



NỘI DUNG

- 1 GIỚI THIỆU CHUNG VỀ THỐNG KÊ HỌC
- 2 ĐIỀU TRA THỐNG KÊ
- 3 TỔNG HỢP THỐNG KÊ
- 4 NGHIÊN CỨU THỐNG KÊ CÁC MỨC ĐỘ CỦA HIỆN TƯỢNG KTXH
- 5 ĐIỀU TRA CHỌN MẪU
- 6 KIỂM ĐỊNH GIẢ THUYẾT THỐNG KÊ

Data analysis

THE GATHERING, DISPLAY, AND SUMMARY OF DATA;

Probability

THE LAWS OF CHANCE, IN AND OUT OF THE CASINO;

Statistical inference

THE SCIENCE OF DRAWING STATISTICAL CONCLUSIONS FROM SPECIFIC DATA, USING A KNOWLEDGE OF PROBABILITY.



CHƯƠNG I: GIỚI THIỆU CHUNG VỀ THỐNG KÊ HỌC



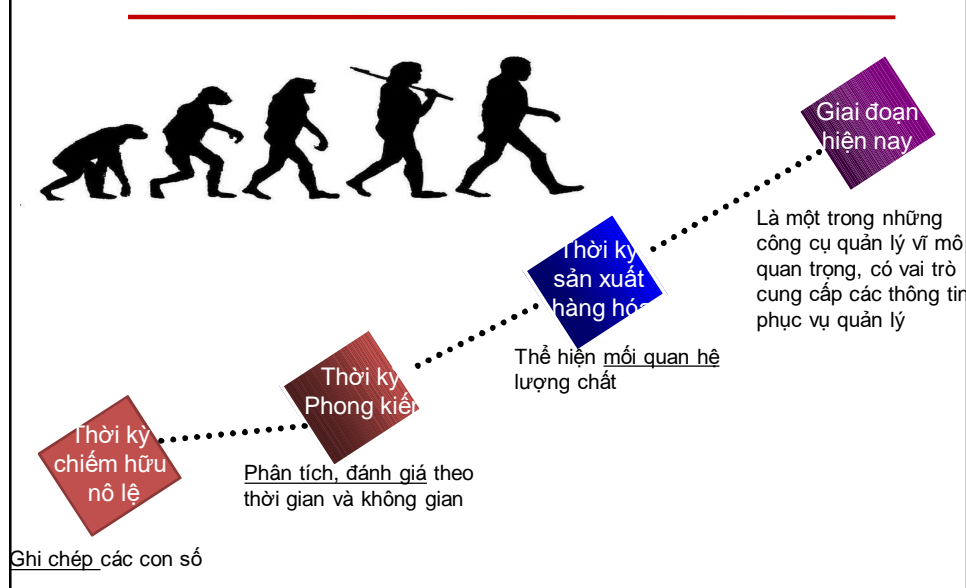
I. Đối tượng nghiên cứu của thống kê học

- Thống kê học là gì?
- Sơ lược lịch sử phát triển thống kê học
- Đối tượng nghiên cứu của thống kê học và các phương pháp trong thống kê

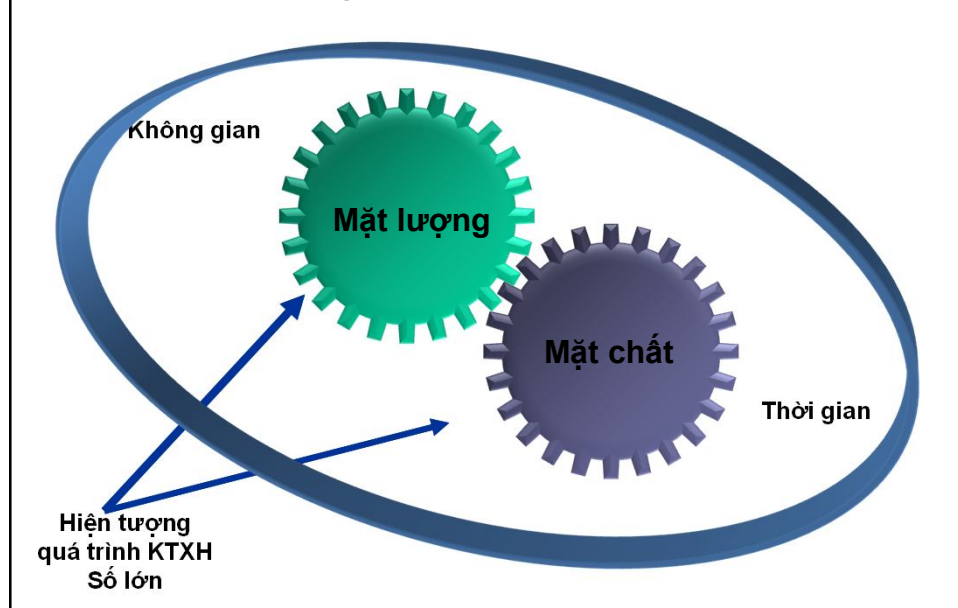
1. Thống kê học:

Thống kê học là khoa học nghiên cứu hệ thống phương pháp (thu thập, xử lý, phân tích) con số (mặt lượng) của các hiện tượng số lớn tìm bản chất và tính quy luật (mặt chất) trong những điều kiện nhất định.

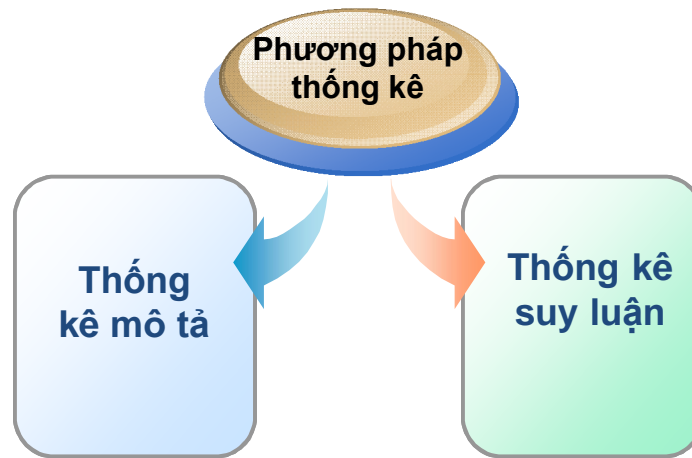
2. Sơ lược về sự ra đời và phát triển của thống kê học



3. Đối tượng nghiên cứu của thống kê học



Các phương pháp thống kê



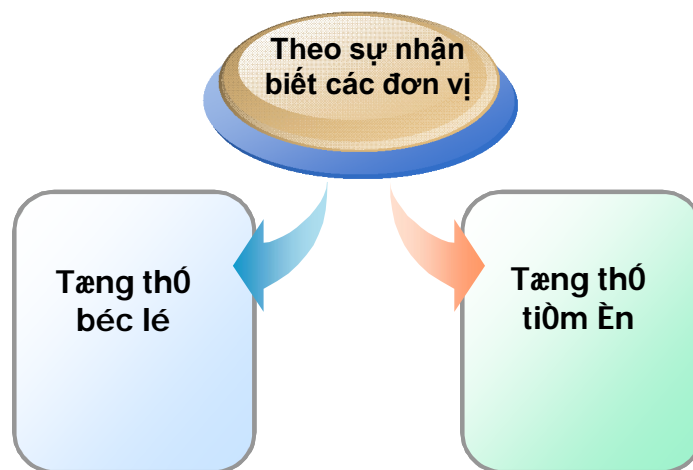
II. Một số khái niệm thường dùng trong thống kê

- Tổng thể thống kê
- Tiêu thức thống kê
- Chỉ tiêu thống kê

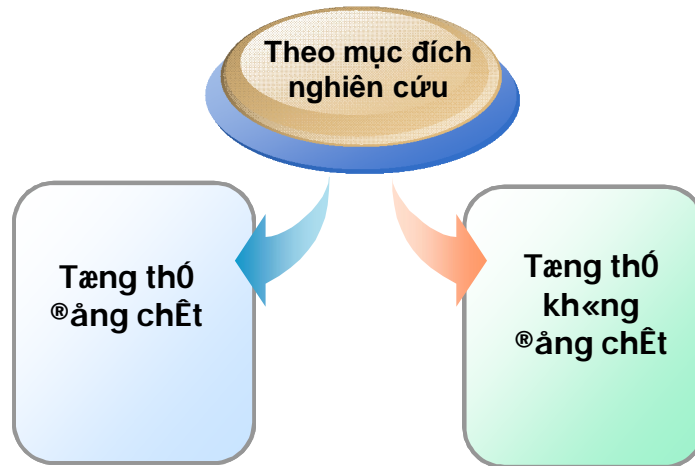
1. Tổng thể thống kê và đơn vị tổng thể

Tổng thể thống kê là hiện tượng số lớn gồm các đơn vị (phần tử) cần quan sát và phân tích mặt lượng.
Các đơn vị (phần tử) - đơn vị tổng thể.

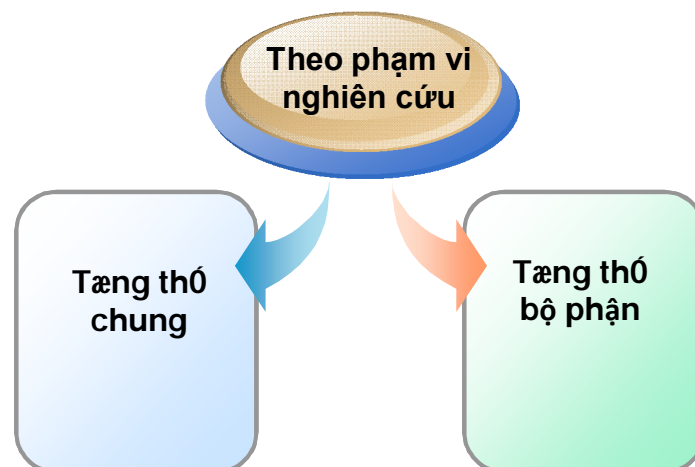
Các loại tổng thể thống kê



Các loại tổng thể thống kê



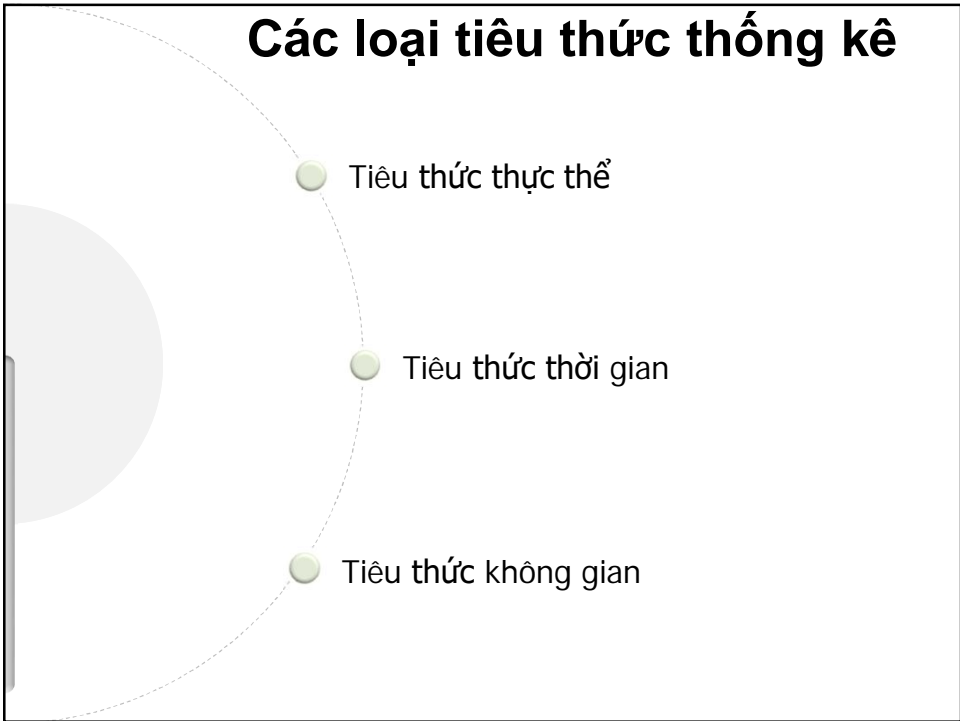
Các loại tổng thể thống kê



2. Tiêu thức thống kê

Tiêu thức thống kê - đặc điểm của đơn vị tổng thể được chọn để nghiên cứu

Các loại tiêu thức thống kê

- 
- Tiêu thức thực thể
 - Tiêu thức thời gian
 - Tiêu thức không gian

Tiêu thức thuộc tính

- Biểu hiện không trực tiếp qua con số
 - Biểu hiện qua đặc điểm, tính chất
- > Tiêu thức có biểu hiện gián tiếp được gọi là chỉ báo thống kê

Tiêu thức số lượng

- Biểu hiện trực tiếp qua con số
- Con số - lượng biến

Tiêu thức thay phiên

Là tiêu thức chỉ có 2 biểu hiện không trùng nhau trên một đơn vị tổng thể

Tiêu thức thời gian

Phản ánh thời gian của hiện tượng nghiên cứu

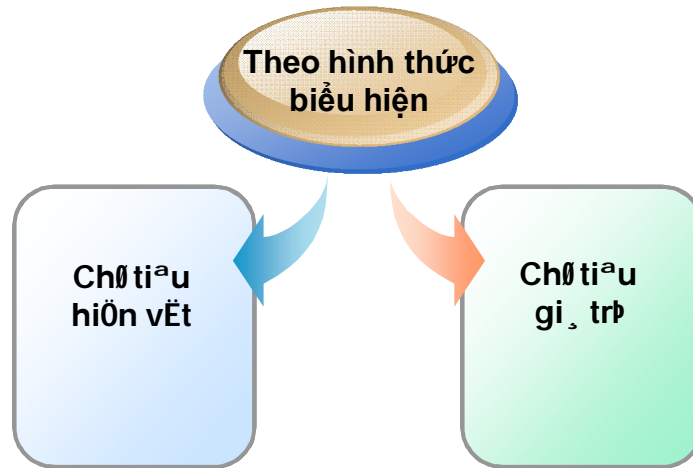
Tiêu thức không gian

Phản ánh phạm vi (lãnh thổ) của hiện tượng

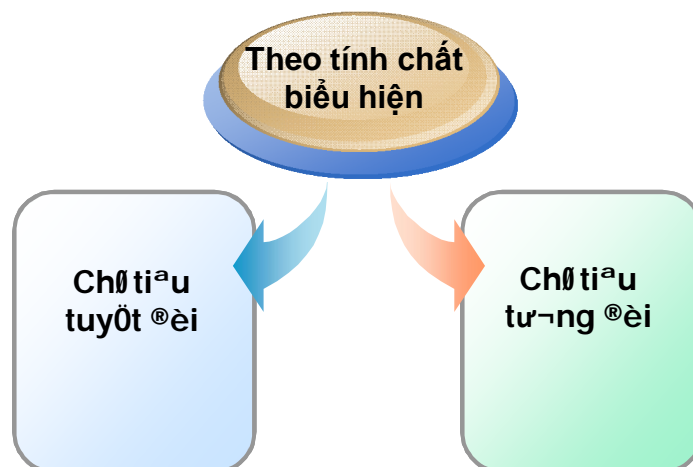
3. Chỉ tiêu thống kê

Chỉ tiêu thống kê phản ánh mặt lượng gắn với chất của các hiện tượng và quá trình KTXH số lớn trong điều kiện thời gian và địa điểm cụ thể.

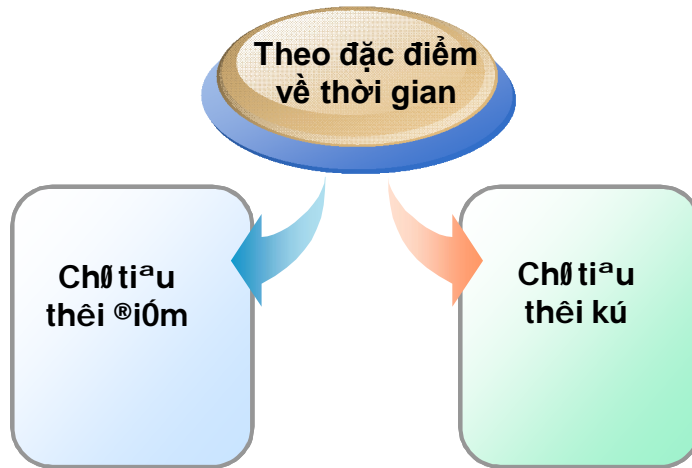
Các loại chỉ tiêu thống kê



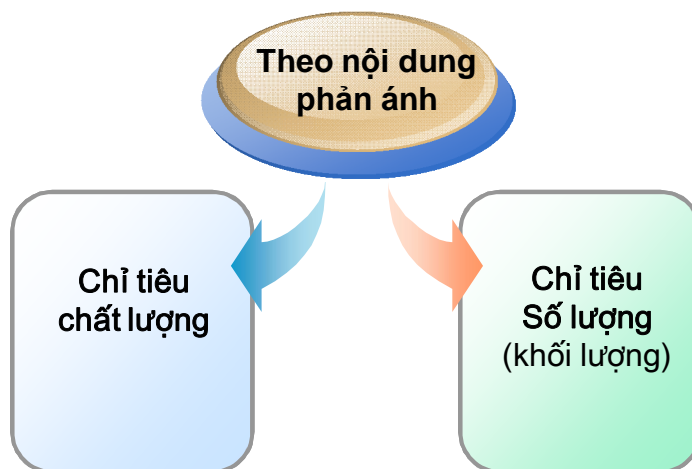
Các loại chỉ tiêu thống kê



Các loại chỉ tiêu thống kê



Các loại chỉ tiêu thống kê



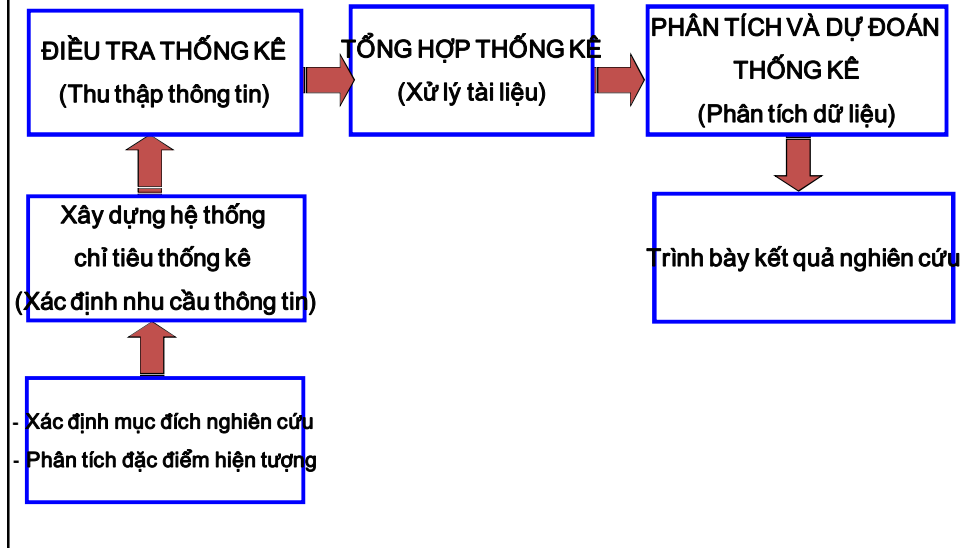
III. THANG ĐO TRONG THỐNG KÊ

- Thang đo định danh
- Thang đo thứ bậc
- Thang đo khoảng
- Thang đo tỷ lệ

MÔ HÌNH MÔ TẢ CÁC THANG ĐO

Tiêu thức Số lượng	THANG ĐO TỶ LỆ (Ratio Scale)		
	THANG ĐO KHOẢNG (Interval Scale)		Có gốc 0
Tiêu thức thuộc tính	THANG ĐO THỨ BẬC (Ordinal Scale)		Có khoảng cách bằng nhau
	THANG ĐO ĐỊNH DANH (Nominal Scale)	Biểu hiệu có thứ tự hơn kém	
	Đánh số các biểu hiện cùng loại của tiêu thức		

IV. Quá trình nghiên cứu thống kê



Xác định mục đích nghiên cứu và phân tích đặc điểm hiện tượng

- **Mục đích nghiên cứu**
 - Các số liệu thống kê phản ánh vấn đề gì?
 - Các thông tin thu thập phục vụ cho đối tượng nào?
- **Phân tích đặc điểm của hiện tượng**
 - Đối tượng nghiên cứu có đặc điểm đặc thù gì
 - Xem đối tượng nằm trong hoàn cảnh không gian và thời gian nào?

Xây dựng hệ thống chỉ tiêu thống kê

- Khái niệm:

Hệ thống chỉ tiêu thống kê là một tập hợp các chỉ tiêu có thể phản ánh các mặt, các tính chất quan trọng nhất, các mối liên hệ cơ bản giữa các mặt trong tổng thể và giữa các hiện tượng nghiên cứu với các hiện tượng có liên quan

Căn cứ xây dựng

- Căn cứ vào mục đích nghiên cứu.
- Căn cứ vào tính chất và đặc điểm của đối tượng nghiên cứu.
- Căn cứ vào khả năng nhân lực và tài chính.

Yêu cầu

- Phải có khả năng nêu được đặc điểm và mối liên hệ cơ bản của hiện tượng nghiên cứu
- Phải có các chỉ tiêu mang tính chất *chung*, các chỉ tiêu mang tính chất *bộ phận* và các chỉ tiêu *nhân tố*.
- Phải đảm bảo tính *thống nhất về nội dung, phương pháp và phạm vi* tính toán của các *chỉ tiêu cùng loại*.

ĐIỀU TRA THỐNG KÊ (Trình bày cụ thể chương II)

TỔNG HỢP THỐNG KÊ (Trình bày cụ thể chương III)

Phân tích và dự đoán thống kê

★ Khái niệm:

Phân tích và dự đoán thống kê là nêu lên một cách tổng hợp bản chất cụ thể và tính quy luật của các hiện tượng số lớn trong điều kiện nhất định qua biểu hiện bằng số lượng và tính toán các mức độ của hiện tượng trong tương lai.

Phân tích và dự đoán thống kê

★ Ý nghĩa:

- Là giai đoạn cuối cùng của quá trình nghiên cứu thống kê và nó biểu hiện tập trung nhất kết quả toàn bộ quá trình đó.
- Phân tích và dự đoán thống kê không chỉ có ý nghĩa nhận thức mà còn góp phần cải tạo hiện tượng.

Phân tích và dự đoán thống kê

★ Yêu cầu:

- Phân tích và dự đoán thống kê phải tiến hành trên cơ sở phân tích lý luận KTXH
- Phân tích và dự đoán thống kê phải căn cứ vào toàn bộ sự kiện và đặt chúng trong mối liên hệ với nhau.
- Đối với những hiện tượng có tính chất khác nhau phải áp dụng các phương pháp khác nhau.

CHƯƠNG II: ĐIỀU TRA THỐNG KÊ

- 1 Khái niệm chung về điều tra thống kê
- 2 Phân loại
- 3 Các hình thức thu thập thông tin
- 4 Phương án điều tra thống kê
- 5 Sai số trong điều tra thống kê

I. Khái niệm chung về điều tra thống kê

★ Khái niệm:

Điều tra thống kê là tổ chức một cách khoa học và theo một kế hoạch thống nhất việc thu thập tài liệu về các hiện tượng và quá trình kinh tế xã hội.

I. Khái niệm chung về điều tra thống kê

★ Ý nghĩa:

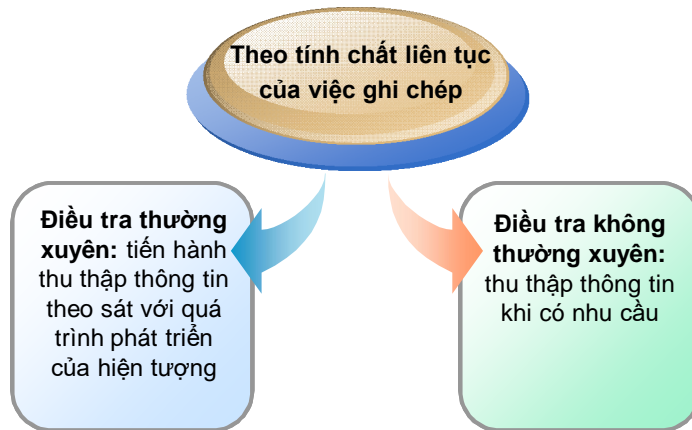
- Là giai đoạn đầu của quá trình nghiên cứu thống kê
- Cung cấp thông tin cho cả quá trình nghiên cứu
- + Là căn cứ để kiểm tra, đánh giá thực trạng
- + Cung cấp luận cứ cho việc phân tích
- + Là căn cứ cho việc phát hiện quy luật

I. Khái niệm chung về điều tra thống kê

★ Yêu cầu:

- Chính xác
- Kịp thời
- Đầy đủ (nội dung, phạm vi)

II. Các loại điều tra thống kê



II. Các loại điều tra thống kê



Điều tra không toàn bộ



III. Các hình thức thu thập thông tin

★ **Báo cáo thống kê định kỳ:** Là hình thức điều tra thường xuyên có định kỳ theo nội dung, phương pháp và chế độ báo cáo chính thức do cơ quan có thẩm quyền quy định

★ **Các loại**

- Báo cáo thống kê **cơ sở**
- Báo cáo **thống kê tổng hợp**

★ **Đặc điểm:** Tổ chức thu thập thông tin có tính chất hành chính

III. Các hình thức thu thập thông tin

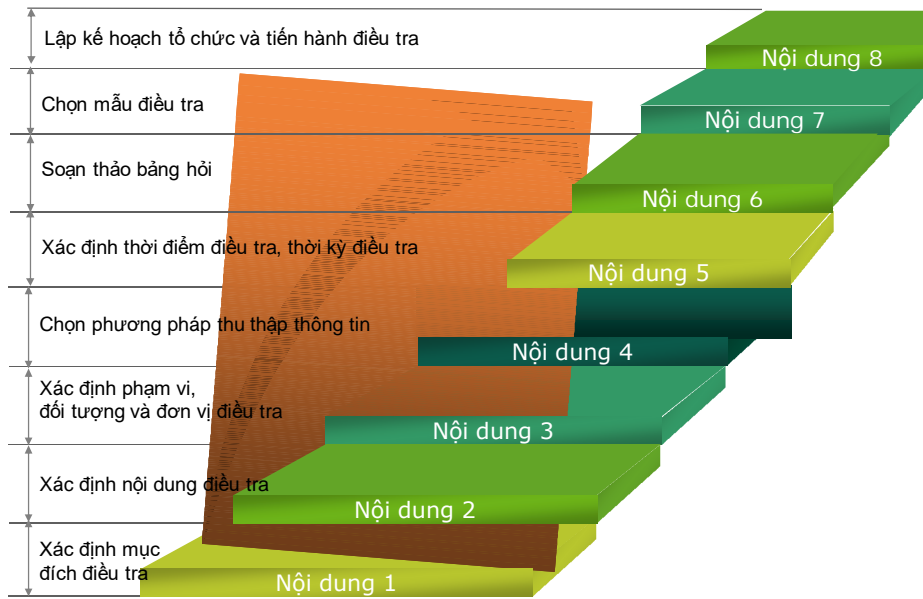
★ **Điều tra chuyên môn:** là hình thức tổ chức điều tra không thường xuyên được tiến hành theo một kế hoạch và phương pháp quy định riêng cho mỗi lần điều tra

★ **Đặc điểm:** Không mang tính hành chính

IV. Phương án điều tra thống kê

Phương án điều tra là văn bản được xây dựng trước khi tiến hành điều tra, quy định rõ về những vấn đề cần giải quyết và hiểu thống nhất trước, trong và sau khi tiến hành điều tra.

NỘI DUNG PHƯƠNG ÁN ĐIỀU TRA THỐNG KÊ



V. Sai số trong điều tra thống kê

★ Là chênh lệch giữa trị số thu được qua điều tra so với trị số thực tế của hiện tượng

★ **Phân loại:**

- Sai số do đăng ký ghi chép
- Sai số do tính chất đại biểu

SAI SỐ DO ĐĂNG KÝ GHI CHÉP

Xảy ra trong tất cả các cuộc điều tra thống kê

- Vô ý khai báo, đăng ký, ghi chép sai
- Cố tình khai báo, đăng ký, ghi chép sai
- Đo lường
- Hiểu sai nội dung câu hỏi
- Ý thức xã hội,...

SAI SỐ DO TÍNH CHẤT ĐẠI BIỂU

Chỉ xảy ra trong điều tra chọn mẫu

- Số lượng đơn vị mẫu không đủ lớn
- Vi phạm nguyên tắc chọn mẫu ngẫu nhiên
- Kết cấu tổng thể mẫu khác tổng thể chung,...

CÁCH KHẮC PHỤC SAI SỐ

- Đối với sai số do đăng ký, ghi chép

+ Làm tốt công tác chuẩn bị điều tra (soạn thảo bảng hỏi,...)

+ Làm tốt công tác kiểm tra, giám sát,...

- Đối với sai số do tính chất đại biểu

+ Lựa chọn phương pháp tổ chức chọn mẫu phù hợp

+ Tăng số đơn vị điều tra

+ Đảm bảo nguyên tắc ngẫu nhiên,...

CHƯƠNG III: TỔNG HỢP THỐNG KÊ



I. Những vấn đề chung về tổng hợp thống kê

☆ Khái niệm:

Tổng hợp thống kê là tiến hành tập trung chỉnh lý và hệ thống hoá một cách khoa học các tài liệu thu thập được trong điều tra thống kê.

I. Những vấn đề chung về tổng hợp thống kê

☆ Ý nghĩa:

- Là giai đoạn trung gian của quá trình nghiên cứu thống kê
- Giúp nhận xét, phân tích đặc trưng cơ bản hiện tượng nghiên cứu
- Là cơ sở để vận dụng các phương pháp phân tích thống kê

I. Những vấn đề chung về tổng hợp thống kê

☆ Nhiệm vụ:

Bước đầu làm cho các đặc trưng riêng của từng đơn vị tổng thể chuyển thành đặc trưng chung của toàn bộ tổng thể.

II. Phương pháp tổng hợp thống kê

-  Sắp xếp dữ liệu và phân tổ thống kê
-  Bảng thống kê
-  Đồ thị thống kê

1. Sắp xếp và phân tổ thống kê

a Sắp xếp dữ liệu

b Phân tổ thống kê

a. Sắp xếp dữ liệu

- Tiêu thức thuộc tính:

- Tiêu thức số lượng:

a. Sắp xếp dữ liệu bằng sơ đồ “thân lá”

Sơ đồ thân lá (stem-and-leaf) nhằm cung cấp một hình ảnh **nhanh** về hình dáng phân bố bao gồm các giá trị dạng số thực trong sơ đồ



a. Sắp xếp dữ liệu bằng sơ đồ “thân lá”

Cách thực hiện:

- Sắp xếp dữ liệu theo thứ tự tăng dần
- Thường chia mỗi quan sát vào một **thân** gồm tất cả các con số ngoại trừ con số cuối cùng và một **lá** – con số cuối cùng.
- Viết các **thân** vào một cột với trị số tăng dần
- Viết từng **lá** vào hàng bên phải **thân** theo trật tự tăng dần

a. Sắp xếp dữ liệu bằng sơ đồ “thân lá”

Kỹ thuật “cắt tỉa” sơ đồ “thân lá”

- Tách mỗi số ở **thân** (khi **thân** nhỏ hơn 5) thành 2 hoặc nhiều số (một với các **lá** từ 0 đến 4 và một từ 5 đến 9 hoặc **nhỏ hơn**).
- Mỗi **lá** có thể đại diện cho nhiều quan sát
- Khi trị số quan sát có nhiều con số, nên “cắt tỉa” các con số bằng cách bỏ bớt một vài con số cuối cùng.

b. Phân tổ thống kê

 Khái niệm, ý nghĩa và nhiệm vụ của phân tổ thống kê

 Các loại phân tổ thống kê

 Các bước tiến hành phân tổ thống kê

***R* Khái niệm phân tổ thống kê**

Phân tổ thống kê là căn cứ vào một (hay một số) tiêu thức nào đó để tiến hành phân chia các đơn vị của hiện tượng nghiên cứu thành các tổ (và các tiểu tổ) có tính chất khác nhau

***R* Ý nghĩa phân tổ thống kê**

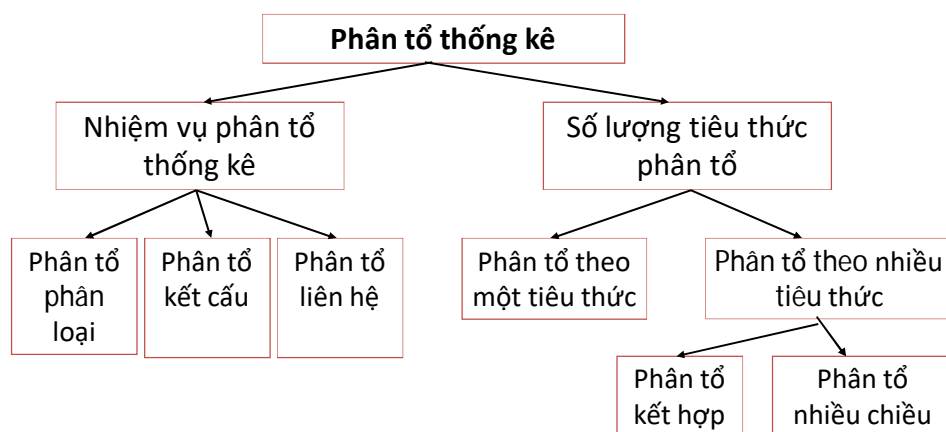
Có ý nghĩa trong cả quá trình nghiên cứu thống kê

- Giai đoạn điều tra thống kê: Cơ sở cho việc lựa chọn các đơn vị điều tra thực tế
- Giai đoạn tổng hợp thống kê: Phương pháp cơ bản của tổng hợp thống kê
- Giai đoạn phân tích thống kê: Cơ sở để vận dụng các phương pháp phân tích thống kê

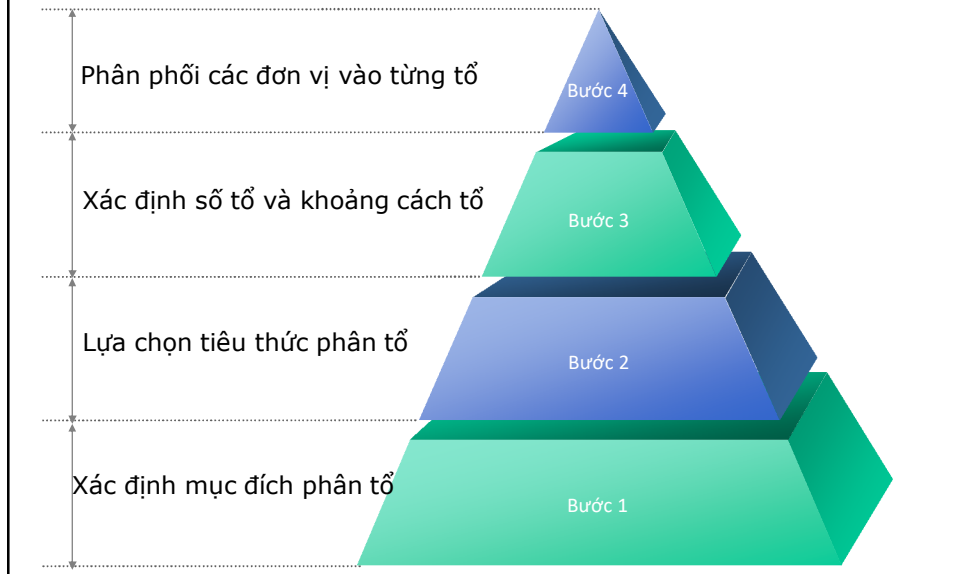
Nhiệm vụ phân tổ thống kê

- Phân chia các loại hình KTXH.
- Biểu hiện kết cấu của hiện tượng nghiên cứu.
- Nghiên cứu mối liên hệ giữa các tiêu thức.

Các loại phân tổ thống kê



* Các bước phân tổ thống kê (phân tổ đơn)



B1: Xác định mục đích phân tổ

Trả lời câu hỏi: phân tổ để làm gì?



B1: Lựa chọn tiêu thức phân tổ

⊕ **Tiêu thức phân tổ** là tiêu thức được chọn làm căn cứ để tiến hành phân tổ thống kê.

***B* Lựa chọn tiêu thức phân tổ**

Căn cứ chọn tiêu thức phân tổ:

- ✓ Dựa vào mục đích nghiên cứu.
- ✓ Căn cứ vào đặc điểm, tính chất và điều kiện lịch sử cụ thể.

B3: Xác định số tổ và khoảng cách tổ

Phân tổ theo tiêu thức thuộc tính

- ✓ Các loại hình tương đối ít: mỗi loại hình có thể hình thành nên 1 tổ
- ✓ Số loại hình thực tế **hiều**: ghép những loại hình giống nhau hoặc gần giống nhau vào một tổ

R Xác định số tổ và khoảng cách tổ

Phân tổ theo tiêu thức số lượng

- ✓ Tiêu thức có ít **lượng biến**: thường cứ mỗi lượng biến là cơ sở để hình thành một tổ, gọi là phân tổ không có khoảng cách tổ

***R* Xác định số tổ và khoảng cách tổ**

Phân tổ theo tiêu thức số lượng

✓ **Tiêu thức có nhiều lượng biến:** mỗi tổ sẽ bao gồm một phạm vi lượng biến, với hai giới hạn (giới hạn trên và giới hạn dưới), gọi là phân tổ có khoảng cách tổ.

Khoảng cách tổ (h) = giới hạn trên – giới hạn dưới

Phân tổ có khoảng cách tổ

+ **Khoảng cách tổ bằng nhau**

$$h = \frac{x_{\max} - x_{\min}}{n}$$

Phân tổ có khoảng cách tổ

+ **Khoảng cách tổ không bằng nhau:** Tùy đặc điểm của hiện tượng và mục đích nghiên cứu để xác định số tổ và khoảng cách tổ phù hợp

B4: Phân phối các đơn vị vào từng tổ

Sắp xếp các đơn vị vào từng tổ tương ứng với biểu hiện của từng tổ

Dãy số phân phối

Dãy số phân phối là kết quả của phân tổ thống kê

Các loại

- ✓ Dãy số phân phối thuộc tính
- ✓ Dãy số phân phối số lượng (dãy số lượng biến)

Thành phần của dãy số lượng biến

Lượng biến (x_i)	Tần số (f_i)	Tần suất ($d_i = f_i / \sum f_i$)	Tần số tích lũy ($S_i = \sum f_i$)
x_1	f_1	d_1	f_1
x_2	f_2	d_2	$f_1 + f_2$
...
x_n	f_n	f_n	$f_1 + f_2 + \dots + f_{n-1} + f_n$
Cộng	$\sum f_i$	$\sum d_i = 1$ (100)	

Khi phân tổ có khoảng cách tổ không bằng nhau, để đánh giá mức độ tập trung từng tổ, sử dụng mật độ phân phối ($m_i = f_i/h_i$)

*** Phân tổ lại**

Phân tổ lại là tiến hành lập ra một số tổ mới trên cơ sở các tổ cũ đã có sẵn từ trước, nhằm đáp ứng mục đích nghiên cứu nào đó.

*** Phân tổ lại**

Các trường hợp sử dụng phân tổ lại

- Các tài liệu trước phân tổ không thống nhất
- Các tài liệu trước được phân thành nhiều tổ quá nhỏ
- Các tài liệu phân tổ cũ chưa hợp lý

*** Phân tổ lại**

Phương pháp phân tổ lại

- Lập các tổ mới bằng cách thay đổi (mở rộng) khoảng cách tổ của phân tổ cũ
- Lập các tổ mới theo tỷ trọng của mỗi tổ chiếm trong tổng thể

*** Các bước tiến hành Phân tổ nhiều chiều**

- *Xác định mục đích phân tổ*
- *Lựa chọn tiêu thức phân tổ*
- *Xác định số tổ và khoảng cách tổ*
- *Phân phối các đơn vị vào từng tổ*

Tiêu thức phân tổ

Chuyển các tiêu thức phân tổ về dạng một **tiêu thức tổng hợp**

Tiêu thức phân tổ

- Gọi x_{ij} ($i=1,n; j=1,k$) là lượng biến của đơn vị thứ i của tiêu thức thứ j

- Đưa các lượng biến của các tiêu thức về dạng tỷ lệ

$$P_{ij} = \frac{x_{ij}}{\bar{x}_j} \quad \text{trong đó} \quad \bar{x}_j = \frac{\sum_{i=1}^n x_{ij}}{n}$$

- Tính **tiêu thức tổng hợp** $\bar{P}_i = \frac{\sum_{j=1}^k P_{ij}}{k}$

Xác định số tổ

Tương tự như phân tổ đơn

Bảng thống kê

Bảng thống kê là một hình thức trình bày các tài liệu thống kê một cách có hệ thống, hợp lý và rõ ràng, nhằm nêu lên các đặc trưng về mặt lượng của hiện tượng nghiên cứu

Tác dụng của bảng thống kê

- Dễ dàng, đối chiếu, so sánh số liệu, có sức thuyết phục
- Giảm thiểu số liệu các giá trị của dữ liệu trong văn bản
- Thu hút sự chú ý của độc giả

Cấu trúc bảng thống kê



Ghi chú (nếu có)
Nguồn thông tin:

Cấu trúc bảng thống kê

Tiêu đề của bảng thống kê phản ánh nội dung, ý nghĩa của bảng

- Ngắn gọn, dễ hiểu, thể hiện rõ ý nghĩa, nội dung
- Phản ánh điều kiện thời gian, không gian

Cấu trúc bảng thống kê

Tiêu đề cột, ở trên cùng của bảng, xác định các dữ liệu được trình bày trong mỗi cột của bảng, đơn vị tính (nếu cần)

Tiêu đề dòng, trong cột đầu tiên của bảng, xác định các dữ liệu được trình bày trong mỗi hàng của bảng

Có thể sử dụng các phân tử trong tiêu đề cột, dòng

Cấu trúc bảng thống kê

Dữ liệu, các số liệu trong bảng là kết quả của quá trình tổng hợp thống kê

Cấu trúc bảng thống kê

- **Chú thích**, ở dưới cùng của bảng, cung cấp thông tin bổ sung cần thiết, các công thức (nếu cần)
- **Nguồn thông tin**, ở dưới cùng của bảng, cung cấp nguồn dữ liệu (cơ quan, đơn vị tạo ra các dữ liệu và phương pháp thu thập dữ liệu)

Các loại bảng thống kê

Bảng giản đơn: là loại bảng thống kê, trong đó hiện tượng chỉ phân tổ theo một tiêu thức nào đó

Bảng kết hợp: là loại bảng thống kê trong đó đối tượng nghiên cứu được phân chia theo từ hai tiêu thức trở lên

Nguyên tắc khi trình bày bảng thống kê

- Quy mô bảng vừa phải
- Tiêu đề bảng, tiêu mục ghi chính xác, ngắn gọn
- Đơn vị tính – nếu tất cả có cùng đơn vị tính thì ghi góc phải phía trên bảng
- Các cột nên cách nhau đều, độ rộng vừa với nội dung
- Các chỉ tiêu được sắp xếp theo thứ tự hợp lý
- Không được để trống ô nào trong bảng, nếu không có dữ liệu thì ghi bằng các ký hiệu

Nguyên tắc ghi ký hiệu

- Nếu hiện tượng không có số liệu, ghi (-)
- Nếu số liệu còn thiếu, có thể bổ sung (...)
- Nếu hiện tượng không liên quan (x)

3. Đồ thị thống kê

Là các hình vẽ hoặc đường nét hình học dùng để miêu tả có tính chất quy ước các tài liệu thống kê

Tác dụng của đồ thị

- Hình tượng hoá các số liệu nhằm so sánh, nghiên cứu kết cấu, xu hướng, mối liên hệ,....
- Giúp đơn giản hoá các mối quan hệ phức tạp
- Có được những phác thảo cơ bản về hiện tượng
- Người đọc ghi nhận thông tin một cách nhanh chóng
- Sinh động, có sức hấp dẫn

Các loại đồ thị

- *Đồ thị phát triển*
- *Đồ thị kết cấu*
- *Đồ thị so sánh*
- *Đồ thị liên hệ*
- *Đồ thị "tháp dân số"*

Tháp dân số

Tháp dân số là sự kết hợp của hai biểu đồ thanh ngang, thường đại diện cho cấu trúc tuổi của dân số nữ và nam giới của một quốc gia hoặc khu vực

Các thành phần của đồ thị thống kê

Các thành phần của dữ liệu dùng để trình bày dữ liệu:
các thanh, đường thẳng, các khu vực hoặc các điểm.

Các thành phần hỗ trợ trong việc tìm hiểu dữ liệu: tiêu đề, ghi chú, nhãn dữ liệu, các đường lưới, chú thích và nguồn dữ liệu.

Các thành phần dùng để trang trí không liên quan đến dữ liệu.

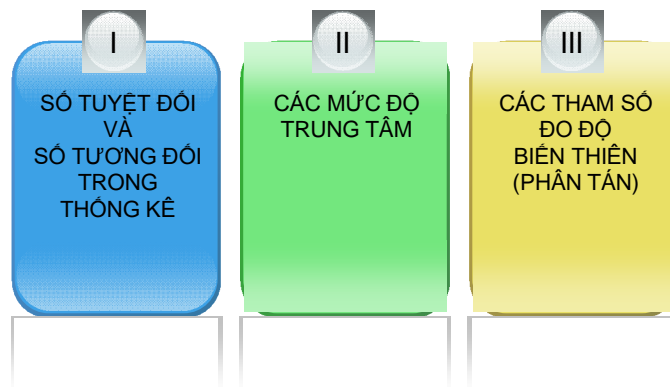
Nguyên tắc trình bày đồ thị

- Quy mô của đồ thị hợp lý (chiều dài, chiều cao).
- Lựa chọn dạng đồ thị phù hợp
- Khoảng cách giữa các cột hợp lý
- Thang đo, tỷ lệ xích phù hợp (tỷ lệ 1: 1,33 hoặc 1:1,5)
- Không nên có quá nhiều hiện tượng trong một đồ thị

Bản đồ thống kê

- Bản đồ thống kê là những công cụ hiệu quả nhất để hình dung mô hình không gian.
- Thường được dùng để biểu thị các cường độ phân bố khác nhau theo vùng địa lý của một chỉ tiêu nào đó (như mật độ dân số của các vùng),...

CHƯƠNG IV: NGHIÊN CỨU THỐNG KÊ CÁC MỨC ĐỘ CỦA HIỆN TƯỢNG KINH TẾ XÃ HỘI



I. Số tuyệt đối và số tương đối trong thống kê

- 1 Số tuyệt đối trong thống kê
- 2 Số tương đối trong thống kê
- 3 Điều kiện vận dụng số tuyệt đối và số tương đối trong thống kê

1, Số tuyệt đối trong thống kê



Khái niệm



Đặc điểm



Đơn vị tính



Tác dụng



Các loại

Khái niệm số tuyệt đối

Số tuyệt đối trong thống kê biểu hiện quy mô, số lượng của hiện tượng nghiên cứu tại thời gian, địa điểm,

Đặc điểm của số tuyệt đối

- ⊕ Bao hàm một nội dung kinh tế xã hội cụ thể trong điều kiện thời gian và địa điểm nhất định,
- ⊕ Phải qua điều tra thực tế và tổng hợp mới xác định được số tuyệt đối trong thống kê,

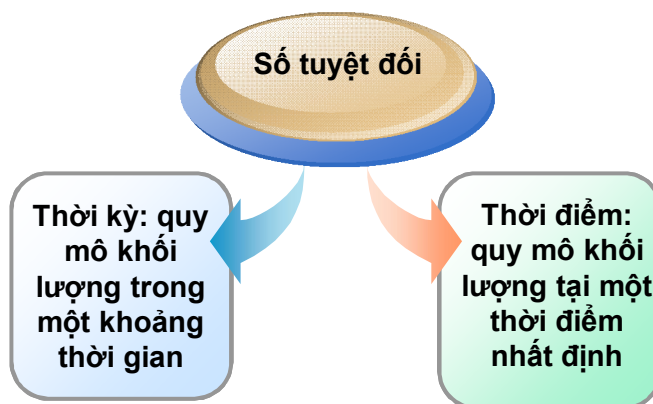
Đơn vị tính số tuyệt đối

- Đơn vị hiện vật: cái, con, quả, chiếc, m, kg,...
- Đơn vị giá trị: VND, USD,...
- Đơn vị kép: tấn-km, kwh,,,

Tác dụng

- ⊕ Có một nhận thức cụ thể về quy mô, khối lượng thực tế của hiện tượng nghiên cứu,
- ⊕ Là cơ sở đầu tiên để tiến hành phân tích thống kê, đồng thời để tính các chỉ tiêu khác,

Các loại số tuyệt đối



2, Số tương đối trong thống kê



Khái niệm



Đặc điểm



Đơn vị tính



Tác dụng



Các loại

Khái niệm số tương đối

Số tương đối trong thống kê biểu hiện quan hệ so sánh giữa hai mức độ nào đó của hiện tượng,

Đặc điểm

Là kết quả của việc so sánh 2 số tuyệt đối

Đơn vị tính

- ✓ **Lần, phần trăm (%) phần nghìn (‰)**
- ✓ **Đơn vị kép:** người/km², triệu đồng/người, sản phẩm/người,...

Tác dụng của số tương đối

- ☞ Phân tích hiện tượng: nêu lên kết cấu, trình độ phổ biến, tốc độ phát triển, đánh giá trình độ hoàn thành kế hoạch;
- ☞ Phản ánh tình hình thực tế trong khi cần bảo đảm được tính chất bí mật của số tuyệt đối,

Các loại số tương đối

- Số tương đối động thái (tốc độ phát triển) $t = \frac{y_1}{y_0} (100)$
- Số tương đối kế hoạch (lập và kiểm tra kế hoạch)
 - Số tương đối nhiệm vụ kế hoạch $K_n = \frac{y_{KH}}{y_0} (100)$
 - Số tương thực hiện kế hoạch $K_T = \frac{y_1}{y_{KH}} (100)$
- Mối quan hệ: $t = K_n \times K_T$ hay $\frac{y_1}{y_0} = \frac{y_{KH}}{y_0} \times \frac{y_1}{y_{KH}}$

Các loại số tương đối

- **Số tương đối kết cấu:** Phản ánh tỷ trọng của từng bộ phận cấu thành trong một tổng thể.

$$d_i = \frac{y_i}{\Sigma y_i} (100)$$

Các loại số tương đối

- **Số tương đối không gian:** so sánh giữa hai hiện tượng cùng loại nhưng khác nhau về không gian hoặc là quan hệ so sánh mức độ giữa hai bộ phận trong một tổng thể

Các loại số tương đối

- **Số tương đối cường độ:** so sánh chỉ tiêu của hai hiện tượng khác nhau nhưng có quan hệ với nhau.






3. Vận dụng chung số tương đối và tuyệt đối trong thống kê

- Phải xét đến đặc điểm của hiện tượng nghiên cứu để rút ra kết luận cho đúng
- Phải vận dụng một cách kết hợp các số tương đối với số tuyệt đối

II. Các mức độ trung tâm

- 1 Số bình quân (trung bình)
- 2 Mốt (Mo)
- 3 Trung vị (Me)

1. Số bình quân (trung bình)

-  Khái niệm chung
-  Các loại số bình quân
-  Đặc điểm của số bình quân
-  Hạn chế của số bình quân
-  Điều kiện vận dụng số bình quân trong thống kê

* Khái niệm

❖ **Số bình quân** trong thống kê là mức độ đại biểu theo một tiêu thức nào đó của một tổng thể bao gồm nhiều đơn vị.

* Tác dụng

- Phản ánh mức độ đại biểu, nêu lên đặc trưng chung nhất của tổng thể
- So sánh các hiện tượng không có cùng quy mô.

* Các loại số bình quân

a. Số bình quân cộng (áp dụng khi các lượng biến có quan hệ tổng)

$$\text{Số trung bình} = \frac{\text{Tổng lượng biến của tiêu thức}}{\text{Tổng số đơn vị}}$$

a. Số bình quân cộng

Số bình quân cộng giản đơn (khi dữ liệu chưa phân tổ)

$$\bar{x} = \frac{X_1 + X_2 + \dots + X_n}{n}$$

Lượng biến (x_i)	Tần số (f_i)
x_1	f_1
x_2	f_2
...	...
x_n	f_n
Cộng	Σf_i

a. Số bình quân cộng

Số bình quân cộng gia quyền

$$\bar{x} = \frac{x_1 f_1 + x_2 f_2 + \dots + x_n f_n}{f_1 + f_2 + \dots + f_n} = \frac{\sum x_i f_i}{\sum f_i}$$

$$\bar{x} = \sum x_i d_i \longrightarrow d_i = \frac{f_i}{\sum f_i}$$

a. Số bình quân cộng

❖ Số bình quân điều hoà gia quyền

$$\bar{x} = \frac{M_1 + M_2 + \dots + M_n}{\frac{M_1}{x_1} + \frac{M_2}{x_2} + \dots + \frac{M_n}{x_n}} = \frac{\sum M_i}{\sum \frac{M_i}{x_i}}$$

$$M_i = x_i f_i \quad \text{Tổng lượng biến tổ thứ } i$$

a. Số bình quân cộng

- ❖ Số bình quân điều hoà giản đơn (áp dụng khi các M_i bằng nhau)

$$\bar{x} = \frac{n}{\sum \frac{1}{x_i}}$$

b. Số bình quân nhân

- ❖ **Số bình quân nhân** (áp dụng khi các lượng biến có quan hệ tích)

– Số bình quân nhân giản đơn

$$\bar{x} = \sqrt[n]{x_1 \times x_2 \times \dots \times x_n} = \sqrt[n]{\prod x_i}$$

- ❖ Số bình quân nhân gia quyền

$$\bar{x} = \sqrt{x_1^{f_1} \times x_2^{f_2} \times \dots \times x_n^{f_n}} = \sqrt{\sum f_i \prod x_i^{f_i}}$$

*** Đặc điểm của số bình quân**

- Mang tính tổng hợp, khái quát cao.
- San bằng các chênh lệch giữa các đơn vị về trị số của tiêu thức nghiên cứu.

*** Hạn chế của số bình quân**

- Chịu ảnh hưởng của các lượng biến động xuất.

*** Điều kiện vận dụng số bình quân**

- Số bình quân chỉ nên tính ra từ tổng thể đồng chất.
- Số bình quân chung cần được vận dụng kết hợp với các số bình quân tổ hoặc dãy số phân phối.

2. Mốt (Mode)



Khái niệm



Cách xác định



Tác dụng

Khái niệm

Mốt là biểu hiện của tiêu thức phổ biến nhất (gặp nhiều nhất) trong một tổng thể hay trong một dãy số phân phối

Cách xác định

✧ Đối với trường hợp phân tổ không có khoảng cách tổ, mốt là lượng biến có tần số lớn nhất.

$$M_o = x_i \text{ (} f_i = \max \text{)}$$

Cách xác định

✧ Đối với trường hợp phân tử có khoảng cách tử

Bước 1: Xác định tử có mốt, là tử có tần số lớn nhất (khi k/c tử bằng nhau, hoặc là tử có mật độ phân phối lớn nhất khi k/c tử không bằng nhau)

Cách xác định

Bước 2: Xác định trị số gần đúng của mốt:

$$M_o = x_{M_o(\min)} + h_{M_o} \frac{\delta_1}{\delta_1 + \delta_2}$$

$$\left. \begin{array}{l} \delta_1 = f_{M_o} - f_{M_o-1} \\ \delta_2 = f_{M_o} - f_{M_o+1} \end{array} \right\} \text{Khoảng cách bằng nhau}$$

$$\left. \begin{array}{l} \delta_1 = m_{M_o} - m_{M_o-1} \\ \delta_2 = m_{M_o} - m_{M_o+1} \end{array} \right\} \text{Khoảng cách không bằng nhau}$$

Tác dụng

- Là mức độ đại biểu nên có thể thay thế hoặc bổ sung cho trung bình cộng trong trường hợp tính trung bình gặp khó khăn
- Có ý nghĩa hơn số bình quân cộng trong trường hợp dãy số có lượng biến đột xuất
- Là một trong những tham số nêu lên đặc trưng phân phối của dãy số
- Có tác dụng trong phục vụ nhu cầu hợp lý

Hạn chế của mốt

- Không xác định được mốt trong trường hợp dãy số phân phối không bình thường.

3. Trung vị (Median)



Khái niệm



Cách xác định



Tác dụng

Khái niệm

Trung vị là lượng biến của đơn vị đứng ở vị trí giữa trong một dãy số, chia dãy số thành hai phần bằng nhau

Cách xác định

✧ Trường hợp phân tử không có khoảng cách tổ

+ Nếu số đơn vị tổng thể lẻ ($\sum f_i = 2m + 1$): $Me = X_{m+1}$

+ Nếu số đơn vị tổng thể chẵn ($\sum f_i = 2m$): $Me = \frac{x_m + X_{m+1}}{2}$

Cách xác định

✧ Trường hợp phân tử có khoảng cách tổ:

- Bước 1: Xác định tổ chứa Me (tổ chứa đơn vị ở vị trí giữa trong dãy số)
- Bước 2: Xác định trị số gần đúng

$$Me = X_{Me(\min)} + h_{Me} \frac{\sum f_i - S_{(Me-1)}}{2 f_{Me}}$$

Tác dụng

- Là mức độ đại biểu nên có thể thay thế hoặc bổ sung cho trung bình cộng trong trường hợp tính trung bình gặp khó khăn
- Có ý nghĩa hơn số bình quân cộng trong trường hợp dãy số có lượng biến động xuất
- Là một trong những tham số nêu lên đặc trưng phân phối của dãy số
- Có tác dụng trong phục vụ công cộng $\sum |x_i - Me| f_i = \min$

Phân vị mức p

Phân vị mức p là giá trị mà có ít nhất p% số quan sát có giá trị nhỏ hơn hoặc bằng giá trị phân vị mức p

Bước 1: Sắp xếp dữ liệu theo thứ tự từ nhỏ đến nhất.

Bước 2: Tính chỉ số i

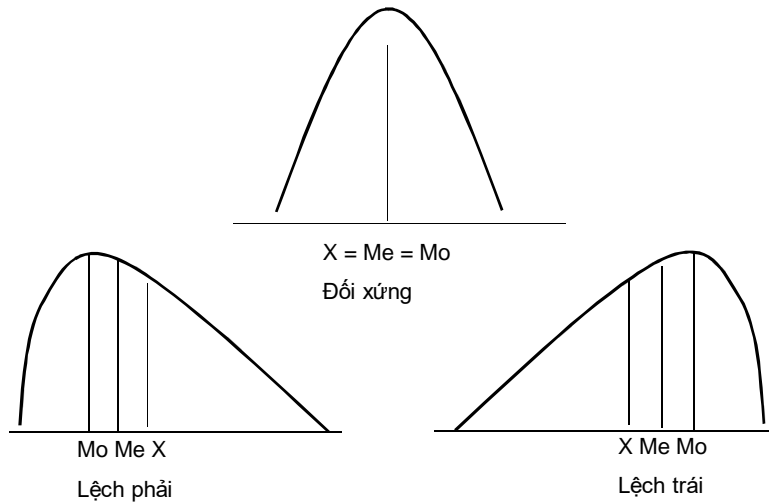
$$i = \frac{p}{100} (n + 1)$$

Bước 3: m là số i được làm tròn xuống

Bước 4: Phân vị mức p

$$P.p = x_m + (i - m) \times (x_{m+1} - x_m)$$

* Đặc trưng phân phối của dãy số



Sơ đồ hộp (box plot)



Tác dụng của box plot

- Nhận biết sự dàn trải của dữ liệu trên cơ sở độ dài của hộp (khoảng tứ phân vị IQR).
- Nhận biết độ lệch phân phối của dữ liệu
- Nhận biết lượng biến động xuất và nghi ngờ là đột xuất

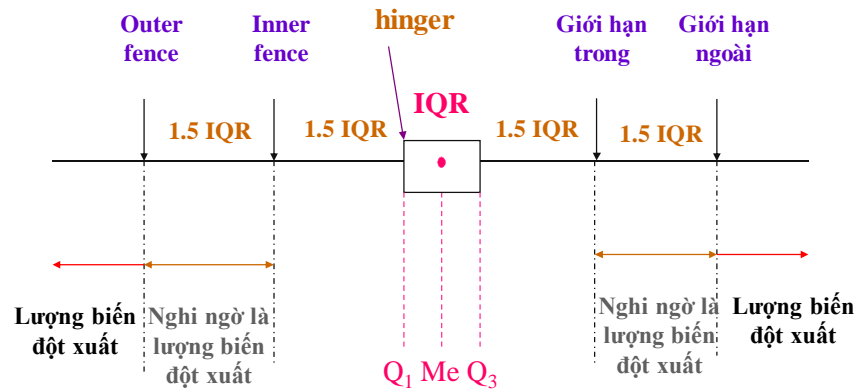
151

Tác dụng của box plot

- Nhận biết sự dàn trải của dữ liệu trên cơ sở độ dài của hộp (khoảng tứ phân vị IQR).
- Nhận biết độ lệch phân phối của dữ liệu
- Nhận biết lượng biến động xuất và nghi ngờ là đột xuất
- So sánh 2 hay nhiều bộ dữ liệu với cùng 1 thước đo

152

Nhận biết lượng biến động xuất



Bài 2

153

III. Các tham số đo độ phân tán (biến thiên)

- 1 Khoảng biến thiên
- 2 Độ lệch tuyệt đối bình quân
- 3 Phương sai
- 4 Độ lệch tiêu chuẩn
- 5 Hệ số biến thiên

ý nghĩa: trị số của các tham số tính ra càng lớn thì lượng biến càng thay đổi, trình độ đại biểu cho số bình quân càng thấp và ngược lại

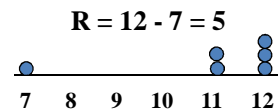
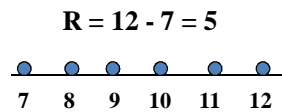
1. Khoảng biến thiên

- Công thức tính:

$$R = X_{\max} - X_{\min}$$

1. Khoảng biến thiên

- Không phụ thuộc vào sự phân bố của dữ liệu:



2. Độ lệch tuyệt đối bình quân

- Công thức tính

$$\bar{d} = \frac{\sum |x_i - \bar{x}|}{n} \qquad \bar{d} = \frac{\sum |x_i - \bar{x}| f_i}{\sum f_i} \text{ (cả quy 0n sè)}$$

3. Phương sai

Công thức tính:

$$S^2 = \frac{\sum (x_i - \bar{x})^2}{n-1} \qquad S^2 = \frac{\sum (x_i - \bar{x})^2 f_i}{\sum f_i - 1} \quad (\text{có quyền số})$$

Công thức thực hành:

$$S^2 = \frac{\sum x_i^2 f_i}{\sum f_i - 1} - \frac{(\sum x_i f_i)^2}{\sum f_i (\sum f_i - 1)} = \frac{\sum f_i}{\sum f_i - 1} \left[\frac{\sum x_i^2 f_i}{\sum f_i} - \left(\frac{\sum x_i f_i}{\sum f_i} \right)^2 \right]$$
$$S^2 = \frac{\sum f_i}{\sum f_i - 1} \left[\overline{x^2} - \bar{x}^2 \right]$$

4. Độ lệch tiêu chuẩn

- Công thức tính: $S = \sqrt{S^2}$

5. Hệ số biến thiên

- Công thức tính: $V = \frac{S}{\bar{x}} \times 100$

* Các tham số phản ánh phân phối

- Hệ số bất đối xứng: $Skewness = \frac{n \sum_{i=1}^n \left(\frac{x_i - \bar{x}}{S} \right)^3}{(n-1)(n-2)}$

Hệ số này có giá trị càng gần 0 thì phân phối của dãy số lại càng đối xứng qua giá trị μ . Khi hệ số này nhỏ hơn 0, dãy số phân phối chuẩn lệch trái. Khi hệ số này lớn hơn 0, dãy số phân phối chuẩn lệch phải.

* Các tham số phản ánh phân phối

$$\text{- Hệ số độ nhọn: Kurtosis} = \frac{n(n+1) \sum_{i=1}^n \left(\frac{x_i - \bar{x}}{S} \right)^4}{(n-1)(n-2)(n-3)}$$

Đối với phân phối chuẩn thì giá trị của hệ số Kurtosis bằng 3.

$$K = \left[\frac{n(n+1) \sum_{i=1}^n \left(\frac{x_i - \bar{x}}{S} \right)^4}{(n-1)(n-2)(n-3)} \right] - \frac{3(n-1)^2}{(n-2)(n-3)}$$

Khi giá trị này bằng 0 thì đó là phân phối chuẩn, nếu giá trị mang dấu dương thì phân phối nhọn hơn so với phân phối chuẩn và ngược

163

CHƯƠNG V: ĐIỀU TRA CHỌN MẪU



I. Những vấn đề chung về ĐTCM

- Khái niệm
- Ưu điểm
- Hạn chế
- Trường hợp vận dụng

Khái niệm

ĐTCM là loại điều tra không toàn bộ trong đó người ta chọn ra *một số* đơn vị đủ lớn để tiến hành điều tra thực tế. Các đơn vị được *chọn theo những quy tắc nhất định* để đảm bảo tính đại biểu, kết quả ĐTCM thường dùng để *suy rộng* cho tổng thể chung

Ưu điểm

- + Tiết kiệm (chi phí, nhân lực, thời gian)
- + Mở rộng nội dung điều tra
- + Tài liệu thu được có độ chính xác cao
- + Tổ chức đơn giản

Hạn chế

- + Không cho biết thông tin đầy đủ về tổng thể chung
- + 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
- Kiểm định giả thuyết thống kê

II. Điều tra chọn mẫu ngẫu nhiên

● Lý luận chung về điều tra chọn mẫu ngẫu nhiên

● Chọn mẫu thời điểm

● Một số phương pháp tổ chức chọn mẫu ngẫu nhiên

1. Lý luận chung về điều tra chọn mẫu ngẫu nhiên

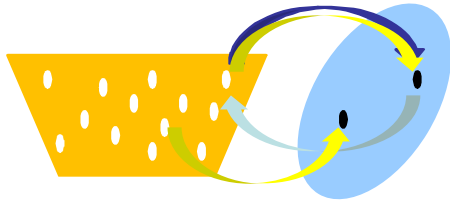
- a Tổng thể chung và tổng thể mẫu
- b Cách chọn
- c Ước lượng (suy rộng) kết quả điều tra
- d Xác định kích thước (quy mô) mẫu

a. Tổng thể chung và tổng thể mẫu

Chỉ tiêu	Tổng thể chung	Tổng thể mẫu
Quy mô	N	n
Số bình quân	$\mu = \frac{\sum x_i}{N}$	$\bar{x} = \frac{\sum x_i}{n}$
Tỷ lệ theo một tiêu thức	p	f
Phương sai	σ^2	s^2
	$p(1-p)$	$f(1-f)$

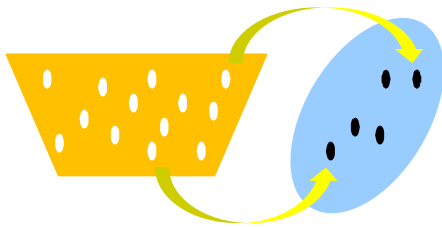
b. Cách chọn mẫu

Chọn hoàn lại (Chọn nhiều lần, chọn lặp)



b. Cách chọn mẫu

Chọn không hoàn lại, (Chọn 1 lần, chọn không lặp)



c. Ước lượng kết quả điều tra

- Với mức ý nghĩa α (độ tin cậy/xác suất $P=1-\alpha$)
- **Ước lượng trung bình**

Khi biết phương sai tổng thể chung (hoặc n lớn)

$$\text{Hai phía} \quad \bar{x} - z_{\alpha/2} \cdot \sigma_{\bar{x}} \leq \mu \leq \bar{x} + z_{\alpha/2} \cdot \sigma_{\bar{x}}$$

$$\text{Vế phải} \quad \bar{x} - z_{\alpha} \cdot \sigma_{\bar{x}} \leq \mu \leq \infty$$

$$\text{Vế trái} \quad -\infty \leq \mu \leq \bar{x} + z_{\alpha} \cdot \sigma_{\bar{x}}$$

Khi chưa biết phương sai tổng thể chung (mẫu nhỏ)

$$\text{Hai phía} \quad \bar{x} - t_{\alpha/2}^{n-1} \cdot \sigma_{\bar{x}} \leq \mu \leq \bar{x} + t_{\alpha/2}^{n-1} \cdot \sigma_{\bar{x}}$$

$$\text{Vế phải} \quad \bar{x} - t_{\alpha}^{n-1} \cdot \sigma_{\bar{x}} \leq \mu \leq \infty$$

$$\text{Vế trái} \quad -\infty \leq \mu \leq \bar{x} + t_{\alpha}^{n-1} \cdot \sigma_{\bar{x}}$$

c. Ước lượng kết quả điều tra

- Với mức ý nghĩa α (độ tin cậy/xác suất $P=1-\alpha$)
- **Ước lượng tỷ lệ**

$$\text{Hai phía} \quad f - z_{\alpha/2} \cdot \sigma_f \leq p \leq f + z_{\alpha/2} \cdot \sigma_f$$

$$\text{Vế phải} \quad f - z_{\alpha} \cdot \sigma_f \leq p \leq \infty$$

$$\text{Vế trái} \quad -\infty \leq p \leq \bar{x} + z_{\alpha} \cdot \sigma_f$$

Ước lượng kết quả điều tra

- ❖ Trong đó Z_{α} , t_{α} được gọi là hệ số tin cậy (giá trị tới hạn mức α của phân phối chuẩn hoá và phân phối Student)
- α – mức ý nghĩa, $P=(1-\alpha)$ là xác suất hay trình độ tin cậy

Ước lượng kết quả điều tra

Một số giá trị đặc biệt của phân phối chuẩn hoá

$Z_{\alpha/2}$	Xác suất tin cậy
1	0,6826
2	0,9544
3	0,9974

Xác suất tin cậy	Mức ý nghĩa	$Z_{\alpha/2}$
0,900	0,100	1,645
0,950	0,050	1,960
0,975	0,025	2,326
0,990	0,010	2,576

- ❖ $\sigma_{\bar{x}}$, σ_f được gọi là sai số bình quân chọn mẫu

Sai số bình quân chọn mẫu

Cách chọn Suy rộng	Hoàn lại (chọn nhiều lần)	Không hoàn lại (chọn một lần)
Số bình quân	$\sigma_{\bar{x}} = \sqrt{\frac{\sigma^2}{n}}$	$\sigma_{\bar{x}} = \sqrt{\frac{\sigma^2}{n} \left(1 - \frac{n}{N}\right)}$
	$\sigma_{\bar{x}} = \sqrt{\frac{S^2}{n}}$	$\sigma_{\bar{x}} = \sqrt{\frac{S^2}{n} \left(1 - \frac{n}{N}\right)}$
Tỷ lệ	$\sigma_f = \sqrt{\frac{f(1-f)}{n}}$	$\sigma_f = \sqrt{\frac{f(1-f)}{n} \left(1 - \frac{n}{N}\right)}$

Các nguyên nhân sai số trong ĐTCM

- Vi phạm nguyên tắc chọn mẫu ngẫu nhiên
- Số lượng đơn vị mẫu không đủ lớn
- Kết cấu tổng thể mẫu khác với kết cấu tổng thể chung
- Sai số do đăng ký, ghi chép

d. Xác định số đơn vị mẫu điều tra

- **Yêu cầu:**
 - + Sai số nhỏ nhất
 - + Chi phí thấp nhất



Công thức xác định

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

$$\varepsilon_{\bar{x}} = z_{\alpha/2} \sigma_{\bar{x}}, \varepsilon_f = z_{\alpha/2} \sigma_f \text{ Phạm vi sai số chọn mẫu}$$

Các nhân tố ảnh hưởng tới kích thước mẫu

- + Cách chọn mẫu
- + Hệ số tin cậy /Trình độ tin cậy
- + Phương sai (độ đồng đều) của tổng thể chung (σ^2)
- + Phạm vi sai số chọn mẫu (ε)

Một số phương pháp xác định phương sai tổng thể chung

- + Lấy phương sai (σ^2) lớn nhất hoặc tỷ lệ (p) gần 0,5 nhất trong các lần điều tra trước (nếu có)
- + 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

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

2. Chọn mẫu thời điểm

Là phương pháp chọn mẫu đặc biệt, trong đó mẫu được xét theo phạm vi thời gian.

Chú ý:

- Khái niệm tổng thể chung và mẫu theo yếu tố thời gian
- Số đơn vị tổng thể chung coi là vô hạn nếu thời gian điều tra ngắn
- Chỉ có thể áp dụng phương pháp chọn không hoàn lại

3. Một số phương pháp tổ chức chọn mẫu ngẫu nhiên

- a. Chọn mẫu ngẫu nhiên đơn giản
- b. Chọn mẫu hệ thống (máy móc)
- c. Chọn mẫu phân loại (phân tổ)
- d. Chọn mẫu cả khối (mẫu chùm)
- e. Chọn mẫu phân tầng (nhiều cấp)

III. Điều tra chọn mẫu phi ngẫu nhiên

➤ Một số phương pháp chọn mẫu phi ngẫu nhiên:

- Chọn mẫu tiện lợi (thuận tiện)
- Chọn mẫu phán đoán
- Chọn mẫu hạn ngạch
- Chọn mẫu tích lũy

“Điểm yếu của tất cả các phương pháp phi ngẫu nhiên là không có sự phát triển về lý thuyết, chọn mẫu phi ngẫu nhiên chỉ có thể được đánh giá bằng chủ quan” Graham Kalton

CHƯƠNG VI: KIỂM ĐỊNH GIẢ THUYẾT THỐNG KÊ



I. Những vấn đề chung về kiểm định

- 1 Giả thuyết thống kê
- 2 Sai lầm và mức ý nghĩa trong kiểm định
- 3 Tiêu chuẩn kiểm định
- 4 Các bước tiến hành một kiểm định giả thuyết thống kê

1. Giả thuyết thống kê

Là giả thuyết về một vấn đề nào đó của tổng thể chung (về các tham số như trung bình, tỷ lệ, phương sai, dạng phân phối,...)

1. Giả thuyết thống kê

Giả thuyết mà ta muốn kiểm định (H_0)

Giả thuyết đối lập (H_a, H_1, \bar{H})

2. Sai lầm và mức ý nghĩa trong kiểm định

- Sai lầm loại I là bác bỏ H_0 khi H_0 đúng
- Sai lầm loại II là chấp nhận H_0 khi H_0 sai

2. Sai lầm và mức ý nghĩa trong kiểm định

Thực tế \ Kết luận	Chấp nhận H_0	Bác bỏ H_0
	H_0 đúng	Kết luận đúng
H_0 sai	Sai lầm loại II	Kết luận đúng

2. Sai lầm và mức ý nghĩa trong kiểm định

Mức ý nghĩa của kiểm định (α) là xác suất mắc sai lầm loại I

$$\alpha = P(\text{Bác bỏ } H_0 / H_0 \text{ đúng})$$

3. Tiêu chuẩn kiểm định

Tiêu chuẩn kiểm định là quy luật phân phối xác suất nào đó dùng để kiểm định.

Trong tập hợp các kiểm định thống kê có cùng mức ý nghĩa α , kiểm định nào có xác suất mắc sai lầm loại 2 nhỏ nhất được xem là “tốt nhất”.

4. Các bước tiến hành kiểm định

- Xây dựng giả thuyết H_0 và giả thuyết đối H_1
- Xác định mức ý nghĩa α
- Chọn tiêu chuẩn kiểm định
- Tính giá trị của tiêu chuẩn kiểm định từ mẫu quan sát
- Kết luận

Kết luận

Quy tắc kiểm định giả thuyết thống kê

- Nếu giá trị tiêu chuẩn kiểm định thuộc miền bác bỏ (W_α), kết luận H_0 sai, có cơ sở để bác bỏ H_0
- Nếu giá trị của tiêu chuẩn kiểm định thuộc miền không bác bỏ, chưa khẳng định H_0 đúng mà kết luận chưa đủ cơ sở để bác bỏ H_0

II. Kiểm định trung bình

1

Kiểm định giả thuyết về giá trị trung bình của một tổng thể chung

2

Kiểm định hai giá trị trung bình của hai tổng thể

3

Kiểm định trung bình thuộc nhiều tổng thể chung

1. Kiểm định giả thuyết về giá trị trung bình của một tổng thể chung

- Giả sử nghiên cứu $X \sim N(\mu, \sigma^2)$
- Chưa biết μ song có cơ sở để giả định nó bằng μ_0
($H_0: \mu = \mu_0$)
- Để kiểm định giả thuyết trên, lấy mẫu n đơn vị từ đó tính trung bình mẫu \bar{x}
- Tiêu chuẩn kiểm định

a. Trường hợp đã biết σ^2

Tiêu chuẩn kiểm định

$$Z = \frac{(\bar{x} - \mu_0)}{\sigma / \sqrt{n}} \sim N(0, 1)$$

a. Trường hợp đã biết σ^2

Miền bác bỏ W_α

- Hai phía: $|Z_{qs}| > z_{\alpha/2}$
- Vế phải: $Z_{qs} > z_\alpha$
- Vế trái: $Z_{qs} < -z_\alpha$

b. Trường hợp chưa biết σ^2

Tiêu chuẩn kiểm định

$$T = \frac{(\bar{x} - \mu_0)}{S / \sqrt{n}} \sim t_{(n-1)}$$

Trong đó
$$S^2 = \frac{\sum (x_i - \bar{x})^2 f_i}{\sum f_i - 1} = \frac{\sum f_i}{\sum f_i - 1} (\overline{x^2} - (\bar{x})^2)$$

b. Trường hợp chưa biết σ^2

Miền bác bỏ W_α

- Hai phía: $|T_{qs}| > t_{\alpha/2(n-1)}$
- Vế phải: $T_{qs} > t_{\alpha(n-1)}$
- Vế trái: $T_{qs} < -t_{\alpha(n-1)}$

b. Trường hợp chưa biết σ^2

Trong trường hợp số quan sát (n) lớn, thống kê t có phân phối xấp xỉ chuẩn, nên có thể viết tiêu chuẩn kiểm định là:

$$Z = \frac{(\bar{x} - \mu_0)}{S / \sqrt{n}} \sim N(0,1)$$

b. Trường hợp chưa biết σ^2

Miền bác bỏ W_α

- Hai phía: $|Z_{qs}| > z_{\alpha/2}$
- Vế phải: $Z_{qs} > z_\alpha$
- Vế trái: $Z_{qs} < -z_\alpha$

2. Kiểm định hai giá trị trung bình của hai tổng thể

- Hai mẫu độc lập
- Hai mẫu phụ thuộc



a. Hai mẫu độc lập

- Giả sử nghiên cứu 2 tổng thể $X_1 \sim N(\mu_1, \sigma_1^2)$ và $X_2 \sim N(\mu_2, \sigma_2^2)$
- Chưa biết μ_1 và μ_2 song có cơ sở để giả định nó bằng nhau ($H_0: \mu_1 = \mu_2$)
- Để kiểm định giả thuyết trên, lấy 2 mẫu ngẫu nhiên độc lập, với kích thước n_1 và n_2 từ đó tính được \bar{x}_1 và \bar{x}_2
- Tiêu chuẩn kiểm định

Trường hợp đã biết σ_1^2 và σ_2^2

Tiêu chuẩn kiểm định

$$Z = \frac{(\bar{x}_1 - \bar{x}_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0, 1)$$

Trường hợp đã biết σ_1^2 và σ_2^2

Miền bác bỏ W_α

- Hai phía: $|Z_{qs}| > z_{\alpha/2}$
- Vế phải: $Z_{qs} > z_\alpha$
- Vế trái: $Z_{qs} < -z_\alpha$

Trường hợp chưa biết σ_1^2 và σ_2^2

Phương sai bằng nhau

Phương sai không bằng nhau

Kiểm định phương sai

-Tiêu chuẩn kiểm định $F = \frac{S_1^2}{S_2^2} \sim f_{\alpha(n_1-1; n_2-1)}$

Miền bác bỏ giả thuyết $H_0: F_{qs} < f_{1-\alpha/2, (n_1-1)(n_2-1)}$
hoặc $F_{qs} > f_{\alpha/2, (n_1-1)(n_2-1)}$

Trường hợp chưa biết σ_1^2 và σ_2^2 và phương sai bằng nhau

Tiêu chuẩn kiểm định

$$T = \frac{(\bar{x}_1 - \bar{x}_2)}{\sqrt{\frac{S^2}{n_1} + \frac{S^2}{n_2}}} \sim t_{\alpha, (n_1+n_2-2)}$$

Trong đó
$$S^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{(n_1 - 1) + (n_2 - 1)}$$

Trường hợp chưa biết σ_1^2 và σ_2^2 và phương sai bằng nhau

Miền bác bỏ W_α

- Hai phía: $|T_{qs}| > t_{\alpha/2, (n_1+n_2-2)}$
- Vế phải: $T_{qs} > t_{\alpha, (n_1+n_2-2)}$
- Vế trái: $T_{qs} < -t_{\alpha, (n_1+n_2-2)}$

Trường hợp chưa biết σ_1^2 và σ_2^2 và phương sai không bằng nhau

Tiêu chuẩn kiểm định

$$T = \frac{(\bar{x}_1 - \bar{x}_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \sim t_{\alpha, v}$$

$$v = \frac{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)^2}{\frac{1}{n_1-1} \left(\frac{S_1^2}{n_1}\right)^2 + \frac{1}{n_2-1} \left(\frac{S_2^2}{n_2}\right)^2}$$

Nếu v lẻ thì làm tròn xuống

Trường hợp chưa biết σ_1^2 và σ_2^2 và phương sai không bằng nhau

Miền bác bỏ W_α

- Hai phía: $|T_{qs}| > t_{\alpha/2(v)}$
- Vế phải: $T_{qs} > t_{\alpha(v)}$
- Vế trái: $T_{qs} < -t_{\alpha(v)}$

b. Hai mẫu phụ thuộc

- Giả sử nghiên cứu 2 tổng thể phụ thuộc

$$X_1 \sim N(\mu_1, \sigma_1^2) \text{ và } X_2 \sim N(\mu_2, \sigma_2^2)$$

Muốn so sánh μ_1 và μ_2 ta xét độ lệch trung bình μ_d (chưa biết) có cơ sở để giả định độ lệch trung bình bằng μ_0 ($H_0: \mu_d = \mu_0$)

- Để kiểm định giả thuyết trên, lấy mẫu ngẫu nhiên phụ thuộc được hình thành bởi n cặp quan sát của 2 mẫu, tính độ lệch giữa 2 mẫu (d_i), từ đó tính trung bình của các độ lệch giữa các cặp của 2 mẫu (\bar{d})

- Tiêu chuẩn kiểm định

b. Hai mẫu phụ thuộc

Tiêu chuẩn kiểm định

$$T = \frac{(\bar{d} - \mu_0)}{S_d / \sqrt{n}} \sim t_{\alpha(n-1)}$$

b. Hai mẫu phụ thuộc

Miền bác bỏ W_α

- Hai phía: $|T_{qs}| > t_{\alpha/2(n-1)}$
- Vế phải: $T_{qs} > t_{\alpha(n-1)}$
- Vế trái: $T_{qs} < -t_{\alpha(n-1)}$

b. Hai mẫu phụ thuộc

Ưu điểm: Thường cho kết quả chính xác hơn vì đã bỏ được các nhân tố ngoại lai

Hạn chế: Tiến hành thực hiện thu thập thông tin phức tạp hơn

3. Kiểm định trung bình thuộc nhiều tổng thể chung (one-way ANOVA)

- Giả sử có k tổng thể đều có phân phối $X_j \sim N(\mu_j, \sigma_j^2)$
- Chưa biết μ_i song có cơ sở để giả định nó bằng nhau ($H_0: \mu_1 = \mu_2 = \dots = \mu_k$)
- Để kiểm định giả thuyết trên, lấy k mẫu với cỡ mẫu tương ứng n_1, n_2, \dots, n_k
- Tiêu chuẩn kiểm định

3. Kiểm định trung bình thuộc nhiều tổng thể chung (one-way ANOVA)

- Tiêu chuẩn kiểm định $F = \frac{MSF}{MSE} \sim f_{\alpha(k-1, n-k)}$

$$MSF = \frac{SSF}{k-1} \quad MSE = \frac{SSE}{n-k}$$

3. Kiểm định trung bình thuộc nhiều tổng thể chung (one-way ANOVA)

$$SST = \sum_{j=1}^k \sum_{i=1}^{n_j} (x_{ij} - \bar{x})^2 \quad \text{Total Sum of Squares}$$

$$SSF = \sum_{j=1}^k (\bar{x}_j - \bar{x})^2 \cdot n_j \quad \text{(Sum of Squares for Factor)}$$

$$SSE = \sum_{j=1}^k \sum_{i=1}^{n_j} (x_{ij} - \bar{x}_j)^2 \quad \text{(Sum of Squares for Error)}$$

$$SST = SSF + SSE$$

3. Kiểm định trung bình thuộc nhiều tổng thể chung (one-way ANOVA)

Miền bác bỏ $W_{\alpha}: F_{qs} > f_{\alpha(k-1; n-k)}$

3. Kiểm định trung bình thuộc nhiều tổng thể chung (one-way ANOVA)

Kết quả (chạy bằng phần mềm) được thể hiện trong bảng sau (ANOVA – Analysis of Variance):

ANOVA

<i>Source of Variation</i>	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>
Between Groups	SSF	k-1	MSF	***
Within Groups	SSE	n-k	MSE	
Total	SST	n-1		

III. Kiểm định tỷ lệ

1

Kiểm định giả thuyết về tỷ lệ của tổng thể chung

2

Kiểm định tỷ lệ của hai tổng thể chung

3

Kiểm định tỷ lệ của nhiều tổng thể chung

1. Kiểm định giả thuyết về tỷ lệ của tổng thể chung

- Giả sử nghiên cứu tiêu thức A nào đó của một tổng thể chung
- Chưa biết p song có cơ sở để giả định nó bằng p_0
($H_0: p = p_0$)
- Để kiểm định giả thuyết trên, lấy mẫu n đơn vị từ đó tính tỷ lệ mẫu f
- Tiêu chuẩn kiểm định

1. Kiểm định giả thuyết về tỷ lệ của tổng thể chung

Khi n đủ lớn ($n \cdot f \geq 5$ hoặc $n(1-f) \geq 5$)

Tiêu chuẩn kiểm định

$$Z = \frac{(f - p_0)}{\sqrt{p_0(1 - p_0) / n}} \sim N(0, 1)$$

1. Kiểm định giả thuyết về tỷ lệ của tổng thể chung

Miền bác bỏ W_α

- Hai phía: $|Z_{qs}| > z_{\alpha/2}$

- Vế phải: $Z_{qs} > z_\alpha$

- Vế trái: $Z_{qs} < -z_\alpha$

2. Kiểm định tỷ lệ của hai tổng thể chung

- Giả sử nghiên cứu 2 tổng thể, tỷ lệ chung theo tiêu thức A lần lượt là p_1 và p_2
- Chưa biết p_1 và p_2 song có cơ sở để giả định nó bằng nhau ($H_0: p_1 = p_2$)
- Để kiểm định giả thuyết trên, lấy 2 mẫu ngẫu nhiên với kích thước n_1 và n_2 từ đó tính được f_1 và f_2
- Tiêu chuẩn kiểm định

2. Kiểm định tỷ lệ của hai tổng thể chung

Khi n_1 và n_2 đủ lớn, tức là $n_1 \cdot f_1 \geq 5$ hoặc $n_1(1-f_1) \geq 5$

Và $n_2 \cdot f_2 \geq 5$ hoặc $n_2(1-f_2) \geq 5$

Tiêu chuẩn kiểm định

$$Z = \frac{f_1 - f_2}{\sqrt{f(1-f)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \sim N(0, 1)$$

Trong đó $f = \frac{n_1 f_1 + n_2 f_2}{n_1 + n_2}$

2. Kiểm định tỷ lệ của hai tổng thể chung

Miền bác bỏ W_α

- Hai phía: $|Z_{qs}| > z_{\alpha/2}$

- Vế phải: $Z_{qs} > z_\alpha$

- Vế trái: $Z_{qs} < -z_\alpha$

3. Kiểm định tỷ lệ của nhiều tổng thể chung

Có ý kiến cho rằng tỷ lệ nghèo ở 3 địa phương (A, B và C) là khác nhau? Từ mỗi địa phương chọn ngẫu nhiên 1 số hộ gia đình và có kết quả như sau:

3. Kiểm định tỷ lệ của nhiều tổng thể chung

Loại hộ \ ĐP	A	B	C
Hộ nghèo	20	50	25
Hộ không nghèo	180	350	95

-Hãy đưa ra kết luận với mức ý nghĩa 5%

-Bảng trên được gọi là bảng ngẫu nhiên 2 dòng ($i=1,2$) và 3 cột ($j=1,3$)

3. Kiểm định tỷ lệ của nhiều tổng thể chung

Gọi tỷ lệ hộ nghèo của địa phương A, B và C lần lượt là p_1 , p_2 và p_3

Cặp giả thuyết cần kiểm định là

$$H_0: p_1 = p_2 = p_3$$

$$H_1: \exists p_i \neq p_j (i \neq j)$$

3. Kiểm định tỷ lệ của nhiều tổng thể chung

Gọi n_{ij} là tần số thực nghiệm (số quan sát ở dòng thứ i và cột thứ j)

Tính tần số lý thuyết (\hat{n}_{ij})

$$\hat{n}_{ij} = \frac{\text{tæng dßng } i \times \text{tæng cét } j}{n}$$

$$n = \sum_i \sum_j n_{ij} \quad \text{Tổng số đơn vị điều tra}$$

3. Kiểm định tỷ lệ của nhiều tổng thể chung

Tiêu chuẩn kiểm định

$$\chi^2 = \sum_i \sum_j \frac{(n_{ij} - \hat{n}_{ij})^2}{\hat{n}_{ij}} \sim \chi_{\alpha, df}^2$$

$$df = (\text{sè dßng} - 1) \cdot (\text{sè cét} - 1)$$