



Thống kê máy tính ứng dụng

Ước lượng - Kiểm định

Ngày 17 tháng 5 năm 2021

Bài 1. (Tham khảo từ [1]) (40 điểm)

Một nhóm sinh viên Khoa Công nghệ Thông tin, Đại học Công nghiệp, tổ chức khảo sát tỉ lệ sinh viên dùng Facebook thường xuyên ở trường Đại học Công nghiệp. Trong 625 sinh viên được khảo sát, tỉ lệ sinh viên dùng Facebook thường xuyên là 80%.

(a) Tìm biên độ sai số (margin of error) E tương ứng với độ tin cậy (confidence level) 95%.

(b) Tìm khoảng ước lượng có độ tin cậy 95% của tỉ lệ quần thể (population proportion)

(c) Từ Câu (b), có thể kết luận là phần lớn sinh viên ĐHCN có dùng Facebook thường xuyên được không?

(d) Viết một câu ngắn gọn trình bày kết quả khảo sát với tất cả các thông tin cần thiết.

Lời giải

(a)

$$E = z_{\alpha/2} \sqrt{\frac{\hat{p}\hat{q}}{n}} \quad (1)$$

Với độ tin cậy 95%, $\alpha = 0.05$, $z_{\alpha/2} = 1.96$. $\hat{p} = 80\% = 0.8$, $\hat{q} = 1 - \hat{p} = 0.2$. $n = 625$.

$$E = 1.96 \sqrt{\frac{0.8 \times 0.2}{625}} = 0.031 \quad (2)$$

(b) Khoảng ước lượng có độ tin cậy 95% của tỉ lệ quần thể là

$$\hat{p} - E < p < \hat{p} + E \quad (3)$$

or

$$0.769 < p < 0.831 \quad (4)$$

(c) Từ Câu (b), có thể kết luận là phần lớn sinh viên ĐHCN có dùng Facebook thường xuyên vì $p > 0.769 > 0.5$

(d) 80 % sinh viên trường Đại học Công nghiệp dùng Facebook thường xuyên. Tỉ lệ này dựa trên một khảo sát của một nhóm sinh viên Khoa Công nghệ Thông tin, Đại học Công nghiệp, khảo sát 625 sinh viên. Trên lý thuyết, trong 95 % khảo sát như vậy, tỉ lệ phần trăm này sẽ chỉ bị lệch nhiều nhất $\pm 3.1\%$ so với tỉ lệ phần trăm khi khảo sát tất cả sinh viên ĐHCN.

Bài 2. (Tham khảo từ [1]) (30 điểm)

Công ty Tiếp thị Markit muốn khảo sát tỉ lệ người trưởng thành ở Việt nam dùng Internet. Giả sử Markit muốn có mức tin cậy 95%. Markit cần khảo sát ít nhất bao nhiêu người để tỉ lệ phần trăm của mẫu có sai số không quá 3%, sử dụng các giả thuyết dưới đây

(a) Giả sử kết quả khảo sát năm 2019 là 70%

(b) Giả sử chúng ta không có thông tin gì về kết quả khảo sát.

p = tỉ lệ của quần thể

\hat{p} = tỉ lệ của mẫu

n = kích thước mẫu

E = biên độ lỗi mong muốn

$z_{\alpha/2} = z$ score tương ứng với diện tích $\alpha/2$ ở đuôi bên phải của phân phối chuẩn tắc.

Khi biết \hat{p} :

$$n = \frac{[z_{\alpha/2}]^2 \hat{p}\hat{q}}{E^2}$$

Khi không biết \hat{p} :

$$n = \frac{[z_{\alpha/2}]^2 0.25}{E^2}$$

(a)

$$\begin{aligned} n &= \frac{[z_{\alpha/2}]^2 \hat{p}\hat{q}}{E^2} = \frac{[1.96]^2 (0.7)(0.3)}{0.03^2} \\ &= 896.37 = 897 \end{aligned}$$

(b)

$$\begin{aligned} n &= \frac{[z_{\alpha/2}]^2 \cdot 0.25}{E^2} = \frac{[1.96]^2 \cdot 0.25}{0.03^2} \\ &= 1067.1111 = 1068 \end{aligned}$$

Bài 3. (Tham khảo từ [1]) (30 điểm)

Một nhóm bác sĩ ở Bệnh viện Hạnh phúc tuyên bố đã tìm ra một phương pháp điều trị (SBG) cho các cặp vợ chồng để tăng tỉ lệ sinh bé gái. Trong một lần thử nghiệm ở Bệnh viện Hạnh phúc có 17 bé gái trong tổng cộng 20 bé. Sử dụng mức ý nghĩa (significance level) là 0.05, có thể kết luận là xác suất sinh bé gái khi sử dụng phương pháp SBG khác 0.5 được không?

$H_0 : p = 0.5$ và $H_1 : p \neq 0.5$

Chúng ta sẽ giả sử giả thuyết H_0 đúng, với $p = 0.5$.

$\hat{p} = 17/20 = 0.85$.

Ta c $p = 0.5$, $\hat{p} = 0.85$, và $n = 20$, có thể tính được test statistic:

$$z = \frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}} = \frac{0.85 - 0.5}{\sqrt{\frac{(0.5)(0.5)}{20}}} = 3.13$$

Giá trị tới hạn $z_c = \pm 1.96$. Vì $z > z_c$, có thể kết luận là xác suất sinh bé gái khi sử dụng phương pháp SBG khác 0.5.

References

- [1] Mario F. Triola, Elementary Statistics, 11th Edition, Addison-Wesley, 2010