

Bài giảng Xác suất Thống kê và ứng dụng

Đại học THĂNG LONG

Ngày 11 tháng 3 năm 2009

Phần II

Thu thập dữ liệu

- 1 Xác định dữ liệu cần thu thập
- 2 Dữ liệu sơ cấp và thứ cấp
- 3 Lấy mẫu từ tổng thể

- 1 Xác định dữ liệu cần thu thập
- 2 Dữ liệu sơ cấp và thứ cấp
- 3 Lấy mẫu từ tổng thể

- 1 Xác định dữ liệu cần thu thập
- 2 Dữ liệu sơ cấp và thứ cấp
- 3 Lấy mẫu từ tổng thể

Xác định dữ liệu cần thu thập

Người nghiên cứu có thể thu thập được nhiều dữ liệu khác nhau liên quan tới các đối tượng nghiên cứu. Vấn đề quan trọng đầu tiên trong việc thu thập dữ liệu là xác định rõ những liệu cần thu thập và thứ tự ưu tiên của các dữ liệu đó để tránh lãng phí thời gian và chi phí cho những dữ liệu ít quan trọng.

- ♣ **Dữ liệu thứ cấp** là dữ liệu thu được từ những nguồn có sẵn, thường đã qua tổng hợp, xử lý.

Nguồn cung cấp: nội bộ, cơ quan thống kê, cơ quan chính phủ, cơ sở nghiên cứu, báo, tạp chí, ...

- ♣ **Dữ liệu sơ cấp** là dữ liệu thu thập trực tiếp từ đối tượng nghiên cứu.

Nguồn cung cấp: Trong nghiên cứu thử nghiệm, người nghiên cứu đo đạc và thu thập dữ liệu trên các kết quả của thử nghiệm. Trong nghiên cứu quan sát, dữ liệu có thể thu được từ nhiều nguồn: cá nhân, gia đình, cơ quan, tổ chức, ... bằng các hình thức như phỏng vấn trực tiếp, thư, điện thoại, ...

Dữ liệu sơ cấp và thứ cấp

- ♣ **Dữ liệu thứ cấp** là dữ liệu thu được từ những nguồn có sẵn, thường đã qua tổng hợp, xử lý.
Nguồn cung cấp: nội bộ, cơ quan thống kê, cơ quan chính phủ, cơ sở nghiên cứu, báo, tạp chí, ...
- ♣ **Dữ liệu sơ cấp** là dữ liệu thu thập trực tiếp từ đối tượng nghiên cứu.
Nguồn cung cấp: Trong nghiên cứu thử nghiệm, người nghiên cứu đo đạc và thu thập dữ liệu trên các kết quả của thử nghiệm. Trong nghiên cứu quan sát, dữ liệu có thể thu được từ nhiều nguồn: cá nhân, gia đình, cơ quan, tổ chức, ... bằng các hình thức như phỏng vấn trực tiếp, thư, điện thoại, ...

Dữ liệu sơ cấp và thứ cấp

- ♣ **Dữ liệu thứ cấp** là dữ liệu thu được từ những nguồn có sẵn, thường đã qua tổng hợp, xử lý.
Nguồn cung cấp: nội bộ, cơ quan thống kê, cơ quan chính phủ, cơ sở nghiên cứu, báo, tạp chí, ...
- ♣ **Dữ liệu sơ cấp** là dữ liệu thu thập trực tiếp từ đối tượng nghiên cứu.
Nguồn cung cấp: Trong nghiên cứu thử nghiệm, người nghiên cứu đo đạc và thu thập dữ liệu trên các kết quả của thử nghiệm. Trong nghiên cứu quan sát, dữ liệu có thể thu được từ nhiều nguồn: cá nhân, gia đình, cơ quan, tổ chức, ... bằng các hình thức như phỏng vấn trực tiếp, thư, điện thoại, ...

Lấy mẫu từ tổng thể

- 1 Mẫu ngẫu nhiên
- 2 Lấy mẫu ngẫu nhiên đơn giản
- 3 Lấy mẫu hệ thống
- 4 Một số phương pháp lấy mẫu khác

Để một mẫu đại diện được cho tổng thể thì mẫu phải được chọn ngẫu nhiên từ tổng thể đó.

- ♣ **Mẫu ngẫu nhiên** là mẫu được chọn sao cho mỗi lần chọn phần tử từ tổng thể, các phần tử còn lại trong tổng thể đều có khả năng được chọn như nhau.

Lấy mẫu ngẫu nhiên đơn giản

Giả sử cần chọn ra một mẫu gồm k phần tử từ một tổng thể có N phần tử. Đầu tiên ta tạo một danh sách các phần tử của tổng thể được đánh số từ 1 đến N .

- ♣ Khi số phần tử của tổng thể là N nhỏ, ta có thể dùng phương pháp bốc thăm như sau: ghi các số 1 đến N vào những mảnh giấy, gập lại, trộn đều, từ đó lấy ngẫu nhiên ra k mảnh giấy, ta sẽ được mẫu tương ứng.
- ♣ Khi số phần tử của tổng thể lớn ta dùng **bảng số ngẫu nhiên**.

Lấy mẫu ngẫu nhiên đơn giản

Giả sử cần chọn ra một mẫu gồm k phần tử từ một tổng thể có N phần tử. Đầu tiên ta tạo một danh sách các phần tử của tổng thể được đánh số từ 1 đến N .

- ♣ Khi số phần tử của tổng thể là N nhỏ, ta có thể dùng phương pháp bốc thăm như sau: ghi các số 1 đến N vào những mảnh giấy, gập lại, trộn đều, từ đó lấy ngẫu nhiên ra k mảnh giấy, ta sẽ được mẫu tương ứng.
- ♣ Khi số phần tử của tổng thể lớn ta dùng **bảng số ngẫu nhiên**.

Lấy mẫu ngẫu nhiên đơn giản

Giả sử cần chọn ra một mẫu gồm k phần tử từ một tổng thể có N phần tử. Đầu tiên ta tạo một danh sách các phần tử của tổng thể được đánh số từ 1 đến N .

- ♣ Khi số phần tử của tổng thể là N nhỏ, ta có thể dùng phương pháp bốc thăm như sau: ghi các số 1 đến N vào những mảnh giấy, gập lại, trộn đều, từ đó lấy ngẫu nhiên ra k mảnh giấy, ta sẽ được mẫu tương ứng.
- ♣ Khi số phần tử của tổng thể lớn ta dùng **bảng số ngẫu nhiên**.

Ví dụ: Một công ty sử dụng máy tính để tạo ra các hóa đơn thanh toán, sau đó các hóa đơn này được gửi tới khách hàng qua đường bưu điện và thường phải mất trên 39 ngày hóa đơn mới được thanh toán.

Công ty muốn đưa vào một hệ thống mới hy vọng thời gian thanh toán sẽ được rút ngắn. Để ước lượng hiệu quả của hệ thống mới, người ta đánh số các hóa đơn được gửi đi trong 12 tuần từ 0001 tới 7823, sau đó dùng bảng số ngẫu nhiên để chọn ra 65 hóa đơn.

51259	77452	16308	60756	92144	49442	53900	70960
60268	89368	19885	55322	44819	01188	65255	64835
94904	31273	04146	18594	29852	71585	85030	51132
58586	23216	14513	83149	98736	23491	64350	94738
09998	42698	06691	76988	13602	51851	46104	88916
14346	09172	30168	90229	04734	59193	22178	30412
74103	47070	25306	76468	26384	58151	06646	21524
24200	13363	38005	94342	28728	35806	06912	17012
87308	58731	00256	45834	15398	46557	41135	10367
07351	19731	92420	60952	61280	50001	67658	32586
96243	24878	82651	66566	14778	76797	14780	13300
26432	46901	20849	89768	81536	86645	12659	92259
66432	84673	40027	32832	61362	98947	96067	64760

Bảng: Một phần của bảng số ngẫu nhiên

Do các hóa đơn được đánh số từ 1 tới 7823 nên ta sẽ chọn từ bảng số ngẫu nhiên ra những nhóm gồm 4 chữ số. Đầu tiên chọn ngẫu nhiên ra một nhóm gồm 4 chữ số trong bảng là 1851, ta được hóa đơn đầu tiên trong mẫu có thứ tự là 1851. Sau đó di chuyển từ nhóm 1851 theo đường thẳng (lên, xuống, trái, phải) một cách tùy ý để thêm những hóa đơn tiếp theo vào mẫu.

Chẳng hạn di chuyển xuống dưới ta có những nhóm gồm 4 chữ số sau

1851, 9193, 8151, 5806, 6557, 0001, 6797

Bỏ đi những nhóm 9193 và 8181 ta thu được những hóa đơn có số thứ tự

1851, 5806, 6557, 0001, 6797

Cứ tiếp tục quá trình trên ta chọn ra được một mẫu ngẫu nhiên gồm 65 hóa đơn , ghi lại thời gian thanh toán của từng hóa đơn ta được mẫu gồm 65 thời gian thanh toán sau:

22	29	16	15	18	17	12	13	17	16	15
19	17	10	21	15	14	17	18	12	20	14
16	15	16	20	22	14	25	19	23	15	19
18	23	22	16	16	19	13	18	24	24	26
13	18	17	15	24	15	17	14	18	17	21
16	21	25	19	20	27	16	17	16	21	

Bảng: Mẫu 65 thời gian thanh toán

Lấy mẫu hệ thống

Trong trường hợp cần lấy mẫu với số phần tử rất nhiều, việc chọn mẫu bằng phương pháp nói trên sẽ khó khăn, thay vào đó người ta dùng cách lấy mẫu theo hệ thống. Các bước thực hiện như sau:

- 1 Tạo một danh sách các phần tử trong tổng thể, đánh số từ 1 đến N .
- 2 Chia tổng thể thành n nhóm với n là số làm tròn của N/k , n được gọi là khoảng cách chọn mẫu.
- 3 Trong n phần tử đầu tiên chọn ngẫu nhiên ra một phần tử, các phần tử tiếp theo có thứ tự cách phần tử đầu tiên này khoảng cách là n , $2n$, $3n, \dots$. Nếu chưa đủ ta quay lại đầu danh sách với quy ước phần tử thứ $N + 1$ là phần tử thứ 1, phần tử thứ $N + 2$ là phần tử thứ 2, \dots

Ví dụ: Với $N = 66$, $k = 10$, ta lấy $n = 7$.

Giả sử số đầu tiên chọn được là 3 thì các phần tử tiếp theo có thứ tự: 10, 17, 24, 31, 38, 45, 52, 59, 66.

Giả sử số đầu tiên chọn được là 7 thì các phần tử tiếp theo có thứ tự: 14, 21, 28, 35, 42, 49, 56, 63, 70 = 4.