

CHƯƠNG VIII: ĐIỀU TRA CHỌN MẪU

I

NHỮNG VẤN ĐỀ
CHUNG VỀ ĐIỀU
TRA CHỌN MẪU

II

ĐIỀU TRA CHỌN
MẪU NGẪU NHIÊN

I. Những vấn đề chung

1

Khái niệm

2

Ưu điểm

3

Hạn chế

4

Trường hợp vận dụng

Khái niệm

ĐTCM: là loại điều tra không toàn bộ trong đó người ta chọn ra một số đủ lớn các đơn vị thuộc đối tượng nghiên cứu để tiến hành điều tra thực tế. Các đơn vị này được chọn theo những quy tắc nhất định để đảm bảo tính đại biểu và kết quả của ĐTCM thường được dùng để suy rộng (bình quân hoặc tỷ lệ) cho tổng thể chung.

Suy rộng (ước lượng): từ các tham số (mức độ) tính toán được trên các đơn vị điều tra suy ra các tham số tương ứng của toàn bộ hiện tượng.

Ưu điểm

- + Tiết kiệm (chi phí, nhân lực)
- + Có phép mở rộng nội dung điều tra
- + Tài liệu thu được trên mẫu có độ chính xác cao
- + Nhanh gọn, đảm bảo tính kịp thời

Hạn chế

- + Không cho biết thông tin đầy đủ về tổng thể
- + Sai số khi suy rộng
- + Kết quả điều tra không thể tiến hành phân tổ theo mọi phạm vi nghiên cứu

Trường hợp vận dụng

- Thay thế cho điều tra toàn bộ
- Kết hợp với điều tra toàn bộ để mở rộng nội dung và tổng hợp nhanh kết quả điều tra
- So sánh các hiện tượng hoặc đưa ra một nhận định nhưng chưa có tài liệu cụ thể (kiểm định giả thiết thống kê)

II. Điều tra chọn mẫu ngẫu nhiên

1

Tổng thể chung và tổng thể mẫu

2

Cách chọn

3

Sai số trong điều tra chọn mẫu

4

Ước lượng (suy rộng) kết quả điều tra

5

Xác định kích thước (quy mô) mẫu

1. Tổng thể chung và tổng thể mẫu

***KN**: - TTC: bao gồm toàn bộ các đơn vị thuộc đối tượng điều tra.
- TTM: bao gồm một số đơn vị nhất định được chọn ra từ TTC để tiến hành điều tra thực tế.

Kí hiệu	Tổng thể chung	Tổng thể mẫu
Quy mô	N	n
Tham số cần ước lượng	θ	θ'
Số bình quân	μ	$\bar{X}$
Tỷ lệ theo một tiêu thức	p	f
Phương sai của số BQ	$\sigma^2 = \overline{X^2} - (\mu)^2$	$\sigma_0^2 = \overline{X^2} - (\bar{X})^2$
Phương sai của tỷ lệ	$p(1 - p)$	$f(1 - f)$

2. Cách chọn

- Chọn hoàn lại (chọn lặp, chọn nhiều lần): Phần tử đã được chọn thì vẫn có thể được chọn lại ở những lần sau. Quy mô tổng thể không đổi

$$\text{Số mẫu có thể có: } k = \overline{A}_N^n = N^n$$

- Chọn không hoàn lại (chọn không lặp, chọn một lần): Phần tử đã được chọn thì không thể chọn lại ở những lần sau. Quy mô tổng thể giảm

$$\text{Số mẫu có thể có: } k = C_N^n = \frac{N!}{n!(N-n)!}$$

3. Sai số trong ĐTCM

Công thức chung xác định các tham số:

$$\theta = \theta' \pm \varepsilon_M \pm \varepsilon_{0M}$$

Trong đó:

θ - các tham số của TTC (parameters)

θ' - các thống kê của TTM (statistics)

ε_M - sai số chọn mẫu (sai số do tính đại biểu): là sai số xảy ra khi các đơn vị mẫu được chọn hoàn toàn ngẫu nhiên

ε_{0M} - sai số phi chọn mẫu: là sai số do vi phạm nguyên tắc chọn mẫu ngẫu nhiên, do số lượng đơn vị mẫu không đủ lớn, sai số do đăng ký, ghi chép.

Sai số bình quân chọn mẫu

Cách chọn Suy rộng		Hoàn lại (chọn nhiều lần)	Không hoàn lại (chọn một lần)
Số bình quân	Biết σ^2	$\sigma_{\bar{x}} = \sqrt{\frac{\sigma^2}{n}}$	$\sigma_{\bar{x}} = \sqrt{\frac{\sigma^2}{n} \left(1 - \frac{n}{N}\right)}$
	Không biết σ^2	$\sigma_{\bar{x}} = \sqrt{\frac{\sigma_0^2}{n-1}}$	$\sigma_{\bar{x}} = \sqrt{\frac{\sigma_0^2}{n-1} \left(1 - \frac{n}{N}\right)}$
Tỷ lệ	Biết p	$\sigma_f = \sqrt{\frac{p(1-p)}{n}}$	$\sigma_f = \sqrt{\frac{p(1-p)}{n} \left(1 - \frac{n}{N}\right)}$
	Không biết p	$\sigma_f = \sqrt{\frac{f(1-f)}{n-1}}$	$\sigma_f = \sqrt{\frac{f(1-f)}{n-1} \left(1 - \frac{n}{N}\right)}$
<i>Nx: SSBQCM của chọn hoàn lại luôn lớn hơn SSBQCM của chọn không hoàn lại</i>			

Phạm vi sai số chọn mẫu (ε)

Suy rộng trung bình: $\varepsilon_{\bar{x}} = Z \cdot \sigma_{\bar{x}}$

Suy rộng tỷ lệ: $\varepsilon_f = Z \cdot \sigma_f$

Trong đó:

z - Hệ số tin cậy, tuân theo quy luật phân phối xác suất $N(0,1)$

$P(\bar{x} - \mu \leq \varepsilon_x) = 2\phi_0(z) = 1 - \alpha$	$P(f - p \leq \varepsilon_f) = 2\phi_0(z) = 1 - \alpha$
Xác suất để trị tuyệt đối của SSCM của số TB ($ \bar{x} - \mu $) không vượt quá phạm vi SSCM ($z \cdot \sigma_{\bar{x}}$) là $2\phi_0(z)$	Xác suất để trị tuyệt đối của SSCM của tỷ lệ ($ f - p $) không vượt quá phạm vi SSCM ($z \cdot \sigma_f$) là $2\phi_0(z)$

z	1	2	3
$2\phi_0(z)$	0.6826	0.9544	0.9974

4. Ước lượng kết quả điều tra

- Ước lượng số bình quân: $\bar{X} - \varepsilon_{\bar{X}} \leq \mu \leq \bar{X} + \varepsilon_{\bar{X}}$

$$\bar{X} - z.\sigma_{\bar{X}} \leq \mu \leq \bar{X} + z.\sigma_{\bar{X}}$$

- Ước lượng tỷ lệ: $f - \varepsilon_f \leq p \leq f + \varepsilon_f$

$$f - z.\sigma_f \leq p \leq f + z.\sigma_f$$

5. Xác định số đơn vị mẫu điều tra

- Yêu cầu: Sai số nhỏ nhất và chi phí thấp nhất
- Cơ sở xác định: Tiêu chuẩn sai số nhỏ nhất

Cách chọn	Chọn hoàn lại (chọn nhiều lần)	Chọn không hoàn lại (chọn một lần)
Suy rộng		
Bình quân	$n = \frac{z^2 \sigma^2}{\varepsilon_{\bar{x}}^2}$	$n = \frac{N \cdot z^2 \cdot \sigma^2}{N \cdot \varepsilon_{\bar{x}}^2 + z^2 \cdot \sigma^2}$
Tỷ lệ	$n = \frac{z^2 \cdot p(1-p)}{\varepsilon_f^2}$	$n = \frac{N \cdot z^2 \cdot p(1-p)}{\varepsilon_f^2 \cdot N + z^2 \cdot p(1-p)}$

Nx: Số đơn vị mẫu điều tra phụ thuộc vào hệ số tin cậy, phương sai (hoặc tỷ lệ) của tổng thể chung và phạm vi sai số chọn mẫu

Một số phương pháp xác định phương sai (hoặc tỷ lệ) của tổng thể chung

- + Lấy phương sai (σ^2) lớn nhất trong các lần điều tra trước hoặc tỷ lệ (p) gần 0,5 nhất
- + Lấy phương sai hoặc tỷ lệ của các cuộc điều tra khác có tính chất tương tự.
- + Điều tra thí điểm để xác định phương sai hoặc tỷ lệ.
- + Ước lượng phương sai dựa vào khoảng biến thiên

$$\sigma \approx \frac{R}{6} = \frac{X_{\max} - X_{\min}}{6}$$