Bài số 12

KIỂM ĐỊNH GIẢ THIẾT VỀ TỶ LỆ

I. KIỂM ĐỊNH MỘT TỶ LỆ

1.Khái niệm. Giả sử một tổng thể có hai loại phần tử: có tính chất $\Im$ và không có tính chất $\Im$, trong đó tỷ lệ phần tử có tính chất $\Im$ là p_0 chưa biết. Bài toán kiểm định giả thiết: $H_0 : p = p_0$ với đối thiết H_1 gọi là bài toán kiểm định tỷ lệ.

Ta sẽ xét bài toán kiểm định giả thuyết tỷ lệ số lần thành công trong phép thử nhị thức bằng giá trị cụ thể nào đó. Tức là, ta kiểm định giả thuyết $H_0 : p = p_0$ với p là tham số trong phân phối nhị thức. Đối thuyết có thể là một phía hoặc hai phía:

$$p < p_0, p > p_0 \text{ hoặc } p \neq p_0.$$

Biến ngẫu nhiên thích hợp mà ta dựa vào để kiểm định là biến ngẫu nhiên nhị thức X. Vì X là biến ngẫu nhiên nhị thức rời rạc, nên không chắc miền bác bỏ có thể được xây dựng với cỡ của nó chính xác bằng một giá trị α cho trước.

1.Đối với mẫu cỡ nhỏ

a. Để kiểm định: $H_0 : p = p_0$,

$H_1 : p < p_0$,

ta sử dụng phân phối nhị thức để tính: $P = P(X \le x \text{ khi } p = p_0)$.

Giá trị x là số lần thành công trong mẫu cỡ n:

Nếu $P \le \alpha$, bài toán sẽ có mức ý nghĩa là α và ta bác bỏ H_0, chấp nhận H_1.

b. Để kiểm định: $H_0 : p = p_0$,

$H_1 : p > p_0$,

tại mức ý nghĩa α, ta tính: $P = P(X \ge x \text{ khi } p = p_0)$

Khi đó ta sẽ bác bỏ H_0 (chấp nhận H_1) nếu $P \le \alpha$.

c. Để kiểm định: $H_0 : p = p_0$,

$H_1 : p \neq p_0$,

tại mức ý nghĩa α, ta tính: $P = 2P(X \le x \text{ khi } p = p_0)$ nếu $x < np_0$

hoặc $P = 2P(X \ge x \text{ khi } p = p_0)$ nếu $x > np_0$

Khi đó ta sẽ bác bỏ H_0 (chấp nhận H_1) nếu $P \le \alpha$.

Các bước kiểm định giả thuyết về tỷ lệ với các đối thuyết khác nhau sử dụng các xác suất nhị thức được cho trong Bảng A.1, được tiến hành như sau:

1. $H_0 : p = p_0$
2. $H_1 : p < p_0, p > p_0$ hoặc $p \neq p_0$.
3. Chọn mức ý nghĩa α.
4. Thống kê tiêu chuẩn: Biến ngẫu nhiên nhị thức X với $p = p_0$.
5. Tính toán: Tìm *x* là số lần thành công, tính giá trị P thích hợp.
6. Kết luận: Đưa ra kết luận phù hợp dựa vào giá trị P.

Ví dụ 1. Một nhà xây dựng khẳng định, máy bơm nhiệt (dùng để sưởi ấm, làm mát tòa nhà) được lắp đặt trong 70% số căn hộ được xây dựng hiện nay bởi công ty Richmond. Có thể đồng ý với khẳng định này không nếu một cuộc điều tra ngẫu nhiên cho thấy, 8 trên 15 căn nhà được lắp máy bơm nhiệt? Dùng mức ý nghĩa $0,1$.

Giải. Gọi p là tỷ lệ căn hộ được lắp bơm nhiệt

Xét bài toán kiểm định giả thiết : $H_0 : p = p_0$

với đối thiết $H_1 : p \neq 0,7$.

Ta có: + mức ý nghĩa: $\alpha = 0,1$

+ Tiêu chuẩn kiểm định giả thiết: Biến ngẫu nhiên nhị thức X với $p = 0,7$ và $n = 15$.

+ Tính toán: $x = 8$ và $np_0 = 15.0,7 = 10,5$, suy ra $x < np_0$.

Do đó, từ Bảng A.1, ta tính được:

$$P = 2P(X \leq 8 \text{ khi } p = p_0) = 2\sum_{x=0}^{8} b(x;15;0,7) = 0,2622 > 0,1.$$

6. **<u>Kết luận</u>**: Không bác bỏ H_0: tức là ta không đủ bằng chứng để nghi ngờ nhà xây dựng.

2. <u>Đối với mẫu cỡ lớn.</u>

Với n lớn, ta cần dùng phương pháp xấp xỉ. Khi giá trị giả thuyết p_0 rất gần 0 hoặc 1, ta sử dụng phân phối Poisson với tham số $\mu = np_0$.

Tuy nhiên, xấp xỉ phân phối chuẩn với tham số $\mu = np_0$ và $\sigma^2 = np_0q_0$ thường được sử dụng nhiều hơn cho trường hợp n lớn và nó chính xác ngay cả khi p_0 không gần 0 hoặc 1.

Nếu ta dùng xấp xỉ phân phối chuẩn tắc để kiểm định $p = p_0$ được cho bởi:

$$z = \frac{x - np_0}{\sqrt{np_0 q_0}}, \; q_0 = 1 - p_0$$

là giá trị của biến ngẫu nhiên chuẩn chuẩn hóa $Z = \dfrac{X - np_0}{\sqrt{np_0 q_0}}$.

Khi đó:

a. Đối với bài toán kiểm định hai phía: $H_0 : p = p_0$

$$H_1 : p \neq p_0$$

tại mức ý nghĩa α, miền bác bỏ là: $(-\infty; -z_{\alpha/2}) \cup (z_{\alpha/2}; +\infty)$ trong đó $z_{\alpha/2}$ được xác định bởi:

$$P(Z > z_{\alpha/2}) = \alpha/2.$$

b. Đối với bài toán kiểm định một phía: $H_0 : p = p_0$

$$H_1 : p < p_0$$

tại mức ý nghĩa α, miền bác bỏ là $(-\infty; -z_{\alpha/2})$ trong đó $z_{\alpha/2}$ được xác định bởi:

$$P(Z > z_{\alpha/2}) = \alpha/2.$$

c. Đối với bài toán kiểm định một phía: $H_0 : p = p_0$,

$$H_1 : p > p_0$$

tại mức ý nghĩa α, miền bác bỏ là $(z_{\alpha/2}; +\infty)$ trong đó $z_{\alpha/2}$ được xác định bởi:

$$P(Z > z_{\alpha/2}) = \alpha/2.$$

Ví dụ 2. Một loại thuốc an thần thường được dùng được tin là chỉ có tác động tới 60% người sử dụng. Kết quả thử nghiệm loại thuốc mới trên 100 người trưởng thành cho thấy 70 người nhận được tác dụng. Có thể tin được hay không rằng loại thuốc mới tốt hơn loại thường dùng? Sử dụng mức ý nghĩa $0,05$.

Giải: Gọi p là tỷ lệ người chịu sự tác động của loại thuốc mới.

Xét bài toán kiểm định giả thiết: $H_0 : p = 0,6$.

với đối thiết $H_1 : p > 0,6$.

Tại mức ý nghĩa: α=0,05 ta có:

+ Miền bác bỏ: $(z_{0,025};+\infty)$ trong đó $z_{0,025}$ được xác định bởi:

$$P(Z > z_{0,025}) = 0,025\text{, tra bảng A3 ta được } z_{0,025} = 1,645$$

Nên miền bác bỏ: $(1,645;+\infty)$.

+ Tiêu chuẩn kiểm định giả thiết: $Z = \dfrac{X - np_0}{\sqrt{np_0q_0}}$.

Tính toán:

$x = 70$; $n = 100$; $np_0 = 100.0,6 = 60$ và:

$$z = \frac{70-60}{\sqrt{100.0,6.0,4}} = 2,04 > 1,645$$

+ ***Kết luận:*** Bác bỏ H_0, tức là loại thuốc mới là tốt hơn.

II. KIỂM ĐỊNH HAI TỶ LỆ

Xét bài toán kiểm định giả thuyết rằng: hai tỷ lệ hoặc hai tham số nhị thức là bằng nhau. Tức là ta muốn kiểm định giả thuyết $H_0 : p_1 = p_2$

với một trong các đối thuyết $p_1 < p_2$, $p_1 > p_2$ hoặc $p_1 \neq p_2$.

Điều này tương đương với giả thuyết $H_0 : p_1 - p_2 = 0$

và các đối thuyết $p_1 - p_2 < 0$, $p_1 - p_2 > 0$ hoặc $p_1 - p_2 \neq 0$.

Thống kê mà ta dựa vào để kiểm định là biến ngẫu nhiên $\hat{P}_1 - \hat{P}_2$. Các mẫu độc lập cỡ n_1, n_2 được chọn ngẫu nhiên từ hai phân phối nhị thức và tỷ lệ thành công $\hat{P}_1, \hat{P}_2$ cho hai mẫu được tính toán.

Trong xây dựng khoảng tin cậy cho p_1, p_2, ta chú ý rằng, với n_1, n_2 đủ lớn, ước lượng điểm $\hat{P}_1 - \hat{P}_2$ xấp xỉ phân phối chuẩn với trung bình:

$$\mu_{\hat{P}_1-\hat{P}_2} = p_1 - p_2 \quad \text{và phương sai} \quad \sigma^2_{\hat{P}_1-\hat{P}_2} = \frac{p_1q_1}{n_1} + \frac{p_2q_2}{n_2}.$$

Do đó, miền chấp nhận và tiêu chuẩn có thể được thành lập dựa vào biến ngẫu nhiên chuẩn chuẩn hóa:

$$Z = \frac{(\hat{P}_1 - \hat{P}_2) - (p_1 - p_2)}{\sqrt{(p_1q_1 / n_1) + (p_2q_2 / n_2)}}.$$

Khi H_0 đúng, ta đặt $p_1 = p_2 = p$ và $q_1 = q_2 = q$ thì công thức trên trở thành:

$$Z = \frac{\hat{P}_1 - \hat{P}_2}{\sqrt{pq[(1/n_1) + (1/n_2)]}}.$$

Để tính Z, ta phải ước lượng được các tham số p và q trong dấu căn. Gộp chung dữ liệu từ cả hai mẫu, **ước lượng gộp chung cho tỷ lệ p** là:

$$\hat{p} = \frac{x_1 + x_2}{n_1 + n_2},$$

với x_1, x_2 là số lần thành công trong mỗi mẫu. Thay $\hat{p}$ cho p và $\hat{q} = 1 - \hat{p}$ cho q, giá trị z để kiểm định $p_1 = p_2$ được xác định qua công thức:

$$z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}\hat{q}[(1/n_1) + (1/n_2)]}}.$$

Khi đó miền bác bỏ:

+ Với đối thuyết $p_1 \neq p_2$ tại mức ý nghĩa α , miền bác bỏ là $(-\infty; -z_{\alpha/2}) \cup (z_{\alpha/2}; +\infty)$.

+ Với đối thuyết $p_1 < p_2$, miền bác bỏ là $(-\infty; -z_{\alpha/2})$.

+ Với đối thuyết $p_1 > p_2$, miền bác bỏ là $(z_{\alpha/2}; +\infty)$.

Ví dụ 3. Một cuộc bỏ phiếu được đưa ra để xác định vị trí xây dựng một nhà máy hóa chất ở trong trị trấn hay ở ngoại vi thị trấn. Có 120 trên 200 cử tri trong thị trấn đồng ý xây dựng nhà máy trong thị trấn và 240 trên 500 cử tri ở ngoại vi đồng ý với đề xuất này. Liệu có thể cho rằng tỷ lệ cử tri trong thị trấn đồng ý với đề xuất lớn hơn tỷ lệ cử tri ở ngoại vi đồng ý hay không? Sử dụng mức ý nghĩa 0,025.

Giải: Gọi p_1, p_2 tương ứng là tỷ lệ cử tri trong thị trấn và ở ngoại vi đồng ý với đề xuất.

Ta xét bài toán kiểm định: $H_0 : p_1 = p_2$.

$H_1 : p_1 > p_2$.

Với mức ý nghĩa: α=0,025, ta có miền bác bỏ là: $(2,24; +\infty)$.

Tính toán với: : $\hat{p}_1 = \frac{x_1}{n_1} = \frac{120}{200} = 0,6$ $\hat{p}_2 = \frac{x_2}{n_2} = \frac{240}{500} = 0,48$

$$\hat{p} = \frac{x_1 + x_2}{n_1 + n_2} = \frac{120 + 240}{200 + 500} = 0,51.$$

Do đó: $z = \frac{0,6 - 0,48}{\sqrt{0,51.0,49.[(1/200) + (1/500)]}} = 2,9 > 2,24$

Kết luận: Bác bỏ H_0, chấp nhận tỷ lệ cử tri trong thị trấn đồng ý với đề xuất lớn hơn tỷ lệ cử tri ở ngoại vi đồng ý.

Về nhà:

Tự đọc: Mục 10.14; 10.15; 10.16 và 10.17

Bài tập: Tr**. 360; 364; 376**

Đọc trước các Mục từ 11.1 đến 11.3 chuẩn bị cho **Bài số 13 :**

Hồi quy tuyến tính đơn