Chương 5
KIỂM ĐỊNH GIẢ THUYẾT THỐNG KÊ

BỘ MÔN TOÁN ỨNG DỤNG[1]

VIỆN TOÁN ỨNG DỤNG VÀ TIN HỌC
ĐẠI HỌC BÁCH KHOA HÀ NỘI

SAMI.HUST – 2023

[1] Phòng BIS.201-D3.5

5.2. KIỂM ĐỊNH GIẢ THUYẾT VỀ THAM SỐ CỦA TỔNG THỂ

Bài toán

Bài toán 1

Cho biến ngẫu nhiên gốc $X \sim \mathcal{N}(\mu;\sigma^2)$ với kỳ vọng $E(X) = \mu$ chưa biết nhưng có cơ sở để nêu lên giả thuyết

$$H_0 : \mu = \mu_0 \quad \text{với} \quad \mu_0 \text{ là số đã biết.}$$

Hãy kiểm định giả thuyết về kỳ vọng ở một trong ba dạng của cặp giả thuyết sau:

1. $H_0 : \mu = \mu_0;\ H_1 : \mu \neq \mu_0$
2. $H_0 : \mu = \mu_0;\ H_1 : \mu > \mu_0$
3. $H_0 : \mu = \mu_0;\ H_1 : \mu < \mu_0$

Phân phối mẫu

- Từ biến ngẫu nhiên gốc X, ta xây dựng một mẫu ngẫu nhiên $W_X = (X_1, X_2, \ldots, X_n)$ kích thước n. Khi đó, $\overline{X} \sim \mathcal{N}(\mu; \sigma^2/n)$ và
$$Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0; 1).$$
- Nếu giả thuyết $H_0 : \mu = \mu_0$ là đúng thì $E(\overline{X}) = \mu_0$ và
$$Z_0 = \frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}} \sim \mathcal{N}(0; 1).$$

Kiểm định giả thuyết về kỳ vọng, phương sai đã biết

Các bước tiến hành

1. Xác định dạng cụ thể của cặp giả thuyết cần kiểm định $\{H_0; H_1\}$.
2. Chọn tiêu chuẩn kiểm định

$$Z_0 = \frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}}. \tag{6}$$

   Nếu giả thuyết $H_0 : \mu = \mu_0$ là đúng thì $Z_0 \sim \mathcal{N}(0; 1)$.

3. Xây dựng miền bác bỏ giả thuyết H_0 phụ thuộc vào đối thuyết H_1.

- Kiểm định hai phía $H_0 : \mu = \mu_0$, $H_1 : \mu \neq \mu_0$. Vì $Z_0 \sim \mathcal{N}(0; 1)$, nên với mức ý nghĩa α cho trước, giả thuyết H_0 bị bác bỏ nếu, xem Hình 2(a),

$$P\Big(|Z_0| > z_{\alpha/2}\Big|(\mu = \mu_0)\Big) = \alpha,$$

  ở đây, $z_{\alpha/2}$ là giá trị tới hạn của phân phối chuẩn tắc mức $\alpha/2$. Do đó, miền bác bỏ giả thuyết H_0 là

$$W_\alpha = (-\infty; -z_{\alpha/2}) \cup (z_{\alpha/2}; +\infty).$$

Kiểm định giả thuyết về kỳ vọng, phương sai đã biết

Các bước tiến hành (tiếp theo)

- Kiểm định một phía phải $H_0 : \mu = \mu_0$, $H_1 : \mu > \mu_0$. Với mức ý nghĩa α cho trước, ta tìm giá trị tới hạn của phân phối chuẩn tắc z_α mức α sao cho, xem Hình 2(b),

$$P\left(Z_0 > z_\alpha \middle| (\mu = \mu_0)\right) = \alpha$$

  và xác định được miền bác bỏ giả thuyết H_0 là

$$W_\alpha = (z_\alpha; +\infty).$$

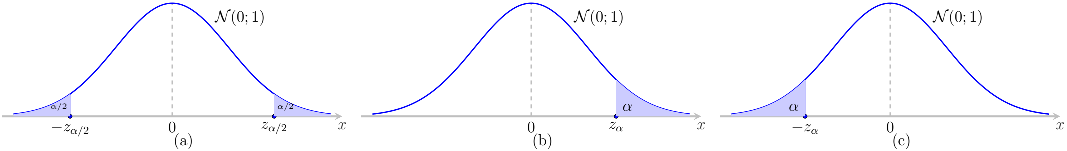
- Kiểm định một phía trái $H_0 : \mu = \mu_0$, $H_1 : \mu < \mu_0$. Với mức ý nghĩa α cho trước, ta tìm giá trị tới hạn của phân phối chuẩn tắc z_α mức α sao cho, xem Hình 2(c),

$$P\left(Z_0 < -z_\alpha \middle| (\mu = \mu_0)\right) = \alpha$$

  và xác định được miền bác bỏ giả thuyết H_0 là

$$W_\alpha = (-\infty; -z_\alpha).$$

Kiểm định giả thuyết về kỳ vọng, phương sai đã biết



Hình 2: Phân phối của Z_0 khi $H_0 : \mu = \mu_0$ là đúng và miền bác bỏ giả thuyết H_0 ứng với (a) $H_1 : \mu \neq \mu_0$, (b) $H_1 : \mu > \mu_0$, (c) $H_1 : \mu < \mu_0$

Kiểm định giả thuyết về kỳ vọng, phương sai đã biết

✍ Tóm lại, miền bác bỏ giả thuyết H_0 được xác định như sau:

H_0	H_1	Miền bác bỏ giả thuyết H_0 (W_α)
$\mu = \mu_0$	$\mu \neq \mu_0$	$(-\infty; -z_{\alpha/2}) \cup (z_{\alpha/2}; +\infty)$
$\mu = \mu_0$	$\mu > \mu_0$	$(z_\alpha; +\infty)$
$\mu = \mu_0$	$\mu < \mu_0$	$(-\infty; -z_\alpha)$

Kiểm định giả thuyết về kỳ vọng, phương sai đã biết

Các bước tiến hành (tiếp theo)

4. Lập mẫu cụ thể $W_x = (x_1, x_2, \ldots, x_n)$, tính giá trị quan sát của tiêu chuẩn kiểm định:

$$z_0 = \frac{\overline{x} - \mu_0}{\sigma/\sqrt{n}}. \tag{7}$$

5. Kiểm tra xem z_0 có thuộc W_α hay không để kết luận.

- Nếu $z_0 \in W_\alpha$ thì bác bỏ giả thuyết H_0.
- Nếu $z_0 \notin W_\alpha$ thì chưa có cơ sở để bác bỏ giả thuyết H_0.

Kiểm định giả thuyết về kỳ vọng, phương sai đã biết

Ví dụ 5

Một hãng bảo hiểm thông báo rằng số tiền trung bình hãng chi trả cho khách hàng bị tai nạn ô tô là 17 triệu đồng. Để kiểm tra lại, người ta điều tra ngẫu nhiên hồ sơ chi trả của 25 khách hàng thì thấy số tiền trung bình chi trả là 17,8 triệu đồng. Giả sử số tiền chi trả là biến ngẫu nhiên có phân phối chuẩn với độ lệch chuẩn là 5,2 triệu đồng. Hãy kiểm định lại thông báo của hãng bảo hiểm trên với mức ý nghĩa 5%.

Giải. Gọi X (triệu đồng) là số tiền hãng bảo hiểm chi trả cho khách hàng, $X \sim \mathcal{N}(\mu; \sigma^2)$ với $\sigma = 5,2$ triệu đồng. Số tiền trung bình hãng chi trả cho khách hàng là $E(X) = \mu$ triệu đồng chưa biết. Đây là bài toán kiểm định giả thuyết cho kỳ vọng của biến ngẫu nhiên có phân phối chuẩn trong trường hợp đã biết phương sai.

Kiểm định giả thuyết về kỳ vọng, phương sai đã biết

- Cặp giả thuyết cần kiểm định là $H_0 : \mu = \mu_0$, $H_1 : \mu \neq \mu_0$ với $\mu_0 = 17$.
- Chọn tiêu chuẩn kiểm định

$$Z_0 = \frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}}.$$

  Nếu giả thuyết H_0 là đúng thì $Z_0 \sim \mathcal{N}(0; 1)$.

- Với $\alpha = 0,05$, $z_{\alpha/2} = z_{0,025} = 1,96$. Miền bác bỏ giả thuyết H_0 là

$$W_\alpha = (-\infty; -1,96) \cup (1,96; +\infty).$$

- Tính giá trị quan sát của tiêu chuẩn kiểm định với $n = 25$, $\overline{x} = 17,8$ và $\sigma = 5,2$,

$$z_0 = \frac{\overline{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{17,8 - 17}{5,2/\sqrt{25}} \simeq 0,7692.$$

- Vì $z_0 = 0,7692 \notin W_\alpha$ nên chưa có cơ sở để bác bỏ giả thuyết H_0. Do đó, chưa có cơ sở để bác bỏ thông báo của hãng bảo hiểm với mức ý nghĩa 5%.

Kiểm định giả thuyết về kỳ vọng, phương sai đã biết

Ví dụ 6

Nếu máy móc hoạt động bình thường thì trọng lượng sản phẩm là biến ngẫu nhiên có phân phối chuẩn với trọng lượng trung bình là 100 gam, độ lệch chuẩn là 2 gam. Qua một thời gian sản xuất người ta nghi ngờ trọng lượng sản phẩm có xu hướng tăng lên, cân thử 100 sản phẩm thì trọng lượng trung bình của chúng là 100,4 gam. Với mức ý nghĩa $\alpha = 5\%$ hãy kết luận về điều nghi ngờ trên.

Giải. Gọi X (gam) là trọng lượng sản phẩm, $X \sim \mathcal{N}(\mu; \sigma^2)$ với $\sigma = 2$ gam. Đây là bài toán kiểm định giả thuyết cho kỳ vọng $E(X) = \mu$ của biến ngẫu nhiên có phân phối chuẩn trong trường hợp đã biết phương sai.

Kiểm định giả thuyết về kỳ vọng, phương sai đã biết

- Cặp giả thuyết cần kiểm đinh là $H_0 : \mu = \mu_0$, $H_1 : \mu > \mu_0$ với $\mu_0 = 100$.
- Chọn tiêu chuẩn kiểm định
$$Z_0 = \frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}}.$$
Nếu giả thuyết H_0 là đúng thì $Z_0 \sim \mathcal{N}(0; 1)$.
- Với $\alpha = 0,05$, $z_\alpha = z_{0,05} = 1,645$. Miền bác bỏ giả thuyết H_0 là
$$W_\alpha = (1,645; +\infty).$$
- Tính giá trị quan sát của tiêu chuẩn kiểm định với $n = 100$, $\overline{x} = 100,4$ và $\sigma = 2$,
$$z_0 = \frac{\overline{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{100,4 - 100}{2/\sqrt{100}} = 2.$$
- Vì $z_0 = 2 \in W_\alpha$ nên bác bỏ giả thuyết H_0. Tức là điều nghi ngờ nói trên là có cơ sở với mức ý nghĩa 5%.

Kiểm định giả thuyết về kỳ vọng, mẫu kích thước lớn

✍

- Trong nhiều tình huống thực tế, phương sai σ^2 của tổng thể không biết, thêm nữa, ta không chắc chắn rằng tổng thể có được mô hình hóa bởi biến ngẫu nhiên có phân phối chuẩn hay không. Trong trường hợp này, khi kích thước mẫu n đủ lớn, ta vẫn dùng tiêu chuẩn kiểm định như mục trên, trong đó, độ lệch chuẩn σ được thay bằng độ lệch chuẩn hiệu chỉnh mẫu S.
- Trong thực hành cho phép vận dụng với $n \geq 30$.

Kiểm định giả thuyết về kỳ vọng, mẫu kích thước lớn

Các bước tiến hành

1. Xác định dạng cụ thể của cặp giả thuyết cần kiểm định $\{H_0; H_1\}$.
2. Chọn tiêu chuẩn kiểm định:

$$Z_0 = \frac{\overline{X} - \mu_0}{S/\sqrt{n}}. \tag{8}$$

   Nếu n đủ lớn và nếu giả thuyết H_0 là đúng thì $Z_0 \sim^{xx} \mathcal{N}(0; 1)$.

3. Xây dựng miền bác bỏ giả thuyết H_0 phụ thuộc vào thuyết đối H_1.

H_0	H_1	Miền bác bỏ giả thuyết H_0 (W_α)
$\mu = \mu_0$	$\mu \neq \mu_0$	$(-\infty; -z_{\alpha/2}) \cup (z_{\alpha/2}; +\infty)$
$\mu = \mu_0$	$\mu > \mu_0$	$(z_\alpha; +\infty)$
$\mu = \mu_0$	$\mu < \mu_0$	$(-\infty; -z_\alpha)$

Kiểm định giả thuyết về kỳ vọng, mẫu kích thước lớn

Các bước tiến hành (tiếp theo)

4. Lập mẫu cụ thể $W_x = (x_1, \ldots, x_n)$, tính giá trị quan sát của tiêu chuẩn kiểm định:

$$z_0 = \frac{\overline{x} - \mu_0}{s/\sqrt{n}}. \tag{9}$$

5. Kieemr tra xem z_0 có thuộc W_α hay không để kết luận.

- Nếu $z_0 \in W_\alpha$ thì bác bỏ giả thuyết H_0.
- Nếu $z_0 \notin W_\alpha$ thì chưa có cơ sở để bác bỏ giả thuyết H_0.

Kiểm định giả thuyết về kỳ vọng, mẫu kích thước lớn

Ví dụ 7

Một công ty có một hệ thống máy tính có thể xử lý 1500 hóa đơn trong một giờ. Công ty mới nhập một hệ thống máy tính mới. Hệ thống này khi chạy kiểm tra trong 40 giờ cho thấy số hóa đơn được xử lý trung bình trong một giờ là 1580 hóa đơn với độ lệch chuẩn mẫu hiệu chỉnh là 215 hóa đơn. Với mức ý nghĩa 5% hãy nhận định xem hệ thống mới có tốt hơn hệ thống cũ hay không?

Giải. Gọi X là số hóa đơn mà hệ thống máy tính mới xử lý được trong vòng một giờ, $E(X) = \mu$ là số hóa đơn trung bình mà hệ thống máy tính mới xử lý được trong một giờ. Đây là bài toán kiểm định giả thuyết cho kỳ vọng μ của biến ngẫu nhiên gốc X chưa biết phân phối xác suất, trường hợp mẫu kích thước $n = 40 > 30$.

Kiểm định giả thuyết về kỳ vọng, mẫu kích thước lớn

- Cặp giả thuyết cần kiểm định là $H_0 : \mu = \mu_0$, $H_1 : \mu > \mu_0$ với $\mu_0 = 1500$.
- Chọn tiêu chuẩn kiểm định:

$$Z_0 = \frac{\overline{X} - \mu_0}{S/\sqrt{n}}.$$

  Vì $n = 40 > 30$, nên nếu giả thuyết H_0 là đúng thì $Z_0 \sim^{xx} \mathcal{N}(0; 1)$.
- Với $\alpha = 0,05$, $z_\alpha = z_{0,05} = 1,645$. Miền bác bỏ giả thuyết H_0 là

$$W_\alpha = (1,645; +\infty).$$

- Tính giá trị quan sát của tiêu chuẩn kiểm định với $n = 40$, $\overline{x} = 1580$ và $s = 215$,

$$z_0 = \frac{\overline{x} - \mu_0}{s/\sqrt{n}} = \frac{1580 - 1500}{215/\sqrt{40}} \simeq 2,3533.$$

- Vì $z_0 = 2,3533 \in W_\alpha$ nên bác bỏ giả thuyết H_0. Nghĩa là, với số liệu này có thể coi hệ thống máy mới tốt hơn hệ thống máy cũ với mức ý nghĩa 5%.

Kiểm định giả thuyết về kỳ vọng, phương sai chưa biết

Phân phối mẫu

- Mục này xem xét bài toán kiểm định giả thuyết về kỳ vọng của biến ngẫu nhiên gốc $X \sim \mathcal{N}(\mu; \sigma^2)$ với σ^2 chưa biết.
- Tương tự như trên, cơ sở lý thuyết quan trọng của quy trình kiểm định này là nếu $\overline{X}$ là kỳ vọng mẫu và S là độ lệch chuẩn hiệu chỉnh mẫu của mẫu ngẫu nhiên $W_X = (X_1, X_2, \ldots, X_n)$ kích thước n được chọn từ biến ngẫu nhiên gốc X, thì thống kê

$$T = \frac{\overline{X} - \mu}{S/\sqrt{n}} \tag{10}$$

có phân phối Student với $n-1$ bậc tự do.

Kiểm định giả thuyết về kỳ vọng, phương sai chưa biết

Các bước tiến hành

1. Xác định dạng cụ thể của cặp giả thuyết cần kiểm định $\{H_0; H_1\}$.
2. Chọn tiêu chuẩn kiểm định:

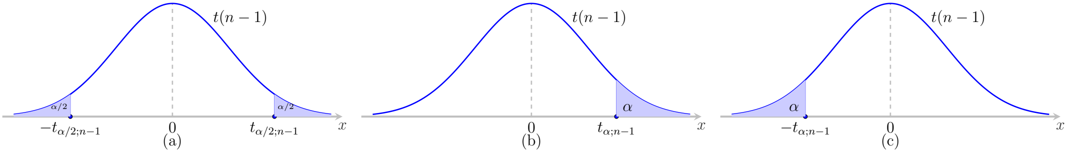
$$T_0 = \frac{\overline{X} - \mu_0}{S/\sqrt{n}}. \tag{11}$$

   Nếu giả thuyết $H_0 : \mu = \mu_0$ là đúng thì $T_0 \sim t(n-1)$.
3. Miền bác bỏ giả thuyết H_0 được xây dựng phụ thuộc vào thuyết đối H_1, xem Hình 3,

H_0	H_1	Miền bác bỏ giả thuyết H_0 (W_α)
$\mu = \mu_0$	$\mu \neq \mu_0$	$\big(-\infty; -t_{\alpha/2;n-1}\big) \cup \big(t_{\alpha/2;n-1}; +\infty\big)$
$\mu = \mu_0$	$\mu > \mu_0$	$\big(t_{\alpha;n-1}; +\infty\big)$
$\mu = \mu_0$	$\mu < \mu_0$	$\big(-\infty; -t_{\alpha;n-1}\big)$

trong đó, $t_{\alpha/2;n-1}$ và $t_{\alpha;n-1}$ là các giá trị tới hạn của phân phối Student mức $\alpha/2$ và α tương ứng với $n-1$ bậc tự do.

Kiểm định giả thuyết về kỳ vọng, phương sai chưa biết



Hình 3: Phân phối xác suất của T_0 khi $H_0 : \mu = \mu_0$ là đúng và miền bác bỏ giả thuyết H_0 ứng với (a) $H_1 : \mu \neq \mu_0$, (b) $H_1 : \mu > \mu_0$, (c) $H_1 : \mu < \mu_0$

Kiểm định giả thuyết về kỳ vọng, phương sai chưa biết

Các bước tiến hành (tiếp theo)

4. Lập mẫu cụ thể $W_x = (x_1, x_2, \dots, x_n)$ và tính giá trị quan sát của tiêu chuẩn kiểm định:

$$t_0 = \frac{\overline{x} - \mu_0}{s/\sqrt{n}}. \tag{12}$$

5. Kiểm tra xem t_0 có thuộc W_α hay không để kết luận.

- Nếu $t_0 \in W_\alpha$ thì bác bỏ giả thuyết H_0.
- Nếu $t_0 \notin W_\alpha$ thì chưa có cơ sở để bác bỏ giả thuyết H_0.

Kiểm định giả thuyết về kỳ vọng, phương sai chưa biết

Ví dụ 8

Một công ty sản xuất hạt giống tuyên bố rằng một loại giống mới của họ có năng suất trung bình là 21,5 tạ/ha. Gieo thử hạt giống mới này tại 16 vườn thí nghiệm và thu được kết quả (đơn vị tính là tạ/ha):

$$19,2;\ 18,7;\ 22,4;\ 20,3;\ 16,8;\ 25,1;\ 17,0;\ 15,8;\ 21,0;\ 18,6;\ 23,7;\ 24,1;\ 23,4;\ 19,8;\ 21,7;\ 18,9.$$

Dựa vào kết quả này hãy xác nhận xem quảng cáo của công ty có đúng không với mức ý nghĩa $\alpha = 0,05$. Biết rằng năng suất giống cây trồng là một biến ngẫu nhiên có phân phối chuẩn.

Giải. Gọi X (tạ/ha) là năng suất giống cây trồng, $X \sim \mathcal{N}(\mu; \sigma^2)$. Đây là bài toán kiểm định giả thuyết về kỳ vọng $E(X) = \mu$ của biến ngẫu nhiên có phân phối chuẩn trong trường hợp phương sai chưa biết, kích thước mẫu $n = 16 < 30$.

Kiểm định giả thuyết về kỳ vọng, phương sai chưa biết

- Cặp giả thuyết cần kiểm định là $H_0 : \mu = \mu_0$, $H_1 : \mu \neq \mu_0$ với $\mu_0 = 21,5$.
- Chọn tiêu chuẩn kiểm định:
$$T_0 = \frac{\overline{X} - \mu_0}{S/\sqrt{n}}.$$
Nếu giả thuyết H_0 là đúng thì $T_0 \sim t(n-1)$.
- Với $\alpha = 0,05$, $t_{\alpha/2;n-1} = t_{0,025;15} = 2,131$. Miền bác bỏ giả thuyết H_0 là
$$W_\alpha = (-\infty; -2,131) \cup (2,131; +\infty).$$
- Tính giá trị quan sát của tiêu chuẩn kiểm định với $n = 16$; $\overline{x} = 20,406$ và $s = 3,038$
$$t_0 = \frac{\overline{x} - \mu_0}{s/\sqrt{n}} = \frac{20,406 - 21,5}{3,038/\sqrt{16}} = -1,4404.$$
- Vì $t_0 = -1,4404 \notin W_\alpha$ nên chưa có cơ sở để bác bỏ giả thuyết H_0. Nghĩa là, với số liệu này có thể chấp nhận lời quảng cáo của công ty với mức ý nghĩa 5%.

Kiểm định giả thuyết về kỳ vọng, phương sai chưa biết

✍ Trong thực hành, khi $n \geq 30$, thì thống kê T trong (10) có phân phối xấp xỉ phân phối chuẩn tắc $\mathcal{N}(0; 1)$. Do đó, nếu $n < 30$ ta áp dụng tiêu chuẩn kiểm định (11), nếu $n \geq 30$ ta sử dụng tiêu chuẩn kiểm định (8).

5.2. KIỂM ĐỊNH GIẢ THUYẾT VỀ THAM SỐ CỦA TỔNG THỂ

Bài toán

Bài toán 2

Cho biến ngẫu nhiên gốc $X \sim \mathcal{N}(\mu; \sigma^2)$ với $V(X) = \sigma^2$ chưa biết, nhưng có cơ sở để nêu lên giả thuyết

$$H_0 : \sigma^2 = \sigma_0^2 \text{ trong đó } \sigma_0^2 \text{ đã biết.}$$

Hãy kiểm định giả thuyết về phương sai σ^2 ở một trong ba dạng của cặp giả thuyết

1. $H_0 : \sigma^2 = \sigma_0^2;\ H_1 : \sigma^2 \neq \sigma_0^2$
2. $H_0 : \sigma^2 = \sigma_0^2;\ H_1 : \sigma^2 > \sigma_0^2$
3. $H_0 : \sigma^2 = \sigma_0^2;\ H_1 : \sigma^2 < \sigma_0^2$

Phân phối mẫu

Phân phối khi-bình phương

- Từ biến ngẫu nhiên gốc X, lập mẫu ngẫu nhiên $W_X = (X_1, X_2, \ldots, X_n)$ kích thước n.
- Với S^2 là phương sai hiệu chỉnh mẫu của mẫu ngẫu nhiên này, thì thống kê

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

có phân phối khi-bình phương với $n-1$ bậc tự do.

Các bước tiến hành

1. Xác định dạng cụ thể của cặp giả thuyết cần kiểm định $\{H_0; H_1\}$.
2. Chọn tiêu chuẩn kiểm định

$$\chi_0^2 = \frac{(n-1)S^2}{\sigma_0^2}. \tag{13}$$

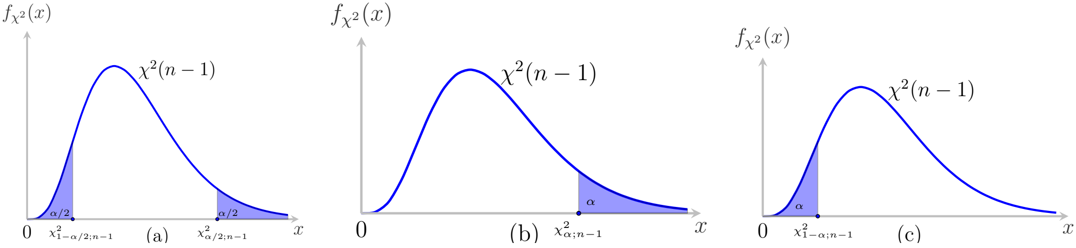
Nếu giả thuyết $H_0 : \sigma^2 = \sigma_0^2$ là đúng, thì $\chi_0^2 \sim \chi^2(n-1)$.

3. Xây dựng miền bác bỏ giả thuyết H_0 phụ thuộc vào đối thuyết H_1, xem Hình 4,

H_0	H_1	Miền bác bỏ giả thuyết H_0 (W_α)
$\sigma^2 = \sigma_0^2$	$\sigma^2 \neq \sigma_0^2$	$(0; \chi^2_{1-\alpha/2;n-1}) \cup (\chi^2_{\alpha/2;n-1}; +\infty)$
$\sigma^2 = \sigma_0^2$	$\sigma^2 > \sigma_0^2$	$(\chi^2_{\alpha;n-1}; +\infty)$
$\sigma^2 = \sigma_0^2$	$\sigma^2 < \sigma_0^2$	$(0; \chi^2_{1-\alpha;n-1})$

trong đó, $\chi^2_{\alpha/2;n-1}$, $\chi^2_{1-\alpha/2;n-1}$, $\chi^2_{1-\alpha;n-1}$ và $\chi^2_{\alpha;n-1}$ là các giá trị tới hạn của phân phối khi-bình phương mức $\alpha/2$, $1-\alpha/2$, $1-\alpha$ và α tương ứng với $n-1$ bậc tự do.

Các bước tiến hành



Hình 4: Minh họa phân phối xác suất cho kiểm định giả thuyết $H_0 : \sigma = \sigma_0$ với miền bác bỏ giả thuyết H_0 ứng với (a) $H_1 : \sigma \neq \sigma_0$, (b) $H_1 : \sigma > \sigma_0$, (c) $H_1 : \sigma < \sigma_0$

Các bước tiến hành

4. Lập mẫu cụ thể $(x_1, \ldots, x_n)$, tính giá trị quan sát của tiêu chuẩn kiểm định:

$$\chi_0^2 = \frac{(n-1)s^2}{\sigma_0^2}. \tag{14}$$

5. Kiểm tra xem χ_0^2 có thuộc W_α hay không để kết luận.

- Nếu $\chi_0^2 \in W_\alpha$ thì bác bỏ giả thuyết H_0.
- Nếu $\chi_0^2 \notin W_\alpha$ thì chưa có cơ sở để bác bỏ giả thuyết H_0.

Ví dụ

Ví dụ 9

Để kiểm tra chất lượng của một loại sản phẩm, người kỹ sư đo đường kính của 12 sản phẩm của một dây chuyền sản xuất tính được độ lệch chuẩn hiệu chỉnh mẫu là $s = 0,3$. Biết rằng, nếu độ biến động của các sản phẩm lớn hơn 0,2 thì dây chuyền sản xuất phải dừng lại để điều chỉnh. Với mức ý nghĩa 5% người kỹ sư có quyết định cho dừng dây chuyền sản xuất để điều chỉnh không? Giả thiết rằng đường kính của các sản phẩm là biến ngẫu nhiên có phân phối chuẩn.

Giải. Gọi X là đường kính của loại sản phẩm nói trên, $X \sim \mathcal{N}(\mu; \sigma^2)$. Đây là bài toán kiểm định giả thuyết về phương sai của biến ngẫu nhiên có phân phối chuẩn.

Ví dụ

- Cặp giả thuyết cần kiểm định là $H_0 : \sigma^2 = \sigma_0^2$, $H_1 : \sigma^2 > \sigma_0^2$ với $\sigma_0^2 = (0,2)^2 = 0,04$.
- Chọn tiêu chuẩn kiểm định
$$\chi_0^2 = \frac{(n-1)S^2}{\sigma_0^2}.$$
Nếu giả thuyết H_0 là đúng thì $\chi_0^2 \sim \chi^2(n-1)$.
- Với $\alpha = 0,05$, $\chi^2_{\alpha;n-1} = \chi^2_{0,05;11} = 19,68$. Miền bác bỏ giả thuyết H_0 là
$$W_\alpha = (19,68; +\infty).$$
- Tính giá trị quan sát của tiêu chuẩn kiểm định với $n = 12$ và $s = 0,3$,
$$\chi_0^2 = \frac{(n-1)s^2}{\sigma_0^2} = \frac{11 \times (0,3)^2}{0,04} = 24,75.$$
- Vì $\chi_0^2 = 24,75 \in W_\alpha$ nên bác bỏ giả thuyết H_0, chấp nhận H_1. Tức là dây chuyền cần điều chỉnh vì độ biến động đã lớn hơn mức cho phép.

5.2. KIỂM ĐỊNH GIẢ THUYẾT VỀ THAM SỐ CỦA TỔNG THỂ

Bài toán

Bài toán 3

Giả sử ta quan tâm đến một dấu hiệu A nào đó mà mỗi cá thể của tổng thể có thể có dấu hiệu này hoặc không. Gọi p là tỷ lệ cá thể có dấu hiệu A của tổng thể. Giả sử p chưa biết, nhưng có cơ sở để nêu lên giả thuyết

$$H_0 : p = p_0 \quad \text{với} \quad p_0 \text{ là một tỷ lệ được giả định.}$$

Hãy kiểm định giả thuyết về tỷ lệ p ở một trong ba dạng của cặp giả thuyết sau

1. $H_0 : p = p_0;\ H_1 : p \neq p_0$
2. $H_0 : p = p_0;\ H_1 : p > p_0$
3. $H_0 : p = p_0;\ H_1 : p < p_0$

Phân phối mẫu

- Giả sử một mẫu ngẫu nhiên có kích thước n được lấy từ một tổng thể có kích thước lớn (có thể là vô hạn) và giả sử X $(\leq n)$ là số cá thể có dấu hiệu A trong mẫu này. Khi đó, $\widehat{P} = X/n$ là một ước lượng điểm không chệch của p.
- Phân phối mẫu của $\widehat{P}$ xấp xỉ phân phối chuẩn với kỳ vọng p và phương sai $p(1-p)/n$, nếu p không quá gần 0 hoặc 1 và nếu n tương đối lớn.
- Để áp dụng xấp xỉ này, yêu cầu $np \geq 5$ và $n(1-p) \geq 5$.

Các bước tiến hành

1. Xác định dạng cụ thể của cặp giả thuyết cần kiểm định $\{H_0; H_1\}$.
2. Chọn tiêu chuẩn kiểm định

$$Z_0 = \frac{\widehat{P} - p_0}{\sqrt{p_0(1-p_0)/n}}. \qquad (15)$$

   Nếu giả thuyết $H_0 : p = p_0$ là đúng và n đủ lớn thì $Z_0 \sim^{xx} \mathcal{N}(0;1)$. Trong thực hành áp dụng với $np_0 \geq 5$ và $n(1-p_0) \geq 5$.

3. Xây dựng miền bác bỏ giả thuyết H_0 phụ thuộc vào đối thuyết H_1.

H_0	H_1	Miền bác bỏ giả thuyết H_0 (W_α)
$p = p_0$	$p \neq p_0$	$(-\infty; -z_{\alpha/2}) \cup (z_{\alpha/2}; +\infty)$
$p = p_0$	$p > p_0$	$(z_\alpha; +\infty)$
$p = p_0$	$p < p_0$	$(-\infty; -z_\alpha)$

Các bước tiến hành

4. Lập mẫu cụ thể, tính giá trị quan sát của tiêu chuẩn kiểm định:

$$z_0 = \frac{\widehat{p} - p_0}{\sqrt{p_0(1-p_0)/n}}. \tag{16}$$

5. Kiểm tra xem z_0 có thuộc W_α hay không để kết luận.

- Nếu $z_0 \in W_\alpha$ thì bác bỏ giả thuyết H_0.
- Nếu $z_0 \notin W_\alpha$ thì chưa có cơ sở để bác bỏ giả thuyết H_0.

Các bước tiến hành

✍ Một tiêu chuẩn kiểm định khác tương đương với (15) được xây dựng như sau. Nếu giả thuyết $H_0 : p = p_0$ là đúng thì $X \sim^{xx} \mathcal{N}(np_0; np_0(1-p_0))$. Khi đó, thống kê

$$Z_0 = \frac{X - np_0}{\sqrt{np_0(1-p_0)}} \tag{17}$$

xấp xỉ phân phối chuẩn tắc $\mathcal{N}(0; 1)$.

Ví dụ

Ví dụ 10

Một công ty A sản xuất bánh kẹo tuyên bố rằng 1/2 số trẻ em thích ăn bánh kẹo của công ty. Trong một mẫu gồm 100 trẻ em được hỏi, có 47 em tỏ ra thích ăn bánh kẹo của công ty. Với mức ý nghĩa 5%, số liệu trên có chứng tỏ là tuyên bố của công ty là đúng hay không?

Giải. Gọi p là tỷ lệ trẻ em thích bánh kẹo của công ty. Đây là bài toán kiểm định giả thuyết về tỷ lệ của tổng thể với kích thước mẫu $n = 100$ đủ lớn. Kiểm tra điều kiện $np_0 = 100 \times 0,5 > 5$ và $n(1 - p_0) = 100 \times (1 - 0,5) > 5$.

Ví dụ

- Cặp giả thuyết cần kiểm định là $H_0 : p = p_0$, $H_1 : p \neq p_0$ với $p_0 = 0,5$.
- Chọn tiêu chuẩn kiểm định

$$Z_0 = \frac{\widehat{P} - p_0}{\sqrt{p_0(1-p_0)/n}}.$$

Nếu giả thuyết H_0 là đúng thì $Z_0 \sim^{xx} \mathcal{N}(0;1)$.

- Với $\alpha = 0,05$, $z_{\alpha/2} = z_{0,025} = 1,96$. Miền bác bỏ giả thuyết H_0 là

$$W_\alpha = (-\infty; -1,96) \cup (1,96; +\infty).$$

- Tính giá trị quan sát của tiêu chuẩn kiểm định với $n = 100$, $x = 47$,

$$z_0 = \frac{\widehat{p} - p_0}{\sqrt{p_0(1-p_0)/n}} = \frac{47/100 - 0,5}{\sqrt{0,5 \times 0,5/100}} = -0,6.$$

- Vì $z_0 = -0,6 \notin W_\alpha$ nên chưa có cơ sở để bác bỏ giả thuyết H_0. Như vậy, tuyên bố của công ty là có cơ sở với mức ý nghĩa 5%.

Tóm tắt bài toán kiểm định về tham số của một tổng thể

Trường hợp	Tiêu chuẩn kiểm định	H_0	H_1	Miền bác bỏ W_α
σ^2 đã biết	$Z_0 = \dfrac{\overline{X} - \mu_0}{\sigma/\sqrt{n}}$	$\mu = \mu_0$	$\mu \neq \mu_0$ $\mu > \mu_0$ $\mu < \mu_0$	$(-\infty; -z_{\alpha/2}) \cup (z_{\alpha/2}; +\infty)$ $(z_\alpha; +\infty)$ $(-\infty; -z_\alpha)$
σ^2 chưa biết $n < 30$	$T_0 = \dfrac{\overline{X} - \mu_0}{S/\sqrt{n}}$	$\mu = \mu_0$	$\mu \neq \mu_0$ $\mu > \mu_0$ $\mu < \mu_0$	$(-\infty; -t_{\alpha/2;n-1}) \cup (t_{\alpha/2;n-1}; +\infty)$ $(t_{\alpha;n-1}; +\infty)$ $(-\infty; -t_{\alpha;n-1})$
σ^2 chưa biết $n \geq 30$	$Z_0 = \dfrac{\overline{X} - \mu_0}{S/\sqrt{n}}$	$\mu = \mu_0$	$\mu \neq \mu_0$ $\mu > \mu_0$ $\mu < \mu_0$	$(-\infty; -z_{\alpha/2}) \cup (z_{\alpha/2}; +\infty)$ $(z_\alpha; +\infty)$ $(-\infty; -z_\alpha)$
$np_0 \geq 5$ $n(1-p_0) \geq 5$	$Z_0 = \dfrac{\widehat{P} - p_0}{\sqrt{p_0(1-p_0)/n}}$	$p = p_0$	$p \neq p_0$ $p > p_0$ $p < p_0$	$(-\infty; -z_{\alpha/2}) \cup (z_{\alpha/2}; +\infty)$ $(z_\alpha; +\infty)$ $(-\infty; -z_\alpha)$
μ chưa biết	$\chi_0^2 = \dfrac{(n-1)S^2}{\sigma_0^2}$	$\sigma^2 = \sigma_0^2$	$\sigma^2 \neq \sigma_0^2$ $\sigma^2 > \sigma_0^2$ $\sigma^2 < \sigma_0^2$	$(0; \chi^2_{1-\alpha/2;n-1}) \cup (\chi^2_{\alpha/2;n-1}; +\infty)$ $(\chi^2_{\alpha;n-1}; +\infty)$ $(0; \chi^2_{1-\alpha;n-1})$

Bài tập

Bài 3

Một công ty sản xuất bóng đèn cho biết tuổi thọ trung bình của loại bóng đèn là 800 giờ và độ lệch tiêu chuẩn là 40 giờ. Hãy kiểm định giả thuyết $\mu = 800$ và đối thuyết $\mu \neq 800$, nếu kiểm tra ngẫu nhiên mẫu 30 bóng đèn thu được tuổi thọ trung bình là 788 giờ, với mức ý nghĩa 5%. Giả sử tuổi thọ của loại bóng đèn trên là biến ngẫu nhiên có phân phối chuẩn $\mathcal{N}(\mu; \sigma^2)$.

Bài 4

Một dây chuyền đóng gói mỳ chính tự động với quy định trọng lượng mỗi gói là 250 gam. Kiểm tra ngẫu nhiên 25 gói được đóng bằng máy tự động trên thì thu được số liệu như sau:

Trọng lượng (gam)	240,7	240,9	250,1	250,3	250,5
Số gói	3	7	8	5	2

Giả sử trọng lượng của các gói mỳ chính có phân phối chuẩn. Với mức ý nghĩa 5%, hỏi trọng lượng trung bình của các gói mỳ chính được đóng gói tự động giống như yêu cầu hay không?

Bài tập

Bài 5

Doanh số bán hàng của một siêu thị là biến ngẫu nhiên X (tỷ VNĐ/ngày). Điều tra về doanh số bán hàng của siêu thị này trong 50 ngày thu được bảng số liệu sau:

Doanh số	[8;10)	[10;12)	[12;14)	[14;16)	[16;18]
Số ngày	5	12	21	8	4

Với mức ý nghĩa 5%, cần kiểm tra xem doanh số bán hàng trung bình một ngày của siêu thị có hơn 12 tỷ VNĐ hay không?

Bài 6

Tại công ty xăng dầu, quá trình nạp nhớt vào bình chỉ chấp nhận độ lệch chuẩn tối đa 10 ml. Để kiểm tra máy nạp M, người ta lấy ra 20 bình để kiểm tra. Kết quả cho thấy độ lệch chuẩn là 12 ml. Với độ tin cậy là 95%, độ biến động của máy M có đạt yêu cầu của công ty hay không?

Bài tập

Bài 7

Trọng lượng của một loại trái cây từ 253 gam trở lên được xếp là trái cây loại I. Kiểm tra ngẫu nhiên 100 trái và thu được kết quả sau:

Trọng lượng (gam)	245	247	248	250	252	253	254
Số trái	8	12	20	32	16	8	4

Có ý kiến cho rằng tỷ lệ trái cây loại I là 15%. Với mức ý nghĩa 5%, có thể kết luận như thế nào về ý kiến nêu trên?