Chương 4: ĐIỀU TRA CHỌN MẪU

Khái niệm

Điều tra chọn mẫu là một điều tra không toàn bộ, người ta chỉ chọn ra một số đơn vị từ tổng thể chung để điều tra thực tế, rồi sau đó bằng các phương pháp khoa học, tính toán suy rộng cho toàn bộ tổng thể.

Trong điều tra chọn mẫu cầu quan tâm đặc biệt 2 vấn đề:

- Quy tắc lựa chọn sao cho các đơn vị có thể đại diện cho toàn bộ tổng thể.
- Dùng công thức suy rộng thành các đặc điểm của tổng thể.

Ta có hai hướng lấy mẫu:

- Lấy mẫu theo xác suất
- Lấy mẫu phi xác suất.

4.1. Chọn mẫu ngẫu nhiên đơn giản

A. Chọn mẫu từ tổng thể hữu hạn

Một mẫu ngẫu nhiên đơn giản cỡ n được chọn từ một tổng thể hữu hạn quy mô N là một mẫu được chọn sao cho mỗi phần tử của tổng thể được chọn vào mẫu với cơ hội (xác suất) như nhau.

Ưu điểm: cách làm đơn giản, tính đại diện cao.

Nhược điểm:

- cần có khung mẫu.
- Các phần tử được chọn có thể phân bố không đều, do vậy gây tốn kém và mất thời gian cho việc thu thập dữ liệu.

Cách làm

Bước 1: tạo khung mẫu

- đánh số thứ tự tất cả các phần tử của tổng thể .

Bước 2: chọn ngẫu nhiên các phần trong danh sách.

- Dùng bảng số ngẫu nhiên: ta có thể tạo bằng chương trình Excel, SPSS.
- Dùng máy tính cầm tay: lấy ngẫu nhiên một số bất kì tứ 1 tới n: Alpha ⇒ RanInt (1,n).
- Bốc thăm.

Ví dụ 72

Cho danh sách 25 đội bóng hàng đầu cho mùa giải 2002 theo NCAA News (ngày 04/01/2003).

1. Ohio State
2. Miami
3. Georgia
4. Southern California
5. Oklahoma
6. Kansas State
7. Texas
8. Iowa
9. Michigan
10. Washington Sate
11. North Carolina State
12. Boise State
13. Maryland
14. Virginia Tech
15. Penn State
16. Auburn
17. Notre Dame
18. Pittsburgh
19. Marshall
20. West Vifginia
21. Colorado
22. TCU
23. Florida State
24. Florida
25. Virginia

Hãy chọn một mẫu *ngẫu nhiên* đơn giản gồm 6 đội bóng trong các độ bóng trên.

B. Chọn mẫu từ tổng thể vô hạn

Một mẫu ngẫu nhiên đơn giản từ một tổng thể vô hạn là một mẫu được chọn sao cho thỏa các điều kiện sau

1. Mỗi phần tử được chọn đều thuộc tổng thể.
2. Mỗi phần tử được chọn một cách độc lập.

Chú ý: Đối với tổng thể vô hạn, quy trình chọn mẫu cần được thiết kế đặc biệt cho từng tình huống cụ thể để chọn các phần tử 1 cách độc lập và do đó tránh được các xu hướng lựa chọn những phần tử đã biết nào đó với xác suất lựa chọn cao hơn.

Ví dụ 73

Một cửa hàng KFC muốn lấy ý kiến khảo sát khách hàng.

Chọn ngẫu nhiên đơn giản các khách hàng tới của hàng dùng bữa, khi đó điều kiện đầu tiên thỏa.

Ta xét đến diều kiện thứ 2. Nếu chọn 3 khách hàng liên tiếp ta có thể gặp trường hợp các khách hàng đó di chung với nhau, khi đó các khách hàng không độc lập sẽ ảnh hưởng tới kết quả khảo sát.

Ta có thể thực hiện khảo sát bằng cách, bất kì khách hàng nào sử dụng thẻ thành viên KFC tới của hàng, khách hàng tiếp theo sẽ được chọn để khảo sát. Do khách hàng sử dụng thể thành viên là ngẫu nhiên và độc lập, cách lấy mẫu này đảm bảo các khách hàng được chọn độc lập.

4.2. Chọn mẫu hệ thống

Trong mẫu hệ thống, các phần tử được chọn theo một khoảng cách đều từ khung mẫu.

Ưu điểm: Mẫu được phân bố đồng đều trong khung mẫu, tính toán nhanh, đơn giản.

Nhược điểm: phần tử mẫu không xếp ngẫu nhiên, thiếu đại diện.

Giả sử tổng thể có N phần tử, ta cần lấy mẫu cỡ n.

A. Lấy mẫu hệ thống

Cách làm

Bước 1: đánh số thứ tự tất cả các phần tử của tổng thể.

Bước 2: xác định khoảng cách chọn mẫu : $k = \dfrac{N}{n}$.

Bước 3:

- với k phần tử đầu tiên của tổng thể, ta lấy mẫu ngẫu nhiên đơn giản chọn ra phần tử thứ nhất của mẫu, $u_1 \in \{1, ..., k\}$

- từ phần tử thứ 2 của mẫu, ta lấy theo công thức $u_i = u_1 + k(i-1)$ với $i = 2, 3, ...$

Ví dụ 74

Lấy mẫu 10 phần tử từ tổng thể gồm 76 phần tử.

Giải.

- Khoảng cách mẫu: $k = \dfrac{76}{10} = 7,6 \Rightarrow$ lấy k =8.

- Chọn ngẫu nhiên một số từ 1 tới 8.

Nếu lấy được số 1, thì các phần tử được chọn là:
1, 9, 17, 25, 33, 41, 49, 57, 65, 73. ⇒ chọn được mẫu n =10.

Nếu lấy được số 7, thì các phần tử được chọn là:
7, 15, 23, 31, 39, 47, 55, 63, 71. ⇒ Ta chọn được mẫu n =9. !!!

Nhận xét: ta thấy rằng với cách làm này, khi N không chia hết cho n thì các phần tử không có cùng 1 xác suất được chọn như nhau, điều này dẫn tới việc khi dùng trung bình mẫu để ước lượng trung bình tổng thể thì rất có khả năng bị chệch.

B. Chọn mẫu hệ thống quay vòng

Cách làm

Bước 1: đánh số thứ tự tất cả các phần tử của tổng thể.

Bước 2: xác định khoảng cách chọn mẫu : $k = \dfrac{N}{n}$

Bước 3: chọn ngẫu nhiên một phần tử trong N phần tử của tổng thể, đây là phần tử đầu tiên của mẫu. ký hiệu $u_1 \in \{1, ..., N\}$.

Bước 4: các phần tử tiếp theo được tính bằng công thức $u_i = u_1 + k.(i-1)$ trong đó $i = 2, 3..., n$

- nếu $u_i \leq N \Rightarrow$ lấy u_i vào mẫu.
- nếu $u_i > N \Rightarrow$ lấy $v_i = u_i - N$ vào mẫu.

Ví dụ 75

Lấy mẫu cỡ n = 4, từ tổng thể 15 phần tử.

Giải.

- Khoảng cách chọn mẫu: $k = \dfrac{15}{4} = 3,75 \Rightarrow$ chọn k = 4.
- Giả sử chọn ngẫu nhiên được $u_1 = 7$, ta được
- $u_2 = 7 + 4 = 11 \leq 15$, lấy 11.
- $u_3 = 7 + 2.4 = 15 \leq 15$, lấy 15.
- $u_4 = 7 + 3.4 = 19 > 15$, lấy $19 - 15 = 4$.

Vậy ta có mẫu gồm các phần tử thứ: 7, 11, 15, 4 của tổng thể.

4.3. Chọn mẫu phân tầng

Chọn mẫu ngẫu nhiên phân tầng được thực hiện dựa trên việc chia tổng thể thành các nhóm riêng biệt, các nhóm riêng biệt này được xem như các tổng thể nhỏ, và các mẫu sẽ được chọn gẫu nhiên từ các tổng thể nhỏ.
Mẫu phân tầng được sử dụng khi:
- Các phần tử được chia thành nhiều lớp.
- Mỗi lớp có tính chất đặc trưng đồng nhất.
- Có thể thực hiện lấy mẫu ngẫu nhiên cho từng tầng.

Ưu điểm: Có thể biết được hình ảnh của từng tầng
Nhược điểm: cần thiết lập khung mẫu chi tiết của từng tầng.
Chú ý: đối với từng mục đích nghiên cứu, ta sẽ có cách phân tầng khác nhau.

Giả sử ta cần lấy n phần tử mẫu từ tổng thể gồm N phần tử.
A. Phân bổ mẫu đều
Cách làm
Bước 1: chia tổng thể thành k tầng.
Bước 2: từ mỗi tầng, ta chọn ra $\frac{n}{k}$ phần tử.
B. Phân bổ mẫu theo tỉ lệ
Cách làm
Bước 1: chia tổng thể thành k tầng, tầng thứ i có N_i phần tử.
Bước 2: số phần tử chọn ra từ tầng i là $n_i = \frac{n}{N}.N_i$.
Chú ý: trong bước 2 , ta chọn các phần tử bằng cách lấy mẫu ngẫu nhiên đơn giản, hoặc hệ thống.

Ví dụ 76

Một trường đại học có 20 000 SV thuộc 5 hệ đào tạo được cho trong bảng sau.

Hệ đào tạo	Số SV
Cử nhân chính quy	10 000
Cử nhân hệ hoàn chỉnh đại học	2 000
Cử nhân hệ văn bằng 2	2 000
Cử nhân tại chức	5 000
Cao học	1 000

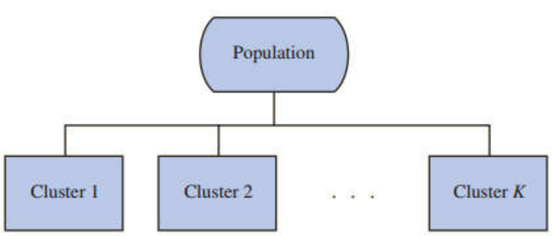
Mỗi hệ đào tạo được coi như một tầng. Hãy chọn một mẫu gồm 1000 SV để tham gia khảo sát.

Giải.
Cách 1.
Chọn 1 000 SV từ 5 hệ đào tạo, mỗi hệ chọn $\frac{1000}{5} = 200$ SV.
Cách 2.
Chọn theo tỉ lệ, ta có số sinh viên mỗi hệ đào tạo được chọn cho bởi bảng sau

Hệ đào tạo	Số SV	Số SV được chọn ($\frac{n}{N}.N_i$)
Cử nhân chính quy	10 000	500
Cử nhân hệ hoàn chỉnh đại học	2 000	100
Cử nhân hệ văn bằng 2	2 000	100
Cử nhân tại chức	5 000	250
Cao học	1 000	50
	20 000	1000

4.4. Chọn mẫu cụm

Lấy mã cụm (Cluster Sampling) là một phương pháp lấy mẫu theo xác suất mà:
- Các phần tử trong tổng thể được chia thành các nhóm riêng biệt gọi là cụm hay khối. Mỗi phần tử trong tổng thể chỉ thuộc duy nhất một cụm.
- Sau đó một mẫu ngẫu nhiên đơn giản các cụm được lấy ra. Tất cả các phần tử trong mỗi cụm mẫu tạo thành mẫu.



Việc lấy mẫu cụm có khuynh hướng tốt nhất khi mỗi cụm là một đại diện thu nhỏ của tổng thể..
Một trong những ứng dụng chính của lấy mẫu cụm là lấy mẫu theo khu vực, diện tích, vùng, trong đó cáccụm chính là các khối nhà trong thành phố hay các khu vực được xác định khác.
Nói chung lấy mẫu cụm đòi hỏi một cỡ mẫu lớn hơn lấy mẫu ng6a4u nhiên đơn giản hay phân tầng. Tuy nhiên nó có thể tiết kiệm chi phí vì thực tế là khi người phỏng vấn đến 1 cụm mẫu (ví dụ 1 tòa nhà), nhiều quan sát mẫu có thể thu được trong khoảng thời gian ngắn.

4.5. Lấy mẫu thuận tiện

Lấy mẫu thuận tiện là một kĩ thuật lấy mẫu phi xác suất. Mẫu được xác định chủ yếu bởi sự thuận tiện. Các phần tử trong mẫu không được biết hay xác định trước về việc được chọn.

Ví dụ, một giáo sư đang nghiên cứu tại một trường đại học có thể sử dụng sinh viên tình nguyện để tạo một mẫu đơn giản bởi họ sẵn lòng tham gia với ít hoặc không tốn chi phí.

Mẫu thuận tiện có lợi thế lựa chọn mẫu và thu thập dữ liệu tương đối dễ dàng, tuy nhiên không thể đánh giá mẫu có đại diện tổng thể là tốt hay không, và không có một phương pháp nào phân tích về chất lượng của các kết quả từ mẫu.

4.6. Lấy mẫu phán đoán

Lấy mẫu phán đoán là kỹ thuật lấy mẫu phi xác suất. Theo phương pháp này, người am hiểu nhất về lĩnh vực nghiên cứu sẽ chọn các phần tử của tổng thể mà người đó cảm thấy đại diện tốt nhất cho mẫu.

Ví dụ, một phón viên có thể chọn phỏng vấn hai hoặc 3 đại biểu quốc hội với đánh giá rằng các đại biểu này phản ánh quan điểm chung của quốc hội.

Tuy nhiên, chất lượng mẫu phụ thuộc vào sự đánh giá, kinh nghiệm của người chọn mẫu.

4.7. Kích thước mẫu

Gọi θ là đại lượng cần ước lượng $\theta \in \{\mu, \sigma^2, p\}$.
Ta cần tính

$$P(\theta \in (a, b)) = \gamma$$

trong đó
- (a, b) khoảng tin cậy.
- $P(\theta \in (a, b))$ xác suất để θ nằm trong đoạn (a, b).
- γ độ tin cậy.
- $\alpha = 1 - \gamma$ mức ý nghĩa.

Ước lượng khoảng cho trung bình μ

Giả thuyết: Cho cỡ mẫu n, tính được $\overline{x}$ và s.
Cho độ tin cậy γ(hoặc mức ý nghĩa α).
Mục tiêu: Tìm sai số ước lượng ε sao cho
$\mu \in (\overline{x} - \varepsilon, \overline{x} + \varepsilon)$: khoảng tin cậy đối xứng
$\mu \in (-\infty, \overline{x} + \varepsilon)$: khoảng tin cậy tối đa.
$\mu \in (\overline{x} - \varepsilon, +\infty)$: khoảng tin cậy tối thiểu.

Phương pháp
Bước 1. xác định σ^2 và n.
Bước 2. tìm C

	$\begin{cases} \left[\begin{array}{l} \sigma^2 \text{ biết} \\ \sigma^2 \text{ chưa biết} \end{array}\right. \\ n \geq 30. \end{cases}$	$\begin{cases} \sigma^2 \text{ chưa biết} \\ n < 30. \end{cases}$
ƯLTB ĐX	$\varphi(C) = \frac{\gamma}{2} \Rightarrow C$	$C = t\left(n-1, \frac{\alpha}{2}\right)$
ƯLTB TĐ	$\varphi(C) = 0,5 - \alpha \Rightarrow C$	$C = t\left(n-1, \alpha\right)$
ƯLTB TT	$\varphi(C) = 0,5 - \alpha \Rightarrow C$	$C = t\left(n-1, \alpha\right)$

Bước 3. Tìm độ chính xác

$$\varepsilon = \frac{C.\square}{\sqrt{n}}$$

trong đó
- $\square = \sigma$ nếu σ biết.
- $\square = s$ nếu σ chưa biết.

Bước 4. $\mu \in (\quad , \quad)$

Một số bài toán thường gặp

$$\varepsilon = \frac{C.\square}{\sqrt{n}}$$

Bài toán 1. Biết $\gamma = 1 - \alpha$ và n. Tìm ε.
Bài toán 2. Biết ε, n tìm $\gamma = 1 - \alpha$.
Bài toán 3. Biết $1 - \alpha$ và ε. Tìm kích thước mẫu cần điều tra thêm.

$$n = \left\lceil \frac{C^2.s^2}{\varepsilon^2} \right\rceil + 1$$

trong đó $\lceil x \rceil$: phần nguyên trên của x. n là kích thước mẫu cần.

Ví dụ 77

Nên chọn cỡ mẫu bằng bao nhiêu để có một khoảng tin cậy 95% với sai số biên 10. Giả sửa độ lệch chuẩn của tổng thể bằng 40.

Ví dụ 78

Phạm vi biến thiên của tập dữ liệu được ước lượng là 36.
1. Giá trị sơ khởi của độ lệch chuẩn tổng thể bằng bao nhiêu? (36/4)
2. Với độ tin cậy 95%, cỡ mẫu là bao nhiêu để sai số biên là 3. (35)
3. Với độ tin cậy 95%, cỡ mẫu là bao nhiêu để sai số biên là 2. (78)

Ví dụ 79

Giá trung bình của 1 galong xăng không chì ở vùng Cincinnati được báo cáo là $2,41$ USD. Trong suốt những kỳ giá thay đổi rất nhanh, tờ báo lấy mẫu các trạm phục vụ và chuẩn bị báo cáo về giá xăng thường xuyên. Giả sử độ lệch chuẩn là $0,15$ USD cho giá một galong xăng không chì thông thường, hãy đề xuất cỡ mẫu thích hợp cho tờ báo nếu họ mong muốn độ tin cậy 95% và
1. sai số biên là $0,07$ USD. (18)
2. sai số biên là $0,05$ USD. (35)
3. sai số biên là $0,03$ USD. (97)

Ví dụ 80

Nghiên cứu của Smith Travel cung cấp thông tin về giá một đêm phòng khách sạn trên nước Mỹ. Sử dụng sai số biên mong muốn 2 USD và giá trị $22,50$USD cho độ lệch chuẩn của tổng thể để tìm cỡ mẫu đề xuất trong các câu sau
1. Ước lượng khoảng tin cậy 90% của trung bình tổng thể giá phòng khách sạn. (343)
2. Ước lượng khoảng tin cậy 95% của trung bình tổng thể giá phòng khách sạn. (487)
3. Ước lượng khoảng tin cậy 99% của trung bình tổng thể giá phòng khách sạn. (840)
4. Khi sai số biên mong muốn là cố định, điều gì xãy ra tới cỡ mẫu khi độ tin cậy tăng. Bạn sẽ đề xuất độ tin cậy 99% cho nghiên cứu của Smith Travel? Giải thích?

Ước lượng tỉ lệ p

Giả thuyết. cho cỡ mẫu n,
tỉ lệ mẫu $f = \frac{m}{n}$, m số phần tử có tính chất A,
độ tin cậy γ.
Mục tiêu: Tìm sai số ước lượng ε sao cho
$p \in (f - \varepsilon, f + \varepsilon)$: khoảng tin cậy đối xứng
$p \in (-\infty, f + \varepsilon)$: khoảng tin cậy tối đa.
$p \in (f - \varepsilon, +\infty)$: khoảng tin cậy tối thiểu.

Phương pháp
Bước 1. xác định n và f.
Bước 2. tìm C

ƯLTL ĐX	$\varphi(C) = \frac{\gamma}{2} \Rightarrow C$
ƯLTL TĐ	$\varphi(C) = 0,5 - \alpha \Rightarrow C$
ƯLTL TT	$\varphi(C) = 0,5 - \alpha \Rightarrow C$

Bước 3. Tìm độ chính xác

$$\varepsilon = C.\sqrt{\frac{f(1-f)}{n}}$$

Bước 4. $p \in (\quad , \quad)$

Một số bài toán thường gặp

$$\varepsilon = C.\sqrt{\frac{f(1-f)}{n}}$$

Bài toán 1. Biết $\gamma = 1 - \alpha$ và n. Tìm ε.
Bài toán 2. Bài toán ngược. Biết ε, n tìm $\gamma = 1 - \alpha$.
Bài toán 3. Biết $1 - \alpha$ và ε. Tìm kích thước mẫu cần điều tra thêm.

$$n = \left\lceil f(1-f).\frac{C^2}{\varepsilon^2} \right\rceil + 1$$

trong đó $\lceil x \rceil$: phần nguyên trên của x. n kích thước mẫu cần điều tra.

Ví dụ 81

Trong một cuộc khảo sát, giá trị sơ khởi của tỉ lệ tổng thể là $p = 0,35$, với độ tin cậy 95% và sai số biên là $0,05$, ta nên chọn cỡ mẫu là bao nhiêu.

Ví dụ 82

Với độ tin cậy 95%, ta nên lấy mẫu có kích thước bao nhiêu để thu được một sai số biên là $0,03$ cho ước lượng tổng thể. Giả sử rằn dữ liệu quá khứ không có hiệu lực để xây dựng giá trị sơ khởi của p. (1068)

Ví dụ 83

Tỉ lệ phần trăm người không được bảo vệ bởi bảo hiểm sức khỏe trong năm 2003 là $15,6\%$. Một uỷ ban quốc hội Mỹ chịu trách nhiệm thực hiện khảo sát mẫu để thu thập thêm thông tin hiện tại.
1. Bạn khuyến cáo cỡ mẫu là bao nhiêu nếu mục tiêu của ủy ban là ước lượng tỉ lệ các ca1 nhân không được bảo hiểm sức khỏe với sai số biên là $0,03$? Sử dụng độ tin cậy 95%.
2. Tương tự câu 1, với độ tin cậy 99%.

Ví dụ 84

Chủ một kho cung cấp sơn muốn ước lượng lượng sơn chứa trong một thùng được sản xuất từ một dây chuyền công nghệ quốc gia. Biết rằng theo tiêu chuẩn của dây chuyền công nghệ đó, độ lệch tiêu chuẩn của lượng sơn là 0,08 thùng. Điều tra một mẫu 50 thùng được lượng sơn trung bình là 0,97 thùng. Với độ tin cậy 99% hãy ước ượng.
1. Lượng sơn trung bình chứa trong 1 thùng.
2. Lượng sơn trung bình tối thiểu chứa trong 1 thùng.
3. Nếu chủ kho muốn ước lượng sơn trung bình trong thùng đảm bảo độ tin cậy 99% và độ chính xác 2% thì cần điều tra thêm bao nhiêu thùng nữa.

Giải

1. $\overline{x} =$ $\sigma =$ $n =$

Gọi μ (thùng) là lượng sơn trung bình chứa trong 1 thùng.

$\gamma =$ $\Rightarrow \alpha =$

$\varphi(C) =$ $\Rightarrow C =$

$\Rightarrow \varepsilon =$

$\Rightarrow \mu \in$

2. $\varphi(C) = 0,5 - \alpha =$

$\Rightarrow C =$

$\Rightarrow \varepsilon =$

$\Rightarrow \mu \in$

3. $\varphi(C) = \frac{\gamma}{2} =$

$\Rightarrow C =$

$\Rightarrow n =$

Ví dụ 85

Giá bán của một loại thiết bị (USD) trên thị trường là biến ngẫu nhiên có phân phối chuẩn. Một người định mua loại thiết bị này, khảo sát ngẫu nhiên tại 8 của hàng được bán giá trung bình 137,75 USD với độ lệch chuẩn 7,98 USD. Với độ tin cậy 90%, hãy ước lượng giá bán trung bình của thiết bị.

Giải

$\overline{x} =$ $s =$ $n =$

$\gamma =$ $\Rightarrow \alpha =$

$C = t\left(n-1, \frac{\alpha}{2}\right) =$

$\Rightarrow \varepsilon =$

$\Rightarrow \mu \in (\quad , \quad)$

Ví dụ 86

Cho mẫu 25 cá thể, đem cân ta có các kết quả

0.63	0.65	0.73	0.60	0.65	0.77	0.82	0.64	0.66
0.72	0.79	0.74	0.45	0.66	0.52	0.62	0.76	0.66
0.84	0.87	0.76	0.54	0.64	0.75	0.77		

1. Tính kỳ vọng, phương sai mẫu.
2. Tìm khoảng tin cậy của cân nặng trung bình với độ tin cậy 95%.
3. Để có độ chính xác 0.01 cần có cỡ mẫu bao nhiêu.

Ví dụ 87

Quan sát ngẫu nhiên 200 lọ thuốc trong một lô hàng lớn, ta thấy có 17 lọ không đạt tiêu chuẩn. Hãy ước lượng tỉ lệ thuốc không đạt chuẩn ở độ tin cậy 95%.

Giải

- $f =$
- $\varphi(C) = \frac{\gamma}{2} = 0,475 \Rightarrow C$
- ε
- $p \in$

Ví dụ 88

Lấy ngẫu nhiên 200 sản phẩm trong một kho hàng thấy có 25 phế phẩm.

a. Nếu muốn độ chính xác của ước lượng là $\varepsilon = 0.035$ thì độ tin cậy của ước lượng là bao nhiêu?

b. Nếu muốn độ chính xác là 0.001, độ tin cậy 95% thì cần kiểm tra thêm bao nhiêu sản phẩm.

Giải

a. $n = 200; \quad f = \frac{25}{200} = 0.125; \quad \varepsilon = 0.035$

$$C = (0.035).\sqrt{\frac{200}{0.125\,(1-0.125)}} = 1.50$$

$$\frac{\gamma}{2} = \varphi(1,50) = 0.4332 \Rightarrow \gamma = 0.8664$$

b. $f = 0.125 \quad ; \varepsilon = 0.001 \quad ; \gamma = 1 - \alpha = 0.95$

$$\Rightarrow C = \varphi(\frac{0.95}{2}) = 1.96$$

$$\Rightarrow n = \left[(0.125)\,(1-0.125)\,.\frac{1.96^2}{0.001^2}\right] + 1 = 420175 + 1$$