

NGUYÊN LÝ THỐNG KÊ KINH TẾ

NGÔ THÁI HÙNG
ĐH Tài Chính – Marketing
Email: ngothai.hung@gmail.com

CHƯƠNG 1

NHỮNG VẤN ĐỀ CHUNG VỀ THỐNG KÊ

GIỚI THIỆU

- Hơn 4.400 người chết ở Philippines vì bão (Theo vnexpress)



GIỚI THIỆU

LỊCH THI ĐẤU		CHÂU Á			TỔNG BÀN THẮNG		
UEFA Champions League							
14/02 02:45	Shakhtar Donetsk vs Dortmund	0 : 0	0.88	-0.95	2 1/2	0.88	-0.98
14/02 02:45	Real Madrid vs Manchester Utd	0 : 1	0.99	0.94	2 3/4	0.82	-0.92

GIỚI THIỆU

Chứng khoán chiều 25/12

VN-Index sụt giảm mạnh thêm về cuối phiên do mất trụ đỡ.

Nguồn: <http://vneconomy.vn>

GIỚI THIỆU

Sau khi bài viết "*Nên để thị trường bất động sản rơi tự do*" nêu quan điểm của TS Alan Phan được đăng tải, đã có nhiều ý kiến xung quanh vấn đề này. Còn quan điểm của bạn ra sao?

TS Alan Phan: "Yêu cầu về bất động sản tại Việt Nam là rất lớn, song phần lớn các sản phẩm hiện nay vẫn chưa phù hợp với túi tiền thì người dân chưa mua mà thôi"

Nguồn: <http://vneconomy.vn>

GIỚI THIỆU



GIỚI THIỆU

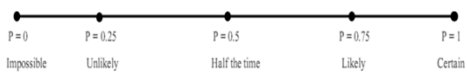
Hàng ngày có các quyết định dựa trên các thông tin không đầy đủ. Ta xét:

Cổ phiếu công ty IBM sẽ (will) cao hơn trong vòng 6 tháng tới.

GIỚI THIỆU

Vì không chắc chắn, các báo cáo cần được sửa đổi:

Giá cổ phiếu IBM có *khả năng (likely)* cao hơn trong sáu tháng tới.



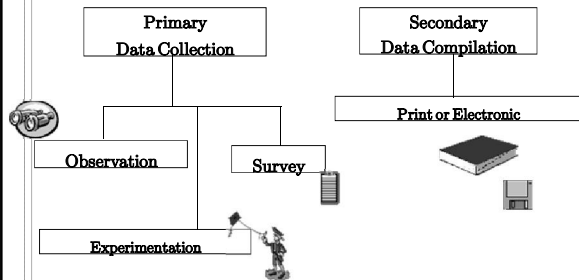
THỐNG KÊ LÀ GÌ?

- **Thống kê (Statistics)** là khoa học giúp con người đưa ra quyết định tốt hơn trong kinh doanh và kinh tế cũng như các lĩnh vực khác
- **Thống kê (Statistics)** hướng dẫn chúng ta cách thức tóm tắt, phân tích và suy diễn từ tập dữ liệu, sau đó dẫn đến quyết định đúng đắn.
- Các quyết định này giúp chúng ta tiến bộ trong điều hành, ví dụ: doanh nghiệp, kinh tế xã hội ...

SỬ DỤNG THỐNG KÊ THEO HAI PHÂN LOẠI

- | | |
|--------------------------|-----------------------------------|
| • Thống kê mô tả | • Thống kê suy diễn |
| • Descriptive Statistics | • Inferential Statistics |
| ✓ Thu thập | ✓ Dự báo giá trị tham số tổng thể |
| ✓ Phân loại | ✓ Kiểm định giá trị tổng thể |
| ✓ Tóm tắt | ✓ Quyết định |
| ✓ Trình bày | |
| ✓ Phân tích | |

NGUỒN DỮ LIỆU Data Sources



1-13

DỮ LIỆU

- **Định tính** (Qualitative- Categorical or Nominal):
- **Định lượng** (Quantitative - Measurable or Countable:)

Ví dụ:

- ✓Màu sắc
- ✓Giới tính
- ✓Quốc tịch

- Ví dụ:**
- ✓Nhiệt độ
- ✓Tiền lương
- ✓Điểm thi
- ✓Thang điểm 0 - 100

1-14

MẪU & TỔNG THỂ

Samples and Populations

- **Tổng thể** thống kê là tập hợp các đơn vị thuộc hiện tượng nghiên cứu, cần quan sát, thu thập và phân tích mặt lượng của chúng theo một hoặc một số tiêu thức nào đó.
- Mỗi đơn vị cá biệt cấu thành tổng thể đó được gọi chung là đơn vị tổng thể.

1-15

MẪU & TỔNG THỂ

Samples and Populations

- **Mẫu** là tập con của tổng thể được lựa chọn để phân tích

Population	Sample
a b c d e f g h i j k l m n o p q r s t u v w x y z	b c g i n o r u y

MẪU & TỔNG THỂ
Samples and Populations

Population (N) Sample (n)

TẠI SAO PHẢI CHỌN MẪU

- Điều tra trên tổng thể có thể:
 - ✓ Không thể
 - ✓ Không thực thi
 - ✓ Chi phí cao

MỘT SỐ KHÁI NIỆM

- **Quan sát (Observation)** là cơ sở để thu thập số liệu và thông tin cần nghiên cứu.
- **Tiêu thức (statistical criteria)** là đặc điểm của đơn vị tổng thể được lựa chọn để phục vụ cho mục đích nghiên cứu

1-19

MỘT SỐ KHÁI NIỆM

- **Tiêu thức thuộc tính:** phản ánh tích chất hay loại hình của đơn vị tổng thể, không có biểu hiện trực tiếp bằng con số
- **Tiêu thức số lượng:** Là tiêu thức biểu hiện đặc trưng của các đơn vị tổng thể bằng con số cụ thể qua cân, đong, đo, đếm

1-20

MỘT SỐ KHÁI NIỆM

- **Chỉ tiêu thống kê (statistical indicator)** là các trị số phản ánh các đặc điểm, các tính chất cơ bản của tổng thể trong điều kiện thời gian và không gian xác định
- **Chỉ tiêu khối lượng:** là các chỉ tiêu biểu hiện qui mô của tổng thể
- **Chỉ tiêu khối lượng:** là các chỉ tiêu biểu hiện qui mô của tổng thể

1-21

MỘT SỐ KHÁI NIỆM

- **Lượng biến (x_i)** là các trị số cụ thể khác nhau của tiêu thức số lượng.
- **Lượng biến rời rạc:** Giá trị có thể nhận một trị số cụ thể, có thể đếm được. *Vi dụ:* Số lượng học sinh trong 1 lớp, số người trong gia đình...
- **Lượng biến liên tục:** Giá trị có thể có của nó có thể lấp đầy một khoảng trên trục số. *Vi dụ:* Chiều cao, cân nặng của các bạn trong lớp...

1-22

MỘT SỐ KHÁI NIỆM

- **Tần số (f_i):** là số lần xuất hiện giá trị (x_i) trong dãy số liệu đã cho.
- **Tần suất của giá trị x_i (F_i):** Tần suất xuất hiện biến cố trong n phép thử là tỷ số giữa số phép thử trong đó biến cố xuất hiện và tổng số phép thử được thực hiện.

1-23

THANG ĐO

Scales of Measurement

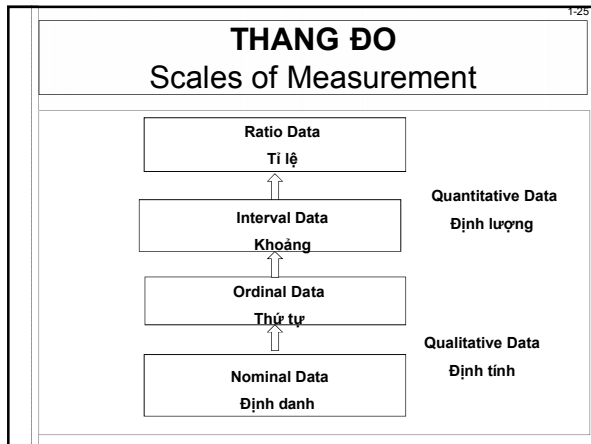
Thang đo là khái niệm dùng để đo lường các mức độ của hiện tượng kinh tế xã hội, trên cơ sở đó mà có thể phân loại, phản ánh mối quan hệ của các đơn vị tổng thể

1-24

THANG ĐO

Scales of Measurement

- **Thang đo định danh (Nominal)**
 - ✓ Giới tính
- **Thứ bậc (Ordinal)**
 - ✓ Xếp hạng (Sinh viên năm mấy?)
- **Khoảng (Interval) – Sự khác nhau hoặc các vấn đề về khoảng cách – Có giá trị 0 tùy ý**
 - ✓ Nhiệt độ ($^{\circ}\text{F}$, $^{\circ}\text{C}$)
- **Tỷ lệ (Ratio)**
 - ✓ Tiền lương



1-26

BẢNG CÂU HỎI

Bản phỏng vấn đơn giản

- Loại điện thoại di động mà bạn sử dụng chính?
 - Nokia
 - Samsung
 - Motorola
 - Khác
- Mức độ hài lòng chung của bạn khi sử dụng loại điện thoại trên?

Rất không hài lòng 1 2 3 4 5 Rất hài lòng
- Chỉ tiêu trung bình một tháng cho việc gọi điện thoại di động ngàn đ
- Bạn theo dõi thông tin về các loại điện thoại mới như thế nào?
 - Không bao giờ
 - Ít khi
 - Thỉnh thoảng
 - Thường xuyên
- Bạn thường sử dụng tính năng nào
 - Nghe - gọi
 - Tin nhắn
 - Nghe nhạc
 - Quay phim, chụp hình
 - Games
 - Khác
- Giới tính: Nam Nữ

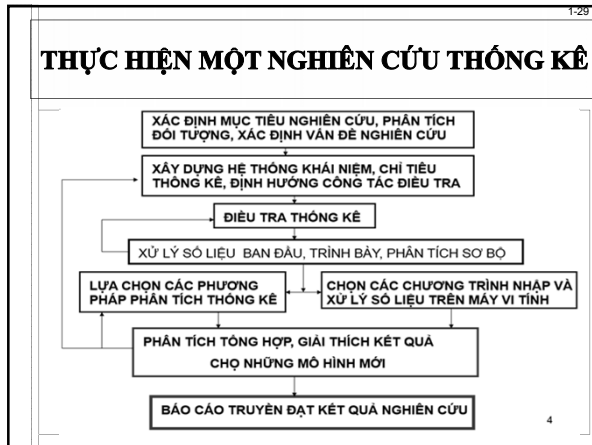
NGUYÊN LÝ THỐNG KÊ KINH TẾ

NGÔ THÁI HƯNG
ĐH Tài Chính – Marketing
Email: ngothai.hung@gmail.com

1-29

CHƯƠNG 2

QUÁ TRÌNH NGHIÊN CỨU THỐNG KÊ



NGUYÊN LÝ THỐNG KÊ KINH TẾ

NGÔ THÁI HƯNG
ĐH Tài Chính – Marketing
Email: ngothai.hung@gmail.com

1-31

CHƯƠNG 3

TỔNG HỢP THỐNG KÊ

1-32

PHÂN TỔ THỐNG KÊ

Phân tổ thống kê là phân chia tổng thể nghiên cứu thành các tổ có tính chất khác nhau dựa trên một hay một số tiêu thức nào đó.

1-33

PHƯƠNG PHÁP PHÂN TỔ

Bước 1. Chọn tiêu thức phân tổ
Bước 2. Xác định số tổ và phạm vi biến thiên của từng tổ
Bước 3. Lựa chọn các tiêu giải thích và sắp xếp các đơn vị vào các tổ tương ứng

1-34

PHƯƠNG PHÁP PHÂN TỔ

Phân tổ theo tiêu thức thuộc tính

- ⊙ Tiêu thức thuộc tính có một vài biểu hiện
- ⊙ Tiêu thức thuộc tính có nhiều biểu hiện

Phân tổ theo tiêu thức số lượng

- ⊙ Phân tổ không có khoảng cách
- ⊙ Phân tổ có khoảng cách tổ

1-35

PHƯƠNG PHÁP PHÂN TỔ

❖ **Phân tổ có khoảng cách tổ đều**

⊙ Trường hợp đối với lượng biến biến thiên liên tục + **Xác định số tổ** (Number off classes): trong thực tế, số tổ **k** được xác định chủ yếu dựa vào **kinh nghiệm** và tùy theo đặc điểm của hiện tượng nghiên cứu. Ngoài ra có thể tham khảo xác định số tổ bằng công thức sau

$$k = (2.n)^{\frac{1}{3}}$$

1-36

PHƯƠNG PHÁP PHÂN TỔ

❖ **Phân tổ có khoảng cách tổ đều**

⊙ Trường hợp đối với lượng biến biến thiên liên tục + **Xác định khoảng cách tổ**

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

1-37

PHƯƠNG PHÁP PHÂN TỔ

■ **Ví dụ:** Lượng vốn đầu tư (triệu USD) của 10 doanh nghiệp nước ngoài đầu tư vào TP Đà Nẵng như sau:

1 2,3 4,6 7 9 3,1 8,3 2 5,1 7,9

Mức vốn (triệu USD)	Số doanh nghiệp
1 – 4,33	4
4,33 – 7,666	3
7,666 – 9	3

1-38

PHƯƠNG PHÁP PHÂN TỔ

❖ **Phân tổ có khoảng cách tổ đều**

⊙ *Trường hợp đối với lượng biến thiên rời rạc:* + **Xác định khoảng cách tổ**

$$h = \frac{(x_{\max} - x_{\min}) - (k - 1)}{k}$$

1-39

PHƯƠNG PHÁP PHÂN TỔ

■ **Ví dụ:** Tuổi nghề của 20 giảng viên trong một trường đại học được cho như sau:

8 5 7 4 9 10 25 28
12 15 2 3 6 5 11 17
1 20 5 2

Tuổi nghề (năm)	Số giảng viên
1 – 7	10
8 – 14	5
15 – 21	3
21 – 28	2

1.40

PHƯƠNG PHÁP PHÂN TỔ

- **Phân tổ có khoảng cách tổ không đều**
Được áp dụng khi lượng biến của tiêu thức biến thiên không đều đặn, hoặc khi mục đích nghiên cứu chỉ nhằm đánh giá hiện tượng về qui mô, mức độ của các loại, theo tiêu chuẩn đã được đề ra

1.41

PHƯƠNG PHÁP PHÂN TỔ

- **Phân tổ mở:** là phân tổ mà tổ đầu tiên không có giới hạn dưới, tổ cuối cùng không có giới hạn trên, các tổ còn lại có thể có khoảng cách tổ đều hoặc tổ không đều.

Doanh thu (triệu đồng)	Số DNTN
< 1000	3
1000 – 2000	8
2000 – 3000	14
3000 – 4000	5
Tổng	30

1.42

TRÌNH BÀY KẾT QUẢ TỔNG HỢP TÀI LIỆU ĐIỀU TRA PHÂN TỔ THỐNG KÊ

- **Bảng thống kê:** là hình thức biểu mẫu thông dụng nhất, được thiết kế với một số cột, một số hàng để trình bày các kết quả tổng hợp số liệu thống kê theo từng nội dung riêng biệt nhằm phục vụ cho yêu cầu của quá trình nghiên cứu thống kê.

1.43

TRÌNH BÀY KẾT QUẢ TỔNG HỢP TÀI LIỆU ĐIỀU TRA PHÂN TỬ THỐNG KÊ

● **Bảng thống kê:**

Phần chủ đề	Phần giải thích				Các chỉ tiêu giải thích (tiêu mục)					Cộng
	Tần số, tần suất, tần số tích lũy, các chỉ tiêu theo nhu cầu phân tích									
a	b	c	...	(1)	(2)	(3)		
Tiểu thức phân tử										
Lượng biến										
Thuộc tính										
Cộng										

1.44

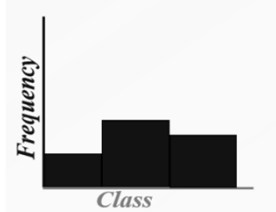
BIỂU ĐỒ

Là một hình thức trình bày có tính chất quy ước các tài liệu thống kê dưới dạng các hình vẽ hay đường nét hình học

1.45

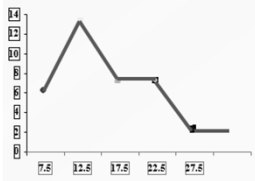
BIỂU ĐỒ

● A Histogram is a graph in which the classes are marked on the horizontal axis and the class frequencies on the vertical axis



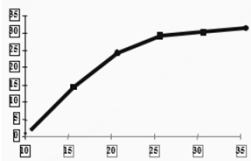
BIỂU ĐỒ

- Frequency Polygons or Line Charts: consists of line segments connecting the points formed by the class midpoint and the class frequency



BIỂU ĐỒ

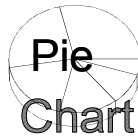
- A cumulative frequency distribution is used to determine how many or what proportion of the data values are below or above a certain value.



BIỂU ĐỒ

Pie chart ... is useful for displaying a Relative Frequency Distribution

A circle is divided proportionally to the relative frequency and portions of the circle are allocated for the different groups.




1-49

BIỂU ĐỒ


Pie chart ...is useful for displaying a Relative Frequency Distribution

A circle is divided proportionally to the relative frequency and portions of the circle are allocated for the different groups.



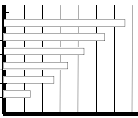
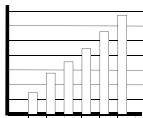
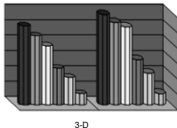
1-50

BIỂU ĐỒ



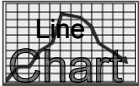
...can be used to depict any of the levels of measurement (nominal, ordinal, interval, or ratio).

Examples of...

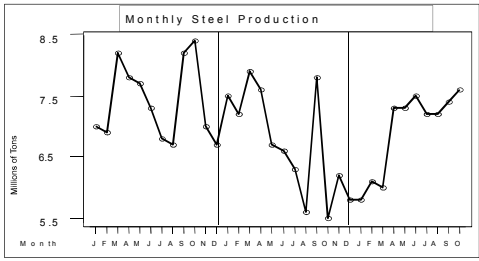




3-D

1-51



Time Plot



Monthly Steel Production

Millions of tons

Month J F M A M J J A S O N D J F M A M J J A S O N D J F M A M J J A S O

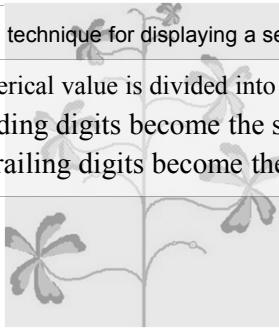
2000 2001 2002

Stem-and-leaf Displays

A statistical technique for displaying a set of data.

Each numerical value is divided into two parts:

1. the leading digits become the stem and
2. the trailing digits become the leaf.



Stem-and-leaf Displays

Question

A student achieved the following scores on the twelve accounting quizzes this semester:

86, 79, 92, 84, 69, 88, 91,
83, 96, 78, 82, 85.

Construct a stem-and-leaf chart to illustrate the results.

Stem-and-leaf Displays

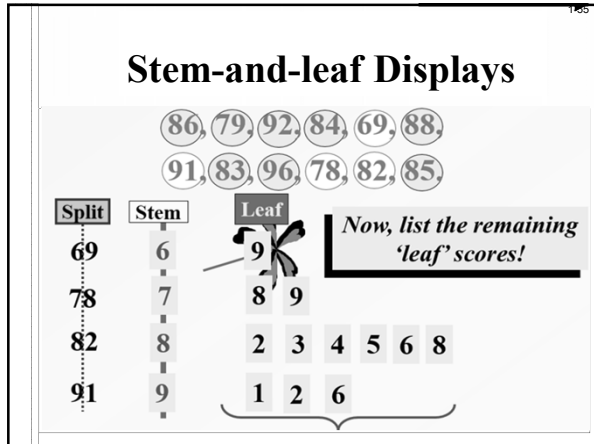
First, find the lowest score

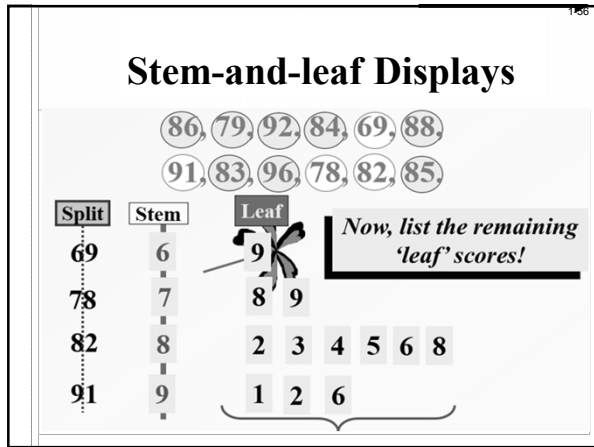
86, 79, 92, 84, **69**, 88,
91, 83, 96, 78, 82, 85.

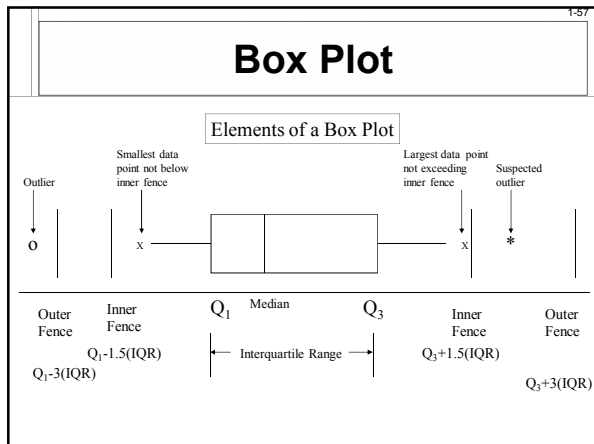
Now list the next scores with the highest leading digits.

You should now have the following **STEMS**:

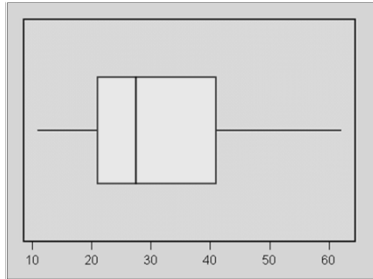
69, 78, 82, 91







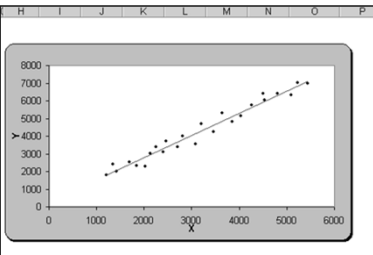
Example: Box Plot



Scatter Plots

- **Scatter Plots** are used to identify and report any underlying relationships among pairs of data sets.
- The plot consists of a scatter of points, each point representing an observation.

Scatter Plots



- Scatter plot with trend line.
- This type of relationship is known as a positive correlation.

Correlation will be discussed in later chapters.

NGUYÊN LÝ THỐNG KÊ KINH TẾ

NGÔ THÁI HƯNG
ĐH Tài Chính – Marketing
Email: ngothai.hung@gmail.com

CHƯƠNG 4

THỐNG KÊ CÁC MỨC ĐỘ CỦA HIỆN TƯỢNG KINH TẾ XH

SỐ TUYỆT ĐỐI

Số tuyệt đối trong thống kê là chỉ tiêu biểu hiện quy mô, khối lượng của hiện tượng kinh tế xã hội trong điều kiện thời gian và địa điểm cụ thể.

■ **Ví dụ:** Tính đến 1/4/2011, dân số Việt Nam là 87.610.947 người

1-64

SỐ TUYỆT ĐỐI

- **Các loại số tuyệt đối**
- **Số tuyệt đối thời điểm:** *Số tuyệt đối thời điểm phản ánh quy mô, khối lượng của hiện tượng nghiên cứu vào một thời điểm nhất định.*
- **Ví dụ:** Số lao động của một xí nghiệp có vào ngày 01 tháng 01 năm 2010 là 350 người; giá trị tài sản cố định của một công ty có vào ngày 02 tháng 02 năm 2011 là 5 tỷ đồng.

1-65

SỐ TUYỆT ĐỐI

- **Các loại số tuyệt đối**
- **Số tuyệt đối thời kỳ:** *Phản ánh mặt lượng của hiện tượng kinh tế xã hội trong một độ dài thời gian nhất định. Nó được hình thành thông qua sự tích lũy về mặt lượng của hiện tượng kinh tế xã hội trong suốt thời gian nghiên cứu và biểu hiện quy mô, khối lượng của hiện tượng nghiên cứu trong thời gian đó*
- **Ví dụ:** Doanh số bán của một công ty năm 2010 là 230 tỷ đồng

1-66

SỐ TƯƠNG ĐỐI

- **Số tương đối trong thống kê là chỉ tiêu biểu hiện quan hệ so sánh giữa hai mức độ của hiện tượng nghiên cứu.**
- **Ví dụ:**

$$\text{Mật độ dân số}_{(\text{người}/\text{km}^2)} = \frac{\text{Dân số trung bình}}{\text{Diện tích đất đai}}$$

$$\text{GDP trung bình}_{(\text{đầu người})} = \frac{\text{GDP}}{\text{Dân số trung bình}}$$

1-67

SỐ TƯƠNG ĐỐI

■ Các loại số tương đối và phương pháp tính

■ **Số tương đối động thái:** *Số tương đối động thái biểu hiện mối quan hệ so sánh giữa hai mức độ cùng loại của hiện tượng nghiên cứu ở hai thời kỳ (thời điểm) khác nhau, và được biểu hiện bằng số lần hay tỷ lệ phần trăm (%)*

$$t = \frac{\text{Mức độ kỳ báo cáo}}{\text{Mức độ kỳ gốc}} = \frac{y_1}{y_0}$$

1-68

SỐ TƯƠNG ĐỐI

■ **Ví dụ:** Số sản phẩm sản xuất của doanh nghiệp A qua hai năm là năm 2010 sản xuất được 1000 sản phẩm, năm 2011 sản xuất được 12000 sản phẩm.

1-69

SỐ TƯƠNG ĐỐI

Tổng quát, khi có một dãy số các mức độ thay đổi theo thời gian:

Thời gian (t_i)	t_1	t_2	t_3	...	t_n
Mức độ (y_i)	y_1	y_2	y_3	...	y_n

Số tương đối động thái liên hoàn $\frac{y_2}{y_1}, \frac{y_3}{y_2}, \dots, \frac{y_n}{y_{n-1}}$

Số tương đối động thái định gốc $\frac{y_2}{y_1}, \frac{y_3}{y_1}, \dots, \frac{y_n}{y_1}$

SỐ TƯƠNG ĐỐI

Ví dụ:

Năm	1999	2000	2001	2002
Doanh số bán (tỷ đồng)	10.00	12.00	14.00	15.84

Tính số tương đối động thái liên hoàn về doanh số của công ty?
 Tính tốc độ phát triển với kỳ gốc 1999?

SỐ TƯƠNG ĐỐI

■Tốc độ tăng (giảm) hay còn gọi là số tương đối tăng (giảm) là chỉ tiêu phản ánh mối quan hệ so sánh giữa mức độ chênh lệch tuyệt đối và mức độ kỳ gốc.

$$\text{Số tương đối tăng (giảm)} = \frac{y_1 - y_0}{y_0}$$

SỐ TƯƠNG ĐỐI

■Số tương đối kế hoạch
 ■Số tương đối nhiệm vụ kế hoạch (%)
 Là tỷ lệ so sánh giữa mức độ kế hoạch với mức độ thực tế của chỉ tiêu ấy ở kỳ gốc.

$$t_{NK} = \frac{y_k}{y_0} \times 100\%$$

1/3

SỐ TƯƠNG ĐỐI

- **Số tương đối kế hoạch**
- **Số tương đối hoàn thành kế hoạch (%)**
Là tỷ lệ so sánh giữa mức độ thực tế đạt được trong kỳ nghiên cứu với mức độ kế hoạch đặt ra cùng kỳ của một chỉ tiêu nào đó.

$$t_{HK} = \frac{y_1}{y_k}$$

1/4

SỐ TƯƠNG ĐỐI

- **Số tương đối kết cấu (%)** nói lên kết cấu tỷ trọng của mỗi bộ phận cấu thành tổng thể.

$$d_i = \frac{y_i}{\sum_{i=1}^n y_i} \times 100\%$$

1/5

SỐ TƯƠNG ĐỐI

- **Số tương đối cường độ** là kết quả so sánh mức độ của hai hiện tượng khác nhau nhưng có liên quan với nhau

$$\text{Số tương đối cường độ (\%)} = \frac{\text{Mức độ của hiện tượng này}}{\text{Mức độ của hiện tượng khác}}$$

- Ví dụ: $\frac{\text{Số lượng sản phẩm theo đầu người}}{\text{Khối lượng sản phẩm Tổng số dân}}$

SỐ TƯƠNG ĐỐI

- Số tương đối không gian** là kết quả so sánh giữa hai mức độ của một hiện tượng nhưng khác nhau về không gian.

Số tương đối không gian (%) = $\frac{\text{Mức độ của bộ phận này trong tổng thể}}{\text{Mức độ của bộ phận khác trong tổng thể}}$

ĐO LƯỜNG XU THẾ TRUNG TÂM

- Trung bình cộng (Arithmetic mean)**
Là mức độ đại diện điển hình cho 1 tiêu thức nào đó của tổng thể mà các đơn vị của tổng thể biểu hiện nhiều mức độ khác nhau
- Số trung bình cộng đơn giản** $x = \frac{\sum_{i=1}^n x_i}{n}$

ĐO LƯỜNG XU THẾ TRUNG TÂM

- Trung bình cộng (Arithmetic mean)**

ĐO LƯỜNG XU THẾ TRUNG TÂM

■ Số bình quân gia quyền (weighted mean)

$$\bar{X} = \frac{\sum_{i=1}^n x_i f_i}{\sum_{i=1}^n f_i}$$

ĐO LƯỜNG XU THẾ TRUNG TÂM

Nếu dãy số lượng biến có khoảng cách tổ, để tính toán cần tìm một trị số đại diện, gọi là **trị số giữa**

$$\text{Trị số giữa} = \frac{X_{\max} + X_{\min}}{2}$$

ĐO LƯỜNG XU THẾ TRUNG TÂM

Số trung bình điều hòa (harmonic mean)

Là số bình quân được tính từ các đại lượng nghịch đảo của các lượng biến

$$\bar{X} = \frac{\sum_{i=1}^n x_i f_i}{\sum_{i=1}^n \frac{x_i f_i}{x_i}}$$

1-82

ĐO LƯỜNG XU THỂ TRUNG TÂM

VD. Tình hình doanh số bán của 3 loại gạo tại một cửa hàng gạo như sau:

Loại gạo	Đơn giá(1000đ/kg)	Doanh thu (1000 đồng)
Loại 1	8	24.000
Loại 2	6	24.000
Loại 3	4	24.000

■ Tính giá trung bình 1kg gạo mà cửa hàng đã bán ra?

1-83

ĐO LƯỜNG XU THỂ TRUNG TÂM

■ **Mốt (Mode), ký hiệu M_0**
Là biểu hiện của một tiêu thức được gặp nhiều nhất trong tổng thể. Đối với một dãy số lượng biến, một là lượng biến có tần số lớn nhất.

1-84

ĐO LƯỜNG XU THỂ TRUNG TÂM

■ **Cách xác định một**

■ **Lượng biến rời rạc**, một là lượng biến có tần số lớn nhất.

1-85

ĐO LƯỜNG XU THẾ TRUNG TÂM

- **Cách xác định một**
- **Lượng biến rời rạc**, một là lượng biến có tần số lớn nhất.

1-86

ĐO LƯỜNG XU THẾ TRUNG TÂM

- **Cách xác định một**
- **Lượng biến có khoảng cách tổ**

Bước 1: Xác định tổ chứa một, là tổ có mật độ phân phối lớn nhất.

Mật độ phân phối là tỷ số giữa các tần số và trị số khoảng cách tổ tương ứng

$$\text{Mật độ phân phối} = \frac{\text{Tần số}}{\text{Trị số khoảng cách tổ}} = \frac{f_i}{d_i} (i = \overline{1, n})$$

1-87

ĐO LƯỜNG XU THẾ TRUNG TÂM

$$M_0 = x_{M_0(\min)} + d_{M_0} \frac{F_{M_0} - F_{M_0-1}}{(F_{M_0} - F_{M_0-1}) + (F_{M_0} - F_{M_0+1})}$$

- $x_{M_0(\min)}$: giới hạn dưới của khoảng cách tổ có Một
- d_{M_0} Trị số khoảng cách tổ có Một
- F_{M_0} Mật độ phân phối của tổ có Một
- F_{M_0-1} Mật độ phân phối của tổ đứng trước tổ có Một
- F_{M_0+1} Mật độ phân phối của tổ đứng sau tổ có Một

1-89

ĐO LƯỜNG XU THẾ TRUNG TÂM

VD: Năng suất lao động của công nhân trong một phân xưởng sau:

Phân tổ công nhân theo mức năng suất lao động (tạ/người)	Trị số Khoảng Cách Tổ (d _i)	Số công nhân (người f _i)	Mật Độ Phân Phối(f _i /d)
20-22	2	5	2.5
22-24	2	10	5
24-26	2	20	10
26-28	2	15	7.5
28-30	2	5	2.5

Tính Một về năng suất lao động?

25.32 tại người

1-89

ĐO LƯỜNG XU THẾ TRUNG TÂM

■ **Số trung vị (Median), ký hiệu M_e**
Là lượng biến của đơn vị tổng thể ở vị trí giữa trong dãy số lượng biến. Số trung vị chia dãy số lượng biến thành hai phần, mỗi phần có số đơn vị tổng thể bằng nhau

1-90

ĐO LƯỜNG XU THẾ TRUNG TÂM

■ **Số trung vị (Median), ký hiệu M_e**
 ■ **Đối với dãy số lượng biến rời rạc**
 - Trường hợp số đơn vị tổng thể lẻ ($n = 2m + 1$), số trung vị sẽ là lượng biến ở đơn vị thứ $m + 1$: $M_e = X_{m+1}$

Ví dụ: Có mức năng suất lao động của năm công nhân trong một tổ là 20,22,25,27,29 tạ/người. $M_e = 25$ tạ/ người

ĐO LƯỜNG XU THẾ TRUNG TÂM

- Số trung vị (Median), ký hiệu M_e
- Đối với dãy số lượng biến rời rạc
 - Trường hợp số đơn vị tổng thể chẵn ($n = 2m$). Số trung vị sẽ là lượng biến giữa lượng biến m và $m + 1$.

$$M_e = \frac{X_m + X_{m+1}}{2}$$

ĐO LƯỜNG XU THẾ TRUNG TÂM

- Số trung vị (Median), ký hiệu M_e
- Đối với dãy số lượng biến rời rạc
- Ví dụ: Có mức năng suất lao động của năm công nhân trong một tổ là:
20, 22, 25, 27, 29, 30 tạ/người.
- $M_e = 26$ tạ/ người

ĐO LƯỜNG XU THẾ TRUNG TÂM

■

$$M_e = XM_{e(\min)} + dM_e \frac{\sum \frac{f}{2} - S_{M_e-1}}{fM_e}$$

1-84

ĐO LƯỜNG XU THẾ TRUNG TÂM

-

1-85

ĐO LƯỜNG XU THẾ TRUNG TÂM

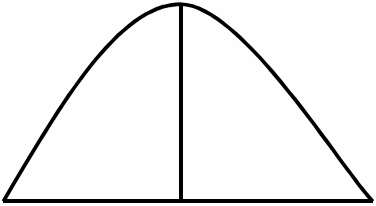
Ví dụ: Về mức lương của công nhân trong một phân xưởng của xí nghiệp X như sau:

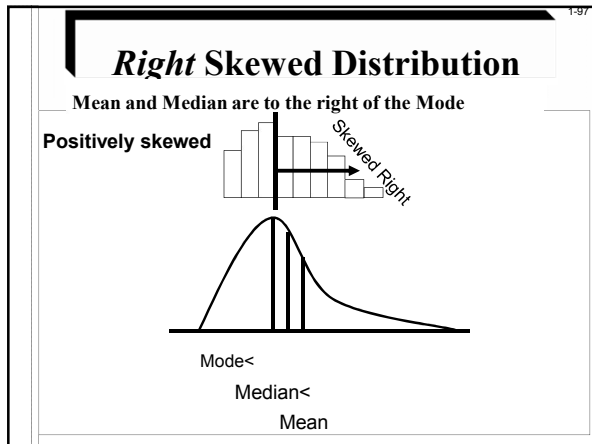
Mức lương một công nhân (triệu đồng / người)	Số công nhân (người)	Tần số Tích lũy
2.7 – 2.9	5	5
2.9 – 3.1	10	15
3.1 – 3.3	20	35
3.3 – 3.5	15	50
3.5 – 3.7	5	55

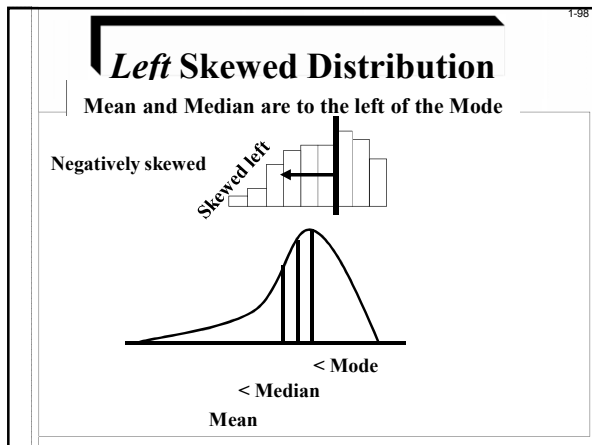
1-86

PHÂN PHỐI ĐỐI XỨNG

- zero skewness
mode = median = mean







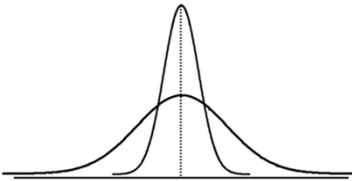
**NGUYÊN LÝ THỐNG KÊ
KINH TẾ**

NGÔ THÁI HƯNG
ĐH Tài Chính – Marketing
Email: ngothai.hung@gmail.com

1-100

ĐO LƯỜNG ĐỘ PHÂN TÁN

(MEASURES OF DISPERSION)



Same center,
different variation

1-101

ĐO LƯỜNG ĐỘ PHÂN TÁN

(MEASURES OF DISPERSION)

- **Độ biến thiên tiêu thức** là chỉ tiêu, dùng để đánh giá tính chất đồng đều của dãy số lượng biến của một tiêu thức nào đó và qua đó đánh giá tính chất đại biểu của số bình quân.
- **Ví dụ:** Mức năng suất lao động của công nhân ở hai tổ như sau: (tạ/ người)
 Tổ 1: 20 25 30 35 40
 Tổ 2: 28 29 30 31 32
 NSLD bình quân là 30 tạ/người, nhưng NSLD của công nhân trong từng tổ thì khác nhau.

1-102

ĐO LƯỜNG ĐỘ PHÂN TÁN

(MEASURES OF DISPERSION)

- **Khoảng biến thiên** (toàn cự) – **Range:** là hiệu số giữa lượng biến lớn nhất và lượng biến nhỏ nhất của tiêu thức nghiên cứu.

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

1-105

ĐO LƯỜNG ĐỘ PHÂN TÁN
(MEASURES OF DISPERSION)

■ **Độ lệch tuyệt đối bình quân** (Mean absolute deviation)

$$MD = \frac{\sum |x - \mu|}{N}$$

1-104

ĐO LƯỜNG ĐỘ PHÂN TÁN
(MEASURES OF DISPERSION)

■ **Phương sai (variance):** Là số bình quân học của bình phương các độ chênh lệch giữa các lượng biến với số bình quân số học của các biến đó.

$$s^2 = \frac{\sum (x - \bar{x})^2}{n - 1}$$

1-105

ĐO LƯỜNG ĐỘ PHÂN TÁN
(MEASURES OF DISPERSION)

■ **Độ lệch chuẩn (Standard deviation)** là căn bậc 2 của phương sai, hay nói cách khác là số bình quân toàn phương của các độ lệch giữa các lượng biến với số bình quân số học của chúng.

$$\sigma = \sqrt{\sigma^2}$$

1-106

ĐO LƯỜNG ĐỘ PHÂN TÁN

(MEASURES OF DISPERSION)

- Hệ số biến thiên(Coefficient of variation)**
là số tương đối, biểu hiện quan hệ so sánh độ lệch tuyệt đối bình quân (hoặc độ lệch chuẩn) với số bình quân học

$$CV = \frac{s}{\bar{x}} (100\%)$$

1-107

ĐO LƯỜNG ĐỘ PHÂN TÁN

(MEASURES OF DISPERSION)

Bell-Shaped Curve
...showing the relationship between σ and μ

The Empirical Rule

- If the data distribution is bell-shaped, the interval:
- $\mu \pm 1\sigma$ contains about 68% of the values in the population or the sample

Skewness

...is the measurement of the lack of symmetry of the distribution

...The coefficient of skewness can range from -3.00 up to +3.00

A value of 0 indicates a symmetric distribution.

It is computed as follows:

$$SK_1 = \frac{3(\text{Mean} - \text{Median})}{\sigma}$$

ĐO LƯỜNG ĐỘ PHÂN TÁN
(MEASURES OF DISPERSION)

■ **Phân vị thứ p** ($0 \leq p \leq 100$) trong một dãy số đã sắp xếp theo thứ tự tăng dần là một giá trị chia dãy số làm hai phần, phần một gồm p% số đơn vị có giá trị nhỏ hơn hoặc bằng giá trị của phân vị thứ p. Công thức xác định phân vị thứ p

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

ĐO LƯỜNG ĐỘ PHÂN TÁN
(MEASURES OF DISPERSION)

■ **TỨ PHÂN VỊ** (Quartiles)

Tứ phân vị chia dãy số thành 4 phần, mỗi phần có số đơn vị bằng nhau.

25%	25%	25%	25%
↑ Q1	↑ Q2	↑ Q3	

1-112

ĐO LƯỜNG ĐỘ PHÂN TÁN

(MEASURES OF DISPERSION)

Cách xác định tứ phân vị

- Tài liệu phân tổ không có khoảng cách tổ

Q_1 tứ phân vị thứ nhất là giá trị đứng ở vị trí $(n+1)/4$, là phân vị thứ 25.

Q_2 tứ phân vị thứ hai chính là trung vị Me đứng ở vị trí $(n+1)/2$, là phân vị thứ 50

Q_3 tứ phân vị thứ ba, là giá trị của đơn vị đứng ở vị trí thứ $3(n+1)/4$, là phân vị thứ 75

1-113

ĐO LƯỜNG ĐỘ PHÂN TÁN

(MEASURES OF DISPERSION)

Cách xác định tứ phân vị

- Tài liệu phân tổ có khoảng cách tổ

Tứ phân vị thứ nhất $Q_1 = X_{q_{\min}} + h_{q_1} \frac{\frac{1}{4} \sum f - S_{q_{i-1}}}{f_{q_1}}$

Tứ phân vị thứ 3 $Q_3 = X_{q_{\min}} + h_{q_3} \frac{\frac{3}{4} \sum f - S_{q_{i-1}}}{f_{q_3}}$

1-114

ĐO LƯỜNG ĐỘ PHÂN TÁN

(MEASURES OF DISPERSION)

■ Ví dụ:

Doanh thu (tr.đ)	Cửa hàng (f_i)	Tần số tích lũy
200-400	8	8
400-500	12	20
500-600	25	45
600-800	25	70
800-1000	9	79
Tổng	79	

■ Tính Q_1 và Q_3 ?

1-115

Box Plots

...is a graphical display, based on quartiles, that helps to picture a set of data

Minimum -- Q1 -- Median -- Q3 -- Maximum

1-116

Box Plots

Left-Skewed

Symmetric

Right-Skewed

1-117

Box Plots

- Below is a Box-and-Whisker plot for the following data:

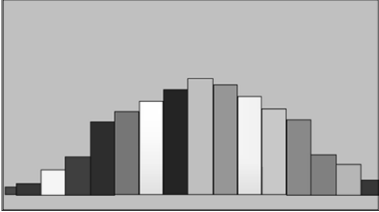
Min	Q1	Q2	Q3	Max
0	2	3	5	27

- This data is very right skewed, as the plot depicts

1-18

Kurtosis

Platykurtic - flat distribution

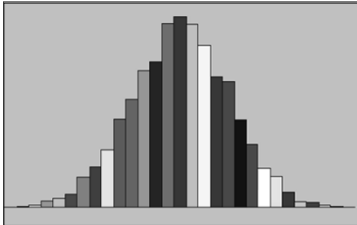


A histogram with approximately 15 bars. The bars are shaded in a gradient from dark grey on the left to light grey on the right. The distribution is relatively flat and wide, with a peak in the center that is not very tall.

1-19

Kurtosis

Mesokurtic - not too flat and not too peaked

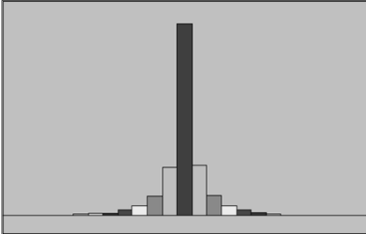


A histogram with approximately 15 bars. The bars are shaded in a gradient from dark grey on the left to light grey on the right. The distribution is bell-shaped and centered, with a peak in the center that is moderately tall.

1-20

Kurtosis

Leptokurtic - peaked distribution



A histogram with approximately 15 bars. The bars are shaded in a gradient from dark grey on the left to light grey on the right. The distribution is very narrow and tall, with a single, very prominent peak in the center.

NGUYÊN LÝ THỐNG KÊ KINH TẾ

NGÔ THÁI HƯNG
ĐH Tài Chính – Marketing
Email: ngothai.hung@gmail.com

CHƯƠNG 5

CHUỖI SỐ THỜI GIAN (Time series)

CHUỖI SỐ THỜI GIAN

■ KHÁI NIỆM

■ *Dãy số thời gian là một phương pháp phân tích thống kê được sử dụng khá phổ biến, nhằm nghiên cứu các đặc điểm, bản chất xu hướng và tính quy luật về sự phát triển của hiện tượng thường xuyên biến động theo thời gian.*

1-124

CHUỖI SỐ THỜI GIAN

■ Dãy số thời gian có 2 thành phần : thời gian: (t_i) và trị số của chỉ tiêu (y_i): giá trị của hiện tượng nghiên cứu

t_i	t_1	t_2	t_n
y_i	y_1	y_2	y_n

1-125

CHUỖI SỐ THỜI GIAN

Ý nghĩa: Qua dãy số thời gian ta có thể phân tích được :

- ✓ Sự biến động của hiện tượng qua thời gian
- ✓ Sự phát triển của hiện tượng
- ✓ Xu hướng phát triển cơ bản của hiện tượng
- ✓ Quy luật phát triển của hiện tượng

1-126

CHUỖI SỐ THỜI GIAN

■ **Dãy số thời kỳ** là dãy số biểu hiện các mức độ của chỉ tiêu ở từng thời kỳ

■ **Ví dụ:** Giá trị sản xuất công nghiệp của một xí nghiệp A (đơn vị tính: triệu đồng)

Năm	2005	2006	2007	2008	2009	2010
GTSX	31.2	35.4	39.47	40.23	46.89	50.12

1-127

CHUỖI SỐ THỜI GIAN

- **Dãy số thời điểm** là dãy số biểu hiện các mức độ của chỉ tiêu ở từng thời điểm nhất định.
- **Ví dụ:** Giá trị hàng nhập khẩu của công ty A. Kiểm kê vào ngày 1 hàng tháng (đơn vị tính: triệu đồng)

Ngày	1/1	1/2	1/3	1/4
Giá trị hàng nhập khẩu	2563	2653	2789	2983

1-128

CHUỖI SỐ THỜI GIAN

- **CÁC CHỈ TIÊU PHÂN TÍCH DSTG**
- **Mức độ bình quân theo thời gian**
- Đối với dãy số thời kỳ $\bar{y} = \frac{\sum_{i=1}^n y_i}{n}$
- Đối với dãy số thời điểm $\bar{y} = \frac{\frac{y_1}{2} + y_2 + y_3 + \dots + y_{n-1} + \frac{y_n}{2}}{n-1}$

1-129

CHUỖI SỐ THỜI GIAN

- **CÁC CHỈ TIÊU PHÂN TÍCH DSTG**
- **Mức độ bình quân theo thời gian**
- Trường hợp dãy số thời điểm có khoảng cách thời gian không đều nhau $\bar{y} = \frac{\sum y_i t_i}{\sum t_i}$

1-130

CHUỖI SỐ THỜI GIAN

- **Ví dụ:** Có tài liệu về số công nhân trong một phân xưởng như sau:
- Từ đầu tháng 6 có 20 người, ngày 10 – 6 bổ sung thêm 4 người, ngày 25 – 6 bổ sung thêm 3 máy, ngày 30 – 6 một công nhân xin nghỉ.

Thời gian	Khoảng cách thời gian (số ngày)	Số công nhân
Từ 1/6 đến 9/6	9	20
Từ 10/6 đến 24/6	15	24
Từ 25/6 đến 29/6	5	27
Ngày 30/6	1	26

1-131

CHUỖI SỐ THỜI GIAN

- **Lượng tăng (giảm) tuyệt đối** là chỉ tiêu thể hiện sự thay đổi về giá trị tuyệt đối của hiện tượng giữa hai thời kỳ hoặc thời điểm.

Lượng tăng (giảm) tuyệt đối liên hoàn:

$$\delta_i = y_i - y_{i-1} \quad i = 2, 3, \dots$$

Lượng tăng (giảm) tuyệt đối định gốc

$$\Delta_i = y_i - y_1 \quad i = 2, 3, \dots \implies \Delta_n = \sum_{i=2}^n \delta_i$$

1-132

CHUỖI SỐ THỜI GIAN

Lượng tăng (giảm) tuyệt đối trung bình

$$\bar{\delta} = \frac{\sum_{i=2}^n \delta_i}{n-1} = \frac{\Delta_n}{n-1} = \frac{y_n - y_1}{n-1}$$

1-133

CHUỖI SỐ THỜI GIAN

■ **Tốc độ phát triển**

Tốc độ phát triển liên hoàn $t_i = \frac{y_i}{y_{i-1}} \quad i=2,3,\dots$

Tốc độ phát triển định góc $T_i = \frac{y_i}{y_1} \quad i=2,3,\dots$

⇒ $\prod_{i=2}^n t_i = T_n \quad \frac{T_i}{T_{i-1}} = t_i \quad i=2,3,\dots$

1-134

CHUỖI SỐ THỜI GIAN

Tốc độ phát triển trung bình
là chỉ tiêu thể hiện nhịp độ phát triển đại diện của
hiện tượng trong suốt thời kỳ nghiên cứu

$$\bar{t} = \sqrt[n-1]{\prod_{i=2}^n t_i} = \sqrt[n-1]{\frac{y_n}{y_1}}$$

Còn gọi là số trung bình nhân (Geometric Mean)

1-135

CHUỖI SỐ THỜI GIAN

■ **Tốc độ tăng (giảm)** là chỉ tiêu phản ánh mức độ
của hiện tượng giữa 2 thời gian nghiên cứu đã
tăng (giảm) bao nhiêu lần (%)

Tốc độ tăng (giảm) liên hoàn

$$a_i = \frac{y_i - y_{i-1}}{y_{i-1}} = \frac{\delta_i}{y_{i-1}} = t_i - 1 \quad (i = \overline{2, n})$$

Tốc độ tăng (giảm) định góc

$$A_i = \frac{y_i - y_1}{y_1} = \frac{\Delta_i}{y_1} = T_i - 1 \quad (i = \overline{2, n})$$

1-136

CHUỖI SỐ THỜI GIAN

Tốc độ tăng (giảm) trung bình

$$\bar{a} = \bar{t} - 1$$

Tính theo số lần

$$\bar{a} = \bar{t} - 100$$

Tính theo số phần trăm

1-137

CHUỖI SỐ THỜI GIAN

■ *Giá trị tuyệt đối của 1% tăng (giảm) phản ánh 1% tăng (giảm) của 2 thời kỳ đứng liền nhau của hiện tượng nghiên cứu tương ứng với một lượng giá trị tuyệt đối là bao nhiêu*

$$g_i = \frac{\delta_i}{a_i(\%)} = \frac{y_i - y_{i-1}}{\frac{y_i - y_{i-1}}{100} \cdot y_{i-1}} = \frac{y_i - y_{i-1}}{y_{i-1} \cdot 100}$$

1-138

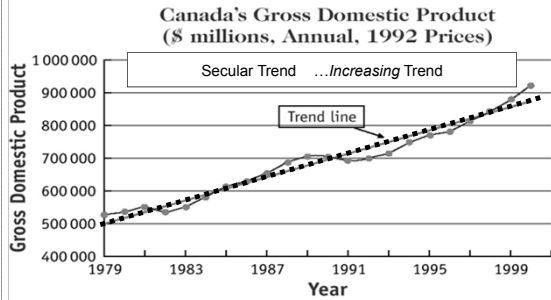
MỘT SỐ PHƯƠNG PHÁP BIỂU HIỆN XU HƯỚNG BIẾN ĐỘNG CƠ BẢN CỦA HIỆN TƯỢNG

Year	GDP
1979	527703
1980	535007
1981	551305
1982	595113
1983	549843
1984	581038
1985	612416
1986	626575
1987	654390
1988	696176
1989	703677
1990	705454
1991	692247
1992	696544
1993	714593
1994	748350
1995	759892
1996	790916
1997	815013
1998	842002
1999	880254
2000	921495

Secular Trend (Xu thế)

MỘT SỐ PHƯƠNG PHÁP BIỂU HIỆN XU HƯỚNG BIẾN ĐỘNG CƠ BẢN CỦA HIỆN TƯỢNG

1-139



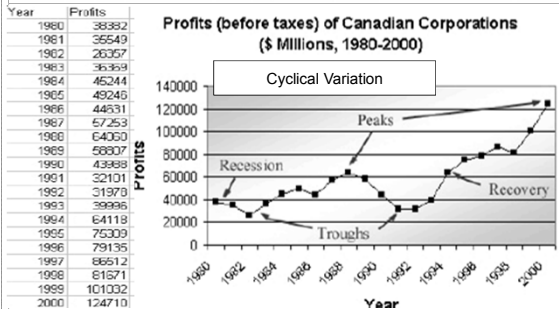
MỘT SỐ PHƯƠNG PHÁP BIỂU HIỆN XU HƯỚNG BIẾN ĐỘNG CƠ BẢN CỦA HIỆN TƯỢNG

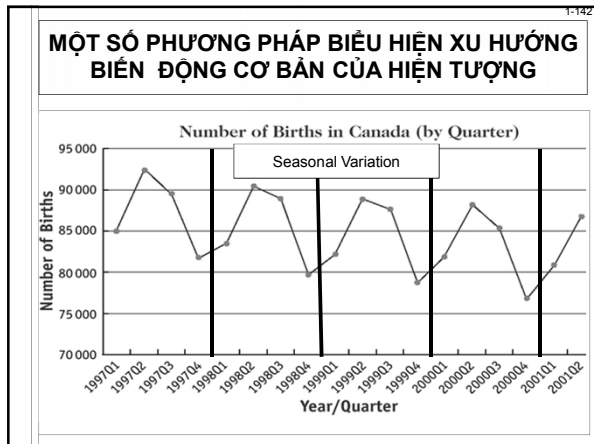
1-140



MỘT SỐ PHƯƠNG PHÁP BIỂU HIỆN XU HƯỚNG BIẾN ĐỘNG CƠ BẢN CỦA HIỆN TƯỢNG

1-141





1-143

MỘT SỐ PHƯƠNG PHÁP BIỂU HIỆN XU HƯỚNG BIẾN ĐỘNG CƠ BẢN CỦA HIỆN TƯỢNG

Phương pháp mở rộng khoảng cách thời gian: Được sử dụng khi 1 dãy số thời gian có khoảng cách quá ngắn, hoặc có nhiều mức độ làm cho ta khó thấy được xu thế phát triển cơ bản của hiện tượng.

Ví dụ: Có tài liệu nói về năng suất thu hoạch loại cây trồng A

Năm	Năng suất thu hoạch (tạ/ha)	Năm	Năng suất thu hoạch (tạ/ha)	Năm	Năng suất thu hoạch (tạ/ha)
1995	1.92	2000	1.98	2005	3.23
1996	1.28	2001	3.06	2006	4.57
1997	1.11	2002	3.18	2007	5.05
1998	1.09	2003	2.99	2008	5.78
1999	1.58	2004	2.89	2009	5.13

1-144

MỘT SỐ PHƯƠNG PHÁP BIỂU HIỆN XU HƯỚNG BIẾN ĐỘNG CƠ BẢN CỦA HIỆN TƯỢNG

Rõ ràng năng suất thu hoạch qua các năm tăng giảm bất thường, nhưng xu thế biến động cơ bản của nó không ngừng tăng lên.

Thời kỳ	1995-1999	2000-2004	2005-2009
Tổng năng suất thu hoạch 5 năm	6.98	14.1	23.73
Năng suất thu hoạch bình quân	1.396	2.820	4.752

1-145

**MỘT SỐ PHƯƠNG PHÁP BIỂU HIỆN XU HƯỚNG
BIẾN ĐỘNG CƠ BẢN CỦA HIỆN TƯỢNG**

■ **Phương pháp số bình quân trượt** (di động) là số bình quân cộng của một nhóm nhất định các mức độ của dãy số được tính bằng cách lần lượt loại trừ dần các mức độ đầu, đồng thời thêm vào các mức độ tiếp theo, sao cho tổng lượng các mức độ tham gia tính số bình quân cộng không thay đổi

Năm	Quý	Giá trị sản xuất (y.)	Số bình quân trượt lần thứ nhất	Số bình quân trượt lần thứ hai	
1991	I	3.18	-	-	Số bình quân trượt (nhóm 4 mức độ)
	II	4.15	3.95	-	
	III	4.56	3.92	3.94	
	IV	3.91	3.92	3.92	
					$y_1 = \frac{3.18+4.15+4.56+3.91}{4} = 3.9$
1992	I	3.07	4.03	3.98	Số bình quân trượt (lần 2 theo nhóm 2 mức độ)
	II	4.14	4.12	4.08	
	III	4.98	4.04	4.08	
	IV	3.29	4.35	4.20	
					$y_2 = \frac{4.15+4.56+3.91+3.07}{4} = 3.9$
1993	I	3.76	4.75	4.55	Số bình quân trượt (lần 2 theo nhóm 2 mức độ)
	II	5.36	5.45	5.10	
	III	6.58	5.88	5.67	
	IV	6.08	6.56	6.22	
					$y_1 = \frac{3.95 + 3.92}{2} = 3.94$
1994	I	5.19	7.24	6.90	Số bình quân trượt (lần 2 theo nhóm 2 mức độ)
	II	8.07	7.90	7.57	
	III	9.30	8.35	8.13	
	IV	8.73	8.97	8.66	
					$y_2 = \frac{3.92 + 3.92}{2} = 3.92$
1995	I	7.30	9.61	9.29	Số bình quân trượt (lần 2 theo nhóm 2 mức độ)
	II	10.53	9.68	9.65	
	III	11.89	-	-	
	IV	9.00	-	-	
					$y_3 = \frac{3.92 + 4.03}{2} = 3.98$

1-147

**MỘT SỐ PHƯƠNG PHÁP BIỂU HIỆN XU HƯỚNG
BIẾN ĐỘNG CƠ BẢN CỦA HIỆN TƯỢNG**

■ **Phương pháp thể hiện xu hướng bằng hàm số Phương trình đường thẳng** $\bar{y}_t = a_0 + a_1 t$
 Trong đó: a_0, a_1 được xác định

$$\begin{cases} \sum y = na_0 + a_1 \sum t \\ \sum yt = a_0 \sum t + a_1 \sum t^2 \end{cases}$$

1-148

MỘT SỐ PHƯƠNG PHÁP BIỂU HIỆN XU HƯỚNG BIẾN ĐỘNG CƠ BẢN CỦA HIỆN TƯỢNG

■ Năm	Năng suất Bình quân (tạ/ha) y	Phần tính toán			
		Thứ tự thời gian (t)	t ²	ty	
2001	30	1	1	30	30.4
2002	32	2	4	64	31.2
2003	31	3	9	93	32.0
2004	34	4	16	136	32.8
2005	33	5	25	165	33.6
Cộng	160	15	55	488	

$$\begin{cases} 160 = 5a_0 + 15a_1 \\ 488 = 15a_0 + 55a_1 \end{cases} \quad \bar{y}_t = 29.6 + 0.8t$$

1-149

MỘT SỐ PHƯƠNG PHÁP BIỂU HIỆN XU HƯỚNG BIẾN ĐỘNG CƠ BẢN CỦA HIỆN TƯỢNG

■ Phân tích biến động thời vụ là phương pháp nghiên cứu xác định sự biến động một cách có quy luật vào những thời kỳ nhất định trong vòng một năm của hiện tượng kinh tế xã hội.

■ $I_s = \frac{\bar{y}_i}{\bar{y}_0}$ I_s Chỉ số thời vụ của thời gian t
 \bar{y}_i Số bình quân các mức độ của các thời gian cùng tên i
 \bar{y}_0 Số bình quân của tất cả các mức độ trong dãy số

1-150

MỘT SỐ PHƯƠNG PHÁP BIỂU HIỆN XU HƯỚNG BIẾN ĐỘNG CƠ BẢN CỦA HIỆN TƯỢNG

■ Có tài liệu về mức tiêu thụ hàng hóa "X" ở một địa phương trong 3 năm.

Năm Tháng	Mức tiêu thụ hàng hóa X (y _i - triệu đồng)			\bar{y}_i	$\frac{\bar{y}_i}{\bar{y}_0} \times 100$
	2000	2001	2002		
1	1495	1500	1490	1495	62.9
2	1461	1490	1480	1477	62.2
3	1533	1599	1604	1578	66.4
4	1922	2210	2005	2046	86.1
5	2746	2804	2745	2765	116.4
6	3289	3282	3250	3274	137.8
7	3523	3620	3700	3614	152.1
8	3330	3300	3215	3282	138.2
9	2597	2604	2599	2597	109.3
10	2249	2205	2304	2253	94.8
11	2144	2200	2190	2178	91.7
12	1983	1889	1950	1941	81.7
Tổng cả năm	28272	28703	28523	2375	

MỘT SỐ PHƯƠNG PHÁP BIỂU HIỆN XU HƯỚNG
BIẾN ĐỘNG CƠ BẢN CỦA HIỆN TƯỢNG

- DỰ BÁO THỐNG KÊ
- Mô hình dự báo tốc độ phát triển bình quân

$$\hat{y}_{n+L} = y_n \times \bar{t}^L$$

y_{n+L} : Mức độ dự báo vào thời gian (n+L)
 y_n : Mức độ cuối cùng của dãy số thời gian
 \bar{t} : Tốc độ phát triển bình quân
L: tầm xa dự báo

MỘT SỐ PHƯƠNG PHÁP BIỂU HIỆN XU HƯỚNG
BIẾN ĐỘNG CƠ BẢN CỦA HIỆN TƯỢNG

- DỰ BÁO THỐNG KÊ
- Mô hình dự báo theo mức độ tăng, giảm tuyệt đối bình quân

$$\hat{y}_{n+L} = y_n + \bar{\delta}_y \cdot L$$

$\bar{\delta}_y = \frac{y_n - y_1}{n - 1}$: Lượng tăng giảm tuyệt đối bình quân
L: Tầm xa dự báo

MỘT SỐ PHƯƠNG PHÁP BIỂU HIỆN XU HƯỚNG
BIẾN ĐỘNG CƠ BẢN CỦA HIỆN TƯỢNG

- DỰ BÁO THỐNG KÊ
- Mô hình dự báo theo mô hình hồi qui đường thẳng:

$$\bar{y}_t = a_0 + a_1 t$$

NGUYÊN LÝ THỐNG KÊ KINH TẾ

NGÔ THÁI HƯNG
ĐH Tài Chính – Marketing
Email: ngothai.hung@gmail.com

CHƯƠNG 6

CHỈ SỐ Index Numbers


KHÁI NIỆM CHỈ SỐ

- *Chỉ số thống kê là chỉ tiêu số tương đối, biểu hiện mối quan hệ so sánh (không gian, thời gian và theo kế hoạch) giữa các mức độ của cùng một hiện tượng kinh tế.*
- *Chỉ số được dùng để nghiên cứu tình hình biến động của những hiện tượng kinh tế phức tạp, không đồng chất bao gồm nhiều yếu tố không thể tổng hợp trực tiếp với nhau được*

1-157

KHÁI NIỆM CHỈ SỐ

- PHÂN LOẠI CHỈ SỐ
 - Chỉ số cá thể
 - Chỉ số tổng hợp



1-158

TÍNH CHỈ SỐ

- Chỉ số cá thể: là chỉ số được lập cho từng yếu tố, từng phần tử trong tổng thể hiện tượng nghiên cứu.
- Chỉ số giá cả cá thể:
$$i_p = \frac{p_1}{p_0} \times 100\%$$
- Chỉ số cá thể khối lượng:
$$i_q = \frac{q_1}{q_0} \times 100\%$$

1-159

TÍNH CHỈ SỐ

- Chỉ số tổng hợp: Chỉ số tổng hợp là loại chỉ số chỉ nghiên cứu sự biến động về một chỉ tiêu nào đó của nhiều đơn vị, nhiều phần tử của hiện tượng phức tạp

1-160

TÍNH CHỈ SỐ

■ Chỉ số tổng hợp giá cả:

$$I_p = \frac{\sum_{i=1}^n P_{i(1)} Q_{i(0)}}{\sum_{i=1}^n P_{i(0)} Q_{i(0)}} \quad \text{Laspeyres...}$$

$$I_p = \frac{\sum_{i=1}^n P_{i(1)} Q_{i(1)}}{\sum_{i=1}^n P_{i(0)} Q_{i(1)}} \quad \text{Paasche ...}$$

1-161

TÍNH CHỈ SỐ

■ Chỉ số tổng hợp khối lượng:

$$I_q = \frac{\sum_{i=1}^n Q_{i(1)} P_{i(0)}}{\sum_{i=1}^n Q_{i(0)} P_{i(0)}} \cdot 100\% \quad \text{Laspeyres...}$$

$$I_q = \frac{\sum_{i=1}^n Q_{i(1)} P_{i(1)}}{\sum_{i=1}^n Q_{i(0)} P_{i(1)}} \cdot 100\% \quad \text{Paasche ...}$$

1-162

CHỈ SỐ KHÔNG GIAN

■ Chỉ số tổng hợp khối lượng không gian

Công thức chỉ số khối lượng tổng hợp ở hai thị trường A và B

$$I_{q(A/B)} = \frac{\sum Q_A \bar{p}}{\sum Q_B \bar{p}}, \quad \bar{p} = \frac{P_A Q_A + P_B Q_B}{Q_A + Q_B}$$

CHỈ SỐ KHÔNG GIAN

■ Chỉ số tổng hợp giá cả không gian

Công thức chỉ số giá cả ở hai thị trường A và B:

$$I_{P(A/B)} = \frac{\sum P_A Q}{\sum P_B Q}, Q = q_A + q_B$$

Ví dụ: Tài liệu giá cả và khối lượng hàng tiêu thụ của 3 mặt hàng tại hai thành phố X và Y trong cùng một chu kỳ như sau:

CHỈ SỐ KHÔNG GIAN

Loại hàng hóa	Thành phố X		Thành phố Y	
	Giá đơn vị (1000đ)	Lượng hàng tiêu thụ (tấn)	Giá đơn vị (1000đ)	Lượng hàng tiêu thụ (tấn)
A	5.0	250	4.8	262
B	4.6	430	4.9	392
C	6.9	187	6.8	213

- Tính chỉ số tổng hợp khối lượng hàng hóa tiêu thụ thành phố X so với thành phố Y?
- Tính chỉ số giá tổng hợp giá cả hàng hóa tiêu thụ thành phố X so với thành phố Y?

HỆ THỐNG CHỈ SỐ

- Là một dãy các chỉ số có mối liên hệ với nhau hợp thành một đẳng thức nhất định. Cơ sở để xây dựng một hệ thống chỉ số là dựa vào các phương trình kinh tế

HỆ THỐNG CHỈ SỐ

■ Tác dụng của hệ thống chỉ số:

- Xác định được vai trò và ảnh hưởng biến động của mỗi nhân tố với biến động của hiện tượng phức tạp
- Tính ra một chỉ số chưa biết khi đã biết các chỉ số còn lại trong hệ thống đó

HỆ THỐNG CHỈ SỐ

Ví dụ: Phân tích tổng mức hàng hóa tiêu thụ biến động qua hai kỳ nghiên cứu trong mối quan hệ giữa hai nhân tố: giá cả và khối lượng hàng hóa tiêu thụ.

$$I_{pq} = I_p \times I_q$$
$$\frac{\sum p_1 q_1}{\sum p_0 q_0} = \frac{\sum p_1 q_1}{\sum p_0 q_1} \times \frac{\sum p_0 q_1}{\sum p_0 q_0}$$

HỆ THỐNG CHỈ SỐ

VẬN DỤNG PHƯƠNG PHÁP C/S ĐỂ PHÂN TÍCH BIẾN ĐỘNG CỦA CHỈ TIÊU TB VÀ TỔNG LƯỢNG BIẾN CỦA TIÊU THỨC:

PHÂN TÍCH BIẾN ĐỘNG CỦA CHỈ TIÊU TB:

CÁC KÝ HIỆU SỬ DỤNG:

x_1, x_0 : LƯỢNG BIẾN CỦA TIÊU THỨC KỲ NGHIÊN CỨU VÀ KỲ GỐC.

\bar{x}_1, \bar{x}_0 : SỐ TB KỲ NGHIÊN CỨU VÀ KỲ GỐC.

f_1, f_0 : SỐ ĐƠN VỊ TỔNG THỂ KỲ NGHIÊN CỨU VÀ KỲ GỐC.

TRONG ĐÓ:

$$\bar{x}_1 = \frac{\sum x_1 f_1}{\sum f_1} \quad \bar{x}_{01} = \frac{\sum x_0 f_1}{\sum f_1} \quad \bar{x}_0 = \frac{\sum x_0 f_0}{\sum f_0}$$

1-169

HỆ THỐNG CHỈ SỐ

TA CÓ HTCS:

$$\bar{I}_x = I_x \times I_f / \sum f$$

$$\frac{\sum x_1 f_1}{\sum f_1} = \frac{\sum x_1 f_1}{\sum x_0 f_1} \times \frac{\sum x_0 f_1}{\sum f_0}$$

NẾU ĐẶT $d_1 = \frac{f_1}{\sum f_1}$ và $d_0 = \frac{f_0}{\sum f_0}$

THÌ: $\frac{\sum x_1 d_1}{\sum x_0 d_0} = \frac{\sum x_1 d_1}{\sum x_0 d_1} \times \frac{\sum x_0 d_1}{\sum x_0 d_0}$

1-170

HỆ THỐNG CHỈ SỐ

$$\frac{\bar{X}_1}{\bar{X}_0} = \frac{\bar{X}_1}{\bar{X}_{01}} \times \frac{\bar{X}_{01}}{\bar{X}_0}$$

(1) (2) (3)

(1): CS CẤU THÀNH KHẢ BIẾN.
 (2): CS CẤU THÀNH CỐ ĐỊNH.
 (3): CS ẢNH HƯỞNG KẾT CẤU.

1-171

HỆ THỐNG CHỈ SỐ

VÍ DỤ: 1 XÍ NGHIỆP CÓ 3 PHÂN XƯỞNG CÙNG SẢN XUẤT 1 LOẠI SẢN PHẨM A.

PX	KỶ GỐC		KỶ BÁO CÁO	
	SẢN LƯỢNG (cái)(q ₀)	Z ĐƠN VỊ (ngđ)(Z ₀)	SẢN LƯỢNG (cái)(q ₁)	Z ĐƠN VỊ (ngđ)(Z)
A	1000	10	8000	9
B	2500	12	3000	11,5
C	4500	13	1000	12,5
Σ	8000		12000	

YÊU CẦU:
 a/ PHÂN TÍCH SỰ BIẾN ĐỘNG CỦA Z TB DO ẢNH HƯỞNG BỞI CÁC NHÂN TỐ CÓ LIÊN QUAN.
 b/ PHÂN TÍCH SỰ BIẾN ĐỘNG CỦA TỔNG CHI PHÍ SX CÓ LIÊN QUAN ĐẾN BIẾN ĐỘNG CỦA Z TB.

1-172

HỆ THỐNG CHỈ SỐ

GIẢI:

a/ $\bar{z}_1 = \frac{\sum z_1 q_1}{\sum q_1} = \frac{119000}{12000} = 9,92 \text{ngđ}$

$\bar{z}_0 = \frac{\sum z_0 q_0}{\sum q_0} = \frac{98500}{8000} = 12,31 \text{ngđ}$

$\bar{z}_{01} = \frac{\sum z_0 q_1}{\sum q_1} = \frac{129000}{12000} = 10,75 \text{ngđ}$

TA CÓ HTCS:

$$\frac{\bar{z}_1}{\bar{z}_0} = \frac{\bar{z}_1}{\bar{z}_{01}} \times \frac{\bar{z}_{01}}{\bar{z}_0}$$

THAY SỐ VÀO:

$$\frac{9,92}{12,31} = \frac{9,92}{10,75} \times \frac{10,75}{12,31}$$

$$0,806 = 0,9228 \times 0,873$$

$$80,6\% = 92,28\% \times 87,3\%$$

$$(-19,4\%) \quad (-7,72) \quad (-12,7)$$

1-173

HỆ THỐNG CHỈ SỐ

PHÂN TÍCH BIẾN ĐỘNG CỦA TỔNG LƯỢNG BIẾN TIÊU THỨC CÓ SỬ DỤNG CHỈ TIÊU TB:
 TRONG NHIỀU TRƯỜNG HỢP CHỈ TIÊU TB CÓ QUAN HỆ VỚI TỔNG LƯỢNG BIẾN TIÊU THỨC.

VÍ DỤ:
TỔNG SẢN PHẨM = NSLD TB 1 CN x SỐ CN
TỔNG CHI PHÍ SX = Z TB 1 ĐƠN VỊ SP x SỐ SP SX

TỔNG QUÁT: $M = \bar{x} \times \sum f$
 SỬ DỤNG HTCS TA CÓ:

$$I_M = I_{\bar{x}} \times I_{\sum f}$$

$$\frac{\bar{x}_1 \sum f_1}{x_0 \sum f_0} = \frac{M_1}{M_0} = \frac{\bar{x}_1}{x_0} \times \frac{\sum f_1}{\sum f_0}$$

1-174

HỆ THỐNG CHỈ SỐ

GIẢI b:
 THEO ĐỀ BÀI TA CÓ HTCS:

$$\frac{z_1 \sum q_1}{z_0 \sum q_0} = \frac{M_1}{M_0} = \frac{z_1}{z_0} \times \frac{\sum q_1}{\sum q_0}$$

$$\frac{9,92 \times 12000}{12,31 \times 8000} = \frac{9,92}{12,31} \times \frac{12000}{8000}$$

$$1,2088 = 0,806 \times 1,5$$

$$(+20,88\%) \quad (-19,4\%) \quad (+50\%)$$

SỐ TUYỆT ĐỐI TĂNG (GIẢM):

$$M_1 - M_0 = (\bar{z}_1 - \bar{z}_0) \sum q_1 + (\sum q_1 - \sum q_0) \bar{z}_0$$

$$(9,92 \times 12000) - (12,31 \times 8000) = (9,92 - 12,31) 12000$$

$$+ (12000 - 8000) 12,31$$

$$20560 \text{ngđ} = (-28680 \text{ngđ}) + (49240 \text{ngđ})$$

HỆ THỐNG CHỈ SỐ

- Giá cả * khối lượng SP sản xuất = Giá trị SX
- Giá cả * lượng hàng hoá tiêu thụ = Mức tiêu thụ hàng hoá
- Năng suất lao động b/q * Số CN = Giá trị SX hoặc lượng SP SX
- Giá thành 1 đ/v SP * số lượng SP = Chi phí SX
- Năng suất b/q lúa 1 ha * Diện tích = Sản lượng lúa

MỘT SỐ CHỈ SỐ THƯỜNG GẶP TRONG THỰC TẾ

■ Chỉ số giá tiêu dùng CPI

Chỉ số giá tiêu dùng ở nước ta hiện nay được tính theo công thức Laspeyres

$$I_p = \frac{\sum p_t q_{2000}}{\sum p_{2000} q_{2000}} \times 100\%$$

Hiện nay, cơ cấu chi tiêu của các hộ gia đình năm 2000 được chọn làm quyền số để tính CPI.

MỘT SỐ CHỈ SỐ THƯỜNG GẶP TRONG THỰC TẾ

■ Chỉ số chứng khoán VN- Index

$$I_{VN-Index} = \frac{\text{Tổng giá trị của các cổ phiếu niêm yết hiện tại}}{\text{Tổng giá trị của các cổ phiếu niêm yết cơ sở}} \cdot 100$$

Ngày được chọn làm cơ sở là 28/7/2000 hay được gọi là kỳ gốc

NGUYÊN LÝ THỐNG KÊ KINH TẾ

NGÔ THÁI HƯNG
ĐH Tài Chính – Marketing
Email: ngothai.hung@gmail.com

CHƯƠNG 8



KIỂM ĐỊNH GIẢ THUYẾT THỐNG KÊ

NỘI DUNG

- KIỂM ĐỊNH TRUNG BÌNH
- KIỂM ĐỊNH PHI THAM SỐ
- KIỂM ĐỊNH CHI BÌNH PHƯƠNG

1-181
XÂY DỰNG GIÁ THUYẾT KHÔNG VÀ GIÁ THUYẾT THAY THẾ

KIỂM ĐỊNH GIÁ THUYẾT NGHIÊN CỨU

Giả thuyết nghiên cứu nên được phát biểu dưới dạng giả thuyết thay thế.

1-182
XÂY DỰNG GIÁ THUYẾT KHÔNG VÀ GIÁ THUYẾT THAY THẾ

KIỂM ĐỊNH TÍNH CHÍNH XÁC CỦA CỬA LỜI PHÁT BIỂU

*Trong các tình huống liên quan đến việc đánh giá tính chính xác của một lời phát biểu hay tuyên bố nào đó, **giả thuyết không** thường được dựa trên giả định là lời phát biểu đó đúng.*

1-183
XÂY DỰNG GIÁ THUYẾT KHÔNG VÀ GIÁ THUYẾT THAY THẾ

KIỂM ĐỊNH TRONG TÌNH HUỐNG RA QUYẾT ĐỊNH

Loại tình huống này xảy ra khi người ta quyết định phải chọn hai loại hành động, một tương ứng với giả thuyết không và một tương ứng với giả thuyết thay thế

1-184

SAI LÂM LOẠI I & II

GIẢ THUYẾT KẾT LUẬN	GIẢ THUYẾT H_0 ĐÚNG	GIẢ THUYẾT H_1 ĐÚNG
CHẤP NHẬN H_0	KẾT LUẬN ĐÚNG	SAI LÂM LOẠI II
BÁC BỎ H_0	SAI LÂM LOẠI I	KẾT LUẬN ĐÚNG

1-185

SAI LÂM LOẠI I & II

Xác suất xảy ra sai lầm loại I được ký hiệu là α :

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

α : gọi là mức ý nghĩa (the level of significance)

1-186

p- Giá trị

p-giá trị (hoặc mức ý nghĩa đạt được) là mức ý nghĩa nhỏ nhất tại đó giả thuyết H_0 bị bác bỏ


$$p\text{-value} = \begin{cases} P(TS < a \mid H_0), & \text{if lower tail test} \\ P(TS > a \mid H_0), & \text{if upper tail test} \\ P(|TS| > a \mid H_0), & \text{if two tail test} \end{cases}$$

1-187

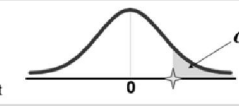
MỨC Ý NGHĨA & MIỀN BÁC BỎ

Mức ý nghĩa = α ★ Giá trị tới hạn

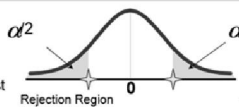
$H_0: \mu \geq 3$
 $H_A: \mu < 3$

Lower tail test 

$H_0: \mu \leq 3$
 $H_A: \mu > 3$

Upper tail test  Miền bác bỏ

$H_0: \mu = 3$
 $H_A: \mu \neq 3$

Two tailed test  Rejection Region

1-188

KIỂM ĐỊNH GIẢ THUYẾT

Hãy nhớ

Bước 1: Phát biểu giả thuyết H_0 và H_a

Bước 2: Chọn mức ý nghĩa α

Bước 3: Xác định kiểm định thống kê

Bước 4: Phát biểu quy tắc quyết định

Bước 5: Tính giá trị kiểm định và quyết định

Không bác bỏ H_0 Bác bỏ H_0 chấp nhận H_a

1-189

KIỂM ĐỊNH SỰ KHÁC BIỆT GIỮA HAI TRUNG BÌNH (t-test)

■ Hai mẫu độc lập

Vấn đề nghiên cứu: So sánh hai giá trị trung bình dựa trên hai mẫu độc lập

1. Cả hai mẫu có kích thước ≥ 30 quan sát

Lý thuyết:

Giả thuyết: $H_0: \mu_x - \mu_y = D_0$
 Đối thuyết: $H_a: \mu_x - \mu_y \neq D_0$

$$t = \frac{\bar{x} - \bar{y} - D_0}{\sqrt{\frac{s_x^2}{n_x} + \frac{s_y^2}{n_y}}}$$

Giá trị kiểm định:

1-190

KIỂM ĐỊNH SỰ KHÁC BIỆT GIỮA HAI TRUNG BÌNH (t-test)

■ Hai mẫu độc lập

Vấn đề nghiên cứu: So sánh hai giá trị trung bình dựa trên hai mẫu độc lập

2. Tôn tại 1 trong 2 mẫu có kích thước < 30

Giả định:

Tổng thể của các trung bình có phân phối chuẩn
Phương sai của các tổng thể đồng nhất

1-191

KIỂM ĐỊNH SỰ KHÁC BIỆT GIỮA HAI TRUNG BÌNH (t-test)

■ Hai mẫu độc lập

Vấn đề nghiên cứu: So sánh hai giá trị trung bình dựa trên hai mẫu độc lập

2. Tôn tại 1 trong 2 mẫu có kích thước < 30

Tính phương sai chung:

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

1-192

KIỂM ĐỊNH SỰ KHÁC BIỆT GIỮA HAI TRUNG BÌNH (t-test)

■ Hai mẫu độc lập

Vấn đề nghiên cứu: So sánh hai giá trị trung bình dựa trên hai mẫu độc lập

2. Tôn tại 1 trong 2 mẫu có kích thước < 30

Giá trị kiểm định

$$T = \frac{\bar{X} - \bar{Y}}{\sqrt{s_p^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}}$$

KIỂM ĐỊNH SỰ KHÁC BIỆT GIỮA HAI TRUNG BÌNH (t-test)

- Hai mẫu phụ thuộc hay phối hợp từng cặp

Giả định:

Độ lớn của hai mẫu phải tương đương nhau

Tổng thể có phân phối chuẩn

Giá trị kiểm định:

$$t = \frac{(\bar{x} - D_0)\sqrt{n}}{s_d}$$

ANOVA

ANOVA là sự mở rộng của kiểm định t, phương pháp này cho phép so sánh giá trị trung bình của 3 nhóm trở lên.

Analysis of Variance

QUÁ TRÌNH THỰC HIỆN ANOVA

- ⇒ Giả thuyết H_0 là trung bình các nhóm trong tổng thể là như nhau.
- ⇒ Giả thuyết thay thế H_a là ít nhất một trung bình có sự khác biệt
- ⇒ Kiểm định bằng thống kê F
- ⇒ Quy tắc quyết định: Bác bỏ H_0 nếu:

$$F_{\text{tính toán}} > F_{\text{tra bảng}}$$

1-196

QUÁ TRÌNH THỰC HIỆN ANOVA

μ_1 μ_2 μ_3
 Tổng thể 1 Tổng thể 2 Tổng thể 3

1-197

QUÁ TRÌNH THỰC HIỆN ANOVA

- **Bước 1: Tính trung bình mẫu**
 Tính trung bình mẫu theo công thức:

$$\bar{x}_i = \frac{\sum_{j=1}^{n_i} x_{ij}}{n_i} \quad (i = 1..k)$$

Tính trung bình của k mẫu
$$\bar{x} = \frac{\sum_{i=1}^k n_i \bar{x}_i}{\sum_{i=1}^k n_i}$$

1-198

QUÁ TRÌNH THỰC HIỆN ANOVA

- **Bước 2: Tính tổng các chênh lệch bình phương**
 Tổng các chênh lệch bình phương trong nội bộ nhóm SSW

$$SS_k = \sum_{j=1}^{n_k} (x_{kj} - \bar{x}_k)^2$$

$$SSW = SS_1 + SS_2 + \dots + SS_k$$

QUÁ TRÌNH THỰC HIỆN ANOVA

- **Bước 2:** Tổng các chênh lệch bình phương giữa các nhóm SSG

$$SSG = \sum_{i=1}^k n_i (\bar{x}_i - \bar{x})^2$$

Tổng các chênh lệch bình phương toàn bộ SST

$$SST = SSW + SSG$$

QUÁ TRÌNH THỰC HIỆN ANOVA

- **Bước 3:** Tính các phương sai:

Tính phương sai nội bộ nhóm: $MSW = \frac{SSW}{n - k}$

Tính phương sai giữa các nhóm $MSG = \frac{SSG}{k - 1}$

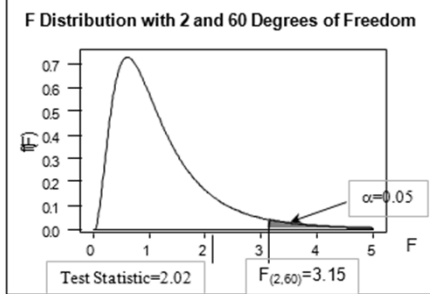
QUÁ TRÌNH THỰC HIỆN ANOVA

- **Bước 4:** Kiểm định giả thuyết.

Giá trị kiểm định $F = \frac{MSG}{MSW}$

Bác bỏ giả thuyết H_0 khi $F > F_{k-1, n-k, \alpha}$

QUÁ TRÌNH THỰC HIỆN ANOVA



QUÁ TRÌNH THỰC HIỆN ANOVA

Bảng tổng quát của ANOVA

Source of Variation	Sum of Squares (SS)	Degree of Freedom (D.f)	Mean Squares (MS)	F ratio
Between-Groups	SSG	$k - 1$	$MSG = \frac{SSG}{k - 1}$	$F = \frac{MSG}{MS}$
Within-Groups	SSW	$n - k$	$MSW = \frac{SSW}{n - k}$	
Total	SST	$n - 1$		

NGUYÊN LÝ THỐNG KÊ KINH TẾ

NGÔ THÁI HÙNG
 ĐH Tài Chính – Marketing
 Email: ngothai.hung@gmail.com

1-205

CHƯƠNG 9

KIỂM ĐỊNH PHI THAM SỐ

1-206

KIỂM ĐỊNH DẤU VÀ HẠNG WILCOXON

■ Hai mẫu liên hệ (phụ thuộc)
 - Trường hợp mẫu nhỏ:
 Đặt giả thuyết: $H_0 : M = m_0$ và $H_a : M > m_0$
 Thay mỗi giá trị quan sát lớn hơn m_0 bằng (+) và nhỏ hơn bằng (-). Nếu giá trị quan sát bằng m_0 , loại bỏ quan sát đó và điều chỉnh lại n .
 Đặt n^+ là số dấu (+) trong mẫu. Với n và $p = 0.5$, tra bảng phân phối nhị thức, tìm:

$$\gamma = P(N^+ \geq n^+)$$

1-207

KIỂM ĐỊNH DẤU VÀ HẠNG WILCOXON

Quyết định: Nếu $\gamma < \alpha$, Bác bỏ H_0 .

Ví dụ: Cho dãy số liệu:
 1.51 1.35 1.69 1.48 1.29 1.27 1.54 1.39 1.45
 Hãy kiểm định giả thuyết sau: $H_0 : M = 1.4$ và $H_a : M > 1.4$
