

Xác suất Thống kê ứng dụng trong Kinh tế Xã hội

Nguyễn Thị Nhung

Bộ môn Toán - Đại học Thăng Long

Ngày 9 tháng 8 năm 2013

Chương II

Thu thập dữ liệu

1 Xác định dữ liệu cần thu thập

2 Dữ liệu thứ cấp và dữ liệu sơ cấp

- Khái niệm dữ liệu thứ cấp và sơ cấp
- Nguồn dữ liệu thứ cấp và sơ cấp
- Thu thập dữ liệu sơ cấp

3 Các kĩ thuật lấy mẫu

- Lí do phải chọn mẫu và điều tra tổng thể
- Một số khái niệm về chọn mẫu
- Chọn mẫu ngẫu nhiên
- Chọn mẫu không ngẫu nhiên
- Lỗi do chọn mẫu và không do chọn mẫu

Chương II

- 1 Xác định dữ liệu cần thu thập
- 2 Dữ liệu thứ cấp và dữ liệu sơ cấp
 - Khái niệm dữ liệu thứ cấp và sơ cấp
 - Nguồn dữ liệu thứ cấp và sơ cấp
 - Thu thập dữ liệu sơ cấp
- 3 Các kỹ thuật lấy mẫu
 - Lí do phải chọn mẫu và điều tra tổng thể
 - Một số khái niệm về chọn mẫu
 - Chọn mẫu ngẫu nhiên
 - Chọn mẫu không ngẫu nhiên
 - Lỗi do chọn mẫu và không do chọn mẫu

- 1 Xác định dữ liệu cần thu thập
- 2 Dữ liệu thứ cấp và dữ liệu sơ cấp
 - Khái niệm dữ liệu thứ cấp và sơ cấp
 - Nguồn dữ liệu thứ cấp và sơ cấp
 - Thu thập dữ liệu sơ cấp
- 3 Các kĩ thuật lấy mẫu
 - Lí do phải chọn mẫu và điều tra tổng thể
 - Một số khái niệm về chọn mẫu
 - Chọn mẫu ngẫu nhiên
 - Chọn mẫu không ngẫu nhiên
 - Lỗi do chọn mẫu và không do chọn mẫu

Nội dung chính của chương

- Giới thiệu về khái niệm, cách thức thu thập hai kiểu dữ liệu thứ cấp và sơ cấp.
- Trình bày về lí do vì sao phải chọn mẫu và một số khái niệm về chọn mẫu.
- Giới thiệu về các kiểu lấy mẫu: lấy mẫu ngẫu nhiên, lấy mẫu không ngẫu nhiên.
- Giới thiệu về những lỗi gặp trong quá trình chọn mẫu.

Yêu cầu đối với sinh viên

- Phân biệt được dữ liệu thứ cấp, sơ cấp. Hiểu được ưu điểm, nhược điểm và cách thức thu thập từng loại dữ liệu.
- Hiểu được vì sao phải điều tra chọn mẫu mà không tổng điều tra.
- Nắm được một số khái niệm về chọn mẫu: chọn mẫu có hoàn lại, không hoàn lại, chọn mẫu ngẫu nhiên, không ngẫu nhiên.
- Nắm được một số kĩ thuật chọn mẫu ngẫu nhiên: chọn mẫu ngẫu nhiên đơn giản, chọn mẫu hệ thống, lấy mẫu phân tầng, lấy mẫu theo khối, lấy mẫu nhiều giai đoạn.
- Nắm được một số kĩ thuật chọn mẫu không ngẫu nhiên: chọn mẫu thuận tiện, chọn mẫu định mức, chọn mẫu phán đoán.
- Hiểu những lỗi có thể gặp trong quá trình chọn mẫu.

Xác định dữ liệu cần thu thập

Dựa vào vấn đề nghiên cứu và mục tiêu nghiên cứu, khi thu thập dữ liệu ta phải xác định rõ:

- Dữ liệu nào cần phải thu thập;
- Thứ tự ưu tiên của các dữ liệu này;
- Xác định rõ giới hạn, phạm vi dữ liệu.

Câu hỏi tình huống

Câu hỏi tình huống

Giả sử bạn là thành viên trong Văn phòng Đoàn của trường Đại học Thăng Long, bạn được giao nhiệm vụ tìm hiểu về điểm thi môn Toán logic của sinh viên Thăng Long để viết bài lên trang Web trường. Bạn có thể lấy được nguồn dữ liệu điểm bằng những cách nào? So sánh ưu nhược điểm của từng cách thức bạn vừa đề cập đến.

Nội dung trình bày

- 1 Xác định dữ liệu cần thu thập
- 2 Dữ liệu thứ cấp và dữ liệu sơ cấp
 - Khái niệm dữ liệu thứ cấp và sơ cấp
 - Nguồn dữ liệu thứ cấp và sơ cấp
 - Thu thập dữ liệu sơ cấp
- 3 Các kĩ thuật lấy mẫu
 - Lí do phải chọn mẫu và điều tra tổng thể
 - Một số khái niệm về chọn mẫu
 - Chọn mẫu ngẫu nhiên
 - Chọn mẫu không ngẫu nhiên
 - Lỗi do chọn mẫu và không do chọn mẫu

Khái niệm dữ liệu thứ cấp và dữ liệu sơ cấp

- Dữ liệu thứ cấp: dữ liệu thu thập từ những nguồn có sẵn, thường là đã qua tổng hợp, xử lý.
 - Ưu điểm: thu thập nhanh, ít tốn kém chi phí và thời gian.
 - Nhược điểm: ít chi tiết, không đáp ứng nhu cầu nghiên cứu.
- Dữ liệu sơ cấp: dữ liệu thu thập trực tiếp từ đối tượng nghiên cứu.
 - Ưu điểm: chi tiết, đáp ứng tốt nhu cầu nghiên cứu.
 - Nhược điểm: tốn kém chi phí và thời gian.

Khái niệm dữ liệu thứ cấp và dữ liệu sơ cấp

- Dữ liệu thứ cấp: dữ liệu thu thập từ những nguồn có sẵn, thường là đã qua tổng hợp, xử lý.
 - Ưu điểm: thu thập nhanh, ít tốn kém chi phí và thời gian.
 - Nhược điểm: ít chi tiết, không đáp ứng nhu cầu nghiên cứu.
- Dữ liệu sơ cấp: dữ liệu thu thập trực tiếp từ đối tượng nghiên cứu.
 - Ưu điểm: chi tiết, đáp ứng tốt nhu cầu nghiên cứu.
 - Nhược điểm: tốn kém chi phí và thời gian.

Nội dung trình bày

- 1 Xác định dữ liệu cần thu thập
- 2 Dữ liệu thứ cấp và dữ liệu sơ cấp
 - Khái niệm dữ liệu thứ cấp và sơ cấp
 - Nguồn dữ liệu thứ cấp và sơ cấp
 - Thu thập dữ liệu sơ cấp
- 3 Các kĩ thuật lấy mẫu
 - Lí do phải chọn mẫu và điều tra tổng thể
 - Một số khái niệm về chọn mẫu
 - Chọn mẫu ngẫu nhiên
 - Chọn mẫu không ngẫu nhiên
 - Lỗi do chọn mẫu và không do chọn mẫu

Nguồn dữ liệu thứ cấp

Các tổ chức xã hội, doanh nghiệp hoặc cá nhân có thể lấy dữ liệu thứ cấp từ các nguồn sau:

- Nguồn nội bộ từ các doanh nghiệp như báo cáo về sản xuất, tài chính, nhân sự,...
- Cơ quan thống kê nhà nước: các số liệu do các cơ quan Thống kê nhà nước (Tổng cục Thống kê, <http://www.gso.gov.vn>) cung cấp trong Niên giám Thống kê.
- Cơ quan chính phủ: số liệu do các cơ quan trực thuộc chính phủ công bố hay cung cấp.
- Báo, tạp chí: số liệu do tạp chí, báo cung cấp. Dữ liệu này mang tính thời sự, cập nhật cao. Tuy nhiên, mức độ tin cậy phụ thuộc vào nguồn số liệu mà các tạp chí, báo này sử dụng hay cách thức điều tra.
- Các tổ chức, hiệp hội, viện nghiên cứu, các công ty và tổ chức nghiên cứu và cung cấp thông tin theo yêu cầu.

Nguồn dữ liệu thứ cấp

Các tổ chức xã hội, doanh nghiệp hoặc cá nhân có thể lấy dữ liệu thứ cấp từ các nguồn sau:

- Nguồn nội bộ từ các doanh nghiệp như báo cáo về sản xuất, tài chính, nhân sự,...
- Cơ quan thống kê nhà nước: các số liệu do các cơ quan Thống kê nhà nước (Tổng cục Thống kê, <http://www.gso.gov.vn>) cung cấp trong Niên giám Thống kê.
- Cơ quan chính phủ: số liệu do các cơ quan trực thuộc chính phủ công bố hay cung cấp.
- Báo, tạp chí: số liệu do tạp chí, báo cung cấp. Dữ liệu này mang tính thời sự, cập nhật cao. Tuy nhiên, mức độ tin cậy phụ thuộc vào nguồn số liệu mà các tạp chí, báo này sử dụng hay cách thức điều tra.
- Các tổ chức, hiệp hội, viện nghiên cứu, các công ty và tổ chức nghiên cứu và cung cấp thông tin theo yêu cầu.

Nguồn dữ liệu thứ cấp

Các tổ chức xã hội, doanh nghiệp hoặc cá nhân có thể lấy dữ liệu thứ cấp từ các nguồn sau:

- Nguồn nội bộ từ các doanh nghiệp như báo cáo về sản xuất, tài chính, nhân sự,...
- Cơ quan thống kê nhà nước: các số liệu do các cơ quan Thống kê nhà nước (Tổng cục Thống kê, <http://www.gso.gov.vn>) cung cấp trong Niên giám Thống kê.
- Cơ quan chính phủ: số liệu do các cơ quan trực thuộc chính phủ công bố hay cung cấp.
- Báo, tạp chí: số liệu do tạp chí, báo cung cấp. Dữ liệu này mang tính thời sự, cập nhật cao. Tuy nhiên, mức độ tin cậy phụ thuộc vào nguồn số liệu mà các tạp chí, báo này sử dụng hay cách thức điều tra.
- Các tổ chức, hiệp hội, viện nghiên cứu, các công ty và tổ chức nghiên cứu và cung cấp thông tin theo yêu cầu.

Nguồn dữ liệu thứ cấp

Các tổ chức xã hội, doanh nghiệp hoặc cá nhân có thể lấy dữ liệu thứ cấp từ các nguồn sau:

- Nguồn nội bộ từ các doanh nghiệp như báo cáo về sản xuất, tài chính, nhân sự,...
- Cơ quan thống kê nhà nước: các số liệu do các cơ quan Thống kê nhà nước (Tổng cục Thống kê, <http://www.gso.gov.vn>) cung cấp trong Niên giám Thống kê.
- Cơ quan chính phủ: số liệu do các cơ quan trực thuộc chính phủ công bố hay cung cấp.
- Báo, tạp chí: số liệu do tạp chí, báo cung cấp. Dữ liệu này mang tính thời sự, cập nhật cao. Tuy nhiên, mức độ tin cậy phụ thuộc vào nguồn số liệu mà các tạp chí, báo này sử dụng hay cách thức điều tra.
- Các tổ chức, hiệp hội, viện nghiên cứu, các công ty và tổ chức nghiên cứu và cung cấp thông tin theo yêu cầu.

Nguồn dữ liệu thứ cấp

Các tổ chức xã hội, doanh nghiệp hoặc cá nhân có thể lấy dữ liệu thứ cấp từ các nguồn sau:

- Nguồn nội bộ từ các doanh nghiệp như báo cáo về sản xuất, tài chính, nhân sự,...
- Cơ quan thống kê nhà nước: các số liệu do các cơ quan Thống kê nhà nước (Tổng cục Thống kê, <http://www.gso.gov.vn>) cung cấp trong Niên giám Thống kê.
- Cơ quan chính phủ: số liệu do các cơ quan trực thuộc chính phủ công bố hay cung cấp.
- Báo, tạp chí: số liệu do tạp chí, báo cung cấp. Dữ liệu này mang tính thời sự, cập nhật cao. Tuy nhiên, mức độ tin cậy phụ thuộc vào nguồn số liệu mà các tạp chí, báo này sử dụng hay cách thức điều tra.
- Các tổ chức, hiệp hội, viện nghiên cứu, các công ty và tổ chức nghiên cứu và cung cấp thông tin theo yêu cầu.

Nguồn dữ liệu thứ cấp

Các tổ chức xã hội, doanh nghiệp hoặc cá nhân có thể lấy dữ liệu thứ cấp từ các nguồn sau:

- Nguồn nội bộ từ các doanh nghiệp như báo cáo về sản xuất, tài chính, nhân sự,...
- Cơ quan thống kê nhà nước: các số liệu do các cơ quan Thống kê nhà nước (Tổng cục Thống kê, <http://www.gso.gov.vn>) cung cấp trong Niên giám Thống kê.
- Cơ quan chính phủ: số liệu do các cơ quan trực thuộc chính phủ công bố hay cung cấp.
- Báo, tạp chí: số liệu do tạp chí, báo cung cấp. Dữ liệu này mang tính thời sự, cập nhật cao. Tuy nhiên, mức độ tin cậy phụ thuộc vào nguồn số liệu mà các tạp chí, báo này sử dụng hay cách thức điều tra.
- Các tổ chức, hiệp hội, viện nghiên cứu, các công ty và tổ chức nghiên cứu và cung cấp thông tin theo yêu cầu.

Nguồn dữ liệu sơ cấp

Dữ liệu sơ cấp thường được thu thập theo một qui trình bài bản tùy theo nghiên cứu thống kê là nghiên cứu thử nghiệm hay nghiên cứu quan sát.

- Trong nghiên cứu thử nghiệm, người nghiên cứu đo đạc và thu thập dữ liệu trên các biến kết quả trong các điều kiện khác nhau của biến nguyên nhân đang cần nghiên cứu.
- Trong nghiên cứu quan sát thì các dữ liệu có thể thu thập từ:
 - Nhiều người cung cấp thông tin khác nhau như: chủ hộ gia đình, khách hàng, chủ doanh nghiệp,...
 - Nguồn nội bộ như phòng kinh doanh của công ty, các bộ phận có chức năng ghi chép số liệu của cơ quan,...
 - Nguồn bên ngoài như công ty nghiên cứu thị trường, các tổ chức chuyên nghiệp,...

Nguồn dữ liệu sơ cấp

Dữ liệu sơ cấp thường được thu thập theo một qui trình bài bản tùy theo nghiên cứu thống kê là nghiên cứu thử nghiệm hay nghiên cứu quan sát.

- Trong nghiên cứu thử nghiệm, người nghiên cứu đo đạc và thu thập dữ liệu trên các biến kết quả trong các điều kiện khác nhau của biến nguyên nhân đang cần nghiên cứu.
- Trong nghiên cứu quan sát thì các dữ liệu có thể thu thập từ:
 - Nhiều người cung cấp thông tin khác nhau như: chủ hộ gia đình, khách hàng, chủ doanh nghiệp,...
 - Nguồn nội bộ như phòng kinh doanh của công ty, các bộ phận có chức năng ghi chép số liệu của cơ quan,...
 - Nguồn bên ngoài như công ty nghiên cứu thị trường, các tổ chức chuyên nghiệp,...

Nguồn dữ liệu sơ cấp

Dữ liệu sơ cấp thường được thu thập theo một qui trình bài bản tùy theo nghiên cứu thống kê là nghiên cứu thử nghiệm hay nghiên cứu quan sát.

- Trong nghiên cứu thử nghiệm, người nghiên cứu đo đạc và thu thập dữ liệu trên các biến kết quả trong các điều kiện khác nhau của biến nguyên nhân đang cần nghiên cứu.
- Trong nghiên cứu quan sát thì các dữ liệu có thể thu thập từ:
 - Nhiều người cung cấp thông tin khác nhau như: chủ hộ gia đình, khách hàng, chủ doanh nghiệp,...
 - Nguồn nội bộ như phòng kinh doanh của công ty, các bộ phận có chức năng ghi chép số liệu của cơ quan,...
 - Nguồn bên ngoài như công ty nghiên cứu thị trường, các tổ chức chuyên nghiệp,...

Nguồn dữ liệu sơ cấp

Dữ liệu sơ cấp thường được thu thập theo một qui trình bài bản tùy theo nghiên cứu thống kê là nghiên cứu thử nghiệm hay nghiên cứu quan sát.

- Trong nghiên cứu thử nghiệm, người nghiên cứu đo đạc và thu thập dữ liệu trên các biến kết quả trong các điều kiện khác nhau của biến nguyên nhân đang cần nghiên cứu.
- Trong nghiên cứu quan sát thì các dữ liệu có thể thu thập từ:
 - Nhiều người cung cấp thông tin khác nhau như: chủ hộ gia đình, khách hàng, chủ doanh nghiệp,...
 - Nguồn nội bộ như phòng kinh doanh của công ty, các bộ phận có chức năng ghi chép số liệu của cơ quan,...
 - Nguồn bên ngoài như công ty nghiên cứu thị trường, các tổ chức chuyên nghiệp,...

Nguồn dữ liệu sơ cấp

Dữ liệu sơ cấp thường được thu thập theo một qui trình bài bản tùy theo nghiên cứu thống kê là nghiên cứu thử nghiệm hay nghiên cứu quan sát.

- Trong nghiên cứu thử nghiệm, người nghiên cứu đo đạc và thu thập dữ liệu trên các biến kết quả trong các điều kiện khác nhau của biến nguyên nhân đang cần nghiên cứu.
- Trong nghiên cứu quan sát thì các dữ liệu có thể thu thập từ:
 - Nhiều người cung cấp thông tin khác nhau như: chủ hộ gia đình, khách hàng, chủ doanh nghiệp,...
 - Nguồn nội bộ như phòng kinh doanh của công ty, các bộ phận có chức năng ghi chép số liệu của cơ quan,...
 - Nguồn bên ngoài như công ty nghiên cứu thị trường, các tổ chức chuyên nghiệp,...

Nội dung trình bày

1 Xác định dữ liệu cần thu thập

2 Dữ liệu thứ cấp và dữ liệu sơ cấp

- Khái niệm dữ liệu thứ cấp và sơ cấp
- Nguồn dữ liệu thứ cấp và sơ cấp
- Thu thập dữ liệu sơ cấp

3 Các kĩ thuật lấy mẫu

- Lí do phải chọn mẫu và điều tra tổng thể
- Một số khái niệm về chọn mẫu
- Chọn mẫu ngẫu nhiên
- Chọn mẫu không ngẫu nhiên
- Lỗi do chọn mẫu và không do chọn mẫu

Thu thập dữ liệu sơ cấp

- Thu thập dữ liệu trong nghiên cứu thực nghiệm;
- Thu thập dữ liệu trong nghiên cứu quan sát:
 - Khảo sát qua điện thoại;
 - Khảo sát dạng viết;
 - Khảo sát trực tiếp và phỏng vấn cá nhân.

Thu thập dữ liệu trong nghiên cứu thực nghiệm

Các công ty và các tổ chức thường thực hiện các thực nghiệm hoặc nhóm các thí nghiệm để thu thập dữ liệu cung cấp cho nhà quản lí đề ra những quyết định. Trong nghiên cứu thực nghiệm, người nghiên cứu đo đạc và thu thập dữ liệu trên các biến kết quả trong các điều kiện khác nhau của các biến nguyên nhân có ảnh hưởng đang nghiên cứu.

Thu thập dữ liệu trong nghiên cứu quan sát

Nội dung chính để thu thập dữ liệu trong nghiên cứu quan sát bao gồm:

- Thiết kế bản câu hỏi hoàn chỉnh;
- Quyết định cách chọn mẫu nếu không khảo sát hết toàn bộ tổng thể;
- Thực hiện việc thu thập dữ liệu: tiếp cận đối tượng, quan sát, ghi nhận dữ liệu.

Khảo sát qua điện thoại

Khi thu thập dữ liệu bằng cách khảo sát qua điện thoại, người khảo sát gọi điện đến đối tượng cần thu thập dữ liệu để thực hiện cuộc trao đổi ngắn về vấn đề cần nghiên cứu.

Nội dung các câu trao đổi trong cuộc khảo sát được thiết kế thành văn bản gồm các phần:

- Phần đầu là những giải thích ngắn gọn về mục đích của cuộc khảo sát;
- Phần thứ hai là những câu hỏi liên quan đến vấn đề cần khảo sát. Các câu hỏi thường là những câu hỏi đóng, tức là yêu cầu người trả lời chọn một câu trả lời từ một số lựa chọn xác định;
- Phần cuối là những câu hỏi liên quan đến thông tin cá nhân của người trả lời phỏng vấn.
 - Ưu điểm: Thu thập dữ liệu qua điện thoại là một cách thức đơn giản, hữu hiệu và ít tốn kém.
 - Nhược điểm: Mẫu dự định chọn ban đầu có thể không hoàn thành theo dự kiến do một số người được hỏi từ chối trả lời hay không có nhà khi gọi đến,...

Khảo sát qua điện thoại

Khi thu thập dữ liệu bằng cách khảo sát qua điện thoại, người khảo sát gọi điện đến đối tượng cần thu thập dữ liệu để thực hiện cuộc trao đổi ngắn về vấn đề cần nghiên cứu.

Nội dung các câu trao đổi trong cuộc khảo sát được thiết kế thành văn bản gồm các phần:

- Phần đầu là những giải thích ngắn gọn về mục đích của cuộc khảo sát;
- Phần thứ hai là những câu hỏi liên quan đến vấn đề cần khảo sát. Các câu hỏi thường là những câu hỏi đóng, tức là yêu cầu người trả lời chọn một câu trả lời từ một số lựa chọn xác định;
- Phần cuối là những câu hỏi liên quan đến thông tin cá nhân của người trả lời phỏng vấn.
 - Ưu điểm: Thu thập dữ liệu qua điện thoại là một cách thức đơn giản, hữu hiệu và ít tốn kém.
 - Nhược điểm: Mẫu dự định chọn ban đầu có thể không hoàn thành theo dự kiến do một số người được hỏi từ chối trả lời hay không có nhà khi gọi đến,...

Khảo sát qua điện thoại

Khi thu thập dữ liệu bằng cách khảo sát qua điện thoại, người khảo sát gọi điện đến đối tượng cần thu thập dữ liệu để thực hiện cuộc trao đổi ngắn về vấn đề cần nghiên cứu.

Nội dung các câu trao đổi trong cuộc khảo sát được thiết kế thành văn bản gồm các phần:

- Phần đầu là những giải thích ngắn gọn về mục đích của cuộc khảo sát;
- Phần thứ hai là những câu hỏi liên quan đến vấn đề cần khảo sát. Các câu hỏi thường là những câu hỏi đóng, tức là yêu cầu người trả lời chọn một câu trả lời từ một số lựa chọn xác định;
- Phần cuối là những câu hỏi liên quan đến thông tin cá nhân của người trả lời phỏng vấn.
 - Ưu điểm: Thu thập dữ liệu qua điện thoại là một cách thức đơn giản, hữu hiệu và ít tốn kém.
 - Nhược điểm: Mẫu dự định chọn ban đầu có thể không hoàn thành theo dự kiến do một số người được hỏi từ chối trả lời hay không có nhà khi gọi đến,...

Khảo sát qua điện thoại

Khi thu thập dữ liệu bằng cách khảo sát qua điện thoại, người khảo sát gọi điện đến đối tượng cần thu thập dữ liệu để thực hiện cuộc trao đổi ngắn về vấn đề cần nghiên cứu.

Nội dung các câu trao đổi trong cuộc khảo sát được thiết kế thành văn bản gồm các phần:

- Phần đầu là những giải thích ngắn gọn về mục đích của cuộc khảo sát;
- Phần thứ hai là những câu hỏi liên quan đến vấn đề cần khảo sát. Các câu hỏi thường là những câu hỏi đóng, tức là yêu cầu người trả lời chọn một câu trả lời từ một số lựa chọn xác định;
- Phần cuối là những câu hỏi liên quan đến thông tin cá nhân của người trả lời phỏng vấn.
 - Ưu điểm: Thu thập dữ liệu qua điện thoại là một cách thức đơn giản, hữu hiệu và ít tốn kém.
 - Nhược điểm: Mẫu dự định chọn ban đầu có thể không hoàn thành theo dự kiến do một số người được hỏi từ chối trả lời hay không có nhà khi gọi đến,...

Khảo sát qua điện thoại

Khi thu thập dữ liệu bằng cách khảo sát qua điện thoại, người khảo sát gọi điện đến đối tượng cần thu thập dữ liệu để thực hiện cuộc trao đổi ngắn về vấn đề cần nghiên cứu.

Nội dung các câu trao đổi trong cuộc khảo sát được thiết kế thành văn bản gồm các phần:

- Phần đầu là những giải thích ngắn gọn về mục đích của cuộc khảo sát;
- Phần thứ hai là những câu hỏi liên quan đến vấn đề cần khảo sát. Các câu hỏi thường là những câu hỏi đóng, tức là yêu cầu người trả lời chọn một câu trả lời từ một số lựa chọn xác định;
- Phần cuối là những câu hỏi liên quan đến thông tin cá nhân của người trả lời phỏng vấn.
 - Ưu điểm: Thu thập dữ liệu qua điện thoại là một cách thức đơn giản, hữu hiệu và ít tốn kém.
 - Nhược điểm: Mẫu dự định chọn ban đầu có thể không hoàn thành theo dự kiến do một số người được hỏi từ chối trả lời hay không có nhà khi gọi đến,...

Khảo sát dạng viết

Thu thập dữ liệu bằng cách khảo sát qua dạng viết là hình thức khảo sát ở dạng một bức thư gửi qua đường bưu điện hoặc các bản câu hỏi phát tận tay người được phỏng vấn. Bản câu hỏi gồm cả các câu hỏi đóng và mở. Câu hỏi mở cho phép người được hỏi trả lời theo đánh giá, nhận định theo quan điểm của chính họ.

- Ưu điểm: Khảo sát dạng viết là phương tiện thu thập dữ liệu ít tốn kém nhất.
- Nhược điểm:
 - Bản câu hỏi yêu cầu chi tiết, dễ hiểu, rõ ràng và hình thức trình bày phải chuyên nghiệp.
 - Tỷ lệ trả lời thường thấp, nhất là hình thức gửi thư qua đường bưu điện.

Khảo sát dạng viết

Thu thập dữ liệu bằng cách khảo sát qua dạng viết là hình thức khảo sát ở dạng một bức thư gửi qua đường bưu điện hoặc các bản câu hỏi phát tận tay người được phỏng vấn. Bản câu hỏi gồm cả các câu hỏi đóng và mở. Câu hỏi mở cho phép người được hỏi trả lời theo đánh giá, nhận định theo quan điểm của chính họ.

- Ưu điểm: Khảo sát dạng viết là phương tiện thu thập dữ liệu ít tốn kém nhất.
- Nhược điểm:
 - Bản câu hỏi yêu cầu chi tiết, dễ hiểu, rõ ràng và hình thức trình bày phải chuyên nghiệp.
 - Tỷ lệ trả lời thường thấp, nhất là hình thức gửi thư qua đường bưu điện.

Khảo sát dạng viết

Thu thập dữ liệu bằng cách khảo sát qua dạng viết là hình thức khảo sát ở dạng một bức thư gửi qua đường bưu điện hoặc các bản câu hỏi phát tận tay người được phỏng vấn. Bản câu hỏi gồm cả các câu hỏi đóng và mở. Câu hỏi mở cho phép người được hỏi trả lời theo đánh giá, nhận định theo quan điểm của chính họ.

- Ưu điểm: Khảo sát dạng viết là phương tiện thu thập dữ liệu ít tốn kém nhất.
- Nhược điểm:
 - Bản câu hỏi yêu cầu chi tiết, dễ hiểu, rõ ràng và hình thức trình bày phải chuyên nghiệp.
 - Tỷ lệ trả lời thường thấp, nhất là hình thức gửi thư qua đường bưu điện.

Khảo sát dạng viết

Thu thập dữ liệu bằng cách khảo sát qua dạng viết là hình thức khảo sát ở dạng một bức thư gửi qua đường bưu điện hoặc các bản câu hỏi phát tận tay người được phỏng vấn. Bản câu hỏi gồm cả các câu hỏi đóng và mở. Câu hỏi mở cho phép người được hỏi trả lời theo đánh giá, nhận định theo quan điểm của chính họ.

- Ưu điểm: Khảo sát dạng viết là phương tiện thu thập dữ liệu ít tốn kém nhất.
- Nhược điểm:
 - Bản câu hỏi yêu cầu chi tiết, dễ hiểu, rõ ràng và hình thức trình bày phải chuyên nghiệp.
 - Tỷ lệ trả lời thường thấp, nhất là hình thức gửi thư qua đường bưu điện.

Khảo sát trực tiếp và phỏng vấn cá nhân

- Khảo sát trực tiếp là phương pháp thu thập dữ liệu dựa vào quan sát và cảm nhận của người thu thập dữ liệu đối với các đối tượng cần nghiên cứu.
- Phỏng vấn cá nhân là phương pháp thu thập dữ liệu từ các đối tượng nghiên cứu thông qua hỏi đáp. Phỏng vấn có thể sử dụng các câu hỏi mở hoặc đóng, có cấu trúc (có bản câu hỏi soạn thảo sẵn) hoặc không có cấu trúc (người phỏng vấn bắt đầu bằng một hoặc nhiều câu hỏi rồi phát triển thành những câu hỏi sau hơn dựa trên câu trả lời trước).

Câu hỏi tình huống

Câu hỏi tình huống

Khi nghiên cứu về việc các chai nước tương có chứa chất 3-MPCD các điều tra viên muốn điều tra toàn bộ các chai nước tương trên thị trường hay chỉ điều tra một số chai nước tương? Hãy nêu một số lí do mà trong nhiều trường hợp điều tra viên chỉ chọn cách điều tra mẫu mà không thể điều tra được tổng thể.

Nội dung trình bày

- 1 Xác định dữ liệu cần thu thập
- 2 Dữ liệu thứ cấp và dữ liệu sơ cấp
 - Khái niệm dữ liệu thứ cấp và sơ cấp
 - Nguồn dữ liệu thứ cấp và sơ cấp
 - Thu thập dữ liệu sơ cấp
- 3 Các kĩ thuật lấy mẫu
 - Lí do phải chọn mẫu và điều tra tổng thể
 - Một số khái niệm về chọn mẫu
 - Chọn mẫu ngẫu nhiên
 - Chọn mẫu không ngẫu nhiên
 - Lỗi do chọn mẫu và không do chọn mẫu

Lí do phải chọn mẫu và điều tra tổng thể

- Lí do phải chọn mẫu:

- Khi số lượng tổng thể lớn, chọn mẫu tiết kiệm thời gian, công sức, tiền của.
- Khi quá trình nghiên cứu phải phá hủy, chọn mẫu tiết kiệm được sản phẩm.
- Khi nghiên cứu tổng thể tiềm ẩn, chọn mẫu là sự lựa chọn duy nhất vì ta không thể nhận biết được hết các đơn vị tổng thể.
- Khi nguồn nghiên cứu cho trước, chọn mẫu giúp mở rộng phạm vi nghiên cứu.

- Lí do phải điều tra tổng thể

- Mẫu ta chọn không đại diện được cho tổng thể.
- Người điều tra không đánh giá cao việc chọn mẫu mà chỉ tin tưởng vào việc điều tra tổng thể.

Lí do phải chọn mẫu và điều tra tổng thể

- Lí do phải chọn mẫu:
 - Khi số lượng tổng thể lớn, chọn mẫu tiết kiệm thời gian, công sức, tiền của.
 - Khi quá trình nghiên cứu phải phá hủy, chọn mẫu tiết kiệm được sản phẩm.
 - Khi nghiên cứu tổng thể tiềm ẩn, chọn mẫu là sự lựa chọn duy nhất vì ta không thể nhận biết được hết các đơn vị tổng thể.
 - Khi nguồn nghiên cứu cho trước, chọn mẫu giúp mở rộng phạm vi nghiên cứu.
- Lí do phải điều tra tổng thể
 - Mẫu ta chọn không đại diện được cho tổng thể.
 - Người điều tra không đánh giá cao việc chọn mẫu mà chỉ tin tưởng vào việc điều tra tổng thể.

Nội dung trình bày

- 1 Xác định dữ liệu cần thu thập
- 2 Dữ liệu thứ cấp và dữ liệu sơ cấp
 - Khái niệm dữ liệu thứ cấp và sơ cấp
 - Nguồn dữ liệu thứ cấp và sơ cấp
 - Thu thập dữ liệu sơ cấp
- 3 Các kĩ thuật lấy mẫu
 - Lí do phải chọn mẫu và điều tra tổng thể
 - Một số khái niệm về chọn mẫu
 - Chọn mẫu ngẫu nhiên
 - Chọn mẫu không ngẫu nhiên
 - Lỗi do chọn mẫu và không do chọn mẫu

Chọn mẫu ngẫu nhiên và không ngẫu nhiên

- Chọn mẫu ngẫu nhiên là quá trình lựa chọn sao cho mỗi đơn vị tổng thể có cơ hội được chọn vào mẫu như nhau.
- Chọn mẫu không ngẫu nhiên là quá trình lựa chọn mà không phải mọi đơn vị tổng thể có cơ hội được chọn vào mẫu như nhau.

Chú ý: Thông tin chứa đựng trong mẫu muốn phản ánh chính xác tổng thể đang nghiên cứu, tức là đại diện được cho tổng thể, thì mẫu phải được chọn một cách ngẫu nhiên từ tổng thể. Trong các phần sau này, mọi mẫu được chọn ra đều giả sử có được từ một quá trình chọn mẫu ngẫu nhiên.

Chọn mẫu ngẫu nhiên và không ngẫu nhiên

- Chọn mẫu ngẫu nhiên là quá trình lựa chọn sao cho mỗi đơn vị tổng thể có cơ hội được chọn vào mẫu như nhau.
- Chọn mẫu không ngẫu nhiên là quá trình lựa chọn mà không phải mọi đơn vị tổng thể có cơ hội được chọn vào mẫu như nhau.

Chú ý: Thông tin chứa đựng trong mẫu muốn phản ánh chính xác tổng thể đang nghiên cứu, tức là đại diện được cho tổng thể, thì mẫu phải được chọn một cách ngẫu nhiên từ tổng thể. Trong các phần sau này, mọi mẫu được chọn ra đều giả sử có được từ một quá trình chọn mẫu ngẫu nhiên.

Chọn mẫu có hoàn lại và không hoàn lại

- Chọn mẫu có hoàn lại là cách chọn mà khi mỗi đơn vị tổng thể được chọn ra lại đặt trả lại tổng thể.
- Chọn mẫu không hoàn lại là cách chọn mà khi mỗi đơn vị tổng thể được chọn ra không được đặt trả lại tổng thể.

Nhận xét: Chọn mẫu không hoàn lại cho ta cái nhìn đầy đủ hơn về tổng thể so với chọn mẫu có hoàn lại vì chọn mẫu không hoàn lại đảm bảo rằng các phần tử trong mẫu là khác nhau, từ đó cho ta nghiên cứu được nhiều phần tử khác nhau nhất có thể trong tổng thể.

Chọn mẫu có hoàn lại và không hoàn lại

- Chọn mẫu có hoàn lại là cách chọn mà khi mỗi đơn vị tổng thể được chọn ra lại đặt trả lại tổng thể.
- Chọn mẫu không hoàn lại là cách chọn mà khi mỗi đơn vị tổng thể được chọn ra không được đặt trả lại tổng thể.

Nhận xét: Chọn mẫu không hoàn lại cho ta cái nhìn đầy đủ hơn về tổng thể so với chọn mẫu có hoàn lại vì chọn mẫu không hoàn lại đảm bảo rằng các phần tử trong mẫu là khác nhau, từ đó cho ta nghiên cứu được nhiều phần tử khác nhau nhất có thể trong tổng thể.

Câu hỏi tình huống

Câu hỏi tình huống

Chọn cách giải quyết hợp lí nếu như bạn gặp các tình huống sau:

- *Tình huống 1: Bạn là nhân viên của một siêu thị, nhân dịp kỉ niệm 5 năm ngày thành lập, siêu thị tổ chức chương trình "Khách hàng may mắn" cho những khách hàng mua hàng ở siêu thị bạn. Bạn được giao nhiệm vụ chọn ra 15 khách hàng may mắn trong mỗi ngày, bạn làm thế nào để chọn một cách khách quan nhất ra những khách hàng này?*
- *Tình huống 2: Công ty bạn làm việc đang thử nghiệm một hình thức thanh toán hóa đơn mới qua mạng nhằm giảm thời gian thanh toán chậm của các hóa đơn. Bạn được yêu cầu chọn ra 60 hóa đơn từ 3250 hóa đơn được gửi đến để ghi lại số ngày thanh toán chậm để xem xét tính hiệu quả của hình thức thanh toán mới.*

Câu hỏi tình huống

Câu hỏi tình huống

Chọn cách giải quyết hợp lí nếu như bạn gặp các tình huống sau:

- *Tình huống 1: Bạn là nhân viên của một siêu thị, nhân dịp kỉ niệm 5 năm ngày thành lập, siêu thị tổ chức chương trình "Khách hàng may mắn" cho những khách hàng mua hàng ở siêu thị bạn. Bạn được giao nhiệm vụ chọn ra 15 khách hàng may mắn trong mỗi ngày, bạn làm thế nào để chọn một cách khách quan nhất ra những khách hàng này?*
- *Tình huống 2: Công ty bạn làm việc đang thử nghiệm một hình thức thanh toán hóa đơn mới qua mạng nhằm giảm thời gian thanh toán chậm của các hóa đơn. Bạn được yêu cầu chọn ra 60 hóa đơn từ 3250 hóa đơn được gửi đến để ghi lại số ngày thanh toán chậm để xem xét tính hiệu quả của hình thức thanh toán mới.*

Câu hỏi tình huống

Câu hỏi tình huống

- *Tình huống 3: Để tìm hiểu về thời gian tự học ở nhà của sinh viên Thăng Long, bạn dự định chọn ra 200 sinh viên để điều tra về thời gian tự học ở nhà của họ. Bạn định làm thế nào để thực hiện được mục đích của mình?*
- *Tình huống 4: Để tìm hiểu về thời gian tự học ở nhà của sinh viên Đại học Kinh tế Quốc dân, bạn dự định chọn ra 200 sinh viên để điều tra về thời gian tự học ở nhà của họ. Bạn định làm thế nào để thực hiện được mục đích của mình?*
- *Tình huống 5: Bạn muốn tìm hiểu về thời gian xem ti vi mỗi ngày của các hộ gia đình ở Hà Nội. Làm thế nào để bạn điều tra được vấn đề này?*

Câu hỏi tình huống

Câu hỏi tình huống

- *Tình huống 3: Để tìm hiểu về thời gian tự học ở nhà của sinh viên Thăng Long, bạn dự định chọn ra 200 sinh viên để điều tra về thời gian tự học ở nhà của họ. Bạn định làm thế nào để thực hiện được mục đích của mình?*
- *Tình huống 4: Để tìm hiểu về thời gian tự học ở nhà của sinh viên Đại học Kinh tế Quốc dân, bạn dự định chọn ra 200 sinh viên để điều tra về thời gian tự học ở nhà của họ. Bạn định làm thế nào để thực hiện được mục đích của mình?*
- *Tình huống 5: Bạn muốn tìm hiểu về thời gian xem ti vi mỗi ngày của các hộ gia đình ở Hà Nội. Làm thế nào để bạn điều tra được vấn đề này?*

Câu hỏi tình huống

Câu hỏi tình huống

- *Tình huống 3: Để tìm hiểu về thời gian tự học ở nhà của sinh viên Thăng Long, bạn dự định chọn ra 200 sinh viên để điều tra về thời gian tự học ở nhà của họ. Bạn định làm thế nào để thực hiện được mục đích của mình?*
- *Tình huống 4: Để tìm hiểu về thời gian tự học ở nhà của sinh viên Đại học Kinh tế Quốc dân, bạn dự định chọn ra 200 sinh viên để điều tra về thời gian tự học ở nhà của họ. Bạn định làm thế nào để thực hiện được mục đích của mình?*
- *Tình huống 5: Bạn muốn tìm hiểu về thời gian xem ti vi mỗi ngày của các hộ gia đình ở Hà Nội. Làm thế nào để bạn điều tra được vấn đề này?*

Nội dung trình bày

- 1 Xác định dữ liệu cần thu thập
- 2 Dữ liệu thứ cấp và dữ liệu sơ cấp
 - Khái niệm dữ liệu thứ cấp và sơ cấp
 - Nguồn dữ liệu thứ cấp và sơ cấp
 - Thu thập dữ liệu sơ cấp
- 3 Các kĩ thuật lấy mẫu
 - Lí do phải chọn mẫu và điều tra tổng thể
 - Một số khái niệm về chọn mẫu
 - **Chọn mẫu ngẫu nhiên**
 - Chọn mẫu không ngẫu nhiên
 - Lỗi do chọn mẫu và không do chọn mẫu

Chọn mẫu ngẫu nhiên đơn giản

- Nếu số lượng đơn vị tổng thể ít (vài chục hoặc vài trăm), ta có thể chọn các đơn vị của mẫu bằng cách bốc thăm hay quay số.

Ví dụ: Nhân dịp kỷ niệm 5 năm ngày thành lập siêu thị, mỗi khách mua hàng được phát phiếu tham dự bốc thăm trúng thưởng có đánh số. Khách hàng điền thông tin vào phiếu, gấp lại và cho vào một cái hộp. 5h chiều nhân viên siêu thị tổ chức bốc thăm ngẫu nhiên ra 15 phiếu để trao quà cho các khách hàng may mắn trúng các giải nhất, nhì, ba.

Chọn mẫu ngẫu nhiên đơn giản

- Nếu số lượng đơn vị tổng thể ít (vài chục hoặc vài trăm), ta có thể chọn các đơn vị của mẫu bằng cách bốc thăm hay quay số.

Ví dụ: Nhân dịp kỷ niệm 5 năm ngày thành lập siêu thị, mỗi khách mua hàng được phát phiếu tham dự bốc thăm trúng thưởng có đánh số. Khách hàng điền thông tin vào phiếu, gấp lại và cho vào một cái hộp. 5h chiều nhân viên siêu thị tổ chức bốc thăm ngẫu nhiên ra 15 phiếu để trao quà cho các khách hàng may mắn trúng các giải nhất, nhì, ba.

Chọn mẫu ngẫu nhiên đơn giản

- Nếu số lượng đơn vị tổng thể là lớn (vài trăm, vài nghìn), việc chọn mẫu bằng bốc thăm trở nên mất thời gian, trong trường hợp này ta có thể chọn mẫu bằng cách dùng bảng số ngẫu nhiên.

Chọn mẫu ngẫu nhiên đơn giản

- Nếu số lượng đơn vị tổng thể là lớn (vài trăm, vài nghìn), việc chọn mẫu bằng bốc thăm trở nên mất thời gian, trong trường hợp này ta có thể chọn mẫu bằng cách dùng bảng số ngẫu nhiên.

Khái niệm

Bảng số ngẫu nhiên là một dãy số được chia thành các nhóm gồm năm chữ số để dễ đọc, trong đó mọi số gồm một chữ số trong bảng được giả sử là chọn ngẫu nhiên từ các chữ số từ 0 đến 9, mọi số gồm hai chữ số trong bảng được giả sử là chọn ngẫu nhiên từ các chữ số từ 00 đến 99, mọi số gồm ba chữ số trong bảng được giả sử là chọn ngẫu nhiên từ các chữ số từ 000 đến 999,...

Bảng số ngẫu nhiên

Bảng số ngẫu nhiên

33276	70997	79936	56865	05859	90106	31595	01547	85590	91610
78188	03427	49626	69445	18663	72695	52180	20847	12234	90511
33703	90322	92737	88974	33488	36320	17617	30015	08272	84115
27156	30613	74952	85689	48237	52267	67689	93394	01511	26358
85140	20285	29975	89868	08178	77233	13916	47564	81056	97735
85977	29372	74461	28551	90707	51259	77452	16308	60756	92144
49442	53900	70960	63990	75601	40719	60268	89368	19885	55322
44819	01188	65255	64835	44919	05944	55157	94904	31273	04146
18594	29852	71585	85030	51132	01915	92747	64951	58586	23216
14513	83149	98736	23495	64350	94738	17752	35156	35749	09998
42698	06691	76988	13602	51851	46104	88916	19509	25625	58104
14346	09172	30168	90229	04734	59193	22178	30421	61666	99904
32812	74103	47070	25306	76468	26384	58151	06646	21524	15227
96909	44592	24200	13363	38005	94342	28728	35806	06912	17012
64161	18296	22851	87308	58731	00256	45834	15398	46557	41135
10367	07684	36188	18510	07351	19731	92420	60952	61280	50001
67658	32586	86679	50720	94953	96423	24878	82651	66566	14778
76797	14780	13300	87074	79666	95725	26432	46901	20849	89768
81536	86645	12659	92259	57102	80428	25280	66432	84673	40027
32832	61362	98947	96067	64760	64584	96096	98253	26422	44407
44048	37937	63904	45766	66134	75470	66520	34693	90449	94305
26766	25940	39972	22209	71500	64568	91402	42416	07844	69618

Chọn mẫu ngẫu nhiên đơn giản

Ví dụ: Để xem xét số ngày thanh toán chậm của các hóa đơn gửi đến một công ty trong vòng một năm, người ta chọn ra 60 hóa đơn từ 3250 hóa đơn được gửi đến (đã được đánh số từ 1 đến 3250). Sử dụng bảng số ngẫu nhiên, chọn 60 hóa đơn này như sau:

- Vì tổng số hóa đơn là 3250 gồm bốn chữ số, nên ta sẽ ngẫu nhiên chọn một nhóm gồm bốn chữ số từ bảng, chẳng hạn 2234, đây là số của hóa đơn đầu tiên được chọn.
- Từ số này, ta di chuyển sang trái (phải, lên hoặc xuống) để thêm được các nhóm bốn chữ số nữa, chẳng hạn ta đi xuống chọn thêm được bảy nhóm gồm bốn chữ số nữa là: 2234, 8272, 1511, 1056, 0756, 9885, 1273. Do chỉ có 3250 hóa đơn nên ta loại nhóm số 8272 và 9885. Như vậy năm nhóm số được chọn đầu tiên là: 2234, 1511, 1056, 0756, 1273. Đây chính là số thứ tự của những hóa đơn sẽ được chọn để đo ngày.
- Cứ tiếp tục quá trình như vậy, ta sẽ thu được số thứ tự của 60 hóa đơn được chọn để đo số ngày thanh toán chậm.

Chọn mẫu ngẫu nhiên đơn giản

Ví dụ: Để xem xét số ngày thanh toán chậm của các hóa đơn gửi đến một công ty trong vòng một năm, người ta chọn ra 60 hóa đơn từ 3250 hóa đơn được gửi đến (đã được đánh số từ 1 đến 3250). Sử dụng bảng số ngẫu nhiên, chọn 60 hóa đơn này như sau:

- Vì tổng số hóa đơn là 3250 gồm bốn chữ số, nên ta sẽ ngẫu nhiên chọn một nhóm gồm bốn chữ số từ bảng, chẳng hạn 2234, đây là số của hóa đơn đầu tiên được chọn.
- Từ số này, ta di chuyển sang trái (phải, lên hoặc xuống) để thêm được các nhóm bốn chữ số nữa, chẳng hạn ta đi xuống chọn thêm được bảy nhóm gồm bốn chữ số nữa là: 2234, 8272, 1511, 1056, 0756, 9885, 1273. Do chỉ có 3250 hóa đơn nên ta loại nhóm số 8272 và 9885. Như vậy năm nhóm số được chọn đầu tiên là: 2234, 1511, 1056, 0756, 1273. Đây chính là số thứ tự của những hóa đơn sẽ được chọn để đo ngày.
- Cứ tiếp tục quá trình như vậy, ta sẽ thu được số thứ tự của 60 hóa đơn được chọn để đo số ngày thanh toán chậm.

Chọn mẫu ngẫu nhiên đơn giản

Ví dụ: Để xem xét số ngày thanh toán chậm của các hóa đơn gửi đến một công ty trong vòng một năm, người ta chọn ra 60 hóa đơn từ 3250 hóa đơn được gửi đến (đã được đánh số từ 1 đến 3250). Sử dụng bảng số ngẫu nhiên, chọn 60 hóa đơn này như sau:

- Vì tổng số hóa đơn là 3250 gồm bốn chữ số, nên ta sẽ ngẫu nhiên chọn một nhóm gồm bốn chữ số từ bảng, chẳng hạn 2234, đây là số của hóa đơn đầu tiên được chọn.
- Từ số này, ta di chuyển sang trái (phải, lên hoặc xuống) để thêm được các nhóm bốn chữ số nữa, chẳng hạn ta đi xuống chọn thêm được bảy nhóm gồm bốn chữ số nữa là: 2234, 8272, 1511, 1056, 0756, 9885, 1273. Do chỉ có 3250 hóa đơn nên ta loại nhóm số 8272 và 9885. Như vậy năm nhóm số được chọn đầu tiên là: 2234, 1511, 1056, 0756, 1273. Đây chính là số thứ tự của những hóa đơn sẽ được chọn để đo ngày.
- Cứ tiếp tục quá trình như vậy, ta sẽ thu được số thứ tự của 60 hóa đơn được chọn để đo số ngày thanh toán chậm.

Chọn mẫu ngẫu nhiên đơn giản

Ví dụ: Để xem xét số ngày thanh toán chậm của các hóa đơn gửi đến một công ty trong vòng một năm, người ta chọn ra 60 hóa đơn từ 3250 hóa đơn được gửi đến (đã được đánh số từ 1 đến 3250). Sử dụng bảng số ngẫu nhiên, chọn 60 hóa đơn này như sau:

- Vì tổng số hóa đơn là 3250 gồm bốn chữ số, nên ta sẽ ngẫu nhiên chọn một nhóm gồm bốn chữ số từ bảng, chẳng hạn 2234, đây là số của hóa đơn đầu tiên được chọn.
- Từ số này, ta di chuyển sang trái (phải, lên hoặc xuống) để thêm được các nhóm bốn chữ số nữa, chẳng hạn ta đi xuống chọn thêm được bảy nhóm gồm bốn chữ số nữa là: 2234, 8272, 1511, 1056, 0756, 9885, 1273. Do chỉ có 3250 hóa đơn nên ta loại nhóm số 8272 và 9885. Như vậy năm nhóm số được chọn đầu tiên là: 2234, 1511, 1056, 0756, 1273. Đây chính là số thứ tự của những hóa đơn sẽ được chọn để đo ngày.
- Cứ tiếp tục quá trình như vậy, ta sẽ thu được số thứ tự của 60 hóa đơn được chọn để đo số ngày thanh toán chậm.

Chọn mẫu ngẫu nhiên đơn giản

- Ta có thể dùng các phần mềm như R, excel, SPSS,... để chọn ra mẫu ngẫu nhiên. Chẳng hạn như trong ví dụ trên, dùng phần mềm thống kê R ta có thể tìm được 60 số ngẫu nhiên bằng lệnh:

```
> sample(1:3250, 60)
```

Hàm `sample` cho ta 60 nhóm số tương ứng với số thứ tự của các hóa đơn từ 1 đến 3250 như sau:

```
[1] 720 760 1163 1415 1451 2281 2558 3068 1808 2652 178 2473 1139  
[16] 1313 1761 759 187 3218 99 2869 2718 2929 288 479 2320 2931  
[31] 198 2046 2033 1602 932 856 564 2446 2886 15 1089 1772 1777  
[46] 1610 1667 2327 2804 2840 1414 2475 1842 1636 537 2689 455 1758
```

Lấy mẫu hệ thống

Khi số đơn vị mẫu khá nhiều hay rất nhiều (vài trăm, vài nghìn đơn vị mẫu), ta có thể sử dụng phương pháp lấy mẫu hệ thống.

- Lấy mẫu hệ thống khi tổng thể không được đánh số.
- Lấy mẫu hệ thống khi tổng thể được đánh số.

Lấy mẫu hệ thống khi tổng thể được đánh số

Giả sử ta cần lấy mẫu gồm n phần tử từ tổng thể có N phần tử đã được đánh số. Chia N đơn vị tổng thể thành k nhóm theo công thức $k = \frac{N}{n}$ (nếu k không phải số nguyên dương thì k được làm tròn đến số nguyên dương gần nhất).

- Trong k đơn vị đầu tiên, chọn ngẫu nhiên ra một phần tử, đây là đơn vị mẫu đầu tiên, các đơn vị mẫu tiếp theo được lấy cách đơn vị này một khoảng $k, 2k, 3k, \dots$
- Nếu đến hết danh sách N đơn vị chưa đủ n đơn vị mẫu ta quay trở lại đầu danh sách với qui ước $N + 1 = 1, N + 2 = 2, \dots$

Lấy mẫu hệ thống

Ví dụ: Giả sử ta cần chọn mẫu gồm $n = 10$ phần tử từ tổng thể có N phần tử:

- Tổng thể gồm $N = 63$ phần tử:
 - Tính $k = \frac{63}{10} = 6.3$, lấy k bằng 6.
 - Trong 6 đơn vị đầu tiên, chọn ngẫu nhiên ra một phần tử chẳng hạn là 4, khi đó các đơn vị mẫu được lấy ra là: 4, 10, 16, 22, 28, 34, 40, 46, 52, 58.
- Tổng thể gồm $N = 67$ phần tử:
 - Tính $k = \frac{67}{10} = 6.7$, lấy k bằng 7.
 - Trong 7 đơn vị đầu tiên, chọn ngẫu nhiên ra một phần tử chẳng hạn là 5, khi đó các đơn vị mẫu được lấy ra là: 5, 12, 19, 26, 33, 40, 47, 54, 61, 1 (61 + 7 - 67).

Lấy mẫu hệ thống

Ví dụ: Giả sử ta cần chọn mẫu gồm $n = 10$ phần tử từ tổng thể có N phần tử:

- Tổng thể gồm $N = 63$ phần tử:
 - Tính $k = \frac{63}{10} = 6.3$, lấy k bằng 6.
 - Trong 6 đơn vị đầu tiên, chọn ngẫu nhiên ra một phần tử chẳng hạn là 4, khi đó các đơn vị mẫu được lấy ra là: 4, 10, 16, 22, 28, 34, 40, 46, 52, 58.
- Tổng thể gồm $N = 67$ phần tử:
 - Tính $k = \frac{67}{10} = 6.7$, lấy k bằng 7.
 - Trong 7 đơn vị đầu tiên, chọn ngẫu nhiên ra một phần tử chẳng hạn là 5, khi đó các đơn vị mẫu được lấy ra là: 5, 12, 19, 26, 33, 40, 47, 54, 61, 1 (61 + 7 - 67).

Lấy mẫu phân tầng

Chọn mẫu phân tầng sử dụng khi các đơn vị tổng thể khác nhau về tính chất liên quan đến vấn đề nghiên cứu, khảo sát. Khi đó tổng thể được chia thành các tầng (lớp), mục tiêu là để các giá trị của các đối tượng tổng thể ta quan tâm thuộc cùng một tầng càng ít khác nhau càng tốt. Sau đó các đơn vị mẫu được chọn từ các tầng này theo các phương pháp lấy mẫu thông thường như lấy mẫu ngẫu nhiên đơn giản hay lấy mẫu hệ thống.

Lấy mẫu theo khối và lấy mẫu nhiều giai đoạn

- Chọn mẫu theo khối là cách chọn liên quan đến việc chia tổng thể thành nhiều khối, mỗi khối xem như một tổng thể con, lấy ngẫu nhiên đơn giản ra m khối, sau đó khảo sát hết các đối tượng trong các khối mẫu được lấy ra.
- Chọn mẫu theo giai đoạn là kiểu chọn mà khi các khối quá lớn, ta phải chia thành các khối nhỏ hơn và lấy các đơn vị mẫu từ các khối nhỏ này.

Lấy mẫu theo khối và lấy mẫu nhiều giai đoạn

- Chọn mẫu theo khối là cách chọn liên quan đến việc chia tổng thể thành nhiều khối, mỗi khối xem như một tổng thể con, lấy ngẫu nhiên đơn giản ra m khối, sau đó khảo sát hết các đối tượng trong các khối mẫu được lấy ra.
- Chọn mẫu theo giai đoạn là kiểu chọn mà khi các khối quá lớn, ta phải chia thành các khối nhỏ hơn và lấy các đơn vị mẫu từ các khối nhỏ này.

Nội dung trình bày

- 1 Xác định dữ liệu cần thu thập
- 2 Dữ liệu thứ cấp và dữ liệu sơ cấp
 - Khái niệm dữ liệu thứ cấp và sơ cấp
 - Nguồn dữ liệu thứ cấp và sơ cấp
 - Thu thập dữ liệu sơ cấp
- 3 Các kĩ thuật lấy mẫu
 - Lí do phải chọn mẫu và điều tra tổng thể
 - Một số khái niệm về chọn mẫu
 - Chọn mẫu ngẫu nhiên
 - **Chọn mẫu không ngẫu nhiên**
 - Lỗi do chọn mẫu và không do chọn mẫu

Chọn mẫu không ngẫu nhiên

- Chọn mẫu thuận tiện: các đơn vị mẫu được chọn ra dựa trên sự thuận tiện của người điều tra.
- Chọn mẫu phán đoán: các đơn vị mẫu được chọn ra theo sự phán đoán của người điều tra.
- Chọn mẫu định mức: người điều tra sẽ quyết định các tổng thể con cần chia ra và tỉ lệ của tổng thể con trong mẫu cần lấy ra.

Câu hỏi: Tại sao các kiểu chọn mẫu trên lại không ngẫu nhiên?

Chọn mẫu không ngẫu nhiên

- Chọn mẫu thuận tiện: các đơn vị mẫu được chọn ra dựa trên sự thuận tiện của người điều tra.
- Chọn mẫu phán đoán: các đơn vị mẫu được chọn ra theo sự phán đoán của người điều tra.
- Chọn mẫu định mức: người điều tra sẽ quyết định các tổng thể con cần chia ra và tỉ lệ của tổng thể con trong mẫu cần lấy ra.

Câu hỏi: Tại sao các kiểu chọn mẫu trên lại không ngẫu nhiên?

Chọn mẫu không ngẫu nhiên

- Chọn mẫu thuận tiện: các đơn vị mẫu được chọn ra dựa trên sự thuận tiện của người điều tra.
- Chọn mẫu phán đoán: các đơn vị mẫu được chọn ra theo sự phán đoán của người điều tra.
- Chọn mẫu định mức: người điều tra sẽ quyết định các tổng thể con cần chia ra và tỉ lệ của tổng thể con trong mẫu cần lấy ra.

Câu hỏi: Tại sao các kiểu chọn mẫu trên lại không ngẫu nhiên?

Chọn mẫu không ngẫu nhiên

- Chọn mẫu thuận tiện: các đơn vị mẫu được chọn ra dựa trên sự thuận tiện của người điều tra.
- Chọn mẫu phán đoán: các đơn vị mẫu được chọn ra theo sự phán đoán của người điều tra.
- Chọn mẫu định mức: người điều tra sẽ quyết định các tổng thể con cần chia ra và tỉ lệ của tổng thể con trong mẫu cần lấy ra.

Câu hỏi: Tại sao các kiểu chọn mẫu trên lại không ngẫu nhiên?

Nội dung trình bày

- 1 Xác định dữ liệu cần thu thập
- 2 Dữ liệu thứ cấp và dữ liệu sơ cấp
 - Khái niệm dữ liệu thứ cấp và sơ cấp
 - Nguồn dữ liệu thứ cấp và sơ cấp
 - Thu thập dữ liệu sơ cấp
- 3 Các kĩ thuật lấy mẫu
 - Lí do phải chọn mẫu và điều tra tổng thể
 - Một số khái niệm về chọn mẫu
 - Chọn mẫu ngẫu nhiên
 - Chọn mẫu không ngẫu nhiên
 - Lỗi do chọn mẫu và không do chọn mẫu

Lỗi do chọn mẫu

- Sai lầm do chọn mẫu xảy ra khi mẫu không đại diện được cho tổng thể. Khi mẫu không đại diện được cho tổng thể thì các tính toán trên mẫu không thể suy diễn cho tổng thể được.

Lỗi do chọn mẫu

- Sai lầm do chọn mẫu xảy ra khi mẫu không đại diện được cho tổng thể. Khi mẫu không đại diện được cho tổng thể thì các tính toán trên mẫu không thể suy diễn cho tổng thể được.

Một ví dụ kinh điển cho sai lầm do chọn mẫu đó là cuộc bầu cử tổng thống của Mỹ năm 1936. Có hai ứng cử viên sáng giá cho chức vụ tổng thống là: Franklin D. Roosevelt (Đảng Dân chủ) và Alf Landon (Đảng Cộng hòa). Có rất nhiều tạp chí và tổ chức tham gia dự đoán kết quả của cuộc bầu cử trong đó có tạp chí Literary Digest và nhóm thăm dò dư luận George Gallup do George Gallup sáng lập.

Lỗi do chọn mẫu

- Tạp chí Literary Digest đã chọn một mẫu gồm 2.4 triệu người chọn từ danh bạ điện thoại và chọn từ tên các thành viên của các câu lạc bộ để điều tra. Tạp chí này đã dự đoán rằng Alf Landon sẽ đánh bại Franklin D.Roosevelt với tỉ lệ là 57% và 43%.
- Nhóm George Gallup đã chọn một mẫu gồm 5000 người theo một phương pháp chọn mẫu kiểu hệ thống và dự đoán Roosevelt sẽ dành chiến thắng trước đối thủ Alf Landon.
- Thực tế Roosevelt đã dành chiến thắng lớn. Như vậy nhóm George Gallup đã dự đoán đúng còn tạp chí Literary Digest đã dự đoán sai. Nguyên nhân vì sao?

Lỗi do chọn mẫu

- Tạp chí Literary Digest đã chọn một mẫu gồm 2.4 triệu người chọn từ danh bạ điện thoại và chọn từ tên các thành viên của các câu lạc bộ để điều tra. Tạp chí này đã dự đoán rằng Alf Landon sẽ đánh bại Franklin D.Roosevelt với tỉ lệ là 57% và 43%.
- Nhóm George Gallup đã chọn một mẫu gồm 5000 người theo một phương pháp chọn mẫu kiểu hệ thống và dự đoán Roosevelt sẽ dành chiến thắng trước đối thủ Alf Landon.
- Thực tế Roosevelt đã dành chiến thắng lớn. Như vậy nhóm George Gallup đã dự đoán đúng còn tạp chí Literary Digest đã dự đoán sai. Nguyên nhân vì sao?

Lỗi do chọn mẫu

- Tạp chí Literary Digest đã chọn một mẫu gồm 2.4 triệu người chọn từ danh bạ điện thoại và chọn từ tên các thành viên của các câu lạc bộ để điều tra. Tạp chí này đã dự đoán rằng Alf Landon sẽ đánh bại Franklin D.Roosevelt với tỉ lệ là 57% và 43%.
- Nhóm George Gallup đã chọn một mẫu gồm 5000 người theo một phương pháp chọn mẫu kiểu hệ thống và dự đoán Roosevelt sẽ dành chiến thắng trước đối thủ Alf Landon.
- Thực tế Roosevelt đã dành chiến thắng lớn. Như vậy nhóm George Gallup đã dự đoán đúng còn tạp chí Literary Digest đã dự đoán sai. Nguyên nhân vì sao?

Lỗi do chọn mẫu

- Sai lầm của Literary Digest là đã chọn những người để điều tra từ danh bạ điện thoại và tên của các thành viên của các câu lạc bộ. Năm 1936, Mỹ chưa phục hồi từ cuộc đại suy thoái nên rất nhiều người thất nghiệp và có thu nhập thấp không có điện thoại và cũng không tham gia vào câu lạc bộ nào. Chương trình chọn mẫu của Literary Digest đã bỏ qua những người này, mà chính họ đã bỏ phiếu ủng hộ Roosevelt. Còn nhóm George Gallup với phương pháp chọn mẫu thích hợp đã dự đoán chính xác chiến thắng của Roosevelt.

Lỗi do chọn mẫu

- Sai lầm của Literary Digest là đã chọn những người để điều tra từ danh bạ điện thoại và tên của các thành viên của các câu lạc bộ. Năm 1936, Mỹ chưa phục hồi từ cuộc đại suy thoái nên rất nhiều người thất nghiệp và có thu nhập thấp không có điện thoại và cũng không tham gia vào câu lạc bộ nào. Chương trình chọn mẫu của Literary Digest đã bỏ qua những người này, mà chính họ đã bỏ phiếu ủng hộ Roosevelt. Còn nhóm George Gallup với phương pháp chọn mẫu thích hợp đã dự đoán chính xác chiến thắng của Roosevelt.

Lỗi do chọn mẫu

- Sai lầm của Literary Digest là đã chọn những người để điều tra từ danh bạ điện thoại và tên của các thành viên của các câu lạc bộ. Năm 1936, Mỹ chưa phục hồi từ cuộc đại suy thoái nên rất nhiều người thất nghiệp và có thu nhập thấp không có điện thoại và cũng không tham gia vào câu lạc bộ nào. Chương trình chọn mẫu của Literary Digest đã bỏ qua những người này, mà chính họ đã bỏ phiếu ủng hộ Roosevelt. Còn nhóm George Gallup với phương pháp chọn mẫu thích hợp đã dự đoán chính xác chiến thắng của Roosevelt.

Lỗi do chọn mẫu

- Sai lầm của Literary Digest là đã chọn những người để điều tra từ danh bạ điện thoại và tên của các thành viên của các câu lạc bộ. Năm 1936, Mỹ chưa phục hồi từ cuộc đại suy thoái nên rất nhiều người thất nghiệp và có thu nhập thấp không có điện thoại và cũng không tham gia vào câu lạc bộ nào. Chương trình chọn mẫu của Literary Digest đã bỏ qua những người này, mà chính họ đã bỏ phiếu ủng hộ Roosevelt. Còn nhóm George Gallup với phương pháp chọn mẫu thích hợp đã dự đoán chính xác chiến thắng của Roosevelt.

Những lỗi không do chọn mẫu

- Những lỗi không phải do chọn mẫu được gọi là lỗi không do chọn mẫu bao gồm: lỗi do mất dữ liệu, do ghi chép sai, do phân tích sai, lỗi do người trả lời phỏng vấn không hiểu, không nói hoặc phóng đại vấn đề được hỏi,... Hầu như không có phương pháp thống kê nào đo hoặc điều khiển được những lỗi kiểu này, tuy nhiên trong các kỹ thuật thống kê sau này, ta đều giả sử rằng không có lỗi kiểu không do chọn mẫu xảy ra.

Bài tập ôn luyện: Thuốc nào tốt hơn

Bài toán (Nghịch lí Simpson)

Một người nghiên cứu muốn xác định xem giữa hai loại thuốc cùng để chữa một loại bệnh, loại nào tốt hơn. Kết quả thống kê về lượng người chữa được khỏi bệnh, phân biệt theo giới tính, được viết dưới đây:

Giới tính: Nữ	Thuốc I	Thuốc II
Chữa được	150	15
Không chữa được	850	285
Giới tính: Nam	Thuốc I	Thuốc II
Chữa được	190	720
Không chữa được	10	180

Hỏi thuốc nào tốt hơn?

Bài toán (Nghịch lí Simpson)

Dựa vào bảng thống kê trên, có hai câu trả lời trái ngược nhau cho câu hỏi thuốc nào tốt hơn như sau:

- 1 Thuốc I đem cho 1200 người dùng, chữa được bệnh cho 340 người. Thuốc II đem cho 1200 người dùng, chữa được 735 người. Như vậy thuốc II tốt hơn.
- 2 Đối với nữ, tỉ lệ chữa được bệnh của thuốc I là 15%, của thuốc II là 5%. Đối với nam, tỉ lệ chữa được bệnh của thuốc I là 95%, của thuốc II là 80%. Trong cả hai trường hợp thì tỉ lệ chữa được bệnh của thuốc I cao hơn, vậy thuốc I tốt hơn.

Nghịch lí nằm ở đâu?

Trong từng tình huống đưa ra trong phần câu hỏi tình huống về chọn mẫu:

- Hãy đưa ra phương pháp chọn mẫu thích hợp.
- Giải thích lí do lựa chọn.
- Mô tả sơ lược cách thực hiện.