

THỐNG KÊ MÔ TẢ

Chương 2

THU THẬP DỮ LIỆU

1. Dữ liệu
2. Các phương pháp thu thập dữ liệu ban đầu
3. Điều tra thống kê
4. Các phương pháp chọn mẫu ngẫu nhiên

1. Dữ liệu

1.2 Xác định dữ liệu cần thu thập

Vấn đề, sự kiện cần nghiên cứu có rất nhiều thông tin, dữ liệu liên quan.

a) Hậu quả nếu không xác định:

- Hao tổn nguồn lực (thời gian, chi phí, công sức) khi thu thập những dữ liệu không quan trọng, không cần thiết, không liên quan đến vấn đề, sự kiện nghiên cứu.
- Thu thập thiếu dữ liệu cần thiết, quan trọng
→ chất lượng nghiên cứu không tốt.

1. Dữ liệu

1.2 Xác định dữ liệu cần thu thập

Ví dụ: nghiên cứu vấn đề sinh viên làm thêm ảnh hưởng đến kết quả học tập.

Những câu hỏi sau có liên quan đến dữ liệu như:

Why?

- Có đi làm thêm không?
- Mức độ thường xuyên công việc làm thêm như thế nào?
- Thời gian làm thêm hằng ngày, hằng tuần?
- Chỗ làm xa hoặc gần nơi học?
- Có thích thú với công việc làm thêm?

1. Dữ liệu

1.1. Khái niệm và phân loại

a) Khái niệm: Dữ liệu là kết quả khảo sát của các biến. Hay gọi cách khác là giá trị của biến.

b) Phân loại: theo phân loại của biến:

- Dữ liệu **định tính**: phản ánh tính chất, thuộc tính hoặc loại hình của các đối tượng nghiên cứu, để thu thập hơn (vì không có quá nhiều giá trị).
- Dữ liệu **định lượng**: phản ánh mức độ, giá trị, để áp dụng phân tích thống kê hơn.

1. Dữ liệu

1.2 Xác định dữ liệu cần thu thập

b) Thu thập dữ liệu cần xác định rõ:

- Những dữ liệu cần thu thập
- Thứ tự ưu tiên các dữ liệu này

c) Nguyên tắc xác định:

- Dựa vào nội dung nghiên cứu
- Dựa vào mục đích nghiên cứu

1. Dữ liệu

1.2 Xác định dữ liệu cần thu thập

Ví dụ: nghiên cứu vấn đề sinh viên làm thêm ảnh hưởng đến kết quả học tập.

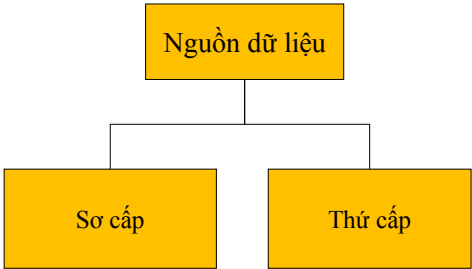
Những câu hỏi sau không liên quan đến dữ liệu như:

Why?

- Có mặc đồng phục khi đi làm thêm không?
- Người quản lý là nam hay nữ?
- Việc làm này do tự kiếm hay được người quen giới thiệu?
- Có bạn học chung làm cùng không?

1. Dữ liệu

1.3. Nguồn dữ liệu



1. Dữ liệu

1.3. Nguồn dữ liệu

1.3.1 Nguồn dữ liệu thứ cấp:

c) Đặc điểm: thu thập nhanh, rẻ nhưng không có đủ dữ liệu theo mong muốn, thường ở tầm vĩ mô.

1.3.2. Nguồn dữ liệu sơ cấp:

Là dữ liệu tự tiến hành, tổ chức thu thập, tự tạo ra dữ liệu.

2. Các PP thu thập dữ liệu ban đầu

a) Theo dõi

- Là thu thập dữ liệu bằng cách **theo dõi** các hành động, thái độ, diễn tiến của đối tượng nghiên cứu trong những tình huống nhất định và ghi nhận lại kết quả xảy ra.
- Ví dụ: quan sát khách đến nơi trưng bày sản phẩm công ty trong hội chợ triển lãm
- Phù hợp sự kiện, vấn đề tính chất tự nhiên, không có sự tác động bên ngoài.

1. Dữ liệu

1.3. Nguồn dữ liệu

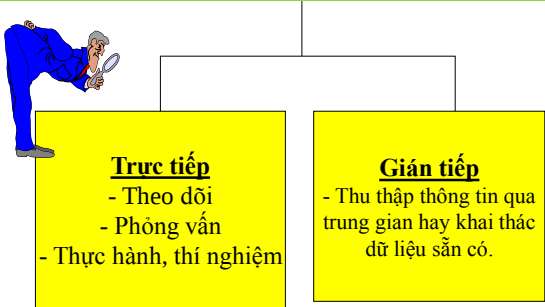
1.3.1. Nguồn dữ liệu thứ cấp:

a) Khái niệm: Là dữ liệu thu thập được từ những nguồn có sẵn.

b) Tính chất: Đa dạng và phong phú. Đối với công ty, doanh nghiệp có thể sử dụng các nguồn sau:

- Nội bộ: báo cáo các phòng, ban công ty
- Cơ quan thống kê nhà nước, cơ quan chính phủ
- Báo đài, tổ chức, hiệp hội cung cấp thông tin

2. Các PP thu thập dữ liệu ban đầu



2. Các PP thu thập dữ liệu ban đầu

b) Phỏng vấn

- Là trực tiếp hỏi đối tượng được chọn điều tra và tự ghi chép dữ liệu vào bản câu hỏi hay phiếu điều tra.
- Phù hợp với những điều tra phức tạp cần thu thập nhiều dữ liệu và đối tượng điều tra là con người.
- Dữ liệu thu thập đầy đủ theo nội dung điều tra và có độ chính xác cao nhưng đòi hỏi chi phí lớn.

2. Các PP thu thập dữ liệu ban đầu

- c) Thực hành, thí nghiệm
- Tiến hành các thí nghiệm, thực hành và ghi nhận lại các dữ liệu kết quả của mỗi lần thực hiện.
 - Phù hợp với những vấn đề, sự kiện nghiên cứu lĩnh vực kĩ thuật, khoa học thực nghiệm, công nghệ.

3. Điều tra thống kê

3.1. Phân loại

- a) Điều tra thường xuyên:
- Thu thập thông tin liên tục có hệ thống theo sát với quá trình biến động của hiện tượng nghiên cứu.
 - Ví dụ:
 - Điều tra biến động nhân khẩu địa phương (sinh, tử, di, đến)
 - Tình hình giá cả thị trường...

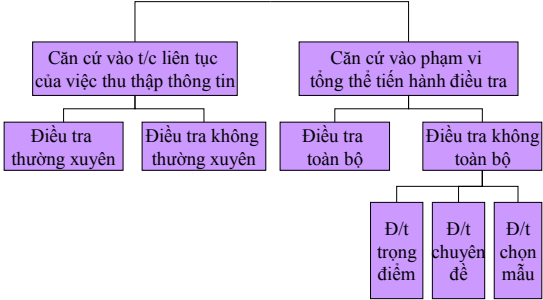
3. Điều tra thống kê

3.1. Phân loại

- c) Điều tra toàn bộ:
- Là tiến hành điều tra tất cả cá thể của tổng thể nên còn gọi là tổng điều tra.
 - Ví dụ: Tổng điều tra dân số
Tổng điều tra nông nghiệp
 - Cung cấp đầy đủ nhất cho nghiên cứu nhưng đòi hỏi chi phí rất lớn về thời gian, công sức, kinh phí.

3. Điều tra thống kê

3.1. Phân loại



3. Điều tra thống kê

3.1. Phân loại

- b) Điều tra không thường xuyên:
- Tiến hành thu thập thông tin không liên tục, phản ánh trạng thái của hiện tượng ở một thời điểm hay thời kỳ nhất định, chỉ tiến hành khi có nhu cầu.
 - Thường dùng cho các hiện tượng cần theo dõi thường xuyên nhưng chi phí điều tra lớn, hoặc các hiện tượng không cần theo dõi thường xuyên.

3. Điều tra thống kê

3.1. Phân loại

- d) Điều tra không toàn bộ:
- Thu thập thông tin của một số cá thể được chọn từ tổng thể hiện tượng nghiên cứu.
 - Mục đích: Có thông tin làm căn cứ nhận định hoặc suy rộng cho tổng thể chung.
 - Gồm :
 - Điều tra trọng điểm
 - Điều tra chuyên đề
 - Điều tra chọn mẫu

3. Điều tra thống kê

3.1. Phân loại

e) Điều tra trong điểm:

- Chỉ tiến hành thu thập thông tin ở bộ phận chủ yếu, chiếm tỷ trọng lớn, tập trung nhất của tổng thể.
- Kết quả điều tra không dùng để suy rộng cho toàn tổng thể nhưng giúp cho việc nắm được những đặc điểm cơ bản của hiện tượng.
- Ví dụ: khi cần nắm bắt nhanh cơ bản về sản xuất cao su, cà phê có thể tiến hành điều tra ở Đông Nam Bộ, Tây Nguyên

3. Điều tra thống kê

3.1. Phân loại

f) Điều tra chuyên đề:

- Kết quả điều tra không dùng để suy rộng hoặc làm căn cứ đánh giá tình hình cơ bản của hiện tượng mà chỉ rút ra kết luận về các cá thể được điều tra.
- Ví dụ nghiên cứu số ít sinh viên có thành tích học xuất sắc.

3. Điều tra thống kê

3.2. Xây dựng kế hoạch điều tra

a) Khái niệm:

Kế hoạch điều tra là tài liệu dạng văn bản đề cập những vấn đề cần giải quyết hoặc được hiểu thống nhất, trình tự và phương pháp tiến hành cuộc điều tra.

3. Điều tra thống kê

3.1. Phân loại

f) Điều tra chuyên đề:

- Là điều tra để thu thập thông tin tiến hành trên số rất ít cá thể nhưng đi sâu nghiên cứu nhiều khía cạnh của cá thể.
- Thường dùng nghiên cứu những diễn hình (tốt, xấu) để tìm hiểu nguyên nhân ảnh hưởng đến hiện tượng nghiên cứu.

3. Điều tra thống kê

3.1. Phân loại

g) Điều tra chọn mẫu:

- Là tiến hành điều tra **ngẫu nhiên** thu thập thông tin trên một số cá thể của tổng thể để thu thập dữ liệu thực tế.
- Kết quả điều tra dùng để suy rộng cho cả tổng thể chung.
- Được dùng nhiều nhất trong nghiên cứu vì tiết kiệm thời gian, chi phí và dữ liệu đáng tin cậy.



3. Điều tra thống kê

3.2. Xây dựng kế hoạch điều tra

b) Nội dung cơ bản của kế hoạch:

Gồm một số vấn đề chủ yếu sau:

- Mô tả mục đích, yêu cầu điều tra
- Xác định đối tượng và cá thể điều tra
- Nội dung điều tra
- Xác định thời gian và thời kì điều tra
- Xác định hình thức, phương pháp điều tra

3. Điều tra thống kê

3.2. Xây dựng kế hoạch điều tra

b) Nội dung cơ bản của kế hoạch:

Gồm một số vấn đề chủ yếu sau:

- Thiết kế phiếu điều tra, XD bảng biểu tổng hợp
- Tuyển chọn điều tra viên, giám sát viên
- Xây dựng chương trình xử lý tổng hợp, phân tích số liệu.

3. Điều tra thống kê

3.3. Sai số trong điều tra

a) Khái niệm:

Là chênh lệch giữa các trị số của tiêu thức điều tra mà ta thu thập được so với trị số thực tế của cá thể điều tra.

Sai số điều tra làm giảm chất lượng của kết quả điều tra và ảnh hưởng đến chất lượng của cả quá trình nghiên cứu thống kê.

Tuy nhiên thực tế khó thể biết được sai số và tránh được hoàn toàn sai số.

3. Điều tra thống kê

3.3. Sai số trong điều tra

b) Phân loại sai số:

- + Sai số chọn mẫu (khách quan)
- + Sai số phi chọn mẫu (chủ quan)

3. Điều tra thống kê

3.3. Sai số trong điều tra

c) Sai số chọn mẫu:

Là sai số do tính chất đại biểu, là sai số xảy ra trong điều tra không toàn bộ, nhất là trong điều tra chọn mẫu.

Nguyên nhân do việc lựa chọn cá thể điều tra thực tế không có tính đại diện cao

Là điều khó tránh khỏi vì việc chọn mẫu có kết cấu giống tổng thể rất khó thực hiện.

Có thể giảm bằng cách tăng cỡ mẫu.

3. Điều tra thống kê

3.3. Sai số trong điều tra

c) Sai số phi chọn mẫu:

Là sai số trong quá trình thực hiện điều tra do các nguyên chủ quan như:

- + Kế hoạch điều tra sai, không khoa học, không sát thực tế
- + Trình độ nhân viên điều tra, không hiểu rõ nội dung, không biết cách khai thác dữ liệu

3. Điều tra thống kê

3.3. Sai số trong điều tra

d) Sai số phi chọn mẫu:

Là sai số trong quá trình thực hiện điều tra do các nguyên chủ quan như:

- + Vô tình ghi chép sai, hiệu chỉnh, nhập liệu sai
- + Cá thể điều tra không hiểu, không muốn cộng tác, cố tình cung cấp dữ liệu sai. ...v...v...

3. Điều tra thống kê

3.3. Sai số trong điều tra

e) Biện pháp hạn chế:

- Làm tốt công tác chuẩn bị điều tra.
- Tiến hành kiểm tra có hệ thống toàn bộ cuộc điều tra
- Làm tốt công tác tuyên truyền với các cá thể được điều tra và nâng cao tinh trách nhiệm đối với nhân viên điều tra.

4. Các phương pháp chọn mẫu ngẫu nhiên

4.2. Đơn giản:

Thực hiện kỹ thuật chọn số ngẫu nhiên nhiều lần đến khi lấy đủ lượng mẫu

4.3. Hệ thống:

- Tính khoảng cách chọn mẫu $k = N/n$ (làm tròn).
- Dùng kỹ thuật chọn số ngẫu nhiên một lần duy nhất cho cá thể đầu tiên. Các cá thể sau chọn cách đều nhau một khoảng cách chọn mẫu k .

4. Các phương pháp chọn mẫu ngẫu nhiên

4.3. Hệ thống:

- Chọn đường thẳng: cá thể đầu tiên lấy trong khối đầu tiên bằng kỹ thuật chọn một số ngẫu nhiên thuộc khoảng **từ 1 đến k** (giả sử có thứ tự **e**). Các cá thể kế tiếp cách lần lượt có thứ tự $e + k, e + 2k, \dots$ cho đến khi lấy đủ mẫu.

4. Các phương pháp chọn mẫu ngẫu nhiên

4.1. Các kỹ thuật liên quan:

- a) Lập dàn mẫu: gán số thứ tự cho tất cả cá thể của tổng thể → yêu cầu có danh sách tổng thể
- b) Chọn số ngẫu nhiên:
 - Bốc thăm, quay số.
 - Bảng số ngẫu nhiên.
 - Hàm Randbetween trong phần mềm Excel:
 - + Cú pháp: “=randbetween(m,n)”
 - + Kết quả: cho một số ngẫu nhiên $x, m \leq x \leq n$

4. Các phương pháp chọn mẫu ngẫu nhiên

4.3. Hệ thống:

- Chọn quay vòng:
 - + Cá thể đầu tiên lấy bằng cách chọn một số thứ tự ngẫu nhiên **từ 1 đến N** (giả sử có thứ tự **m**)
 - + Các cá thể kế tiếp lần lượt có thứ tự: $m + k, m + 2k, \dots$
 - + Nếu chọn đến cá thể có số thứ tự $m + ik > N$, lấy quay vòng trở lại trong khối đầu tiên cá thể có số thứ tự $(m + ik - N)$ và tiếp tục tương tự cho đủ số lượng mẫu.

4. Các phương pháp chọn mẫu ngẫu nhiên

4.4. Phân tầng:

- Tổng thể chia thành nhiều **m tầng** theo tính chất, đặc điểm tự nhiên. Gọi N_i là số lượng tầng thứ i .
- Phân bổ cỡ mẫu:
 - + Phân đều: lượng mẫu mọi tầng = $n/m \rightarrow$ tỷ trọng tầng thứ $i: f_i = \frac{N_i}{N} . m$
 - + Theo tỷ lệ kết cấu: lượng mẫu tầng $i: n_i = \frac{N_i}{N} . n$

4. Các phương pháp chọn mẫu ngẫu nhiên

4.5. Cả khối/cum:
Tổng thể được chia ngẫu nhiên thành M khối, chọn ngẫu nhiên m khối và khảo sát tất cả cá thể trong m khối được chọn.
Nếu tổng thể có nhiều cấp thì tiếp tục thực hiện tương tự cho các cấp nhỏ hơn của nó gọi là chọn mẫu nhiều giai đoạn.

4. Các phương pháp chọn mẫu ngẫu nhiên

Ví dụ: Tổng thể khối kinh tế có 12000 sinh viên, trong đó ngành Quản trị có 4000SV, ngành Kế toán có 5000SV, ngành Tài – ngân có 3000SV. Hãy trình bày cách lấy mẫu 4000SV theo hai phương pháp phân tầng và hệ thống kết hợp.

5. Bài tập thực hành nghiên cứu thống kê

CÔNG VIỆC THỰC HIỆN:

1. Lập kế hoạch nghiên cứu

- Mục đích, yêu cầu
- Đối tượng, phạm vi
- Tiến độ thực hiện
- Tổ chức thực hiện

2. Thiết kế phiếu điều tra / bảng hỏi.

Ít nhất 10 câu hỏi và có ít nhất 2 câu hỏi dữ liệu định lượng không tính các câu hỏi thông tin cá nhân.

4. Các phương pháp chọn mẫu ngẫu nhiên

4.6. Đặc điểm:

	Biết cỡ tổng thể	Lập dàn mẫu	Phân biệt tính chất cá thể	Chia khối tổng thể
Đơn giản	Có	Có	Không	Không
Hệ thống	Có	Có	Không	Không
Phân tầng	Có	Không	Có	Có
Cả khối	Không	Không	Không	Có

Tùy (mục đích, tính chất) vấn đề, sự kiện nghiên cứu cụ thể, các phương pháp có thể vận dụng phối hợp để hỗ trợ cho nhau.

5. Bài tập thực hành nghiên cứu thống kê

TỔ CHỨC THỰC HIỆN:

- Tự chọn thành lập tổ, số lượng không quá 6 sinh viên / tổ.
- Các tổ lựa chọn một đề tài trong cuộc sống. Khuyến khích các đề tài gần gũi với sinh viên.
- Đăng kí đề tài với giáo viên. Đề tài trùng nhau thì tổ đăng kí sau phải thay đổi đề tài khác.

5. Bài tập thực hành nghiên cứu thống kê

CÔNG VIỆC THỰC HIỆN:

3. Thu thập dữ liệu (mẫu).

- Điều tra thực tế: Không ít hơn 20 cá thể
- Nhập liệu (phần mềm Excel)

4. Phân tích:

- Biến nghiên cứu định tính: lập bảng tần số, vẽ biểu đồ, xác định mode
- Biến nghiên cứu định lượng: Tính trung bình, trung vị, mode, độ lệch chuẩn, ước lượng trung bình 95%.

5. Bài tập thực hành nghiên cứu thống kê

- CÔNG VIỆC THỰC HIỆN:**
- 5. Kết luận: Dựa vào kết quả phân tích được, nhận xét, đánh giá, dự đoán các vấn đề nghiên cứu của đề tài.
 - 6. Trình bày tích hợp tất cả công việc thành tập báo cáo và nộp giáo viên hạn chót buổi học thứ 13.

5. Bài tập thực hành nghiên cứu thống kê

- MỘT SỐ ĐỀ TÀI GỢI Ý:**
- 6. Nghiên cứu mối quan hệ giữa thời gian làm thêm và kết quả học tập của SV.
 - 7. Nghiên cứu mối quan hệ giữa tự học và kết quả học tập của SV.
 - 8. Nghiên cứu về cách sử dụng thời gian rảnh của các bạn SV.
 - 9. Nghiên cứu về mong muốn, hoài bão của sinh viên về việc làm và thu nhập sau khi ra trường và sau này.

5. Bài tập thực hành nghiên cứu thống kê

- MỘT SỐ ĐỀ TÀI GỢI Ý:**
- 1. Nghiên cứu đánh giá khả năng rèn luyện của sinh viên.
 - 2. Nghiên cứu việc sử dụng điện thoại của sinh viên.
 - 3. Nghiên cứu về chi tiêu của sinh viên.
 - 4. Nghiên cứu về việc làm bán thời gian của sinh viên.
 - 5. Nghiên cứu về việc sử dụng phương tiện giao thông của sinh viên.