Bài số 10

BÀI TOÁN ƯỚC LƯỢNG TỶ LỆ

I. ƯỚC LƯỢNG TỶ LỆ.

1.Định nghĩa. Giả sử tổng thể được chia làm hai loại phần tử. Tỷ lệ phần tử có dấu hiệu $\Im$ là p chưa biết. Ước lượng tỷ lệ là chỉ ra khoảng (p_1, p_2) chứa p sao cho $P(p_1 < p < p_2) = 1 - \alpha$.

Một ước lượng điểm có tỷ lệ p trong một phép thử nhị thức được xác định bằng thống kê $\hat{P} = X / n$, trong đó X là số các thành công trong n lần thử nghiệm. Vì thế, tỷ lệ mẫu $\hat{p} = x / n$ sẽ được dùng làm ước lượng điểm của tham số p.
Ký hiệu thất bại trong thử nghiệm nhị thức là 0 và thành công là 1, số thành công x có thể được hiểu là tổng n các giá trị chỉ bao gồm các số 0 và số 1, và $\hat{p}$ chỉ là trung bình mẫu của n các giá trị này. Vì thế, theo Định lý giới hạn trung tâm, với n đủ lớn $\hat{P}$ có phân bố chuẩn tắc với kỳ vọng là:

$$\mu_{\bar{p}} = E(\hat{P}) = E\left[\frac{X}{n}\right] = \frac{np}{n} = p$$

và phương sai

$$\sigma_{\hat{P}}^2 = \sigma_{X/m}^2 = \frac{\sigma_x^2}{n^2} = \frac{npq}{n^2} = \frac{pq}{n}$$

vì thế, chúng ta có thể kết luận rằng

$$P(-z_{\alpha/2} < Z < z_{\alpha/2}) = 1 - \alpha$$

trong đó

$$Z = \frac{\hat{P} - p}{\sqrt{pq / n}}$$

và $z_{\alpha/2}$ là giá trị của biểu đồ chuẩn trên đó chúng ta có thể xác định được một diện tích *α/2*. Thay cho *Z*, chúng ta viết

$$P\left(-z_{\alpha/2} < \frac{\hat{P} - p}{\sqrt{pq / n}} < z_{\alpha/2}\right) = 1 - \alpha$$

Nhân mỗi số hạng của bất đẳng thức với $\sqrt{pq / n}$, và sau đó trừ với $\hat{P}$ và nhân với -1, chúng ta thu được:

$$P\left(\hat{P} - z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}} < p < \hat{P} + z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}}\right) = 1 - \alpha; \quad \hat{q} = 1 - \hat{p}$$

Khi n lớn, rất ít sai số xảy ra khi thay thế ước lượng điểm $\hat{p} = x / n$ cho *p* dưới dấu căn. Khi đó chúng ta có thể viết

$$P\left(\hat{P} - z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}} < p < \hat{P} + z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}}\right) \simeq 1-\alpha$$

Đối với mẫu ngẫu nhiên có cỡ n, tỷ lệ mẫu $\hat{p} = x / n$ được xác định và khoảng tin cậy (1-α) % ước lượng sau đối với *p* được xác định.

2. Khoảng tin cậy đối với p khi mẫu cỡ lớn.

Định lý 1. Nếu $\hat{p}$ là tỷ lệ của các thành công trong một mẫu ngẫu nhiên có kích thước n và $\hat{q} = 1 - \hat{p}$, khoảng tin cậy $(1-\alpha)$ cho tham số p được xác định bởi

$$\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}} < p < \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}}$$

trong đó $z_{\alpha/2}$ là giá trị sao cho: $P\left(Z > z_{\alpha/2}\right) = \dfrac{\alpha}{2}$.

Chú ý.

+ Khi n nhỏ, tỷ lệ p chưa biết nhưng gần 0 hoặc 1, thì khoảng tin cậy được xác định sẽ không chính xác.

+ Tuy nhiên, nếu cả $n\hat{p}$ hay $n\hat{q}$ lớn hơn hoặc bằng 5 thì ta vẫn nhận được kết quả tốt.

Phương pháp xác định khoảng tin cậy cho tham số *p* cũng có thể áp dụng khi phân bố nhị thức được sử dụng để ước lượng phân bố siêu bội, nghĩa là khi n tương đối nhỏ so với N.

Ví dụ 1. Trong một mẫu ngẫu nhiên $N = 500$ gia đình có ti vi tại thành phố Hamilton, Canada, xác định được rằng $X = 340$ đăng ký HBO. Tìm khoảng tin cậy 95% cho tỷ lệ các gia đình trong thành phố này đăng ký sử dụng HBO.

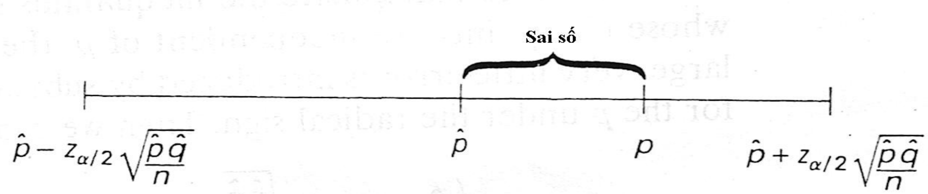
Giải. Ước lượng điểm của p là $\hat{p} = 340 / 500 = 0,68$.

+ Sử dụng Bảng A.3, chúng ta xác định được rằng $z_{0,25} = 1,96$, vì thế, khoảng tin cậy 95% cho p là

$$0.68 - 196\sqrt{\frac{(0.68)(0.32)}{500}} < p < 0.68 + 1.96\sqrt{\frac{(0.68)(0.32)}{500}}$$

tức là: $0,64 < p < 0,72$.

Nhận xét. Nếu *p* là giá trị tâm của khoảng tin cậy (1-α)%, khi đó $\hat{p}$ ước lượng p không có sai số. Tuy nhiên, thường thì $\hat{p}$ sẽ không chính xác bằng p và ước lượng điểm có sai số. Cỡ sai số này sẽ là sai số dương tách p và $\hat{p}$, và chúng ta có thể tin cậy (1-α)% rằng sai số này sẽ không vượt quá $z_{\alpha/2}\sqrt{\dfrac{\hat{p}\hat{q}}{n}}$.



Định lý 2. Nếu $\hat{p}$ là một ước lượng của p, với độ tin cậy (1-α) chúng ta có thể khẳng định rằng sai số của ước lượng không vượt quá $z_{\alpha/2}\sqrt{\dfrac{\hat{p}\hat{q}}{n}}$, tức là $\left|\hat{p}-p\right| \le z_{\alpha/2}\sqrt{\dfrac{\hat{p}\hat{q}}{n}}$.

Bây giờ cần xác định một mẫu cần có độ lớn là bao nhiêu để đảm bảo sai số $\left|\hat{p}-p\right|$ khi ước lượng p sẽ không nhỏ hơn giá trị *e*.

Định lý 3. Nếu $\hat{p}$ là ước lượng của *p* với độ tin cậy $1-\alpha$, sai số $\left|\hat{p}-p\right|$ sẽ nhỏ hơn giá trị xác định e khi kích thước mẫu gần bằng $n = z^2_{\alpha/2}\dfrac{\hat{p}\hat{q}}{e}$.

Ví dụ 2. Trong Ví dụ 1 **cần mẫu lớn bao nhiêu** nếu chúng ta muốn độ tin cậy 95% sao cho ước lượng $\left|\hat{p}-p\right|$ của p nằm trong khoảng 0,02?

Giải. Chúng ta coi 500 gia đình là một mẫu ngẫu nhiên có ước lượng điểm $\hat{p} = 0{,}68$. Khi đó theo Định lý 3:

$$n = \frac{(1.96)^2(0.68).(0.32)}{(0.02)^2} = 2090.$$

Vì thế, nếu căn cứ vào ước lượng của p trên một biến ngẫu nhiên có kích thước 2090, chúng ta có thể tin tưởng 95% rằng tỷ lệ mẫu sẽ không khác so với tỷ lệ chân thực một khoảng lớn hơn 0,02.

Ta vẫn có thể đưa ra cỡ của mẫu mà trong công thức tính không cần đến $\hat{q}$ như trong Định lý sau:

Định lý 4. Nếu $\hat{p}$ là một ước lượng của p với độ tin cậy $(1-\alpha)$, khi đó sai số $\left|\hat{p}-p\right| \le e$ khi kích thước mẫu là $n = \dfrac{z^2_{\alpha/2}}{4e^2}$.

Ví dụ 3. Trong Ví dụ 2 cần **môt mẫu lớn đến cỡ** nào nếu chúng ta muốn độ tin cậy ít nhất 95% và ước lượng $\left|\hat{p} - p\right|$ của chúng ta nằm trong khoảng 0,02?

Giải. Chúng ta có thể tin cậy 95% rằng tỷ lệ mẫu này sẽ không khác so với tỷ lệ chân thực lớn hơn 0,02 nếu chúng ta lựa chọn mẫu kích thước.

$$n = \frac{\left(1.96\right)^2}{\left(4\right)(0.02)^2} = 2401.$$

II. ƯỚC LƯỢNG HIỆU HAI TỶ LỆ

1. Bài toán: Xét hai tổng thể Ω_1 và Ω_2 mà mỗi phần tử trong các tổng thể đó đều có thể mang dấu hiệu $\Im$. Gọi $p_1; p_2$ tương ứng là tỷ lệ các cá thể mang dấu hiệu đang xét trong hai tổng thể đó. Ta sẽ tìm khoảng tin cậy $\left(1-\alpha\right)$ cho $p_1 - p_2$.

2. Khoảng tin cậy cho $p_1 - p_2$ khi cỡ mẫu lớn

Nếu $\hat{p}_1$và $\hat{p}_2$ là tỷ lệ thành công trong các mẫu ngẫu nhiên có cỡ n_1 và n_2 tương ứng. Đặt $\hat{q}_1 = 1 - \hat{p}_1$; $\hat{q}_2 = 1 - \hat{p}_2$. Khoảng tin cậy với độ tin cậy (1- α) cho sự sai khác $p_1 - p_2$ được xác định bằng công thức:

$$\left(\hat{p}_1 - \hat{p}_2\right) - z_{\alpha/2}\sqrt{\frac{\hat{p}_1\hat{q}_2}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_1}} < p_1 - p_2 < \left(\hat{p}_1 - \hat{p}_2\right) + z_{\alpha/2}\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}},$$

trong đó $z_{\alpha/2}$ là giá trị z sinh ra một diện tích $\alpha/2$ về bên phía phải, tức là $P\left(Z > z_{\alpha/2}\right) = \alpha/2$.

Chú ý. Nếu $\begin{cases} p_1 - p_2 \in (a,b) \\ a.b < 0 \end{cases}$ ta sẽ không so sánh tỷ lệ của tổng thể này với tỷ lệ của tổng thể kia.

Ví dụ 4. Xét sự thay đổi nhất định trong quá trình sản xuất.. Các mẫu thu được sử dụng cả quá trình mới và quá trình hiện tại để xác định liệu quá trình mới có hiệu quả hơn hay không. Giả sử 75 trong số 1500 linh kiện trong quy trình hiện tại được xác định có lỗi và 80 trong số 2000 linh kiện của quá trình mới được xác định có lỗi. Tìm khoảng tin cậy 90% cho sai số chân thực trong phần các sản phẩm bị lỗi giữa quá trình hiện tại và quá trình mới.

Giải. Lấy p_1 và p_2 là các tỷ lệ chân thực của các thiết bị có lỗi cho các quy trình hiện tại và mới tương ứng. Khi đó $\hat{p}_1 = \dfrac{75}{1500} = 0,05$; $\hat{p}_2 = \dfrac{80}{2000} = 0,04$ và ước lượng điểm p_1 - p_2 bằng.

$$\hat{p}_1 - \hat{p}_2 = 0,05 - 0,04 = 0,01$$

Sử dụng Bảng A.3, chúng ta xác định được $Z_{0,05} = 1,645$. Vì vậy thay vào công thức ta được khoảng tin cậy 90% là:

$$0.01 - 1.645\sqrt{\frac{(0.5)(0.95}{1500} + \frac{(0.04)(0.96)}{2000}} < p_1 - p_2 < 0.01 + 1.645\sqrt{\frac{(0.05)(0.95)}{1500} + \frac{(0.04)(0.96)}{2000}}$$

Rút gọn thành $-0,0017 < p_1 - p_2 < 0,0217$. Vì khoảng này chứa giá trị 0, cho nên quá trình mới không giảm nhiều về tỷ lệ các sản phẩm có lỗi so với phương pháp hiện đại.

III.ƯỚC LƯỢNG PHƯƠNG SAI.

1. Khoảng tin cậy cho σ²

Nếu s^2 là phương sai của một mẫu ngẫu nhiên kích thước n của một tổng thể với phân phối chuẩn, thì khoảng tin cậy $(1-\alpha)$ cho σ^2 là:

$$\frac{(n-1)\sigma^2}{X^2_{\alpha/2}} < \sigma^2 < \frac{(n-1)s^2}{X^2_{1-\alpha/2}}$$

trong đó $X^2_{\alpha/2}$ và $X^2_{1-\alpha/2}$ được xác định bởi $P\left(X^2 > X^2_{1-\alpha/2}\right) = 1 - \alpha/2$, $P\left(X^2 > X^2_{\alpha/2}\right) = \alpha/2$ và hỗ trợ bởi Bảng A.5.

2. Khoảng tin cậy cho δ_1^2 / δ_2^2

Nếu s_1^2 và s_2^2 là các phương sai của các mẫu độc lập có cỡ n_1 và n_2 từ các tổng thể có phân phối chuẩn, khi đó khoảng tin cậy $(1-\alpha)$ cho δ_1^2 / δ_2^2 là:

$$\frac{S_1^2}{S_2^2}\frac{1}{f_{\alpha/2}(\upsilon_1,\upsilon_2)} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{3,07^2}{0,80^2}(3,87)$$

trong đó $f_{\alpha/2}(\upsilon_1,\upsilon_2)$ là giá trị f với $\upsilon_1 = n_1 - 1, \upsilon_2 = n_2 - 1$ bậc tự do.

Về nhà:

Tự đọc: Mục Mục 9.13; 9.14

Bài tập: Tr. **299; 304; 316**

Đọc trước các Mục từ 10.1 đến 10.9 chuẩn bị cho **Bài số 11:**

Kiểm định giả thiết về trung bình