

CÁC PHƯƠNG PHÁP ĐỊNH LƯỢNG 1

LỜI GIẢI GỢI Ý BÀI TẬP 3

Ngày phát: 16/11/2023

Hạn nộp: 8h20, 21/11/2023

Bài làm được yêu cầu chỉ nộp bản điện tử trên Microsoft Teams

Câu 1: Giải thích ngắn cách bạn hiểu về các khái niệm hoặc trả lời ngắn cho các câu hỏi sau đây:

a. Giả định I.I.D cho một mẫu gồm 3 biến ngẫu nhiên

Phân phối của 3 biến ngẫu nhiên X_1, X_2, X_3 này phải độc lập và đồng nhất

- Độc lập: sự xảy ra của một biến này không có ảnh hưởng đến sự xảy ra của biến kia, nghĩa 3 cặp biến $(X_1, X_2), (X_1, X_3), (X_2, X_3)$ đều độc lập.
- Đồng nhất: Các biến ngẫu nhiên X_1, X_2, X_3 có cùng một phân phối xác suất, nghĩa là hàm phân phối xác suất của mỗi biến trong tập hợp là giống nhau.

b. Một thống kê (a statistic): là một hàm số bất kỳ của các biến ngẫu nhiên hình thành nên mẫu, hay $g(X_1, X_2, \dots, X_n)$, có thể hiểu đơn giản theo cách khác một thống kê (a statistic) là một cách lấy giá trị đại diện một bộ dữ liệu, ví dụ: trung bình, phương sai, độ lệch chuẩn.

c. Phân phối mẫu (sampling distribution): là hàm phân phối xác suất của một thống kê, ví dụ hàm phân phối xác suất của trung bình, hàm phân phối xác suất của phương sai.

d. Hai tính chất tạo nên tính tối ưu của ước lượng điểm là gì? Tại sao chúng ta cần có hai tính chất đó?

Hai tính chất tạo nên tính tối ưu của ước lượng điểm là:

- Không chệch: Nếu như ta lặp lại quá trình ước lượng nhiều lần, trung bình của ước lượng sẽ gần như hội tụ tại tham số chính xác của cả tổng thể.
- Hiệu quả toàn diện: Ước lượng điểm có phương sai nhỏ nhất trong tất cả các cách ước lượng khác.

Chúng ta cần có 2 tính chất này để chọn phương pháp tối ưu nhất trong rất nhiều (hay vô số) các phương pháp ước lượng điểm khác nhau.

Câu 2: Một công ty sản xuất nước giải khát muốn kiểm tra sự ổn định dung tích thực của chai sản xuất ra thị trường trong lô sản xuất gần nhất. Dung tích ghi trên bao bì đưa ra thị trường là 2 lít. Công ty lấy một mẫu ngẫu nhiên gồm 100 chai và đo dung tích. Mẫu này cho trung bình mẫu là 1.99 lít. Số liệu quan sát cho thấy độ lệch chuẩn tổng thể là 0.05 lít. Hãy trả lời những câu hỏi sau:

a. Ước lượng khoảng với độ tin cậy 95% cho trung bình tổng thể của lô này.

Từ đề bài, ta có: $\sigma = 0.05$ và $n = 100$

Với $\alpha = 5\% \rightarrow Z_{\frac{\alpha}{2}} = NORM.S.INV(0.975,1) = 1.96$

Khoảng tin cậy 95% của dung tích trung bình lô sản phẩm này nằm trong khoảng:

$$\mu \pm Z_{\frac{\alpha}{2}} \times \sqrt{\frac{\sigma^2}{n}} = 1.99 \pm 1.96 \times \sqrt{\frac{0.05^2}{100}} = [1.9802; 1.9998] \text{ (lít)}$$

b. Ước lượng với độ tin cậy 99%, khoảng ước lượng thay đổi thế nào?

Với $\alpha = 1\% \rightarrow Z_{\frac{\alpha}{2}} = NORM.S.INV(0.995,1) = 2.57583$

Khoảng tin cậy 99% của dung tích trung bình lô sản phẩm này nằm trong khoảng:

$$\mu \pm Z_{\frac{\alpha}{2}} \times \sqrt{\frac{\sigma^2}{n}} = 1.99 \pm 2.57583 \times \sqrt{\frac{0.05^2}{100}} = [1.97712; 2.00288] \text{ (lít)}$$

→ Khoảng ước lượng 99% sẽ rộng hơn khoảng ước lượng 95%

c (*). Một chai dung tích 2.02 lít có phải khác thường so với các khoảng tin cậy bạn tính toán hay không?

Với trường hợp trung bình tổng thể rơi vào mức giới hạn giới dưới là 1.98 lít (độ tin cậy 95%), thì quan sát 2.02 lít cách trung bình tổng thể là 0.4 lít. 0.4 lít là nhỏ hơn MỘT độ lệch chuẩn tổng thể (bằng 0.5 lít). Như vậy rất ít khả năng dung tích này nằm quá xa khỏi trung bình của phân phối. Giả định là phân phối tổng thể là phân phối chuẩn, thì mức 2.02 lít vẫn nằm trong trong khu vực xảy ra 68.2% của 1 độ lệch chuẩn. Nói ngắn gọn, quan sát 2.02 lít không thể là quan sát khác thường.

Câu 3: Cho một mẫu 15 bạn sinh viên nam có cân nặng (kg) như sau:

65, 67, 72, 80, 62, 77, 71, 68, 75, 64, 70, 85, 67, 73, 71

a. Hãy ước lượng khoảng với độ tin cậy 95% cho trung bình tổng thể.

Với tập mẫu của đề bài, ta tính được các giá trị của mẫu như sau:

Giá trị trung bình về cân nặng $\mu = 71.13333$ và độ lệch chuẩn cân nặng $\sigma^2 = 38.69524$

Với $\alpha = 5\% \rightarrow t_{\frac{\alpha}{2},df} = t_{0.025,14} = T.INV(0.975,14) = 2.144$

Khoảng tin cậy 95% cân nặng của trung bình tổng thể sinh viên nam nằm trong khoảng:

$$\mu \pm t_{0.025,14} \times \sqrt{\frac{\sigma^2}{n}} = 71.13333 \pm 2.144 \times \sqrt{\frac{38.69524}{15}} = [67.688; 74.578] \text{ (kg)}$$

b. Chúng ta có cần giả định phân phối chuẩn cho cân nặng của tổng thể để ước lượng khoảng hay không?

Chúng ta không cần phải giả định phân phối chuẩn cho cân nặng cho tổng thể để ước lượng khoảng vì theo định lý giới hạn trung tâm phân phối nào khi lấy cỡ mẫu đủ lớn cũng sẽ xấp xỉ phân phối chuẩn.

c. Giải thích cách hiểu của bạn về ước lượng khoảng cho giá trị trung bình của tổng thể nhìn từ góc độ xác suất?

Nhìn từ góc độ xác suất, khi lặp đi lặp lại việc lấy mẫu quan sát nhiều lần thì 95% giá trị trung bình của tổng thể sẽ nằm trong khoảng ước lượng tính từ mẫu quan sát.

---HẾT---