



GIÁO TRÌNH MÔN HỌC NGUYÊN LÝ THỐNG KÊ

TRÌNH ĐỘ TRUNG CẤP

NGHỀ: KHAI THÁC VẬN TẢI ĐƯỜNG BỘ

Ban hành theo Quyết định số 407/QĐ-CDNGTVT-TWI-ĐT ngày
31/03/2017 của Hiệu trưởng Trường Cao đẳng GTVT Trung ương I

Hà Nội, 2017

BỘ GIAO THÔNG VẬN TẢI
TRƯỜNG CAO ĐẲNG GIAO THÔNG VẬN TẢI TRUNG ƯƠNG I

GIÁO TRÌNH
MÔN HỌC: NGUYÊN LÝ THỐNG KÊ
NGHỀ: KHAI THÁC VẬN TẢI ĐƯỜNG BỘ
TRÌNH ĐỘ: TRUNG CẤP



Hà Nội- 2017

MỤC LỤC

	Mục lục	Trang
CHƯƠNG I. MỘT SỐ VẤN ĐỀ CHUNG VỀ THỐNG KÊ		
I. NGUỒN GỐC MÔN HỌC		
II. THỐNG KÊ LÀ GÌ?		
1. Định nghĩa		
2. Chức năng của thống kê		
3. Phương pháp thống kê		
CHƯƠNG II. QUÁ TRÌNH NGHIÊN CỨU THỐNG KÊ		
1. Tổng thể thống kê		
2. Mẫu		
3. Quan sát		
4. Tiêu thức thống kê		
5. Tham số tổng thể		
6. Tham số mẫu		
IV. CÁC LOẠI THANG ĐO		
1. Khái niệm		
2. Các loại thang đo		
V. THU THẬP THÔNG TIN		
1. Xác định nội dung thông tin		
2. Nguồn số liệu		
2.1. Dữ liệu thứ cấp		
2.2. Dữ liệu sơ cấp		
4.3. Các phương pháp thu thập thông tin		
CHƯƠNG III. PHÂN TỔ THỐNG KÊ		
I. PHÂN TỔ THỐNG KÊ		
1. Khái niệm		
2. Nguyên tắc phân tổ		
3. Phân tổ theo tiêu thức thuộc tính		
4. Phân tổ theo tiêu thức số lượng		
5. Bảng phân phối tần số		
6. Các loại phân tổ thống kê		
II. BẢNG THỐNG KÊ		
1. Khái niệm		
2. Cấu thành bảng thống kê		
3. Các yêu cầu và qui ước xây dựng bảng thống kê		
III. TỔNG HỢP BẢNG ĐỒ THỊ		
1. Biểu đồ hình cột		
2. Biểu đồ diện tích		
3. Biểu đồ tượng hình		
4. Đồ thị đường gấp khúc		
5. Biểu đồ hình màn gập		
CHƯƠNG IV. CÁC MỨC ĐỘ CỦA HIỆN TƯỢNG KINH TẾ-XÃ HỘI		
I. SỐ TUYỆT ĐỐI		
II. SỐ TƯƠNG ĐỐI		
1. Số tương đối động thái		

2. Số tương đối so sánh
3. Số tương đối kế hoạch
4. Số tương đối kết cấu
5. Số tương đối cường độ
- III. SỐ ĐO ĐỘ TẬP TRUNG – SỐ BÌNH QUÂN
 1. Số trung bình cộng
 2. Số trung bình gia quyền
 3. Số trung bình điều hòa
 4. Số trung bình nhân
 5. Số trung vị - Me
 6. Mốt – Mo
- IV. SỐ ĐO ĐỘ PHÂN TÁN
 1. Khoảng biến thiên
 2. Độ lệch tuyệt đối trung bình
 3. Phương sai
 4. Độ lệch chuẩn
 5. Hệ số biến thiên
- V. PHƯƠNG PHÁP CHỈ SỐ
 1. Chỉ số cá thể
 2. Chỉ số tổng hợp
 - 2.1. Chỉ số tổng hợp giá cả
 - 2.2. Chỉ số tổng hợp khối lượng
 3. Chỉ số trung bình tính từ chỉ số tổng hợp
 - 3.1. Chỉ số trung bình điều hòa về biến động của chỉ tiêu chất lượng
 - 3.2. Chỉ số trung bình số học về biến động của chỉ tiêu khối lượng
 4. Chỉ số không gian
 - 4.1. Chỉ số tổng hợp nghiên cứu sự biến động của chỉ tiêu chất lượng ở hai thị trường A và B.
 - 4.2. Chỉ số tổng hợp nghiên cứu sự biến động của chỉ tiêu khối lượng ở hai thị trường A và B
 5. Hệ thống chỉ số liên hoàn 2 nhân tố

CHƯƠNG V. SỰ BIẾN ĐỘNG CỦA CÁC HIỆN TƯỢNG KINH TẾ - XÃ HỘI

- I. PHÂN PHỐI CHUẨN
 1. Định nghĩa
 2. Phân phối chuẩn tắc (đơn giản)
 3. Bảng phân phối chuẩn tắc (đơn giản)
 4. Khái niệm Z_{α}
 5. Một vài công thức xác suất thường dùng
- II. PHÂN PHỐI CỦA ĐẠI LƯỢNG THỐNG KÊ
 1. Phân phối Chi bình phương
 2. Phân phối Student
 3. Phân phối Fisher (F)
- III. PHÂN PHỐI MẪU
 1. Khái niệm
 2. Định lý giới hạn trung tâm
 3. Các tính chất của phân phối mẫu

LỜI NÓI ĐẦU

Thống kê là một ngành khoa học có vai trò quan trọng trong hầu hết các lĩnh vực kinh tế xã hội. Nguyên lý thống kê kinh tế, lý thuyết thống kê theo hướng ứng dụng trong lĩnh vực kinh tế và quản trị kinh doanh, là công cụ không thể thiếu được trong hoạt động nghiên cứu và quản lý. Nguyên lý thống kê kinh tế đã trở thành một môn học cơ sở trong hầu hết các ngành đào tạo thuộc khối kinh tế.

Trong bối cảnh đào tạo đại học theo tín chỉ hóa, thời gian lên lớp được giới hạn và sinh viên được khuyến khích tự tham khảo tài liệu và tự học có hướng dẫn của giảng viên. Nhu cầu về một tài liệu giảng dạy và học tập môn nguyên lý thống kê kinh tế, vừa phù hợp với chương trình đào tạo theo tín chỉ, vừa nhất quán với các môn học định lượng trong chương trình đào tạo bậc đại học là cần thiết. Giáo trình này được biên soạn nhằm mục đích giúp cho bạn đọc am hiểu các vấn đề về lý thuyết, chuẩn bị cho những tiết thực hành trên máy tính có hiệu quả, là cơ sở quan trọng cho người học tiếp cận các môn học chuyên ngành kinh tế.

Để đáp ứng nhu cầu trên, Tác giả thực hiện biên soạn quyển sách giáo trình thống kê kinh tế. Tài liệu này được viết trên cơ sở bạn đọc đã có kiến thức về xác suất thống kê toán, cho nên cuốn sách không đi sâu về mặt toán học mà chú trọng đến kết quả và ứng dụng trong lĩnh vực kinh tế và quản trị kinh doanh với các ví dụ gần gũi với thực tế.

Với kinh nghiệm giảng dạy được tích lũy qua nhiều năm, tham gia thực hiện các đề tài nghiên cứu trong lĩnh vực kinh tế xã hội; cùng với sự phối hợp và hỗ trợ của đồng nghiệp, đặc biệt của ThS. Nguyễn Ngọc Lam, Tác giả hy vọng quyển sách này đáp ứng được nhu cầu học tập của các sinh viên và nhu cầu tham khảo của các bạn đọc có quan tâm đến nguyên lý thống kê kinh tế trong nghiên cứu kinh tế xã hội.

Trong quá trình biên soạn chắc chắn không tránh khỏi những thiếu sót, Tác giả rất mong nhận được những ý kiến đóng góp quý báu của bạn đọc để lần tái bản sau quyển sách được hoàn thiện hơn. Xin chân thành cảm ơn.

Nhóm tác giả

CHƯƠNG I

MỘT SỐ VẤN ĐỀ CHUNG VỀ THỐNG KÊ

I. NGUỒN GỐC MÔN HỌC

Nếu thống kê được hiểu theo nghĩa thông thường thì ngay từ thời cổ đại con người đã đã chú ý đến việc này thông qua việc ghi chép đơn giản.

Cuối thế kỷ XVII, lực lượng sản xuất phát triển mạnh mẽ làm cho phương thức sản xuất của chủ nghĩa tư bản ra đời. Kinh tế hàng hóa phát triển dẫn đến các ngành sản xuất riêng biệt tăng thêm, phân công lao động xã hội ngày càng phát triển. Tính chất xã hội của sản xuất ngày càng cao, thị trường được mở rộng không chỉ trong một nước mà toàn thế giới. Để phục vụ cho mục đích kinh tế, chính trị và quân sự nhà nước tư bản và các chủ tư bản cần rất nhiều thông tin thường xuyên về thị trường, giá cả, sản xuất, nguyên liệu, dân số,... Do đó, công tác thống kê phát triển nhanh chóng. Chúng ta có thể đưa ra 3 nhóm tác giả được gọi là những người khai sáng cho ngành khoa học thống kê:

- Những người đầu tiên đưa ngành khoa học thống kê đi vào thực tiễn, đại diện cho những tác giả này là nhà kinh tế học người Đức H.Conhring (1606 - 1681), năm 1660 ông đã giảng dạy tại trường đại học Halmsted về phương pháp nghiên cứu hiện tượng xã hội dựa vào số liệu điều tra cụ thể.

- Với những thành quả của người đi trước, bổ sung hoàn chỉnh thành môn học chính thống, đại diện là William Petty, một nhà kinh tế học của người Anh, là tác giả cuốn “Số học chính trị” xuất bản năm 1682, một số tác phẩm có tính chất phân tích thống kê đầu tiên ra đời.

- Thống kê được gọi với nhiều tên khác nhau thời bấy giờ, sau đó năm 1759 một giáo sư người Đức, Achenwall (1719-1772) lần đầu tiên dùng danh từ “Statistics” (một thuật ngữ gốc La tinh “Status”, có nghĩa là Nhà nước hoặc trạng thái của hiện tượng) - sau này người ta dịch ra là “Thống kê”.

Kể từ đó, thống kê có sự phát triển rất mạnh mẽ và ngày càng hoàn thiện, gắn liền với nhiều nhà toán học - thống kê học nổi tiếng như: M.V.Lomonoxop (nga, 1711-1765), Laplace (Pháp, 1749-1827), I.Fisher, W.M.Pearsons,...

II. THỐNG KÊ LÀ GÌ?

1. Định nghĩa

Thống kê là một hệ thống các phương pháp bao gồm thu thập, tổng hợp, trình bày số liệu, tính toán các đặc trưng của đối tượng nghiên cứu nhằm phục vụ cho quá trình phân tích, dự đoán và ra quyết định.

2. Chức năng của thống kê

Thống kê thường được phân thành 2 lĩnh vực:

- **Thống kê mô tả** (*Descriptive statistics*): là các phương pháp có liên quan đến việc thu thập số liệu, tóm tắt, trình bày, tính toán và mô tả các đặc trưng khác nhau để phản ánh một cách tổng quát đối tượng nghiên cứu.

- **Thống kê suy luận** (*Inferential statistics*): là bao gồm các phương pháp ước lượng các đặc trưng của tổng thể, phân tích mối liên hệ giữa các hiện tượng nghiên cứu, dự đoán hoặc ra quyết định trên cơ sở thông tin thu thập từ kết quả quan sát mẫu.

3. Phương pháp thống kê

- Thu thập và xử lý số liệu:

Số liệu thu thập thường rất nhiều và hỗn độn, các dữ liệu đó chưa đáp ứng cho quá trình nghiên cứu. Để có hình ảnh tổng quát về tổng thể nghiên cứu, số liệu thu thập phải được xử lý tổng hợp, trình bày, tính toán các số đo; kết quả có được sẽ giúp khái quát được đặc trưng của tổng thể.

- Nghiên cứu các hiện tượng trong hoàn cảnh không chắc chắn:

Trong thực tế, có nhiều hiện tượng mà thông tin liên quan đến đối tượng nghiên cứu không đầy đủ mặc dù người nghiên cứu đã có sự cố gắng. Ví dụ như nghiên cứu về nhu cầu của thị trường về một sản phẩm ở mức độ nào, tình trạng của nền kinh tế ra sao, để nắm được các thông tin này một cách rõ ràng quả là một điều không chắc chắn.

- Điều tra chọn mẫu:

Trong một số trường hợp để nghiên cứu toàn bộ tất cả các quan sát của tổng thể là một điều không hiệu quả, xét cả về tính kinh tế (chi phí, thời gian) và tính kịp thời, hoặc không thực hiện được. Chính điều này đã đặt ra cho thống kê xây dựng các phương pháp chỉ cần nghiên cứu một bộ phận của tổng thể mà có thể suy luận cho hiện tượng tổng quát mà vẫn đảm bảo độ tin cậy cho phép, đó là phương pháp điều tra chọn mẫu.

- Nghiên cứu mối liên hệ giữa các hiện tượng:

Giữa các hiện tượng nghiên cứu thường có mối liên hệ với nhau. Ví dụ như mối liên hệ giữa chi tiêu và thu nhập; mối liên hệ giữa lượng vốn vay và các yếu tố tác động đến lượng vốn vay như chi tiêu, thu nhập, trình độ học vấn; mối liên hệ giữa tốc độ phát triển với tốc độ phát triển của các ngành, lạm phát, tốc độ phát triển dân số,... Sự hiểu biết về mối liên hệ giữa các hiện tượng rất có ý nghĩa, phục vụ cho quá trình dự đoán

- Dự đoán:

Dự đoán là một công việc cần thiết trong tất cả các lĩnh vực hoạt động. Trong hoạt động dự đoán người ta có thể chia ra thành nhiều loại:

(1). Dự đoán dựa vào định lượng và dựa vào định tính. Tuy nhiên, trong thống kê chúng ta chủ yếu xem xét về mặt định lượng với mục đích cung cấp cho những nhà quản lý có cái nhìn mang tính khoa học hơn và cụ thể hơn trước khi ra quyết định phù hợp.

(2). Dự đoán dựa vào nội suy và dựa vào ngoại suy.

- Dự đoán nội suy là chúng ta dựa vào bản chất của hiện tượng để suy luận, ví dụ như chúng ta xem xét một liên hệ giữa lượng sản phẩm sản xuất ra phụ thuộc các yếu tố đầu vào như vốn, lao động và trình độ khoa học kỹ thuật.

- Dự đoán dựa vào ngoại suy là chúng ta chỉ quan sát sự biến động của hiện tượng trong thực tế, tổng hợp lại thành qui luật và sử dụng qui luật này để suy luận, dự đoán sự phát triển của hiện tượng. Ví dụ như để đánh giá kết quả hoạt động của một công ty người ta xem xét kết quả hoạt động kinh doanh của họ qua nhiều năm.

Ngoài ra, người ta còn có thể phân chia dự báo thống kê ra thành nhiều loại khác.

CHƯƠNG II

CÁC KHÁI NIỆM THƯỜNG DÙNG TRONG THỐNG KÊ

1. Tổng thể thống kê (Populations)

Tổng thể thống kê là tập hợp các đơn vị cá biệt về sự vật, hiện tượng trên cơ sở một đặc điểm chung nào đó cần được quan sát, phân tích mặt lượng của chúng. Các đơn vị, phần tử tạo nên hiện tượng được gọi là các đơn vị tổng thể.

Như vậy muốn xác định được một tổng thể thống kê, ta cần phải xác định được tất cả các đơn vị tổng thể của nó. Thực chất của việc xác định tổng thể thống kê là việc xác định các đơn vị tổng thể.

Trong nhiều trường hợp, các đơn vị của tổng thể được biểu hiện một cách rõ ràng, dễ xác định. Ta gọi nó là tổng thể bộ lộ. Ngược lại, một tổng thể mà các đơn vị của nó không được nhận biết một cách trực tiếp, ranh giới của tổng thể không rõ ràng được gọi là tổng thể tiềm ẩn.

Đối với tổng thể tiềm ẩn, việc tìm được đầy đủ, chính xác gặp nhiều khó khăn. Việc nhầm lẫn, bỏ sót các đơn vị trong tổng thể dễ xảy ra. Ví dụ như tổng thể là những nghệ sĩ nhạc cổ điển, tổng thể người mê tín dị đoan,...

2. Mẫu (Samples)

Mẫu là một bộ phận của tổng thể, đảm bảo được tính đại diện và được chọn ra để quan sát và dùng để suy diễn cho toàn bộ tổng thể. Như vậy, tất cả các phần tử của mẫu phải thuộc tổng thể, nhưng ngược lại các phần tử của tổng thể thì chưa chắc thuộc mẫu. Điều này tưởng chừng là đơn giản, tuy nhiên trong một số trường hợp việc xác định mẫu cũng có thể dẫn đến nhầm lẫn, đặc biệt là trong trường hợp tổng thể ta nghiên cứu là tổng thể tiềm ẩn.

Ngoài ra, chọn mẫu như thế nào để làm cơ sở suy diễn cho tổng thể, tức là mẫu phải mang tính đại diện cho tổng thể. Điều này thực sự không dễ dàng, ta chỉ cố gắng hạn chế tối đa sự sai biệt này mà thôi chứ không thể khắc phục được hoàn toàn.

3. Quan sát (Observations)

Là mỗi đơn vị của mẫu ; trong một số tài liệu còn được gọi là quan trắc.

4. Tiêu thức thống kê

Các đơn vị tổng thể thường có nhiều đặc điểm khác nhau, tuy nhiên trong thống kê người ta chỉ chọn một số đặc điểm để nghiên cứu, các đặc điểm này người ta gọi là tiêu thức thống kê. Như vậy, tiêu thức thống kê là khái niệm chỉ các đặc điểm của đơn vị tổng thể. Mỗi tiêu thức thống kê đều có các giá trị biểu hiện của nó, dựa vào sự biểu hiện của nó người ta chia ra làm hai loại:

a) Tiêu thức thuộc tính: là tiêu thức phản ánh loại hoặc tính chất của đơn vị. Ví dụ như ngành kinh doanh, nghề nghiệp,...

b) Tiêu thức số lượng: là đặc trưng của đơn vị tổng thể được thể hiện bằng con số. Ví dụ, năng suất của một loại cây trồng.

Tiêu thức số lượng được chia làm 2 loại:

- Loại rời rạc: là loại các giá trị có thể của nó là hữu hạn hay vô hạn và có thể đếm được.

- Loại liên tục: là loại mà giá trị của nó có thể nhận bất kỳ một trị số nào đó trong một khoảng nào đó.

5. Tham số tổng thể

Là giá trị quan sát được của tổng thể và dùng để mô tả đặc trưng của hiện tượng nghiên cứu. Trong xác suất thống kê toán chúng ta đã biết các tham số tổng thể như trung bình tổng thể (μ), tỷ lệ tổng thể (p), phương sai tổng thể (σ^2). Ngoài ra, trong quá trình nghiên cứu sâu môn thống kê chúng ta còn có thêm nhiều tham số tổng thể nữa như: tương quan tổng thể (ρ), hồi qui tuyến tính tổng thể,...

6. Tham số mẫu

Tham số mẫu là giá trị tính toán được của một mẫu và dùng để suy rộng cho tham số tổng thể. Đó là cách giải thích mang tính chất thông thường, còn đối với xác suất thống kê thì tham số mẫu là ước lượng điểm của tham số tổng thể, trong trường hợp chúng ta chưa biết tham số tổng thể chúng ta có thể sử dụng tham số mẫu để ước lượng tham số tổng thể. Chúng ta có thể liệt kê vài tham số mẫu như sau: trung bình mẫu ($\bar{x}$), tỷ lệ mẫu ($\hat{p}$), phương sai mẫu (S^2), hệ số tương quan mẫu (r),...

III. CÁC LOẠI THANG ĐO (Scales of Measurement)

Đứng trên quan điểm của nhà nghiên cứu, chúng ta cần xác định các phương pháp phân tích thích hợp dựa vào mục đích nghiên cứu và bản chất của dữ liệu. Do vậy, đầu tiên chúng ta tìm hiểu bản chất của dữ liệu thông qua khảo sát các cấp độ đo lường khác nhau vì mỗi cấp độ sẽ chỉ cho phép một số phương pháp nhất định mà thôi.

1. Khái niệm

- **Số đo:** là việc gán những dữ kiện lượng hoá hay những ký hiệu cho những hiện tượng quan sát. Chẳng hạn như những đặc điểm của khách hàng về sự chấp nhận, thái độ, thị hiếu hoặc những đặc điểm có liên quan khác đối với một sản phẩm mà họ tiêu dùng.

- **Thang đo:** là tạo ra một thang điểm để đánh giá đặc điểm của đối tượng nghiên cứu thể hiện qua sự đánh giá, nhận xét.

2. Các loại thang đo

- Thang đo danh nghĩa (Nominal scale):

Là loại thang đo sử dụng cho dữ liệu thuộc tính mà các biểu hiện của dữ liệu không có sự hơn kém, khác biệt về thứ bậc. Các con số không có mối quan hệ hơn kém, không thực hiện được các phép tính đại số. Các con số chỉ mang tính chất mã hoá. Ví dụ, tiêu thức giới tính ta có thể đánh số 1 là nam, 2 là nữ.

- Thang đo thứ bậc (Ordinal scale):

Là loại thang đo dùng cho các dữ liệu thuộc tính. Tuy nhiên trường hợp này biểu hiện của dữ liệu có sự so sánh. Ví dụ, trình độ thành thạo của công nhân được phân chia ra các bậc thợ từ 1 đến 7. Phân loại giảng viên trong các trường đại học: Giáo sư, P.Giáo sư, Giảng viên chính, Giảng viên. Thang đo này cũng không thực hiện được các phép tính đại số.

- Thang đo khoảng (Interval scale):

Là loại thang đo dùng cho các dữ liệu số lượng. Là loại thang đo cũng có thể dùng để xếp hạng các đối tượng nghiên cứu nhưng khoảng cách bằng nhau trên thang đo đại diện cho khoảng cách bằng nhau trong đặc điểm của đối tượng. Với thang đo này ta có thể thực hiện các phép tính đại số trừ phép chia không có ý nghĩa. Ví dụ như điểm môn học của sinh viên. Sinh viên A có điểm thi là 8 điểm, sinh viên B có điểm là 4 thì không thể nói rằng sinh viên A giỏi gấp hai lần sinh viên B.

- Thang đo tỷ lệ (Ratio scale):

Là loại thang đo cũng có thể dùng dữ liệu số lượng. Trong các loại thang đo đây là loại thang đo cao nhất. Ngoài đặc tính của thang đo khoảng, phép chia có thể thực hiện

được. Ví dụ, thu nhập trung bình 1 tháng của ông A là 2 triệu đồng và thu nhập của bà B là 4 triệu đồng, ta có thể nói rằng thu nhập trung bình trong một tháng của bà B gấp đôi thu nhập của ông A.

Tuỳ theo thang đo chúng ta có thể có một số phương pháp phân tích phù hợp, ta có thể tóm tắt như sau:

Phương pháp phân tích thống kê thích hợp với các thang đo

Loại thang đo	Đo lường độ tập trung	Đo lường độ phân tán	Đo lường tính tương quan	Kiểm định
1. Thang biểu danh	Mốt	Không có	Hệ số ngẫu nhiên	Kiểm định χ^2
2. Thang thứ tự	Trung vị	Sô phần trăm	Dãy tương quan	Kiểm định dấu
3. Thang khoảng	Trung bình	Độ lệch chuẩn	Hệ số tương quan	Kiểm định t, F
4. Thang tỷ lệ	Trung bình tỷ lệ	Hệ số biến thiên	Tất cả các phép trên	Sử dụng tất cả các phép trên

IV. THU THẬP THÔNG TIN

Về nguyên tắc, thống kê mô tả chắc hẳn có từ lâu đời cũng gần như chữ viết. Nó liên quan chặt chẽ với nhu cầu của con người muốn sắp xếp lại một cách có trật tự trong vô vàn thông tin sự kiện đã đến với họ để hiểu hơn thực tại hơn nhằm tác động lên nó tốt hơn. Khi nghiên cứu bất kỳ hiện tượng kinh tế xã hội nào công việc đầu tiên là thu thập dữ liệu, sau đó là trình bày dữ liệu và phân tích.

1. Xác định nội dung thông tin

Nói chung, tuỳ thuộc vào mục đích nghiên cứu để xác định những nội dung thông tin cần thu thập. Thông tin sử dụng cho quá trình nghiên cứu phải đảm bảo các yêu cầu cơ bản sau:

- **Thích đáng:** Số liệu thu thập phải phù hợp, đáp ứng được mục đích nghiên cứu. Số liệu đáp ứng được mục tiêu nghiên cứu có tính chất trực tiếp hoặc gián tiếp. Đối với những thông tin dễ tiếp cận thường thì ta sử dụng số liệu trực tiếp, ví dụ muốn biết được nhu cầu của khách hàng chúng ta có thể hỏi trực tiếp khách hàng. Tuy nhiên, một số nội dung nghiên cứu mang tính chất nhạy cảm hoặc khó thu thập thì chúng ta có thể thu nhập những số liệu gián tiếp có liên quan, ví dụ để thu thập thu nhập của cá nhân chúng ta có thể thu thập những nội dung có liên quan như nghề nghiệp, đơn vị công tác, chức vụ, nhà ở, phương tiện đi lại...

- **Chính xác:** Các thông tin trong quá trình nghiên cứu phải có giá trị, đáng tin cậy để các phân tích kết luận phản ánh được đặc điểm bản chất của hiện tượng.

- **Kịp thời:** Yêu cầu thông tin không những đáp ứng yêu cầu phù hợp, chính xác mà giá trị thông tin còn thể hiện ở chỗ nó có phục vụ kịp thời cho công tác quản lý và tiến trình ra các quyết định hay không.

- **Khách quan:** Tức là số liệu thu thập được không bị ảnh hưởng vào tính chủ quan của người thu thập cũng như người cung cấp số liệu và ngay cả trong thiết kế bảng câu hỏi. Yếu tố khách quan tường chừng thực hiện rất dễ dàng nhưng thực tế thì chúng ta khó có thể khắc phục vấn đề này một cách trọn vẹn, chúng ta chỉ có thể hạn chế yếu tố chủ quan một cách tối đa. Ví dụ chỉ cần một hành động đơn giản là tiếp cận với đáp viên là ít nhiều cũng ảnh hưởng đến kết quả trả lời của họ.

2. Nguồn số liệu

Khi nghiên cứu một hiện tượng cụ thể, người nghiên cứu có thể sử dụng từ nguồn số liệu đã có sẵn đã được công bố hay chưa công bố hay tự mình thu thập các dữ liệu cần thiết cho nghiên cứu. Dựa vào cách thức này người ta chia dữ liệu thành 2 nguồn: dữ liệu thứ cấp và dữ liệu sơ cấp.

2.1. Dữ liệu thứ cấp (Secondary data):

Dữ liệu thứ cấp là các thông tin đã có sẵn và đã qua tổng hợp, xử lý. Loại dữ kiện này có thể thu thập từ các nguồn sau:

(1) Số liệu nội bộ: là loại số liệu đã được ghi chép cập nhật trong đơn vị hoặc được thu thập từ các cuộc điều tra trước đây.

(2) Số liệu từ các ấn phẩm của nhà nước: Các dữ liệu do các cơ quan thống kê nhà nước phát hành định kỳ như niên giám thống kê, các thông tin cập nhật hàng năm về tình hình dân số lao động, kết quả sản xuất của các ngành trong nền kinh tế, số liệu về văn hoá xã hội.

(3) Báo, tạp chí chuyên ngành: Các báo và tạp chí đề cập đến vấn đề có tính chất chuyên ngành như tạp chí thống kê, giá cả thị trường,...

(4) Thông tin của các tổ chức, hiệp hội nghề nghiệp: Viện nghiên cứu kinh tế, phòng thương mại

(5) Các công ty chuyên tổ chức thu thập thông tin, nghiên cứu và cung cấp thông tin theo yêu cầu.

Số liệu thứ cấp có ưu điểm là có thể chia sẻ chi phí, do đó nó có tính kinh tế hơn, số liệu được cung cấp kịp thời hơn. Tuy nhiên, dữ liệu thứ cấp thường là các thông tin cơ bản, số liệu đã được tổng hợp đã qua xử lý cho nên không đầy đủ hoặc không phù hợp cho quá trình nghiên cứu. Số liệu thứ cấp thường ít được sử dụng để dự báo trong thống kê, số liệu này thường được sử dụng trong trình bày tổng quan nội dung nghiên cứu, là cơ sở để phát hiện ra vấn đề nghiên cứu. Ngoài ra, số liệu thứ cấp còn được sử dụng để đối chiếu lại kết quả nghiên cứu để nhằm kiểm tra lại tính đúng đắn hoặc phát hiện ra những vấn đề mới để có hướng nghiên cứu tiếp.

2.2. Dữ liệu sơ cấp (Primary data):

Là các thông tin thu thập từ các cuộc điều tra. Căn cứ vào phạm vi điều tra có thể chia thành 2 loại: Điều tra toàn bộ và điều tra chọn mẫu.

a) Điều tra toàn bộ: Là tiến hành thu thập thông tin trên tất cả các đơn vị thuộc tổng thể nghiên cứu.

Ưu điểm của điều tra toàn bộ là thu thập được thông tin về tất cả các đơn vị tổng thể. Tuy nhiên, loại điều tra này thường gặp phải một số trở ngại sau:

- Số lượng đơn vị thuộc tổng thể chung thường rất lớn cho nên tiến hành điều tra toàn bộ mất nhiều thời gian và tốn kém.

- Trong một số trường hợp do thời gian kéo dài dẫn đến số liệu kém chính xác do hiện tượng tự biến động qua thời gian.

- Trong một số trường hợp điều tra toàn bộ sẽ không thực hiện được, ví dụ như kiểm tra chất lượng sản phẩm phải phá huỷ các đơn vị thuộc đối tượng nghiên cứu.

b) Điều tra chọn mẫu: Để nghiên cứu tổng thể, ta chỉ cần lấy ra một số phần tử đại diện để nghiên cứu và từ đó suy ra kết quả cho tổng thể bằng các phương pháp thống kê.

Điều tra chọn mẫu thường được sử dụng vì các lý do sau:

- Tiết kiệm chi phí

- Cung cấp thông tin kịp thời cho quá trình nghiên cứu

- Đáng tin cậy. Đây là yếu tố rất quan trọng, nó làm cho điều tra chọn mẫu trở nên có hiệu quả và được chấp nhận. Tuy nhiên, để có sự đáng tin cậy này chúng ta phải có phương pháp khoa học để đảm bảo tính chính xác để chỉ cần chọn ra một số quan sát mà có thể suy luận cho cả tổng thể rộng lớn – đó là nhờ vào các lý thuyết thống kê.

Việc sử dụng điều tra toàn bộ hay điều tra chọn mẫu phụ thuộc vào nhiều yếu tố có liên quan: kích thước tổng thể, thời gian nghiên cứu, khả năng về tài chính và nguồn lực, đặc điểm của nội dung nghiên cứu.

3. Các phương pháp thu thập thông tin

Để thu thập dữ liệu ban đầu, tùy theo nguồn kinh phí và đặc điểm của đối tượng cần thu thập thông tin, ta có các phương pháp sau đây:

a) Quan sát: Là phương pháp thu thập dữ liệu bằng cách quan sát hành động, hành vi thái độ của đối tượng được điều tra. Ví dụ, nghiên cứu trẻ con yêu thích màu sắc nào, quan sát thái độ khách hàng khi dùng thử loại sản phẩm. Phương pháp này tỏ ra hiệu quả đối với các trường hợp đối tượng khó tiếp cận và tăng tính khách quan của đối tượng. Tuy nhiên, phương pháp này tỏ ra khá tốn kém nhưng lượng thông tin thu thập được ít.

b) Phương pháp gửi thư: Theo phương pháp này nhân viên điều tra gửi bằng câu hỏi đến đối tượng cung cấp thông tin qua đường bưu điện. Phương pháp gửi thư có thể thu thập thông tin với khối lượng lớn, tiết kiệm chi phí so với các phương pháp khác. Tuy nhiên tỷ lệ trả lời bằng phương pháp này tương đối thấp, đây là một nhược điểm rất lớn của phương pháp này.

c) Phòng vấn bằng điện thoại: Phương pháp thu thập thông tin bằng cách phỏng vấn qua điện thoại. Phương pháp này thu thập được thông tin một cách nhanh chóng, tuy nhiên phương pháp này có nhược điểm: tốn kém, nội dung thu thập thông tin bị hạn chế.

d) Phòng vấn trực tiếp:

Phương pháp phỏng vấn trực tiếp thích hợp cho những cuộc điều tra cần thu thập nhiều thông tin, nội dung của thông tin tương đối phức tạp cần thu thập một cách chi tiết. Phương pháp phỏng vấn trực tiếp có 2 hình thức:

(1) Phỏng vấn cá nhân. Nhân viên điều tra tiếp xúc với đối tượng cung cấp thông tin thường tại nhà riêng hoặc nơi làm việc. Thông thường phỏng vấn trực tiếp được áp dụng khi chúng ta cho tiến hành điều tra chính thức.

(2) Phỏng vấn nhóm. Nhân viên điều tra phỏng vấn từng nhóm để thảo luận về một vấn đề nào đó. Trường hợp này người ta thường sử dụng khi điều tra thử để kiểm tra lại nội dung của bảng câu hỏi được hoàn chỉnh chưa hoặc nhằm tìm hiểu một vấn đề phức tạp mà bản thân người nghiên cứu chưa nắm được một cách đầy đủ mà cần phải có ý kiến cụ thể từ những người am hiểu.

Sau đây ta có bảng tổng hợp một số ưu nhược điểm của các phương pháp thu thập thông tin.

Đặc điểm của các phương pháp thu thập thông tin

Tính chất	Phương pháp gửi thư	Phòng vấn qua điện thoại	Phòng vấn trực tiếp
Linh hoạt	Kém	Tốt	Tốt
Khối lượng thông tin	Đầy đủ	Hạn chế	Đầy đủ
Tốc độ thu thập thông tin	Chậm	Nhanh	Nhanh
Tỷ lệ câu hỏi được trả lời	Thấp	Cao	Cao
Chi phí	Tiết kiệm	Tốn kém	Tốn kém

CHƯƠNG III

PHÂN TỔ THỐNG KÊ

Thông tin ban đầu có tính rời rạc, dữ liệu hỗn độn không theo một trật tự nào và có thể quá nhiều nếu nhìn vào đây chúng ta không thể phát hiện được điều gì để phục vụ cho quá trình nghiên cứu. Do đó, chúng ta cần phải trình bày một cách có thể thống với hai mục đích là làm cho bảng dữ liệu gọn lại, hai là thể hiện được tính chất của nội dung nghiên cứu.

I. PHÂN TỔ THỐNG KÊ

1. Khái niệm

Phân tổ còn được gọi là phân lớp thống kê là căn cứ vào một hay một số tiêu thức để chia các đơn vị tổng thể ra thành nhiều tổ (lớp, nhóm) có tính chất khác nhau.

2. Nguyên tắc phân tổ

Một cách tổng quát tổng thể phải được phân chia một cách trọn vẹn, tức là một đơn vị của tổng thể chỉ thuộc một tổ duy nhất và một đơn vị thuộc một tổ nào đó phải thuộc tổng thể.

3. Phân tổ theo tiêu thức thuộc tính

- Trường hợp tiêu thức thuộc tính chỉ có một vài biểu hiện thì mỗi biểu hiện của tiêu thức thuộc tính có thể chia thành một tổ. Ví dụ, tiêu thức giới tính.
- Trường hợp tiêu thức thuộc tính có nhiều biểu hiện, ta ghép nhiều nhóm nhỏ lại với nhau theo nguyên tắc các nhóm ghép lại với nhau có tính chất giống nhau hoặc gần giống nhau. Ví dụ phân tổ trong công nghiệp chế biến: Thực phẩm và đồ uống, thuốc lá, dệt,...

4. Phân tổ theo tiêu thức số lượng

- *Trường hợp tiêu thức số lượng có ít biểu hiện*, thì cứ mỗi một lượng biến có thể thành lập một tổ.

Ví dụ 1.1: phân tổ công nhân trong một xí nghiệp dệt theo số máy do mỗi công nhân thực hiện.

Số máy/Công nhân	Số công nhân
10	3
11	7
12	20
13	50
14	35
15	15
Tổng	130

- *Trường hợp tiêu thức số lượng có nhiều biểu hiện*, ta phân tổ khoảng cách mỗi tổ và mỗi tổ có một giới hạn:

- Giới hạn trên: lượng biến nhỏ nhất của tổ.

- Giới hạn dưới: lượng biến lớn nhất của tổ.

Tuỳ theo mục đích nghiên cứu, người ta phân ra 2 loại phân tổ đều và phân tổ không đều.

- Phân tổ đều: Là phân tổ có khoảng cách tổ bằng nhau. Thông thường nếu chỉ vì mục đích nghiên cứu phân phối của tổng thể hoặc làm cho bảng thống kê gọn lại thì ta thường dùng phương pháp này.

Để xác định số tổ hình như không có một tiêu chuẩn tối ưu nó phụ thuộc vào kinh nghiệm. Dưới đây là một cách phân chia tổ mang tính chất tham khảo.

- Xác định số tổ (Number off classes):

$$\text{Số tổ} = (2 \times n)^{0,3333} \quad n: \text{Số đơn vị tổng thể}$$

- Xác định khoảng cách tổ (Class interval):

$$k = \frac{X_{\max} - X_{\min}}{\text{Số tổ}}$$

- Xác định tần số (Frequency) của mỗi tổ: bằng cách đếm các quan sát rơi vào giới hạn của tổ đó.

- Một số qui ước khi lập bảng phân tổ:

- Trường hợp phân tổ theo tiêu thức số lượng rời rạc thì giới hạn trên và giới hạn dưới của 2 tổ kế tiếp nhau không được trùng nhau.

Ví dụ 1.2: Các xí nghiệp ở tỉnh X được phân tổ theo tiêu số lượng công nhân:

Số lượng công nhân	Số xí nghiệp
≤ 100	80
101 – 200	60
201 – 500	6
501 – 1.000	4
1.001 – 2.000	1
Tổng	151

- Trường hợp phân tổ theo tiêu thức số lượng loại liên tục, thường có qui ước sau:

- * Giới hạn trên và giới hạn dưới của 2 tổ kế tiếp trùng nhau.

- * Quan sát có lượng biến bằng đúng giới hạn trên của một tổ nào đó thì đơn vị đó được xếp vào tổ kế tiếp.

Ví dụ 1.3: phân tổ các tổ chức thương nghiệp theo doanh thu.

Doanh thu (triệu đồng)	Số tổ chức thương nghiệp
≤ 1.000	2
1.000-2.000	9
2.000-3.000	12
3.000-4.000	7
Tổng	30

5. Bảng phân phối tần số (Frequency table)

Sau khi phân tổ chúng ta có thể trình bày số liệu bằng cách sử dụng bảng phân phối tần số để biết được một số tính chất cơ bản của hiện tượng nghiên cứu.

Lượng biến	Tần số	Tần số tương đối	Tần số tích lũy
x_1	f_1	f_1/n	f_1
x_2	f_2	f_2/n	$f_1 + f_2$
...	...	...	...
x_i	f_i	f_i/n	$f_1 + f_2 + \dots + f_i$
...	...	...	...
x_k	f_k	f_k/n	$f_1 + f_2 + \dots + f_k$
Cộng	$\sum_{i=1}^k f_i = n$	1	

Trong đó lượng biến có thể là giá trị cụ thể hoặc là một khoảng.

6. Các loại phân tổ thống kê

- *Phân tổ kết cấu:*

Trong công tác nghiên cứu thống kê, các bảng phân tổ kết cấu được sử dụng rất phổ biến nhằm mục đích nêu lên bản chất của hiện tượng trong điều kiện nhất định và để nghiên cứu xu hướng phát triển của hiện tượng qua thời gian.

Ví dụ 1.4: Để xem xét cơ cấu giữa các nhóm ngành trong một quốc gia nào đó ta lập bảng như sau:

Bảng 1.1. Cơ cấu tổng sản phẩm của quốc gia X theo nhóm ngành, 2003 -2007

Đơn vị tính: %.

Tổng sản phẩm theo nhóm ngành	2003	2004	2005	2006	2007
Nông, lâm nghiệp và thủy sản	24,53	23,24	23,03	22,54	21,76
Công nghiệp và xây dựng	36,73	38,13	38,49	39,47	40,09
Dịch vụ	38,74	38,63	38,48	37,99	38,15
Tổng	100,00	100,00	100,00	100,00	100,00

Qua bảng kết cấu trên, ta thấy có thấy sự thay đổi về dịch chuyển cơ cấu ngành: Nhóm ngành công nghiệp và xây dựng có xu hướng tăng, nhóm ngành nông, lâm, thủy sản có xu hướng giảm,...

- *Phân tổ liên hệ:*

Khi tiến hành phân tổ liên hệ, các tiêu thức có liên hệ với nhau được phân biệt thành 2 loại tiêu thức nguyên nhân và tiêu thức kết quả. Phân tổ liên hệ có thể được vận dụng để nghiên cứu mối liên hệ giữa nhiều tiêu thức: mối liên hệ giữa năng suất với lượng phân bón, nghiên cứu giữa năng suất lao động của công nhân với tuổi nghề, bậc thợ, trình độ trang bị kỹ thuật,...

Ví dụ 1.5: Ta có bảng phân tổ liên hệ sau:

Bảng 1.2. Mối liên hệ giữa năng suất lao động với trình độ kỹ thuật nghề nghiệp của quốc gia X năm 2007

Trình độ kỹ thuật	Tuổi nghề (Năm)	Số công nhân	Sản lượng cả năm (tấn)	Năng suất lao động bình quân (tấn)
Đã được đào tạo kỹ thuật	dưới 5	15	1.125	75
	5-10	40	3.750	94
	10-15	40	4.200	105
	15-20	15	1.725	115
	trên 20	10	1.200	120
Cả tổ	-	120	12.000	100
Chưa được đào tạo kỹ thuật	dưới 5	10	510	51
	5-10	30	2.140	71
	10-15	20	1.540	79
	15-20	10	860	86
	trên 20	10	910	91
Cả tổ	-	80	6.000	75
Chung cho cả doanh nghiệp	-	200	18.000	90

II. BẢNG THỐNG KÊ (Statistical table)

Sau khi tổng hợp các tài liệu điều tra thống kê, muốn phát huy tác dụng của nó đối với phân tích thống kê, cần thiết phải trình bày kết quả tổng hợp theo một hình thức thuận lợi nhất cho việc sử dụng sau này.

1. Khái niệm

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. Đặc điểm chung của tất cả các bảng thống kê là bao giờ cũng có những con số của từng bộ phận và có mối liên hệ mật thiết với nhau.

2. Cấu thành bảng thống kê

a) Về hình thức: Bảng thống kê bao gồm các hàng, cột, các tiêu đề, tiêu mục và các con số.

Các hàng cột thể hiện qui mô của bảng, số hàng và cột càng nhiều thì bảng thống kê càng lớn và càng phức tạp.

Tiêu đề của bảng thống kê phản ánh nội dung, ý nghĩa của bảng và của từng chi tiết trong bảng. Trước hết ta có tiêu đề chung, sau đó là các tiêu đề nhỏ (tiêu mục) là tên riêng của mỗi hàng, cột phản ánh ý nghĩa của cột đó.

b) Phần nội dung: Bảng thống kê gồm 2 phần: Phần chủ đề và phần giải thích.

Phần chủ đề nói lên tổng thể được trình bày trong bảng thống kê, tổng thể này được phân thành những đơn vị, bộ phận. Nó giải đáp: đối tượng nghiên cứu là những đơn vị nào, những loại hình gì. Có khi phần chủ đề phản ánh các địa phương hoặc các thời gian nghiên cứu khác nhau của một hiện tượng.

Phần giải thích gồm các chỉ tiêu giải thích các đặc điểm của đối tượng nghiên cứu, tức là giải thích phần chủ đề của bảng.

Phần chủ đề thường được đặt bên trái của bảng thống kê, còn phần giải thích được đặt ở phía trên của bảng. Cũng có trường hợp ta thay đổi vị trí.

Cấu thành của bảng thống kê có thể biểu hiện bằng sơ đồ sau:

Phần chủ đề	Phần giải thích	Các chỉ tiêu giải thích (tên cột)			
		(2)	(3)	(4)	(5)
Tên chủ đề					

3. Các yêu cầu và qui ước xây dựng bảng thống kê

- **Quy mô của bảng thống kê:** không nên quá lớn, tức là quá nhiều hàng, cột và nhiều phân tổ kết hợp. Một bảng thống kê ngắn, gọn một cách hợp lý sẽ tạo điều kiện dễ dàng cho việc phân tích. Nếu thấy cần thiết nên xây dựng hai, ba,... bảng thống kê nhỏ thay cho một bảng thống kê quá lớn

- **Số hiệu bảng:** nhằm giúp cho người đọc dễ dàng xác định vị trí của bảng khi tham khảo, đặc biệt là đối với các tài liệu nghiên cứu người ta thường lập mục lục biểu bảng để người đọc dễ tham khảo và người trình bày dễ dàng hơn. Nếu số hiệu bảng không nhiều thì chúng ta chỉ cần đánh số theo thứ tự xuất hiện của biểu bảng, nếu tài liệu được chia thành nhiều chương và số liệu biểu bảng nhiều thì ta có thể đánh số theo chương và theo số thứ tự xuất hiện của biểu bảng trong chương. Ví dụ, Bảng II.5 tức là bảng ở chương II và là bảng thứ 5.

- **Tên bảng:** yêu cầu ngắn gọn, đầy đủ, rõ ràng, đặt trên đầu bảng và phải chứa đựng nội dung, thời gian, không gian mà số liệu được biểu hiện trong bảng. Tuy nhiên yêu cầu này chỉ mang tính chất tương đối không có tiêu chuẩn rõ ràng nhưng thông thường người ta cố gắng trình bày trong một hàng hoặc tối đa là hai hàng.

• Đơn vị tính:

- Đơn vị tính dùng chung cho toàn bộ số liệu trong bảng thống kê, trường hợp này đơn vị tính được ghi bên góc phải của bảng.

- Đơn vị tính theo từng chỉ tiêu trong cột, trong trường hợp này đơn vị tính sẽ được đặt dưới chỉ tiêu của cột.

- Đơn vị tính theo từng chỉ tiêu trong hàng, trong trường hợp này đơn vị tính sẽ được đặt sau chỉ tiêu theo mỗi hàng hoặc tạo thêm một cột ghi đơn vị tính.

• Cách ghi số liệu trong bảng:

- Số liệu trong từng hàng (cột) có đơn vị tính phải nhận cùng một số lẻ, số liệu ở các hàng (cột) khác nhau không nhất thiết có cùng số lẻ với hàng (cột) tương ứng.

- Một số ký hiệu qui ước:

- + Nếu không có tài liệu thì trong ô ghi dấu gạch ngang “-”

- + Nếu số liệu còn thiếu, sau này sẽ bổ sung sau thì trong ô ghi dấu ba chấm “...”

- + Ký hiệu gạch chéo “x” trong ô nào đó thì nói lên hiện tượng không có liên quan đến chỉ tiêu đó, nếu ghi số liệu vào đó sẽ vô nghĩa hoặc thừa.

- **Phần ghi chú ở cuối bảng:** được dùng để giải thích rõ các nội dung chỉ tiêu trong bảng, nói rõ nguồn tài liệu đã sử dụng hoặc các chỉ tiêu cần thiết khác. Đối với các tài liệu khoa học, việc ghi rõ nguồn số liệu được coi như là bắt buộc không thể thiếu được trong biểu bảng.

III. TỔNG HỢP BẢNG ĐỒ THỊ

Phương pháp đồ thị thống kê là phương pháp trình bày và phân tích các thông tin thống kê bằng các biểu đồ, đồ thị và bản đồ thống kê. Phương pháp đồ thị thống kê sử dụng con số kết hợp với các hình vẽ, đường nét và màu sắc để trình bày các đặc điểm số lượng của hiện tượng. Chính vì vậy, ngoài tác dụng phân tích giúp ta nhận thức được những đặc điểm cơ bản của hiện tượng bằng trực quan một cách dễ dàng và nhanh chóng, đồ thị thống kê còn là một phương pháp trình bày các thông tin thống kê một cách khái quát và sinh động, chứa đựng tính mỹ thuật; thu hút sự chú ý của người đọc, giúp người xem dễ hiểu, dễ nhớ nên có tác dụng tuyên truyền cổ động rất tốt. Đồ thị thống kê có thể biểu thị:

- Kết cấu của hiện tượng theo tiêu thức nào đó và sự biến đổi của kết cấu.
- Sự phát triển của hiện tượng theo thời gian.
- So sánh các mức độ của hiện tượng.
- Mối liên hệ giữa các hiện tượng.
- Trình độ phổ biến của hiện tượng.
- Tình hình thực hiện kế hoạch.

Trong công tác thống kê thường dùng các loại đồ thị: Biểu đồ hình cột, biểu đồ tượng hình, biểu đồ diện tích (hình vuông, hình tròn, hình chữ nhật), đồ thị đường gấp khúc và biểu đồ hình mảng nhện.

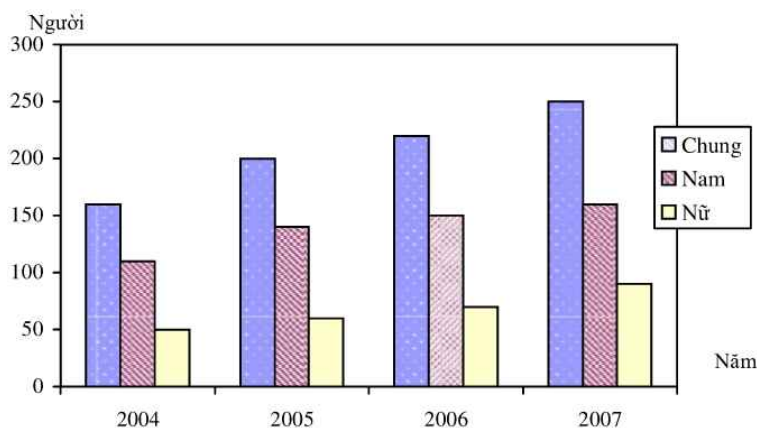
1. Biểu đồ hình cột

Biểu đồ hình cột là loại biểu đồ biểu hiện các tài liệu thống kê bằng các hình chữ nhật hay khối chữ nhật thẳng đứng hoặc nằm ngang có chiều rộng và chiều sâu bằng nhau, còn chiều cao tương ứng với các đại lượng cần biểu hiện.

Biểu đồ hình cột được dùng để biểu hiện quá trình phát triển, phản ánh cơ cấu và thay đổi cơ cấu hoặc so sánh cũng như biểu hiện mối liên hệ giữa các hiện tượng.

Ví dụ 1.6: Biểu diễn số lượng cán bộ khoa học công nghệ của một quốc gia nào đó chia theo nam nữ của 4 năm: 2004, 2005, 2006 và 2007 qua biểu đồ 1.1.

Biểu đồ 1.1: Hình cột phản ánh số lượng cán bộ khoa học công nghệ của quốc gia X, 2004 - 2007



Đồ thị trên vừa phản ánh quá trình phát triển của cán bộ khoa học công nghệ vừa so sánh cũng như phản ánh mối liên hệ giữa cán bộ là nam và nữ.

2. Biểu đồ diện tích

Biểu đồ diện tích là loại biểu đồ, trong đó các thông tin thống kê được biểu hiện bằng các loại diện tích hình học như hình vuông, hình chữ nhật, hình tròn, hình ô van,...

Biểu đồ diện tích thường được dùng để biểu hiện kết cấu và biến động cơ cấu của hiện tượng.

Tổng diện tích của cả hình là 100%, thì diện tích từng phần tương ứng với mỗi bộ phận phản ánh cơ cấu của bộ phận đó.

Biểu đồ diện tích hình tròn còn có thể biểu hiện được cả cơ cấu, biến động cơ cấu kết hợp thay đổi mức độ của hiện tượng. Trong trường hợp này số đo của góc các hình quạt phản ánh cơ cấu và biến động cơ cấu, còn diện tích toàn hình tròn phản ánh quy mô của hiện tượng.

Khi vẽ đồ thị ta tiến hành như sau:

- Lấy giá trị của từng bộ phận chia cho giá trị chung của chỉ tiêu nghiên cứu để xác định tỷ trọng (%) của từng bộ phận đó. Tiếp tục lấy 360 (360^0) chia cho 100 rồi nhân với tỷ trọng của từng bộ phận sẽ xác định được góc độ tương ứng với cơ cấu của từng bộ phận.

- Xác định bán kính của mỗi hình tròn có diện tích tương ứng là $S: R = \sqrt{S : \pi}$ vì diện tích hình tròn: $S = \pi.R^2$. Khi có độ dài của bán kính mỗi hình tròn, ta sẽ dễ dàng vẽ được các hình tròn đó.

Ví dụ 1.7: Có số lượng về học sinh phổ thông phân theo cấp học 3 năm 2005, 2006 và 2007 của địa phương X như bảng 1.3:

Bảng 1.3: Học sinh phổ thông phân theo cấp học của địa phương X, 2005 - 2007

	2005		2006		2007	
	Số lượng (Người)	Cơ cấu (%)	Số lượng (Người)	Cơ cấu (%)	Số lượng (Người)	Cơ cấu (%)
Tổng số học sinh	1.000	100,0	1.140	100,0	1.310	100,0
<i>Chia ra:</i>						
Tiểu học	500	50,0	600	53,0	700	53,5
Trung học cơ sở	300	30,0	320	28,0	360	27,5
Trung học phổ thông	200	20,0	220	19,0	250	19,0

Từ số liệu bảng 1.3 ta tính các bán kính tương ứng:

$$\text{Năm 2005: } R = \sqrt{1000 / 3,14} = 17,84$$

$$\text{Năm 2006: } R = \sqrt{1140 / 3,14} = 19,05$$

$$\text{Năm 2007: } R = \sqrt{1310 / 3,14} = 20,42$$

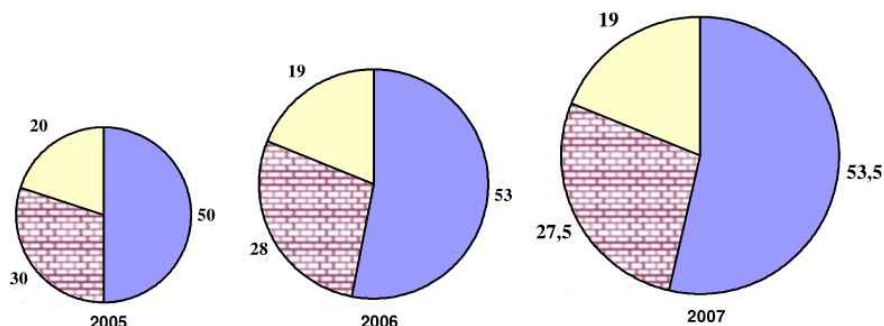
Nếu năm 2005 lấy $R = 1,00$

Thì năm 2006 có $R = 19,05 : 17,84 = 1,067$

Năm 2007 có $R = 20,42 : 17,84 = 1,144$

Kết quả 3 hình tròn được vẽ phản ánh cả quy mô học sinh phổ thông lẫn cơ cấu và biến động cơ cấu theo cấp học của học sinh qua các năm 2005, 2006 và 2007.

Biểu đồ 1.2: Biểu đồ cơ cấu học sinh phổ thông địa phương X từ 2005 – 2007



Tuy nhiên, nếu chúng ta chỉ vẽ biểu đồ mang tính đơn lẻ thì không cần phải xác định độ lớn của đường kính.

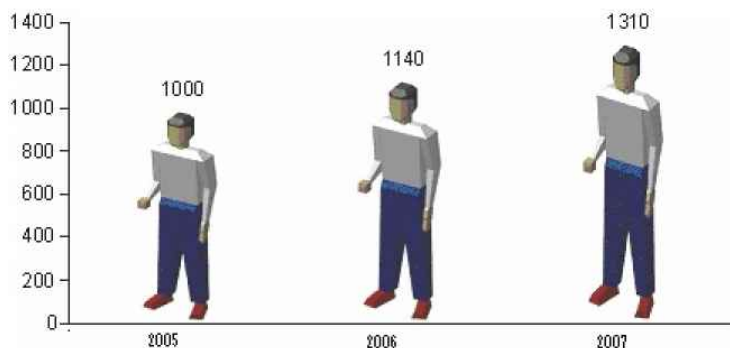
3. Biểu đồ tượng hình

Biểu đồ tượng hình là loại đồ thị thống kê, trong đó các tài liệu thống kê được thể hiện bằng các hình vẽ tượng trưng. Biểu đồ tượng hình được dùng rộng rãi trong việc tuyên truyền, phổ biến thông tin trên các phương tiện sử dụng rộng rãi. Biểu đồ hình tượng có nhiều cách vẽ khác nhau, tùy theo sáng kiến của người trình bày mà lựa chọn loại hình vẽ tượng hình cho phù hợp và hấp dẫn.

Tuy nhiên khi sử dụng loại biểu đồ này phải theo nguyên tắc: cùng một chỉ tiêu phải được biểu hiện bằng cùng một loại hình vẽ, còn chỉ tiêu đó ở các trường hợp nào có trị số lớn nhỏ khác nhau thì sẽ biểu hiện bằng hình vẽ có kích thước lớn nhỏ khác nhau theo tỷ lệ tương ứng.

Trở lại ví dụ trên số lượng học sinh phổ thông được biểu diễn như sau:

Biểu đồ 1.3: Biểu đồ cơ cấu học sinh phổ thông địa phương X từ 2005 – 2007



4. Đồ thị đường gấp khúc

Đồ thị đường gấp khúc là loại đồ thị thống kê biểu hiện các tài liệu bằng một đường gấp khúc nối liền các điểm trên một hệ tọa độ, thường là hệ tọa độ vuông góc.

Đồ thị đường gấp khúc được dùng để biểu hiện quá trình phát triển của hiện tượng, biểu hiện tình hình phân phối các đơn vị tổng thể theo một tiêu thức nào đó hoặc biểu thị tình hình thực hiện kế hoạch theo từng thời gian của các chỉ tiêu nghiên cứu.

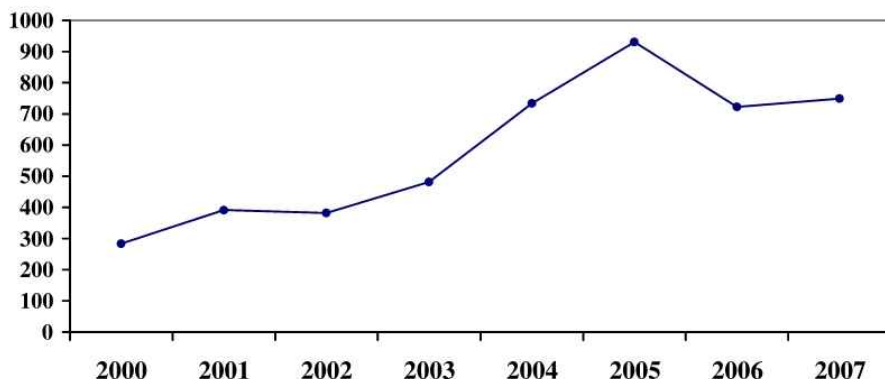
Trong một đồ thị đường gấp khúc, trục hoành thường được biểu thị thời gian, trục tung biểu thị mức độ của chỉ tiêu nghiên cứu. Cũng có khi các trục này biểu thị hai chỉ tiêu có liên hệ với nhau hoặc lượng biến và các tần số (hay tần suất) tương ứng. Độ phân chia trên các trục cần được xác định cho thích hợp vì có ảnh hưởng trực tiếp đến độ dốc của đồ thị. Mặt khác, cần chú ý là trên mỗi trục tọa độ chiều dài của các khoảng phân chia tương ứng với sự thay đổi về lượng của chỉ tiêu nghiên cứu phải bằng nhau.

Ví dụ 1.8: Sản lượng cà phê xuất khẩu của quốc gia X qua các năm từ 2000 - 2007 (ngàn tấn) có kết quả như sau:

Năm	2000	2001	2002	2003	2004	2005	2006	2007
Sản lượng (ngàn tấn)	283,3	391,6	382,0	482,0	733,9	931,0	722,0	749,0

Số liệu trên được biểu diễn qua đồ thị đường gấp khúc sau:

Đồ thị 1.4: Sản lượng cà phê xuất khẩu của quốc gia X từ 2000 – 2007



5. Biểu đồ hình màng nhện

Biểu đồ hình màng nhện là loại đồ thị thống kê dùng để phản ánh kết quả đạt được của hiện tượng lập đi lập lại về mặt thời gian, ví dụ phản ánh về biến động thời vụ của một chỉ tiêu nào đó qua 12 tháng trong năm. Để lập đồ thị hình màng nhện ta vẽ một hình tròn bán kính R, sao cho R lớn hơn trị số lớn nhất của chỉ tiêu nghiên cứu (lớn hơn bao nhiêu lần không quan trọng, miễn là đảm bảo tỷ lệ nào đó để hình vẽ được cân đối, kết quả biểu diễn của đồ thị dễ nhận biết). Sau đó chia đường tròn bán kính R thành các phần đều nhau theo số kỳ nghiên cứu (ở đây là 12 tháng) bởi các đường thẳng đi qua tâm đường tròn. Nối các giao điểm của bán kính cắt đường tròn ta được đa giác đều nội tiếp đường tròn. Đó là giới hạn phạm vi của đồ thị. Độ dài đo từ tâm đường tròn đến các điểm xác định theo các đường phân chia đường tròn nói trên chính là các đại lượng cần biểu hiện của hiện tượng tương ứng với mỗi thời kỳ. Nối các điểm xác định sẽ được hình vẽ của đồ thị hình màng nhện.

Ví dụ 1.9: Có số liệu về trị giá xuất, nhập khẩu hải sản của tỉnh X 2 năm 2006 và 2003 như sau:

Bảng 1.4. Giá trị xuất khẩu hải sản trong 12 tháng tỉnh X năm 2006 - 2007

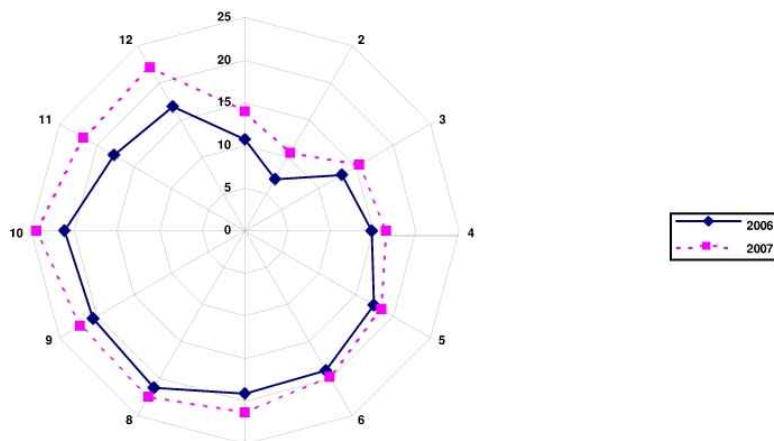
ĐVT: Triệu đồng

Tháng \ Năm		Năm		Tháng \ Năm	
		2006	2007		
1	1	10,7	14,0	7	19,1
	2	7,0	10,5	8	21,2
	3	13,1	15,4	9	20,5
	4	14,8	16,5	10	21,1
	5	17,4	18,4	11	17,7
	6	18,9	19,8	12	16,8
					22,1

Chia đường tròn thành 12 phần đều nhau, vẽ các đường thẳng tương ứng cắt đường

tròn tại 12 điểm. Nối các điểm lại có đa giác đều 12 cạnh nội tiếp đường tròn. Căn cứ số liệu của bảng ta xác định các điểm tương ứng với giá trị xuất khẩu đạt được của các tháng trong từng năm rồi nối các điểm đó lại thành đường liền ta được đồ thị hình màng nhện biểu diễn kết quả xuất khẩu qua các tháng trong 2 năm của tỉnh X.

Đồ thị 1.5: Đồ thị Giá trị xuất khẩu hải sản trong 12 tháng tỉnh X năm 2006 - 2007



Sự mô tả của đồ thị hình màng nhện cho phép ta quan sát và so sánh không chỉ kết quả xuất khẩu giữa các tháng khác nhau trong cùng một năm, mà cả kết quả sản xuất giữa các tháng cùng tên của các năm khác nhau cũng như xu thế biến động chung về xuất khẩu của các năm.

CHƯƠNG IV

CÁC MỨC ĐỘ CỦA HIỆN TƯỢNG KINH TẾ XÃ HỘI

Nghiên cứu các mức độ của hiện tượng kinh tế xã hội là yêu cầu quan trọng của việc tổng hợp, tính toán và phân tích thống kê nhằm biểu hiện mật lượng trong quan hệ mật thiết với mật chất của hiện tượng nghiên cứu trong điều kiện thời gian và không gian cụ thể nhờ vào sự trợ giúp của các phương pháp thống kê.

Dưới đây là nội dung, phương pháp tính và điều kiện vận dụng của các đại lượng đó.

I. SỐ TUYỆT ĐỐI

Số tuyệt đối là chỉ tiêu biểu hiện quy mô, khối lượng của hiện tượng hoặc quá trình kinh tế - xã hội trong điều kiện thời gian và không gian cụ thể.

Số tuyệt đối trong thống kê bao gồm các con số phản ánh quy mô của tổng thể hay của từng bộ phận trong tổng thể (số doanh nghiệp, số nhân khẩu, số học sinh đi học, số lượng cán bộ khoa học,...) hoặc tổng các trị số theo một tiêu thức nào đó (tiền lương của công nhân, giá trị sản xuất công nghiệp, tổng sản phẩm trong nước (GDP), v.v...).

Số tuyệt đối dùng để đánh giá và phân tích thống kê, là căn cứ không thể thiếu được trong việc xây dựng chiến lược phát triển kinh tế, tính toán các mật cân đối, nghiên cứu các mối quan hệ kinh tế - xã hội, là cơ sở để tính toán các chỉ tiêu tương đối và bình quân.

Có hai loại số tuyệt đối: Số tuyệt đối thời kỳ và số tuyệt đối thời điểm.

Số tuyệt đối thời kỳ: Phản ánh quy mô, khối lượng của hiện tượng trong một thời kỳ nhất định. Ví dụ: Giá trị sản xuất công nghiệp trong 1 tháng, quý hoặc năm; Sản lượng lương thực năm 2005, năm 2006, năm 2007,...

Số tuyệt đối thời điểm: Phản ánh quy mô, khối lượng của hiện tượng ở một thời điểm nhất định như: dân số của một địa phương nào đó có đến 0 giờ ngày 01/04/2005; giá trị tài sản cố định có đến 31/12/2007; lao động làm việc của doanh nghiệp vào thời điểm 1/7/2007,...

II. SỐ TƯƠNG ĐỐI

Số tương đối là chỉ tiêu biểu hiện quan hệ so sánh giữa hai chỉ tiêu thống kê cùng loại nhưng khác nhau về thời gian hoặc không gian hoặc giữa hai chỉ tiêu khác loại nhưng có quan hệ với nhau. Trong hai chỉ tiêu để so sánh của số tương đối, sẽ có một số được chọn làm gốc (chuẩn) để so sánh.

Số tương đối có thể được biểu hiện bằng số lần, số phần trăm (%) hoặc phần nghìn (‰), hay bằng các đơn vị kép (người/km², người/1000 người; đồng/1000đồng, ...).

Trong công tác thống kê, số tương đối được sử dụng rộng rãi để phản ánh những đặc điểm về kết cấu, quan hệ tỷ lệ, tốc độ phát triển, mức độ hoàn thành kế hoạch, mức độ phổ biến của hiện tượng kinh tế - xã hội được nghiên cứu trong điều kiện thời gian và không gian nhất định.

Số tương đối phải được vận dụng kết hợp với số tuyệt đối. Số tương đối thường là kết quả của việc so sánh giữa hai số tuyệt đối. Số tương đối tính ra có thể rất khác nhau, tùy thuộc vào việc lựa chọn gốc so sánh. Có khi số tương đối có giá trị rất lớn nhưng ý nghĩa của nó không đáng kể vì trị số tuyệt đối tương ứng của nó lại rất nhỏ. Ngược lại, có số tương đối tính ra khá nhỏ nhưng lại mang ý nghĩa quan trọng vì trị số tuyệt đối tương ứng của nó có quy mô đáng kể. Ví dụ: 1% dân số Việt Nam tăng lên trong những năm 1960 đồng nghĩa với dân số tăng thêm 300 nghìn người, nhưng 1% dân số tăng lên trong những năm 2000 lại đồng nghĩa với dân số tăng thêm 800 nghìn người.

Căn cứ vào nội dung mà số tương đối phản ánh, có thể phân biệt: số tương đối động

thái, số tương đối kế hoạch, số tương đối kết cấu, số tương đối cường độ, và số tương đối không gian.

1. Số tương đối động thái

Số tương đối động thái là chỉ tiêu phản ánh biến động theo thời gian về mức độ của chỉ tiêu kinh tế - xã hội. Số tương đối này tính được bằng cách so sánh hai mức độ của chỉ tiêu được nghiên cứu ở hai thời gian khác nhau. Mức độ của thời kỳ được tiến hành nghiên cứu thường gọi là mức độ của kỳ báo cáo, còn mức độ của một thời kỳ nào đó được dùng làm cơ sở so sánh thường gọi là mức độ kỳ gốc.

Trong hai mức độ đó, mức độ từ số (y_1) là mức độ cần nghiên cứu (hay còn gọi là mức độ kỳ báo cáo), mức độ ở mẫu số (y_0) là mức độ kỳ gốc (hay mức độ dùng làm cơ sở so sánh).

- Nếu y_0 cố định qua các kỳ nghiên cứu ta có kỳ gốc cố định: dùng để so sánh một chỉ tiêu nào đó ở hai thời kỳ tương đối xa nhau. Thông thường người ta chọn năm gốc là năm đầu tiên của dãy số.

- Nếu y_0 thay đổi theo kỳ nghiên cứu ta có kỳ gốc liên hoàn: dùng để nói lên sự biến động của hiện tượng liên tiếp nhau qua các kỳ nghiên cứu.

Ví dụ 2.1: Sản lượng hàng hóa tiêu thụ (1.000 tấn) của một công ty X qua các năm như sau:

Năm	2001	2002	2003	2004	2005	2006	2007
Sản lượng hàng hóa (1.000 tấn)	240,0	259,2	282,5	299,5	323,4	355,8	387,8
Tốc độ phát triển liên hoàn (lần)		1,08	1,09	1,06	1,08	1,10	1,09

- Mỗi liên hệ giữa tốc độ phát triển định gốc và tốc độ phát triển liên hoàn. Nếu ta có dãy số sau:

Thời kỳ	1	2	3	...	n-1	n
y_i	y_1	y_2	y_3	...	y_{n-1}	y_n

thì mối liên hệ giữa tốc độ phát triển định gốc và tốc độ phát triển liên hoàn được thể hiện

qua công thức sau: $\frac{y_2 y_3 \dots y_n}{y_1 y_2 \dots y_{n-1}} = \frac{y_n}{y_1}$.

2. Số tương đối so sánh

Số tương đối so sánh là chỉ tiêu phản ánh quan hệ so sánh giữa hai bộ phận trong một tổng thể hoặc giữa hai hiện tượng cùng loại nhưng khác nhau về điều kiện không gian. Ví dụ: Dân số thành thị so với dân số nông thôn, dân số là nam so với dân số là nữ; giá trị tăng thêm của doanh nghiệp ngoài quốc doanh so với giá trị tăng thêm của doanh nghiệp quốc doanh; năng suất lúa của tỉnh X so với năng suất lúa của tỉnh Y; số học sinh đạt kết quả học tập khá giỏi so với số học sinh đạt kết quả trung bình,...

3. Số tương đối kế hoạch

Số tương đối kế hoạch là chỉ tiêu phản ánh mức cần đạt tới trong kỳ kế hoạch hoặc mức đã đạt được so với kế hoạch được giao về một chỉ tiêu kinh tế - xã hội nào đó. Số tương đối kế hoạch được chia thành hai loại:

+ Số tương đối nhiệm vụ kế hoạch: Phản ánh quan hệ so sánh giữa mức độ đề ra trong kỳ kế hoạch với mức độ thực tế ở kỳ gốc của một chỉ tiêu kinh tế - xã hội.

$$KH = \frac{\text{Mức cần đạt}}{\text{Mức thực tế kỳ gốc}} \times 100 = \frac{y_{KH}}{y_0} \times 100$$

+ Số tương đối hoàn thành kế hoạch: Phản ánh quan hệ so sánh giữa mức thực tế đã

đạt được với mức kế hoạch trong kỳ về một chỉ tiêu kinh tế - xã hội.

$$HT = \frac{\text{Mức thực hiện}}{\text{Mức kế hoạch}} \times 100 = \frac{Y_I}{Y_{KH}} \times 100$$

Mối liên hệ giữa số tương đối động thái và số tương đối kế hoạch:

$$\frac{Y_I}{Y_0} = \frac{Y_{KH}}{Y_0} \times \frac{Y_I}{Y_{KH}}$$

4. Số tương đối kết cấu

Số tương đối kết cấu là chỉ tiêu phản ánh tỷ trọng của mỗi bộ phận chiếm trong tổng thể, tính được bằng cách đem so sánh mức độ tuyệt đối của từng bộ phận với mức độ tuyệt đối của toàn bộ tổng thể.

Số tương đối kết cấu thường được biểu hiện bằng số phần trăm. Ví dụ: Tỷ trọng của GDP theo từng ngành trong tổng GDP của nền kinh tế quốc dân; tỷ trọng dân số của từng giới nam hoặc nữ trong tổng số dân,...

$$\text{Số tương đối kết cấu} = \frac{\text{Số tuyệt đối của bộ phận}}{\text{Số tuyệt đối của tổng thể}} \times 100$$

Ví dụ 2.2: Trong công ty A có 500 công nhân, trong đó có 300 công nhân nam và 200 công nhân nữ.

$$\text{Tỷ trọng nam trong tổng công nhân} = \frac{300}{500} \cdot 100\% = 60\%$$

$$\text{Tỷ trọng nữ trong tổng công nhân} = \frac{200}{500} \cdot 100\% = 40\%$$

5. Số tương đối cường độ

Số tương đối cường độ là chỉ tiêu biểu hiện trình độ phổ biến của một hiện tượng trong các điều kiện thời gian và không gian cụ thể.

Số tương đối cường độ tính được bằng cách so sánh mức độ của hai chỉ tiêu khác nhau nhưng có quan hệ với nhau. Số tương đối cường độ biểu hiện bằng đơn vị kép, do đơn vị tính ở tử số và ở mẫu số hợp thành. Số tương đối cường độ được tính toán và sử dụng rất phổ biến trong công tác thống kê. Các số tương đối trong số liệu thống kê thường gặp như mật độ dân số bằng tổng số dân (người) chia cho diện tích tự nhiên (km²) với đơn vị tính là người /km²; GDP bình quân đầu người bằng tổng GDP (nghìn đồng) chia cho dân số trung bình (người) với đơn vị tính là 1000đ/người; số bác sĩ tính bình quân cho một vạn dân bằng tổng số bác sĩ chia cho tổng số dân tính bằng vạn người với đơn vị tính là người /10.000 người,...

$$\text{Mật độ dân số} = \frac{\text{Số dân}}{\text{Diện tích}} \text{ (người/Km}^2\text{)}$$

$$\text{Năng suất lao động} = \frac{\text{Tổng giá phẩm}}{\text{Tổng số lao động}} \text{ (Giá phẩm / người)}$$

III. SỐ ĐO ĐỘ TẬP TRUNG – SỐ BÌNH QUÂN (Measures of central tendency):

Số bình quân là chỉ tiêu biểu hiện mức độ điển hình của một tổng thể gồm nhiều đơn vị cùng loại được xác định theo một tiêu thức nào đó. Số bình quân được sử dụng phổ biến trong thống kê để nêu lên đặc điểm chung nhất, phổ biến nhất của hiện tượng kinh tế - xã hội trong các điều kiện thời gian và không gian cụ thể. Ví dụ: Tiền lương bình quân một công nhân trong doanh nghiệp là mức lương phổ biến nhất, đại diện cho các mức lương khác nhau của công nhân trong doanh nghiệp; thu nhập bình quân đầu người của

một địa bàn là mức thu nhập phổ biến nhất, đại diện cho các mức thu nhập khác nhau của mọi người trong địa bàn đó.

Số bình quân còn dùng để so sánh đặc điểm của những hiện tượng không có cùng một quy mô hay làm căn cứ để đánh giá trình độ đồng đều của các đơn vị tổng thể.

Xét theo vai trò đóng góp khác nhau của các thành phần tham gia bình quân hoá, số bình quân chung được chia thành số bình quân giản đơn và số bình quân gia quyền.

+ Số trung bình giản đơn: Được tính trên cơ sở các thành phần tham gia bình quân hoá có vai trò về qui mô (tần số) đóng góp như nhau.

+ Số trung bình gia quyền (trung bình có trọng số): Được tính trên cơ sở các thành phần tham gia bình quân hoá có vai trò về qui mô (tần số) đóng góp khác nhau.

Để tính được số trung bình chính xác và có ý nghĩa, điều kiện chủ yếu là nó phải được tính cho những đơn vị cùng chung một tính chất (thường gọi là tổng thể đồng chất). Muốn vậy, phải dựa trên cơ sở phân tổ thống kê một cách khoa học và chính xác. Đồng thời phải vận dụng kết hợp giữa số bình quân tổ với số bình quân chung.

Có nhiều loại số bình quân khác nhau. Trong thống kê kinh tế - xã hội thường dùng các loại sau: Số trung bình số học, số trung bình điều hoà, số trung bình hình học (số trung bình nhân), mốt và trung vị.

Dưới đây là từng loại số bình quân nêu trên.

1. Số trung bình cộng (Mean)

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

x_i : Giá trị lượng biến quan sát

n : Số quan sát

2. Số trung bình gia quyền (Weighted mean)

Với mỗi lượng biến x_i có tần số tương ứng f_i , số trung bình được xác định theo công thức sau:

$$\bar{x} = \frac{\sum_{i=1}^k x_i f_i}{\sum_{i=1}^k f_i}$$

x_i : Giá trị lượng biến quan sát

f_i : Tần số lượng biến quan sát

Ví dụ 2.3: Có tài liệu về mức thu nhập của các hộ theo tháng

Thu nhập hàng tháng (triệu đồng)	Số hộ
5.000	3
5.250	8
5.400	9
5.450	10
5.600	12
6.000	30
6.200	15
6.300	7
6.500	6
Tổng	100

Ta lập bảng tổng thu nhập hàng tháng của các hộ

Thu nhập hàng tháng (triệu đồng) (x_i)	Số hộ (f_i)	$x_i f_i$
5.000	3	15.000
5.250	8	42.000
5.400	9	48.600
5.450	10	54.500
5.600	12	67.200
6.000	30	180.000
6.200	15	93.000
6.300	7	44.100
6.500	6	39.000
Tổng	100	583.400

Ví dụ 2.4: Có số liệu thu nhập hàng tháng (ngàn đồng) của nhân viên một công ty như sau:

Thu nhập hàng tháng (ngàn đồng)	Số nhân viên
500-520	8
520-540	12
540-560	20
560-580	56
580-600	18
600-620	16
Trên 620	10
Tổng	140

Chú ý, trường hợp dãy số được phân tổ thì lượng biến x_i là trị số giữa của các tổ. Nếu dãy số có tổ mở thì lấy khoảng cách tổ của tổ gần tổ mở nhất để tính giới hạn trên của tổ mở từ đó xác định được giá trị x_i .

- Đối với tổ không có giới hạn trên: giới hạn dưới của tổ mở cộng với khoảng cách tổ của tổ trước đó mở rồi chia hai.

- Đối với tổ không có giới hạn dưới: giới hạn trên của tổ mở trừ khoảng cách tổ của tổ sau đó mở rồi chia hai. Tùy theo tính chất của nội dung nghiên cứu mà có thể chọn giá trị x_i phù hợp.

Từ bảng trên ta có bảng sau:

Thu nhập hàng tuần (1.000đ)	x_i	f_i	$x_i f_i$
500-520	510	8	4.080
520-540	530	12	6.360
540-560	550	20	11.000
560-580	570	56	31.920
580-600	590	18	10.620
600-620	610	16	9.760
Trên 620	630	10	6.300
Tổng		140	80.040

Áp dụng công thức ta có:

$$\bar{x} = \frac{80.040}{140} = 571,71$$

Tuy nhiên, việc ước lượng các giá trị x_i có chính xác hay không còn phụ thuộc vào phân phối của từng tổ. Nếu phân phối của từng tổ có tính chất đối xứng thì việc ước lượng x_i có thể chấp nhận được, tuy nhiên đối với các trường hợp phân phối của tổ lệch trái hoặc lệch phải thì kết quả đó khó có thể chấp nhận được. Do đó, trong quá trình tính toán với sự hỗ trợ của các phần mềm máy tính ta nên sử dụng số liệu điều tra và tính với công thức trung bình đơn giản để đảm bảo tính chính xác.

3. Số trung bình điều hòa (Harmonic mean)

Số trung bình điều hòa được sử dụng trong trường hợp biết các lượng biến x_i và tích $x_i f_i$ mà chưa biết tần số f_i .

$$\bar{x} = \frac{\sum_{i=1}^k M_i}{\sum_{i=1}^k \frac{M_i}{x_i}}, M_i = x_i f_i$$

Ví dụ 2.5: Có số liệu giá thành sản và chi phí sản xuất của 3 phân xưởng của một doanh nghiệp:

Phân xưởng	Giá thành 1 tấn sản phẩm (1.000đ)	Chi phí sản xuất(1.000đ)
Số 1	18,5	740
Số 2	19,0	855
Số 3	19,4	970

Đặt x_i : Giá thành của phân xưởng i

M_i : Chi phí của phân xưởng i

Giá thành trung bình một tấn sản phẩm của doanh nghiệp được xác định bởi công thức:

$$\bar{x} = \frac{740 + 855 + 970}{\frac{740}{18,5} + \frac{855}{19,0} + \frac{970}{19,4}} = 19 \text{ (1.000\text{đ})}$$

4. Số trung bình nhân (Geometric mean)

Số trung nhân hay số trung bình hình học sử dụng để tính tốc độ phát triển trung bình nói riêng và dùng để tính số trung bình trong trường hợp các giá trị x_i có mối liên hệ tích.

$$\bar{x} = \sqrt[n]{x_1 \cdot x_2 \cdot \dots \cdot x_n} = \sqrt[n]{\prod_{i=1}^n x_i}$$

Ví dụ 2.6: Hãy tính tốc độ phát triển sản lượng hàng hóa tiêu thụ (1.000 tấn) của một công ty qua các năm như sau:

Năm	2001	2002	2003	2004	2005	2006	2007
Sản lượng hàng hóa (1.000 tấn)	240,0	259,2	282,5	299,5	323,4	355,8	387,8
Tốc độ phát triển liên hoàn (lần)		1,08	1,09	1,06	1,08	1,10	1,09

Giữa các tốc độ phát triển liên hoàn có mối quan hệ nhân, do đó ta áp dụng công thức trung bình nhân:

$$\bar{x} = \sqrt[6]{x_1 \cdot x_2 \cdot x_3 \cdot x_4 \cdot x_5 \cdot x_6} = \sqrt[6]{1,6158} = 1,08$$

Như vậy, trung bình mỗi một năm sản lượng hàng hóa năm sau sẽ bằng 1,08 lần năm trước.

5. Số trung vị - Me (Median)

Trong một số trường hợp đặc biệt, nếu dữ liệu có sự biến động lớn hay có sự chênh lệch bất thường thì số trung bình tỏ ra không đại diện cho tổng thể vì các giá trị quá nhỏ hay quá lớn sẽ làm lệch kết quả của số trung bình. Số trung vị là một giá trị bình quân có vẻ đại diện tốt hơn cho số trung bình trong trường hợp này, bởi vì nó sẽ chia tổng thể ra thành hai nhóm có số quan sát bằng nhau: một nhóm có giá trị nhỏ hơn, một nhóm có giá trị lớn hơn.

5.1. Định nghĩa: Số trung vị là lượng biến đứng ở vị trí giữa trong dãy số đã được sắp xếp theo thứ tự tăng dần hay giảm dần.

5.2. Phương pháp xác định số trung vị:

Trước tiên ta sắp xếp lượng biến theo thứ tự tăng dần.

• Tài liệu không phân tổ:

- Trường hợp n lẻ: số trung vị là lượng biến ở vị trí thứ $(n+1)/2$

$$Me = x_{(n+1)/2}$$

- Trường hợp n chẵn: số trung vị rơi vào giữa hai lượng biến $x_{n/2}$ và $x_{(n+2)/2}$. Trường hợp này qui ước số trung vị là trung bình cộng của hai lượng biến đó.

Ví dụ 2.7: thu nhập hàng tháng của số công nhân sau:

500, 520, 530, 550, 560, 570, 590, 600, 610, 670

Số trung vị là: $Me = (560+570)/2 = 565$

• Tài liệu phân tổ có khoảng cách tổ:

Trong trường hợp này ta tìm tổ chứa số trung vị. Trước hết ta tính $((f_i/2))$ và đem so sánh với tần số tích lũy của tổ. Giá trị $((f_i/2))$ thuộc tổ nào thì tổ đó chứa số trung vị.

$$Me = x_{Me(min)} + k_{Me} \frac{\sum f_i / 2 - S_{Me-1}}{f_{Me}}$$

$x_{Me(min)}$: Giới hạn dưới của tổ chứa số trung vị

k_{Me} : Trị số khoảng cách tổ chứa số trung vị

f_{Me} : Tần số của tổ chứa số trung vị

S_{Me-1} : Tần số tích lũy trước tổ chứa số trung vị

Ví dụ 2.8: Sử dụng số liệu của ví dụ trước ta tìm số trung vị. Ta có bảng:

Thu nhập hàng tháng (ngàn đồng)	Số nhân viên	Tần số tích lũy
500-520	8	8
520-540	12	20
540-560	20	40= S_{Me-1}
560-580	56= f_{Me}	96
580-600	18	114
600-620	16	130
Trên 620	10	140
Tổng	140	

Như vậy số trung vị rơi vào tổ: 560-580

$$X_{Me(min)} = 560$$

$$f_{Me} = 56$$

$$S_{Me-1} = 40$$

Thay vào công thức, ta có:

$$Me = 560 + 20 \frac{\sum 140/2 - 40}{56} = 570,714$$

6. Mốt – Mo (mode)

• **Định nghĩa:** Mốt là lượng biến có tần số xuất hiện lớn nhất trong tổng thể. Số Mo là giá trị thể hiện tính phổ biến của hiện tượng, tức là dữ liệu tập trung nhiều ở một khoảng giá trị nào đó. Trong thực tế người ta có thể sử dụng giá trị này trong sản xuất giày, quần áo may sẵn,...

• **Phương pháp xác định Mo:**

Ta phân biệt 2 trường hợp:

- Trường hợp tài liệu phân tổ không có khoảng cách tổ: (Phân tổ thuộc tính) thì đại lượng là Mo lượng biến có tần số lớn nhất.

- Trường hợp tài liệu phân tổ có khoảng cách tổ: trước hết ta xác định tổ chứa M_o , tổ chứa M_o là tổ có tần số lớn nhất, sau đó trị số gần đúng của M_o được xác định theo công thức sau:

$$M_o = x_{M_o(\min)} + k_{M_o} \frac{f_{M_o} - f_{M_o-1}}{(f_{M_o} - f_{M_o-1}) + (f_{M_o} - f_{M_o+1})}$$

$x_{M_o(\min)}$: Giới hạn dưới của tổ chứa M_o

f_{M_o} : Tần số của tổ chứa M_o

f_{M_o-1} : Tần số của tổ đứng trước tổ chứa M_o

f_{M_o+1} : Tần số của tổ đứng sau tổ chứa M_o

k_{M_o} : Trị số khoảng cách tổ chứa M_o

Trở lại ví dụ trước ta tính M_o về thu nhập:

$$M_o = 560 + 20 \frac{56 - 20}{(56 - 20) + (56 - 18)} = 569,73$$

Chúng ta đã nghiên cứu các số đo tập trung biểu thị khuynh hướng tập trung của tổng thể, tức là nghiên cứu đại lượng mang tính chất đại diện cho tổng thể. Không có một số đo duy nhất nào có thể mô tả một cách đầy đủ cho một tổng thể. Tùy theo mục đích nghiên cứu ta cần xem xét để vận dụng các số đo cho phù hợp. Tuy nhiên, trong thực tế số trung bình được sử dụng rộng rãi vì dựa vào số trung bình người ta phát triển nhiều cơ sở suy luận để xây dựng các lý thuyết và tính các số đo khác.

IV. SỐ ĐO ĐỘ PHÂN TÁN (Measure of dispersion)

Độ biến thiên của tiêu thức dùng để đánh giá mức độ đại diện của số bình quân đối với tổng thể được nghiên cứu. Trị số này tính ra càng lớn, độ biến thiên của tiêu thức càng lớn do đó mức độ đại diện của số bình quân đối với tổng thể càng thấp và ngược lại.

Quan sát độ biến thiên tiêu thức trong dãy số lượng biến sẽ thấy nhiều đặc trưng về phân phối, kết cấu, tính đồng đều của tổng thể.

Độ biến thiên của tiêu thức được sử dụng nhiều trong nghiên cứu thống kê như phân tích biến thiên cũng như mối liên hệ của hiện tượng, dự đoán thống kê, điều tra chọn mẫu,...

Khi nghiên cứu độ biến thiên của tiêu thức, thống kê thường dùng các chỉ tiêu như khoảng biến thiên, độ lệch tuyệt đối bình quân, phương sai, độ lệch tiêu chuẩn và hệ số biến thiên. Dưới đây là nội dung và phương pháp tính của các chỉ tiêu đó.

1. Khoảng biến thiên (Range)

Khoảng biến thiên (còn gọi là toàn cự) là chỉ tiêu được tính bằng hiệu số giữa lượng biến lớn nhất và lượng biến nhỏ nhất của một dãy số lượng biến. Khoảng biến thiên càng lớn, mức độ biến động của chỉ tiêu càng lớn. Ngược lại, khoảng biến thiên nhỏ, mức độ biến động của chỉ tiêu thấp, tức là mức độ đồng đều của chỉ tiêu cao.

Công thức:

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

Trong đó:

R - Toàn cự;

$X_{\max}$ - Lượng biến có trị số lớn nhất

$X_{\min}$ - Lượng biến có trị số nhỏ nhất

Ví dụ 2.9: Thu nhập của hộ gia đình như sau:

Hộ	1	2	3	4	5	6	7	8
Thu nhập (1000 đồng)	6.000	7.000	85.000	86.000	9.000	9.100	9.500	10.000

Từ số liệu bảng, sử dụng công thức ở trên ta tính được khoảng biến thiên:

$$R = 10.000 - 6.000 = 4000 \text{ (nghìn đồng)}$$

Khoảng biến thiên phản ánh khoảng cách biến động của tiêu thức tuy tính toán đơn giản song phụ thuộc vào lượng biến lớn nhất và nhỏ nhất của tiêu thức, tức là không tính gì đến mức độ khác nhau của các lượng biến còn lại trong dãy số.

2. Độ lệch tuyệt đối trung bình (Mean Absolute Deviation)

Độ lệch tuyệt đối bình quân là số bình quân số học của các độ lệch tuyệt đối giữa các lượng biến với số bình quân số học của các lượng biến đó.

Công thức:

Trường hợp tính giản đơn
$$\bar{d} = \frac{\sum_{i=1}^n |x_i - \bar{x}|}{n};$$

Trường hợp có quyền số
$$\bar{d} = \frac{\sum_{i=1}^k |x_i - \bar{x}| f_i}{\sum_{i=1}^k f_i};$$

Trong đó:

$\bar{d}$ - Độ lệch tuyệt đối bình quân;

x_i - Các trị số của lượng biến;

$\bar{x}$ - Số trung bình số học;

f_i - Quyền số của từng lượng biến x_i ;

n - Tổng số lượng biến ($n = \sum_{i=1}^k f_i$).

Chỉ tiêu này biểu hiện độ biến thiên của tiêu thức nghiên cứu một cách đầy đủ hơn khoảng biến thiên. Qua đó phản ánh rõ nét hơn tính chất đồng đều của tổng thể: vì nó tính đến độ lệch của tất cả các lượng biến. Về cách tính cũng tương đối đơn giản, nhưng có đặc điểm là phải lấy giá trị tuyệt đối (giá trị dương) của chênh lệch.

Ví dụ 2.10: Có số liệu về năng suất lao động năm của công nhân trong một doanh nghiệp:

STT	Năng suất lao động năm (Triệu đồng /người)	Số công nhân (Ngàn người)	STT	Năng suất lao động năm (Triệu đồng /người)	Số công nhân (Ngàn người)
A	1	2	A	1	2
1	10	10	4	25	10
2	15	20	5	35	10
3	20	50			

a. Số bình quân

$$\bar{x} = \frac{(10.10) + (15.20) + (20.50) + (25.10) + (35.10)}{10 + 20 + 50 + 10 + 10} = 20$$

b. Độ lệch tuyệt đối bình quân

$$\begin{aligned} \bar{d} &= \frac{|10 - 20|10 + |15 - 20|20 + |20 - 20|50 + |25 - 20|10 + |35 - 20|10}{10 + 20 + 50 + 10 + 10} \\ &= \frac{400}{100} = 4 \end{aligned}$$

3. Phương sai (Variance)

Phương sai là số bình quân số học của bình phương các độ lệch giữa các lượng biến với số bình quân số học của các lượng biến đó.

Phương sai là sai số trung bình bình phương giữa các lượng biến và số trung bình số học của các lượng biến đó.

3.1. Phương sai tổng thể:

$$\sigma_x^2 = \frac{\sum_{i=1}^N (x_i - \mu)^2}{N}$$

x_i : Giá trị lượng biến thứ i

μ : Trung bình tổng thể

N : Số đơn vị tổng thể

3.2. Phương sai mẫu:

$$S_x^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1} = \frac{\sum_{i=1}^n x_i^2 - n\bar{x}^2}{n - 1}$$

Nếu dãy số có tần số f_i thì:

$$S_x^2 = \frac{\sum_{i=1}^k (x_i - \bar{x})^2 f_i}{\sum_{i=1}^k f_i - 1}$$

Trong công thức phương sai mẫu người ta gọi tử số là tổng độ lệch bình phương và mẫu số là bậc tự do.

Chú ý, đối với công thức phương sai mẫu, theo toán học người ta chia ra thành 2 loại là phương sai mẫu và phương sai mẫu điều chỉnh. Tuy nhiên phương sai mẫu (bậc tự do là n) là ước lượng chệch của phương sai của tổng thể, còn phương sai mẫu là ước lượng không chệch. Chính vì vậy, để cho đơn giản chúng ta hiểu phương sai mẫu ở đây là phương sai mẫu đã điều chỉnh theo quan điểm của toán học.

4. Độ lệch chuẩn (Standard deviation)

4.1. Độ lệch chuẩn của tổng thể:

$$\sigma_x = \sqrt{\sigma_x^2} = \sqrt{\frac{\sum (x_i - \mu)^2}{N}}$$

4.2. Độ lệch chuẩn của mẫu:

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

5. Hệ số biến thiên (Coefficient of Variation)

Hệ số biến thiên là chỉ tiêu tương đối phản ánh mối quan hệ so sánh giữa độ lệch chuẩn với số bình quân số học.

Công thức:

$$V = \frac{\sigma}{\bar{x}}$$

Trong đó:

V - Hệ số biến thiên;

σ - Độ lệch chuẩn;

$\bar{x}$ - Số bình quân số học.

Hệ số biến thiên cũng dùng để đánh giá độ biến thiên của tiêu thức và tính chất đồng đều của tổng thể. Hệ số này biểu hiện bằng số tương đối nên còn có thể được dùng để so sánh cả những chỉ tiêu cùng loại nhưng ở các quy mô khác nhau như so sánh độ đồng đều về thu nhập bình quân của hộ gia đình ở khu vực nông thôn (có thu nhập thấp và số hộ ít hơn) với thu nhập bình quân của hộ gia đình ở thành thị (có mức thu nhập cao hơn và số hộ nhiều hơn), đặc biệt để so sánh được những chỉ tiêu của các hiện tượng khác nhau và có đơn vị đo lường khác nhau như so sánh hệ số biến thiên về bậc thợ với hệ số biến thiên về tiền lương bình quân, hệ số biến thiên về năng suất lao động bình quân, so sánh hệ số biến thiên về chỉ tiêu thu nhập của hộ gia đình với hệ số biến thiên về chỉ tiêu của hộ gia đình,...

Hệ số biến thiên còn có thể tính theo độ lệch tuyệt đối bình quân, nhưng hệ số biến thiên tính theo độ lệch chuẩn thường được sử dụng rộng rãi hơn, tuy phần tính toán có phức tạp hơn phải sử dụng độ lệch tuyệt đối trung bình.

Hệ số biến thiên tính theo độ lệch tuyệt đối bình quân có công thức tính:

$$V = \frac{\bar{d}}{\bar{x}}$$

Trong đó: $\bar{d}$ - Độ lệch tuyệt đối bình quân.

V. PHƯƠNG PHÁP CHỈ SỐ

Hiện nay các nhà doanh nghiệp có thể nắm bắt thông tin trên nhiều phương tiện thông tin khác nhau, họ quan tâm đến giá cả hay khối lượng sản phẩm từng mặt hàng hay nhiều mặt hàng tăng lên hay giảm xuống qua thời gian trên một thị trường hay nhiều thị trường khác nhau. Những thông tin này được tính toán thông qua phương pháp chỉ số. Một cách tổng quát, chỉ số đo lường sự thay đổi của hiện tượng kinh tế qua hai thời gian và không gian nghiên cứu.

Một số ký hiệu thường dùng trong phương pháp chỉ số:

p: Giá hàng hóa nói chung

z: Giá thành

q: Khối lượng sản phẩm (chỉ tiêu số lượng)

i: Chỉ số cá thể

I: Chỉ số chung, chỉ số tổng hợp

(0): Thể hiện kỳ gốc

(1): Thể hiện kỳ báo cáo hay kỳ nghiên cứu

Các loại chỉ số và cách tính:

Căn cứ vào phạm vi tính toán, có 2 loại chỉ số tương ứng với việc nghiên cứu hai loại chỉ tiêu chất lượng và số lượng.

- Chỉ tiêu chất lượng: giá cả, giá thành sản phẩm, năng suất thu hoạch, năng suất lao động, mức nguyên liệu cần thiết để sản xuất một thành phẩm,...

- Chỉ tiêu khối lượng: lượng hàng hóa tiêu thụ, lượng hàng hóa sản xuất, số lượng lao động,...

1. Chỉ số cá thể

Là loại chỉ số nghiên cứu sự biến động về một chỉ tiêu nào đó của từng loại đơn vị, từng phần tử của hiện tượng.

- Chỉ số cá thể giá cả (chỉ số chất lượng):

$$i_{pi} = \frac{p_{i(1)}}{p_{i(0)}}$$

$p_{i(1)}$: là giá cả mặt hàng thứ i kỳ nghiên cứu

$p_{i(0)}$: là giá cả mặt hàng thứ i kỳ gốc.

- Chỉ số cá thể khối lượng:

$$i_{qi} = \frac{q_{i(1)}}{q_{i(0)}}$$

$q_{i(1)}$: là khối lượng mặt hàng thứ i kỳ nghiên cứu

$q_{i(0)}$: là khối lượng mặt hàng thứ i kỳ gốc.

2. Chỉ số tổng hợp

Là loại chỉ tiêu nghiên cứu sự biến động về một chỉ tiêu nào đó của nhiều đơn vị, nhiều phần tử của hiện tượng phức tạp.

2.1. Chỉ số tổng hợp giá cả:

- Chỉ số tổng hợp giá cả đơn giản:

$$I_r = \frac{\sum_{i=1}^n p_{i(1)}}{\sum_{i=1}^n p_{i(0)}}$$

$p_{i(1)}$: là giá cả mặt hàng thứ i kỳ nghiên cứu

$p_{i(0)}$: là giá cả mặt hàng thứ i kỳ gốc.

- Chỉ số tổng hợp giá cả có quyền số:

$$I_p = \frac{\sum_{i=1}^n p_{i(1)} q_i}{\sum_{i=1}^n p_{i(0)} q_i}$$

$p_{i(1)}$: là giá cả mặt hàng thứ i kỳ nghiên cứu

$p_{i(0)}$: là giá cả mặt hàng thứ i kỳ gốc.

q_i : là quyền số của mặt hàng thứ i

Quyền số của mặt hàng thứ i có 2 giá trị: giá trị chọn ở kỳ định gốc hoặc là chọn ở kỳ nghiên cứu. Tùy theo cách chọn quyền số ta có 2 phương pháp tính:

a) Phương pháp Laspeyres:

Nếu trọng số là lượng hàng hóa tiêu thụ chọn ở kỳ gốc làm căn bản để so sánh, ta có công thức chỉ số giá cả của Laspeyres:

$$I_p = \frac{\sum_{i=1}^n p_{i(1)} q_{i(0)}}{\sum_{i=1}^n p_{i(0)} q_{i(0)}}$$

Ví dụ 2.11: Ta có số liệu về giá cả và lượng hàng tiêu thụ của 4 mặt hàng sau, tính chỉ số giá theo phương pháp Laspeyres:

TT	Mặt hàng	Giá cả (1.000đ)		Lượng hàng tiêu thụ		$p_{i(0)}q_{i(0)}$	$p_{i(1)}q_{i(0)}$
		2000 ($p_{i(0)}$)	2005 ($p_{i(1)}$)	2000 ($q_{i(0)}$)	2005 ($q_{i(1)}$)		
1	Sữa (hộp)	3,0	5,0	50.000	190.000	150.000	250.000
2	Lúa (kg)	1,6	2,4	100.000	120.000	160.000	240.000
3	Dầu (lít)	2,4	3,6	200.000	360.000	480.000	720.000
4	Máy tính bỏ túi (chiếc)	40,0	25,0	4.000	4.200	160.000	100.000
	Tổng					950.000	1.310.000

Về số tương đối:

$$I_p = \frac{\sum_{i=1}^n p_{i(1)} q_{i(0)}}{\sum_{i=1}^n p_{i(0)} q_{i(0)}} = \frac{1.310.000}{950.000} = 137,9\%$$

Về số tuyệt đối:

$$\sum_{i=1}^n p_{i(1)} q_{i(0)} - \sum_{i=1}^n p_{i(0)} q_{i(0)} = 1.310.000 - 950.000 = 360.000$$

Kết luận: Nhìn chung giá cả 4 mặt hàng trên năm 2005 so với năm 2000 bằng 137,9%, tăng 37,9% làm tăng giá trị tiêu thụ (hay doanh số tiêu thụ) một lượng là 360.000 (ngàn đồng).

Cách tính chỉ số giá theo phương pháp Laspeyres có những ưu điểm sau:

Thứ nhất, với trọng số là lượng hàng hóa tiêu thụ ở thời kỳ gốc; ta chỉ cần số liệu lượng hàng hóa tiêu thụ và giá cả của một thời kỳ căn bản nào đó để làm căn cứ so sánh.

Thứ hai, trên cơ sở chỉ số giá cả tính toán so với cùng kỳ gốc, phương pháp Laspeyres cho phép ta xác định sự thay đổi giá cả giữa hai thời gian nghiên cứu bất kỳ.

Tuy nhiên, phương pháp Laspeyres có nhược điểm là không phản ánh, cập nhật được những thay đổi về khuynh hướng, thói quen của người tiêu dùng. Một số mặt hàng nào đó vài năm trước được người tiêu thụ mua với số lượng lớn, nhưng có thể ngày nay không còn quan trọng đối với họ nữa.

b) Phương pháp tính chỉ số Paasche:

Ngược lại với chỉ số Laspeyres, chỉ số Paasche chọn lượng sản phẩm tiêu thụ ở kỳ nghiên cứu làm trọng số. Ta có công thức tính chỉ số giá của Paasche:

$$I_p = \frac{\sum_{i=1}^n p_{i(1)} q_{i(1)}}{\sum_{i=1}^n p_{i(0)} q_{i(1)}}$$

Ví dụ 2.12: Từ số liệu ở ví dụ 2.11, tính chỉ số giá theo phương pháp Paasche:

TT	Mặt hàng	Giá cả (1.000đ)		Lượng hàng tiêu thụ		$p_{i(0)}q_{i(1)}$	$p_{i(1)}q_{i(1)}$
		2000 ($p_{i(0)}$)	2005 ($p_{i(1)}$)	2000 ($q_{i(0)}$)	2005 ($q_{i(1)}$)		
1	Sữa (hộp)	3,0	5,0	50.000	190.000	570.000	950.000
2	Lúa (kg)	1,6	2,4	100.000	120.000	192.000	288.000
3	Dầu (lít)	2,4	3,6	200.000	360.000	860.000	1.296.000
4	Máy tính bỏ túi (chiếc)	40,0	25,0	4.000	4.200	168.00	105.000
	Tổng					1.794.000	2.639.000

Về số tương đối:

$$I_p = \frac{\sum_{i=1}^n p_{i(1)} q_{i(1)}}{\sum_{i=1}^n p_{i(0)} q_{i(1)}} = \frac{2.639.000}{1.794.000} = 147,1\%$$

Về số tuyệt đối:

$$\sum_{i=1}^n p_{i(1)} q_{i(1)} - \sum_{i=1}^n p_{i(0)} q_{i(1)} = 2.639.000 - 1.794.000 = 845.000$$

Kết luận: Nhìn chung giá cả 4 mặt hàng trên năm 2005 so với năm 2000 bằng 147,1%, tăng 47,1% làm tăng doanh thu một lượng là 845.000 (ngàn đồng).

Cách tính chỉ số giá theo phương pháp này khắc phục được nhược điểm của phương pháp Laspeyres. Ngoài ra phương pháp của Paasche còn cho thấy ảnh hưởng một cách cụ thể sự thay đổi giá cả đến người tiêu thụ. Nhưng một khó khăn khi sử dụng phương pháp Paasche là phải thường xuyên thu thập lại lượng hàng hóa tiêu thụ ở kỳ nghiên cứu.

Tuỳ theo mức độ số liệu thể thu thập được mà chúng ta có thể áp dụng một trong 3 trường hợp trên một cách linh hoạt nhưng phương pháp Paasche là phương pháp tỏ ra ưu việt hơn bởi số liệu mang tính cập nhật hơn và đây là phương pháp mà người ta thường dùng. Ngoài ra, người ta còn đưa ra một phương pháp khác mang tính dung hoà hơn đó là tính quyền số là trung bình của lượng hàng hoá tiêu thụ ở hai thời điểm.

2.2. Chỉ số tổng hợp khối lượng:

Chỉ số tổng hợp khối lượng về căn bản giống như chỉ số tổng hợp giá cả nhưng ngược lại nhân tố xem xét là khối lượng sản phẩm còn giá cả đóng vai trò là trọng số.

- Chỉ số tổng hợp khối lượng đơn giản:

$$I_q = \frac{\sum_{i=1}^n q_{i(1)}}{\sum_{i=1}^n q_{i(0)}}$$

$q_{i(1)}$: là khối lượng mặt hàng thứ i kỳ nghiên cứu.

$q_{i(0)}$: là khối lượng mặt hàng thứ i kỳ gốc.

a) Phương pháp Laspeyres:

$$I_q = \frac{\sum_{i=1}^n q_{i(1)} P_{i(0)}}{\sum_{i=1}^n q_{i(0)} P_{i(0)}}$$

Ví dụ 2.13: Từ số liệu ở ví dụ 2.11, tính chỉ số khối lượng theo phương pháp Laspeyres.

TT	Mặt hàng	Giá cả (1.000đ)		Lượng hàng tiêu thụ		$q_{i(0)}P_{i(0)}$	$q_{i(1)}P_{i(0)}$
		2000 ($P_{i(0)}$)	2005 ($P_{i(1)}$)	2000 ($q_{i(0)}$)	2005 ($q_{i(1)}$)		
1	Sữa (hộp)	3,0	5,0	50.000	190.000	150.000	570.000
2	Lúa (kg)	1,6	2,4	100.000	120.000	160.000	192.000
3	Dầu (lít)	2,4	3,6	200.000	360.000	480.000	864.000
4	Máy tính bỏ túi (chiếc)	40,0	25,0	4.000	4.200	160.000	168.000
	Tổng					950.000	1.794.000

Về số tương đối:

$$I_q = \frac{\sum_{i=1}^n q_{i(1)} P_{i(0)}}{\sum_{i=1}^n q_{i(0)} P_{i(0)}} = \frac{1.794.000}{950.000} = 188,8\%$$

Về số tuyệt đối:

$$\sum_{i=1}^n q_{i(1)} P_{i(0)} - \sum_{i=1}^n q_{i(0)} P_{i(0)} = 1.794.000 - 950.000 = 844.000$$

Kết luận: Nhìn chung lượng hàng tiêu thụ 4 mặt hàng trên năm 2005 so với năm 2000 bằng 188,8%, tăng 88,8% làm tăng giá trị tiêu thụ một lượng là 844.000 (ngàn đồng).

Phương pháp chỉ số tổng hợp khối lượng theo phương pháp Laspeyres thường được chọn vì các lý do sau:

Thứ nhất, với cách chọn giá cả ở kỳ gốc làm quyền số, phương pháp này giúp ta có thể tính toán chỉ số tổng hợp khối lượng một cách nhanh chóng.

Thứ hai, trên cơ sở chỉ số khối lượng tính toán so với kỳ gốc, phương pháp này cho phép ta xác định sự thay đổi khối lượng giữa hai thời kỳ nghiên cứu.

b) Phương pháp tính chỉ số Paasche:

$$I_q = \frac{\sum_{i=1}^n q_{i(1)} P_{i(1)}}{\sum_{i=1}^n q_{i(0)} P_{i(1)}}$$

Ví dụ 2.14: Từ số liệu ở ví dụ 2.11, tính chỉ số khối lượng theo phương pháp Paasche

TT	Mặt hàng	Giá cả (1.000đ)		Lượng hàng tiêu thụ		$q_{i(0)}P_{i(1)}$	$q_{i(1)}P_{i(1)}$
		2000 ($p_{i(0)}$)	2005 ($p_{i(1)}$)	2000 ($q_{i(0)}$)	2005 ($q_{i(1)}$)		
1	Sữa (hộp)	3,0	5,0	50.000	190.000	250.000	950.000
2	Lúa (kg)	1,6	2,4	100.000	120.000	240.000	288.000
3	Dầu (lít)	2,4	3,6	200.000	360.000	720.000	1.296.000
4	Máy tính bỏ túi (chiếc)	40,0	25,0	4.000	4.200	100.000	105.000
	Tổng					1.310.000	2.639.000

Về số tương đối:

$$I_q = \frac{\sum_{i=1}^n q_{i(1)} P_{i(1)}}{\sum_{i=1}^n q_{i(0)} P_{i(1)}} = \frac{2.639.000}{1.310.000} = 201,5\%$$

Về số tuyệt đối:

$$\sum_{i=1}^n q_{i(1)} P_{i(1)} - \sum_{i=1}^n q_{i(0)} P_{i(1)} = 2.639.000 - 1.310.000 = 1.329.000$$

Kết luận: Nhìn chung lượng hàng tiêu thụ 4 mặt hàng trên năm 2005 so với năm 2000 bằng 201,5%, tăng 101,5% làm tăng giá trị tiêu thụ một lượng là 1.329.000 (ngàn đồng).

Theo phương pháp Paasche, xét về nội dung có ý nghĩa kinh tế hơn là chỉ số tổng hợp theo phương pháp Laspeyres. Tuy nhiên, nếu xét trong mối tương quan giữa mặt lượng và mặt chất tác động đến một hiện tượng (Doanh thu phụ thuộc vào giá bán và lượng hàng hoá tiêu thụ) thì người ta sử dụng phương pháp Laspeyres, do đó đây là phương pháp thường được sử dụng.

Nếu chỉ nghiên cứu riêng chỉ số tổng hợp khối lượng thì tùy theo số liệu thu thập mà ta có thể chọn phương pháp phù hợp nhưng với công thức đơn giản chúng ta cần thận trọng trong nhiều trường hợp tính toán công thức này không có ý nghĩa.

3. Chỉ số trung bình tính từ chỉ số tổng hợp

3.1. Chỉ số trung bình điều hòa về biến động của chỉ tiêu chất lượng: Trong trường hợp tài liệu chỉ có giá trị ở kỳ báo cáo và chỉ số giá cả cá thể:

$$I_p = \frac{\sum_{i=1}^n P_{i(1)} q_{i(1)}}{\sum_{i=1}^n P_{i(0)} q_{i(1)}} = \frac{\sum_{i=1}^n P_{i(1)} q_{i(1)}}{\sum_{i=1}^n \frac{P_{i(0)}}{P_{i(1)}} P_{i(1)} q_{i(1)}} = \frac{\sum_{i=1}^n P_{i(1)} q_{i(1)}}{\sum_{i=1}^n P_{i(0)} q_{i(1)}}$$

Ví dụ 2.15: Có số liệu sau đây của một công ty:

Mặt hàng	Doanh thu năm 2007	Thay đổi giá bán năm 2007 so với năm 2006
A	5.408	+4
B	6.175	-5
C	9.996	+2

Tính chỉ số giá của 3 mặt hàng trên.

Từ số liệu trên ta có bảng:

Mặt hàng	Doanh thu năm 2007 ($p_{i(1)}q_{i(1)}$)	Thay đổi giá bán năm 2007 so với năm 2006 (i_{pi})	($p_{i(1)}q_{i(1)}$)/ (i_{pi})
A	5.408	104	5.200
B	6.175	95	6.500
C	9.996	102	9.800
Tổng	21.579		21.500

Thay vào công thức ta có:

$$I_p = \frac{\sum_{i=1}^n p_{i(1)} q_{i(1)}}{\sum_{i=1}^n p_{i(0)} q_{i(1)}} = \frac{\sum_{i=1}^n p_{i(1)} q_{i(1)}}{\sum_{i=1}^n \frac{1}{i_{pi}} p_{i(1)} q_{i(1)}} = \frac{21.579}{21.500} = 100,37\%$$

Kết luận: Giá cả 3 mặt hàng năm 2007 so với năm 2006 bằng 100,37%, tăng 0,37%.

3.2. Chỉ số trung bình số học về biến động của chỉ tiêu khối lượng: Trong trường hợp tài liệu chỉ cho giá trị kỳ gốc và chỉ số khối lượng cá thể.

$$I_q = \frac{\sum_{i=1}^n q_{i(1)} p_{i(0)}}{\sum_{i=1}^n q_{i(0)} p_{i(0)}} = \frac{\sum_{i=1}^n \frac{q_{i(1)}}{q_{i(0)}} q_{i(0)} p_{i(0)}}{\sum_{i=1}^n q_{i(0)} p_{i(0)}} = \frac{\sum_{i=1}^n i_{qi} q_{i(0)} p_{i(0)}}{\sum_{i=1}^n q_{i(0)} p_{i(0)}}$$

Ví dụ 2.16: Có số liệu sau đây của một công ty:

Mặt hàng	Doanh thu năm 2006	Tỷ lệ lượng hàng bán năm 2007 so với năm 2006
A	2.000	104
B	5.000	96
C	3.000	105

Tính chỉ số khối lượng hàng hóa tiêu thụ của 3 mặt hàng trên.

Từ số liệu trên ta có bảng:

Mặt hàng	Doanh thu năm 2006 ($p_{i(0)}q_{i(0)}$)	Thay đổi giá bán năm 2007 so với năm 2006 (i_{qi})	$i_{pi} p_{i(0)} q_{i(0)}$
A	2.000	104	2.080
B	5.000	96	4.800
C	3.000	105	3.150
Tổng	10.000		10.030

Thay vào công thức ta có:

$$I_q = \frac{\sum_{i=1}^n q_{i(1)} P_{i(0)}}{\sum_{i=1}^n q_{i(0)} P_{i(0)}} = \frac{\sum_{i=1}^n i_{qi} q_{i(0)} P_{i(0)}}{\sum_{i=1}^n q_{i(0)} P_{i(0)}} = \frac{10.030}{10.000} = 100,3\%$$

Kết luận: Khối lượng hàng hóa tiêu thụ của công ty tính chung 3 mặt hàng năm 2007 so với năm 2006 bằng 100,3%, tăng 0,3%.

4. Chỉ số không gian

Là chỉ số so sánh các hiện tượng cùng loại nhưng qua các điều kiện không gian khác nhau. Ví dụ, nghiên cứu sự biến động về lượng hàng bán ra và giá cả các mặt hàng ở hai thị trường khác nhau.

4.1. Chỉ số tổng hợp nghiên cứu sự biến động của chỉ tiêu chất lượng ở hai thị trường A và B.

$$I_p(A/B) = \frac{\sum_{i=1}^n p_{Ai} q_i}{\sum_{i=1}^n p_{Bi} q_i}$$

Trong đó, $q_i = q_{Ai} + q_{Bi}$: Khối lượng sản phẩm cùng loại ở hai thị trường A và B.

Ví dụ 2.17: Tình hình tiêu thụ mặt hàng X và Y tại hai chợ A và B. Hãy tính sự biến động về giá cả ở hai thị trường trên.

Mặt hàng	Thị trường A		Thị trường B	
	Lượng bán	Giá đơn vị	Lượng bán	Giá đơn vị
X	480	12.000	520	10.000
Y	300	10.000	200	18.000

Ta có:

$$Q_X = q_{AX} + q_{BX} = 480 + 520 = 1.000$$

$$Q_Y = q_{AY} + q_{BY} = 300 + 200 = 500$$

$$I_p(A/B) = \frac{\sum p_{Ai} q_i}{\sum p_{Bi} q_i} = \frac{p_{AX} q_X + p_{AY} q_Y}{p_{BX} q_X + p_{BY} q_Y} = \frac{(12.000 \times 1.000) + (10.000 \times 500)}{(10.000 \times 1.000) + (18.000 \times 500)} = \frac{17.000.000}{19.000.000} = 89,5\%$$

$$\text{Về số tuyệt đối: } (17.000.000 - 19.000.000) = -2.000.000$$

Kết luận: Nói chung giá cả hai mặt hàng ở thị trường A thấp hơn thị trường B là 10,5%. giá trị tiêu thụ ở thị trường A thấp hơn thị trường B 2.000.000.

4.2. Chỉ số tổng hợp nghiên cứu sự biến động của chỉ tiêu khối lượng ở hai thị trường A và B: Trong trường hợp này có thể có các quyền số khác nhau là các chỉ tiêu chất lượng, chẳng hạn như giá cố định cho từng mặt hàng (p) hoặc tính giá trung bình từng mặt hàng ở hai thị trường).

$$I_q(A/B) = \frac{\sum_{i=1}^n q_{Ai} p_i}{\sum_{i=1}^n q_{Bi} p_i}$$

Trong đó: p_i là giá cố định cho mặt hàng i

Nếu ta chọn giá cố định là giá trung bình của hai mặt hàng ở từng thị trường, ta có:

$$I_q(A/B) = \frac{\sum q_{Ai} \bar{p}_i}{\sum q_{Bi} \bar{p}_i}$$

$\bar{p}_i$: Giá trung bình của mặt hàng i ở hai thị trường.

Ví dụ 2.18: Sử dụng số liệu của ví dụ 2.17, hãy tính sự biến động về khối lượng hàng tiêu thụ ở hai thị trường trên.

Mặt hàng	Thị trường A		Thị trường B	
	Lượng bán	Giá đơn vị	Lượng bán	Giá đơn vị
X	480	12.000	520	10.000
Y	300	10.000	200	18.000

Sử dụng số bình quân gia quyền ta tính giá cố định của hai mặt hàng.

$$p_x = \frac{(12.000 \times 480) + (10.000 \times 520)}{480 + 520} = 10.960$$

$$p_y = \frac{(10.000 \times 300) + (18.000 \times 200)}{300 + 200} = 13.200$$

$$I_q(A/B) = \frac{\sum q_{Ai} \bar{p}_i}{\sum q_{Bi} \bar{p}_i} = \frac{q_{AX} \bar{p}_x + q_{AY} \bar{p}_y}{q_{BX} \bar{p}_x + q_{BY} \bar{p}_y}$$

$$\frac{(480 \times 10.960) + (300 \times 13.200)}{(300 \times 10.960) + (200 \times 13.200)} = \frac{9.220.800}{8.339.200} = 110,6\%$$

Về số tuyệt đối: $9.220.800 - 8.339.200 = 881.600$.

Kết luận: Nói chung lượng hàng tiêu thụ ở thị trường A cao hơn thị trường B là 10,6%, làm tăng giá trị tiêu thụ của thị trường A cao hơn thị trường B là 881.600.

5. Hệ thống chỉ số liên hoàn 2 nhân tố

Phương pháp chỉ số giúp nghiên cứu sự thay đổi của hiện tượng kinh tế qua thời gian. Ngoài phương pháp chỉ số còn có thể phân tích mối liên hệ giữa các nhân tố và mức độ ảnh hưởng của các nhân tố đến sự thay đổi một chỉ tiêu kinh tế tổng hợp. Ví dụ, doanh số bán của một công ty biến động là do ảnh hưởng của hai nhân tố: giá bán và khối lượng hàng tiêu thụ, sản lượng thu hoạch của một loại cây trồng do ảnh hưởng của hai loại nhân tố: năng suất thu hoạch và diện tích gieo trồng,...

Giả sử cần phân tích tổng mức hàng hóa tiêu thụ biến động qua hai thời kỳ nghiên cứu trong một liên hệ giữa hai nhân tố: giá cả và khối lượng hàng hóa tiêu thụ. Hệ thống chỉ số thể hiện mối liên hệ giữa các chỉ tiêu trên như sau:

$$I_{pq} = I_p \times I_q$$

Trong đó:

I_{pq} : Chỉ số tổng mức hàng hóa tiêu thụ

I_p : Chỉ số giá được xác định theo phương pháp Paasche

I_q : Chỉ số khối lượng được xác định theo phương pháp Laspeyres

$$\frac{\sum_{i=1}^n p_{i(1)} q_{i(1)}}{\sum_{i=1}^n p_{i(0)} q_{i(0)}} = \frac{\sum_{i=1}^n p_{i(1)} q_{i(1)}}{\sum_{i=1}^n p_{i(0)} q_{i(1)}} \times \frac{\sum_{i=1}^n p_{i(0)} q_{i(1)}}{\sum_{i=1}^n p_{i(0)} q_{i(0)}}$$

Về số tuyệt đối:

$$\sum_{i=1}^n p_{i(1)} q_{i(1)} - \sum_{i=1}^n p_{i(0)} q_{i(0)} = \left(\sum_{i=1}^n p_{i(1)} q_{i(1)} - \sum_{i=1}^n p_{i(0)} q_{i(1)} \right) + \left(\sum_{i=1}^n p_{i(0)} q_{i(1)} - \sum_{i=1}^n p_{i(0)} q_{i(0)} \right)$$

Về số tương đối so với giá trị tiêu thụ kỳ gốc:

$$\frac{\sum_{i=1}^n p_{i(1)} q_{i(1)} - \sum_{i=1}^n p_{i(0)} q_{i(0)}}{\sum_{i=1}^n p_{i(0)} q_{i(0)}} = \frac{\sum_{i=1}^n p_{i(1)} q_{i(1)} - \sum_{i=1}^n p_{i(0)} q_{i(1)}}{\sum_{i=1}^n p_{i(0)} q_{i(0)}} + \frac{\sum_{i=1}^n p_{i(0)} q_{i(1)} - \sum_{i=1}^n p_{i(0)} q_{i(0)}}{\sum_{i=1}^n p_{i(0)} q_{i(0)}}$$

Ví dụ 2.19: Từ số liệu ví dụ 2.11, nghiên cứu sự ảnh hưởng của giá bán và lượng bán đến doanh số bán.

Ta có bảng số liệu:

TT	Mặt hàng	Giá cả (1.000đ)		Lượng hàng tiêu thụ		$p_{i(1)} q_{i(1)}$	$p_{i(0)} q_{i(0)}$	$p_{i(0)} q_{i(1)}$
		2000 ($p_{i(0)}$)	2005 ($p_{i(1)}$)	2000 ($q_{i(0)}$)	2005 ($q_{i(1)}$)			
1	Sữa (hộp)	3,0	5,0	50.000	190.000	950.000	150.000	570.000
2	Lúa (kg)	1,6	2,4	100.000	120.000	288.000	160.000	192.00
3	Dầu (lít)	2,4	3,6	200.000	360.000	1.296.000	480.000	864.000
4	Máy tính bỏ túi (chiếc)	40,0	25,0	4.000	4.200	105.000	160.000	168.00
	Tổng					2.639.000	950.000	1.794.000

Nhận xét về số tương đối:

$$\frac{\sum_{i=1}^n p_{i(1)} q_{i(1)}}{\sum_{i=1}^n p_{i(0)} q_{i(0)}} = \frac{\sum_{i=1}^n p_{i(1)} q_{i(1)}}{\sum_{i=1}^n p_{i(0)} q_{i(1)}} \times \frac{\sum_{i=1}^n p_{i(0)} q_{i(1)}}{\sum_{i=1}^n p_{i(0)} q_{i(0)}}$$

$$\frac{2.639.000}{950.000} = \frac{2.639.000}{1.794.000} \times \frac{1.794.000}{950.000}$$

$$277,8\% = 147,1\% \times 188,8\%$$

(tăng 177,8%) (tăng 47%) (tăng 88,8%)

Nhận xét về số tuyệt đối:

$$\sum_{i=1}^n p_{i(1)} q_{i(1)} - \sum_{i=1}^n p_{i(0)} q_{i(0)} = \left(\sum_{i=1}^n p_{i(1)} q_{i(1)} - \sum_{i=1}^n p_{i(0)} q_{i(1)} \right) + \left(\sum_{i=1}^n p_{i(0)} q_{i(1)} - \sum_{i=1}^n p_{i(0)} q_{i(0)} \right)$$

$$(2.639.000 - 950.000) = (2.639.000 - 1.794.000) + (1.794.000 - 950.000)$$

$$1.689.000 = 845.000 + 844.000$$

Kết luận: Tổng mức tiêu thụ hàng hóa tính chung 4 mặt hàng năm 2005 so với năm 2000 bằng 277,8% (tăng 117,8%), mức tăng là 1.689.000 ngàn đồng là do ảnh hưởng của 2 nhân tố liên quan:

- Do giá cả năm 2005 so với năm 2000 tăng 47,1% làm tăng giá trị tiêu thụ là 845.000 ngàn đồng.

- Do khối lượng các mặt hàng bán ra nói chung tăng 88,8% làm tăng giá trị tiêu thụ là 844.000 ngàn đồng.

Hệ thống liên hoàn hai nhân tố người ta có thể mở rộng ra thành hệ thống liên hoàn nhiều nhân tố và chúng ta cũng thực hiện theo nguyên tắc chỉ số chất lượng thì sử dụng phương pháp Paasche còn chỉ số số lượng thì sử dụng phương pháp Laspeyres. Tuy nhiên, việc mở rộng này tỏ ra không hiệu quả cao vì nó có thể làm cho việc phân tích quá phức tạp.

CHƯƠNG V

SỰ BIẾN ĐỘNG CỦA CÁC HIỆN TƯỢNG KINH TẾ XÃ HỘI

Đối với các khái niệm và tính chất có liên quan đến phân phối của tổng thể chúng ta đã nghiên cứu ở môn học xác suất thống kê toán. Ở đây chỉ mang tính chất nhắc lại một cách khái quát nhất.

I. PHÂN PHỐI CHUẨN

Phân phối chuẩn chiếm một vị trí rất quan trọng trong lý thuyết thống kê nó liên quan đến các kết luận thống kê suy luận sau này. Trong thực tế, nhiều biến ngẫu nhiên tuân theo qui luật phân phối chuẩn hoặc gần chuẩn, chẳng hạn như trọng lượng và chiều cao của người lớn, mức độ thông minh của trẻ em, điểm thi của các thí sinh, lực chịu đựng của một thanh sắt, các sai số đo đạc, Do đó, việc nhắc lại là rất cần thiết.

1. Định nghĩa

Phân phối chuẩn là phân phối của đại lượng ngẫu nhiên liên tục X có miền xác định từ $-\infty$ đến $+\infty$ với hàm mật độ xác suất:

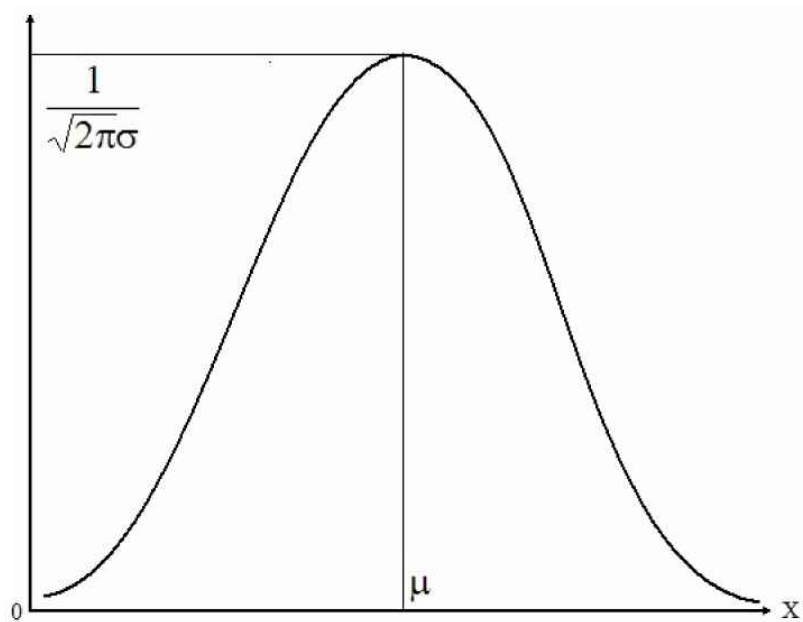
$$f_x = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \text{ trong đó: } e = 2,71828, \pi = 3,1415$$

Ký hiệu: $X \sim N(\mu, \sigma^2)$

- *Tính chất của hàm phân phối chuẩn:*

$$\int_{-\infty}^{+\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = 1. \text{ Chính là diện tích giới hạn bởi đồ thị } f(x) \text{ và trục hoành.}$$

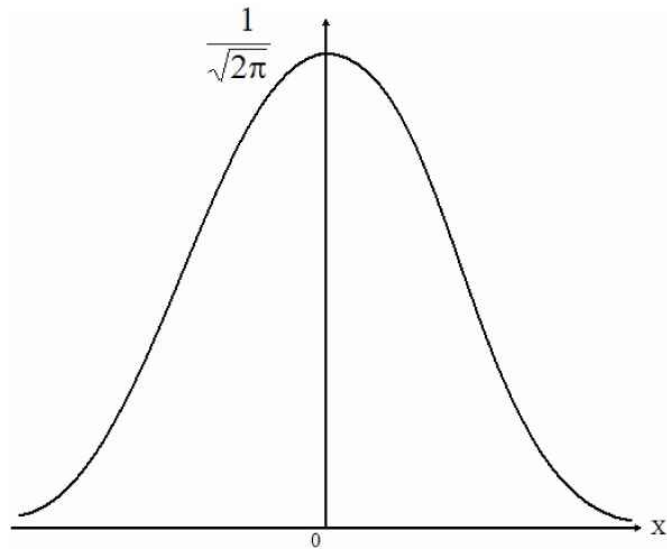
- Đồ thị đối xứng với nhau qua đường thẳng $x=\mu$
- X có trung bình là μ và phương sai là σ^2



2. Phân phối chuẩn tắc (đơn giản)

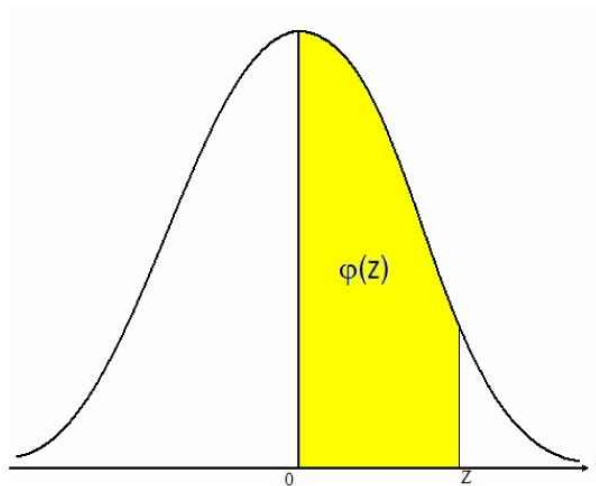
Định nghĩa: là phân phối chuẩn có $\mu=0$ và $\sigma=1$. $f_t = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}$

Ta có thể dùng phương pháp đổi biến $t=(x-\mu)/\sigma$ đổi với phân phối chuẩn thành phân phối chuẩn tắc.



Bảng phân phối chuẩn tắc (đơn giản):

Hàm số $\varphi(z) = \int_0^z \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}$ được gọi là hàm tích phân Laplace



- Tính chất của hàm tích phân Laplace:

- $\varphi(z) = p(0 < t < z)$

- $\varphi(z)$ là hàm số lẻ: $\varphi(-z) = -\varphi(z)$

- $\int_0^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} = 0,5$

Để tìm được $\varphi(Z)$ ta có thể tra bảng ở phụ lục 1 bằng phương pháp tọa độ. Ví dụ để tìm $\varphi(1,08)$ ta thực hiện như sau:

Z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,0000	0,0040	0,0080	0,0120	0,0160	0,0199	0,0239	0,0279	0,0319	0,0359
0,1	0,0398	0,0438	0,0478	0,0517	0,0557	0,0596	0,0636	0,0675	0,0714	0,0753
0,2	0,0793	0,0832	0,0871	0,0910	0,0948	0,0987	0,1026	0,1064	0,1103	0,1141
0,3	0,1179	0,1217	0,1255	0,1293	0,1331	0,1368	0,1406	0,1443	0,1480	0,1517
0,4	0,1554	0,1591	0,1628	0,1664	0,1700	0,1736	0,1772	0,1808	0,1844	0,1879
0,5	0,1915	0,1950	0,1985	0,2019	0,2054	0,2088	0,2123	0,2157	0,2190	0,2224
0,6	0,2257	0,2291	0,2324	0,2357	0,2389	0,2422	0,2454	0,2486	0,2517	0,2549
0,7	0,2580	0,2611	0,2642	0,2673	0,2704	0,2734	0,2764	0,2794	0,2823	0,2852
0,8	0,2881	0,2910	0,2939	0,2967	0,2995	0,3023	0,3051	0,3078	0,3106	0,3133
0,9	0,3159	0,3186	0,3212	0,3238	0,3264	0,3289	0,3315	0,3340	0,3365	0,3389
1,0	0,3413	0,3438	0,3461	0,3485	0,3508	0,3531	0,3554	0,3577	0,3599	0,3621
1,1	0,3643	0,3665	0,3686	0,3708	0,3729	0,3749	0,3770	0,3790	0,3810	0,3830
1,2	0,3849	0,3869	0,3888	0,3907	0,3925	0,3944	0,3962	0,3980	0,3997	0,4015

$$\varphi(1,08) = 0,3599$$

Chúng ta có thể sử dụng hàm NORMSDIST() trong Excel, tuy nhiên trong Excel là bảng 1, còn phụ lục của ta là bảng 0,5. Do đó để có kết quả chính xác ta sử dụng (NORMSDIST(Z) - 0,5).

3. Khái niệm Z_α

Z_α là một số sao cho $p(Z > Z_\alpha) = \alpha$. Đây chính là xác suất sai lầm mà ta thường dùng trong thống kê. Ta có thể tìm giá trị Z_α bằng hàm NORMSINV(1 - α) trong Excel. Một vài giá trị đặc biệt:

α	0,005	0,010	0,025	0,050	0,100
Z_α	2,575	2,330	1,960	1,645	1,280

4. Một vài công thức xác suất thường dùng

Cách tính xác suất của biến ngẫu nhiên X có phân phối chuẩn tắc:

a) $p(X > a) = 0,5 - \varphi(a)$

b) $p(X < b) = 0,5 + \varphi(b)$

d) $p(a < X < b) = ((b) - ((a))$, với $a < b$

e) $p(X < a, X > b) = p(X < a) + p(X > b)$, với $a < b$, X - phân phối chuẩn

II. PHÂN PHỐI CỦA MỘT VÀI ĐẠI LƯỢNG THỐNG KÊ

1. Phân phối Chi bình phương

Nếu $x_1, x_2, \dots, x_n$ là đại lượng ngẫu nhiên được chọn từ tổng thể phân phối chuẩn thì $(n-1) \frac{S^2}{\sigma^2}$ có phân phối Chi bình phương bậc tự do (n-1).

Ký hiệu: $(n-1) \frac{S^2}{\sigma^2} \sim \chi^2_{n-1}$

Không giống như hàm phân phối chuẩn, hàm chi bình phương không đối xứng. Nó có các tính chất sau:

$$E(\chi^2_{df}) = df$$

$$\text{Var}(\chi^2_{df}) = 2df$$

Hàm phân phối χ^2 có đồ thị gần giống như đồ thị hàm phân phối chuẩn nhưng không đối xứng nhau qua cực trị. Tuy nhiên, nếu n lớn thì phân phối Chi bình phương sẽ sắp xỉ phân phối chuẩn. Phân phối χ^2 do Karl Pearson đưa ra vào năm 1900.

- **Khái niệm $\chi^2_{n-1;\alpha}$:** $\chi^2_{n-1;\alpha}$ là một số sao cho $p(\chi^2 > \chi^2_{n-1;\alpha}) = \alpha$. Định nghĩa này tương tự như đối với Z_α . Muốn tìm giá trị này ta có thể tra bảng ở phụ lục 3 cũng bằng phương pháp tọa độ hoặc sử dụng hàm CHINV(α, df) trong Excel. Trong đó, df là bậc tự do.

PEARSON Karl

London 1857 – London 1936

Nhà toán học thống kê người Anh này đã từng học tại King's College ở Oxford. Năm 1884 ông được bổ nhiệm phụ trách bộ môn Toán ứng dụng và Cơ học tại University College ở London. Năm 1901 ông thành lập phòng thí nghiệm về sinh trắc và cùng năm ấy với sự tài trợ của Galton, ông cho ra đời tạp chí Biometrika và ông đã điều hành tạp chí này cho đến khi mất. Năm 1911 ông phụ trách luôn bộ môn Eugénisme (Giáo sư Đào Duy Anh dịch là Nhân chủng cải lương học, Ưu sinh học) cho đến khi ông về hưu năm 1933. Kế tục công trình của Galton, ông là một trong những người sáng lập ngành Toán học thống kê hiện đại. Ông nghiên cứu lý thuyết tiến hoá theo mô hình thống kê toán học của ông. Ngoài những giấy phức tạp mê mẩn với khoa học, ông còn dành thời gian cho thi ca và hoạt động chính trị.

2. Phân phối Student

Nếu $x_1, x_2, \dots, x_n$ là đại lượng ngẫu nhiên được chọn từ tổng thể phân phối chuẩn thì $t = \frac{\bar{x} - \mu}{S/\sqrt{n}}$ có phân phối Student bậc tự do $(n-1)$.

Ký hiệu: $t = \frac{\bar{x} - \mu}{S/\sqrt{n}} \sim t_{n-1}$

- **Khái niệm $t_{n-1;\alpha}$:** $t_{n-1;\alpha}$ là một số sao cho $p(t > t_{n-1;\alpha}) = \alpha$. Định nghĩa này tương tự như đối với Z_α . Muốn tìm giá trị này ta có thể tra bảng ở phụ lục 2 cũng bằng phương pháp tọa độ hoặc sử dụng hàm TINV($2\alpha, df$) trong Excel.

Hàm phân phối t có đồ thị gần giống như đồ thị hàm phân phối chuẩn nhưng độ nhọn thấp hơn. Tuy nhiên, nếu n lớn thì phân phối Student sẽ sắp xỉ phân phối chuẩn.

GOSSET William Sealy

Canterbury 1876 - Beaconsfield (Ngoại ô Luân Đôn) 1937

Thường thường người ta biết ông dưới dạng biệt hiệu Student. Ông là nhà thống kê người Anh. Thời trẻ, Ông học toán ở New college (Oxford). Năm 1899 ông vào làm việc xưởng sản xuất bia cho Guinness Brewery ở Dublin. Nhà máy muốn giảm giá thành sản xuất, nâng cao chất lượng lúa mạch và cây Hublon để sản xuất bia nên thành lập một phòng thí nghiệm nghiên cứu. Một bài toán đặt ra là làm sao từ những mẫu nghiên cứu ở phòng thí nghiệm rút ra được những kết luận xác đáng. Gosset tự nguyện tham gia nghiên cứu bài toán này: lý thuyết các mẫu nhỏ để từ đó rút ra kết luận. Gosset lao vào đề tài đặt ra, cùng Karl Pearson ở Luân Đôn miệt mài trong 2 năm 1906, 1907. Năm 1908 ông đưa ra một test - sau này gọi là test Gosset - dùng để lựa chọn đại lúa mạch. Để bảo mật, nhà máy yêu cầu Gosset dấu tên thật, chỉ dùng biệt hiệu Pupil hay Student. Và như chúng ta biết, Gosset dùng hiệu thứ 2.

2.3. Phân phối Fisher (F)

Giả sử có hai mẫu độc lập có n_x, n_y quan sát lấy từ hai tổng thể có phân phối chuẩn, phương sai tổng thể và phương sai mẫu lần lượt là $\sigma_x^2, \sigma_y^2, S_x^2, S_y^2$ thì khi đó

$F = \frac{S_x^2/\sigma_x^2}{S_y^2/\sigma_y^2}$ có phân phối Fisher bậc tự do của tử (n_x-1) và bậc tự do của mẫu (n_y-1) .

Ký hiệu: $F \sim F_{v_1, v_2}$

Trong thực tế khi kiểm định sự bằng nhau của hai phương sai tổng thể và khi đó:

$$F = \frac{S_x^2}{S_y^2}$$

• **Khái niệm $F_{v_1; v_2; \alpha}$:** $F_{v_1, v_2; \alpha}$ là một số sao cho $p(F > F_{v_1, v_2; \alpha}) = \alpha$. Định nghĩa này tương tự như đối với Z_α . Muốn tìm giá trị này ta có thể tra bảng ở phụ lục 4 cũng bằng phương pháp tọa độ hoặc sử dụng hàm $\text{FINV}(\alpha, df_1, df_2)$ trong Excel.

FISHER Ronald Aylmer

Londres 1890 - Adélaide 1962

Ông là một chuyên gia về Lý thuyết tiến hóa, nhà Di truyền học người Anh này đã có công phát triển các phương pháp thống kê so sánh những trung bình của hai mẫu nhờ đó xác định sự khác biệt của chúng có ý nghĩa hay không.

III. PHÂN PHỐI MẪU

1. Khái niệm

Mục đích của phân tích thống kê là sử dụng số liệu thu thập từ mẫu như trung bình và tỷ lệ mẫu để ước lượng giá trị thực của tổng thể. Quá trình khái quát hóa kết quả nghiên cứu của mẫu cho tổng thể chung được gọi là suy luận thống kê.

Về lý thuyết, để có thể sử dụng các thông tin mẫu để suy luận các tham số của tổng thể chung, ta nên dựa vào kết quả nghiên cứu của nhiều mẫu nếu có thể. Nếu điều này được thực hiện, phân phối của các kết quả từ mẫu được gọi là phân phối mẫu. Nhưng về mặt thực hành việc ước lượng cho các tham số của tổng thể chỉ căn cứ vào kết quả cụ thể của một mẫu. Cho nên khái niệm về phân phối mẫu cần phải được xem xét để có thể ứng dụng lý thuyết về xác suất cho quá trình suy luận thống kê.

Phân phối mẫu có 3 trường hợp:

- Phân phối của trung bình mẫu để ước lượng trung bình của tổng thể;
- Phân phối của phương sai mẫu để ước lượng phương sai tổng thể;
- Phân phối của tỷ lệ mẫu để ước lượng tỷ lệ của tổng thể.

2. Định lý giới hạn trung tâm

Trong thực tế thường gặp là ta không biết về phân phối của tổng thể hoặc tổng thể không có phân phối chuẩn. Trong những trường hợp đó định lý giới hạn trung tâm giúp ta giải quyết vấn đề khi xem xét phân phối mẫu.

Định lý: Khi cỡ mẫu n đủ lớn thì phân phối của trung bình mẫu $\bar{X}$ sẽ xấp xỉ phân phối chuẩn, bất chấp hình dáng phân phối của tổng thể.

Định lý: Một biến ngẫu nhiên là tổng của của nhiều biến ngẫu nhiên khác sẽ có phân phối xấp xỉ phân phối chuẩn.

Điều này rất hữu ích trong kinh tế lượng bởi vì ta có thể coi sai số của một mô hình là tổng của nhiều tác động ngẫu nhiên.

3. Các tính chất của phân phối mẫu

- Nếu X có phân phối χ^2_m thì $n\bar{X}$ cũng có phân phối χ^2_{nm}

- Nếu X có phân phối chuẩn $N(\mu, \sigma^2)$ thì:

1). $n\bar{X}$ cũng có phân phối chuẩn $N(n\mu, n\sigma^2)$ và $\bar{X} \sim N(\mu, \sigma^2/n)$

2). Với kích thước mẫu khá lớn ($n \geq 30$), thì phân phối của trung bình mẫu sẽ xấp xỉ phân phối chuẩn và $z = \frac{\bar{x} - \mu}{\sigma_x / \sqrt{n}}$ có phân phối chuẩn tắc.

$$3). \frac{(n-1)S^2}{\sigma^2} \sim \chi_{n-1}^2$$

4). $\bar{X}$ và S^2 độc lập với nhau.

TÀI LIỆU THAM KHẢO

- Đào Hữu Hồ (2001), *Xác suất thống kê*, NXB Đại học quốc gia Hà Nội.
- Hoàng Trọng, Chu Nguyễn Mộng Ngọc (2007), *Thống kê ứng dụng trong kinh tế xã hội*, NXB Thống kê.
- Khoa dự báo và phát triển Trường Đại học Kinh tế Quốc dân (2003), *Giáo trình Dự báo phát triển kinh tế - xã hội*, NXB Thống kê.
- Nguyễn Đình Hương (1999), *Thống kê ứng dụng trong quản lý*, NXB Thanh niên.
- Nguyễn Khắc Minh (2002), *Các phương pháp phân tích và dự báo trong kinh tế*, NXB Khoa học và kỹ thuật.
- Nguyễn Thành Long (2005), *Giáo trình dự đoán kinh tế*, Đại học Đà Nẵng.
- Trần Bá Nhân, Đinh Thái Hoàng (1998), *Lý thuyết thống kê ứng dụng trong quản trị, kinh doanh và nghiên cứu kinh tế*, NXB Thống kê.
- Trần Ngọc Phát, Trần thị Kim Thu (2006), *Giáo trình Lý thuyết thống kê*, NXB Thống kê.
- Viện Khoa học thống kê (2005), *Một số vấn đề phương pháp luận thống kê*.
- Võ Thị Thanh Lộc (2000), *Thống kê ứng dụng và dự báo trong kinh doanh và kinh tế*, NXB Thống kê.
- Joseph F. Healy (2002), *Statistics: A tool for Social Research*, Wadsworth Publishing Company (An International Thomson Publishing Company).



TRƯỜNG CAO ĐẲNG GIAO THÔNG VẬN TẢI TRUNG ƯƠNG I

✉ : Thụy An, Ba Vì, Hà Nội

☎ : (024) 33.863.050

🌐 : [http:// gtvttw1.edu.vn](http://gtvttw1.edu.vn)

✉ : info@gtvttw1.edu.vn