai lầm loại I và loại II

Thực tế / Quyết định	H_0 đúng	H_0 sai
Không bác bỏ H_0	Không có sai lầm $(1-\alpha)$	Sai lầm loại II β
Bác bỏ H_0	Sai lầm loại I α	Không có sai lầm $(1-\beta)$

Hình: Sai lầm loại I và sai lầm loại II

Sai lầm loại I và loại II

Ví dụ 1

Khảo sát tốc độ cháy của một loại nhiên liệu rắn dùng để đẩy tên lửa ra khỏi giàn phóng. Giả sử biến ngẫu nhiên $X =$ *tốc độ cháy của nhiên liệu (cm/s) có phân phối chuẩn với kỳ vọng* μ *và độ lệch chuẩn* $\sigma = 2.5$ *.*

Sai lầm loại I và loại II

Ví dụ 1

Khảo sát tốc độ cháy của một loại nhiên liệu rắn dùng để đẩy tên lửa ra khỏi giàn phóng. Giả sử biến ngẫu nhiên $X =$ tốc độ cháy của nhiên liệu (cm/s) có phân phối chuẩn với kỳ vọng μ và độ lệch chuẩn $\sigma = 2.5$.
Ta cần kiểm định giả thuyết

$$\begin{cases} H_0 : \mu = 50 \\ H_1 : \mu \neq 50 \end{cases}$$

Sai lầm loại I và loại II

Ví dụ 1

Khảo sát tốc độ cháy của một loại nhiên liệu rắn dùng để đẩy tên lửa ra khỏi giàn phóng. Giả sử biến ngẫu nhiên $X =$ *tốc độ cháy của nhiên liệu (cm/s) có phân phối chuẩn với kỳ vọng* μ và *độ lệch chuẩn* $\sigma = 2.5$.
Ta cần kiểm định giả thuyết

$$\begin{cases} H_0 : \mu = 50 \\ H_1 : \mu \neq 50 \end{cases}$$

Giả sử ta bác bỏ H_0 *khi* $\bar{x} < 48.5$ *hoặc* $\bar{x} > 51.5$. *Các giá trị 48.5 và 51.5 gọi là giá trị tới hạn (Critial value). Giả sử khảo sát mẫu ngẫu nhiên cỡ* $n = 10$, *ta tìm xác suất sai lầm loại I .*

$$\alpha = \mathbb{P}(\text{ Bác bỏ } H_0 \text{ khi } H_0 \text{ đúng })$$

Sai lầm loại I và loại II

Ví dụ 1

Khảo sát tốc độ cháy của một loại nhiên liệu rắn dùng để đẩy tên lửa ra khỏi giàn phóng. Giả sử biến ngẫu nhiên $X =$ *tốc độ cháy của nhiên liệu (cm/s) có phân phối chuẩn với kỳ vọng* μ và *độ lệch chuẩn* $\sigma = 2.5$.
Ta cần kiểm định giả thuyết

$$\begin{cases} H_0 : \mu = 50 \\ H_1 : \mu \neq 50 \end{cases}$$

Giả sử ta bác bỏ H_0 *khi* $\bar{x} < 48.5$ *hoặc* $\bar{x} > 51.5$. *Các giá trị 48.5 và 51.5 gọi là giá trị tới hạn (Critial value). Giả sử khảo sát mẫu ngẫu nhiên cỡ* $n = 10$, *ta tìm xác suất sai lầm loại I .*

$$\alpha = \mathbb{P}(\text{ Bác bỏ } H_0 \text{ khi } H_0 \text{ đúng })$$

Sai lầm loại I và loại II -Ví dụ (tt)

Tức là,

$$\begin{aligned}\alpha &= \mathbb{P}(\bar{X} < 48.5|\mu = 50) + \mathbb{P}(\bar{X} > 51.5|\mu = 50)\\ &= \mathbb{P}\left(\frac{\bar{X}-50}{2.5/\sqrt{10}} < \frac{48.5-50}{2.5/\sqrt{10}}\right) + \mathbb{P}\left(\frac{\bar{X}-50}{2.5/\sqrt{10}} > \frac{51.5-50}{2.5/\sqrt{10}}\right)\\ &= \mathbb{P}(Z < -1.90) + \mathbb{P}(Z > 1.90) = 0.0287 + 0.0287 = 0.0574\end{aligned}$$

Nghĩa là có 5.74% số mẫu ngẫu nhiên khảo sát được sẽ dẫn đến kết luận bác bỏ giả thuyết $H_0 : \mu = 50(cm/s)$ khi tốc độ trung bình thực sự là 50 (cm/s).

Sai lầm loại I và loại II -Ví dụ (tt)

Tức là,

$$\begin{aligned}\alpha &= \mathbb{P}(\bar{X} < 48.5|\mu = 50) + \mathbb{P}(\bar{X} > 51.5|\mu = 50)\\ &= \mathbb{P}\left(\frac{\bar{X}-50}{2.5/\sqrt{10}} < \frac{48.5-50}{2.5/\sqrt{10}}\right) + \mathbb{P}\left(\frac{\bar{X}-50}{2.5/\sqrt{10}} > \frac{51.5-50}{2.5/\sqrt{10}}\right)\\ &= \mathbb{P}(Z < -1.90) + \mathbb{P}(Z > 1.90) = 0.0287 + 0.0287 = 0.0574\end{aligned}$$

Nghĩa là có 5.74% số mẫu ngẫu nhiên khảo sát được sẽ dẫn đến kết luận bác bỏ giả thuyết $H_0 : \mu = 50(cm/s)$ khi tốc độ trung bình thực sự là 50 (cm/s).

Sai lầm loại I và loại II -Ví dụ (tt)

Ta có thể giảm sai lầm α bằng cách mở rộng miền chấp nhận.

Sai lầm loại I và loại II -Ví dụ (tt)

Ta có thể giảm sai lầm α bằng cách mở rộng miền chấp nhận.
Giả sử với cỡ mẫu $n = 10$, miền chấp nhận là $48 < \bar{x} < 52$, khi đó giá trị của α là

$$\begin{aligned} \alpha &= \mathbb{P}\left(Z < \frac{48-50}{2.5\sqrt{10}}\right) + \mathbb{P}\left(Z > \frac{52-50}{2.5\sqrt{10}}\right) \\ &= 0.0057 + 0.0057 = 0.0114 \end{aligned}$$

Sai lầm loại I và loại II

Cách thứ hai để giảm α là tăng cỡ mẫu khảo sát.

Sai lầm loại I và loại II

Cách thứ hai để giảm α là tăng cỡ mẫu khảo sát.

Giả sử cỡ mẫu $n = 16$, ta có $\frac{\sigma}{\sqrt{n}} = \frac{2.5}{\sqrt{16}} = 0.625$, với miền bác bỏ là $\bar{x} < 48.5$ hoặc $\bar{x} > 51.5$, ta có

Sai lầm loại I và loại II

Cách thứ hai để giảm α là tăng cỡ mẫu khảo sát.

Giả sử cỡ mẫu $n = 16$, ta có $\frac{\sigma}{\sqrt{n}} = \frac{2.5}{\sqrt{16}} = 0.625$, với miền bác bỏ là $\bar{x} < 48.5$ hoặc $\bar{x} > 51.5$, ta có

$$\begin{aligned}
\alpha &= \mathbb{P}(\bar{X} < 48.5 | \mu = 50) + \mathbb{P}(\bar{X} > 51.5 | \mu = 50) \\
&= \mathbb{P}\left(Z < \frac{48.5 - 50}{0.625}\right) + \mathbb{P}\left(Z > \frac{51.5 - 50}{0.625}\right) \\
&= 0.0082 + 0.0082 = 0.0164.
\end{aligned}$$

Xác suất sai lầm loại II , β ,được tính như sau

$$\beta = \mathbb{P}(\text{Không bác bỏ } H_0 \text{ khi } H_0 \text{ sai}).$$

Sai lầm loại I và loại II

Cách thứ hai để giảm α là tăng cỡ mẫu khảo sát.
Giả sử cỡ mẫu $n = 16$, ta có $\frac{\sigma}{\sqrt{n}} = \frac{2.5}{\sqrt{16}} = 0.625$, với miền bác bỏ là $\bar{x} < 48.5$ hoặc $\bar{x} > 51.5$, ta có

$$\begin{aligned}
\alpha &= \mathbb{P}(\bar{X} < 48.5|\mu = 50) + \mathbb{P}(\bar{X} > 51.5|\mu = 50) \\
&= \mathbb{P}\left(Z < \frac{48.5 - 50}{0.625}\right) + \mathbb{P}\left(Z > \frac{51.5 - 50}{0.625}\right) \\
&= 0.0082 + 0.0082 = 0.0164.
\end{aligned}$$

Xác suất sai lầm loại II , β ,được tính như sau

$$\beta = \mathbb{P}(\text{Không bác bỏ } H_0 \text{ khi } H_0 \text{ sai}).$$

Để tính β ta cần chỉ ra một giá trị cụ thể cho tham số trong đối thuyết H_1.

Sai lầm loại I và loại II

Cách thứ hai để giảm α là tăng cỡ mẫu khảo sát.

Giả sử cỡ mẫu $n = 16$, ta có $\frac{\sigma}{\sqrt{n}} = \frac{2.5}{\sqrt{16}} = 0.625$, với miền bác bỏ là $\bar{x} < 48.5$ hoặc $\bar{x} > 51.5$, ta có

$$\begin{aligned}\alpha &= \mathbb{P}(\bar{X} < 48.5|\mu = 50) + \mathbb{P}(\bar{X} > 51.5|\mu = 50) \\ &= \mathbb{P}\left(Z < \frac{48.5 - 50}{0.625}\right) + \mathbb{P}\left(Z > \frac{51.5 - 50}{0.625}\right) \\ &= 0.0082 + 0.0082 = 0.0164.\end{aligned}$$

Xác suất sai lầm loại II , β ,được tính như sau

$$\beta = \mathbb{P}(\text{Không bác bỏ } H_0 \text{ khi } H_0 \text{ sai}).$$

Để tính β ta cần chỉ ra một giá trị cụ thể cho tham số trong đối thuyết H_1.

Sai lầm loại I và loại II

Giả sử với cỡ mẫu $n = 10$, miền chấp nhận của giả thuyết H_0 là $48.5 \leq \bar{X} \leq 51.5$ trong khi giá trị thực sự của $\mu = 52$. Sai lầm β cho bởi

Sai lầm loại I và loại II

Giả sử với cỡ mẫu $n = 10$, miền chấp nhận của giả thuyết H_0 là $48.5 \leq \bar{X} \leq 51.5$ trong khi giá trị thực sự của $\mu = 52$. Sai lầm β cho bởi

$$
\begin{aligned}
\beta &= \mathbb{P}\left(48.5 \leq \bar{X} \leq 51.5 | \mu = 52\right) \\
&= \mathbb{P}\left(\frac{48.5 - 52}{2.5/\sqrt{10}} \leq \frac{\bar{X} - 52}{2.5/\sqrt{10}} \leq \frac{51.5 - 52}{2.5/\sqrt{10}} | \mu = 52\right) \\
&= \mathbb{P}\left(-4.43 \leq Z \leq -0.63\right) = \mathbb{P}(Z \leq -0.63) - \mathbb{P}(Z \leq -4.43) \\
&= 0.2643 - 0.0000 = 0.2463
\end{aligned}
$$

Sai lầm loại I và loại II

Giả sử với cỡ mẫu $n = 10$, miền chấp nhận của giả thuyết H_0 là $48.5 \leq \bar{X} \leq 51.5$ trong khi giá trị thực sự của $\mu = 52$. Sai lầm β cho bởi

$$
\begin{aligned}
\beta &= \mathbb{P}\left(48.5 \leq \bar{X} \leq 51.5 | \mu = 52\right) \\
&= \mathbb{P}\left(\frac{48.5 - 52}{2.5/\sqrt{10}} \leq \frac{\bar{X} - 52}{2.5/\sqrt{10}} \leq \frac{51.5 - 52}{2.5/\sqrt{10}} | \mu = 52\right) \\
&= \mathbb{P}\left(-4.43 \leq Z \leq -0.63\right) = \mathbb{P}(Z \leq -0.63) - \mathbb{P}(Z \leq -4.43) \\
&= 0.2643 - 0.0000 = 0.2463
\end{aligned}
$$

Sai lầm loại I và loại II

Giả sử giá trị thực sự $\mu = 50.5$, khi đó

$$\begin{aligned}\beta &= \mathbb{P}\left(48.5 \leq \bar{X} \leq 51.5 | \mu = 50.5\right) \\ &= \mathbb{P}\left(\frac{48.5 - 50.5}{2.5/\sqrt{10}} \leq \frac{\bar{X} - 50.5}{2.5/\sqrt{10}} \leq \frac{51.5 - 50.5}{2.5/\sqrt{10}} | \mu = 52\right) \\ &= \mathbb{P}\left(-2.53 \leq Z \leq 1.27\right) \\ &= 0.8980 - 0.0057 = 0.8923\end{aligned}$$

Tương tự, α, tăng cỡ mẫu sẽ làm giảm sai lầm β, với cỡ mẫu $n = 16$ và miền chấp nhận là $48 < \bar{X} < 52$, ta tính được $\beta = 0.229$.

Sai lầm loại I và loại II

Nhận xét 1

1 *Ta có thể giảm kích thước của miền bác bỏ (tương ứng tăng kích thước miền chấp nhận),xác suất sai lầm loại I,* α*, bằng cách chọn những điểm tới hạn thích hợp.*

Sai lầm loại I và loại II

Nhận xét 1

1. *Ta có thể giảm kích thước của miền bác bỏ (tương ứng tăng kích thước miền chấp nhận),xác suất sai lầm loại I, α, bằng cách chọn những điểm tới hạn thích hợp.*
2. *Xác suất sai lầm loại I và loại II có liên quan với nhau. Một cỡ mẫu cố định, việc giảm sai lầm loại này sẽ tăng sai lầm loại kia.*

Sai lầm loại I và loại II

Nhận xét 1

1 *Ta có thể giảm kích thước của miền bác bỏ (tương ứng tăng kích thước miền chấp nhận),xác suất sai lầm loại I,* α*, bằng cách chọn những điểm tới hạn thích hợp.*

2 *Xác suất sai lầm loại I và loại II có liên quan với nhau. Một cỡ mẫu cố định, việc giảm sai lầm loại này sẽ tăng sai lầm loại kia.*

3 *Cố định các điểm tới hạn, tăng cỡ mẫu n sẽ làm giảm xác suất sai lầm loại I,* α*, và loại II,* β*.*

Sai lầm loại I và loại II

Nhận xét 1

1. *Ta có thể giảm kích thước của miền bác bỏ (tương ứng tăng kích thước miền chấp nhận),xác suất sai lầm loại I, α, bằng cách chọn những điểm tới hạn thích hợp.*
2. *Xác suất sai lầm loại I và loại II có liên quan với nhau. Một cỡ mẫu cố định, việc giảm sai lầm loại này sẽ tăng sai lầm loại kia.*
3. *Cố định các điểm tới hạn, tăng cỡ mẫu n sẽ làm giảm xác suất sai lầm loại I, α, và loại II, β.*
4. *Nếu H_0 sai, sai lầm β sẽ tăng khi giá trị thực của tham số tiến gần đến giá trị được phát biểu trong giả thuyết H_0.*

Sai lầm loại I và loại II

Nhận xét 1

1. *Ta có thể giảm kích thước của miền bác bỏ (tương ứng tăng kích thước miền chấp nhận),xác suất sai lầm loại I, α, bằng cách chọn những điểm tới hạn thích hợp.*
2. *Xác suất sai lầm loại I và loại II có liên quan với nhau. Một cỡ mẫu cố định, việc giảm sai lầm loại này sẽ tăng sai lầm loại kia.*
3. *Cố định các điểm tới hạn, tăng cỡ mẫu n sẽ làm giảm xác suất sai lầm loại I, α, và loại II, β.*
4. *Nếu H_0 sai, sai lầm β sẽ tăng khi giá trị thực của tham số tiến gần đến giá trị được phát biểu trong giả thuyết H_0.*

$p-$ giá trị

Định nghĩa 4.

Tương ứng với một giá trị thống kê kiểm định tính trên một mẫu các giá trị quan trắc xác định, $p-$ giá trị là mức ý nghĩa nhỏ nhất dùng để bác bỏ giả thuyết H_0.

$p-$ giá trị

Định nghĩa 4.

Tương ứng với một giá trị thống kê kiểm định tính trên một mẫu các giá trị quan trắc xác định, $p-$ giá trị là mức ý nghĩa nhỏ nhất dùng để bác bỏ giả thuyết H_0.

Dựa vào đối thuyết H_1, các bước tính $p-$ giá trị như sau:

1. Xác định thống kê kiểm định: TS. Tính giá trị thống kê kiểm định dựa trên mẫu $(x_1, \ldots, x_n)$, giả sử bằng a.

$p-$ giá trị

Định nghĩa 4.

Tương ứng với một giá trị thống kê kiểm định tính trên một mẫu các giá trị quan trắc xác định, $p-$ giá trị là mức ý nghĩa nhỏ nhất dùng để bác bỏ giả thuyết H_0.

Dựa vào đối thuyết H_1, các bước tính $p-$ giá trị như sau:

1. Xác định thống kê kiểm định: TS. Tính giá trị thống kê kiểm định dựa trên mẫu $(x_1, \ldots, x_n)$, giả sử bằng a.
2. $p-$ giá trị cho bởi

$$p - \text{giá trị} = \begin{cases} \mathbb{P}\left(|TS| > |a| | H_0\right), & \text{kiểm định hai phía} \\ \mathbb{P}\left(TS < a | H_0\right), & \text{kiểm định một phía - bên trái} \\ \mathbb{P}\left(TS > a | H_0\right), & \text{kiểm định một phía - bên phải} \end{cases}$$

Nội dung

Kiểm định giả thuyết cho trường hợp một mẫu

- Kiểm định giả thuyết cho kỳ vọng

Kiểm định giả thuyết cho trường hợp một mẫu

- Kiểm định giả thuyết cho kỳ vọng
    - Trường hợp biết phương sai,

Kiểm định giả thuyết cho trường hợp một mẫu

- Kiểm định giả thuyết cho kỳ vọng
  - Trường hợp biết phương sai,
  - Trường hợp không biết phương sai, mẫu nhỏ,

Kiểm định giả thuyết cho trường hợp một mẫu

- Kiểm định giả thuyết cho kỳ vọng
  - Trường hợp biết phương sai,
  - Trường hợp không biết phương sai, mẫu nhỏ,
  - Trường hợp không biết phương sai, mẫu lớn.

Kiểm định giả thuyết cho trường hợp một mẫu

- Kiểm định giả thuyết cho kỳ vọng
  - Trường hợp biết phương sai,
  - Trường hợp không biết phương sai, mẫu nhỏ,
  - Trường hợp không biết phương sai, mẫu lớn.
- Kiểm định giả thuyết cho tỷ lệ.

Kiểm định giả thuyết cho kỳ vọng. TH biết σ^2

- Các giả định:

Kiểm định giả thuyết cho kỳ vọng. TH biết σ^2

- Các giả định:
  - Mẫu ngẫu nhiên $X_1, \ldots, X_n$ được chọn từ tổng thể có phân phối chuẩn $\mathcal{N}(\mu, \sigma^2)$ với kỳ vọng μ chưa biết.

Kiểm định giả thuyết cho kỳ vọng. TH biết σ^2

- Các giả định:
  - Mẫu ngẫu nhiên $X_1, \ldots, X_n$ được chọn từ tổng thể có phân phối chuẩn $\mathcal{N}(\mu, \sigma^2)$ với kỳ vọng μ chưa biết.
  - Phương sai σ^2 đã biết.

Kiểm định giả thuyết cho kỳ vọng. TH biết σ^2

- Các giả định:
  - Mẫu ngẫu nhiên $X_1, \ldots, X_n$ được chọn từ tổng thể có phân phối chuẩn $\mathcal{N}(\mu, \sigma^2)$ với kỳ vọng μ chưa biết.
  - Phương sai σ^2 đã biết.
  - Cho trước giá trị μ_0, cần so sánh kỳ vọng μ với μ_0.

Kiểm định giả thuyết cho kỳ vọng. TH biết σ^2

- Các giả định:
  - Mẫu ngẫu nhiên $X_1, \ldots, X_n$ được chọn từ tổng thể có phân phối chuẩn $\mathcal{N}(\mu, \sigma^2)$ với kỳ vọng μ chưa biết.
  - Phương sai σ^2 đã biết.
  - Cho trước giá trị μ_0, cần so sánh kỳ vọng μ với μ_0.
- Bài toán kiểm định có 3 trường hợp, với mức ý nghĩa α cho trước:

$$(a)\begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu \neq \mu_0 \end{cases} \qquad (b)\begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu < \mu_0 \end{cases} \qquad (c)\begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu > \mu_0 \end{cases}$$

Kiểm định giả thuyết cho kỳ vọng.
Các bước kiểm định TH biết σ^2

1. Phát biểu giả thuyết không và đối thuyết,

Kiểm định giả thuyết cho kỳ vọng.
Các bước kiểm định TH biết σ^2

1. Phát biểu giả thuyết không và đối thuyết,
2. Dưới giả định H_0 đúng, thống kê kiểm định

$$Z_0 = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \sim \mathcal{N}(0,1),$$

Kiểm định giả thuyết cho kỳ vọng.
Các bước kiểm định TH biết σ^2

1. Phát biểu giả thuyết không và đối thuyết,
2. Dưới giả định H_0 đúng, thống kê kiểm định

$$Z_0 = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \sim \mathcal{N}(0,1),$$

3. Giá trị thống kê kiểm định $z_0 = \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}$,

Kiểm định giả thuyết cho kỳ vọng.
Các bước kiểm định TH biết σ^2

1. Phát biểu giả thuyết không và đối thuyết,
2. Dưới giả định H_0 đúng, thống kê kiểm định

$$Z_0 = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \sim \mathcal{N}(0,1),$$

3. Giá trị thống kê kiểm định $z_0 = \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}$,
4. Xác định mức ý nghĩa α, xác định miền bác bỏ W_α, bảng bên dưới,

Kiểm định giả thuyết cho kỳ vọng.
Các bước kiểm định TH biết σ^2

1. Phát biểu giả thuyết không và đối thuyết,
2. Dưới giả định H_0 đúng, thống kê kiểm định
$$Z_0 = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \sim \mathcal{N}(0, 1),$$
3. Giá trị thống kê kiểm định $z_0 = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$,
4. Xác định mức ý nghĩa α, xác định miền bác bỏ W_α, bảng bên dưới,
5. Kết luận: Nếu $z_0 \in W_\alpha$, bác bỏ giả thuyết H_0, ngược lại $z_0 \notin W_\alpha$, chưa đủ cơ sở bác bỏ giả thuyết H_0.

Kiểm định giả thuyết cho kỳ vọng. TH biết σ^2

Giả Thuyết	Miền bác bỏ
$H_0: \mu = \mu_0$ $H_1: \mu \neq \mu_0$	$W_\alpha = \left\{z_0 : \lvert z_0 \rvert > z_{1-\alpha/2}\right\}$
$H_0: \mu = \mu_0$ $H_1: \mu < \mu_0$	$W_\alpha = \{z_0 : z_0 < -z_{1-\alpha}\}$
$H_0: \mu = \mu_0$ $H_1: \mu > \mu_0$	$W_\alpha = \{z_0 : z_0 > z_{1-\alpha}\}$

Bảng: Miền bác bỏ với đối thuyết tương ứng

Kiểm định giả thuyết cho kỳ vọng. TH biết σ^2

- **Sử dụng $p-$ giá trị (p-value):** tính $p-$giá trị dựa theo đối thuyết và kết luận bác bỏ H_0 khi $p-$giá trị $\leq \alpha$, với mức ý nghĩa α cho trước. Công thức tính $p-$giá trị theo các trường hợp xem ở bảng bên dưới.

Giả Thuyết	$p-$giá trị
$H_0: \mu = \mu_0$ $H_1: \mu \neq \mu_0$	$p = 2[1 - \Phi(\lvert z_0 \rvert)]$
$H_0: \mu = \mu_0$ $H_1: \mu < \mu_0$	$p = \Phi(z_0)$
$H_0: \mu = \mu_0$ $H_1: \mu > \mu_0$	$p = 1 - \Phi(z_0)$

Bảng: $p-$giá trị với đối thuyết tương ứng

Kiểm định giả thuyết cho kỳ vọng. TH biết σ^2

Ví dụ 2:

Dây chuyền sản xuất kem đánh răng P/S được thiết kế để đóng hộp những tuýp kem có trọng lượng trung bình là 6 oz (1 oz $=$ 28g). Một mẫu gồm 30 tuýp kem được chọn ngẫu nhiên để kiểm tra định kỳ. Bộ phận điều khiển dây chuyền phải đảm bảo để trọng lượng trung bình mỗi tuýp kem là 6 oz; nếu nhiều hơn hoặc ít hơn, dây chuyền phải được điều chỉnh lại.
Giả sử trung bình mẫu của 30 tuýp kem là 6.1 oz và độ lệch tiêu chuẩn của tổng thể $\sigma = 0.2$ oz.
Thực hiện kiểm định giả thuyết với mức ý nghĩa 3% để xác định xem dây chuyền sản xuất có vận hành tốt hay không?

So sánh kỳ vọng với một số cho trước, TH biết phương sai

Bài giải 1

Gọi $X(oz)$ là trọng lượng tuýp kem đánh răng. $X \sim \mathcal{N}(\mu, \sigma^2)$, $\sigma = 0.2$

- $H_0 : \mu = 6, \quad H_1 : \mu \neq 6$

So sánh kỳ vọng với một số cho trước, TH biết phương sai

Bài giải 1

Gọi $X(oz)$ là trọng lượng tuýp kem đánh răng. $X \sim \mathcal{N}(\mu, \sigma^2)$, $\sigma = 0.2$

- $H_0 : \mu = 6, \quad H_1 : \mu \neq 6$
- *Dưới giả định H_0 đúng, thống kê kiểm định* $Z_0 = \frac{\bar{X} - \mu_0}{\sigma / \sqrt{n}} \sim \mathcal{N}(0, 1)$

So sánh kỳ vọng với một số cho trước, TH biết phương sai

Bài giải 1

Gọi $X(oz)$ là trọng lượng tuýp kem đánh răng. $X \sim \mathcal{N}(\mu, \sigma^2)$, $\sigma = 0.2$

- $H_0 : \mu = 6, \quad H_1 : \mu \neq 6$
- *Dưới giả định H_0 đúng, thống kê kiểm định* $Z_0 = \frac{\bar{X} - \mu_0}{\sigma / \sqrt{n}} \sim \mathcal{N}(0, 1)$
- *Ta có* $\bar{x} = 6.1, \mu_0 = 6, n = 30, \sigma = 0.2$ *suy ra* $z_0 = \frac{6.1 - 6}{0.2/\sqrt{30}} = 2.739$

So sánh kỳ vọng với một số cho trước, TH biết phương sai

Bài giải 1

Gọi $X(oz)$ là trọng lượng tuýp kem đánh răng. $X \sim \mathcal{N}(\mu, \sigma^2)$, $\sigma = 0.2$

- $H_0 : \mu = 6, \quad H_1 : \mu \neq 6$
- *Dưới giả định H_0 đúng, thống kê kiểm định* $Z_0 = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \sim \mathcal{N}(0,1)$
- *Ta có* $\bar{x} = 6.1, \mu_0 = 6, n = 30, \sigma = 0.2$ *suy ra* $z_0 = \frac{6.1-6}{0.2/\sqrt{30}} = 2.739$
- *Miền bác bỏ* $W_{0.03} = \{_0 \colon |z_0| > z_{1-\alpha/2} = z_{0.985} = 2.17\}$
  Ta có $|z_0| = 2.739 > 2.17 \Rightarrow z_0 \in W_{0.03}$ *Vậy với mức ý nghĩa 3%, bác bỏ H_0, chấp nhận H_1, nghĩa là dây chuyền hoạt động không bình thường.*

So sánh kỳ vọng với một số cho trước, TH biết phương sai

Bài giải 1

Gọi $X(oz)$ là trọng lượng tuýp kem đánh răng. $X \sim \mathcal{N}(\mu, \sigma^2)$, $\sigma = 0.2$

- $H_0 : \mu = 6, \quad H_1 : \mu \neq 6$
- *Dưới giả định H_0 đúng, thống kê kiểm định* $Z_0 = \frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}} \sim \mathcal{N}(0,1)$
- *Ta có* $\bar{x} = 6.1, \mu_0 = 6, n = 30, \sigma = 0.2$ *suy ra* $z_0 = \frac{6.1-6}{0.2/\sqrt{30}} = 2.739$
- *Miền bác bỏ* $W_{0.03} = \{_0\colon |z_0| > z_{1-\alpha/2} = z_{0.985} = 2.17\}$
  Ta có $|z_0| = 2.739 > 2.17 \Rightarrow z_0 \in W_{0.03}$ *Vậy với mức ý nghĩa 3%, bác bỏ H_0, chấp nhận H_1, nghĩa là dây chuyền hoạt động không bình thường.*
- *Hoặc*
  $p-$*giá trị* $= 2 \times [1-\phi(|z_0|)] = 2 \times (1-\phi(2.74)) = 2(1-0.9969) = 0.0062 < 0.03.$
  Vậy với mức ý nghĩa 3%, bác bỏ H_0, chấp nhận H_1, nghĩa là dây chuyền hoạt động không bình thường.

Kiểm định giả thuyết cho kỳ vọng. TH biết σ^2

Ví dụ 3:

Một bệnh viện tại trung tâm thành phố cung cấp dịch vụ cấp cứu tại nhà. Với khoảng 20 xe cấp cứu, mục tiêu của trung tâm là cung cấp dịch vụ cấp cứu trong khoảng thời gian trung bình là 12 phút sau khi nhận được điện thoại yêu cầu. Một mẫu ngẫu nhiên gồm thời gian đáp ứng khi có yêu cầu của 40 ca cấp cứu được chọn. Trung bình mẫu là 13.25 phút. Biết rằng độ lệch tiêu chuẩn của tổng thể là $\sigma = 3.2$ phút. Giám đốc EMS muốn thực hiện một kiểm định, với mức ý nghĩa 5%, để xác định xem liệu thời gian một ca cấp cứu có bé hơn hoặc bằng 12 phút hay không?

So sánh kỳ vọng với một số cho trước, TH biết phương sai

Bài giải 2

Gọi X (phút) là thời gian thực hiện một ca cấp cứu. $X \sim \mathcal{N}(\mu, \sigma^2)$, $\sigma = 3.2$

- $H_0 : \mu \leq 12, \quad H_1 : \mu > 12$

So sánh kỳ vọng với một số cho trước, TH biết phương sai

Bài giải 2

Gọi X (phút) là thời gian thực hiện một ca cấp cứu. $X \sim \mathcal{N}(\mu, \sigma^2)$, $\sigma = 3.2$

- $H_0 : \mu \leq 12, \quad H_1 : \mu > 12$
- *Dưới giả định H_0 đúng, thống kê kiểm đinh* $Z_0 = \frac{\bar{X} - \mu_0}{\sigma / \sqrt{n}} \sim \mathcal{N}(0, 1)$

So sánh kỳ vọng với một số cho trước, TH biết phương sai

Bài giải 2

Gọi X (phút) là thời gian thực hiện một ca cấp cứu. $X \sim \mathcal{N}(\mu, \sigma^2)$, $\sigma = 3.2$

- $H_0 : \mu \leq 12, \quad H_1 : \mu > 12$
- *Dưới giả định* H_0 *đúng, thống kê kiểm đinh* $Z_0 = \frac{\bar{X} - \mu_0}{\sigma / \sqrt{n}} \sim \mathcal{N}(0, 1)$
- *Ta có* $\bar{x} = 13.25, \mu_0 = 12, n = 40, \sigma = 3.2$ *suy ra* $z_0 = \frac{13.25 - 12}{3.2/\sqrt{40}} = 2.47$

So sánh kỳ vọng với một số cho trước, TH biết phương sai

Bài giải 2

Gọi X (phút) là thời gian thực hiện một ca cấp cứu. $X \sim \mathcal{N}(\mu, \sigma^2)$, $\sigma = 3.2$

- $H_0 : \mu \leq 12, \quad H_1 : \mu > 12$
- *Dưới giả định* H_0 *đúng, thống kê kiểm đinh* $Z_0 = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \sim \mathcal{N}(0, 1)$
- *Ta có* $\bar{x} = 13.25, \mu_0 = 12, n = 40, \sigma = 3.2$ *suy ra* $z_0 = \frac{13.25 - 12}{3.2/\sqrt{40}} = 2.47$
- *Miền bác bỏ* $W_{0.05} = \{z_0 : z_0 > z_{1-\alpha} = z_{0.95} = 1.65\}$
  Ta có $z_0 = 2.47 > 1.65 \Rightarrow z_0 \in W_{0.05}$. *Vậy với mức ý nghĩa 5%, bác bỏ* H_0, *chấp nhận* H_1, *nghĩa là thời gian một ca cấp cứu là hơn 12 phút.*

So sánh kỳ vọng với một số cho trước, TH biết phương sai

Bài giải 2

Gọi X (phút) là thời gian thực hiện một ca cấp cứu. $X \sim \mathcal{N}(\mu, \sigma^2)$, $\sigma = 3.2$

- $H_0 : \mu \leq 12, \quad H_1 : \mu > 12$
- *Dưới giả định* H_0 *đúng, thống kê kiểm đinh* $Z_0 = \frac{\bar{X} - \mu_0}{\sigma / \sqrt{n}} \sim \mathcal{N}(0, 1)$
- *Ta có* $\bar{x} = 13.25, \mu_0 = 12, n = 40, \sigma = 3.2$ *suy ra* $z_0 = \frac{13.25 - 12}{3.2/\sqrt{40}} = 2.47$
- *Miền bác bỏ* $W_{0.05} = \{z_0 : z_0 > z_{1-\alpha} = z_{0.95} = 1.65\}$
  Ta có $z_0 = 2.47 > 1.65 \Rightarrow z_0 \in W_{0.05}$. *Vậy với mức ý nghĩa* 5%, *bác bỏ* H_0, *chấp nhận* H_1, *nghĩa là thời gian một ca cấp cứu là hơn 12 phút.*
- *Hoặc* $p - \text{giá trị} = 1 - \phi(z_0) = 1 - \phi(2.47) = 1 - 0.9932 = 0.0068 < 0.05$. *Vậy với mức ý nghĩa* 5%, *bác bỏ* H_0, *chấp nhận* H_1, *nghĩa là thời gian một ca cấp cứu là hơn 12 phút.*

Kiểm định giả thuyết cho kỳ vọng. TH không biết σ^2, mẫu nhỏ

- Các giả định:

Kiểm định giả thuyết cho kỳ vọng. TH không biết σ^2, mẫu nhỏ

- Các giả định:
  - Mẫu ngẫu nhiên $X_1, \ldots, X_n$ được chọn từ tổng thể có phân phối chuẩn $\mathcal{N}(\mu, \sigma^2)$ với kỳ vọng μ và phương sai σ^2 chưa biết.

Kiểm định giả thuyết cho kỳ vọng. TH không biết σ^2, mẫu nhỏ

- Các giả định:
  - Mẫu ngẫu nhiên $X_1, \ldots, X_n$ được chọn từ tổng thể có phân phối chuẩn $\mathcal{N}(\mu, \sigma^2)$ với kỳ vọng μ và phương sai σ^2 chưa biết.
  - Sử dụng ước lượng không chệch S thay cho σ.

Kiểm định giả thuyết cho kỳ vọng.
TH không biết σ^2, mẫu nhỏ

- Các giả định:
  - Mẫu ngẫu nhiên $X_1, \ldots, X_n$ được chọn từ tổng thể có phân phối chuẩn $\mathcal{N}(\mu, \sigma^2)$ với kỳ vọng μ và phương sai σ^2 chưa biết.
  - Sử dụng ước lượng không chệch S thay cho σ.
  - Cỡ mẫu nhỏ: $n \leq 30$.

Kiểm định giả thuyết cho kỳ vọng. TH không biết σ^2, mẫu nhỏ

- Các giả định:
  - Mẫu ngẫu nhiên $X_1, \dots, X_n$ được chọn từ tổng thể có phân phối chuẩn $\mathcal{N}(\mu, \sigma^2)$ với kỳ vọng μ và phương sai σ^2 chưa biết.
  - Sử dụng ước lượng không chệch S thay cho σ.
  - Cỡ mẫu nhỏ: $n \leq 30$.
  - Cho trước giá trị μ_0, cần so sánh kỳ vọng μ với μ_0.

Kiểm định giả thuyết cho kỳ vọng.
TH không biết σ^2, mẫu nhỏ

- Các giả định:
  - Mẫu ngẫu nhiên $X_1, \ldots, X_n$ được chọn từ tổng thể có phân phối chuẩn $\mathcal{N}(\mu, \sigma^2)$ với kỳ vọng μ và phương sai σ^2 chưa biết.
  - Sử dụng ước lượng không chệch S thay cho σ.
  - Cỡ mẫu nhỏ: $n \leq 30$.
  - Cho trước giá trị μ_0, cần so sánh kỳ vọng μ với μ_0.
- Bài toán kiểm định có 3 trường hợp, với mức ý nghĩa α cho trước:

$$(a)\begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu \neq \mu_0 \end{cases} \qquad (b)\begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu < \mu_0 \end{cases} \qquad (c)\begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu > \mu_0 \end{cases}$$

Kiểm định giả thuyết cho kỳ vọng. TH không biết σ^2, mẫu nhỏ

Các bước kiểm định

1. Phát biểu giả thuyết không và đối thuyết,

Kiểm định giả thuyết cho kỳ vọng. TH không biết σ^2, mẫu nhỏ

Các bước kiểm định

1. Phát biểu giả thuyết không và đối thuyết,
2. Dưới giả định H_0 đúng, thống kê kiểm định

$$T_0 = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} \sim T(n-1),$$

Kiểm định giả thuyết cho kỳ vọng. TH không biết σ^2, mẫu nhỏ

Các bước kiểm định

1. Phát biểu giả thuyết không và đối thuyết,
2. Dưới giả định H_0 đúng, thống kê kiểm định

$$T_0 = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} \sim T(n-1),$$

3. Giá trị thống kê kiểm định $t_0 = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$,

Kiểm định giả thuyết cho kỳ vọng. TH không biết σ^2, mẫu nhỏ

Các bước kiểm định

1. Phát biểu giả thuyết không và đối thuyết,
2. Dưới giả định H_0 đúng, thống kê kiểm định

$$T_0 = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} \sim T(n-1),$$

3. Giá trị thống kê kiểm định $t_0 = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$,
4. Xác định mức ý nghĩa α, xác định miền bác bỏ W_α, bảng bên dưới,

Kiểm định giả thuyết cho kỳ vọng. TH không biết σ^2, mẫu nhỏ

Các bước kiểm định

1. Phát biểu giả thuyết không và đối thuyết,
2. Dưới giả định H_0 đúng, thống kê kiểm định

$$T_0 = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} \sim T(n-1),$$

3. Giá trị thống kê kiểm định $t_0 = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$,
4. Xác định mức ý nghĩa α, xác định miền bác bỏ W_α, bảng bên dưới,
5. Kết luận: Nếu $t_0 \in W_\alpha$, bác bỏ giả thuyết H_0, ngược lại $t_0 \notin W_\alpha$, chưa đủ cơ sở bác bỏ giả thuyết H_0.

Kiểm định giả thuyết cho kỳ vọng. TH không biết σ^2, mẫu nhỏ

Giả Thuyết	Miền bác bỏ
$H_0: \mu = \mu_0$ $H_1: \mu \neq \mu_0$	$W_\alpha = \left\{ t_0 : \lvert t_0 \rvert > t^{n-1}_{1-\alpha/2} \right\}$
$H_0: \mu = \mu_0$ $H_1: \mu < \mu_0$	$W_\alpha = \left\{ t_0 : t_0 < -t^{n-1}_{1-\alpha} \right\}$
$H_0: \mu = \mu_0$ $H_1: \mu > \mu_0$	$W_\alpha = \left\{ t_0 : t_0 > t^{n-1}_{1-\alpha} \right\}$

Bảng: Miền bác bỏ với đối thuyết tương ứng (TH mẫu nhỏ)

Kiểm định giả thuyết cho kỳ vọng. TH không biết σ^2, mẫu nhỏ

- **Sử dụng $p-$ giá trị (p-value):** tính $p-$giá trị dựa theo đối thuyết và kết luận bác bỏ H_0 khi $p-$giá trị $\leq \alpha$, với mức ý nghĩa α cho trước. Công thức tính $p-$giá trị theo các trường hợp xem ở bảng bên dưới.

Giả Thuyết	$p-$giá trị
$H_0: \mu = \mu_0$ $H_1: \mu \neq \mu_0$	$p = 2\mathbb{P}(t^{n-1} \geq \lvert t_0 \rvert)$
$H_0: \mu = \mu_0$ $H_1: \mu < \mu_0$	$p = \mathbb{P}(t^{n-1} \leq t_0)$
$H_0: \mu = \mu_0$ $H_1: \mu > \mu_0$	$p = \mathbb{P}(t^{n-1} \geq t_0)$

Bảng: $p-$giá trị với đối thuyết tương ứng (TH mẫu nhỏ)

So sánh kỳ vọng với một số cho trước, TH biết phương sai

Bài giải 3

Gọi X*(gram) là trọng lượng của con lợn khi mới sinh .* $X \sim \mathcal{N}(\mu, \sigma^2)$, σ^2 *chưa biết,* $n = 27 < 30$.

- $H_0 : \mu = 300, \quad H_1 : \mu \neq 300$

So sánh kỳ vọng với một số cho trước, TH biết phương sai

Bài giải 3

Gọi $X(gram)$ *là trọng lượng của con lợn khi mới sinh .* $X \sim \mathcal{N}(\mu, \sigma^2)$, σ^2 *chưa biết,* $n = 27 < 30$.

- $H_0 : \mu = 300, \quad H_1 : \mu \neq 300$
- *Dưới giả định* H_0 *đúng, thống kê kiểm định* $T_0 = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} \sim T(n-1)$

So sánh kỳ vọng với một số cho trước, TH biết phương sai

Bài giải 3

Gọi X*(gram) là trọng lượng của con lợn khi mới sinh .* $X \sim \mathcal{N}(\mu, \sigma^2)$, σ^2 *chưa biết,* $n = 27 < 30$.

- $H_0 : \mu = 300, \quad H_1 : \mu \neq 300$
- *Dưới giả định* H_0 *đúng, thống kê kiểm định* $T_0 = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} \sim T(n-1)$
- *Ta có* $\bar{x} = 325.496, s = 198.79, \mu_0 = 300, n = 27$ *suy ra* $t_0 = \frac{325.496 - 300}{198.79/\sqrt{27}} = 0.666$

So sánh kỳ vọng với một số cho trước, TH biết phương sai

Bài giải 3

Gọi X*(gram) là trọng lượng của con lợn khi mới sinh .* $X \sim \mathcal{N}(\mu, \sigma^2)$, σ^2 *chưa biết,* $n = 27 < 30$.

- $H_0 : \mu = 300, \quad H_1 : \mu \neq 300$
- *Dưới giả định* H_0 *đúng, thống kê kiểm định* $T_0 = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} \sim T(n-1)$
- *Ta có* $\bar{x} = 325.496, s = 198.79, \mu_0 = 300, n = 27$ *suy ra* $t_0 = \frac{325.496 - 300}{198.79/\sqrt{27}} = 0.666$
- *Miền bác bỏ* $|t_0| > t^{n-1}_{1-\alpha/2} = t^{26}_{0.975} = 2.0555$
  Ta có $|t_0| = 0.666 < 2.0555$. *Vậy với mức ý nghĩa* 5%, *chưa đủ cơ sở bác bỏ* H_0, *nghĩa là trọng lượng trung bình là 300 gram.*

Kiểm định giả thuyết cho kỳ vọng. TH không biết σ^2, mẫu lớn

- Các giả định:

Kiểm định giả thuyết cho kỳ vọng. TH không biết σ^2, mẫu lớn

- Các giả định:
  - Mẫu ngẫu nhiên $X_1, \ldots, X_n$ được chọn từ tổng thể có phân phối chuẩn $\mathcal{N}(\mu, \sigma^2)$ với kỳ vọng μ và phương sai σ^2 chưa biết.

Kiểm định giả thuyết cho kỳ vọng. TH không biết σ^2, mẫu lớn

- Các giả định:
  - Mẫu ngẫu nhiên $X_1, \ldots, X_n$ được chọn từ tổng thể có phân phối chuẩn $\mathcal{N}(\mu, \sigma^2)$ với kỳ vọng μ và phương sai σ^2 chưa biết.
  - Sử dụng ước lượng không chệch S thay cho σ.

Kiểm định giả thuyết cho kỳ vọng. TH không biết σ^2, mẫu lớn

- Các giả định:
  - Mẫu ngẫu nhiên $X_1, \ldots, X_n$ được chọn từ tổng thể có phân phối chuẩn $\mathcal{N}(\mu, \sigma^2)$ với kỳ vọng μ và phương sai σ^2 chưa biết.
  - Sử dụng ước lượng không chệch S thay cho σ.
  - Cỡ mẫu lớn: $n > 30$.

Kiểm định giả thuyết cho kỳ vọng. TH không biết σ^2, mẫu lớn

- Các giả định:
  - Mẫu ngẫu nhiên $X_1, \ldots, X_n$ được chọn từ tổng thể có phân phối chuẩn $\mathcal{N}(\mu, \sigma^2)$ với kỳ vọng μ và phương sai σ^2 chưa biết.
  - Sử dụng ước lượng không chệch S thay cho σ.
  - Cỡ mẫu lớn: $n > 30$.
  - Cho trước giá trị μ_0, cần so sánh kỳ vọng μ với μ_0.

Kiểm định giả thuyết cho kỳ vọng. TH không biết σ^2, mẫu lớn

- Các giả định:
  - Mẫu ngẫu nhiên $X_1, \ldots, X_n$ được chọn từ tổng thể có phân phối chuẩn $\mathcal{N}(\mu, \sigma^2)$ với kỳ vọng μ và phương sai σ^2 chưa biết.
  - Sử dụng ước lượng không chệch S thay cho σ.
  - Cỡ mẫu lớn: $n > 30$.
  - Cho trước giá trị μ_0, cần so sánh kỳ vọng μ với μ_0.
- Bài toán kiểm định có 3 trường hợp tương tự hai trường hợp trên, với mức ý nghĩa α cho trước.

Kiểm định giả thuyết cho kỳ vọng. TH không biết σ^2, mẫu lớn

- Các giả định:
  - Mẫu ngẫu nhiên $X_1, \ldots, X_n$ được chọn từ tổng thể có phân phối chuẩn $\mathcal{N}(\mu, \sigma^2)$ với kỳ vọng μ và phương sai σ^2 chưa biết.
  - Sử dụng ước lượng không chệch S thay cho σ.
  - Cỡ mẫu lớn: $n > 30$.
  - Cho trước giá trị μ_0, cần so sánh kỳ vọng μ với μ_0.
- Bài toán kiểm định có 3 trường hợp tương tự hai trường hợp trên, với mức ý nghĩa α cho trước.

Kiểm định giả thuyết cho kỳ vọng. TH không biết σ^2, mẫu lớn

- Các giả định:
  - Mẫu ngẫu nhiên $X_1, \ldots, X_n$ được chọn từ tổng thể có phân phối chuẩn $\mathcal{N}(\mu, \sigma^2)$ với kỳ vọng μ và phương sai σ^2 chưa biết.
  - Sử dụng ước lượng không chệch S thay cho σ.
  - Cỡ mẫu lớn: $n > 30$.
  - Cho trước giá trị μ_0, cần so sánh kỳ vọng μ với μ_0.
- Bài toán kiểm định có 3 trường hợp tương tự hai trường hợp trên, với mức ý nghĩa α cho trước.Khi cỡ mẫu lớn biến ngẫu nhiên

$$Z_0 = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} \tag{1}$$

Kiểm định giả thuyết cho kỳ vọng. TH không biết σ^2, mẫu lớn

- Bài toán kiểm định có 3 trường hợp tương tự hai trường hợp trên, với mức ý nghĩa α cho trước.

Kiểm định giả thuyết cho kỳ vọng. TH không biết σ^2, mẫu lớn

- Bài toán kiểm định có 3 trường hợp tương tự hai trường hợp trên, với mức ý nghĩa α cho trước.
- Khi cỡ mẫu lớn biến ngẫu nhiên

$$Z_0 = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} \tag{2}$$

sẽ hội tụ về phân phối chuẩn hóa $\mathcal{N}(0, 1)$. Khi đó các bước kiểm định cũng như miền bác bỏ W_α hoặc $p-$giá trị sẽ được tính tương tụ như trong trường hợp biết phương sai, chỉ thay thế $\frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}$ bằng Z_0 ở phương trình (**??**).

Kiểm định giả thuyết cho kỳ vọng.
Các bước kiểm định TH không biết $\sigma^2, n \geq 30$

1. Phát biểu giả thuyết không và đối thuyết,

Kiểm định giả thuyết cho kỳ vọng.
Các bước kiểm định TH không biết $\sigma^2, n \geq 30$

1. Phát biểu giả thuyết không và đối thuyết,
2. Dưới giả định H_0 đúng, thống kê kiểm định

$$Z_0 = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} \sim \mathcal{N}(0, 1),$$

Kiểm định giả thuyết cho kỳ vọng.

Các bước kiểm định TH không biết σ^2, $n \geq 30$

1. Phát biểu giả thuyết không và đối thuyết,
2. Dưới giả định H_0 đúng, thống kê kiểm định

$$Z_0 = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} \sim \mathcal{N}(0, 1),$$

3. Giá trị thống kê kiểm định $z_0 = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$,

Kiểm định giả thuyết cho kỳ vọng.
Các bước kiểm định TH không biết $\sigma^2, n \geq 30$

1. Phát biểu giả thuyết không và đối thuyết,
2. Dưới giả định H_0 đúng, thống kê kiểm định

$$Z_0 = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} \sim \mathcal{N}(0, 1),$$

3. Giá trị thống kê kiểm định $z_0 = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$,
4. Xác định mức ý nghĩa α, xác định miền bác bỏ W_α, bảng bên dưới,

Kiểm định giả thuyết cho kỳ vọng.
Các bước kiểm định TH không biết $\sigma^2, n \geq 30$

1. Phát biểu giả thuyết không và đối thuyết,
2. Dưới giả định H_0 đúng, thống kê kiểm định

$$Z_0 = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} \sim \mathcal{N}(0,1),$$

3. Giá trị thống kê kiểm định $z_0 = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$,
4. Xác định mức ý nghĩa α, xác định miền bác bỏ W_α, bảng bên dưới,
5. Kết luận: Nếu $z_0 \in W_\alpha$, bác bỏ giả thuyết H_0, ngược lại $z_0 \notin W_\alpha$, chưa đủ cơ sở bác bỏ giả thuyết H_0.

Kiểm định giả thuyết cho kỳ vọng. TH không biết σ^2, mẫu lớn

Ví dụ 4:

Trạm CSGT trên đường cao tốc sẽ thực hiện việc bắn tốc độ định kỳ tại các địa điểm khác nhau để kiểm tra tốc độ của các PTGT. Một mẫu về tốc độ của các loại xe được chọn để thực hiện kiểm định giả thuyết sau $\begin{cases} H_0: & \mu = 65, \\ H_1: & \mu > 65. \end{cases}$

Những vị trí mà bác bỏ H_0 là những vị trí tốt nhất được chọn để đặt radar kiểm soát tốc độ.

Tại địa điểm F, một mẫu gồm tốc độ của 64 phương tiện được bắn tốc độ ngẫu nhiên có trung bình là 66.2 mph và độ lệch tiêu chuẩn 4.2 mph. Sử dụng $\alpha = 5\%$ để kiểm định giả thuyết.

So sánh kỳ vọng với một số cho trước, TH không biết phương sai

Bài giải 4

Gọi $X(mph)$ *là tốc độ của phương tiện giao thông tại địa điểm F.* $X \sim \mathcal{N}(\mu, \sigma^2)$, σ^2 *chưa biết,* $n = 64 > 30$.

- $H_0 : \mu = 65, \quad H_1 : \mu > 65$

So sánh kỳ vọng với một số cho trước, TH không biết phương sai

Bài giải 4

Gọi $X(mph)$ là tốc độ của phương tiện giao thông tại địa điểm F. $X \sim \mathcal{N}(\mu, \sigma^2)$, σ^2 *chưa biết,* $n = 64 > 30$.

- $H_0 : \mu = 65, \quad H_1 : \mu > 65$
- *Dưới giả định H_0 đúng, thống kê kiểm định* $Z_0 = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} \sim \mathcal{N}(0, 1)$

So sánh kỳ vọng với một số cho trước, TH không biết phương sai

Bài giải 4

Gọi $X(mph)$ *là tốc độ của phương tiện giao thông tại địa điểm F.* $X \sim \mathcal{N}(\mu, \sigma^2)$, σ^2 *chưa biết,* $n = 64 > 30$.

- $H_0 : \mu = 65, \quad H_1 : \mu > 65$
- *Dưới giả định* H_0 *đúng, thống kê kiểm định* $Z_0 = \frac{\bar{X}-\mu_0}{S/\sqrt{n}} \sim \mathcal{N}(0,1)$
- *Ta có* $\bar{x} = 66.2, \mu_0 = 65, n = 64, s = 4.2$ *suy ra* $z_0 = \frac{66.2-65}{4.2/\sqrt{64}} = 2.286$

So sánh kỳ vọng với một số cho trước, TH không biết phương sai

Bài giải 4

Gọi $X(mph)$ là tốc độ của phương tiện giao thông tại địa điểm F. $X \sim \mathcal{N}(\mu, \sigma^2)$, σ^2 chưa biết, $n = 64 > 30$.

- $H_0 : \mu = 65, \quad H_1 : \mu > 65$
- *Dưới giả định H_0 đúng, thống kê kiểm định* $Z_0 = \frac{\bar{X}-\mu_0}{S/\sqrt{n}} \sim \mathcal{N}(0,1)$
- *Ta có* $\bar{x} = 66.2, \mu_0 = 65, n = 64, s = 4.2$ *suy ra* $z_0 = \frac{66.2-65}{4.2/\sqrt{64}} = 2.286$
- $\alpha = 0.05$, *miền bác bỏ* $W_\alpha = \{z_0 : z_0 > z_{1-\alpha} = z_{0.95} = 1.65\}$

So sánh kỳ vọng với một số cho trước, TH không biết phương sai

Bài giải 4

Gọi $X(mph)$ là tốc độ của phương tiện giao thông tại địa điểm F. $X \sim \mathcal{N}(\mu, \sigma^2)$, σ^2 chưa biết, $n = 64 > 30$.

- $H_0 : \mu = 65, \quad H_1 : \mu > 65$
- *Dưới giả định H_0 đúng, thống kê kiểm định* $Z_0 = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} \sim \mathcal{N}(0, 1)$
- *Ta có* $\bar{x} = 66.2, \mu_0 = 65, n = 64, s = 4.2$ *suy ra* $z_0 = \frac{66.2-65}{4.2/\sqrt{64}} = 2.286$
- $\alpha = 0.05$, *miền bác bỏ* $W_\alpha = \{z_0 : z_0 > z_{1-\alpha} = z_{0.95} = 1.65\}$
- *Ta có* $z_0 = 2.286 > 1.65 \to z_0 \in W_\alpha$. *Bác bỏ H_0, chấp nhận H_1 với* $\alpha = 0.05$. *Với độ tin cậy 95% địa điểm F là một nơi tốt để đặt radar kiểm soát tốc độ.*

Kiểm định giả thuyết cho tỷ lệ

- **Bài toán:**
  Cho tổng thể X, trong đó tỷ lệ phần tử mang đặc tính A nào đó là trong tổng thể là p (p chưa biết). Từ mẫu ngẫu nhiên $(X_1, X_2, \ldots, X_n)$ hãy kiểm định

$$\begin{cases} H_0 : p = p_0 \\ H_1 : p \neq p_0 \end{cases} \quad (b) \begin{cases} H_0 : p = p_0 \\ H_1 : p < p_0 \end{cases} \quad (c) \begin{cases} H_0 : p = p_0 \\ H_1 : p > p_0 \end{cases}$$

  với mức ý nghĩa α.

Kiểm định giả thuyết cho tỷ lệ

Ví dụ 5:

Trong kỳ nghỉ giáng sinh và đầu năm mới, Cục An toàn giao thông đã thống kê được rằng có 500 người chết và 25000 người bị thương do các vụ tai nạn giao thông trên toàn quốc. Theo thông cáo của Cục ATGT thì khoảng 50% số vụ tai nạn có liên quan đến rượu bia.
Khảo sát ngẫu nhiên 120 vụ tai nạn thấy có 67 vụ do ảnh hưởng của rượu bia. Sử dụng số liệu trên để kiểm định lời khẳng định của Cục ATGT với mức ý nghĩa $\alpha = 5\%$.

Kiểm định giả thuyết cho tỷ lệ

- Giả định: Cỡ mẫu lớn, để phân phối nhị thức xấp xỉ phân phối chuẩn tốt nhất cần có $np_0 \geq 5$ và $n(1 - p_0) \geq 5$.

Kiểm định giả thuyết cho tỷ lệ

- Giả định: Cỡ mẫu lớn, để phân phối nhị thức xấp xỉ phân phối chuẩn tốt nhất cần có $np_0 \geq 5$ và $n(1 - p_0) \geq 5$.
- Quan sát sự xuất hiện của biến cố "phần tử mang đặc tính A" trong n phép thử độc lập. Gọi Y là số lần xuất hiện biến cố trên thì $Y \sim B(n, p)$. Và

$$\hat{P} = \frac{Y}{n}$$

là một ước lượng không chệch cho p.

Kiểm định giả thuyết cho tỷ lệ

- Giả định: Cỡ mẫu lớn, để phân phối nhị thức xấp xỉ phân phối chuẩn tốt nhất cần có $np_0 \geq 5$ và $n(1-p_0) \geq 5$.
- Quan sát sự xuất hiện của biến cố "phần tử mang đặc tính A" trong n phép thử độc lập. Gọi Y là số lần xuất hiện biến cố trên thì $Y \sim B(n, p)$. Và

$$\hat{P} = \frac{Y}{n}$$

là một ước lượng không chệch cho p.
- Nếu H_0 đúng, thống kê

$$Z_0 = \frac{\hat{P} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \sim \mathcal{N}(0, 1).$$

Chọn z_0 làm tiêu chuẩn kiểm định.

Kiểm định giả thuyết cho tỷ lệ

Các bước kiểm định

1. Phát biểu giả thuyết không và đối thuyết,

Kiểm định giả thuyết cho tỷ lệ

Các bước kiểm định

1. Phát biểu giả thuyết không và đối thuyết,
2. Dưới giả định H_0, thống kê kiểm định

$$Z_0 = \frac{\hat{P} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \sim \mathcal{N}(0,1),$$

Kiểm định giả thuyết cho tỷ lệ

Các bước kiểm định

1. Phát biểu giả thuyết không và đối thuyết,
2. Dưới giả định H_0, thống kê kiểm định

$$Z_0 = \frac{\hat{P} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \sim \mathcal{N}(0,1),$$

3. Tính giá trị thống kê kiểm định

$$z_0 = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}, \tag{3}$$

Kiểm định giả thuyết cho tỷ lệ

Các bước kiểm định

1. Phát biểu giả thuyết không và đối thuyết,
2. Dưới giả định H_0, thống kê kiểm định

$$Z_0 = \frac{\hat{P} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \sim \mathcal{N}(0,1),$$

3. Tính giá trị thống kê kiểm định

$$z_0 = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}, \tag{3}$$

4. Xác định mức ý nghĩa α,, xác định miền bác bỏ W_α : tương tự kiểm định kỳ vọng trường hợp biết phương sai.

Kiểm định giả thuyết cho tỷ lệ

Các bước kiểm định

1. Phát biểu giả thuyết không và đối thuyết,
2. Dưới giả định H_0, thống kê kiểm định
$$Z_0 = \frac{\hat{P} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \sim \mathcal{N}(0,1),$$
3. Tính giá trị thống kê kiểm định
$$z_0 = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}, \tag{3}$$
4. Xác định mức ý nghĩa α,, xác định miền bác bỏ W_α : tương tự kiểm định kỳ vọng trường hợp biết phương sai.
5. Kết luận: Bác bỏ H_0/ Chưa đủ cơ sở bác bỏ H_0.

Kiểm định giả thuyết cho tỷ lệ

Giả Thuyết	Miền bác bỏ
$H_0: \quad p = p_0$ $H_1: \quad p \neq p_0$	$W_\alpha = \left\{z_0 : \lvert z_0 \rvert > z_{1-\alpha/2}\right\}$
$H_0: \quad p = p_0$ $H_1: \quad p < p_0$	$W_\alpha = \left\{z_0 : z_0 < -z_{1-\alpha}\right\}$
$H_0: \quad p = p_0$ $H_1: \quad p > p_0$	$W_\alpha = \left\{z_0 : z_0 > z_{1-\alpha}\right\}$

Bảng: Miền bác bỏ với đối thuyết tương ứng

So sánh tỷ lệ với một số cho trước

Ví dụ 5:

Trong kỳ nghỉ giáng sinh và đầu năm mới, Cục An toàn giao thông đã thống kê được rằng có 500 người chết và 25000 người bị thương do các vụ tai nạn giao thông trên toàn quốc. Theo thông cáo của Cục ATGT thì khoảng 50% số vụ tai nạn có liên quan đến rượu bia.
Khảo sát ngẫu nhiên 120 vụ tai nạn thấy có 67 vụ do ảnh hưởng của rượu bia. Sử dụng số liệu trên để kiểm định lời khẳng định của Cục ATGT với mức ý nghĩa $\alpha = 5\%$.

So sánh tỷ lệ với một số cho trước

Bài giải 5

Gọi Y là số vụ tai nạn giao thông có liên quan đến rượu bia trong kỳ nghỉ Giáng sinh và đầu năm mới. $Y \sim B(n, p)$*,n=120 .*

- $H_0 : p = 0.5 \quad H_1 : p \neq 0.5$

So sánh tỷ lệ với một số cho trước

Bài giải 5

Gọi Y là số vụ tai nạn giao thông có liên quan đến rượu bia trong kỳ nghỉ Giáng sinh và đầu năm mới. $Y \sim B(n, p)$, $n=120$.

- $H_0 : p = 0.5 \quad H_1 : p \neq 0.5$
- *Dưới giả định H_0 đúng, thống kê kiểm định* $Z_0 = \frac{\hat{P} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \sim \mathcal{N}(0, 1)$

So sánh tỷ lệ với một số cho trước

Bài giải 5

Gọi Y là số vụ tai nạn giao thông có liên quan đến rượu bia trong kỳ nghỉ Giáng sinh và đầu năm mới. $Y \sim B(n, p)$,*n=120 .*

- $H_0 : p = 0.5 \quad H_1 : p \neq 0.5$
- *Dưới giả định* H_0 *đúng, thống kê kiểm định* $Z_0 = \frac{\hat{P} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \sim \mathcal{N}(0, 1)$
- *Ta có* $\hat{p} = \frac{67}{120} = 0.5583$, $p_0 = 0.5$, *suy ra* $z_0 = \frac{0.5583 - 0.5}{\sqrt{\frac{0.5 \times 0.5}{120}}} = 1.277$

So sánh tỷ lệ với một số cho trước

Bài giải 5

Gọi Y là số vụ tai nạn giao thông có liên quan đến rượu bia trong kỳ nghỉ Giáng sinh và đầu năm mới. $Y \sim B(n,p)$, $n=120$.

- $H_0: p = 0.5 \quad H_1: p \neq 0.5$
- *Dưới giả định* H_0 *đúng, thống kê kiểm định* $Z_0 = \frac{\hat{P} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \sim \mathcal{N}(0,1)$
- *Ta có* $\hat{p} = \frac{67}{120} = 0.5583$, $p_0 = 0.5$, *suy ra* $z_0 = \frac{0.5583 - 0.5}{\sqrt{\frac{0.5 \times 0.5}{120}}} = 1.277$
- $\alpha = 0.05$, *miền bác bỏ* $W_{0.05} = \{z_0 : |z_0| > z_{1-\alpha/2} = z_{0.975} = 1.96\}$

So sánh tỷ lệ với một số cho trước

Bài giải 5

Gọi Y là số vụ tai nạn giao thông có liên quan đến rượu bia trong kỳ nghỉ Giáng sinh và đầu năm mới. $Y \sim B(n,p)$,*n=120* .

- $H_0 : p = 0.5 \quad H_1 : p \neq 0.5$
- *Dưới giả định* H_0 *đúng, thống kê kiểm định* $Z_0 = \frac{\hat{P} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \sim \mathcal{N}(0,1)$
- *Ta có* $\hat{p} = \frac{67}{120} = 0.5583, p_0 = 0.5,$ *suy ra* $z_0 = \frac{0.5583 - 0.5}{\sqrt{\frac{0.5 \times 0.5}{120}}} = 1.277$
- $\alpha = 0.05,$ *miền bác bỏ* $W_{0.05} = \{z_0 : |z_0| > z_{1-\alpha/2} = z_{0.975} = 1.96\}$
- *Ta có* $|z_0| = 1.277 < 1.96 \to z_0 \notin W_\alpha$*. Chưa đủ cơ sở bác bỏ* H_0*, với* $\alpha = 0.05$. *Với độ tin cậy 95%, có thể tin lời phát biểu của cục ATGT.*

Nội dung

So sánh hai tỷ lệ

- Khảo sát những phần tử thỏa một tính chất A nào đó trên hai tổng thể độc lập với tỷ lệ tương ứng là p_1 và p_2; từ hai tổng thể chọn ra hai mẫu với cỡ lần lượt là n và m. Gọi X và Y là số phần tử thỏa tính chất A trong mẫu 1 và mẫu 2. Khi đó, ta có $X \sim B(n, p_1)$ và $Y \sim B(m, p_2)$.

So sánh hai tỷ lệ

- Khảo sát những phần tử thỏa một tính chất A nào đó trên hai tổng thể độc lập với tỷ lệ tương ứng là p_1 và p_2; từ hai tổng thể chọn ra hai mẫu với cỡ lần lượt là n và m. Gọi X và Y là số phần tử thỏa tính chất A trong mẫu 1 và mẫu 2. Khi đó, ta có $X \sim B(n, p_1)$ và $Y \sim B(m, p_2)$.
- Bài toán: so sánh hai tỷ lệ p_1 và p_2.

So sánh hai tỷ lệ

- Các giả định

So sánh hai tỷ lệ

- Các giả định
  - Hai mẫu độc lập,

So sánh hai tỷ lệ

- Các giả định
  - Hai mẫu độc lập,
  - Cỡ mẫu lớn, $np_1 > 5; n(1-p_1) > 5$ và $mp_2 > 5; m(1-p_2) > 5$

So sánh hai tỷ lệ

- Các giả định
  - Hai mẫu độc lập,
  - Cỡ mẫu lớn, $np_1 > 5; n(1 - p_1) > 5$ và $mp_2 > 5; m(1 - p_2) > 5$
- Bài toán kiểm định giả thuyết, với mức ý nghĩa α gồm các trường hợp sau:

$$(a)\begin{cases} H_0 : p_1 - p_2 = D_0 \\ H_1 : p_1 - p_2 \neq D_0 \end{cases} \quad (b)\begin{cases} H_0 : p_1 - p_2 = D_0 \\ H_1 : p_1 - p_2 < D_0 \end{cases}$$

$$(c)\begin{cases} H_0 : p_1 - p_2 = D_0 \\ H_1 : p_1 - p_2 > D_0 \end{cases}$$

So sánh hai tỷ lệ

Các bước kiểm định

1. Phát biểu giả thuyết không và đối thuyết,
2. Dưới giả định H_0 đúng, thống kê kiểm định

$$Z_0 = \frac{\hat{P}_1 - \hat{P}_2 - D_0}{\sqrt{\hat{P}(1-\hat{P})\left(\frac{1}{n} + \frac{1}{m}\right)}} \sim \mathcal{N}(0,1),$$

   với $\hat{P}_1 = \frac{X}{n}$; $\hat{P}_2 = \frac{Y}{m}$; $\hat{P} = \frac{X+Y}{n+m}$

3. Tính thống kê kiểm định

$$z_0 = \frac{\hat{p}_1 - \hat{p}_2 - D_0}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n} + \frac{1}{m}\right)}}, \tag{4}$$

So sánh hai tỷ lệ

4. Xác định mức ý nghĩa α, xác định miền bác bỏ

Đối thuyết	**Miền bác bỏ**	**$p-$giá trị**
$H_1 : p_1 - p_2 \neq D_0$	$\lvert z_0 \rvert > z_{1-\alpha/2}$	$p = 2[1 - \Phi(\lvert z_0 \rvert)]$
$H_1 : p_1 - p_2 < D_0$	$z_0 < -z_{1-\alpha}$	$p = \Phi(z_0)$
$H_1 : p_1 - p_2 > D_0$	$z_0 > z_{1-\alpha}$	$p = 1 - \Phi(z_0)$

5. Kết luận: Nếu bác bỏ H_0, ta kết luận H_1 đúng với độ tin cậy $(1-\alpha)100\%$. Ngược lại ta kết luận chưa đủ cơ sở để bác bỏ H_0 với α cho trước.

So sánh hai tỷ lệ

Ví dụ 6:

Một công ty sản xuất thuốc cần kiểm tra một loại thuốc có tác dụng làm giảm việc xuất hiện cơn đau ngực ở các bệnh nhân. Công ty thực hiện thí nghiệm trên 400 người, chia làm hai nhóm: nhóm 1 gồm 200 được uống thuốc và nhóm 2 gồm 200 người được uống giả dược. Theo dõi thấy ở nhóm 1 có 8 người lên cơn đau ngực và nhóm 2 có 25 người lên cơn đau ngực. Với mức ý nghĩa $\alpha = 0.05$, hãy cho kết luận về hiệu quả của thuốc mới sản xuất.

So sánh hai tỷ lệ

Bài giải 6

Gọi X,Y lần lượt là số người lên cơn đau ngực trong nhóm dùng thuốc và nhóm dùng giả dược. $X \sim B(n, p_1), Y \sim B(m, p_2)$.

1. $H_0 : p_1 - p_2 = 0 \quad H_1 : p_1 - p_2 < 0$,

So sánh hai tỷ lệ

Bài giải 6

Gọi X,Y lần lượt là số người lên cơn đau ngực trong nhóm dùng thuốc và nhóm dùng giả dược. $X \sim B(n, p_1)$, $Y \sim B(m, p_2)$.

1. $H_0 : p_1 - p_2 = 0 \quad H_1 : p_1 - p_2 < 0$,
2. *Dưới giả định* H_0 *đúng, thống kê kiểm định*

$$Z_0 = \frac{\hat{P}_1 - \hat{P}_2 - D_0}{\sqrt{\hat{P}(1 - \hat{P})\left(\frac{1}{n} + \frac{1}{m}\right)}} \sim \mathcal{N}(0, 1),$$

So sánh hai tỷ lệ

Bài giải 6

Gọi X,Y lần lượt là số người lên cơn đau ngực trong nhóm dùng thuốc và nhóm dùng giả dược. $X \sim B(n, p_1)$, $Y \sim B(m, p_2)$.

1. $H_0 : p_1 - p_2 = 0$ $H_1 : p_1 - p_2 < 0$,
2. *Dưới giả định* H_0 *đúng, thống kê kiểm định*

$$Z_0 = \frac{\hat{P}_1 - \hat{P}_2 - D_0}{\sqrt{\hat{P}(1-\hat{P})\left(\frac{1}{n} + \frac{1}{m}\right)}} \sim \mathcal{N}(0,1),$$

3. $\hat{p}_1 = \frac{8}{200} = 0.04; \hat{p}_2 = \frac{25}{200} = 0.125; \hat{p} = \frac{8+25}{400} = 0.0825, D_0 = 0$

$$z_0 = \frac{0.04 - 0.125 - 0}{\sqrt{0.0825 \times (1 - 0.0825)\left(\frac{1}{200} + \frac{1}{200}\right)}} = -3.0895, \tag{5}$$

p_value=phi(z_0) =phi(-3.0895)=0.001<0.05 -> Bác bỏ H0....

So sánh hai tỷ lệ

4. Mức ý nghĩa $\alpha = 0.05$, miền bác bỏ

$$W_\alpha = \{z_0 : z_0 < -z_{1-\alpha} = -z_{0.95} = -1.65\}$$

$p_{value} = \phi(z_0) \approx \phi(-3.09) = 1 - \phi(3.09) = 0.001$

So sánh hai tỷ lệ

4. Mức ý nghĩa $\alpha = 0.05$, miền bác bỏ

$$W_\alpha = \{z_0 : z_0 < -z_{1-\alpha} = -z_{0.95} = -1.65\}$$

   $p_{value} = \phi(z_0) \approx \phi(-3.09) = 1 - \phi(3.09) = 0.001$

5. Kết luận: Ta có $z_0 = -3.0895 < -1.65$, suy ra $z_0 \in W_\alpha$. (Hoặc $p_{value} = 0.001 < 0.05$) Bác bỏ H_0, chấp nhận H_1, với mức ý ngĩa $\alpha = 0.05$. Nghĩa là thuốc hiệu quả, làm giảm tỷ lệ lên cơn đau ngực với độ tin cậy 95%.

So sánh hai tỷ lệ

Bài 5.126 Gọi X,Y lần lượt là số sản sản phẩm bị lỗi từ mẫu lấy từ máy 1 và máy 2. $X \sim B(n, p_1)$, $Y \sim B(m, p_2)$.

1. $H_0 : p_1 - p_2 = 0 \quad H_1 : p_1 - p_2 \neq 0$,

So sánh hai tỷ lệ

Bài 5.126 Gọi X,Y lần lượt là số sản sản phẩm bị lỗi từ mẫu lấy từ máy 1 và máy 2. $X \sim B(n, p_1)$, $Y \sim B(m, p_2)$.

1. $H_0 : p_1 - p_2 = 0 \quad H_1 : p_1 - p_2 \neq 0$,
2. Dưới giả định H_0 đúng, thống kê kiểm định

$$Z_0 = \frac{\hat{P}_1 - \hat{P}_2 - D_0}{\sqrt{\hat{P}(1-\hat{P})\left(\frac{1}{n} + \frac{1}{m}\right)}} \sim \mathcal{N}(0,1),$$

So sánh hai tỷ lệ

Bài 5.126 Gọi X,Y lần lượt là số sản sản phẩm bị lỗi từ mẫu lấy từ máy 1 và máy 2. $X \sim B(n, p_1)$, $Y \sim B(m, p_2)$.

1. $H_0 : p_1 - p_2 = 0 \quad H_1 : p_1 - p_2 \neq 0$,
2. Dưới giả định H_0 đúng, thống kê kiểm định

$$Z_0 = \frac{\hat{P}_1 - \hat{P}_2 - D_0}{\sqrt{\hat{P}(1-\hat{P})\left(\frac{1}{n}+\frac{1}{m}\right)}} \sim \mathcal{N}(0,1),$$

3. $\hat{p}_1 = \frac{15}{300} = 0.05$; $\hat{p}_2 = \frac{8}{300} \approx 0.03$; $\hat{p} = \frac{15+8}{300+300} \approx 0.04$, $D_0 = 0$

$$z_0 = \frac{0.05 - 0.03 - 0}{\sqrt{0.04 \times (1-0.04)\left(\frac{1}{300}+\frac{1}{300}\right)}} = 1.25,$$

So sánh hai tỷ lệ

4. Mức ý nghĩa $\alpha = 0.05$, miền bác bỏ

$$W_{0.05} = \{z_0 : |z_0| > z_{1-\alpha/2} = z_{0.975} = 1.96\}$$

So sánh hai tỷ lệ

4. Mức ý nghĩa $\alpha = 0.05$, miền bác bỏ

$$W_{0.05} = \{z_0 : |z_0| > z_{1-\alpha/2} = z_{0.975} = 1.96\}$$

5. Kết luận: Ta có $|z_0| = 1.25 < 1.96$, suy ra $z_0 \notin W_{0.05}$. Chưa đủ cơ sở bác bỏ H_0 với mức ý nghĩa $\alpha = 0.05$. Nghĩa là có thể xem như tỷ lệ lỗi của 2 máy là như nhau với độ tin cậy 95%.

So sánh hai tỷ lệ

4. Mức ý nghĩa $\alpha = 0.05$, miền bác bỏ

$$W_{0.05} = \{z_0 : |z_0| > z_{1-\alpha/2} = z_{0.975} = 1.96\}$$

5. Kết luận: Ta có $|z_0| = 1.25 < 1.96$, suy ra $z_0 \notin W_{0.05}$. Chưa đủ cơ sở bác bỏ H_0 với mức ý nghĩa $\alpha = 0.05$. Nghĩa là có thể xem như tỷ lệ lỗi của 2 máy là như nhau với độ tin cậy 95%.

- $p_{value} = 2(1 - \phi(|z_0|) = 2(1 - \phi(1.25) = 2(1 - 0.8944) = 0.2112$

So sánh hai tỷ lệ

Bài 8.33 Gọi X,Y lần lượt là số trẻ có cân nặng hơn 3000 gram ở thành thị và nông thôn. $X \sim B(n, p_1)$, $Y \sim B(m, p_2)$.

1. $H_0 : p_1 - p_2 = 0 \quad H_1 : p_1 - p_2 \neq 0$,

So sánh hai tỷ lệ

Bài 8.33 Gọi X,Y lần lượt là số trẻ có cân nặng hơn 3000 gram ở thành thị và nông thôn. $X \sim B(n, p_1)$, $Y \sim B(m, p_2)$.

1. $H_0 : p_1 - p_2 = 0 \quad H_1 : p_1 - p_2 \neq 0$,
2. Dưới giả định H_0 đúng, thống kê kiểm định

$$Z_0 = \frac{\hat{P}_1 - \hat{P}_2 - D_0}{\sqrt{\hat{P}(1-\hat{P})\left(\frac{1}{n} + \frac{1}{m}\right)}} \sim \mathcal{N}(0, 1),$$

So sánh hai tỷ lệ

Bài 8.33 Gọi X,Y lần lượt là số trẻ có cân nặng hơn 3000 gram ở thành thị và nông thôn. $X \sim B(n, p_1)$, $Y \sim B(m, p_2)$.

1. $H_0 : p_1 - p_2 = 0 \quad H_1 : p_1 - p_2 \neq 0$,
2. Dưới giả định H_0 đúng, thống kê kiểm định

$$Z_0 = \frac{\hat{P}_1 - \hat{P}_2 - D_0}{\sqrt{\hat{P}(1-\hat{P})\left(\frac{1}{n} + \frac{1}{m}\right)}} \sim \mathcal{N}(0,1),$$

3. $\hat{p}_1 = \frac{100}{150} = 0.67$; $\hat{p}_2 = \frac{98}{200} \approx 0.49$; $\hat{p} = \frac{100+98}{150+200} \approx 0.56$, $D_0 = 0$

$$z_0 = \frac{0.67 - 0.49 - 0}{\sqrt{0.56 \times (1-0.56)\left(\frac{1}{150} + \frac{1}{200}\right)}} \approx 3.36,$$

So sánh hai tỷ lệ

4. Mức ý nghĩa $\alpha = 0.05$, miền bác bỏ

$$W_{0.05} = \{z_0 : |z_0| > z_{1-\alpha/2} = z_{0.975} = 1.96\}$$

So sánh hai tỷ lệ

4. Mức ý nghĩa $\alpha = 0.05$, miền bác bỏ

$$W_{0.05} = \{z_0 : |z_0| > z_{1-\alpha/2} = z_{0.975} = 1.96\}$$

5. Kết luận: Ta có $|z_0| = 3.36 > 1.96$, suy ra $z_0 \in W_{0.05}$. Bác bỏ H_0, chấp nhận H_1 với mức ý nghĩa $\alpha = 0.05$. Nghĩa là tỷ lệ trẻ cân nặng hơn 3000 gram ở thành thị và nông thôn có sự khác biệt với độ tin cậy 95%.

So sánh hai tỷ lệ

4. Mức ý nghĩa $\alpha = 0.05$, miền bác bỏ

$$W_{0.05} = \{z_0 : |z_0| > z_{1-\alpha/2} = z_{0.975} = 1.96\}$$

5. Kết luận: Ta có $|z_0| = 3.36 > 1.96$, suy ra $z_0 \in W_{0.05}$. Bác bỏ H_0, chấp nhận H_1 với mức ý nghĩa $\alpha = 0.05$. Nghĩa là tỷ lệ trẻ cân nặng hơn 3000 gram ở thành thị và nông thôn có sự khác biệt với độ tin cậy 95%.

- $H_1 : p_1 - p_2 > 0$

$$W_{0.05} = \{z_0 : z_0 > z_{1-\alpha} = z_{0.95} = 1.645\}$$

Ta có $z_0 = 3.36 > 1.645$, suy ra $z_0 \in W_{0.05}$. Bác bỏ H_0, chấp nhận H_1 với mức ý nghĩa $\alpha = 0.05$. Nghĩa là tỷ lệ trẻ cân nặng hơn 3000 gram ở thành thị cao hơn ở nông thôn với độ tin cậy 95%

Kiểm định giả thuyết về tỷ lệ cho trường hợp hai mẫu

Ví dụ 2

Kiểm tra ngẫu nhiên sản phẩm sản xuất từ hai cơ sở ta có số liệu

- *Cơ sở 1: Có 20 phế phẩm trong 1000 sản phẩm kiểm tra.*
- *Cơ sở 2: Có 30 phế phẩm trong 900 sản phẩm kiểm tra.*

Với mức ý nghĩa $\alpha = 0.02$ *có thể coi rằng tỉ lệ phế phẩm của hai cơ sở sản xuất trên như nhau hay không?*

So sánh hai tỷ lệ

Bài giải 7

Gọi X,Y lần lượt là số phế phẩm của cơ sở 1 và co sở 2.
$X \sim B(n, p_1), Y \sim B(m, p_2)$.

1. $H_0 : p_1 - p_2 = 0$ *tỷ lệ phế phẩm ở 2 cở sở như nhau*
   $H_1 : p_1 - p_2 \neq 0$ *tỷ lệ phế phẩm ở 2 cở sở khác nhau ,*

So sánh hai tỷ lệ

Bài giải 7

Gọi X,Y lần lượt là số phế phẩm của cơ sở 1 và co sở 2.
$X \sim B(n, p_1), Y \sim B(m, p_2)$.

1. $H_0 : p_1 - p_2 = 0$ *tỷ lệ phế phẩm ở 2 cở sở như nhau*
   $H_1 : p_1 - p_2 \neq 0$ *tỷ lệ phế phẩm ở 2 cở sở khác nhau ,*
2. *Dưới giả định* H_0 *đúng, thống kê kiểm định*

$$Z_0 = \frac{\hat{P}_1 - \hat{P}_2 - D_0}{\sqrt{\hat{P}(1-\hat{P})\left(\frac{1}{n}+\frac{1}{m}\right)}} \sim \mathcal{N}(0,1),$$

So sánh hai tỷ lệ

Bài giải 7

Gọi X,Y lần lượt là số phế phẩm của cơ sở 1 và co sở 2.
$X \sim B(n, p_1), Y \sim B(m, p_2)$.

1. $H_0 : p_1 - p_2 = 0$ *tỷ lệ phế phẩm ở 2 cở sở như nhau*
   $H_1 : p_1 - p_2 \neq 0$ *tỷ lệ phế phẩm ở 2 cở sở khác nhau ,*
2. *Dưới giả định* H_0 *đúng, thống kê kiểm định*

$$Z_0 = \frac{\hat{P}_1 - \hat{P}_2 - D_0}{\sqrt{\hat{P}(1-\hat{P})\left(\frac{1}{n} + \frac{1}{m}\right)}} \sim \mathcal{N}(0,1),$$

3. $\hat{p}_1 = \frac{20}{1000} = 0.02; \hat{p}_2 = \frac{30}{900} = \frac{1}{30} \approx 0.033; \hat{p} = \frac{20+30}{1900} \approx 0.0263, D_0 = 0$

So sánh hai tỷ lệ

$$z_0 = \frac{0.02 - 0.033 - 0}{\sqrt{0.0263 \times (1 - 0.0263)\left(\frac{1}{1000} + \frac{1}{900}\right)}} = -1.7681,$$

4. Mức ý nghĩa $\alpha = 0.02$, miền bác bỏ

$$W_{0.02} = \{z_0 : |z_0| > z_{1-\alpha/2} = z_{0.99} = 2.325\}$$

So sánh hai tỷ lệ

$$z_0 = \frac{0.02 - 0.033 - 0}{\sqrt{0.0263 \times (1 - 0.0263)\left(\frac{1}{1000} + \frac{1}{900}\right)}} = -1.7681,$$

4. Mức ý nghĩa $\alpha = 0.02$, miền bác bỏ

$$W_{0.02} = \{z_0 : |z_0| > z_{1-\alpha/2} = z_{0.99} = 2.325\}$$

5. Kết luận: Ta có $|z_0| = 1.7681 < 2.325$, suy ra $z_0 \notin W_{0.02}$. Chưa đủ cơ sở bác bỏ H_0với mức ý nghĩa $\alpha = 0.02$. Vậy có thể xem như tỷ lệ phế phẩm của 2 cơ sở là như nhau với độ tin cậy 95%.

So sánh hai kỳ vọng, TH biết phương sai

- Các giả định:

So sánh hai kỳ vọng, TH biết phương sai

- Các giả định:
  - $X_1, \ldots, X_n$ là mẫu ngẫu nhiên được chọn từ tổng thể 1 có phân phối chuẩn với kỳ vọng μ_1 và phương sai σ_1^2.

So sánh hai kỳ vọng, TH biết phương sai

- Các giả định:
  - $X_1, \ldots, X_n$ là mẫu ngẫu nhiên được chọn từ tổng thể 1 có phân phối chuẩn với kỳ vọng μ_1 và phương sai σ_1^2.
  - $Y_1, \ldots, Y_m$ là mẫu ngẫu nhiên được chọn từ tổng thể 2 có phân phối chuẩn với kỳ vọng μ_2 và phương sai σ_2^2.

So sánh hai kỳ vọng, TH biết phương sai

- Các giả định:
  - $X_1, \ldots, X_n$ là mẫu ngẫu nhiên được chọn từ tổng thể 1 có phân phối chuẩn với kỳ vọng μ_1 và phương sai σ_1^2.
  - $Y_1, \ldots, Y_m$ là mẫu ngẫu nhiên được chọn từ tổng thể 2 có phân phối chuẩn với kỳ vọng μ_2 và phương sai σ_2^2.
  - Tổng thể 1 và 2 (đại diện bởi X và Y) độc lập với nhau.

So sánh hai kỳ vọng, TH biết phương sai

- Các giả định:
  - $X_1, \ldots, X_n$ là mẫu ngẫu nhiên được chọn từ tổng thể 1 có phân phối chuẩn với kỳ vọng μ_1 và phương sai σ_1^2.
  - $Y_1, \ldots, Y_m$ là mẫu ngẫu nhiên được chọn từ tổng thể 2 có phân phối chuẩn với kỳ vọng μ_2 và phương sai σ_2^2.
  - Tổng thể 1 và 2 (đại diện bởi X và Y) độc lập với nhau.
  - Các phương sai σ_1^2 và σ_2^2 đã biết.

So sánh hai kỳ vọng, TH biết phương sai

- Bài toán kiểm định giả thuyết trên hai mẫu độc lập với mức ý nghĩa α cho trước gồm các dạng sau:

$$(a)\begin{cases} H_0 : \mu_1 - \mu_2 = D_0 \\ H_1 : \mu_1 - \mu_2 \neq D_0 \end{cases} \quad (b)\begin{cases} H_0 : \mu_1 - \mu_2 = D_0 \\ H_1 : \mu_1 - \mu_2 < D_0 \end{cases}$$

$$(c)\begin{cases} H_0 : \mu_1 - \mu_2 = D_0 \\ H_1 : \mu_1 - \mu_2 > D_0 \end{cases}$$

So sánh hai kỳ vọng, TH biết phương sai

Các bước kiểm định

1. Phát biểu giả thuyết không và đối thuyết,

So sánh hai kỳ vọng, TH biết phương sai

Các bước kiểm định

1. Phát biểu giả thuyết không và đối thuyết,
2. Dưới giả định H_0 đúng, thống kê kiểm định

$$Z_0 = \frac{\bar{X} - \bar{Y} - D_0}{\sqrt{\left(\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}\right)}} \sim \mathcal{N}(0, 1),$$

So sánh hai kỳ vọng, TH biết phương sai

Các bước kiểm định

1. Phát biểu giả thuyết không và đối thuyết,
2. Dưới giả định H_0 đúng, thống kê kiểm định

$$Z_0 = \frac{\bar{X} - \bar{Y} - D_0}{\sqrt{\left(\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}\right)}} \sim \mathcal{N}(0,1),$$

3. Tính thống kê kiểm định

$$z_0 = \frac{\bar{x} - \bar{y} - D_0}{\sqrt{\left(\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}\right)}} \tag{6}$$

So sánh hai kỳ vọng, TH biết phương sai

4. Xác định mức ý nghĩa α,,xác định miền bác bỏ

Đối thuyết	**Miền bác bỏ**	**$p-$giá trị**
$H_1 : \mu_1 - \mu_2 \neq D_0$	$\lvert z_0 \rvert > z_{1-\alpha/2}$	$p = 2[1 - \Phi(\lvert z_0 \rvert)]$
$H_1 : \mu_1 - \mu_2 < D_0$	$z_0 < -z_{1-\alpha}$	$p = \Phi(z_0)$
$H_1 : \mu_1 - \mu_2 > D_0$	$z_0 > z_{1-\alpha}$	$p = 1 - \Phi(z_0)$

So sánh hai kỳ vọng, TH biết phương sai

4. Xác định mức ý nghĩa α,,xác định miền bác bỏ

Đối thuyết	**Miền bác bỏ**	**$p-$giá trị**
$H_1 : \mu_1 - \mu_2 \neq D_0$	$\lvert z_0 \rvert > z_{1-\alpha/2}$	$p = 2[1 - \Phi(\lvert z_0 \rvert)]$
$H_1 : \mu_1 - \mu_2 < D_0$	$z_0 < -z_{1-\alpha}$	$p = \Phi(z_0)$
$H_1 : \mu_1 - \mu_2 > D_0$	$z_0 > z_{1-\alpha}$	$p = 1 - \Phi(z_0)$

5. Kết luận: Nếu bác bỏ H_0, ta kết luận H_1 đúng với độ tin cậy $(1-\alpha)100\%$. Ngược lại ta kết luận chưa đủ cơ sở để bác bỏ H_0 với α cho trước.

So sánh hai kỳ vọng, TH biết phương sai

Ví dụ 7:

Một công ty sản xuất sơn nghiên cứu về 1 loại phụ gia làm giảm thời gian khô của sơn. Thực hiên thí nghiệm trên 2 mẫu: mẫu thứ nhất gồm 10 vật mẫu được sơn bằng loại sơn bình thường; mẫu thứ hai gồm 10 mẫu vật được sơn với sơn có chất phụ gia mới. Trong những nghiên cứu trước, biết rằng độ lệch tiêu chuẩn của thời gian khô sau khi quét sơn là 8 phút và không thay đổi khi thêm phụ gia vào. Trung bình của mẫu 1 và mẫu 2 lần lượt là $\bar{x} = 121$ phút và $\bar{y} = 112$ phút. Với mức ý nghĩa 5%, hãy cho kết luận về loại sơn với chất phụ gia mới.

So sánh hai kỳ vọng, TH biết phương sai

Bài giải 8

Gọi X *(phút) là thời gian khô của sơn bình thường.* $X \sim \mathcal{N}(\mu_1, \sigma_1^2)$.
Y *(phút) là thời gian khô của sơn có thêm chất phụ gia.* $Y \sim \mathcal{N}(\mu_2, \sigma_2^2)$,.

- $H_0 : \mu_1 - \mu_2 = 0$
  Sơn thêm chất phụ gia không làm giảm thời gian khô
  $H_1 : \mu_1 - \mu_2 > 0$ *Sơn thêm chất phụ gia làm giảm thời gian khô*

So sánh hai kỳ vọng, TH biết phương sai

Bài giải 8

Gọi X *(phút) là thời gian khô của sơn bình thường.* $X \sim \mathcal{N}(\mu_1, \sigma_1^2)$.
Y *(phút) là thời gian khô của sơn có thêm chất phụ gia.* $Y \sim \mathcal{N}(\mu_2, \sigma_2^2)$,.

- $H_0 : \mu_1 - \mu_2 = 0$
  Sơn thêm chất phụ gia không làm giảm thời gian khô
  $H_1 : \mu_1 - \mu_2 > 0$ *Sơn thêm chất phụ gia làm giảm thời gian khô*
- *Dưới giả định* H_0 *đúng, thống kê kiểm định*

$$Z_0 = \frac{\bar{X} - \bar{Y} - D_0}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}} \sim \mathcal{N}(0, 1)$$

So sánh hai kỳ vọng, TH biết phương sai

Bài giải 8

Gọi X *(phút) là thời gian khô của sơn bình thường.* $X \sim \mathcal{N}(\mu_1, \sigma_1^2)$.
Y *(phút) là thời gian khô của sơn có thêm chất phụ gia.* $Y \sim \mathcal{N}(\mu_2, \sigma_2^2)$,.

- $H_0 : \mu_1 - \mu_2 = 0$
  Sơn thêm chất phụ gia không làm giảm thời gian khô
  $H_1 : \mu_1 - \mu_2 > 0$ *Sơn thêm chất phụ gia làm giảm thời gian khô*
- *Dưới giả định* H_0 *đúng, thống kê kiểm định*

$$Z_0 = \frac{\bar{X} - \bar{Y} - D_0}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}} \sim \mathcal{N}(0, 1)$$

- *Ta có* $\bar{x} = 121, \bar{y} = 112, \ \sigma_1 = \sigma_2 = 8, D_0 = 0,$ *suy ra*
  $z_0 = \frac{121 - 112 - 0}{\sqrt{\frac{8^2}{10} + \frac{8^2}{10}}} = 2.515 \approx 2.52$

So sánh hai kỳ vọng, TH biết phương sai

- $\alpha = 0.05$, miền bác bỏ $W_\alpha = \{z_0 : z_0 > z_{1-\alpha} = z_{0.95} = 1.65\}$
  Hoặc
  $$p_{value} = 1 - \phi(z_0) = 1 - \phi(2.52) = 0.0051$$

So sánh hai kỳ vọng, TH biết phương sai

- $\alpha = 0.05$, miền bác bỏ $W_\alpha = \{z_0 : z_0 > z_{1-\alpha} = z_{0.95} = 1.65\}$
  Hoặc
  $$p_{value} = 1 - \phi(z_0) = 1 - \phi(2.52) = 0.0051$$
- Ta có $z_0 = 2.515 > 1.65 \rightarrow z_0 \in W_\alpha$. Bác bỏ H_0, chấp nhận H_1 với $\alpha = 0.05$. Với độ tin cậy 95%, sơn có thể chất phụ gia làm giảm thời gian khô của sơn.
  Hoặc $p_{value} = 0.0051 < 0.05$ nên Bác bỏ H_0, chấp nhận H_1 với $\alpha = 0.05$. Với độ tin cậy 95%, sơn có thể chất phụ gia làm giảm thời gian khô của sơn

So sánh hai kỳ vọng, TH biết phương sai

Bài giải 9

Gọi $X(cm/s)$ *là tốc độ cháy của nhiên liệu 1.* $X \sim \mathcal{N}(\mu_1, \sigma_1^2)$.
$Y(cm/s)$ *là tốc độ cháy của nhiên liệu 1.* $Y \sim \mathcal{N}(\mu_2, \sigma_2^2)$,.

- $H_0 : \mu_1 - \mu_2 = 0 \quad H_1 : \mu_1 - \mu_2 \neq 0$

So sánh hai kỳ vọng, TH biết phương sai

Bài giải 9

Gọi $X(cm/s)$ *là tốc độ cháy của nhiên liệu 1.* $X \sim \mathcal{N}(\mu_1, \sigma_1^2)$.
$Y(cm/s)$ *là tốc độ cháy của nhiên liệu 1.* $Y \sim \mathcal{N}(\mu_2, \sigma_2^2)$,.

- $H_0 : \mu_1 - \mu_2 = 0 \quad H_1 : \mu_1 - \mu_2 \neq 0$
- *Dưới giả định* H_0 *đúng, thống kê kiểm định*

$$Z_0 = \frac{\bar{X} - \bar{Y} - D_0}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}} \sim \mathcal{N}(0, 1)$$

So sánh hai kỳ vọng, TH biết phương sai

Bài giải 9

Gọi $X(cm/s)$ là tốc độ cháy của nhiên liệu 1. $X \sim \mathcal{N}(\mu_1, \sigma_1^2)$.
$Y(cm/s)$ là tốc độ cháy của nhiên liệu 1. $Y \sim \mathcal{N}(\mu_2, \sigma_2^2)$,.

- $H_0 : \mu_1 - \mu_2 = 0 \quad H_1 : \mu_1 - \mu_2 \neq 0$
- *Dưới giả định H_0 đúng, thống kê kiểm định*

$$Z_0 = \frac{\bar{X} - \bar{Y} - D_0}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}} \sim \mathcal{N}(0, 1)$$

- *Ta có* $\bar{x} = 18, \bar{y} = 24, \ \sigma_1 = \sigma_2 = 3, D_0 = 0, n = m = 20$ *suy ra* $z_0 = \frac{18-24-0}{\sqrt{\frac{3^2}{20}+\frac{3^2}{20}}} = -6.325$

So sánh hai kỳ vọng, TH biết phương sai

- $\alpha = 0.05$, miền bác bỏ $W_\alpha = \{z_0 : |z_0| > z_{1-\alpha/2} = z_{0.975} = 1.96\}$

So sánh hai kỳ vọng, TH biết phương sai

- $\alpha = 0.05$, miền bác bỏ $W_\alpha = \{z_0 : |z_0| > z_{1-\alpha/2} = z_{0.975} = 1.96\}$
- Ta có $|z_0| = 6.235 > 1.96 \to z_0 \in W_\alpha$. Bác bỏ H_0, chấp nhận H_1 với $\alpha = 0.05$. Tốc độ cháy của hai nhiên liệu khác nhau,với độ tin cậy 95%.

So sánh kỳ vọng, TH không biết phương sai, mẫu lớn

- Các giả định:

So sánh kỳ vọng, TH không biết phương sai, mẫu lớn

- Các giả định:
  - $X_1, \ldots, X_n$ là mẫu ngẫu nhiên được chọn từ tổng thể 1 có phân phối chuẩn với kỳ vọng μ_1 và phương sai σ_1^2.

So sánh kỳ vọng, TH không biết phương sai, mẫu lớn

- Các giả định:
  - $X_1, \ldots, X_n$ là mẫu ngẫu nhiên được chọn từ tổng thể 1 có phân phối chuẩn với kỳ vọng μ_1 và phương sai σ_1^2.
  - $Y_1, \ldots, Y_m$ là mẫu ngẫu nhiên được chọn từ tổng thể 2 có phân phối chuẩn với kỳ vọng μ_2 và phương sai σ_2^2.

So sánh kỳ vọng, TH không biết phương sai, mẫu lớn

- Các giả định:
  - $X_1, \ldots, X_n$ là mẫu ngẫu nhiên được chọn từ tổng thể 1 có phân phối chuẩn với kỳ vọng μ_1 và phương sai σ_1^2.
  - $Y_1, \ldots, Y_m$ là mẫu ngẫu nhiên được chọn từ tổng thể 2 có phân phối chuẩn với kỳ vọng μ_2 và phương sai σ_2^2.
  - Tổng thể 1 và 2 (đại diện bởi X và Y) độc lập với nhau.

So sánh kỳ vọng, TH không biết phương sai, mẫu lớn

- Các giả định:
  - $X_1, \ldots, X_n$ là mẫu ngẫu nhiên được chọn từ tổng thể 1 có phân phối chuẩn với kỳ vọng μ_1 và phương sai σ_1^2.
  - $Y_1, \ldots, Y_m$ là mẫu ngẫu nhiên được chọn từ tổng thể 2 có phân phối chuẩn với kỳ vọng μ_2 và phương sai σ_2^2.
  - Tổng thể 1 và 2 (đại diện bởi X và Y) độc lập với nhau.
  - Các phương sai σ_1^2 và σ_2^2 chưa biết.

So sánh hai kỳ vọng, TH biết phương sai

- Bài toán kiểm định giả thuyết trên hai mẫu độc lập với mức ý nghĩa α cho trước gồm các dạng sau:

$$(a)\begin{cases} H_0: \mu_1 - \mu_2 = D_0 \\ H_1: \mu_1 - \mu_2 \neq D_0 \end{cases} \quad (b)\begin{cases} H_0: \mu_1 - \mu_2 = D_0 \\ H_1: \mu_1 - \mu_2 < D_0 \end{cases}$$

$$(c)\begin{cases} H_0: \mu_1 - \mu_2 = D_0 \\ H_1: \mu_1 - \mu_2 > D_0 \end{cases}$$

So sánh hai kỳ vọng, TH không biết phương sai, mẫu lớn

Các bước kiểm định

1. Phát biểu giả thuyết không và đối thuyết,

So sánh hai kỳ vọng, TH không biết phương sai, mẫu lớn

Các bước kiểm định

1. Phát biểu giả thuyết không và đối thuyết,
2. Dưới giả định H_0 đúng, thống kê kiểm định

$$Z_0 = \frac{\bar{X} - \bar{Y} - D_0}{\sqrt{\left(\frac{S_1^2}{n} + \frac{S_2^2}{m}\right)}} \sim \mathcal{N}(0, 1). \tag{7}$$

So sánh hai kỳ vọng, TH không biết phương sai, mẫu lớn

Các bước kiểm định

1. Phát biểu giả thuyết không và đối thuyết,
2. Dưới giả định H_0 đúng, thống kê kiểm định

$$Z_0 = \frac{\bar{X} - \bar{Y} - D_0}{\sqrt{\left(\frac{S_1^2}{n} + \frac{S_2^2}{m}\right)}} \sim \mathcal{N}(0, 1). \tag{7}$$

3. Tính thống kê kiểm định

$$z_0 = \frac{\bar{x} - \bar{y} - D_0}{\sqrt{\left(\frac{s_1^2 1}{n} + \frac{s_2^2 2}{m}\right)}} \tag{8}$$

So sánh hai kỳ vọng, TH biết phương sai

4. Xác định mức ý nghĩa α,,xác định miền bác bỏ

Đối thuyết	**Miền bác bỏ**	**$p-$giá trị**
$H_1 : \mu_1 - \mu_2 \neq D_0$	$\lvert z_0\rvert > z_{1-\alpha/2}$	$p = 2[1 - \Phi(\lvert z_0\rvert)]$
$H_1 : \mu_1 - \mu_2 < D_0$	$z_0 < -z_{1-\alpha}$	$p = \Phi(z_0)$
$H_1 : \mu_1 - \mu_2 > D_0$	$z_0 > z_{1-\alpha}$	$p = 1 - \Phi(z_0)$

So sánh hai kỳ vọng, TH biết phương sai

4. Xác định mức ý nghĩa α,,xác định miền bác bỏ

Đối thuyết	**Miền bác bỏ**	**$p-$giá trị**
$H_1 : \mu_1 - \mu_2 \neq D_0$	$\lvert z_0 \rvert > z_{1-\alpha/2}$	$p = 2[1 - \Phi(\lvert z_0 \rvert)]$
$H_1 : \mu_1 - \mu_2 < D_0$	$z_0 < -z_{1-\alpha}$	$p = \Phi(z_0)$
$H_1 : \mu_1 - \mu_2 > D_0$	$z_0 > z_{1-\alpha}$	$p = 1 - \Phi(z_0)$

5. Kết luận: Nếu bác bỏ H_0, ta kết luận H_1 đúng với độ tin cậy $(1-\alpha)100\%$. Ngược lại ta kết luận chưa đủ cơ sở để bác bỏ H_0 với α cho trước.

So sánh kỳ vọng, TH không biết phương sai, mẫu lớn

Ví dụ 8:

Khảo sát về chiều cao của sinh hai khoa Toán và CNTT: chọn ngẫu nhiên 50 sinh viên khoa Toán, tính được chiều cao trung bình là 163 cm và độ lệch tiêu chuẩn 5 cm. Đo chiều cao 50 sinh viên khoa CNTT, có trung bình mẫu là 166 cm và độ lệch tiêu chuẩn 8 cm. Với mức ý nghĩa $\alpha = 1\%$, hãy cho kết luận về chiều cao của sinh viên hai khoa.

So sánh hai kỳ vọng, TH không biết phương sai, mẫu lớn

Bài giải 10

Gọi $X(cm)$ là chiều cao của sinh viên khoa Toán, $Y(cm)$ là chiều cao của sinh viên khoa CNTT. $n = m = 50 > 30$.

- $H_0 : \mu_X - \mu_T = 0 \quad H_1 : \mu_X - \mu_Y \neq 0$

So sánh hai kỳ vọng, TH không biết phương sai, mẫu lớn

Bài giải 10

Gọi X (cm) là chiều cao của sinh viên khoa Toán, Y (cm) là chiều cao của sinh viên khoa CNTT. $n = m = 50 > 30$.

- $H_0 : \mu_X - \mu_T = 0 \quad H_1 : \mu_X - \mu_Y \neq 0$
- *Dưới giả định* H_0 *đúng, thống kê kiểm định*

$$Z_0 = \frac{\bar{X} - \bar{Y} - D_0}{\sqrt{\frac{S_X^2}{n} + \frac{S_Y^2}{m}}} \sim \mathcal{N}(0, 1)$$

So sánh hai kỳ vọng, TH không biết phương sai, mẫu lớn

Bài giải 10

Gọi X(cm) là chiều cao của sinh viên khoa Toán, Y(cm) là chiều cao của sinh viên khoa CNTT. $n = m = 50 > 30$.

- $H_0 : \mu_X - \mu_T = 0 \quad H_1 : \mu_X - \mu_Y \neq 0$
- *Dưới giả định* H_0 *đúng, thống kê kiểm định*

$$Z_0 = \frac{\bar{X} - \bar{Y} - D_0}{\sqrt{\frac{S_X^2}{n} + \frac{S_Y^2}{m}}} \sim \mathcal{N}(0,1)$$

- *Ta có* $\bar{x} = 163, \bar{y} = 166, s_X = 5, s_Y = 8, D_0 = 0$, *suy ra* $z_0 = \frac{163-166-0}{\sqrt{\frac{5^2}{50}+\frac{8^2}{50}}} = -2.248$

So sánh hai kỳ vọng, TH không biết phương sai, mẫu lớn

- $\alpha = 0.01$, miền bác bỏ $W_\alpha = \{z_0 : |z_0| > z_{1-\alpha/2} = z_{0.995} = 2.58\}$

So sánh hai kỳ vọng, TH không biết phương sai, mẫu lớn

- $\alpha = 0.01$, miền bác bỏ $W_\alpha = \{z_0 : |z_0| > z_{1-\alpha/2} = z_{0.995} = 2.58\}$
- Ta có $|z_0| = 2.25 < 2.58 \rightarrow z_0 \notin W_\alpha$. Chưa đủ cơ sở bác bỏ H_0, với $\alpha = 0.01$. Với độ tin cậy 99%, có thể xem như chiều cao của sinh viên hai khoa là như nhau.

So sánh hai kỳ vọng, TH không biết phương sai, mẫu nhỏ

- Các giả định:

So sánh hai kỳ vọng, TH không biết phương sai, mẫu nhỏ

- Các giả định:
  - $X_1, \ldots, X_n$ là mẫu ngẫu nhiên được chọn từ tổng thể 1 có phân phối chuẩn với kỳ vọng μ_1 và phương sai σ_1^2.

So sánh hai kỳ vọng, TH không biết phương sai, mẫu nhỏ

- Các giả định:
  - $X_1, \ldots, X_n$ là mẫu ngẫu nhiên được chọn từ tổng thể 1 có phân phối chuẩn với kỳ vọng μ_1 và phương sai σ_1^2.
  - $Y_1, \ldots, Y_m$ là mẫu ngẫu nhiên được chọn từ tổng thể 2 có phân phối chuẩn với kỳ vọng μ_2 và phương sai σ_2^2.

So sánh hai kỳ vọng, TH không biết phương sai, mẫu nhỏ

- Các giả định:
  - $X_1, \ldots, X_n$ là mẫu ngẫu nhiên được chọn từ tổng thể 1 có phân phối chuẩn với kỳ vọng μ_1 và phương sai σ_1^2.
  - $Y_1, \ldots, Y_m$ là mẫu ngẫu nhiên được chọn từ tổng thể 2 có phân phối chuẩn với kỳ vọng μ_2 và phương sai σ_2^2.
  - Tổng thể 1 và 2 (đại diện bởi X và Y) độc lập với nhau.

So sánh hai kỳ vọng, TH không biết phương sai, mẫu nhỏ

- Các giả định:
  - $X_1, \ldots, X_n$ là mẫu ngẫu nhiên được chọn từ tổng thể 1 có phân phối chuẩn với kỳ vọng μ_1 và phương sai σ_1^2.
  - $Y_1, \ldots, Y_m$ là mẫu ngẫu nhiên được chọn từ tổng thể 2 có phân phối chuẩn với kỳ vọng μ_2 và phương sai σ_2^2.
  - Tổng thể 1 và 2 (đại diện bởi X và Y) độc lập với nhau.
  - Các phương sai σ_1^2 và σ_2^2 chưa biết.

So sánh hai kỳ vọng, TH không biết phương sai, mẫu nhỏ

- Các giả định:
  - $X_1, \ldots, X_n$ là mẫu ngẫu nhiên được chọn từ tổng thể 1 có phân phối chuẩn với kỳ vọng μ_1 và phương sai σ_1^2.
  - $Y_1, \ldots, Y_m$ là mẫu ngẫu nhiên được chọn từ tổng thể 2 có phân phối chuẩn với kỳ vọng μ_2 và phương sai σ_2^2.
  - Tổng thể 1 và 2 (đại diện bởi X và Y) độc lập với nhau.
  - Các phương sai σ_1^2 và σ_2^2 chưa biết.
  - Cỡ mẫu nhỏ : $n \leq 30$ và $m \leq 30$.

So sánh hai kỳ vọng, TH không biết phương sai, mẫu nhỏ

- Các giả định:
  - $X_1, \ldots, X_n$ là mẫu ngẫu nhiên được chọn từ tổng thể 1 có phân phối chuẩn với kỳ vọng μ_1 và phương sai σ_1^2.
  - $Y_1, \ldots, Y_m$ là mẫu ngẫu nhiên được chọn từ tổng thể 2 có phân phối chuẩn với kỳ vọng μ_2 và phương sai σ_2^2.
  - Tổng thể 1 và 2 (đại diện bởi X và Y) độc lập với nhau.
  - Các phương sai σ_1^2 và σ_2^2 chưa biết.
  - Cỡ mẫu nhỏ : $n \leq 30$ và $m \leq 30$.
- Ta xét hai trường hợp:

So sánh hai kỳ vọng, TH không biết phương sai, mẫu nhỏ

- Các giả định:
  - $X_1, \dots, X_n$ là mẫu ngẫu nhiên được chọn từ tổng thể 1 có phân phối chuẩn với kỳ vọng μ_1 và phương sai σ_1^2.
  - $Y_1, \dots, Y_m$ là mẫu ngẫu nhiên được chọn từ tổng thể 2 có phân phối chuẩn với kỳ vọng μ_2 và phương sai σ_2^2.
  - Tổng thể 1 và 2 (đại diện bởi X và Y) độc lập với nhau.
  - Các phương sai σ_1^2 và σ_2^2 chưa biết.
  - Cỡ mẫu nhỏ : $n \leq 30$ và $m \leq 30$.
- Ta xét hai trường hợp:
  1. Trường hợp phương sai bằng nhau $\sigma_1^2 = \sigma_2^2$,

So sánh hai kỳ vọng, TH không biết phương sai, mẫu nhỏ

- Các giả định:
  - $X_1, \ldots, X_n$ là mẫu ngẫu nhiên được chọn từ tổng thể 1 có phân phối chuẩn với kỳ vọng μ_1 và phương sai σ_1^2.
  - $Y_1, \ldots, Y_m$ là mẫu ngẫu nhiên được chọn từ tổng thể 2 có phân phối chuẩn với kỳ vọng μ_2 và phương sai σ_2^2.
  - Tổng thể 1 và 2 (đại diện bởi X và Y) độc lập với nhau.
  - Các phương sai σ_1^2 và σ_2^2 chưa biết.
  - Cỡ mẫu nhỏ : $n \leq 30$ và $m \leq 30$.
- Ta xét hai trường hợp:
  1. Trường hợp phương sai bằng nhau $\sigma_1^2 = \sigma_2^2$,
  2. Trường hợp phương sai khác nhau $\sigma_1^2 \neq \sigma_2^2$.

So sánh hai kỳ vọng, TH không biết phương sai, mẫu nhỏ

- Giả sử $X_1, \dots, X_n$ và $Y_1, \dots, Y_m$ lần lượt là hai mẫu ngẫu nhiên chọn từ hai tổng thể độc lập và có phân phối chuẩn với kỳ vọng và phương sai là (μ_1, σ_1^2) và (μ_2, σ_2^2). Ta cần kiểm định giả thuyết

$$\begin{cases} H_0: & \sigma_1^2 = \sigma_2^2 \\ H_1: & \sigma_1^2 \neq \sigma_2^2 \end{cases}$$

So sánh hai kỳ vọng, TH không biết phương sai, mẫu nhỏ

- Giả sử $X_1, \ldots, X_n$ và $Y_1, \ldots, Y_m$ lần lượt là hai mẫu ngẫu nhiên chọn từ hai tổng thể độc lập và có phân phối chuẩn với kỳ vọng và phương sai là (μ_1, σ_1^2) và (μ_2, σ_2^2). Ta cần kiểm định giả thuyết

$$\begin{cases} H_0: & \sigma_1^2 = \sigma_2^2 \\ H_1: & \sigma_1^2 \neq \sigma_2^2 \end{cases}$$

- Nếu S_1^2 là phương sai mẫu ngẫu nhiên $(X_1, \ldots, X_n)$ thì

$$\frac{(n-1)S_1^2}{\sigma_1^2} \sim \chi^2(n-1)$$

tương tự, ta có

$$\frac{(m-1)S_2^2}{\sigma_2^2} \sim \chi^2(m-1)$$

So sánh hai kỳ vọng, TH không biết phương sai, mẫu nhỏ

- Khi đó, đại lượng
$$F = \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2}$$
sẽ có phân phối $\mathfrak{F}$ với $(n-1, m-1)$ bậc tự do.

So sánh hai kỳ vọng, TH không biết phương sai, mẫu nhỏ

- Khi đó, đại lượng
$$F = \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2}$$
sẽ có phân phối $\mathfrak{F}$ với $(n-1, m-1)$ bậc tự do.
- Xét biến ngẫu nhiên $F \sim \mathfrak{F}(u, v)$ có hàm mật độ xác suất là $f(x)$, phân vị trên mức α của F là $f_{\alpha,u,v)}$ được định nghĩa như sau
$$\mathbb{P}(F > f_{\alpha,u,v}) = \int_{f_{\alpha,u,v}}^{\infty} f(x)dx = \alpha$$

So sánh hai kỳ vọng, TH không biết phương sai, mẫu nhỏ

- Khi đó, đại lượng
$$F = \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2}$$
sẽ có phân phối $\mathfrak{F}$ với $(n-1, m-1)$ bậc tự do.
- Xét biến ngẫu nhiên $F \sim \mathfrak{F}(u, v)$ có hàm mật độ xác suất là $f(x)$, phân vị trên mức α của F là $f_{\alpha,u,v)}$ được định nghĩa như sau
$$\mathbb{P}(F > f_{\alpha,u,v}) = \int_{f_{\alpha,u,v}}^{\infty} f(x)dx = \alpha$$
- Phân vị mức $1-\alpha$ của F cho bởi
$$f_{1-\alpha,u,v} = \frac{1}{f_{\alpha,u,v}}$$

So sánh hai kỳ vọng, TH không biết phương sai, mẫu nhỏ

Các bước kiểm định

1. Phát biểu giả thuyết $H_0 : \sigma_1^2 = \sigma_2^2$ và đối thuyết $H_1 : \sigma_1^2 \neq \sigma_2^2$

So sánh hai kỳ vọng, TH không biết phương sai, mẫu nhỏ

Các bước kiểm định

1. Phát biểu giả thuyết $H_0 : \sigma_1^2 = \sigma_2^2$ và đối thuyết $H_1 : \sigma_1^2 \neq \sigma_2^2$
2. Xác định mức ý nghĩa α

So sánh hai kỳ vọng, TH không biết phương sai, mẫu nhỏ

Các bước kiểm định

1. Phát biểu giả thuyết $H_0 : \sigma_1^2 = \sigma_2^2$ và đối thuyết $H_1 : \sigma_1^2 \neq \sigma_2^2$
2. Xác định mức ý nghĩa α
3. Khi H_0 đúng, thống kê

$$f = \frac{S_1^2}{S_2^2}$$

có phân phối $\mathfrak{F}$ với $(n-1, m-1)$ bậc tự do.

So sánh hai kỳ vọng, TH không biết phương sai, mẫu nhỏ

Các bước kiểm định

1. Phát biểu giả thuyết $H_0 : \sigma_1^2 = \sigma_2^2$ và đối thuyết $H_1 : \sigma_1^2 \neq \sigma_2^2$
2. Xác định mức ý nghĩa α
3. Khi H_0 đúng, thống kê
$$f = \frac{S_1^2}{S_2^2}$$
có phân phối $\mathfrak{F}$ với $(n-1, m-1)$ bậc tự do.
4. Xác định miền bác bỏ: Bác bỏ H_0 khi $|f| > f_{1-\alpha/2, n-1, m-1}$

So sánh hai kỳ vọng, TH không biết phương sai, mẫu nhỏ

Các bước kiểm định

1. Phát biểu giả thuyết $H_0 : \sigma_1^2 = \sigma_2^2$ và đối thuyết $H_1 : \sigma_1^2 \neq \sigma_2^2$
2. Xác định mức ý nghĩa α
3. Khi H_0 đúng, thống kê
$$f = \frac{S_1^2}{S_2^2}$$
có phân phối $\mathfrak{F}$ với $(n-1, m-1)$ bậc tự do.
4. Xác định miền bác bỏ: Bác bỏ H_0 khi $|f| > f_{1-\alpha/2, n-1, m-1}$
5. Kết luận: Nếu bác bỏ H_0, ta kết luận H_1 đúng với $100(1-\alpha)\%$ độ tin cậy. Ngược lại kết luận chưa đủ cơ sở để bác bỏ H_0.

So sánh hai kỳ vọng, TH không biết phương sai, mẫu nhỏ, TH $\sigma_1^2 = \sigma_2^2 = \sigma^2$

- Ta sử dụng một ước lượng chung cho cả σ_1^2 và σ_2^2 là S_p^2 phương sai mẫu chung

$$S_p^2 = \frac{(n-1)S_1^2 + (m-1)S_2^2}{n+m-2}$$

So sánh hai kỳ vọng, TH không biết phương sai, mẫu nhỏ, TH $\sigma_1^2 = \sigma_2^2 = \sigma^2$

- Ta sử dụng một ước lượng chung cho cả σ_1^2 và σ_2^2 là S_p^2 phương sai mẫu chung

$$S_p^2 = \frac{(n-1)S_1^2 + (m-1)S_2^2}{n+m-2}$$

- Thống kê

$$T_0 = \frac{\bar{X} - \bar{Y} - D_0}{S_p\sqrt{\frac{1}{n} + \frac{1}{m}}}$$

có phân phối Student với $n+m-2$ bậc tự do.

So sánh hai kỳ vọng, TH không biết phương sai, mẫu nhỏ, TH $\sigma_1^2 = \sigma_2^2 = \sigma^2$

- Đặt $df = n + m - 2$, miền bác bỏ và $p-$ giá trị trong trường hợp này có dạng

Đối thuyết	**Miền bác bỏ**	$p-$**giá trị**
$H_1 : \mu_1 - \mu_2 \neq D_0$	$\lvert t_0 \rvert > t_{1-\alpha/2}^{df}$	$p = 2\mathbb{P}(t^{df} \geq \lvert t_0 \rvert)$
$H_1 : \mu_1 - \mu_2 < D_0$	$t_0 < -t_{1-\alpha}^{df}$	$p = \mathbb{P}(t^{df} \leq t_0)$
$H_1 : \mu_1 - \mu_2 > D_0$	$t_0 > t_{1-\alpha}^{df}$	$p = \mathbb{P}(t^{df} \geq t_0)$

So sánh hai kỳ vọng, TH không biết phương sai, mẫu nhỏ, TH $\sigma_1^2 = \sigma_2^2 = \sigma^2$

- Đặt $df = n + m - 2$, miền bác bỏ và $p-$ giá trị trong trường hợp này có dạng

Đối thuyết	**Miền bác bỏ**	$p-$**giá trị**
$H_1 : \mu_1 - \mu_2 \neq D_0$	$\|t_0\| > t^{df}_{1-\alpha/2}$	$p = 2\mathbb{P}(t^{df} \geq \|t_0\|)$
$H_1 : \mu_1 - \mu_2 < D_0$	$t_0 < -t^{df}_{1-\alpha}$	$p = \mathbb{P}(t^{df} \leq t_0)$
$H_1 : \mu_1 - \mu_2 > D_0$	$t_0 > t^{df}_{1-\alpha}$	$p = \mathbb{P}(t^{df} \geq t_0)$

- Kết luận: Bác bỏ H_0/ Chưa đủ cơ sở để bác bỏ H_0.

So sánh hai kỳ vọng, TH không biết phương sai, mẫu nhỏ,TH $\sigma_1^2 \neq \sigma_2^2$

- Ta sử dụng thống kê

$$T_0 = \frac{\bar{X} - \bar{Y} - D_0}{\sqrt{\frac{S_1^2}{n} + \frac{S_2^2}{m}}}$$

khi đó t_0 có phân phối Student với bậc tự do df được xác định như sau

$$df = \frac{[(s_1^2/n) + (s_2^2/m)]^2}{\frac{(s_1^2/n)^2}{n-1} + \frac{(s_2^2/m)^2}{m-1}} \tag{9}$$

So sánh hai kỳ vọng, TH không biết phương sai, mẫu nhỏ,TH $\sigma_1^2 \neq \sigma_2^2$

- Ta sử dụng thống kê

$$T_0 = \frac{\bar{X} - \bar{Y} - D_0}{\sqrt{\frac{S_1^2}{n} + \frac{S_2^2}{m}}}$$

khi đó t_0 có phân phối Student với bậc tự do df được xác định như sau

$$df = \frac{[(s_1^2/n) + (s_2^2/m)]^2}{\frac{(s_1^2/n)^2}{n-1} + \frac{(s_2^2/m)^2}{m-1}} \tag{9}$$

- Miền bác bỏ trong trường hợp này giống như trong trường hợp phương sai bằng nhau, chỉ thay bậc tự do df cho bởi phương trình (**??**)

So sánh hai kỳ vọng, TH không biết phương sai, mẫu nhỏ

Ví dụ 9:

Tại một thành phố, khu vực A, người ta chọn ngẫu nhiên 17 sinh viên và cho làm bài kiểm tra để đo chỉ số IQs, thu được trung bình mẫu là 106 và độ lệch tiêu chuẩn bằng 10; tại khu vực B, chỉ số IQs trung bình của một mẫu gồm 14 sinh viên bằng 109 với độ lệch tiêu chuẩn là 7. Giả sử phương sai bằng nhau. Có sự khác biệt về chỉ số IQs của sinh viên ở hai khu vực A và B hay không? $\alpha = 0.02$.

So sánh hai kỳ vọng, TH không biết phương sai, mẫu nhỏ

Bài giải 11

Gọi X là chỉ số IQs của sinh viên khu vực A, $X \sim \mathcal{N}(\mu_1, \sigma_1^2)$. *Y là chỉ số IQs của sinh viên khu vực B,* $Y \sim \mathcal{N}(\mu_2, \sigma_2^2)$, $n = 17 < 30$, $m = 14 < 30$.

- $H_0 : \mu_1 - \mu_2 = 0 \quad H_1 : \mu_1 - \mu_2 \neq 0$

So sánh hai kỳ vọng, TH không biết phương sai, mẫu nhỏ

Bài giải 11

Gọi X là chỉ số IQs của sinh viên khu vực A, $X \sim \mathcal{N}(\mu_1, \sigma_1^2)$. *Y là chỉ số IQs của sinh viên khu vực B,* $Y \sim \mathcal{N}(\mu_2, \sigma_2^2)$, $n = 17 < 30$, $m = 14 < 30$.

- $H_0 : \mu_1 - \mu_2 = 0 \quad H_1 : \mu_1 - \mu_2 \neq 0$
- *Dưới giả định* H_0 *đúng, thống kê kiểm định*

$$T_0 = \frac{\bar{X} - \bar{Y} - D_0}{S_p\sqrt{\frac{1}{n} + \frac{1}{m}}} \sim T(29)$$

So sánh hai kỳ vọng, TH không biết phương sai, mẫu nhỏ

Bài giải 11

Gọi X là chỉ số IQs của sinh viên khu vực A, $X \sim \mathcal{N}(\mu_1, \sigma_1^2)$. *Y là chỉ số IQs của sinh viên khu vực B,* $Y \sim \mathcal{N}(\mu_2, \sigma_2^2)$, $n = 17 < 30, m = 14 < 30$.

- $H_0 : \mu_1 - \mu_2 = 0 \quad H_1 : \mu_1 - \mu_2 \neq 0$
- *Dưới giả định* H_0 *đúng, thống kê kiểm định*

$$T_0 = \frac{\bar{X} - \bar{Y} - D_0}{S_p\sqrt{\frac{1}{n} + \frac{1}{m}}} \sim T(29)$$

- *Ta có* $\bar{x} = 106, \bar{y} = 109, s_1 = 10, s_2 = 7, D_0 = 0,$ *suy ra*

$$s_p = \sqrt{\frac{(n-1)s_1^2 + (m-1)s_2^2}{n+m-2}} = 8.783$$

So sánh hai kỳ vọng, TH không biết phương sai, mẫu nhỏ

$t_0 = \frac{106-109-0}{8.783\sqrt{\frac{1}{17}+\frac{1}{14}}} = -0.95$

- $\alpha = 0.02$, miền bác bỏ $W_\alpha = \{t_0 : |t_0| > t^{29}_{1-\alpha/2} = t^{29}_{0.99} = 2.4620\}$
  $p_{value} = 2\left(1-(P)(T^{29} \leq |-0.95|\right) \Rightarrow 2\times(1-0.90) < p_{value} < 2\times(1-0.75) \Rightarrow$

So sánh hai kỳ vọng, TH không biết phương sai, mẫu nhỏ

$t_0 = \frac{106-109-0}{8.783\sqrt{\frac{1}{17}+\frac{1}{14}}} = -0.95$

- $\alpha = 0.02$, miền bác bỏ $W_\alpha = \{t_0 : |t_0| > t^{29}_{1-\alpha/2} = t^{29}_{0.99} = 2.4620\}$
  $p_{value} = 2\left(1 - (P)(T^{29} \leq |-0.95|\right) \Rightarrow 2 \times (1 - 0.90) < p_{value} < 2 \times (1 - 0.75) \Rightarrow$
- Ta có $|t_0| = 0.95 < t^{29}_{0.99} = 2.4620 \rightarrow t_0 \notin W_\alpha$. (Hoặc $p_{value} \in (0.2; 0.5) \Rightarrow p_{value} > 0.02$). Chưa đủ cơ sở bác bỏ H_0, với $\alpha = 0.02$. Với độ tin cậy 98%, có thể xem như chỉ số IQs của sinh viên hai khu vực là như nhau.

So sánh hai kỳ vọng, TH không biết phương sai, mẫu nhỏ

Ví dụ 10:

Hàm lượng thạch tín (Đv:ppb) trong nước càng cao càng có hại cho sức khỏe. Người ta kiểm tra hàm lượng thạch tín ở hai khu vực trung tâm thành phố Biên Hòa và khu vực gần sân bay Biên Hòa. Tại mỗi khu vực, người ta đo ngẫu nhiên hàm lượng thạch tín trong nước ứng với 10 điểm khác nhau. Số liệu cho bởi bảng thống kê bên dưới

Trung tâm TP	3	7	25	10	15	6	12	25	15	7
KV gần sân bay	48	44	40	38	33	21	20	12	1	18

Với $\alpha = 0.05$, hãy kiểm tra xem có sự khác biệt về hàm lượng thạch tín ở hai khu vực này. Giả sử phương sai khác nhau.

So sánh hai kỳ vọng, TH không biết phương sai, mẫu nhỏ

Bài giải 12

Gọi X là hàm lượng thạch tín ở khu vực trung tâm thành phố, $X \sim \mathcal{N}(\mu_1, \sigma_1^2)$. *Y là hàm lượng thạch tín ở khu vực gần sân bay,* $Y \sim \mathcal{N}(\mu_2, \sigma_2^2)$, $n = m = 10 < 30, \sigma_1 \neq \sigma_2$

- $H_0 : \mu_1 - \mu_2 = 0 \quad H_1 : \mu_1 - \mu_2 \neq 0$

So sánh hai kỳ vọng, TH không biết phương sai, mẫu nhỏ

Bài giải 12

Gọi X là hàm lượng thạch tín ở khu vực trung tâm thành phố, $X \sim \mathcal{N}(\mu_1, \sigma_1^2)$. *Y là hàm lượng thạch tín ở khu vực gần sân bay,* $Y \sim \mathcal{N}(\mu_2, \sigma_2^2)$, $n = m = 10 < 30, \sigma_1 \neq \sigma_2$

- $H_0 : \mu_1 - \mu_2 = 0 \quad H_1 : \mu_1 - \mu_2 \neq 0$
- *Dưới giả định* H_0 *đúng, thống kê kiểm định*

$$T_0 = \frac{\bar{X} - \bar{Y} - D_0}{\sqrt{\frac{S_1^2}{n} + \frac{S_2^2}{m}}} \sim T(df)$$

So sánh hai kỳ vọng, TH không biết phương sai, mẫu nhỏ

Bài giải 12

Gọi X là hàm lượng thạch tín ở khu vực trung tâm thành phố, $X \sim \mathcal{N}(\mu_1, \sigma_1^2)$. Y là hàm lượng thạch tín ở khu vực gần sân bay, $Y \sim \mathcal{N}(\mu_2, \sigma_2^2)$,$n = m = 10 < 30, \sigma_1 \neq \sigma_2$

- $H_0 : \mu_1 - \mu_2 = 0 \quad H_1 : \mu_1 - \mu_2 \neq 0$
- *Dưới giả định H_0 đúng, thống kê kiểm định*

$$T_0 = \frac{\bar{X} - \bar{Y} - D_0}{\sqrt{\frac{S_1^2}{n} + \frac{S_2^2}{m}}} \sim T(df)$$

- *Ta có* $\bar{x} = 12.5, \bar{y} = 27.5, s_1 = 7.634, s_2 = 15.35, D_0 = 0,$

$$df = \frac{(7.634^2/10 + 15.35^2/10)^2}{\frac{(7.634^2/10)^2}{9} + \frac{(15.35^2/10)^2}{9}} \simeq 13.201$$

Vậy $T_0 \sim T(13)$

So sánh hai kỳ vọng, TH không biết phương sai, mẫu nhỏ

$t_0 = -2.77$

- $\alpha = 0.05$, miền bác bỏ $W_\alpha = \{t_0 : |t_0| > t^{13}_{1-\alpha/2} = t^{13}_{0.975} = 2.1604\}$
  $p_{value} = 2\left(1 - (P)(T^{13} \leq |-2.77|\right) \Rightarrow 2 \times (1 - 0.995) < p_{value} < 2 \times (1 - 0.99) \Rightarrow p_{value} \in (0.01; 0.02)$

So sánh hai kỳ vọng, TH không biết phương sai, mẫu nhỏ

$t_0 = -2.77$

- $\alpha = 0.05$, miền bác bỏ $W_\alpha = \{t_0 : |t_0| > t^{13}_{1-\alpha/2} = t^{13}_{0.975} = 2.1604\}$
  $p_{value} = 2\left(1 - (P)(T^{13} \leq |-2.77|\right) \Rightarrow 2 \times (1 - 0.995) < p_{value} < 2 \times (1 - 0.99) \Rightarrow p_{value} \in (0.01; 0.02)$
- Ta có $|t_0| = 2.77 > t^{29}_{0.975} = 2.1604 \rightarrow t_0 \in W_\alpha$. (Hoặc $p_{value} \in (0.01; 0.02) \Rightarrow p_{value} < 0.05$) Bác bỏ H_0, chấp nhận H_1 với $\alpha = 0.05$. Với độ tin cậy 95%, hàm lượng thạch tín ở hai khu vực có sự khác biệt.

Nội dung

So sánh hai kỳ vọng hai mẫu không độc lập (paired $t-$test)

- Khi hai mẫu không độc lập thì mỗi giá trị quan trắc được trong mẫu một có mối liên hệ tương ứng với một giá trị quan trắc ở mẫu thứ hai. Như vậy, ta có thể ghép cặp tương ứng giá trị trong hai mẫu với nhau.

So sánh hai kỳ vọng hai mẫu không độc lập (paired $t-$test)

- Khi hai mẫu không độc lập thì mỗi giá trị quan trắc được trong mẫu một có mối liên hệ tương ứng với một giá trị quan trắc ở mẫu thứ hai. Như vậy, ta có thể ghép cặp tương ứng giá trị trong hai mẫu với nhau.
- Việc ghép cặp là kết quả của việc:

So sánh hai kỳ vọng hai mẫu không độc lập (paired $t-$test)

- Khi hai mẫu không độc lập thì mỗi giá trị quan trắc được trong mẫu một có mối liên hệ tương ứng với một giá trị quan trắc ở mẫu thứ hai. Như vậy, ta có thể ghép cặp tương ứng giá trị trong hai mẫu với nhau.
- Việc ghép cặp là kết quả của việc:
    - quan trắc giá trị trước và sau khi thực hiện 1 thí nghiệm. Chẳng hạn như đo trọng lượng trước và sau khi thực hiện một chế độ ăn kiêng.

So sánh hai kỳ vọng hai mẫu không độc lập (paired $t-$test)

- Khi hai mẫu không độc lập thì mỗi giá trị quan trắc được trong mẫu một có mối liên hệ tương ứng với một giá trị quan trắc ở mẫu thứ hai. Như vậy, ta có thể ghép cặp tương ứng giá trị trong hai mẫu với nhau.
- Việc ghép cặp là kết quả của việc:
  - quan trắc giá trị trước và sau khi thực hiện 1 thí nghiệm. Chẳng hạn như đo trọng lượng trước và sau khi thực hiện một chế độ ăn kiêng.
  - so sánh cùng một đặc tính.

So sánh hai kỳ vọng hai mẫu không độc lập (paired $t-$test)

- Khi hai mẫu không độc lập thì mỗi giá trị quan trắc được trong mẫu một có mối liên hệ tương ứng với một giá trị quan trắc ở mẫu thứ hai. Như vậy, ta có thể ghép cặp tương ứng giá trị trong hai mẫu với nhau.
- Việc ghép cặp là kết quả của việc:
  - quan trắc giá trị trước và sau khi thực hiện 1 thí nghiệm. Chẳng hạn như đo trọng lượng trước và sau khi thực hiện một chế độ ăn kiêng.
  - so sánh cùng một đặc tính.
  - thí nghiệm trên cùng 1 địa điểm.

So sánh hai kỳ vọng hai mẫu không độc lập (paired t−test)

- Khi hai mẫu không độc lập thì mỗi giá trị quan trắc được trong mẫu một có mối liên hệ tương ứng với một giá trị quan trắc ở mẫu thứ hai. Như vậy, ta có thể ghép cặp tương ứng giá trị trong hai mẫu với nhau.
- Việc ghép cặp là kết quả của việc:
  - quan trắc giá trị trước và sau khi thực hiện 1 thí nghiệm. Chẳng hạn như đo trọng lượng trước và sau khi thực hiện một chế độ ăn kiêng.
  - so sánh cùng một đặc tính.
  - thí nghiệm trên cùng 1 địa điểm.
  - thí nghiệm với cùng thời gian.

So sánh hai kỳ vọng hai mẫu không độc lập (paired $t-$test)

- Xét (X_{1i}, X_{2i}), với $i = 1, 2, \ldots, n$ là tập gồm n cặp gía trị quan trắc với giả sử rằng kỳ vọng của hai tổng thể đại diện bởi X_1 và X_2 lần lượt là μ_1 và σ_1^2; μ_2 và σ_2^2.

So sánh hai kỳ vọng hai mẫu không độc lập (paired $t-$test)

- Xét (X_{1i}, X_{2i}), với $i = 1, 2, \ldots, n$ là tập gồm n cặp gía trị quan trắc với giả sử rằng kỳ vọng của hai tổng thể đại diện bởi X_1 và X_2 lần lượt là μ_1 và σ_1^2; μ_2 và σ_2^2.
- Định nghĩa độ sai khác giữa mỗi cặp trong tập hợp các giá trị quan trắc là

$$D_i = X_{1i} - X_{2i},\ i = 1, 2, \ldots, n$$

So sánh hai kỳ vọng hai mẫu không độc lập (paired $t-$test)

- Xét (X_{1i}, X_{2i}), với $i = 1, 2, \ldots, n$ là tập gồm n cặp gía trị quan trắc với giả sử rằng kỳ vọng của hai tổng thể đại diện bởi X_1 và X_2 lần lượt là μ_1 và σ_1^2; μ_2 và σ_2^2.
- Định nghĩa độ sai khác giữa mỗi cặp trong tập hợp các giá trị quan trắc là

$$D_i = X_{1i} - X_{2i},\ i = 1, 2, \ldots, n$$

- Các $D_i,\ i = 1, 2, \ldots, n$ được giả sử có phân phối chuẩn.

So sánh hai mẫu không độc lập (paired t−test)

- Gọi $\mu_D = E(D_i)$, bởi vì $D_1, \ldots, D_n$ là những biến ngẫu nhiên độc lập và có cùng phân phối, nếu $d_1, \ldots, d_n$ là những giá trị của $D_1, \ldots, D_n$, ta định nghĩa

$$
\begin{aligned}
\bar{d} &= \frac{1}{n}\sum_{i=1}^{n} d_i \\
s_d^2 &= \frac{1}{n-1}\sum_{i=1}^{n}(d_i - d)^2 = \frac{1}{n-1}\sum_{i=1}^{n} d_i^2 - \frac{n}{n-1}(\bar{d})^2
\end{aligned}
$$

So sánh hai mẫu không độc lập (paired $t-$test)

- Gọi $\mu_D = E(D_i)$, bởi vì $D_1, \ldots, D_n$ là những biến ngẫu nhiên độc lập và có cùng phân phối, nếu $d_1, \ldots, d_n$ là những giá trị của $D_1, \ldots, D_n$, ta định nghĩa

$$\begin{aligned} \bar{d} &= \frac{1}{n}\sum_{i=1}^{n} d_i \\ s_d^2 &= \frac{1}{n-1}\sum_{i=1}^{n}(d_i - d)^2 = \frac{1}{n-1}\sum_{i=1}^{n} d_i^2 - \frac{n}{n-1}(\bar{d})^2 \end{aligned}$$

- Ta cần kiểm định các giả thuyết và đối thuyết, với mức ý nghĩa α sau:

$$(a)\begin{cases} H_0 : \mu_D = D_0 \\ H_1 : \mu_D \neq D_0 \end{cases} \quad (b)\begin{cases} H_0 : \mu_D = D_0 \\ H_1 : \mu_D < D_0 \end{cases} \quad (c)\begin{cases} H_0 : \mu_D = D_0 \\ H_1 : \mu_D > D_0 \end{cases}$$

So sánh hai kỳ vọng hai mẫu không độc lập (paired $t-$test)

Các bước kiểm định

- TH mẫu nhỏ
  1. Phát biểu giả thuyết không và đối thuyết,
  2. Thống kê kiểm định
$$T_0 = \frac{\bar{D} - D_0}{S_d/\sqrt{n}} \sim T(n-1)$$
  3. Tính thống kê kiểm định
$$t_0 = \frac{\bar{d} - D_0}{s_d/\sqrt{n}} \tag{10}$$

So sánh hai kỳ vọng hai mẫu không độc lập (paired t−test)

So sánh hai kỳ vọng hai mẫu không độc lập (paired $t-$test)

4. Xác định mức ý nghĩa α,, xác định miền bác bỏ / tính p-giá trị

Đối thuyết	**Miền bác bỏ**	**$p-$giá trị**
$H_1 : \mu_D \neq D_0$	$\lvert t_0\rvert > t^{n-1}_{1-\alpha/2}$	$p = 2\mathbb{P}(t^{n-1} \geq \lvert t_0\rvert)$
$H_1 : \mu_D < D_0$	$t_0 < -t^{n-1}_{1-\alpha}$	$p = \mathbb{P}(t^{n-1} \leq t_0)$
$H_1 : \mu_D > D_0$	$t_0 > t^{n-1}_{1-\alpha}$	$p = \mathbb{P}(t^{n-1} \geq t_0)$

So sánh hai kỳ vọng hai mẫu không độc lập (paired $t-$test)

4. Xác định mức ý nghĩa α,, xác định miền bác bỏ / tính p-giá trị

Đối thuyết	**Miền bác bỏ**	$p-$**giá trị**
$H_1 : \mu_D \neq D_0$	$\|t_0\| > t^{n-1}_{1-\alpha/2}$	$p = 2\mathbb{P}(t^{n-1} \geq \|t_0\|)$
$H_1 : \mu_D < D_0$	$t_0 < -t^{n-1}_{1-\alpha}$	$p = \mathbb{P}(t^{n-1} \leq t_0)$
$H_1 : \mu_D > D_0$	$t_0 > t^{n-1}_{1-\alpha}$	$p = \mathbb{P}(t^{n-1} \geq t_0)$

5. Kết luận: Nếu bác bỏ H_0, ta kết luận H_1 đúng với độ tin cậy $(1-\alpha)100\%$. Ngược lại ta kết luận chưa đủ cơ sở để bác bỏ H_0 với α cho trước.

So sánh hai kỳ vọng hai mẫu không độc lập (paired $t-$test)

4. Xác định mức ý nghĩa α,, xác định miền bác bỏ / tính p-giá trị

Đối thuyết	**Miền bác bỏ**	$p-$**giá trị**
$H_1 : \mu_D \neq D_0$	$\|t_0\| > t^{n-1}_{1-\alpha/2}$	$p = 2\mathbb{P}(t^{n-1} \geq \|t_0\|)$
$H_1 : \mu_D < D_0$	$t_0 < -t^{n-1}_{1-\alpha}$	$p = \mathbb{P}(t^{n-1} \leq t_0)$
$H_1 : \mu_D > D_0$	$t_0 > t^{n-1}_{1-\alpha}$	$p = \mathbb{P}(t^{n-1} \geq t_0)$

5. Kết luận: Nếu bác bỏ H_0, ta kết luận H_1 đúng với độ tin cậy $(1-\alpha)100\%$. Ngược lại ta kết luận chưa đủ cơ sở để bác bỏ H_0 với α cho trước.

- TH cỡ mẫu lớn ($n > 30$), bài toán kiểm định hai mẫu phụ thuộc thực hiện tương tự như trong trường hợp một mẫu ngẫu nhiên $(D_1, \ldots, D_n)$.

So sánh hai kỳ vọng hai mẫu không độc lập (paired $t-$test)

Ví dụ 11:

Một bác sĩ dinh dưỡng nghiên cứu một chế độ ăn kiêng và tập thể dục mới để làm giảm lượng đường trong máu của các bệnh nhân bị bệnh tiểu đường. 10 bệnh nhân bị bệnh tiểu đường được chọn để thử nghiệm chương trình này, bảng kết quả bên dưới cho biết lượng đường trong máu trước và sau khi các bệnh nhân tham gia chương trình

Trước	268	225	252	192	307	228	246	298	231	185
Sau	106	186	223	110	203	101	211	176	194	203

Số liệu được cung cấp có đủ bằng chứng để kết luận rằng chế độ ăn kiêng và tập thể dục làm giảm lượng đường trong máu không ? $\alpha = 0.05$.

Nội dung

Kiểm định giả thuyết về phân phối

- **Bài toán:** Khảo sát biến ngẫu nhiên X liên quan đến một tổng thể có phân phối chưa biết. Cần kiểm định xem phân phối của tổng thể có phải là $F(x, \theta)$ hay không? Ví dụ, ta cần kiểm định phân phối của tổng thể đang xét là phân phối chuẩn.

Kiểm định giả thuyết về phân phối

Các bước kiểm định

1. Chọn mẫu ngẫu nhiên cỡ $n : (X_1, \ldots, X_n)$. Chia miền giá trị của các biến ngẫu nhiên X_i thành k khoảng không trùng nhau $I_1, I_2, \ldots, I_k$. (TH X là biến ngẫu nhiên rời rạc, ta chia thành k điểm :$x_1, x_2, \ldots, x_k$).
2. Gọi O_j là số các giá trị mẫu nằm trong khoảng I_j, $j = 1, 2, , kc$
3. Phát biểu giả thuyết $H_0 : X$ tuân theo luật phân phối $F(x, \theta)$. Khi đó, tính $p_j = \mathbb{P}(X \in I_j)$ (hoặc $\mathbb{P}(X = x_j)$ nếu X rời rạc). Đặt $E_j = np_j$, E_j gọi là tần số lý thuyết. Điều kiện: $E_j \geq 5, j = 1, 2, \ldots k$.

Kiểm định giả thuyết về phân phối

4. Thống kê kiểm định Q^2 cho bởi công thức

$$Q^2 = \sum_{i=1}^{k} \frac{(O_j - E_j)^2}{E_j}$$

Q^2 xấp xỉ phân phối χ^2 với $k-1$ bậc tự do.

5. Bác bỏ H_0 nếu

$$Q^2 \geq \chi^2_{1-\alpha,k-r-1}$$

với r là số tham số ước lượng. Tìm $\chi^2_{1-\alpha,k-r-1}$:tra bảng Chi - bình phương.

Kiểm định giả thuyết về phân phối

Ví dụ 12:

Bảng thống kê số vụ tai nạn xe máy/ ngày ở một quận của thành phố trong 80 ngày

Số vụ tai nạn	Số ngày
0	34
1	25
2	11
3	7
4	3

Với mức ý nghĩa 5%, hãy kiểm tra xem số vụ tai nạn xe máy hàng ngày có tuân theo luật phân phối Poisson hay không?

Kiểm định giả thuyết về phân phối

Ví dụ 13:

Điểm thi của 200 sinh viên trong một lớp học cho bởi bảng bên dưới. Có ý kiến cho rằng điểm thi của sinh viên là đại lượng ngẫu nhiên có phân phối chuẩn với điểm trung bình bằng 75 và độ lệch chuẩn bằng 8. Với $\alpha = 0.05$, hãy kiểm tra ý kiến này.

Điểm thi	(0,60]	(60,70]	(70,80]	(80,90]	(90,100]
Số sinh viên	12	36	90	44	18

Kiểm định giả thuyết về phân phối

Ví dụ 14:

Nhóm máu của 500 người chọn ngẫu nhiên từ một khu vực cho bởi bảng sau:

A	B	AB	O
75	150	15	260

Theo từ điển y khoa thì tỷ lệ nhóm máu trong dân số là $0.18, 0.28, 0.05, 0.49$. Hỏi nhóm máu trong dân số có phù hợp với từ điển y khoa hay không? mức ý nghĩa 1%.

Kiểm định giả thuyết về tính độc lập

- Bài toán:
  - Giả sử mỗi phần tử trong một tổng thể có thể được phân loại theo hai đặc tính khác nhau, gọi là đặc tính X và đặc tính Y. X có r giá trị và Y có s giá trị. Gọi
  $$P_{ij} = \mathbb{P}(X = x_i, Y = y_j)$$
  với $i = 1, \ldots, r$ và $j = 1, \ldots, s$. p_{ij} là các xác suất chọn được một phần tử trong tổng thể có đặc tính X bằng i và đặc tính Y bằng j.
  - Gọi
  $$p_i = \mathbb{P}(X = x_i) = \sum_{j=1}^{s} P_{ij},\ i = 1, \ldots, r$$
  và
  $$q_j = \mathbb{P}(Y = y_j) = \sum_{i=1}^{r} P_{ij},\ j = 1, \ldots, s$$

Kiểm định giả thuyết về tính độc lập

p_i là xác suất chọn được một phần tử của tổng thể có đặc tính X bằng x_i, q_j là xác suất chọn được một phần tử của tổng thể có đặc tính Y bằng y_j,

- Ta cần kiểm định xem X có độc lập với Y hay không?
Phát biểu gải thuyết

$$H_0 : P_{ij} = p_i q_j \ \forall i = 1, \ldots, r; j = 1, \ldots, s$$

và đối thuyết

$$H_1 : \exists (i, j) \text{ sao cho } P_{ij} \neq p_i q_j$$

Kiểm định giả thuyết về tính độc lập

- Khảo sát N phần tử, ta được bảng kết quả, trong bài toán này gọi là bảng ngẫu nhiên.

	y_1	y_2	$\cdots$	y_s	Tổng hàng
x_1	n_{11}	n_{12}	$\cdots$	n_{1s}	n_1
x_2	n_{21}	n_{22}	$\cdots$	n_{2s}	n_2
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
x_r	n_{r1}	n_{r2}	$\cdots$	n_{rs}	n_r
Tổng cột	m_1	m_2	$\cdots$	m_s	N

Kiểm định giả thuyết về tính độc lập

- Ước lượng của p_i và q_j lần lượt bằng

$$\hat{p}_i = \frac{n_i}{N}, \quad i = 1, \dots, r$$

$$\hat{q}_j = \frac{m_j}{N}, \quad j = 1, \dots, s$$

- Gọi N_{ij} là số phần tử có đặc tính (x_i, y_j) trong N phần tử khảo sát, thì $N_{ij} \sim B(N, P_{ij})$. Khi đó,

$$\mathbb{E}(N_{ij}) = NP_{ij} = Np_i q_j \text{ khi } H_0 \text{ đúng}$$

Đặt

$$e_{ij} = N\hat{p}_i\hat{q}_j = \frac{n_i m_j}{N}$$

e_{ij} gọi là tần số lý thuyết.

Kiểm định giả thuyết về tính độc lập

Định lý 1 (Pearson)

Với N_{ij} *và* $E_{ij} = NP_{ij}$, *biến ngẫu nhiên*

$$\sum_{i=1}^{r}\sum_{j=1}^{s}\frac{(N_{ij}-E_{ij})^2}{E_{ij}}$$

sẽ hội tụ theo phân phối về biến ngẫu nhiên Chi - bình phương $\chi^2_{(r-1)(s-1)}$ *bậc tự do.*

Kiểm định giả thuyết về tính độc lập

Các bước kiểm định

1. Phát biểu giả thuyết $H_0 : X$ và Y độc lập
2. Xác định tần số thực nghiệm n_{ij} và tần số lý thuyết

$$e_{ij} = \frac{n_i m_j}{N}$$

với n_i và m_j là tổng hàng i và tổng cột j tương ứng.
Điều kiện: $e_{ij} \geq 5$.

Kiểm định giả thuyết về tính độc lập

3. Tính thống kê kiểm định

$$Q^2 = \sum_{i=1}^{r} \sum_{j=1}^{s} \frac{(n_{ij} - e_{ij})^2}{e_{ij}} = \sum_{i=1}^{r} \sum_{j=1}^{s} \frac{n_{ij}^2}{e_{ij}} - N$$

Nếu H_0 đúng, thống kê Q^2 có phân phối Chi bình phương với $(r-1)(s-1)$ bậc tự do.

4. Bác bỏ H_0 khi

$$Q^2 > \chi^2_{(r-1)(s-1)}(1-\alpha)$$

4b. Sử dụng $p-$giá trị:

$$p = \mathbb{P}(\chi^2_{(r-1)(s-1)} \geq Q^2$$

Bác bỏ H_0 khi: $p \leq \alpha$.

Kiểm định giả thuyết về tính độc lập

Ví dụ 15:

Một báo cáo khoa học trong y khoa tuyên bố rằng việc sở hữu một thú cưng trong nhà (chó hoặc mèo) sẽ làm tăng khả năng sống sót của những người chủ mà thường bị lên cơn đau tim. Mọt mẫu ngẫu nhiên gồm 95 người đã lên cơn đau tim được chọn để khảo sát. Dữ liệu của mỗi người khảo sát được chia làm 2 loại:

- ♣ Những người sống sót/tử vong 1 năm sau khi lên cơn đau tim.
- ♣ Người sống sót/tử vong có nuôi thú cưng trong nhà hay không.

Kết quả cho bởi bảng sau:

	Có nuôi thú cưng	Không nuôi thú cưng
Sống sót	28	44
Tử vong	8	15

Kiểm định giả thuyết về tính độc lập

1. Phát biểu giả thuyết H_0 : Bệnh lên cơn đau tim độc lập với việc nuôi thú cưng.
2. Tính tần số thực nghiệm: với $n_1 = 72, n_2 = 23, m_1 = 36, m_2 = 59$

$$e_{11} = \frac{n_1 m_1}{N} = \frac{72 \times 36}{95} = 27.284; \;\; e_{12} = \frac{n_1 m_2}{N} = \frac{72 \times 59}{95} = 44.716$$

$$e_{21} = \frac{n_2 m_1}{N} = \frac{23 \times 36}{95} = 8.716; \;\; e_{22} = \frac{n_2 m_2}{N} = \frac{23 \times 59}{95} = 14.284$$

3. Tính giá trị thống kê Q^2

$$Q^2 = \sum_{i=1}^{2} \sum_{j=1}^{2} \frac{n_{ij}^2}{e_{ij}} - n = \left(\frac{28^2}{27.284} + \frac{44^2}{44.716} + \frac{8^2}{8.716} + \frac{15^2}{15.284} \right) - 95 = 0.$$

Kiểm định giả thuyết về tính độc lập

4. Bác bỏ H_0 khi: $Q^2 > \chi^2_{(r-1)(s-1)}(1-\alpha) = \chi^2_1(1-0.05)$.
   Tra bảng Chi-bình phương, ta được $\chi^2_1(0.95) = 3.841$.
   $Q^2 = 0.125$, suy ra $Q^2 < 3.841$. Ta kết luận chưa đủ cơ sở để bác bỏ H_0 tức là bệnh lên cơn đau tim độc lập với việc nuôi thú cưng.

Kiểm định giả thuyết về tính độc lập

Ví dụ 16:

Vé máy bay của hãng hàng không Việt Nam Airline được chia làm 3 loại: Hạng thường (C), hạng trung (B) và hạng doanh nhân (A). Hành khách đi máy bay của VN Airlines nằm trong 1 trong 2 dạng sau: bay nội địa hoặc quốc tế. Khảo sát 920 hành khách đã bay của hãng, cho kết quả sau:

	Loại chuyến bay	
Loại vé	Nội địa	Quốc tế
Hạng thường	29	22
Hạng trung	95	121
Hạng doanh nhân	518	135

Có ý kiến cho rằng hành khách mua loại vé nào (A, B, C) sẽ phụ thuộc vào việc người đó bay nội địa hay quốc tế. Với mức ý nghĩa 5%, hãy kiểm tra ý kiến trên.