

# ĐIỀU TRA CHỌN MẪU TRONG THỐNG KÊ

Phạm Thành Đạo  
Viện Khoa học Thống kê

## 1. Phương pháp điều tra chọn mẫu trong thống kê

Để xây dựng phương pháp điều tra chọn mẫu cho riêng ngành Thống kê trước tiên ta cần nghiên cứu đặc điểm của các cuộc điều tra thống kê.

**a. Đặc điểm của các cuộc điều tra thống kê:** Ngành Thống kê thường xuyên tổ chức phối hợp các cuộc điều tra toàn bộ với các cuộc điều tra chọn mẫu.

+ Cùng một hệ thống chỉ tiêu năm nay điều tra toàn bộ, sang năm hoặc một số năm sau đó lại điều tra chọn mẫu.

+ Trong cùng một năm một số chỉ tiêu dùng điều tra toàn bộ, một số chỉ tiêu khác dùng điều tra chọn mẫu.

Ký hiệu các chỉ tiêu điều tra chọn mẫu là  $X_i$  ( $i = 1, 2, 3, \dots, k$ )

Và ký hiệu các chỉ tiêu điều tra toàn bộ là  $Y_j$  ( $j = 1, 2, 3, \dots, h$ )

### b. Công thức sai số và cỡ mẫu

Để có thể so sánh mức độ sai số của nhiều chỉ tiêu ta nên dùng tỷ lệ sai số thay cho mức sai số. Quan hệ giữa mức sai số và tỷ lệ sai số như sau:

$$s = \frac{\Delta_s}{m} \cdot 100 = \frac{\sigma}{m} \sqrt{\frac{1}{n}} \cdot 100$$

với  $\Delta_s$  là mức sai số của số bình quân mẫu

$\sigma$  là phương sai của tổng thể

$m$  là số bình quân mẫu

$s$  là tỷ lệ sai số (tính bằng phần trăm)

Và công thức xác định cỡ mẫu cần thiết:

$$n \geq \left[ \frac{100 \cdot \frac{\sigma}{m} \cdot t}{s} \right]^2$$

với  $\Delta_s$  là mức sai số của số bình quân mẫu

$\sigma^2$  là phương sai của tổng thể

$n$  là số đơn vị mẫu

$t$  là hệ số tin cậy

Như vậy mức sai số và cỡ mẫu cần thiết đều phụ thuộc vào độ lớn của đại lượng ( $\frac{\sigma}{m}$ )

và đại lượng ( $\frac{\sigma}{m}$ ) này càng nhỏ thì sai số và cỡ mẫu càng nhỏ.

### c. Đặc điểm của số liệu thống kê

Nếu ta tính độ lệch tương đối của bản thân các chỉ tiêu cần điều tra và chỉ tiêu trung gian là thương của chỉ tiêu điều tra mẫu với chỉ tiêu điều tra toàn bộ ta thấy:

Lấy số liệu điều tra doanh nghiệp 2001 làm ví dụ:

+ Các chỉ tiêu điều tra mẫu là: tiền lương, giá trị tài sản, giá trị sản xuất, thu của người sản xuất.

+ Các chỉ tiêu điều tra toàn bộ gồm: Lao động, giá trị TSCĐ, doanh thu

Ta tính được giá trị của đại lượng  $(\frac{\sigma}{m})$  của bản thân chỉ tiêu điều tra và các chỉ tiêu trung gian như sau:

### GIÁ TRỊ CỦA ĐẠI LƯỢNG $(\frac{\sigma}{m})$

Các cách tiếp cận	Các chỉ tiêu điều tra mẫu			
	Tiền Lương	Tài sản	Giá trị sản xuất	Thu của người sx
Bản thân chỉ tiêu điều tra mẫu	3,06	4,37	4,47	3,11
Chỉ tiêu trung gian: Chỉ tiêu đt chia lao động	<b>0,73</b>	2,67	2,10	<b>0,70</b>
Chỉ tiêu trung gian: Chỉ tiêu đt chia TSCĐ	2,54	<b>0,69</b>	2,71	2,53
Chỉ tiêu trung gian: Chỉ tiêu đt chia Doanh thu	1,37	3,07	<b>0,12</b>	1,35

#### Nhận xét:

+ Giá trị của đại lượng  $(\frac{\sigma}{m})$  của chỉ tiêu trung gian  $\frac{X_i}{Y_j}$  nhỏ hơn nhiều giá trị số của đại lượng  $(\frac{\sigma}{m})$  của bản thân chỉ tiêu cần điều tra  $X_i$ . (Nếu  $X_i$  và  $Y_j$  là những số bất kỳ thì giá trị của đại lượng  $(\frac{\sigma}{m})$  của tỷ lệ  $\frac{X_i}{Y_j}$  không nhỏ hơn giá trị của đại lượng  $(\frac{\sigma}{m})$  của chỉ tiêu  $X_i$ ).

+ Có những giá trị của đại lượng  $(\frac{\sigma}{m})$  rất nhỏ, vấn đề đặt ra là ta chọn được chỉ tiêu điều tra toàn bộ  $Y_j$  sao cho giá trị của đại lượng  $(\frac{\sigma}{m})$  là nhỏ nhất. Sau khi lựa chọn được chỉ tiêu điều tra toàn bộ làm mẫu số cho chỉ tiêu điều tra trung gian, công thức ước lượng của chỉ tiêu điều tra như sau.

#### d. Công thức ước lượng

Sau khi tiến hành điều tra ta thu thập được số liệu như sau

+ Số liệu của chỉ tiêu điều tra của các đơn vị mẫu:  $x^1, x^2, \dots, x^n$

+ Số liệu của chỉ tiêu điều tra toàn bộ của các đơn vị mẫu:  $y^1, y^2, \dots, y^n$

Tổng chỉ tiêu cần điều tra của các đơn vị mẫu  $x = \sum_{i=1}^n x^i$

Tổng chỉ tiêu điều tra toàn bộ của các đơn vị mẫu  $y = \sum_{i=1}^n y^i$

Chỉ tiêu cần điều tra ước lượng cho toàn tổng thể ký hiệu là  $X^{\text{Tổng thể}}$

Tổng của chỉ tiêu điều tra toàn bộ trên toàn tổng thể ký hiệu là  $Y^{\text{Tổng thể}}$  ta có

+ Giá trị trung bình (có quyền số) của các tỷ lệ  $\frac{x_i}{y_j}$  của các đơn vị mẫu (m):

$$m = \frac{\sum_{i=1}^n y_i \cdot \frac{x_i}{y_i}}{\sum_{j=1}^n y_j} = \frac{\sum_{i=1}^n x_i}{\sum_{j=1}^n y_j} = \frac{\sum_{i=1}^n x_i}{\sum_{j=1}^n y_j} = \frac{x}{y}$$

+ Chỉ tiêu cần điều tra ước lượng cho toàn tổng thể:

$$X_{\text{Tổng thể}} = \frac{X}{y} \cdot Y_{\text{Tổng thể}}$$

**e. Tỷ lệ sai số điều tra mẫu**

$$s = \frac{\sigma}{m} \sqrt{\frac{1}{n}} \cdot 100 \text{ (tính bằng \%)}$$

$$\text{với } \sigma^2 = \frac{\sum_{i=1}^n y^j \left( \frac{x^i}{y^i} - \frac{x}{y} \right)^2}{\sum_{j=1}^n y^j} \text{ và } m = \frac{x}{y}$$

**3. Vận dụng phương pháp chọn mẫu trong thống kê vào thực tiễn**

A. Trường hợp cùng một hệ thống chỉ tiêu năm nay điều tra toàn bộ, sang năm

hoặc một số năm sau đó điều tra chọn mẫu. Ví dụ cuộc điều tra các đơn vị hành chính, sự nghiệp của Vụ hệ thống tài khoản quốc gia năm 2002 điều tra toàn bộ, năm 2003 điều tra mẫu trên các chỉ tiêu tổng số thu và tổng số chi của từng đơn vị. Với trường hợp này chỉ tiêu trung gian  $\frac{X_i}{Y_j}$  thành tốc độ tăng thu hoặc chi của các đơn vị điều tra. Nếu vận dụng phương pháp điều tra mẫu chung vào cuộc điều tra này thì cỡ mẫu có tỷ lệ sai số 5% cho các tỉnh là gần như điều tra toàn bộ, còn nếu vận dụng phương pháp điều tra mẫu trong thống kê thì cụ thể cỡ mẫu cho từng tỉnh như sau:

a) Thành phố Hà Nội: Số đơn vị tổng thể 3629

Chỉ tiêu ước lượng từ mẫu	Đại lượng ( $\frac{\sigma}{m}$ )	Cỡ mẫu cần chọn (Sai số 5%)	Cỡ mẫu dự kiến
Tốc độ tăng thu	0,662	176	200
Tốc độ tăng chi	0,511	105	200

b) Thành phố Hải Phòng: Số đơn vị tổng thể 3253

Chỉ tiêu ước lượng từ mẫu	Đại lượng ( $\frac{\sigma}{m}$ )	Cỡ mẫu cần chọn (Sai số 5%)	Cỡ mẫu dự kiến
Tốc độ tăng thu	0,400	64	200
Tốc độ tăng chi	0,415	69	200

c) Thành phố Hồ Chí Minh: Số đơn vị tổng thể 6522

Chỉ tiêu ước lượng từ mẫu	Đại lượng ( $\frac{\sigma}{m}$ )	Cỡ mẫu cần chọn (Sai số 5%)	Cỡ mẫu dự kiến
Tốc độ tăng thu	0,405	66	250
Tốc độ tăng chi	0,635	161	250

d) Thành phố Đà Nẵng: Số đơn vị tổng thể 1219

Chỉ tiêu ước lượng từ mẫu	Đại lượng ( $\frac{\sigma}{m}$ )	Cỡ mẫu cần chọn (Sai số 5%)	Cỡ mẫu dự kiến
Tốc độ tăng thu	0,648	168	250
Tốc độ tăng chi	0,907	329	250

Như vậy, nếu vận dụng phương pháp mẫu trong thống kê thì chỉ cần điều tra khoảng 200 đơn vị /tỉnh đối với các tỉnh phía Bắc và khoảng 250 đơn vị/tỉnh đối với các tỉnh phía Nam là ta sẽ được kết quả điều tra với sai số dưới 5%. Cụ thể đối với Hà Nội dùng tốc độ tăng chi tốt hơn dùng tốc độ tăng thu còn Thành phố Hồ Chí Minh thì ngược lại. Tỉnh Hải Phòng có thể giảm cỡ mẫu ít nhiều, trong khi đối với Tỉnh Đà Nẵng thì nên tăng cỡ mẫu chút ít.

B. Trường hợp trong cùng một năm một số chỉ tiêu điều tra toàn bộ, một số chỉ tiêu thu thập bằng điều tra chọn mẫu. Ví dụ điều tra doanh nghiệp ta có

+ Các chỉ tiêu điều tra toàn bộ: Số lao động, giá trị TSCĐ, doanh thu

+ Các chỉ tiêu điều tra mẫu: Tiền lương, giá trị tài sản, giá trị sản xuất, thu của người sản xuất.

Để ước lượng các chỉ tiêu điều tra mẫu ta dùng chỉ tiêu điều tra toàn bộ có quan hệ gần gũi với nó làm chỉ tiêu mẫu số ví dụ.

Để ước lượng các chỉ tiêu điều tra mẫu là tiền lương ta chọn mẫu số là chỉ tiêu lao động.

Để ước lượng các chỉ tiêu điều tra mẫu là giá trị tài sản ta chọn mẫu số là chỉ tiêu giá trị TSCĐ.

Để ước lượng các chỉ tiêu điều tra mẫu là giá trị sản xuất ta chọn mẫu số là chỉ tiêu doanh thu.

Để ước lượng các chỉ tiêu điều tra mẫu là thu của người sản xuất ta chọn mẫu số là chỉ tiêu lao động, v.v...

Tính toán cỡ mẫu cần điều tra cho từng chỉ tiêu điều tra mẫu với tỷ lệ sai số 5% cụ thể như sau

Chỉ tiêu tiền lương cần cỡ mẫu là 213

Chỉ tiêu giá trị tài sản cần cỡ mẫu là 190

Chỉ tiêu giá trị sản xuất cần cỡ mẫu là 60

Chỉ tiêu thu của người sản xuất cần cỡ mẫu là 196

Như vậy với một mẫu khoảng 250 doanh nghiệp, mỗi chỉ tiêu điều tra mẫu có một cách ước lượng riêng, ta sẽ có số liệu của cả tổng thể với tỷ lệ sai số dưới 5% ■