

BỘ XÂY DỰNG

GIÁO TRÌNH
LÝ THUYẾT
THỐNG KÊ



NHÀ XUẤT BẢN XÂY DỰNG

BỘ XÂY DỰNG

GIÁO TRÌNH
Lý thuyết thống kê

NHÀ XUẤT BẢN XÂY DỰNG
HÀ NỘI - 2004

LỜI NÓI ĐẦU

Lý thuyết thống kê là những vấn đề lý luận cơ sở nằm trong khối kiến thức chuyên ngành thuộc chương trình đào tạo cao đẳng kế toán. Những kiến thức của thống kê học mà Lý thuyết thống kê cung cấp rất cần thiết cho sinh viên khối kinh tế, cũng như đối với những người đang làm công tác quản lý.

Trước yêu cầu đó, tập thể giáo viên thuộc Khoa Kinh tế trường Cao đẳng Xây dựng số 3, do Giảng viên chính Phạm Đình Văn - Chủ nhiệm Khoa Kinh tế làm chủ biên, đã biên soạn Giáo trình Lý thuyết thống kê, nhằm phục vụ cho công tác giảng dạy và làm tài liệu học tập cho sinh viên cao đẳng ngành kế toán. Giáo trình được biên soạn theo đề cương chương trình giảng dạy cho sinh viên cao đẳng ngành Kế toán do Bộ Xây dựng ban hành, đã được Hội đồng khoa học nhà trường xét duyệt và thông qua. Trong quá trình biên soạn, chúng tôi đã nhận được nhiều ý kiến đóng góp quý báu của các đồng nghiệp trong và ngoài trường, các chuyên gia đầu ngành ở các Cục, Vụ, Viện... trong cơ quan Bộ.

Lần đầu tiên biên soạn giáo trình dành cho hệ cao đẳng chắc khó tránh khỏi những thiếu sót, chúng tôi rất mong nhận được ý kiến đóng góp của bạn đọc để kịp thời bổ sung và chỉnh lý cho giáo trình được hoàn thiện hơn.

Xin chân thành cảm ơn.

Trường Cao đẳng Xây dựng số 3

Chương I

ĐỐI TƯỢNG NGHIÊN CỨU CỦA THỐNG KÊ HỌC

I. ĐỐI TƯỢNG NGHIÊN CỨU CỦA THỐNG KÊ HỌC

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

Thống kê học là một môn khoa học xã hội có lịch sử phát triển lâu đời. Người ta đã tìm thấy một số di tích cổ tại Trung Quốc, cổ Hy Lạp, La Mã, Ai Cập,... chứng tỏ rằng ngay từ thời cổ đại con người đã biết làm công việc đăng ký và ghi chép số liệu. Tuy nhiên, các công việc này còn đơn giản, tiến hành trong phạm vi nhỏ hẹp, chưa mang tính thống kê rõ nét.

Đến thời chiếm hữu nô lệ, các chủ nô thường tìm cách ghi chép và tính toán để nắm được tài sản của mình như: số nô lệ, số súc vật,... Dưới chế độ phong kiến, công tác thống kê đã có nhiều bước phát triển. Hầu hết các quốc gia châu Á, châu Âu đều đã tổ chức các cuộc đăng ký, kê khai ở phạm vi rộng, nội dung phong phú và có tính chất thống kê rõ rệt như: đăng ký nhân khẩu, kê khai ruộng đất,... Tuy đã có tính chất thống kê, nhưng các hoạt động này chưa đúc kết thành lý luận và chỉ dừng lại ở thống kê mô tả.

Mãi đến cuối thế kỷ thứ XVII, lực lượng sản xuất phát triển mạnh mẽ và phương thức sản xuất tư bản chủ nghĩa ra đời. Tính chất xã hội của sản xuất ngày càng cao, thị trường được mở rộng và cạnh tranh ngày càng gay gắt đòi hỏi những người quản lý kinh doanh, quản lý nhà nước, những nhà khoa học phải đi sâu nghiên cứu lý luận và phương pháp thực tiễn để thu thập, tính toán, phân tích số liệu. Thống kê học thực sự ra đời và chuyển sang giai đoạn thống kê phân tích. Các tài liệu, sách báo về thống kê ra đời và ở một số trường học bắt đầu giảng dạy lý luận thống kê.

Trong thời kỳ này, một số tác phẩm có tính chất phân tích thống kê đầu tiên ra đời như cuốn “Số học chính trị” (1676) của nhà kinh tế học người Anh William Petty (1623 - 1687). ông đã có ý nghĩ về việc sử dụng thống kê để nghiên cứu các hiện tượng kinh tế - xã hội. Năm 1660, nhà kinh tế học người Đức H. Conhring (1606 - 1681) đã giảng dạy tại trường Đại học Heimsted phương pháp nghiên cứu hiện tượng xã hội dựa vào số liệu điều tra cụ thể. Đến năm 1759, một Giáo sư người Đức G. Achenwall (1719 - 1772) lần đầu tiên dùng từ “Statistik” để chỉ phương pháp nghiên cứu trên. Sau này người ta dịch là “Thống kê”.

Những thành tựu khoa học tự nhiên trong thời kỳ này, đặc biệt là sự ra đời của lý thuyết xác suất thống kê toán đã có ảnh hưởng rất quan trọng đến sự phát triển của thống kê học.

Những người sáng lập ra chủ nghĩa Mác - Lênin cũng nhiều lần nhấn mạnh đến ý nghĩa khoa học và thực tiễn của thống kê. Trong các tác phẩm của Mác, Ăngghen, Lênin kiến thức thống kê được sử dụng để phân tích những vấn đề phức tạp và phong phú.

Nghiên cứu sơ lược quá trình hình thành và phát triển của thống kê học, có thể thấy thống kê học ra đời và phát triển do nhu cầu của hoạt động thực tiễn xã hội.

1.2. Đối tượng nghiên cứu của thống kê học

Thống kê học là một môn khoa học xã hội, nó ra đời và phát triển gắn liền với sự phát triển của nền sản xuất xã hội. Các hiện tượng mà thống kê học nghiên cứu là các hiện tượng và quá trình xã hội, chủ yếu là các hiện tượng và quá trình kinh tế - xã hội, bao gồm:

- Các hiện tượng về quá trình tái sản xuất mở rộng của cải vật chất xã hội.
- Các hiện tượng về dân số và nguồn lao động.
- Các hiện tượng về đời sống vật chất, văn hóa và tinh thần của dân cư.
- Các hiện tượng về sinh hoạt chính trị, xã hội.

Thống kê học nghiên cứu các hiện tượng xã hội, không nghiên cứu các hiện tượng tự nhiên. Nhưng giữa các hiện tượng xã hội và hiện tượng tự nhiên có mối quan hệ chặt chẽ với nhau, cho nên khi nghiên cứu các hiện tượng xã hội thống kê cũng phải nghiên cứu ảnh hưởng của các nhân tố tự nhiên (thời tiết, khí hậu, địa lý) và kỹ thuật (cải tiến công cụ, áp dụng kỹ thuật mới...) đến các hiện tượng và quá trình xã hội. Khác với các môn khoa học xã hội khác, thống kê học không trực tiếp nghiên cứu bản chất và quy luật của các hiện tượng xã hội. Thống kê học nghiên cứu mặt lượng trong mối quan hệ chặt chẽ với mặt chất của hiện tượng xã hội. Mặt lượng trong quan hệ chặt chẽ với mặt chất của hiện tượng xã hội là những biểu hiện bằng số lượng về bản chất và tính quy luật của hiện tượng trong điều kiện thời gian và địa điểm cụ thể. Những biểu hiện số lượng đó được thể hiện bằng quy mô, kết cấu, quan hệ tỷ lệ, tốc độ phát triển,... của các hiện tượng. Mặt lượng đó không phải là con số trừu tượng, mà là những số liệu có ý nghĩa, gắn liền với nội dung kinh tế - xã hội nhất định, chúng giúp ta nhận thức được cụ thể bản chất của hiện tượng. Các con số thống kê phản ánh được mặt chất của hiện tượng vì chất và lượng là hai mặt không thể tách rời nhau của sự vật và hiện tượng.

Khi nghiên cứu các hiện tượng xã hội thống kê học coi tập hợp gồm nhiều hiện tượng cá biệt là một tổng thể hoàn chỉnh để nghiên cứu và dùng phương pháp quan sát số lớn để loại trừ những ảnh hưởng mang tính chất ngẫu nhiên, qua đó nêu lên đầy đủ và nổi

bật những đặc trưng của bản chất và tính quy luật của hiện tượng nghiên cứu. Nhưng khi nghiên cứu các hiện tượng số lớn thống kê học cũng không bỏ qua nghiên cứu các hiện tượng cá biệt nhằm giúp cho nhận thức hiện tượng xã hội được toàn diện, phong phú và sâu sắc hơn.

Hiện tượng kinh tế - xã hội bao giờ cũng tồn tại trong những điều kiện cụ thể về thời gian và không gian. Trong những điều kiện lịch sử khác nhau, hiện tượng kinh tế - xã hội có đặc điểm về chất và biểu hiện về lượng khác nhau. Do vậy, khi sử dụng tài liệu thống kê vào phân tích tình hình kinh tế - xã hội, phải luôn xét tới điều kiện thời gian và không gian cụ thể của hiện tượng mà tài liệu phản ánh.

Tóm lại, thống kê học là một môn khoa học xã hội, nó nghiên cứu mặt lượng trong mối quan hệ chặt chẽ với mặt chất của hiện tượng và quá trình kinh tế - xã hội số lớn trong điều kiện thời gian và địa điểm cụ thể.

II. MỘT SỐ KHÁI NIỆM THƯỜNG DÙNG TRONG THỐNG KÊ HỌC

2.1. Tổng thể thống kê

Tổng thể thống kê (gọi tắt là tổng thể) là tập hợp nhiều đơn vị cá biệt trên cơ sở một hoặc một số đặc điểm chung. Ví dụ: toàn bộ dân số nước ta có vào 0 giờ ngày 01-4-1999 là 76.323.173 người là một tổng thể thống kê, vì đây là một tập hợp những người Việt Nam không phân biệt già hay trẻ, nam hay nữ, nghề nghiệp, trình độ văn hóa, ...

Để cấu thành tổng thể, các đơn vị tổng thể chỉ cần có một hoặc một số đặc điểm chung. Nhưng mỗi đơn vị tổng thể lại có nhiều đặc điểm khác nhau. Do vậy, tùy theo mục đích nghiên cứu người ta còn phân biệt tổng thể đồng chất và tổng thể không đồng chất.

Tổng thể đồng chất bao gồm các đơn vị giống nhau về một số đặc điểm chủ yếu có liên quan đến mục đích nghiên cứu. Còn tổng thể không đồng chất bao gồm các đơn vị khác nhau về các đặc điểm, các loại hình. Ví dụ: các doanh nghiệp sản xuất công nghiệp có thể được coi là một tổng thể đồng chất, nếu đem so sánh chúng với các đơn vị sản xuất nông nghiệp, giao thông vận tải, thương nghiệp.... nhưng nếu đi sâu nghiên cứu các loại hình trong tổng thể các doanh nghiệp sản xuất công nghiệp, thì đây lại là một tổng thể không đồng chất, vì các doanh nghiệp này thuộc các thành phần kinh tế khác nhau, sản xuất ra các loại sản phẩm khác nhau. Như vậy, tổng thể thống kê được coi là đồng chất trong trường hợp nghiên cứu này, lại là tổng thể không đồng chất trong trường hợp nghiên cứu khác.

2.2. Đơn vị tổng thể thống kê

Đơn vị tổng thể là từng đơn vị cá biệt cấu thành nên tổng thể. Ví dụ: với tổng thể là toàn bộ dân số cả nước thì đơn vị tổng thể là từng người dân. Còn trong tổng thể các doanh nghiệp sản xuất công nghiệp thì từng doanh nghiệp là một đơn vị tổng thể.

Đặc điểm của đơn vị tổng thể là những phần tử không thể chia nhỏ được nữa. Ngoài đặc điểm giống nhau để cấu thành tổng thể, mỗi đơn vị tổng thể còn có nhiều đặc điểm riêng.

2.3. Tiêu thức thống kê

Tiêu thức thống kê là từng đặc điểm của đơn vị tổng thể. *Ví dụ:* từng nhân khẩu có các tiêu thức như: họ và tên, tuổi, giới tính, trình độ văn hóa, tình trạng hôn nhân, ...

Tùy theo tính chất lượng hóa của tiêu thức, có thể chia ra hai loại tiêu thức là: tiêu thức thuộc tính và tiêu thức số lượng.

- Tiêu thức thuộc tính (còn gọi là tiêu thức chất lượng) là tiêu thức không biểu hiện giá trị của nó bằng con số cụ thể. *Ví dụ:* dân tộc, giới tính, nghề nghiệp, ...

- Tiêu thức số lượng là tiêu thức có thể biểu hiện giá trị của nó bằng những con số cụ thể. *Ví dụ:* tuổi đời, tuổi nghề, chiều cao, sức nặng, mức lương, ...

2.4. Chỉ tiêu thống kê

Chỉ tiêu thống kê là khái niệm phản ánh một cách tổng hợp mặt lượng trong quan hệ chặt chẽ với mặt chất của tổng thể trong điều kiện thời gian và địa điểm cụ thể.

Trong bản thân mỗi chỉ tiêu thống kê có hai mặt: khái niệm và con số. Mặt khái niệm quy định nội dung kinh tế - xã hội của chỉ tiêu. Con số chỉ tiêu nêu lên mức độ của chỉ tiêu trong điều kiện thời gian và địa điểm cụ thể.

Chỉ tiêu thống kê thường mang tính chất tổng hợp, biểu hiện đặc điểm của cả tổng thể. Đây cũng là căn cứ để phân biệt chỉ tiêu và tiêu thức. Tuy nhiên do đối tượng nghiên cứu cụ thể của thống kê thường thay đổi khác nhau, cho nên trong trường hợp nghiên cứu này một đặc điểm nào đó được coi là một chỉ tiêu, còn trong trường hợp nghiên cứu khác cũng đặc điểm đó được coi là tiêu thức. *Ví dụ:* số lao động trong một doanh nghiệp sản xuất công nghiệp là chỉ tiêu của doanh nghiệp, nhưng cũng số lao động đó lại là tiêu thức của đơn vị tổng thể là các doanh nghiệp sản xuất công nghiệp.

Chương II

QUÁ TRÌNH NGHIÊN CỨU THỐNG KÊ

I. XÁC ĐỊNH HỆ THỐNG CHỈ TIÊU THỐNG KÊ

1.1. Khái niệm hệ thống chỉ tiêu thống kê

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

Hệ thống chỉ tiêu thống kê có tác dụng lượng hóa các mặt quan trọng nhất, cơ cấu khách quan, mối liên hệ cơ bản của đối tượng nghiên cứu. Đó là cơ sở để nhận thức được bản chất, tính quy luật và xu hướng phát triển của hiện tượng nghiên cứu.

Trong quản lý kinh tế - xã hội thường sử dụng hai loại hệ thống chỉ tiêu thống kê: hệ thống chỉ tiêu chung cho toàn bộ nền kinh tế quốc dân và hệ thống chỉ tiêu cho từng ngành kinh tế nghiệp vụ.

Hệ thống chỉ tiêu chung cho toàn bộ nền kinh tế quốc dân là hệ thống chỉ tiêu phản ánh toàn diện các điều kiện vật chất của đời sống xã hội, tình hình phân bổ lực lượng sản xuất, quá trình tái sản xuất mở rộng, hiệu quả kinh tế của nền sản xuất xã hội.

Hệ thống chỉ tiêu cho từng ngành kinh tế nghiệp vụ phù hợp với đặc điểm của từng ngành, được các đơn vị báo cáo theo chế độ quy định.

1.2. Các yêu cầu cơ bản để xác định hệ thống chỉ tiêu thống kê

Xây dựng hệ thống chỉ tiêu thống kê phải thỏa mãn các yêu cầu như:

- Phải có khả năng nêu được mối liên hệ giữa các bộ phận, các mặt của hiện tượng nghiên cứu.
- Phải có các chỉ tiêu tổng hợp và các chỉ tiêu phân tích phản ánh sâu về từng mặt của hiện tượng nghiên cứu.
- Phải đảm bảo sự thống nhất về nội dung, phương pháp và phạm vi tính toán của các chỉ tiêu thống kê cùng loại...

II. ĐIỀU TRA THỐNG KÊ

2.1. Khái niệm, ý nghĩa, nhiệm vụ điều tra thống kê

Điều tra thống kê là việc tổ chức một cách khoa học và theo một kế hoạch thống nhất việc thu thập, ghi chép tài liệu ban đầu về các hiện tượng và quá trình kinh tế - xã hội để phục vụ cho những mục đích nhất định.

Ví dụ: khi cần nghiên cứu tình hình dân số cả nước, thống kê phải tổ chức thu thập tài liệu ban đầu trên từng người dân về: tên, tuổi, giới tính, trình độ văn hoá, chuyên môn... hoặc khi cần nghiên cứu tình hình sản xuất của các doanh nghiệp, thống kê phải tổ chức thu thập tài liệu trong từng doanh nghiệp về: số lao động sử dụng, số giờ máy hoạt động, số nguyên liệu tiêu dùng vào sản xuất,... việc thu thập tài liệu ban đầu như vậy được gọi là điều tra thống kê.

Nhiệm vụ của điều tra thống kê là thu thập tài liệu ban đầu cần thiết dùng làm căn cứ cho tổng hợp và phân tích thống kê, phục vụ cho việc xây dựng các kế hoạch phát triển kinh tế, văn hóa xã hội, kiểm tra tình hình thực hiện kế hoạch, nêu lên các nguồn tài nguyên và khả năng tiềm tàng của đất nước, ngoài ra còn phục vụ cho các yêu cầu nghiên cứu khác.

Để tài liệu điều tra thống kê có chất lượng, phải thoả mãn các yêu cầu sau đây:

- Chính xác: tài liệu điều tra thống kê phải phản ánh đúng thực tế của hiện tượng nghiên cứu. Đây là yêu cầu cơ bản nhất của điều tra thống kê.
- Kịp thời: điều tra thống kê phải thu thập và cung cấp đúng lúc các tài liệu mà người sử dụng cần. Tài liệu điều tra thu thập và cung cấp không kịp thời sẽ mất tác dụng.
- Đầy đủ: tài liệu điều tra thống kê phải được thu thập theo đúng nội dung và số đơn vị cần điều tra.

2.2. Các loại điều tra thống kê

2.2.1. Căn cứ vào tính chất thường xuyên, liên tục của thu thập số liệu ban đầu: có thể chia thành hai loại là điều tra thường xuyên và điều tra không thường xuyên.

a) Điều tra thường xuyên: là việc thu thập tài liệu ban đầu một cách thường xuyên, liên tục gắn liền với quá trình phát sinh, phát triển của hiện tượng.

Ví dụ: ghi chép tình hình biến động nhân khẩu tại một địa phương, hoặc ghi chép hàng ngày số lao động có mặt, số nguyên vật liệu tiêu dùng cho sản xuất sản phẩm, số sản phẩm sản xuất, số sản phẩm tiêu thụ,... tại một doanh nghiệp sản xuất.

Điều tra thường xuyên giúp theo dõi tỉ mỉ tình hình phát triển của hiện tượng theo thời gian. Nó được áp dụng đối với hiện tượng biến động nhanh. Tài liệu điều tra thường xuyên là cơ sở để lập báo cáo thống kê định kỳ, là căn cứ để kiểm tra tình hình thực hiện kế hoạch, phản ánh kết quả tích lũy của hiện tượng trong một thời kỳ.

b) Điều tra không thường xuyên: là việc thu thập tài liệu ban đầu về hiện tượng không thường xuyên, liên tục, không gắn liền với quá trình phát sinh, phát triển của hiện tượng.

Điều tra không thường xuyên có thể chia thành điều tra không thường xuyên định kỳ và không định kỳ. Điều tra không thường xuyên định kỳ được tiến hành lặp đi, lặp lại theo một chu kỳ nhất định như: tổng điều tra dân số, kiểm kê hàng hóa tồn kho định kỳ,... Điều tra không thường xuyên không định kỳ như: điều tra nghiên cứu thị trường, thăm dò ý kiến khách hàng, điều tra thiên tai, ...

Điều tra không thường xuyên áp dụng cho những trường hợp không xảy ra thường xuyên hoặc xảy ra thường xuyên nhưng không đòi hỏi theo dõi thường xuyên. Điều tra không thường xuyên thường đi sâu vào khía cạnh chuyên môn cần nghiên cứu.

2.2.2. Căn cứ vào phạm vi thu thập tài liệu ban đầu: có thể chia thành hai loại là điều tra toàn bộ và điều tra không toàn bộ.

a) Điều tra toàn bộ: là việc thu thập tài liệu ban đầu trên tất cả các đơn vị hoặc các bộ phận của tổng thể.

Ví dụ: tổng điều tra dân số, tổng điều tra tồn kho vật tư hàng hoá, tổng điều tra gia súc,... là điều tra toàn bộ.

Điều tra toàn bộ cung cấp tài liệu đầy đủ nhất cho nghiên cứu thống kê. Điều tra toàn bộ giúp cho việc tính toán các chỉ tiêu phản ánh quy mô, khối lượng của hiện tượng được chính xác, làm cơ sở đề ra các quyết định trong quản lý. Tuy nhiên chi phí cho điều tra toàn bộ rất lớn, đòi hỏi phải tổ chức chỉ đạo khoa học chặt chẽ. Tuy vậy, có những trường hợp nhất thiết phải điều tra toàn bộ như tổng điều tra dân số.

b) Điều tra không toàn bộ: là việc thu thập tài liệu ban đầu trên một số đơn vị hoặc bộ phận của tổng thể.

Ví dụ: các cuộc điều tra thu nhập và chi tiêu gia đình, điều tra dư luận xã hội, điều tra giá cả hàng hoá trên thị trường, ... là điều tra không toàn bộ.

Điều tra không toàn bộ được sử dụng nhiều trong nghiên cứu thống kê vì có nhiều ưu điểm như: tiến hành gọn nhẹ, nhanh chóng, đáp ứng yêu cầu quản lý, tiết kiệm chi phí điều tra. Ngoài ra, do phạm vi điều tra được thu hẹp, nên có thể mở rộng nội dung điều tra để nghiên cứu hiện tượng toàn diện hơn, chi tiết hơn.

Điều tra không toàn bộ có các loại sau đây: điều tra chọn mẫu, điều tra trọng điểm, điều tra chuyên đề.

- *Điều tra chọn mẫu:* là thu thập tài liệu ban đầu trên một số đơn vị được chọn ra từ tổng thể chung. Sau đó căn cứ vào kết quả thu thập được từ tổng thể mẫu để tính toán suy rộng thành đặc điểm chung của toàn bộ tổng thể.

Ví dụ: điều tra chất lượng đồ hộp, điều tra năng suất lúa, điều tra mức sống dân cư,...

Để suy rộng được kết quả điều tra, tổng thể mẫu được chọn phải đại biểu được cho toàn bộ tổng thể nghiên cứu.

- *Điều tra trọng điểm*: là thu thập tài liệu ban đầu ở một bộ phận chủ yếu nhất của tổng thể. Bộ phận chủ yếu nhất thường là bộ phận chiếm tỷ trọng lớn trong toàn bộ tổng thể nghiên cứu.

Kết quả điều tra giúp ta nhận thức được tình hình cơ bản của hiện tượng nghiên cứu, nhưng không dùng để tính toán suy rộng thành các đặc điểm chung của tổng thể. Điều tra trọng điểm chỉ thích hợp cho những hiện tượng có từng bộ phận tương đối tập trung.

Ví dụ: điều tra các vùng chuyên canh trong nông nghiệp như cây chè ở Phú Thọ, Hà Giang, Lâm Đồng, cao su ở Đồng Nai, Bình Dương, Phước Long, cà phê ở Đắk Lắk....

- *Điều tra chuyên đề*: là thu thập tài liệu ban đầu trên một số rất ít, thậm chí chỉ một đơn vị của tổng thể nghiên cứu nhưng lại đi sâu nghiên cứu chi tiết nhiều khía cạnh khác nhau của đơn vị đó.

Ví dụ: điều tra các điển hình tiên tiến hoặc lạc hậu.

Mục đích điều tra chuyên đề là để nghiên cứu các nhân tố mới hay những thiếu sót trong xu hướng phát triển của hiện tượng. Kết quả điều tra chuyên đề không dùng để suy rộng hoặc làm căn cứ để đánh giá tình hình cơ bản của toàn bộ hiện tượng nghiên cứu.

2.3. Các phương pháp thu thập tài liệu điều tra thống kê

Tùy theo đặc điểm của đối tượng điều tra và mục đích điều tra, có thể dùng các phương pháp điều tra như: đăng ký trực tiếp, phỏng vấn và đăng ký qua chứng từ sổ sách.

2.3.1. Đăng ký trực tiếp

Là phương pháp thu thập tài liệu ban đầu trong đó nhân viên điều tra phải tiếp xúc trực tiếp với đơn vị được điều tra, trực tiếp tiến hành hoặc giám sát việc cân, đo, đóng, đếm và tự ghi chép tài liệu vào phiếu điều tra.

Ví dụ: điều tra tồn kho, điều tra năng suất lúa,...

Tài liệu thu thập từ phương pháp này có mức độ chính xác cao, nhưng lại tốn kém chi phí. Tuy nhiên, trên thực tế có những hiện tượng không thể quan sát trực tiếp được như các khoản thu chi gia đình trong điều tra mức sống dân cư... nên phạm vi áp dụng phương pháp này có hạn chế.

2.3.2. Phỏng vấn

Là phương pháp mà nhân viên điều tra thu thập tài liệu ban đầu qua sự trả lời của người hoặc đơn vị được điều tra. Có nhiều hình thức phỏng vấn:

- *Phương pháp phái viên điều tra*: nhân viên điều tra đến tận địa điểm điều tra, gặp người được điều tra, đặt câu hỏi, nghe trả lời và ghi chép lại.

Ví dụ: điều tra dân số, điều tra dư luận xã hội, điều tra mức sống dân cư, điều tra nghiên cứu thị trường,...

- *Phương pháp tự ghi báo*: người được điều tra tự ghi chép vào phiếu điều tra và giao trả lại cho cơ quan điều tra.

Ví dụ: điều tra cán bộ khoa học kỹ thuật, điều tra lao động,....

- *Phương pháp gửi thư*: là phương pháp thu thập tài liệu qua đường bưu điện, được thực hiện bằng cách đơn vị điều tra phát các phiếu điều tra cho người được thăm dò, hỏi ý kiến hoặc có người làm thông tin viên gửi thư từ, tài liệu cho cơ quan thống kê để phản ánh các vấn đề cần nghiên cứu.

2.3.3. Đăng ký qua chứng từ sổ sách

Là phương pháp thu thập tài liệu dựa vào các chứng từ sổ sách đã được lưu trữ và ghi chép một cách có hệ thống.

2.4. Các hình thức tổ chức điều tra thống kê

2.4.1. Báo cáo thống kê định kỳ

Báo cáo thống kê định kỳ là hình thức tổ chức điều tra thống kê một cách thường xuyên, có định kỳ theo nội dung, phương pháp và chế độ báo cáo do Nhà nước quy định thống nhất.

Ví dụ: định kỳ hàng tháng, quý, năm các doanh nghiệp nhà nước, các cơ quan thuộc quyền quản lý của Nhà nước phải lập và gửi các báo cáo thống kê lên cơ quan cấp trên, cơ quan hữu quan.

Báo cáo thống kê định kỳ là hình thức tổ chức điều tra mang tính chất hành chính bắt buộc, phạm vi áp dụng chủ yếu cho các doanh nghiệp nhà nước và cơ quan nhà nước. Đối với doanh nghiệp tư nhân cũng áp dụng nhưng nội dung báo cáo còn rất hạn chế.

Nội dung của báo cáo thống kê định kỳ cần giải quyết tốt những vấn đề chủ yếu sau đây:

a) Ghi chép ban đầu

Là việc ghi chép lần đầu tiên theo chế độ quy định về các hiện tượng kinh tế - xã hội phát sinh ở đơn vị cơ sở.

Ví dụ: trong một doanh nghiệp sản xuất phải ghi chép hàng ngày số lao động có mặt, số nguyên vật liệu sử dụng, số sản phẩm sản xuất ra,...

Phạm vi ghi chép ban đầu không bao gồm việc tính toán tổng hợp để vào sổ trung gian hoặc lập các báo cáo thống kê.

Tài liệu ghi chép ban đầu là cơ sở để tổng hợp, tính toán các chỉ tiêu trong báo cáo thống kê định kỳ. Ghi chép ban đầu là cơ sở để thống nhất ba loại hạch toán trong mỗi đơn vị (hạch toán nghiệp vụ kỹ thuật, hạch toán kế toán, hạch toán thống kê).

b) Hệ thống chỉ tiêu thống kê và hệ thống biểu mẫu

Trong báo cáo thống kê định kỳ, hệ thống chỉ tiêu và trị số của chỉ tiêu là nội dung của các biểu mẫu báo cáo thống kê.

Biểu mẫu Báo cáo thống kê là loại bảng thống kê được lập sẵn theo mẫu quy định để các đơn vị ghi số liệu vào và gửi lên cấp trên và các cơ quan hữu quan.

Về hình thức, mỗi biểu mẫu Báo cáo thống kê đều có các chi tiết sau: tên biểu, số hiệu của biểu, cơ quan lập biểu, ngày tháng quyết định phê chuẩn biểu, thời gian tính toán các chỉ tiêu trong biểu, tên đơn vị lập báo cáo, nơi nhận báo cáo, các chỉ tiêu cần báo cáo (đây là nội dung chính của biểu), chữ ký của người lập biểu, kế toán trưởng, thủ trưởng đơn vị.

2.4.2. Điều tra chuyên môn

Điều tra chuyên môn là hình thức tổ chức điều tra không thường xuyên được tiến hành theo nội dung và phương pháp quy định riêng cho mỗi lần điều tra.

Phạm vi áp dụng của điều tra chuyên môn là những hiện tượng mà chế độ báo cáo thống kê định kỳ không thể thu thập được (điều tra dân số, điều tra nhu cầu nhà ở...), hoặc những hiện tượng xảy ra bất thường như (thiên tai, tai nạn lao động,...).

Nội dung của điều tra chuyên môn hay nội dung của một phương án điều tra bao gồm những vấn đề chủ yếu sau đây:

a) Mục đích điều tra

Là xác định rõ cuộc điều tra nhằm tìm hiểu vấn đề gì, phục vụ cho yêu cầu nghiên cứu nào. Xác định mục đích điều tra là căn cứ để xác định đối tượng, đơn vị và nội dung điều tra. Mục đích điều tra rõ ràng cụ thể tạo điều kiện thu thập tài liệu đúng yêu cầu nghiên cứu, đầy đủ, tránh lãng phí.

Ví dụ: mục đích điều tra tồn kho vật tư trong một doanh nghiệp là thu thập, tổng hợp, cung cấp những số liệu về số lượng của từng loại vật tư hiện có trong kho một cách có hệ thống, chính xác làm căn cứ cho việc:

- Xây dựng kế hoạch, kiểm tra tình hình thực hiện kế hoạch vật tư ở doanh nghiệp.
- Đối chiếu với định mức, với kết quả sản xuất để phát hiện những sai sót trong quá trình sử dụng vật tư.

b) Đối tượng điều tra

Là tổng thể các đơn vị của hiện tượng nghiên cứu cần thu thập tài liệu. Xác định đối tượng điều tra nhằm: quy định phạm vi điều tra, tránh nhầm lẫn khi thu thập tài liệu ban đầu. Muốn xác định đối tượng điều tra phải căn cứ vào phân tích lý luận kinh tế - xã hội và mục đích điều tra.

Ví dụ: đối tượng của cuộc điều tra dân số là toàn bộ dân số của địa phương có hộ khẩu thường trú và hộ khẩu tạm trú và là người Việt Nam.

Hoặc trong cuộc điều tra tồn kho vật tư của doanh nghiệp đối tượng điều tra là tất cả các loại vật tư có trong kho.

c) Đơn vị điều tra

Là nơi phát sinh các tài liệu ban đầu cần thu thập trong mỗi cuộc điều tra. Đơn vị tổng thể và đơn vị điều tra có thể trùng hoặc không trùng nhau.

Ví dụ: trong điều tra dân số thì đơn vị điều tra có thể là từng hộ gia đình hoặc là từng người dân. Muốn xác định đúng đơn vị điều tra phải căn cứ vào đối tượng điều tra và mục đích điều tra.

d) Nội dung điều tra

Là những tiêu thức cần thu thập trong cuộc điều tra. Nội dung điều tra được thể hiện thành các câu hỏi ngắn gọn, rõ ràng mà đơn vị điều tra sẽ trả lời hoặc tự viết vào phiếu điều tra. Muốn xác định một nội dung điều tra phải căn cứ vào mục đích điều tra.

Ví dụ: nội dung điều tra trong tổng điều tra dân số là các tiêu thức: họ tên, quan hệ với chủ hộ, giới tính, ngày tháng năm sinh, dân tộc, nơi ở thường xuyên, trình độ văn hóa, trình độ chuyên môn kỹ thuật, tình trạng hôn nhân,....

Hoặc nội dung điều tra tồn kho vật tư là các tiêu thức: tên vật tư, đơn vị tính, số lượng (trong đó theo sổ sách, thực kiểm kê), chất lượng,....

e) Thời điểm và thời kỳ điều tra

Thời điểm điều tra là mốc thời gian quy định để thu thập tài liệu tất cả các đơn vị điều tra.

Ví dụ: thời điểm điều tra của tổng điều tra dân số của nước ta năm 1999 là 0 giờ ngày 01/4/1999.

Quy định thời điểm điều tra để tránh đăng ký trùng lặp hoặc bỏ sót đơn vị điều tra khi thu thập tài liệu. Yêu cầu về thời điểm điều tra đối với mỗi loại hiện tượng cũng có khác nhau. Có hiện tượng đòi hỏi quy định thời điểm điều tra chính xác đến giờ như điều tra dân số. Có hiện tượng chỉ cần chính xác đến ngày như điều tra tồn kho vật tư.

Thời kỳ điều tra là độ dài thời gian quy định để thu thập tài liệu tất cả các đơn vị điều tra. Thời kỳ điều tra dài hay ngắn phụ thuộc vào tính chất phức tạp của hiện tượng nghiên cứu, mục đích và nội dung điều tra.

2.5. Sai số trong điều tra thống kê

Sai số trong điều tra thống kê là chênh lệch giữa trị số của tiêu thức điều tra thu thập được so với trị số thực tế của hiện tượng.

Có thể phân biệt hai loại sai số: sai số do đăng ký và sai số do tính chất đại biểu.

- Sai số do đăng ký là loại sai số phát sinh do việc ghi chép thu thập tài liệu ban đầu không chính xác. Nguyên nhân có thể do nhân viên điều tra vô tình cân đong, đo, đếm, ghi sai hoặc cố tình ghi sai, hoặc do người hay đơn vị điều tra khai báo sai do vô tình hay cố ý.

- Sai số do tính chất đại biểu là loại sai số trong điều tra chọn mẫu. Nguyên nhân của sai số này là do việc lựa chọn đơn vị điều tra không đủ tính chất đại biểu.

Để khắc phục và hạn chế các sai số này, cần có biện pháp như: chuẩn bị điều tra tốt, tiến hành kiểm tra tài liệu thu thập được và kiểm tra tính chất đại diện của các đơn vị điều tra trong điều tra chọn mẫu.

III. TỔNG HỢP THỐNG KÊ

3.1. Khái niệm, ý nghĩa, nhiệm vụ tổng hợp thống kê

Tổng hợp thống kê là tiến hành chính lý, hệ thống hoá một cách khoa học các tài liệu ban đầu thu thập được trong điều tra thống kê.

Nhiệm vụ cơ bản của tổng hợp thống kê là chuyển những đặc trưng riêng biệt của từng đơn vị tổng thể thành những đặc trưng chung của toàn bộ tổng thể.

Ví dụ: sau khi tiến hành tổng điều tra dân số, cơ quan thống kê đã thu thập được một số lớn tài liệu về tiêu thức điều tra trên từng người dân như: tuổi tác, giới tính, nghề nghiệp, dân tộc,... Qua tổng hợp các kết quả điều tra trên, thống kê sẽ nêu lên một số chỉ tiêu tổng hợp phản ánh đặc điểm của toàn bộ dân số nước ta như: quy mô, kết cấu, sự phân bố dân số, nguồn lao động...

Hoặc tài liệu điều tra tồn kho vật tư của doanh nghiệp qua tổng hợp cho biết tổng giá trị vật tư tồn kho vào thời điểm điều tra, tổng khối lượng vật tư từng loại,...

Tổng hợp thống kê là giai đoạn thứ hai của quá trình nghiên cứu thống kê, nó có ý nghĩa rất lớn đối với kết quả nghiên cứu thống kê. Tổng hợp thống kê một cách đúng đắn và khoa học sẽ làm cho các tài liệu thu thập trong điều tra thống kê có ý nghĩa hơn và sẽ cung cấp số liệu chính xác cho giai đoạn phân tích thống kê.

3.2. Những vấn đề chủ yếu của tổng hợp thống kê

3.2.1. Mục đích tổng hợp thống kê

Mục đích của tổng hợp thống kê là khái quát các đặc trưng chung của tổng thể nghiên cứu bằng các chỉ tiêu thống kê. Kết quả của tổng hợp thống kê là căn cứ để phân tích thống kê. Vì vậy khi xác định mục đích của tổng hợp thống kê phải dựa vào yêu cầu phân tích hiện tượng nghiên cứu để nêu ra các chỉ tiêu tổng hợp cần đạt được.

3.2.2. Nội dung tổng hợp thống kê

Nội dung tổng hợp thống kê là danh mục của một hệ thống chỉ tiêu tổng hợp.

3.2.3. Tổ chức và kỹ thuật tổng hợp thống kê

Có hai hình thức tổ chức tổng hợp thống kê là: tổng hợp từng cấp và tổng hợp tập trung.

- Tổng hợp từng cấp là tổng hợp tài liệu điều tra từ cấp dưới lên cấp trên theo một kế hoạch đã định sẵn.

- Tổng hợp tập trung là toàn bộ tài liệu điều tra được tập trung về một cơ quan để tiến hành tổng hợp.

Kỹ thuật tổng hợp thống kê có hai loại là: tổng hợp thủ công và tổng hợp bằng máy.

- Tổng hợp thủ công là tổng hợp bằng tay hay dùng một số phương tiện tính toán đơn giản, khi tài liệu ban đầu không nhiều. Nó được tiến hành theo 3 bước: sắp xếp tài liệu vào từng tổ, tính số đơn vị mỗi tổ, tính các số cộng và tổng cộng của hàng và cột.

- Tổng hợp bằng máy là sử dụng hệ thống máy móc chuyên môn để tổng hợp thống kê. Đây là biện pháp quan trọng để nâng cao năng suất lao động trong tổng hợp, đảm bảo số liệu tổng hợp nhanh chóng chính xác.

3.3. Bảng thống kê và đồ thị thống kê

3.3.1. Bảng thống kê

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

Đặc điểm chung của bảng thống kê là bao giờ cũng có những con số cộng và tổng cộng. Các con số này có mối liên hệ mật thiết với nhau.

Bảng thống kê có tác dụng rất lớn trong công tác nghiên cứu và phân tích thống kê:

- Giúp ta so sánh đối chiếu, phân tích theo các phương pháp khác nhau.
- Bảng thống kê nếu được trình bày tốt thì có sức thuyết phục rất lớn.

Cấu thành của bảng thống kê về mặt hình thức bao gồm: các hàng ngang, cột dọc phản ánh quy mô của bảng thống kê; các ô số liệu để điền số liệu thống kê; các tiêu đề gồm tiêu đề chung là tên gọi chung của bảng thống kê và các tiêu đề của các hàng ngang, cột dọc phản ánh nội dung, ý nghĩa của các hàng và cột.

Về nội dung của bảng thống kê gồm hai phần: phần chủ đề và phần giải thích.

Phần chủ đề thường đặt bên trái bảng, nêu lên đối tượng nghiên cứu của bảng thống kê chỉ rõ bảng thống kê nói về cái gì, gồm những đơn vị nào, tổ nào. Có khi là các địa phương hoặc các thời gian nghiên cứu khác nhau của hiện tượng nghiên cứu.

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

Khi lập bảng thống kê cần chú ý một số quy tắc sau:

- Quy mô bảng không nên quá lớn, quá phức tạp.
- Các tiêu đề cần ghi chính xác, rõ ràng, dễ hiểu.
- Các hàng và cột được ký hiệu bằng chữ hoặc số.
- Các chỉ tiêu giải thích cần được sắp xếp theo thứ tự hợp lý, các chỉ tiêu thực hiện kế hoạch bố trí gần chỉ tiêu kế hoạch, chỉ tiêu tương đối bố trí gần chỉ tiêu tuyệt đối.

Cách ghi số liệu vào bảng thống kê:

- Đơn vị tính của các số liệu phải rõ ràng, tránh bỏ sót.
- Trường hợp không ghi số liệu vào các ô trong bảng thì dùng các ký hiệu quy ước như: dấu gạch ngang (-) nếu ô không có số liệu, dấu ba chấm (...) nếu ô thiếu số liệu, sau này có thể bổ sung, dấu gạch chéo (x) nếu ô không có liên quan giữa chỉ tiêu với hiện tượng nghiên cứu.

Các loại bảng thống kê gồm có:

- Bảng giản đơn: là loại bảng thống kê trong đó phần chủ đề không phân tổ mà chỉ liệt kê các đơn vị tổng thể, các địa phương hoặc các thời gian khác nhau của hiện tượng nghiên cứu.
- Bảng phân tổ: là loại bảng thống kê trong đó đối tượng nghiên cứu trong phần chủ đề được phân tổ theo một tiêu thức nào đó.
- Bảng kết hợp: là loại bảng thống kê trong đó đối tượng nghiên cứu trong phần chủ đề được phân tổ theo 2, 3, ... tiêu thức khác nhau.

3.3.2. Đồ thị thống kê

Đồ thị thống kê là dùng các hình vẽ, đường nét hình học cùng với màu sắc thích hợp để biểu hiện các đặc trưng về mặt lượng của hiện tượng.

Các loại đồ thị thống kê thường dùng như: biểu đồ hình cột, biểu đồ diện tích (hình vuông, hình tròn, hình quạt...), biểu đồ đường gấp khúc....

IV. PHÂN TÍCH VÀ DỰ ĐOÁN THỐNG KÊ

4.1. Phân tích thống kê

4.1.1. Khái niệm, ý nghĩa, nhiệm vụ phân tích thống kê

Phân tích thống kê là nêu lên một cách tổng hợp bản chất và tính quy luật của hiện tượng kinh tế - xã hội thông qua biểu hiện bằng số lượng của hiện tượng trong điều kiện địa điểm và thời gian cụ thể nhằm đưa ra những căn cứ cho quyết định quản lý.

Phân tích thống kê là giai đoạn cuối cùng của quá trình nghiên cứu thống kê, nó biểu hiện tập trung kết quả của toàn bộ quá trình nghiên cứu thống kê. Các tài liệu của điều tra và tổng hợp thống kê chỉ có qua phân tích sâu sắc và toàn diện mới nêu lên biểu hiện về lượng bản chất và tính quy luật của hiện tượng nghiên cứu. Nhiệm vụ chung của phân tích thống kê là nêu rõ bản chất, tính quy luật và sự phát triển của hiện tượng kinh tế - xã hội.

4.1.2. Những yêu cầu có tính nguyên tắc cần được tuân thủ trong phân tích thống kê

- Khi phân tích thống kê phải dựa trên cơ sở phân tích lý luận của chủ nghĩa duy vật biện chứng và chủ nghĩa duy vật lịch sử để thấy được bản chất của hiện tượng trong quá trình phát sinh và phát triển.

- Khi phân tích phải căn cứ vào toàn bộ sự kiện và đặt chúng trong mối liên hệ tác động với nhau, để tìm ra mối liên quan giữa các hiện tượng trong một tổng thể chung.

- Khi phân tích đối với các hiện tượng có tính chất và hình thức khác nhau, phải áp dụng các phương pháp khác nhau, không thể áp dụng một phương pháp chung cho tất cả các hiện tượng.

4.1.3. Những vấn đề chủ yếu trong phân tích thống kê

- *Mục đích của phân tích thống kê:* là xác định những vấn đề mà phân tích thống kê cần giải quyết trong một phạm vi nhất định.

Ví dụ: khi phân tích tình hình sản xuất kinh doanh của một doanh nghiệp, mục đích cụ thể được xác định là: nêu rõ ưu nhược điểm, tìm nguyên nhân và biện pháp khắc phục trong kỳ tới để đưa sản xuất phát triển.

- *Lựa chọn đánh giá tài liệu dùng cho phân tích:* căn cứ vào mục đích phân tích để lựa chọn những tài liệu thật cần thiết gồm các tài liệu chính và các tài liệu có liên quan.

Chất lượng tài liệu có ảnh hưởng đến chất lượng phân tích, do đó các tài liệu cần phải được đánh giá trên các mặt sau đây:

+ Tài liệu thu thập được có đảm bảo các yêu cầu chính xác, kịp thời, đầy đủ không? Phương pháp thu thập tài liệu có khoa học hay không?

+ Tài liệu có được chỉnh lý, hệ thống hóa khoa học hay không? Có đáp ứng yêu cầu mục đích phân tích không?

+ Các chỉ tiêu được tính toán theo phương pháp nào? Có thống nhất với phương pháp thống kê hay không?

- *Xác định các phương pháp, các chỉ tiêu phân tích:* Thống kê học có nhiều phương pháp để phân tích như: nhóm các phương pháp nghiên cứu các mức độ của hiện tượng (số tuyệt đối, số tương đối, số bình quân), nhóm các phương pháp nghiên cứu sự biến động của hiện tượng (dãy số thời gian, chỉ số, hệ thống chỉ số), nhóm các phương pháp nghiên cứu mối liên hệ giữa các hiện tượng (phân tổ, phương pháp tương quan,...).

Để lựa chọn phương pháp phân tích cho từng trường hợp cụ thể phải chú ý các đặc điểm sau đây:

+ Phải căn cứ vào mục đích phân tích và đặc điểm tính chất của hiện tượng nghiên cứu để chọn phương pháp thích hợp.

+ Phải hiểu rõ ưu nhược điểm và điều kiện áp dụng của từng phương pháp.

+ Phải sử dụng kết hợp nhiều phương pháp để làm cho phân tích sâu sắc và toàn diện.

Khi phân tích phải xác định các chỉ tiêu cần thiết phù hợp với mục đích nghiên cứu. Khi xác định các chỉ tiêu cần chú ý:

+ Phải lựa chọn chỉ tiêu quan trọng nhất phản ánh đúng đắn bản chất của hiện tượng nghiên cứu.

+ Các chỉ tiêu thống kê phải có sự liên hệ bổ sung cho nhau.

- *So sánh, đối chiếu các chỉ tiêu*

So sánh đối chiếu các chỉ tiêu với nhau sẽ thấy được các đặc điểm, bản chất, xu hướng phát triển và tính quy luật của hiện tượng nghiên cứu. Khi so sánh đối chiếu các chỉ tiêu cần chú ý đảm bảo tính chất có thể so sánh được.

- *Dự đoán các mức độ tương lai của hiện tượng*

Dự đoán thống kê là căn cứ vào tài liệu thống kê về hiện tượng nghiên cứu trong thời gian đã qua và sử dụng các phương pháp thích hợp để tính toán mức độ tương lai hiện tượng.

- *Đề xuất các quyết định quản lý*

Các quyết định quản lý được đề xuất trên cơ sở phân tích và khẳng định được ưu, nhược điểm và các tồn tại cần quan tâm giải quyết. Các ý kiến đề xuất cho công tác quản lý phải có ý nghĩa thiết thực, phù hợp với hoàn cảnh thực tế và có khả năng thực hiện được.

Nội dung và phương pháp phân tích thống kê sẽ được đề cập tiếp trong các chương sau.

4.2. Dự đoán thống kê

Dự đoán thống kê là sự tiếp tục của quá trình phân tích thống kê, trong đó sử dụng các phương pháp sẵn có của thống kê để đánh giá về tương lai của các hiện tượng kinh tế - xã hội bằng những con số cụ thể.

Dự đoán là một công việc cần thiết và phù hợp với khả năng của thống kê, vì thống kê nắm được phần lớn các thông tin thực hiện về mọi mặt hoạt động kinh tế - xã hội và có một hệ thống các phương pháp chuyên môn thích hợp để dự đoán.

Tuy nhiên, không phải bất kỳ hiện tượng kinh tế - xã hội nào cũng đòi hỏi phải dự đoán. Thông thường, thống kê tiến hành dự đoán những hiện tượng kinh tế - xã hội quan trọng nhất gắn liền với việc xây dựng và thực hiện kế hoạch, hoặc những hiện tượng và quá trình chịu nhiều ảnh hưởng của các nhân tố ngẫu nhiên, hoặc có sự xuất hiện và phát triển của các hiện tượng và quá trình mới.

Đặc điểm lớn nhất của dự đoán thống kê là tính chất nhiều phương án với những xác suất tin cậy nhất định. Dựa trên cơ sở đó, các cơ quan sử dụng có thể chọn lấy một phương án dự đoán nào mà qua phân tích bổ sung thấy là tốt nhất.

Tuỳ theo tính chất của hiện tượng nghiên cứu và nhiệm vụ cụ thể của dự đoán, thống kê có thể có những dự đoán khác nhau: dài hạn, trung hạn và ngắn hạn.

Dự đoán dài hạn thường được lập cho khoảng thời gian 10 - 30 năm và lâu hơn nữa. Dự đoán dài hạn nhằm phục vụ cho việc xây dựng các kế hoạch và chương trình dài hạn. Thông thường, người ta chỉ dự đoán dài hạn cho các chỉ tiêu kinh tế quan trọng nhất, các

mối liên hệ liên ngành, các mục tiêu chiến lược và xác định các hiện tượng và quá trình kinh tế mới sẽ xuất hiện.

Dự đoán trung hạn thường được lập cho 3 - 5 năm hoặc dưới 10 năm. Dự đoán trung hạn chủ yếu nhằm phục vụ cho việc xây dựng các kế hoạch 5 năm hoặc các chương trình phát triển kinh tế trung hạn.

Dự đoán ngắn hạn có thời hạn dự đoán rất linh hoạt từ vài ngày, tuần, tháng, đến 1 năm hoặc 2, 3 năm. Dự đoán ngắn hạn phục vụ cho việc lập các kế hoạch và chương trình ngắn hạn. Ngoài ra, dự đoán ngắn hạn còn có tác dụng tích cực trong việc quản lý kinh tế, kịp thời tác động để điều khiển và điều chỉnh các hoạt động sản xuất kinh doanh ở các đơn vị cơ sở.

Thông kê trong các đơn vị kinh tế cơ sở thường làm các dự đoán ngắn hạn, vì nó thiết thực phục vụ cho các yêu cầu lãnh đạo và quản lý kinh tế đơn vị.

Chương III

ĐIỀU TRA CHỌN MẪU

I. KHÁI NIỆM, Ý NGHĨA CỦA ĐIỀU TRA CHỌN MẪU

Điều tra chọn mẫu là loại điều tra không toàn bộ, trong đó người ta chọn ra một số đơn vị thuộc tổng thể nghiên cứu để tiến hành điều tra thực tế, sau đó dùng kết quả này để tính toán suy rộng thành các đặc điểm chung của tổng thể.

Điều tra chọn mẫu ra đời từ lâu trên cơ sở vận dụng lý thuyết xác suất thống kê toán. Phương pháp này có những ưu điểm sau:

- Tiết kiệm thời gian và chi phí, đảm bảo tính kịp thời.
- Nội dung điều tra có thể mở rộng thêm, có thể đi sâu nghiên cứu nhiều mặt của hiện tượng.
- Tài liệu thu thập có thể đạt mức độ chính xác cao.

Tuy nhiên, điều tra chọn mẫu cũng có những mặt nhược điểm hạn chế. Do chỉ điều tra ở một số ít đơn vị rồi suy rộng, nên kết quả luôn có sai số nhất định không thể tránh khỏi.

Điều tra chọn mẫu sử dụng thích hợp nhất với những đối tượng nghiên cứu không cho phép tiến hành điều tra toàn bộ như: kiểm tra chất lượng sản phẩm, điều tra năng suất, điều tra mức sống của dân cư...

II. NHỮNG VẤN ĐỀ LÝ LUẬN VỀ ĐIỀU TRA CHỌN MẪU

2.1. Tổng thể chung và tổng thể mẫu

Tổng thể chung là tổng thể bao gồm tất cả các đơn vị thuộc đối tượng điều tra, ký hiệu là N .

Tổng thể mẫu là tổng thể bao gồm một số đơn vị nhất định được chọn ra từ tổng thể chung để điều tra thu thập tài liệu. Tổng thể mẫu có kích thước nhỏ hơn tổng thể chung, ký hiệu là n .

Ví dụ: trong một đợt sản xuất 120.000 sản phẩm, người ta chọn 1.000 sản phẩm để điều tra chất lượng sản phẩm. Vậy tổng thể chung $N = 120.000$ và số đơn vị tổng thể mẫu $n = 1.000$.

Sau khi tiến hành điều tra tổng thể mẫu, ta có thể tính ra các chỉ tiêu sau đây làm căn cứ suy rộng thành các chỉ tiêu tương ứng của tổng thể chung.

- Từ số bình quân mẫu $\bar{x}$ có thể suy rộng ra số bình quân chung $\bar{x}$.

Ngoài ra, căn cứ vào các số bình quân trên có thể tính ra phương sai mẫu σ_0^2 và phương sai chung σ^2 .

$$\sigma_0^2 = \frac{\sum (x_i - \bar{x})^2}{n}, \text{ nếu có tần số } \sigma_0^2 = \frac{\sum (x_i - \bar{x})^2 n_i}{\sum n_i}$$

$$\sigma^2 = \frac{\sum (x_i - \bar{x})^2}{N}, \text{ nếu có tần số } \sigma^2 = \frac{\sum (x_i - \bar{x})^2 N_i}{\sum N_i}$$

- Từ tỷ lệ mẫu w có thể suy rộng ra tỷ lệ cùng loại chiếm trong tổng thể chung, gọi là tỷ lệ chung p .

Trong trường hợp này phương sai mẫu là $w(1-w)$ và phương sai chung là pq .

Ví dụ: trong một đợt sản xuất 120.000 sản phẩm, trong đó có 114.000 sản phẩm loại 1. Người ta chọn ra 1.000 sản phẩm để điều tra chất lượng sản phẩm, thì thấy có 970 sản phẩm loại 1.

Vậy tỷ lệ sản phẩm loại 1 điều tra được là: $\frac{970}{1.000} \times 100 = 97\%$ là tỷ lệ mẫu theo kết quả điều tra chọn mẫu.

Tỷ lệ sản phẩm đạt loại 1 là: $\frac{114.000}{120.000} \times 100 = 95\%$ là tỷ lệ chung của tổng số sản phẩm sản xuất.

2.2. Sai số trong điều tra chọn mẫu

Sai số trong điều tra chọn mẫu là chênh lệch giữa giá trị của tổng thể mẫu với giá trị thực tế của tổng thể chung, tức là chênh lệch giữa các số bình quân $(\bar{x} - \bar{x})$ và giữa các tỷ lệ $(w - p)$.

Qua ví dụ trên ta thấy, giữa tỷ lệ mẫu và tỷ lệ chung chênh lệch là $97\% - 95\% = 2\%$, đây chính là sai số trong chọn mẫu khi suy rộng tài liệu. Sai số thường là không biết, vì ta không biết giá trị của tổng thể chung,

Sai số chọn mẫu tồn tại trong bản thân điều tra chọn mẫu, phụ thuộc vào tính chất đại biểu của tổng thể mẫu đối với tổng thể chung. Tính chất đại biểu của tổng thể mẫu càng thấp thì sai số chọn mẫu càng lớn. Tính chất đại biểu của tổng thể mẫu lại phụ thuộc vào các yếu tố như: quy mô của tổng thể mẫu, tính chất đồng đều của tổng thể chung và các phương pháp chọn mẫu.

Sai số chọn mẫu là một trị số không cố định, vì từ một tổng thể chung có thể thành lập nhiều tổng thể mẫu khác nhau. Sai số bình quân chọn mẫu là kết quả tổng hợp các sai số khác nhau do việc lựa chọn tổng thể mẫu có kết cấu thay đổi.

Trường hợp chọn mẫu ngẫu nhiên nhiều lần (chọn lặp), tức là chọn ngẫu nhiên từng đơn vị mẫu để ghi chép, đăng ký số liệu, sau đó đơn vị được chọn lại được trả về tổng thể chung và tiếp tục chọn đơn vị kế tiếp cho đến khi đủ số đơn vị mẫu mới thôi. Như vậy, mỗi đơn vị tổng thể chung đều có khả năng được chọn ra đăng ký nhiều lần. Công thức tính sai số bình quân chọn mẫu trong trường hợp chọn mẫu này như sau:

- Nếu để suy rộng chỉ tiêu bình quân:

$$\mu_x = \sqrt{\frac{\sigma^2}{n}}$$

- Nếu để suy rộng chỉ tiêu tương đối:

$$\mu_p = \sqrt{\frac{pq}{n}} = \sqrt{\frac{p(1-p)}{n}}$$

Trường hợp chọn mẫu ngẫu nhiên một lần (chọn không lặp) tức là chọn ngẫu nhiên đơn thuần từng đơn vị để điều tra, sau đó đơn vị được chọn không được trả lại vào tổng thể chung và tiếp tục chọn đơn vị tiếp theo. Như vậy, sau mỗi lần chọn, tổng thể chung giảm đi một đơn vị và mỗi đơn vị tổng thể chỉ được chọn một lần. Công thức tính sai số bình quân chọn mẫu trong trường hợp này như sau:

- Nếu để suy rộng chỉ tiêu bình quân:

$$\mu_x = \sqrt{\frac{\sigma^2}{n} \left(1 - \frac{n}{N}\right)}$$

- Nếu để suy rộng chỉ tiêu tương đối:

$$\mu_p = \sqrt{\frac{pq}{n} \left(1 - \frac{n}{N}\right)}$$

Vì $n < N$ nên $1 - \frac{n}{N} < 1$: như vậy sai số bình quân trong trường hợp chọn mẫu một lần nhỏ hơn sai số bình quân chọn mẫu nhiều lần.

Nếu số đơn vị tổng thể mẫu chiếm tỷ lệ không đáng kể so với tổng thể chung thì $1 - \frac{n}{N} \approx 1$ và sai số bình quân trong hai cách chọn một lần và nhiều lần không chênh lệch nhiều. Vì vậy, nếu hiện tượng nghiên cứu có nhiều đơn vị tổng thể, cho dù chọn mẫu theo phương pháp nào người ta vẫn tính sai số bình quân chọn mẫu theo công thức chọn nhiều lần.

Nếu không có tài liệu về phương sai, thì phải tính gần đúng sai số bình quân chọn mẫu bằng cách thay thế phương sai chung bằng phương sai mẫu.

Trường hợp chọn mẫu ngẫu nhiên nhiều lần:

- Nếu để suy rộng chỉ tiêu bình quân:
$$\mu_x = \sqrt{\frac{\sigma_0^2}{n}}$$

- Nếu để suy rộng chỉ tiêu tương đối:
$$\mu_p = \sqrt{\frac{w(1-w)}{n}}$$

Trường hợp chọn mẫu ngẫu nhiên một lần:

- Nếu để suy rộng chỉ tiêu bình quân:
$$\mu_x = \sqrt{\frac{\sigma_0^2}{n} \left(1 - \frac{n}{N}\right)}$$

- Nếu để suy rộng chỉ tiêu tương đối:
$$\mu_p = \sqrt{\frac{w(1-w)}{n} \left(1 - \frac{n}{N}\right)}$$

2.3. Phạm vi sai số chọn mẫu

Sai số chọn mẫu có thể nằm trong phạm vi chênh lệch nhiều hơn hoặc ít hơn so với các chỉ tiêu của tổng thể chung. Vì vậy, chênh lệch của $\bar{x} - \bar{x}$ và $w - p$ nằm trong phạm vi $\pm\mu$. Theo chứng minh của lý thuyết xác suất, nếu sai số chọn mẫu nằm trong khoảng $\pm\mu$ thì xác suất hay trình độ tin cậy của việc suy rộng tài liệu là 0.6827. Tức là trong 10.000 trường hợp chọn mẫu, thì có 6.827 trường hợp chắc chắn sai số chọn mẫu không vượt quá phạm vi $\pm\mu$.

Để nâng cao trình độ tin cậy của việc suy rộng tài liệu, người ta mở rộng thêm phạm vi sai số chọn mẫu.

Nếu phạm vi này được mở rộng lên $\pm 2\mu$ thì xác suất của việc suy rộng lên đều 0.9545. Tức là trong 10.000 trường hợp chọn mẫu thì có 9.545 trường hợp chắc chắn sai số chọn mẫu không vượt quá $\pm 2\mu$. Tương tự nếu mở rộng phạm vi chọn mẫu lên $\pm 3\mu$ thì xác suất của việc suy rộng lên đến 0.9973...

$$P(|\bar{x} - \bar{x}| < 3) = 2\Phi(3) = 0,9973$$

$$P(|\bar{x} - \bar{x}| < 2) = 2\Phi(2) = 0,9545$$

Trong đó $\Phi(t)$ là hàm Laplace Gauss và:

$$\Phi(t) = \frac{1}{\sqrt{2\pi}} \int_{-t}^t e^{-\frac{t^2}{2}} dt$$

Vậy phạm vi sai số chọn mẫu càng mở rộng thì xác suất của việc suy rộng tài liệu càng tăng, điều đó dẫn đến sai số chọn mẫu cũng tăng theo phạm vi sai số chọn mẫu được tính theo công thức:

$$\varepsilon = t.\mu$$

trong đó:

ε - phạm vi sai số chọn mẫu.

t - hệ số tin cậy.

μ - sai số bình quân chọn mẫu.

$\Phi(t)$ - trình độ tin cậy.

Trong thực tế thường sử dụng bảng tính sẵn trị số của hàm $\Phi(t)$, nếu biết t suy ra $\Phi(t)$ và ngược lại.

2.4. Xác định số đơn vị tổng thể mẫu

Ta có thể dựa vào các công thức xác định phạm vi sai số chọn mẫu để tính ra số đơn vị tổng thể mẫu.

- Nếu để suy rộng chỉ tiêu bình quân:

+ Chọn nhiều lần:
$$n = \frac{t^2 \cdot \sigma^2}{\varepsilon^2}$$

+ Chọn một lần:
$$n = \frac{t^2 \cdot \sigma^2 \cdot N}{\varepsilon^2 \cdot N + t^2 \cdot \sigma^2}$$

- Nếu để suy rộng chỉ tiêu tương đối:

+ Chọn nhiều lần:
$$n = \frac{t^2 \cdot p \cdot (1-p)}{\varepsilon^2}$$

+ Chọn một lần:
$$n = \frac{t^2 \cdot p \cdot (1-p) \cdot N}{\varepsilon^2 \cdot N + t^2 \cdot p \cdot (1-p)}$$

Trong các công thức trên ta thấy số mẫu (n) phụ thuộc vào các nhân tố sau: phạm vi sai số chọn mẫu (ε), hệ số tin cậy (t) và phương sai của tổng thể chung (σ^2).

Trong thực tế khi xác định số mẫu (n) cần điều tra thường khó khăn vì không có phương sai chung. Có thể giải quyết theo các hướng sau đây:

- Tính ra phương sai dùng cho lần điều tra này, căn cứ vào tài liệu điều tra của nhiều cuộc điều tra chọn mẫu tương tự trước đây đối với hiện tượng nghiên cứu.

- Chọn phương sai nào lớn nhất cho điều tra lần này, căn cứ vào tài liệu điều tra của nhiều cuộc điều tra chọn mẫu tương tự trước đây đối với hiện tượng nghiên cứu. Khi điều tra để tính chỉ tiêu tương đối, nên chọn chỉ tiêu nào có trị số gần 0,5 nhất, như vậy

phương sai tính ra sẽ lớn hơn cả, để số mẫu n chọn ra phải tăng lên mới đảm bảo tính đại biểu của mẫu.

- Chọn phương sai của các cuộc điều tra chọn mẫu ở nơi khác, nếu hiện tượng nghiên cứu nơi khác có những đặc điểm và điều kiện tương tự.

- Điều tra chọn mẫu thí điểm trong phạm vi nhỏ để tính toán gần đúng chỉ tiêu cần thiết.

Ví dụ 1: để nghiên cứu nhu cầu tiêu dùng vải của nhân dân tại một địa phương, người ta tiến hành điều tra chọn mẫu. Biết rằng, độ lệch tiêu chuẩn về mức tiêu dùng vải là 25,5 mét, yêu cầu sai số không quá 1,5 mét đối với mỗi hộ gia đình và độ tin cậy là 95,45%. Xác định số đơn vị cần điều tra để đạt được các yêu cầu trên?

Ta có: $\Phi(t) = 0,9545 \Rightarrow t = 2$

$\sigma = 25,5$ mét ; $\varepsilon = 1,5$ mét

$$\text{Số hộ cần điều tra là: } n = \frac{t^2 \cdot \sigma^2}{\varepsilon^2} = \frac{2^2 \times (25,5)^2}{(1,5)^2} = 1.156 \text{ hộ}$$

Ví dụ 2: để kiểm tra chất lượng của một loại sản phẩm sản xuất tại doanh nghiệp, người ta sử dụng phương pháp điều tra chọn mẫu. Yêu cầu trình độ tin cậy của suy rộng tài liệu là 95,45%, phạm vi sai số cho phép 4%. Biết rằng trong các lần điều tra trước tỷ lệ sản phẩm loại I là 60% , 70% và 65%. Xác định số đơn vị cần điều tra để đạt được các yêu cầu trên.

Ta có: $\Phi(t) = 0,9545 \Rightarrow t = 2$

$p = 60\%$ hay 0,6 (là trị số gần 0,5 nhất) ; $\varepsilon = 0,04$

Số sản phẩm cần điều tra là:

$$n = \frac{t^2 \cdot p \cdot (1-p)}{\varepsilon^2} = \frac{2^2 \times 0,6 \times (1-0,6)}{(0,04)^2} = 600 \text{ sản phẩm}$$

2.5. Một số phương pháp chọn mẫu trong thống kê

Sai số chọn mẫu phụ thuộc vào các phương pháp chọn mẫu. Cách chọn mẫu khác nhau dẫn đến các sai số chọn mẫu khác nhau. Phương pháp chọn mẫu khoa học là phương pháp đáp ứng yêu cầu của lý thuyết chọn mẫu, nghĩa là chọn ngẫu nhiên, khách quan, đồng thời phù hợp với đặc điểm của tổng thể chung. Trong thống kê thường dùng các phương pháp chọn mẫu chủ yếu sau đây:

2.5.1. Chọn ngẫu nhiên đơn thuần

Là phương pháp lấy các số đơn vị mẫu một cách hoàn toàn ngẫu nhiên, không qua bất kỳ một sự sắp xếp nào như: rút thăm, quay số, chọn theo bảng số ngẫu nhiên... Mỗi đơn vị của tổng thể chung có thể được chọn một lần (chọn không lặp) hoặc chọn nhiều lần (chọn lặp).

Chọn ngẫu nhiên đơn thuần có ưu điểm là tạo ra một tổng thể mẫu khách quan, cho kết quả chính xác nếu độ biến thiên tiêu thức của các đơn vị tổng thể không lớn lắm, kết cấu tổng thể không phức tạp. Tuy nhiên, nếu tổng thể chung có quá nhiều đơn vị, kết cấu phức tạp thì cách chọn này có thể sai sót và tốn công sức.

2.5.2. Chọn mẫu máy móc

Là phương pháp chọn mẫu một cách ngẫu nhiên, căn cứ vào khoảng cách nhất định từ danh sách các đơn vị của tổng thể chung được sắp xếp theo một thứ tự nào đó như: theo thứ tự vần A, B, C ... của tên gọi, theo thứ tự địa phương, theo quy mô từ nhỏ đến lớn v.v...

$$\text{Khoảng cách chọn một đơn vị điều tra} = \frac{\text{Số đơn vị tổng thể chung}}{\text{Số đơn vị tổng thể mẫu}}$$

Trong chọn mẫu máy móc, các đơn vị mẫu được chọn theo một khoảng cách nhất định, nên được phân phối đều trong tổng thể chung theo tiêu thức làm căn cứ lập danh sách, vì vậy tính chất đại biểu của mẫu được nâng cao. Tuy nhiên, nếu quy mô tổng thể chung quá lớn thì việc lập danh sách có thể gặp nhiều khó khăn.

2.5.3. Chọn mẫu phân loại (chọn mẫu có phân tổ)

Là phương pháp chọn các đơn vị điều tra từ tổng thể chung đã được phân tổ theo tiêu thức liên quan đến nội dung nghiên cứu. Mỗi tổ thường đại diện cho một loại hình kinh tế - xã hội. Sau đó trong phạm vi mỗi tổ tiến hành chọn ra một số đơn vị nhất định theo cách chọn ngẫu nhiên đơn thuần hoặc chọn máy móc. Số đơn vị được chọn từ mỗi tổ có thể tương ứng hoặc không tương ứng với tỷ trọng của tổ trong tổng thể chung.

Chọn mẫu phân loại theo tỷ lệ đảm bảo kết cấu của mẫu tương tự kết cấu của tổng thể chung xét theo tiêu thức phân tổ. Từ đó tính chất đại biểu của mẫu được nâng cao. Theo phương pháp này, có thể sử dụng tài liệu của số mẫu từng tổ để nghiên cứu đặc điểm tổ đó, nên giúp cho việc nghiên cứu tổng thể được sâu sắc, toàn diện hơn.

2.5.4. Chọn cả khối

Là phương pháp chọn những đơn vị tổng thể mẫu từ tổng thể chung, bằng cách chọn từng khối đơn vị mẫu cùng một lúc theo cách chọn ngẫu nhiên đơn thuần hay chọn máy móc. Như vậy tổng thể chung được chia thành nhiều khối và chọn ra một số khối mẫu để điều tra toàn bộ các đơn vị trong khối mẫu đó.

Chọn cả khối có ưu điểm là tiến hành đơn giản, nhưng tính chất đại biểu của tổng thể mẫu không cao, sai số chọn mẫu có khả năng lớn hơn so với các phương pháp khác.

III. SUY RỘNG KẾT QUẢ ĐIỀU TRA CHỌN MẪU

Suy rộng kết quả điều tra chọn mẫu là tính toán các chỉ tiêu của tổng thể chung, trên cơ sở của tài liệu đã thu thập được từ tổng thể mẫu.

Công thức tính:

- Để suy rộng ra chỉ tiêu bình quân:

$$\bar{x} - \varepsilon \leq \bar{x} \leq \bar{x} + \varepsilon$$

- Để suy rộng ra chỉ tiêu tương đối:

$$w - \varepsilon \leq p \leq w + \varepsilon$$

Ví dụ 1: năng suất lao động bình quân một công nhân tính được trong điều tra của tổng thể mẫu ở doanh nghiệp A là $\bar{x} = 20$ sản phẩm/ngày/người, với xác suất 0.9545 phạm vi sai số chọn mẫu $\varepsilon = 0,2$ sản phẩm/ngày/người.

Như vậy, suy ra năng suất lao động bình quân một công nhân của doanh nghiệp nằm trong phạm vi là:

$$20 - 0,2 \leq \bar{x} \leq 20 + 0,2$$

hay $19,8 \leq \bar{x} \leq 20,2$ sản phẩm/ngày/người.

Ví dụ 2: tỷ lệ sản phẩm loại I điều tra được của tổng thể mẫu $w = 50\%$ với xác suất 0.9545, phạm vi sai số cho phép là $\varepsilon = 4\%$. Như vậy, tỷ lệ sản phẩm loại I của tổng thể chung nằm trong phạm vi là:

$$50\% - 4\% \leq p \leq 50\% + 4\%$$

hay $46\% \leq p \leq 54\%$

Chương IV

PHÂN TỔ THỐNG KÊ

I. KHÁI NIỆM, Ý NGHĨA, NHIỆM VỤ PHÂN TỔ THỐNG KÊ

1.1. Khái niệm

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

Ví dụ: khi nghiên cứu tình hình sản xuất của các doanh nghiệp, có thể chia các doanh nghiệp thành các tổ theo các tiêu thức như: “thành phần kinh tế”, “số lượng lao động”, “giá trị sản xuất”, “thu nhập bình quân của một lao động”,...

1.2. Ý nghĩa và nhiệm vụ của phân tổ thống kê

- Phân tổ thống kê có nhiều ý nghĩa trong nghiên cứu thống kê

+ Trong một số trường hợp điều tra thống kê người ta phải dùng đến phương pháp phân tổ.

Ví dụ: khi điều tra doanh thu của những người buôn bán trước hết phải chia số người buôn bán theo ngành hàng, nhóm hàng kinh doanh để thu thập số liệu của những người buôn bán theo từng ngành hàng, nhóm hàng đó.

+ Phân tổ thống kê là phương pháp cơ bản để tiến hành hệ thống hóa tài liệu một cách khoa học trong tổng hợp thống kê. Đây là công việc rất yếu khách quan, vì hiện tượng kinh tế xã hội rất phức tạp. Khi tổng hợp thống kê, các đơn vị trong tổng thể có cùng đặc điểm, tính chất được sắp xếp chung vào một tổ. Sau đó nghiên cứu đặc điểm riêng từng tổ và rút ra các đặc điểm chung của tổng thể.

+ Phân tổ thống kê là một trong những phương pháp quan trọng của phân tích thống kê và là cơ sở để áp dụng các phương pháp phân tích thống kê khác như phương pháp chỉ số, tương quan, bảng cân đối,...

- Phân tổ thống kê cần phải giải quyết các nhiệm vụ cơ bản sau:

+ Phân chia hiện tượng nghiên cứu theo các loại hình kinh tế - xã hội (phân tổ này được gọi là phân tổ phân loại).

Ví dụ: phân tổ các doanh nghiệp sản xuất công nghiệp theo ngành hoạt động, theo thành phần kinh tế,...

Các loại hình kinh tế - xã hội tồn tại khách quan. Sự vận động và phát triển của toàn bộ hiện tượng là kết quả đấu tranh giữa các loại hình đối lập tồn tại ngay trong bản thân hiện tượng. Do vậy, việc nêu rõ những loại hình tồn tại trong hiện tượng có ý nghĩa quan trọng.

+ Biểu hiện kết cấu của hiện tượng nghiên cứu (phân tổ kết cấu)

Mỗi hiện tượng kinh tế xã hội thường bao gồm nhiều bộ phận, nhiều nhóm đơn vị có tính chất khác nhau hợp thành. Các bộ phận, các nhóm này chiếm những tỷ trọng và biểu hiện tầm quan trọng của chúng trong tổng thể. Tỷ trọng của các bộ phận phản ánh kết cấu của tổng thể theo một tiêu thức nào đó.

Phân tổ kết cấu là xác định chính xác các bộ phận có tính chất khác nhau trong tổng thể, sau đó tính toán các tỷ trọng của các bộ phận này.

Ví dụ: phân tổ kết cấu dân số bình quân ở nước ta năm 1999 theo thành thị và nông thôn.

Bảng 3-1

Khu vực	Số dân (1.000 người)	Tỷ trọng (%)
Thành thị	18.076.823	23,7
Nông thôn	58.246.350	76,3
Cả nước	76.323.173	100,0

+ Biểu hiện mối liên hệ giữa các tiêu thức (phân tổ liên hệ)

Giữa các hiện tượng kinh tế xã hội hoặc giữa các tiêu thức thống kê thường có mối liên hệ và phụ thuộc lẫn nhau theo những quy luật nhất định. Sự biến động của hiện tượng (hoặc tiêu thức) này là kết quả tác động của các hiện tượng (hoặc tiêu thức) khác có liên quan. Nghiên cứu tính chất và mức độ của mối liên hệ giữa các hiện tượng (hoặc tiêu thức) là một trong những nhiệm vụ quan trọng của phân tổ thống kê. Theo tiêu thức nguyên nhân, phân tổ thống kê chia hiện tượng nghiên cứu thành các tổ và tính toán các mức độ của các bộ phận tương ứng theo tiêu thức kết quả. Qua đó xác định mối liên hệ có tính quy luật giữa hai tiêu thức.

Ví dụ: có tài liệu điều tra 10 công nhân tại một doanh nghiệp được sắp xếp theo tuổi nghề và sau đó tính mức năng suất lao động của các công nhân như sau:

Bảng 3-2

Tên công nhân	Tuổi nghề (năm)	Năng suất lao động (sản phẩm)
A	1	25
B	2	40
C	3	45
D	4	60
E	5	65
G	6	70
H	7	90
I	8	90
K	9	100
L	10	100

Qua đó, ta thấy tính quy luật về mối liên hệ giữa tuổi nghề (tiêu thức nguyên nhân) và năng suất lao động (tiêu thức kết quả) của 10 công nhân trên là tuổi nghề càng cao thì năng suất lao động càng cao.

II. TIÊU THỨC PHÂN TỔ

2.1. Khái niệm

Tiêu thức phân tổ là tiêu thức được chọn làm căn cứ để phân chia tổng thể hiện tượng nghiên cứu thành các tổ, các bộ phận có tính chất và đặc điểm khác nhau.

2.2. Căn cứ để lựa chọn tiêu thức phân tổ

Chọn tiêu thức phân tổ là vấn đề đầu tiên phải giải quyết khi tiến hành phân tổ thống kê. Do mỗi đơn vị tổng thể có nhiều tiêu thức khác nhau, tiêu thức nào cũng có thể dùng để phân tổ được. Tuy nhiên, có tiêu thức phản ánh đúng bản chất của hiện tượng, cũng có những tiêu thức không nêu rõ bản chất, đặc điểm của hiện tượng. Vì vậy, phải chọn tiêu thức phân tổ phản ánh đúng bản chất của hiện tượng nghiên cứu và phù hợp với mục đích nghiên cứu.

Để lựa chọn tiêu thức phân tổ phải căn cứ vào các yêu cầu sau đây:

- Phải dựa vào phân tích lý luận để chọn ra tiêu thức bản chất nhất phù hợp với mục đích nghiên cứu. Tiêu thức bản chất là tiêu thức nói lên được bản chất của hiện tượng nghiên cứu trong điều kiện thời gian và địa điểm cụ thể.

Ví dụ: khi dùng phương pháp phân tổ thống kê để nghiên cứu quy mô sản xuất kinh doanh của các doanh nghiệp, thì ta có thể chọn một số tiêu thức phân tổ như: số lượng lao động, số lượng thiết bị sản xuất, giá trị sản xuất,... Nhưng chọn tiêu thức nào là bản chất nhất phải dựa vào phân tích lý luận cụ thể. Đối với những doanh nghiệp mà quá trình sản xuất chủ yếu vẫn dựa vào sức lao động của công nhân, thì tiêu thức phân tổ có thể chọn là số lượng lao động. Còn đối với doanh nghiệp đã được cơ giới hóa, tự động hóa cao thì có thể chọn tiêu thức phân tổ là giá trị sản xuất hoặc số lượng thiết bị sản xuất.

- Phải căn cứ điều kiện lịch sử cụ thể của hiện tượng nghiên cứu để chọn ra tiêu thức phân tổ phù hợp. Cùng một loại hiện tượng nghiên cứu nhưng phát sinh trong những điều kiện thời gian và địa điểm khác nhau, thì bản chất có thể thay đổi khác nhau. Vì vậy tiêu thức phân tổ cũng mang ý nghĩa khác nhau. Một tiêu thức phân tổ không thể dùng chung cho mọi trường hợp, vì trong điều kiện này tiêu thức đó giúp ta nghiên cứu chính xác, nhưng trong điều kiện khác lại không có ý nghĩa.

III. PHÂN TỔ THỐNG KÊ

3.1. Phân tổ theo tiêu thức thuộc tính

Trong phân tổ này, số tổ được hình thành bằng số các loại hình khác nhau của hiện tượng nghiên cứu. Có hai trường hợp:

- Nếu số loại hình tương đối ít, có thể coi mỗi loại hình là một tổ.

Ví dụ: phân tổ dân số theo giới tính, phân tổ các doanh nghiệp theo thành phần kinh tế,...

- Nếu số loại hình thực tế có nhiều, nếu coi mỗi loại hình là một tổ thì số tổ sẽ rất nhiều không giúp ta nghiên cứu được các đặc trưng của tổng thể từ sự khác nhau giữa các tổ. Trong trường hợp này, phải ghép nhiều tổ nhỏ thành một số tổ lớn theo nguyên tắc các tổ nhỏ ghép lại phải giống nhau (hoặc gần giống nhau) về tính chất.

Trong thực tế, thống kê thường phân tổ theo bảng danh mục hay bảng phân loại do Nhà nước quy định thống nhất và ổn định trong một thời gian dài.

Ví dụ: Bảng danh mục hàng hóa, Bảng danh mục nghề nghiệp, Bảng phân ngành kinh tế quốc dân,...

3.2. Phân tổ theo tiêu thức số lượng

Tiêu thức số lượng là tiêu thức mà biểu hiện cụ thể của nó là những con số, những con số đó được gọi là lượng biến. Trong phân tổ này, phải căn cứ vào số lượng biến khác nhau của tiêu thức mà xác định các tổ khác nhau về tính chất. Có hai trường hợp:

- Trường hợp phân tổ không có khoảng cách tổ:

Được áp dụng khi lượng biến thay đổi ít, nghĩa là chênh lệch về lượng giữa các đơn vị không nhiều như: số người trong gia đình, số máy do một công nhân phụ trách,... thì số tổ được hình thành bằng số lượng biến.

Ví dụ: phân tổ số công nhân của một doanh nghiệp dệt theo số máy dệt mỗi công nhân phụ trách ở bảng sau:

Bảng 3-3

Số máy dệt mỗi CN phụ trách	Số công nhân (người)
11	6
12	14
13	40
14	100
15	80
16	30
Cộng	270

- Trường hợp phân tổ có khoảng cách tổ:

Được áp dụng khi lượng biến của tiêu thức này thay đổi lớn. Nếu mỗi lượng biến hình thành một tổ thì số tổ sẽ quá nhiều, đồng thời không nói rõ sự khác nhau về chất giữa các tổ. Trong trường hợp này cần chú ý tới mối liên hệ giữa lượng và chất của hiện tượng, xem lượng biến tích lũy đến mức độ nào thì chất của hiện tượng mới thay đổi và

làm nảy sinh một tổ khác. Như vậy, mỗi tổ sẽ bao gồm một phạm vi lượng biến, có hai giới hạn là “giới hạn trên” và “giới hạn dưới”. Giới hạn trên là lượng biến lớn nhất của tổ, giới hạn dưới là lượng biến nhỏ nhất của tổ. Trị số chênh lệch giữa hai giới hạn đó gọi là khoảng cách tổ. Khoảng cách tổ có thể đều nhau hoặc không đều nhau.

Khoảng cách tổ đều nhau được áp dụng khi hiện tượng biến động tương đối đồng đều. Trị số khoảng cách tổ đều được xác định như sau:

+ Đối với lượng biến liên tục, thành lập các tổ theo quy định sau: giới hạn dưới của tổ sau trùng với giới hạn trên của tổ trước và trị số của khoảng cách tổ được xác định theo công thức:

$$d = \frac{x_{\max} - x_{\min}}{n} \quad (1)$$

trong đó:

d – trị số khoảng cách tổ.

$x_{\max}$ – lượng biến lớn nhất của tiêu thức.

$x_{\min}$ – lượng biến nhỏ nhất của tiêu thức.

n – số tổ.

Ví dụ: phân tổ 30 công nhân tại một doanh nghiệp theo tiêu thức mức thu nhập tháng của một công nhân trong năm 2003. Biết rằng thu nhập lớn nhất là 1.040.000đ/tháng, thấp nhất là 940.000đ/tháng. Dự kiến chia thành 5 tổ, nên:

$$d = \frac{1.040.000 - 940.000}{5} = 20.000đ$$

Dựa vào $d = 20.000đ$ ta thành lập các tổ và sắp xếp số công nhân vào các tổ thích hợp. Ta có bảng phân tổ công nhân theo mức thu nhập tháng:

Bảng 3-4

Mức thu nhập tháng của 1 CN (đồng)	Số công nhân (người)
940.000 - 960.000	2
960.000 - 980.000	3
980.000 - 1.000.000	5
1.000.000 - 1.020.000	8
1.020.000 - 1.040.000	12
Cộng	30

+ Đối với lượng biến rời rạc, thành lập các tổ theo quy định sau: giới hạn dưới của tổ sau lớn hơn giới hạn trên của tổ trước và trị số của khoảng cách tổ được xác định theo công thức:

$$d = \frac{x_{\max} - x_{\min} - (n - 1)}{n} \quad (2)$$

Nội dung các ký hiệu giống như công thức (1).

Ví dụ: có tài liệu về số công nhân của 20 doanh nghiệp sản xuất công nghiệp trong năm 2003 như sau:

Bảng 3-5

DN	Số CN (người)	DN	Số CN (người)	DN	Số CN (người)	DN	Số CN (người)
1	1.200	6	1.430	11	1.650	16	2.883
2	1.304	7	1.350	12	2.050	17	2.540
3	1.500	8	1.240	13	2.120	18	2.760
4	1.670	9	1.700	14	1.980	19	2.300
5	1.400	10	1.800	15	2.400	20	2.130

Giả sử chia 20 doanh nghiệp này thành 4 tổ có khoảng cách đều nhau theo tiêu thức số công nhân.

Tiêu thức số công nhân là tiêu thức có lượng biến rời rạc, nên dùng công thức (2) để xác định trị số khoảng cách tổ:

$$d = \frac{(2.883 - 1.200) - (4 - 1)}{4} = 420 \text{ công nhân}$$

Dựa vào $d = 420$ công nhân ta thành lập các tổ và sắp xếp các đơn vị tổng thể vào các tổ thích hợp. Ta có bảng phân tổ các doanh nghiệp theo số công nhân:

Bảng 3-6

Số công nhân của doanh nghiệp	Số doanh nghiệp
1.200 - 1.620	7
1.621 - 2.041	5
2.042 - 2.462	5
2.463 - 2.883	3
Cộng	20

Khoảng cách tổ không đều được áp dụng khi hiện tượng biến động không đều, làm cho tính chất khác nhau giữa các tổ cũng không đều và còn tùy theo mục đích nghiên cứu mà xác định khoảng cách tổ đều hay không đều.

Ngoài việc xác định khoảng cách tổ ta còn gặp trường hợp tổ có khoảng cách tổ kín và tổ mở.

Tổ kín là tổ có đầy đủ hai giới hạn. Nếu thiếu một giới hạn gọi là tổ mở. Tổ mở thường gặp ở tổ đầu tiên (thiếu giới hạn dưới) hoặc tổ cuối cùng (thiếu giới hạn trên).

Ví dụ: phân tổ dân số tại một địa phương trong năm 2003 theo độ tuổi như sau:

Bảng 3-7

Độ tuổi	Số dân (1.000 người)	Ghi chú
- Dưới 1 tuổi	15	Còn bú mẹ
- Từ 1 - 3 tuổi	80	Nhà trẻ
- Từ 4 - 6 tuổi	70	Mẫu giáo
- Từ 7 - 18 tuổi	515	Học phổ thông
- Từ 19 - 60 tuổi	1.200	Tuổi lao động
- Từ 61 tuổi trở lên	120	Tuổi nghỉ ngơi
Cộng	2.000	

3.3. Phân tổ liên hệ

Phân tổ liên hệ là dùng phương pháp phân tổ để biểu hiện mối liên hệ giữa các tiêu thức. Các tiêu thức có liên hệ với nhau được chia làm hai loại là tiêu thức nguyên nhân và tiêu thức kết quả.

- Tiêu thức nguyên nhân là tiêu thức được coi là nguyên nhân chủ yếu làm cho tiêu thức liên quan biến động.

- Tiêu thức kết quả là tiêu thức biến động do ảnh hưởng của tiêu thức nguyên nhân và cần tập trung nghiên cứu sự biến động của nó.

Theo phương pháp phân tổ liên hệ, các đơn vị tổng thể được phân tổ theo tiêu thức nguyên nhân, sau đó trong mỗi tổ tính trị số bình quân của tiêu thức kết quả. Quan sát sự biến động của hai tiêu thức này sẽ rút ra nhận xét về mối liên hệ giữa tiêu thức nguyên nhân và tiêu thức kết quả.

Khi phân tổ liên hệ có thể gặp các trường hợp sau đây:

3.3.1. Phân tổ để nghiên cứu mối liên hệ giữa một tiêu thức nguyên nhân và một tiêu thức kết quả

Phân tổ liên hệ trong trường hợp này gọi là phân tổ giản đơn (phân tổ theo một tiêu thức). Trong trường hợp này sẽ thu được bảng phân tổ giản đơn nói lên mối liên hệ giữa hai tiêu thức.

Ví dụ: phân tổ để nghiên cứu mối liên hệ giữa mức độ cơ giới hoá và năng suất lao động của công nhân được phản ánh ở bảng sau:

Bảng 3-8

Phân theo mức độ cơ giới hoá (%)	Số công nhân (người)	NSLĐ 1CN (1000đ)	Giá trị sản xuất (1000đ)
< 30	30	540	16.200
30 - 50	50	700	35.000
> 50	20	800	16.000
Toàn doanh nghiệp	100	672	67.200

3.3.2. Phân tổ để nghiên cứu mối liên hệ giữa nhiều tiêu thức nguyên nhân và một tiêu thức kết quả

Trong trường hợp này tổng thể nghiên cứu được phân tổ theo tiêu thức kết hợp. Theo cách này tổng thể nghiên cứu trước hết được phân tổ theo tiêu thức nguyên nhân thứ nhất, sau đó mỗi tổ lại được phân thành các tiểu tổ theo tiêu thức nguyên nhân thứ hai,... Cuối cùng tính trị số bình quân của tiêu thức kết quả của từng tổ và tiểu tổ.

IV. CHỈ TIÊU GIẢI THÍCH

4.1. Khái niệm

Sau khi xác định được số tổ cần thiết còn phải xác định được các chỉ tiêu giải thích. Chỉ tiêu giải thích là các chỉ tiêu dùng để giải thích các đặc điểm riêng của từng tổ và toàn bộ tổng thể.

Ví dụ: sau khi phân tổ số công nhân theo thu nhập có thể xác định một số chỉ tiêu giải thích như: số công nhân, tổng thu nhập,... trong mỗi tổ.

4.2. Tác dụng của chỉ tiêu giải thích

Các chỉ tiêu giải thích giúp ta thấy rõ đặc trưng riêng về mặt lượng của từng tổ và của toàn bộ tổng thể, làm căn cứ so sánh các tổ với nhau và để tính toán hàng loạt các chỉ tiêu phân tích khác.

Muốn xác định chỉ tiêu giải thích phải căn cứ vào mục đích nghiên cứu và nhiệm vụ chủ yếu của phân tổ để chọn ra các chỉ tiêu có liên hệ và bổ sung cho nhau.

Ngoài ra cần chú ý mối quan hệ giữa chỉ tiêu giải thích với tiêu thức phân tổ.

Ví dụ: khi phân tổ các doanh nghiệp theo quy mô thì nên chọn các chỉ tiêu giải thích như: giá trị sản xuất, giá trị tài sản cố định, số lao động,... sẽ giúp ta hiểu rõ thêm về quy mô của doanh nghiệp.

Các chỉ tiêu giải thích cần được sắp xếp theo trình tự hợp lý để thuận tiện cho việc so sánh, nhận thức hiện tượng. Các chỉ tiêu có ý nghĩa quan trọng trong việc so sánh nên bố trí gần nhau.

V. DÂY SỐ PHÂN PHỐI

5.1. Khái niệm

Sau khi xác định được số tổ và sắp xếp các đơn vị tổng thể vào các tổ tương ứng ta có một dãy số phân phối. Dãy số phân phối là dãy số trình bày có thứ tự số lượng đơn vị tổng thể của từng tổ trong một tổng thể đã được phân tổ theo một tiêu thức nhất định.

5.2. Tác dụng

Trong thống kê dãy số phân phối được dùng để nghiên cứu kết cấu của tổng thể và sự biến động của kết cấu đó, nghiên cứu mối liên hệ giữa các bộ phận của tổng thể, tính toán một số chỉ tiêu bình quân đặc trưng cho tổng thể.

5.3. Các loại dãy số phân phối

Có hai loại dãy số phân phối:

- Dãy số thuộc tính (dãy số phân phối theo tiêu thức thuộc tính) phản ánh kết cấu của tổng thể theo một tiêu thức thuộc tính nào đó.

Ví dụ: dãy số phân phối dân số theo giới tính, dãy số phân phối giá trị sản xuất theo từng ngành kinh tế,...

- Dãy số lượng biến (dãy số phân phối theo tiêu thức số lượng) phản ánh kết cấu tổng thể theo một tiêu thức số lượng nào đó.

Ví dụ: dãy số phân phối dân số theo độ tuổi, dãy số phân phối số công nhân theo mức thu nhập bình quân một công nhân,...

Nếu ký hiệu x_i ($i = 1, 2, \dots, n$) là các trị số lượng biến, ta thấy ứng với mỗi x_i được phân phối một số đơn vị tổng thể nhất định được gọi là tần số, ký hiệu là n_i ($i = 1, 2, \dots, n$). Dãy số lượng biến có dạng chung như sau:

Lượng biến (x_i)	Tần số (n_i)
x_1	n_1
x_2	n_2
.	.
.	.
.	.
x_n	n_n

Từ dãy số phân phối có thể tính ra:

- Tần suất của x_i , ký hiệu f_i , phản ánh tỷ trọng của từng x_i chiếm trong tổng thể.

$$f_i = \frac{n_i}{n}$$

- Tần số tích lũy (hay tần suất tích lũy) là tổng các đơn vị tổng thể (hay các tần suất) tính dồn từ lượng biến thứ nhất đến lượng biến thứ i .

Chương V

CÁC MỨC ĐỘ CỦA HIỆN TƯỢNG KINH TẾ - XÃ HỘI

Các mức độ của hiện tượng kinh tế - xã hội phản ánh quy mô, khối lượng, các quan hệ tỷ lệ so sánh, đặc điểm điển hình về mặt lượng của hiện tượng nghiên cứu bao gồm nhiều đơn vị cùng loại, đánh giá độ biến thiên của tiêu thức, tình hình phân phối các đơn vị tổng thể.

Các mức độ của hiện tượng kinh tế - xã hội được thể hiện bằng các chỉ tiêu chủ yếu sau đây:

- Số tuyệt đối.
- Số tương đối.
- Số bình quân.
- Chỉ tiêu đánh giá độ biến thiên của tiêu thức.

I. SỐ TUYỆT ĐỐI TRONG THỐNG KÊ

1.1. Khái niệm, ý nghĩa và đặc điểm số tuyệt đối

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

Số tuyệt đối có thể được biểu hiện bằng số đơn vị tổng thể của một tổng thể nào đó như: số nhân khẩu, số doanh nghiệp, số công nhân, số học sinh, số mét vuông diện tích gieo trồng.... hoặc là trị số của một tiêu thức như: tổng sản lượng, tổng chi phí sản xuất, tổng số tiền lương....

Số tuyệt đối có ý nghĩa quan trọng đối với công tác quản lý kinh tế - xã hội. Số tuyệt đối chính xác đó là sự thật khách quan có sức thuyết phục không thể phủ nhận được.

Số tuyệt đối là số liệu đầu tiên của hiện tượng, là cơ sở để tính các chỉ tiêu khác như số tương đối, số bình quân.

Quy mô các nguồn tài nguyên của đất nước, các khả năng tiềm tàng trong nền kinh tế quốc dân, các kết quả phát triển kinh tế, văn hóa, xã hội đều được phản ánh bằng số tuyệt đối.

Đặc điểm của số tuyệt đối trong thống kê là gắn liền với hiện tượng kinh tế - xã hội cụ thể, trong điều kiện thời gian và địa điểm cụ thể. Số tuyệt đối trong thống kê phải thông qua các giai đoạn điều tra thu thập, tổng hợp thực tế mà có và phải có đơn vị tính cụ thể.

1.2. Đơn vị tính số tuyệt đối

1.2.1. Đơn vị hiện vật

Là dùng đơn vị đo lường tự nhiên phù hợp với đặc điểm vật lý của hiện tượng, hay dùng đơn vị đo lường tiêu chuẩn để biểu hiện đặc trưng của hiện tượng. Ví dụ: cái, con, m, lít, kg,...

1.2.2. Đơn vị hiện vật quy ước

Được sử dụng khi hiện tượng có các phần tử có cùng giá trị sử dụng (công dụng kinh tế) nhưng khác nhau về quy cách, phẩm chất,... Ví dụ: đơn vị hiện vật quy ước để tính toán và tổng hợp các loại xi măng là loại xi măng P300; đơn vị hiện vật quy ước để tính toán và tổng hợp các loại nhiên liệu (than đá, xăng dầu...) là loại nhiên liệu có năng suất tỏa nhiệt là 7.000 kilô calo/kg...

1.2.3. Đơn vị tiền tệ

Được sử dụng rộng rãi nhất trong thống kê để biểu hiện giá trị sản phẩm. Nó giúp cho việc tổng hợp và so sánh nhiều loại sản phẩm có giá trị sử dụng khác nhau. Tuy nhiên, đơn vị tiền tệ có nhược điểm lớn là chịu ảnh hưởng của giá cả, nên việc tính số tuyệt đối theo đơn vị tiền tệ sẽ không có tính chất so sánh được qua thời gian. Để khắc phục nhược điểm ảnh hưởng của giá cả, thống kê dùng giá so sánh hay giá cố định là giá thực tế của kỳ được chọn làm gốc khi so sánh giá trị khối lượng sản phẩm qua hai kỳ.

1.2.4. Đơn vị thời gian lao động

Dùng để tính lượng lao động hao phí để sản xuất những sản phẩm không thể tổng hợp, so sánh bằng các đơn vị tính toán khác hoặc những sản phẩm phức tạp do nhiều người thực hiện qua nhiều giai đoạn khác nhau. Đơn vị thời gian lao động như: ngày công, giờ công.

1.3. Các loại số tuyệt đối

1.3.1. Số tuyệt đối thời điểm

Số tuyệt đối thời điểm phản ánh quy mô, khối lượng của hiện tượng tại một thời điểm nhất định.

Ví dụ: tổng số dân nước ta có lúc 0 giờ ngày 01/4/1999 là 76.323.173 người.

Đặc điểm của số tuyệt đối thời điểm là không có sự tích lũy về lượng, trị số của chỉ tiêu lớn hay nhỏ không phụ thuộc vào thời gian dài hay ngắn.

1.3.2. Số tuyệt đối thời kỳ

Số tuyệt đối thời kỳ phản ánh quy mô, khối lượng của hiện tượng nghiên cứu trong một độ dài thời gian nhất định.

Ví dụ: sản lượng lương thực quy thóc nước ta năm 1999 là 29 triệu tấn, tổng sản phẩm trong nước (GDP) năm 1997 là 52.198 nghìn tỷ đồng.

Đặc điểm của số tuyệt đối thời kỳ là sự tích lũy về lượng của hiện tượng trong cả thời gian nghiên cứu, nên có thể cộng dồn các số tuyệt đối thời kỳ. Thời kỳ tính toán càng dài, trị số của chỉ tiêu càng lớn.

II. SỐ TƯƠNG ĐỐI TRONG THỐNG KÊ

2.1. Khái niệm, ý nghĩa và đặc điểm số tương đối

Số tương đối trong thống kê là chỉ tiêu biểu hiện quan hệ so sánh giữa hai mức độ của hiện tượng cùng loại nhưng khác nhau về điều kiện thời gian hoặc không gian, hoặc so sánh giữa hai mức độ của hai hiện tượng khác loại nhưng lại có liên quan với nhau, hoặc so sánh bộ phận với tổng thể và giữa các bộ phận trong cùng một tổng thể với nhau.

Số tương đối là một trong những chỉ tiêu phân tích thống kê. Nó phân tích được các đặc điểm của hiện tượng, nghiên cứu các hiện tượng trong mối quan hệ so sánh với nhau. Số tương đối cũng cần thiết trong công tác lập và kiểm tra tình hình thực hiện kế hoạch. Số tương đối còn sử dụng để công bố khi muốn giữ bí mật của số tuyệt đối.

Đặc điểm của số tương đối là có gốc so sánh. Tùy theo mục đích nghiên cứu mà gốc so sánh được chọn khác nhau. Việc chọn gốc so sánh khi tính số tương đối là quan trọng, vì cùng một trị số tuyệt đối như nhau, nhưng sử dụng gốc so sánh khác nhau sẽ có kết quả, kết luận khác nhau.

Hình thức biểu hiện theo số lần, phần trăm (%), phần nghìn (‰) hoặc người/km², đ/người, ...

2.2. Các loại số tương đối

2.2.1. Số tương đối động thái (phát triển)

Số tương đối động thái là kết quả so sánh giữa hai mức độ của hiện tượng cùng loại nhưng khác nhau về thời gian.

- Mức độ được nghiên cứu gọi là mức độ kỳ nghiên cứu, hay còn gọi là mức độ kỳ báo cáo (y_1).

- Mức độ được dùng làm cơ sở so sánh, được gọi là mức độ kỳ gốc (y_0).

Số tương đối động thái phản ánh sự biến động của hiện tượng nghiên cứu theo thời gian, nên nó còn được gọi là tốc độ phát triển hay chỉ số phát triển.

Công thức tính:

$$t = \frac{y_1}{y_0} \quad (\text{nếu tính bằng lần})$$

$$t = \frac{y_1}{y_0} \times 100\% \quad (\text{nếu tính bằng phần trăm})$$

trong đó:

t - số tương đối động thái.

y_1 - mức độ kỳ nghiên cứu.

y_0 - mức độ kỳ gốc.

Ví dụ: doanh số bán hàng của Công ty X năm 2001 là 10 tỷ đồng, năm 2002 là 12 tỷ đồng. Vậy số tương đối động thái là:

$$t = \frac{12}{10} \times 100\% = 120\%$$

Như vậy, doanh số bán hàng của Công ty X năm 2002 so với năm 2001 tăng 20% tương ứng tăng 2 tỷ đồng.

2.2.2. Số tương đối kế hoạch

Số tương đối kế hoạch được dùng để xây dựng kế hoạch và kiểm tra tình hình thực hiện kế hoạch kinh tế - xã hội.

Số tương đối kế hoạch có hai loại:

- Số tương đối nhiệm vụ kế hoạch

Là so sánh giữa mức độ nhiệm vụ kế hoạch (y_k) với mức độ thực tế kỳ gốc (y_0) của một chỉ tiêu. Số tương đối nhiệm vụ kế hoạch được sử dụng trong công tác xây dựng kế hoạch. Công thức tính:

$$t_{NV} = \frac{y_k}{y_0}$$

trong đó:

t_{NV} - số tương đối nhiệm vụ kế hoạch.

y_k - mức độ kế hoạch của kỳ nghiên cứu.

y_0 - mức độ thực tế kỳ gốc.

- Số tương đối thực hiện kế hoạch

Là quan hệ so sánh giữa mức độ thực tế đạt được trong kỳ nghiên cứu (y_1) với mức độ kế hoạch đặt ra cùng kỳ (y_k) của một chỉ tiêu. Số tương đối thực hiện kế hoạch được dùng để kiểm tra tình hình thực hiện nhiệm vụ kế hoạch.

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

trong đó:

t_{TH} - số tương đối thực hiện kế hoạch.

y_1, y_k - như ký hiệu trên.

Giữa các số tương đối động thái và số tương đối kế hoạch của cùng một chỉ tiêu có mối liên hệ như sau:

$$\frac{y_1}{y_0} = \frac{y_k}{y_0} \times \frac{y_1}{y_k}$$

Ví dụ: sản lượng của Công ty Y năm 2001 là 25.000 sản phẩm, kế hoạch dự kiến sản lượng năm 2002 là 30.000 sản phẩm, thực tế năm 2002 công ty đã sản xuất được 33.000 sản phẩm. Như vậy:

Số tương đối động thái:

$$t = \frac{y_1}{y_0} = \frac{33.000}{25.000} = 1,32$$

Số tương đối nhiệm vụ kế hoạch 2002:

$$t_{NV} = \frac{y_k}{y_0} = \frac{30.000}{25.000} = 1,2$$

Số tương đối thực hiện kế hoạch:

$$t_{TH} = \frac{y_1}{y_0} = \frac{33.000}{30.000} = 1,1$$

Như vậy, mối liên hệ giữa ba số tương đối trên là:

$$1,32 = 1,2 \times 1,1$$

hay:

$$\text{Số tương đối động thái} = \text{Số tương đối nhiệm vụ kế hoạch} \times \text{Số tương đối thực hiện kế hoạch}$$

2.2.3. Số tương đối kết cấu

Số tương đối kết cấu xác định tỷ trọng của mỗi bộ phận chiếm trong tổng thể. Công thức tính:

$$d_i = \frac{y_{BPI}}{y_{TT}} \times 100\%$$

trong đó:

d_i - tỷ trọng của bộ phận thứ i.

y_{BPI} - mức độ của bộ phận thứ i.

y_{TT} - mức độ của tổng thể.

Phân tổ thống kê chính xác là cơ sở bảo đảm tính chính xác của số tương đối kết cấu. Muốn có số tương đối kết cấu chính xác, các bộ phận của tổng thể phải được phân biệt rõ ràng, giữa các bộ phận có sự khác nhau về tính chất. Như vậy, việc tính số tương đối kết cấu có liên quan mật thiết với phương pháp phân tổ thống kê.

2.2.4. Số tương đối cường độ

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

Mức độ của hiện tượng cần nghiên cứu được đặt ra ở tử số, còn mức độ của hiện tượng có quan hệ được đặt ở mẫu số.

Ví dụ: mật độ dân số, GDP bình quân đầu người, số bác sĩ trên 1.000 dân,...

Số tương đối cường độ được dùng để phản ánh trình độ phát triển sản xuất, trình độ bảo đảm mức sống vật chất và văn hóa của dân cư trong phạm vi từng vùng, từng khu vực hoặc cả nước. Chỉ tiêu này thường được dùng để so sánh trình độ phát triển sản xuất, đời sống giữa các địa phương, các vùng, các khu vực hoặc giữa các nước với nhau.

2.2.5. Số tương đối so sánh

Số tương đối so sánh là kết quả so sánh giữa các bộ phận trong cùng một tổng thể.

Ví dụ: so sánh số lao động nữ với số lao động nam, số lao động gián tiếp với số lao động trực tiếp trong một doanh nghiệp.

Số tương đối so sánh còn là kết quả so sánh giữa các hiện tượng cùng loại nhưng khác nhau về không gian.

Ví dụ: so sánh giá thành của cùng một loại sản phẩm được sản xuất ở hai doanh nghiệp khác nhau, so sánh giá cả một loại hàng giữa hai địa phương,...

Khi tính số tương đối so sánh có thể tính hai số tương đối có trị số nghịch đảo nhau, nếu ta so sánh A/B và ngược lại B/A .

III. SỐ BÌNH QUÂN TRONG THỐNG KÊ

3.1. Khái niệm, ý nghĩa và đặc điểm số bình quân

Số bình quân trong thống kê là chỉ tiêu biểu hiện mức độ đại biểu theo một tiêu thức số lượng nào đó của một tổng thể bao gồm nhiều đơn vị cùng loại.

Đặc điểm của số bình quân là chỉ dùng một trị số để nói lên đặc điểm điển hình của cả một tổng thể hiện tượng nghiên cứu. Số bình quân san bằng mọi chênh lệch về lượng giữa các đơn vị tổng thể.

Qua số bình quân, có thể so sánh về không gian các hiện tượng không có cùng quy mô như: so sánh giá thành bình quân, năng suất lao động bình quân, tiền lương bình quân giữa các doanh nghiệp. Theo dõi sự biến động của số bình quân theo thời gian, có thể thấy được xu hướng phát triển và tính quy luật của hiện tượng nghiên cứu. Số bình quân còn được dùng để xây dựng và kiểm tra tình hình thực hiện kế hoạch. Số bình quân còn có ý nghĩa quan trọng trong việc vận dụng nhiều phương pháp phân tích thống kê như: phân tích biến động, phân tích mối liên hệ, điều tra chọn mẫu, dự đoán thống kê,...

3.2. Các loại số bình quân

3.2.1. Số bình quân cộng (Số bình quân số học)

Số bình quân cộng được tính bằng cách đem tổng số các lượng biến của tiêu thức chia cho tổng số đơn vị tổng thể. Số bình quân cộng có hai loại:

- *Số bình quân cộng giản đơn*: áp dụng khi mỗi lượng biến chỉ có một đơn vị tổng thể tương ứng.

Công thức tính:

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n} = \frac{\sum x_i}{n}$$

trong đó:

$\bar{x}$ - số bình quân.

x_i - các trị số lượng biến.

n - tổng số đơn vị tổng thể.

Ví dụ: tính tiền lương bình quân 1 công nhân của một tổ sản xuất gồm 4 công nhân, với tiền lương 1 công nhân trong tháng lần lượt là: 550.000đ, 650.000đ, 750.000đ, 850.000đ.

$$\bar{x} = \frac{550.000 + 650.000 + 750.000 + 850.000}{4} = 700.000 \text{ đồng}$$

- *Số bình quân cộng gia quyền*: trường hợp ứng với các lượng biến x_i có số đơn vị tổng thể n_i (tức là tần số) khác nhau, thì tổng lượng của tiêu thức bằng tổng số của lượng biến (x_i) nhân với số đơn vị tổng thể có lượng biến tương ứng ($\sum x_i n_i$), và tổng số đơn vị tổng thể là $\sum n_i$.

Công thức tính:

$$\bar{x} = \frac{\sum x_i n_i}{\sum n_i}$$

trong đó:

$\bar{x}$ - số bình quân.

x_i - các trị số lượng biến.

n_i - các tần số còn được gọi là quyền số.

Ví dụ: tính tiền lương bình quân 1 công nhân căn cứ vào tài liệu về tiền lương trong tháng của công nhân tại một phân xưởng sản xuất gồm các mức lương sau:

Bảng 5-1

Tiền lương tháng 1 công nhân (đồng)	Số công nhân (người)
650.000	15
750.000	20
850.000	15
Cộng	50

$$\bar{x} = \frac{650.000 \times 15 + 750.000 \times 20 + 850.000 \times 15}{15 + 20 + 15} = 750.000 \text{ đồng}$$

Trường hợp tài liệu phân tổ có khoảng cách tổ, thì x_i là trị số giữa của tổ tính theo công thức:

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

trong đó:

$x_{\min}$, $x_{\max}$ - giới hạn dưới (lượng biến nhỏ nhất của tổ) và giới hạn trên (lượng biến lớn nhất của tổ) của từng khoảng cách tổ.

Trường hợp phân tổ có tổ mở (ở tổ đầu tiên và tổ cuối cùng), giả định rằng khoảng cách tổ của tổ mở bằng khoảng cách tổ của tổ đứng kế ngay bên nó, ta sẽ tính trị số giả thiết của giới hạn dưới (hoặc giới hạn trên) rồi tìm trị số giữa.

Ví dụ: tính tiền lương bình quân 1 công nhân căn cứ vào tài liệu về tiền lương trong tháng của công nhân tại một doanh nghiệp sản xuất gồm các mức lương sau:

Bảng 5-2

Tiền lương tháng 1 công nhân (đồng)	Trị số giữa (x_i)	Số công nhân (người) (n_i)
Dưới 700.000	650.000	25
700.000 - 800.000	750.000	35
800.000 - 900.000	850.000	30
900.000 - 1.000.000	950.000	5
Trên 1.000.000	1.050.000	5
Cộng	Σ	100

$$\bar{x} = \frac{650.000 \times 25 + 750.000 \times 35 + 850.000 \times 30 + 950.000 \times 5 + 1.050.000 \times 5}{25 + 35 + 30 + 5 + 5} = 780.000 \text{ đ}$$

Trong công thức tính số bình quân cộng gia quyền, có thể rút ra một công thức khác tính số bình quân như sau:

$$\bar{x} = \Sigma x_i d_i$$

trong đó:

d_i - số tương đối kết cấu xác định tỷ trọng của từng lượng biến hay từng tổ trong tổng thể, tính bằng lần.

Ví dụ: có tài liệu về tiền lương của công nhân tại một doanh nghiệp sản xuất gồm 3 phân xưởng sản xuất như sau:

Bảng 5-3

Phân xưởng	Tiền lương tháng 1 công nhân (đồng)	Tỷ trọng công nhân (%)
I	650.000	30
II	750.000	40
III	850.000	30
Cộng	x	100

Vậy tiền lương bình quân 1 công nhân trong toàn doanh nghiệp là:

$$\bar{x} = 650.000 \times 0,3 + 750.000 \times 0,4 + 850.000 \times 0,3 = 750.000 \text{ đồng}$$

3.2.2. Số bình quân điều hòa

Số bình quân điều hòa là số bình quân được tính từ các đại lượng nghịch đảo của các lượng biến. Có hai loại số bình quân điều hòa:

- Số bình quân điều hòa gia quyền:

Được áp dụng trong trường hợp không có tài liệu về số đơn vị tổng thể (n_i), mà chỉ có tài liệu tổng lượng của từng nhóm lượng biến ($M_i = x_i n_i$).

$$\text{Do: } M_i = x_i n_i \rightarrow n_i = \frac{M_i}{x_i}$$

Ta có công thức tính số bình quân điều hòa gia quyền:

$$\bar{x} = \frac{\sum M_i}{\sum \frac{M_i}{x_i}}$$

Ví dụ: có tài liệu về tiền lương của công nhân tại một doanh nghiệp sản xuất gồm 3 phân xưởng sản xuất như sau:

Bảng 5-4

Phân xưởng	Tiền lương tháng 1 công nhân (đồng)	Tổng số tiền lương (đồng)
I	650.000	19.500.000
II	750.000	30.000.000
III	850.000	25.500.000
Cộng	x	75.000.000

Vậy tiền lương bình quân 1 công nhân trong toàn doanh nghiệp là:

$$\bar{x} = \frac{\sum M_i}{\sum \frac{M_i}{x_i}} = \frac{19.500.000 + 30.000.000 + 25.500.000}{\frac{19.500.000}{650.000} + \frac{30.000.000}{750.000} + \frac{25.500.000}{850.000}} = 750.000 \text{ đồng}$$

- *Số bình quân điều hòa giản đơn:*

Trong công thức số bình quân điều hòa giản đơn, nếu các M_i bằng nhau ($M_1 = M_2 = \dots = M_n = M$), ta có:

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

Ta có công thức tính số bình quân điều hòa giản đơn:

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

Ví dụ: một nhóm 3 công nhân cùng sản xuất một loại sản phẩm trong khoảng thời gian như nhau. Người thứ nhất hoàn thành 1 sản phẩm mất 24 phút, người thứ hai mất 30 phút, người thứ ba mất 40 phút. Vậy thời gian bình quân để hoàn thành 1 sản phẩm của 1 công nhân cả nhóm trên là:

$$\bar{x} = \frac{3}{\frac{1}{24} + \frac{1}{30} + \frac{1}{40}} = 30 \text{ phút}$$

3.2.3. Số bình quân nhân

Số bình quân nhân được sử dụng trong trường hợp các lượng biến có mối quan hệ tích số với nhau và được dùng để tính các tốc độ phát triển bình quân. Số bình quân nhân gồm có hai loại:

- Số bình quân nhân giản đơn:

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

trong đó:

$\bar{x}$ - tốc độ phát triển bình quân.

x_i - tốc độ phát triển của các năm.

n - số tốc độ phát triển.

Ví dụ: có tài liệu về tốc độ phát triển doanh số bán hàng của một công ty thương mại từ năm 1999 đến năm 2002 (tính bằng lần so với năm trước).

Bảng 5-5

Năm	1999	2000	2001	2002
Tốc độ phát triển	1,10	1,24	1,22	1,19

Vậy tốc độ phát triển bình quân hàng năm qua 4 năm về doanh số của công ty trên là:

$$\bar{x} = \sqrt[4]{\prod x_i} = \sqrt[4]{1,10 \times 1,24 \times 1,22 \times 1,19} = 1,18 \quad \text{hay } 118\%$$

- *Số bình quân nhân gia quyền:*

Áp dụng trong trường hợp mỗi tốc độ phát triển x_i có tần số (f_i) xuất hiện khác nhau. Công thức tính số bình quân nhân gia quyền:

$$\bar{x} = \frac{\sum f_i \sqrt[4]{\prod x_i^{f_i}}}{\sum f_i}$$

trong đó:

$\bar{x}$ - tốc độ phát triển bình quân.

f_i - số tốc độ phát triển liên hoàn x_i

$$\sum f_i = n$$

Ví dụ: trong thời gian 10 năm ($\sum f_i = 10$), tốc độ phát triển đàn gia súc của nông trường M như sau: 5 năm đầu mỗi năm phát triển 110%, trong 3 năm tiếp theo mỗi năm phát triển 115%, trong 2 năm cuối cùng mỗi năm phát triển 125%.

Vậy tốc độ phát triển bình quân 1 năm trong 10 năm trên là:

$$\bar{x} = \frac{\sum f_i \sqrt[4]{\prod x_i^{f_i}}}{\sum f_i} = \frac{10 \sqrt[4]{(1,1)^5 \times (1,15)^3 \times (1,25)^2}}{10} = 1,14 \quad \text{hay } 114\%$$

3.3. Mod (ký hiệu M_0)

Mod là lượng biến có tần số lớn nhất trong một dãy số phân phối thống kê.

Ví dụ: có tài liệu về phân tổ các gia đình tại một địa phương theo số lượng con trong gia đình.

Bảng 5-6

Số con trong gia đình	Số gia đình
0	10
1	30
2	75
3	45
4	20
5	15
6	5
Cộng	200

Trong ví dụ trên, Mod là gia đình có hai con, vì ứng với giá trị của lượng biến này có số gia đình nhiều nhất (có tần số lớn nhất).

Trị số của Mod có đặc điểm là không được tính từ tất cả các lượng biến trong dãy số, mà được xác định căn cứ vào sự phân phối các đơn vị tổng thể vào các lượng biến.

Phương pháp xác định Mod như sau:

- Trong trường hợp một dãy phân phối rời rạc không khoảng cách tổ, chỉ cần tìm trong phân phối lượng biến x_i nào có tần số lớn nhất, thì lượng biến đó chính là số Mod.

- Trong trường hợp một dãy phân phối lượng biến có khoảng cách tổ, xác định Mod gồm 2 bước:

Bước 1: Xác định tổ chứa Mod (tổ có tần số lớn nhất).

Bước 2: Tính Mod theo công thức:

$$M_0 = x_0 + d \frac{f_m - f_{m-1}}{(f_m - f_{m-1}) + (f_m - f_{m+1})}$$

trong đó:

M_0 - Mod

x_0 - giới hạn dưới của tổ có Mod

d - khoảng cách tổ của tổ có Mod

f_m - tần số của tổ có Mod

f_{m-1} - tần số của tổ đứng trước tổ có Mod

f_{m+1} - tần số của tổ đứng sau tổ có Mod

Ví dụ: có số liệu về năng suất lao động tại một doanh nghiệp sản xuất như sau:

Bảng 5-7

Năng suất lao động (kg)	Số công nhân (người)
150-155	4
155-160	10
160-165	61
165-170	100
170-175	130
175-180	114
180-185	62
185-190	11
190-195	8
Cộng	500

Theo số liệu trong bảng trên ta thấy Mod sẽ rơi vào tổ 170 - 175 vì tổ này có tần số lớn nhất (130 người). Khoảng cách tổ của tổ có Mod là $d = 5$, giới hạn dưới của tổ có

Mod là 170, tần số của tổ trước tổ có Mod là 100, tổ đứng sau Mod là 114. Dùng công thức trên ta tính được số Mod là:

$$M_0 = 170 + 5 \frac{130 - 100}{(130 - 100) + (130 - 114)} = 173,26\text{kg}$$

Mod là chỉ tiêu được dùng nhiều trong nghiên cứu kinh tế. Nó là lượng biến thường gặp nhiều nhất trong một tổng thể. Do đó tác dụng của nó rất lớn như: xác định cỡ, loại hàng hóa được người tiêu dùng ưa chuộng nhất,...Trong quy hoạch, người ta dùng Mod để xác định vị trí người, xe qua lại nhiều để bố trí kho hàng,...

3.4. Số trung vị (ký hiệu M_e)

Số trung vị là lượng biến nằm ở vị trí chính giữa trong dãy số phân phối. Như vậy, số trung vị phân chia dãy số lượng biến thành hai phần, mỗi phần có số đơn vị tổng thể bằng nhau.

Đặc điểm của số trung vị là không tính từ tất cả các lượng biến mà được tính từ vị trí của các lượng biến. Trung vị có tính chất là tổng các độ lệch tuyệt đối giữa các lượng biến với số trung vị là một trị số nhỏ nhất:

$$\sum |x_i - M_e| = \min$$

Đặc điểm trên làm cho số trung vị có tác dụng như: xác định vị trí của các công trình công cộng (bến xe, trường học, bệnh viện...) phục vụ thuận lợi cho các khu dân cư.

Phương pháp xác định số trung vị như sau:

- Nếu dãy số phân phối rời rạc, không có khoảng cách tổ và có số đơn vị tổng thể lẻ ($n = 2p + 1$), thì số trung vị là lượng biến ở vị trí $p + 1$.

Ví dụ: ta có 97 lượng biến sắp xếp theo thứ tự từ nhỏ đến lớn, trung vị sẽ là lượng biến thứ 49 ($p = 48$).

Nếu n chẵn, ta có khoảng trung vị là $(p, p + 1)$, người ta quy ước lấy giá trị trung bình của khoảng đó làm trung vị

$$M_e = \frac{x_p + x_{p+1}}{2}$$

Trong trường hợp này, nếu các x_i chỉ lấy số nguyên thì không thể lấy giá trị trung bình, ta chỉ có khoảng trung vị.

- Nếu dãy phân phối lượng biến có khoảng cách tổ, xác định trung vị gồm 2 bước:

Bước 1: Xác định tổ chứa trung vị là tổ đầu tiên có tần số tích lũy $S \geq \frac{\sum f}{2}$

Bước 2: Tính số trung vị theo công thức:

$$M_c = x_0 + d \frac{\frac{\sum f}{2} - S_{m-1}}{f_m}$$

trong đó:

M_c - số trung vị.

x_0 - giới hạn dưới của tổ có trung vị.

d - khoảng cách tổ của tổ chứa số trung vị.

$\sum f$ - tổng các tần số.

S_{m-1} - tần số tích lũy của tổ đứng trước tổ có số trung vị.

f_m - tần số của tổ có trung vị.

Ví dụ: theo số liệu của bảng 5-7 ta lập bảng tính toán sau:

Bảng 5-8

Năng suất lao động (kg)	Số công nhân (người)	Tần số tích lũy (S_i)
150-155	4	4
155-160	10	14
160-165	61	75
165-170	100	175
170-175	130	305
175-180	114	419
180-185	62	481
185-190	11	492
190-195	8	500
Cộng	500	-

Tổ có chứa số trung vị là tổ có mức năng suất lao động 170 - 175 ($S_i = 305 > 500/2$).

Theo công thức xác định số trung vị ta có:

$$M_c = 170 + 5 \frac{\frac{500}{2} - 175}{130} = 172,8 \text{ kg}$$

3.5. Điều kiện tính số bình quân

- Trước hết, số bình quân phải được tính ra từ một tổng thể đồng chất.

Tổng thể đồng chất là tổng thể mà trong đó các đơn vị của tổng thể có cùng chung một tính chất, thuộc cùng một loại hình kinh tế - xã hội theo một tiêu thức nào đó.

- Muốn tính được số bình quân phải căn cứ vào phương trình kinh tế để xác định sự đúng đắn của các yếu tố tham gia vào việc tính số bình quân.

Ví dụ: muốn tính giá thành bình quân một đơn vị sản phẩm dựa vào phương trình kinh tế:

$$\text{Giá thành toàn bộ sản phẩm} = \frac{\text{Giá thành}}{\text{đơn vị sản phẩm}} \times \frac{\text{Số lượng}}{\text{sản phẩm sản xuất}}$$

Muốn tính giá thành bình quân cho một đơn vị sản phẩm phải dựa vào phương trình kinh tế:

$$\frac{\text{Sản lượng}}{\text{sản phẩm}} = \frac{\text{Năng suất lao động}}{\text{một công nhân}} \times \frac{\text{Số lượng}}{\text{công nhân}}$$

Các phương trình kinh tế trên đều có dạng tổng quát: $M_i = x_i \times n_i$

IV. CÁC CHỈ TIÊU ĐÁNH GIÁ ĐỘ BIẾN THIÊN CỦA TIÊU THỨC

Trong thống kê, chênh lệch giữa các lượng biến với nhau hoặc giữa các lượng biến với số bình quân gọi là độ biến thiên của tiêu thức.

Độ biến thiên của tiêu thức giúp ta đánh giá tính chất đại biểu của số bình quân. Trị số của các chỉ tiêu đánh giá độ biến thiên của tiêu thức càng lớn, thì mức độ đại biểu của số bình quân càng thấp và ngược lại.

Các chỉ tiêu đánh giá độ biến thiên của tiêu thức bao gồm các chỉ tiêu sau đây:

4.1. Khoảng biến thiên của tiêu thức

Khoảng biến thiên là hiệu số giữa lượng biến lớn nhất và lượng biến nhỏ nhất của tiêu thức nghiên cứu.

Công thức tính: $R = x_{\max} - x_{\min}$

trong đó: R - khoảng biến thiên của tiêu thức.

$x_{\max} - x_{\min}$ - lượng biến lớn nhất và nhỏ nhất của tiêu thức nghiên cứu.

Khoảng biến thiên R càng lớn thì tính đại biểu của số bình quân càng thấp và ngược lại.

4.2. Độ lệch tuyệt đối bình quân

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

Công thức tính: $\bar{d} = \frac{\sum |x_i - \bar{x}|}{n}$ (trường hợp giản đơn)

$$\bar{d} = \frac{\sum |x_i - \bar{x}| f_i}{\sum f_i} \quad (\text{trường hợp có quyền số})$$

trong đó: $\bar{d}$ - độ chênh lệch tuyệt đối bình quân.

$\bar{x}$ - số bình quân của các lượng biến x_i .

f_i - tần số.

4.3. Phương sai

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

Công thức tính:

$$\sigma_x^2 = \frac{\sum (x_i - \bar{x})^2}{n} \quad (\text{trường hợp giản đơn})$$

$$\sigma_x^2 = \frac{\sum (x_i - \bar{x})^2 \cdot f_i}{\sum f_i} \quad (\text{trường hợp có quyền số})$$

trong đó: σ_x^2 - phương sai.

4.4. Độ lệch tiêu chuẩn

Độ lệch tiêu chuẩn là số bình quân toàn phương của các độ lệch giữa các lượng biến với số bình quân của chúng, tức là căn bậc hai của phương sai.

Công thức tính:

$$\sigma_x = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n}} \quad (\text{trường hợp giản đơn})$$

$$\sigma_x = \sqrt{\frac{\sum (x_i - \bar{x})^2 \cdot f_i}{\sum f_i}} \quad (\text{trường hợp có quyền số})$$

trong đó: σ_x - độ lệch tiêu chuẩn.

Phương sai và độ lệch tiêu chuẩn tính ra càng nhỏ thì chênh lệch giữa các đơn vị càng nhỏ, tính đại biểu của số bình quân càng cao và ngược lại.

4.5. Hệ số biến thiên

Phương sai và độ lệch tiêu chuẩn có ưu điểm là xét đến toàn bộ các độ lệch giữa các lượng biến với số bình quân của các lượng biến đó. Tuy nhiên do phương sai và độ lệch tiêu chuẩn biểu hiện bằng số tuyệt đối nên khi so sánh các tổng thể có quy mô khác nhau, ta không thể so sánh trực tiếp các phương sai. Vì vậy cần tính ra một chỉ tiêu tương đối biểu hiện quan hệ so sánh giữa độ lệch tiêu chuẩn với số bình quân.

Công thức tính:

$$V = \frac{\sigma_x}{\bar{x}} \cdot 100\%$$

trong đó: V - hệ số biến thiên.

Hệ số biến thiên càng nhỏ thì độ chênh lệch giữa các đơn vị càng nhỏ và tính đại biểu của số bình quân càng cao và ngược lại.

Ví dụ: có tài liệu về mức năng suất lao động của công nhân ở hai tổ sản xuất (tổ 1 và tổ 2 như sau):

Tổ 1: 20, 25, 30, 35, 40 (sản phẩm/ người).

Tổ 2: 28, 29, 30, 31, 32 (sản phẩm/ người).

Căn cứ tài liệu trên, ta lập bảng tính toán như sau:

Bảng 5-9

		Tổ 1		Tổ 2			
X_i	$ x_i - \bar{x} $	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$	X_i	$ x_i - \bar{x} $	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$
20	10	- 10	100	28	2	- 2	4
25	5	5	25	29	1	1	1
30	0	0	0	30	0	0	0
35	5	+5	25	31	1	+1	1
40	10	+10	100	32	2	+2	4
150	30	0	250	150	6	0	10

* Năng suất lao động bình quân:

$$\bar{x}_1 = \frac{150}{5} = 30 \text{ sản phẩm/người}$$

$$\bar{x}_2 = \frac{150}{5} = 30 \text{ sản phẩm/người}$$

* Khoảng biến thiên:

$$R_1 = 40 - 20 = 20 \text{ sản phẩm/người}$$

$$R_2 = 32 - 28 = 4 \text{ sản phẩm/người}$$

* Độ lệch tuyệt đối bình quân:

$$d_1 = \frac{30}{5} = 6 \text{ sản phẩm/người}$$

$$d_2 = \frac{6}{5} = 1,2 \text{ sản phẩm/người}$$

* Phương sai:

$$\sigma_{x_1}^2 = \frac{250}{5} = 50 \text{ sản phẩm/người}$$

$$\sigma_{x_2}^2 = \frac{10}{5} = 2 \text{ sản phẩm/người}$$

* Độ lệch tiêu chuẩn:

$$\sigma_{x_1} = \sqrt{50} = 7,07 \text{ sản phẩm/ người}$$

$$\sigma_{x_2} = \sqrt{2} = 1,414 \text{ sản phẩm/ người}$$

*Hệ số biến thiên:

$$V_1 = \frac{7,07}{30} \times 100 = 23,57\%$$

$$V_2 = \frac{1,414}{30} \times 100 = 4,71\%$$

Kết quả trên cho thấy các chỉ tiêu đánh giá độ biến thiên của tiêu thức ở tổ 1 lớn hơn tổ 2 nên tính đại biểu của số bình quân (năng suất lao động bình quân) ở tổ 1 thấp hơn tổ 2.

4.6. Một số vấn đề tính toán và vận dụng phương sai

Nếu số lượng biến nhiều và lẻ thì việc tính toán phương sai theo công thức nêu ở trên khá phức tạp. Vì vậy cần biến đổi công thức trên thành dạng khác để tính toán dễ dàng hơn chỉ cần dựa vào bảng số liệu đã tính sẵn.

Ta có:
$$\sigma_x^2 = \frac{\sum (x_i - \bar{x})^2 \cdot f_i}{\sum f_i}$$

Khai triển tử số của công thức trên ta có:

$$\sum (x_i - \bar{x})^2 \cdot f_i = \sum x_i^2 f_i + (\bar{x})^2 \sum f_i - 2 \cdot \bar{x} \cdot \sum x_i \cdot f_i$$

Chia cả 2 vế cho $\sum f_i$ là số đơn vị tổng thể:

$$\frac{\sum (x_i - \bar{x})^2 \cdot f_i}{\sum f_i} = \frac{\sum x_i^2 f_i}{\sum f_i} + (\bar{x})^2 - 2 \bar{x} \cdot \frac{\sum x_i \cdot f_i}{\sum f_i}$$

hay:
$$\sigma_x^2 = \overline{x^2} + (\bar{x})^2 - 2 \bar{x} \cdot \bar{x}$$

Vậy:
$$\sigma_x^2 = \overline{x^2} - (\bar{x})^2$$

trong đó:

$\overline{x^2}$ - số bình quân của bình phương các lượng biến

$(\bar{x})^2$ - bình phương của số bình quân của các lượng biến.

Công thức trên là công thức tính phương sai được sử dụng rất phổ biến.

Chương VI

HỒI QUY VÀ TƯƠNG QUAN

I. PHƯƠNG PHÁP HỒI QUY VÀ TƯƠNG QUAN

1.1. Liên hệ hàm số và liên hệ tương quan

Giữa các hiện tượng kinh tế - xã hội hoặc giữa các tiêu thức thường có mối liên hệ và phụ thuộc lẫn nhau theo những quy luật nhất định. Nghiên cứu mối liên hệ giữa các hiện tượng, các tiêu thức, xác định tính chất và trình độ chặt chẽ của mối liên hệ là nhiệm vụ cụ thể của thống kê. Có hai loại liên hệ phụ thuộc giữa các hiện tượng kinh tế - xã hội là: liên hệ hàm số và liên hệ tương quan.

- Liên hệ hàm số là loại liên hệ hết sức chặt chẽ giữa các đại lượng biến thiên nhưng xác định. Khi đại lượng này biến thiên thì làm cho đại lượng kia cũng biến thiên theo tỷ lệ tương ứng chặt chẽ.

Ví dụ: liên hệ giữa diện tích và bán kính hình tròn ($S = \pi.R^2$) là liên hệ hàm số.

Liên hệ hàm số ít có trong mối liên hệ giữa các hiện tượng kinh tế - xã hội.

- Liên hệ tương quan là liên hệ không hoàn toàn chặt chẽ như liên hệ hàm số. Trong liên hệ tương quan, cứ mỗi trị số của tiêu thức này có thể ứng với phân phối các trị số khác. Nói cách khác, có liên hệ tương quan khi giữa các tiêu thức có ràng buộc nhau không có sự tương ứng chặt chẽ. Khi trị số của tiêu thức này tăng (hay giảm) thì trị số của tiêu thức kia (tiêu thức có liên hệ) cũng tăng (hay giảm) nhưng không theo tỷ lệ chặt chẽ, vì ngoài tiêu thức nguyên nhân này còn có nguyên nhân khác tác động.

Ví dụ: giữa mức trang bị TSCĐ cho lao động và năng suất lao động, giữa quy mô sản xuất kinh doanh với hiệu quả đồng vốn,... là liên hệ tương quan.

Xét theo chiều hướng của liên hệ tương quan, chia làm hai loại: liên hệ tương quan thuận và liên hệ tương quan nghịch. Liên hệ tương quan thuận khi trị số của tiêu thức này tăng lên hay giảm đi thì trị số của tiêu thức kia cũng tăng lên hay giảm đi; liên hệ nghịch khi trị số của tiêu thức này tăng thì trị số của tiêu thức kia giảm và ngược lại.

Ta còn có mối liên hệ tương quan theo đường thẳng (liên hệ tương quan tuyến tính) và mối liên hệ tương quan theo đường cong (liên hệ tương quan phi tuyến tính).

1.2. Phương pháp hồi quy và tương quan (phương pháp phân tích tương quan)

Hồi quy và tương quan là các phương pháp toán học được áp dụng trong thống kê để phân tích mối liên hệ giữa các hiện tượng kinh tế xã hội.

Công việc chủ yếu của phương pháp phân tích tương quan gồm các bước:

- 1- Phát hiện ra mối liên hệ tương quan giữa các hiện tượng, các tiêu thức.
- 2- Xác định quy luật liên hệ dưới dạng hàm số. Dùng hàm số này để đặc trưng cho mối liên hệ tương quan nghiên cứu.
- 3- Xác định mức độ chặt chẽ của mối liên hệ tương quan.
- 4- Đánh giá mối liên hệ tương quan vừa nghiên cứu.

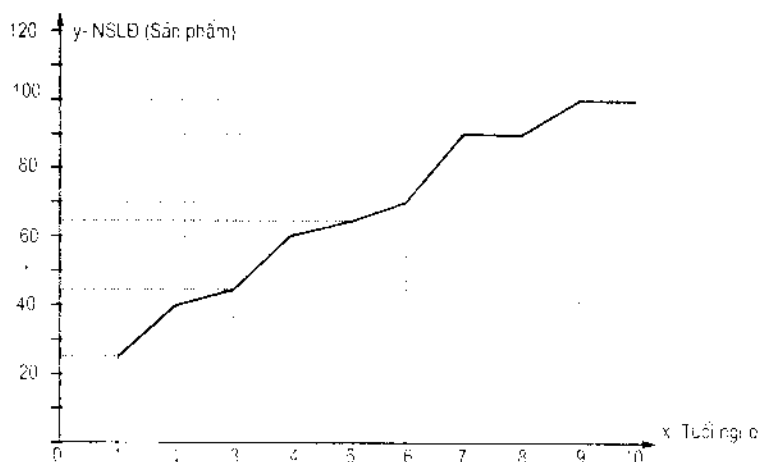
II. LIÊN HỆ TƯƠNG QUAN TUYẾN TÍNH GIỮA HAI TIÊU THỨC SỐ LƯỢNG

2.1. Phương trình hồi quy

Để thuận tiện cho việc xem xét phương pháp này chúng ta hãy xét mối quan hệ giữa tuổi nghề và năng suất lao động của công nhân theo ví dụ của bảng sau:

Bảng 6-1

Tên công nhân được chọn để điều tra	Tuổi nghề (năm) x	Năng suất lao động (sản phẩm) y	x^2	$x.y$
A	1	25	1	25
B	2	40	4	80
C	3	45	9	135
D	4	60	16	240
E	5	65	25	325
G	6	70	36	420
H	7	90	49	630
I	8	90	64	720
K	9	100	81	900
L	10	100	100	1.000
Cộng	55	685	385	4.475



Hình 6.1: Đồ thị thực nghiệm về mối liên hệ giữa năng suất lao động và tuổi nghề của công nhân.

Căn cứ vào tài liệu bảng 6-1 và đồ thị thực nghiệm hình 6-1 ta thấy tuổi nghề của công nhân càng cao thì năng suất lao động càng cao, chứng tỏ giữa hai tiêu thức này có mối liên hệ tương quan. Trên đồ thị ta thấy những điểm tương ứng x-y rải theo một chiều hướng gần thẳng. Như vậy giữa năng suất lao động và tuổi nghề công nhân có mối quan hệ tương quan tuyến tính. Tuy nhiên đường gấp khúc thực nghiệm y_x mới cho biết tính chất của mối liên hệ như đã nói nhưng chưa thật rõ ràng cụ thể. Vì vậy để mô tả khái quát và chính xác nhất quy luật tương quan tuyến tính giữa hai tiêu thức nguyên nhân x (tuổi nghề) và tiêu thức kết quả y (năng suất lao động) người ta dùng đường hồi quy lý thuyết có dạng:

$$\bar{y}_x = a + bx$$

trong đó:

$\bar{y}_x$ - trị số điều chỉnh của tiêu thức kết quả y theo mối liên hệ phụ thuộc với tiêu thức nguyên nhân x.

x - trị số của tiêu thức nguyên nhân.

a, b - các tham số.

Yêu cầu đặt ra cho đường hồi quy lý thuyết là phải nằm gần nhất với đường gấp khúc thực nghiệm y_x , sao cho tổng các mức chênh lệch giữa đường hồi quy lý thuyết và đường gấp khúc thực nghiệm phải có giá trị nhỏ nhất, tức là:

$$S = \sum (\bar{y}_x - y)^2 = \min$$

Bằng cách lấy đạo hàm riêng của S theo a, b và cho triệt tiêu ta được hệ phương trình:

$$\begin{cases} \sum y = na + b\sum x & (1) \end{cases}$$

$$\begin{cases} \sum xy = a\sum x + b\sum x^2 & (2) \end{cases}$$

Giải hệ ta được:

$$a = \bar{y} - b\bar{x}$$

$$b = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\sigma_x^2}$$

Với: $\sigma_x^2 = \bar{x}^2 - (\bar{x})^2$

Từ bảng 6-1 ta tính được:

$$\bar{x} = \frac{55}{10} = 5,5$$

$$(\bar{x})^2 = (5,5)^2 = 30,25$$

$$\bar{y} = \frac{685}{10} = 68,5$$

$$\overline{xy} = \frac{4.475}{10} = 447,5$$

$$\overline{x^2} = \frac{385}{10} = 38,5$$

$$\sigma_x^2 = 38,5 - 30,25 = 8,25$$

$$b = \frac{447,5 - (5,5 \times 68,5)}{8,25} = 8,576$$

$$a = 68,5 - (8,576 \times 5,5) = 21,332$$

Vậy phương trình hồi quy lý thuyết là:

$$y_x = 21,332 + 8,576x$$

Kết quả trên đây có nghĩa là:

- Tham số $b = 8,576$ cho biết cứ thêm 1 tuổi nghề thì năng suất lao động tăng 8,576 sản phẩm.
- Tham số $a = 21,332$ cho biết mức năng suất tối thiểu đối với công nhân mới bước vào nghề.

2.2. Hệ số tương quan

Hệ số tương quan là một chỉ tiêu dùng để đánh giá trình độ chặt chẽ của mối liên hệ tương quan tuyến tính.

Hệ số tương quan được xác định trên cơ sở so sánh cùng một biểu hiện của mối liên hệ trong trường hợp có liên hệ tương quan và liên hệ hàm số.

Hệ số tương quan được tính theo công thức:

$$r = \frac{\Sigma(x - \bar{x})(y - \bar{y})}{\sqrt{\Sigma(x - \bar{x})^2 \cdot \Sigma(y - \bar{y})^2}}$$

Hệ số tương quan có các tính chất sau:

- 1- Hệ số tương quan có trị số: $-1 \leq r \leq +1$
- 2- Hệ số tương quan mang dấu trừ (-) ta có tương quan nghịch, mang dấu cộng (+) ta có tương quan thuận.
- 3- Trị số của hệ số tương quan càng gần ± 1 thì mối liên hệ tương quan càng chặt chẽ, càng gần 0 thì ngược lại.

Ví dụ: tính hệ số tương quan tuyến tính để đánh giá mức độ chặt chẽ của mối liên hệ tương quan giữa tuổi nghề và năng suất lao động của công nhân. Từ số liệu bảng 6-1, ta lập bảng tính toán như sau:

Bảng 6-2

Tuổi nghề x_i	Năng suất lao động y_i	$(x - \bar{x})$	$(y - \bar{y})$	$(x - \bar{x}).(y - \bar{y})$	$(x - \bar{x})^2$	$(y - \bar{y})^2$
1	25	- 4,5	- 43,5	+ 195,75	20,25	1.892,25
2	40	- 3,5	- 28,5	+ 99,75	12,25	812,25
3	45	- 2,5	- 23,5	+ 58,75	6,25	552,25
4	60	- 1,5	- 8,5	+ 12,75	2,25	72,25
5	65	- 0,5	- 3,5	+ 1,75	0,25	12,25
6	70	+ 0,5	+ 1,5	+ 0,75	0,25	2,25
7	90	+ 1,5	+ 21,5	+ 32,25	2,25	462,25
8	90	+ 2,5	+ 21,5	+ 53,75	6,25	462,25
9	100	+ 3,5	+ 31,5	+ 110,25	12,25	992,25
10	100	+ 4,5	+ 31,5	+ 141,75	20,25	992,25
55	685	0	0	+ 707,50	82,50	6.252,50

$$r = \frac{+707,5}{\sqrt{82,5 \times 6.252,5}} = +0,985$$

Kết quả trên cho thấy giữa tuổi nghề và năng suất lao động có mối liên hệ tương quan khá chặt chẽ.

III. LIÊN HỆ TƯƠNG QUAN PHI TUYẾN GIỮA HAI TIÊU THỨC SỐ LƯỢNG

3.1. Phương trình hồi quy

Các hiện tượng kinh tế - xã hội chẳng những có mối liên hệ tương quan tuyến tính (theo dạng đường thẳng) mà còn có thể có mối liên hệ tương quan phi tuyến tính (theo các dạng đường cong).

Về mặt lý thuyết, ta có thể sử dụng nhiều dạng phương trình hồi quy phi tuyến như dạng hypebôn, parabol, lôgarit, hàm số mũ,... để biểu hiện những mối liên hệ tương quan phi tuyến tính.

- Phương trình parabol bậc 2:

$$\bar{y}_x = a + bx + cx^2$$

Với hệ phương trình chuẩn (rút ra bằng phương pháp bình phương bé nhất):

$$\begin{cases} \Sigma y = na + b\Sigma x + c\Sigma x^2 & (1) \\ \Sigma xy = a\Sigma x + b\Sigma x^2 + c\Sigma x^3 & (2) \\ \Sigma x^2y = a\Sigma x^2 + b\Sigma x^3 + c\Sigma x^4 & (3) \end{cases}$$

Giải hệ ta tìm được các tham số a, b, c

- Phương trình hypebôn:

$$\bar{y}_x = a + \frac{b}{x}$$

Các tham số được tính từ hệ phương trình:

$$\begin{cases} \Sigma y = na + b \Sigma \frac{1}{x} & (1) \\ \Sigma \frac{1}{x} y = a \Sigma \frac{1}{x} + b \Sigma \left(\frac{1}{x} \right)^2 & (2) \end{cases}$$

Bằng cách làm tương tự, ta có thể xây dựng được các hệ phương trình chuẩn cho các dạng đường cong khác.

Ví dụ: điều tra 10 doanh nghiệp cùng sản xuất một loại sản phẩm, thu thập được tài liệu về hai tiêu thức là khối lượng sản phẩm sản xuất và giá thành một đơn vị sản phẩm theo bảng sau:

Bảng 6-3

Tên doanh nghiệp	Khối lượng sản phẩm sản xuất (1.000 cái) x_i	Giá thành một đơn vị sản phẩm (1.000 đ) y_i
A	0,5	45,0
B	1,0	25,0
C	2,5	20,0
D	4,0	15,0
E	4,5	15,0
G	5,0	12,5
H	5,0	10,0
I	7,5	7,5
K	8,0	8,0
L	10,0	7,0

Quan sát tài liệu của bảng 6-3, ta thấy khi khối lượng sản phẩm tăng lên thì giá thành giảm xuống, chứng tỏ hai tiêu thức đó có mối liên hệ tương quan. Nếu đưa số liệu lên đồ thị, ta thấy đường gấp khúc thực tế biến động dạng một đường hypebôn (bạn đọc có thể tự vẽ với trục hoành biểu diễn khối lượng sản phẩm sản xuất X, trục tung biểu diễn giá thành đơn vị sản phẩm Y).

Ta có đường hồi quy lý thuyết: $\bar{y}_x = a + \frac{b}{x}$

Các tham số a, b của phương trình được tính từ hệ phương trình:

$$\begin{cases} \Sigma y = na + b \Sigma \frac{1}{x} & (1) \end{cases}$$

$$\begin{cases} \Sigma \frac{1}{x} y = a \Sigma \frac{1}{x} + b \Sigma \left(\frac{1}{x} \right)^2 & (2) \end{cases}$$

Theo tài liệu của bảng 6-3, ta có thể lập bảng tính toán các số liệu cần thiết:

Bảng 6-4

Khối lượng sản phẩm sản xuất x_i	Giá thành một đơn vị sản phẩm y_i	$\frac{1}{x}$	$\left(\frac{1}{x} \right)^2$	$\frac{1}{x} y$
0,5	45,0	2,0000	4,0000	90,0000
1,0	25,0	1,0000	1,0000	25,0000
2,5	20,0	0,4000	0,1600	8,0000
4,0	15,0	0,2500	0,0625	3,7500
4,5	15,0	0,2222	0,0494	3,3333
5,0	12,5	0,2000	0,0400	2,5000
5,0	10,0	0,2000	0,0400	2,0000
7,5	7,5	0,1333	0,0178	0,9995
8,0	8,0	0,1250	0,0156	1,0000
10,0	7,0	0,1000	0,0100	0,7000
Cộng	165,0	4,6305	5,3953	137,2828

Thay kết quả tính toán trong bảng vào hệ phương trình chuẩn, ta có:

$$\begin{cases} 165 = 10a + 4,6305b \\ 137,2828 = 4,6305a + 5,3953b \end{cases}$$

Giải ra được: $a = 7,8$; $b = 18,72$

$$\Rightarrow \bar{y}_x = 7,8 + \frac{18,72}{x}$$

3.2. Tỷ số tương quan

Tỷ số tương quan là một chỉ tiêu dùng để đánh giá trình độ chặt chẽ của mối liên hệ tương quan phi tuyến tính.

Tỷ số tương quan được xác định dựa vào phương pháp luận cho rằng: Nếu tiêu thức nguyên nhân (x) càng có ảnh hưởng mạnh đến tiêu thức kết quả (y) thì phần phương sai do ảnh hưởng của tiêu thức x đối với y càng chiếm phần lớn so với phương sai chung của

y (do tất cả các nguyên nhân tác động). Điều đó có nghĩa là phương sai của tiêu thức nguyên nhân (x) là cơ sở để xác định tỷ lệ tương quan.

Tỷ lệ tương quan được tính theo công thức:

$$\eta = \sqrt{\frac{\sigma_{yx}^2}{\sigma_y^2}}$$

trong đó:

η (êta) - tỷ suất tương quan.

σ_y^2 - phương sai của tiêu thức kết quả y, phản ánh sai lệch bình quân giữa trị số thực tế và số bình quân chung.

$$\sigma_y^2 = \frac{\sum(y - \bar{y})^2}{n} = \bar{y}^2 - (\bar{y})^2$$

σ_{yx}^2 - phương sai của tiêu thức kết quả y do ảnh hưởng của tiêu thức nguyên nhân x, phản ánh sai lệch bình quân giữa trị số lý thuyết và số bình quân chung.

$$\sigma_{yx}^2 = \frac{\sum(\bar{y}_x - \bar{y})^2}{n}$$

Theo nguyên tắc cộng phương sai:

$$\sigma_y^2 = \sigma_{yx}^2 + \sigma_{(y-y_x)}^2$$

trong đó:

$\sigma_y^2, \sigma_{yx}^2$ - như trên.

$\sigma_{(y-y_x)}^2$ - phương sai của tiêu thức kết quả y, phản ánh sai lệch bình quân giữa trị số thực tế và trị số lý thuyết.

$$\sigma_{(y-y_x)}^2 = \frac{\sum(y - \bar{y})^2}{n}$$

Do mối liên hệ giữa các loại phương sai trên ta có:

$$\eta = \sqrt{\frac{\sigma_{yx}^2}{\sigma_y^2}} = \sqrt{\frac{\sigma_y^2 - \sigma_{(y-y_x)}^2}{\sigma_y^2}} = \sqrt{1 - \frac{\sigma_{(y-y_x)}^2}{\sigma_y^2}}$$

① Tỷ số tương quan có những tính chất sau:

Tỷ số tương quan có trị số từ 0 → 1

+ Nếu $\eta = 1$, có mối liên hệ hàm số giữa tiêu thức nguyên nhân x và tiêu thức kết quả y.

+ Nếu $\eta = 0$, không có mối liên hệ nào giữa các tiêu thức được khảo sát (x, y,...).

② Tỷ số tương quan càng gần 1 thì mối liên hệ tương quan càng chặt chẽ, càng gần 0 thì ngược lại. Tỷ số tương quan không cho biết tính chất thuận, nghịch của mối liên hệ đang xét.

Ví dụ: tính tỷ số tương quan để đánh giá mức độ chặt chẽ của mối liên hệ tương quan giữa khối lượng sản phẩm và giá thành đơn vị. Từ số liệu của bảng 6-4, ta lập bảng tính toán sau:

Bảng 6-5

x	y	y ²	$\bar{y}_x$	y - $\bar{y}_x$	(y - $\bar{y}_x$) ²
0,5	45,0	2.025,00	45,24	-0,24	0,058
1,0	25,0	625,00	26,52	-1,52	2,310
2,5	20,0	400,00	15,29	+4,71	22,184
4,0	15,0	225,00	12,48	+2,52	6,350
4,5	15,0	225,00	11,96	+3,04	9,242
5,0	12,5	156,25	11,54	+0,96	0,922
5,0	10,0	100,00	11,54	-1,54	2,372
7,5	7,5	56,25	10,30	-2,80	7,840
8,0	8,0	64,00	10,14	-2,14	4,580
10,0	7,0	49,00	9,67	-2,67	7,129
Cộng	165,0	3.925,50	-	-	62,987 ≈ 63

$$\bar{y} = \frac{165}{10} = 16,5$$

$$(\bar{y})^2 = (16,5)^2 = 272,25$$

$$\bar{y}^2 = \frac{3.925,5}{10} = 392,55$$

$$\sigma_y^2 = \bar{y}^2 - (\bar{y})^2 = 392,55 - 272,25 = 120,3$$

$$\sigma_{(y-\bar{y}_x)}^2 = \frac{63}{10} = 6,3$$

$$\eta = \sqrt{1 - \frac{63}{10}} = 0,9735$$

Kết quả trên cho ta thấy giữa khối lượng sản phẩm và giá thành đơn vị sản phẩm có mối liên hệ tương quan khá chặt chẽ.

IV. LIÊN HỆ TƯƠNG QUAN GIỮA HAI TIÊU THỨC “THAY PHIÊN”

Tiêu thức “thay phiên” là những tiêu thức có hai biểu hiện trái ngược (chẳng hạn tốt và không tốt, có và không có, biết chữ và mù chữ...). Ta lập bảng phân tổ kết hợp sau:

Bảng 6-6

I \ II	B	$\bar{B}$	Cộng
A	a	b	a + b
$\bar{A}$	c	d	c + d
Cộng	a + c	b + d	n

Trong bảng trên:

+ I, II: Các tiêu thức thay phiên có liên hệ nhau.

+ A, $\bar{A}$, B, $\bar{B}$: Các biểu hiện của tiêu thức.

+ a, b, c, d: Là tần số của mỗi tổ được phân tổ theo I và II.

Chỉ tiêu biểu hiện trình độ chặt chẽ của mối liên hệ giữa hai tiêu thức thay phiên gọi là hệ số kết hợp (K). Hệ số này được xác định theo công thức:

$$K = \frac{ad - bc}{\sqrt{(a+b)(c+d)(a+c)(b+d)}}$$

Hệ số kết hợp K cũng có trị số từ -1 đến +1. Khi hệ số của K = ± 1 có liên hệ hàm số giữa hai tiêu thức thay phiên. Ngược lại, nếu K = 0 không có mối liên hệ phụ thuộc giữa các tiêu thức đang xét.

Ví dụ: khảo sát mối quan hệ giữa hai tiêu thức giới tính và trình độ văn hoá của dân số tại một địa phương ở nước ta trong hai thời kỳ 1945 và 1960 theo tài liệu sau:

Bảng 6-7

(Đvt: %)

Trình độ VH \ Giới tính	Năm 1945			Năm 1960		
	Biết chữ	Mù chữ	Cộng	Biết chữ	Mù chữ	Cộng
Nam	49,5	50,5	100	99,3	0,7	100
Nữ	5,3	94,7	100	97,8	2,2	100
Cộng	54,8	145,2	-	197,1	2,9	-

$$K_{1945} = \frac{(49 \times 94,7) - (50,5 \times 5,3)}{\sqrt{100 \times 100 \times 54,8 \times 145,2}} = 0,490$$

$$K_{1960} = \frac{(99,3 \times 2,2) - (0,7 \times 97,8)}{\sqrt{100 \times 100 \times 197,1 \times 2,9}} = 0,063$$

Kết quả trên cho ta thấy:

- Vào thời kỳ 1945 dân ta mù chữ với tỷ lệ khá cao (nam với tỷ lệ 50,5%- nữ 94,7%). Nạn mù chữ biểu hiện rõ rệt ở giới nữ (tệ trọng nam khinh nữ). Điều đó cho thấy có mối liên hệ rõ rệt giữa giới tính và trình độ văn hoá trong thời kỳ này. Hệ số kết hợp $K = 0,490$ cũng đã cho thấy mối quan hệ nói trên.

- Ngược lại, vào thời kỳ 1960 dưới chế độ mới dân ta đã thoát khỏi nạn mù chữ. Mối quan hệ giữa hai tiêu thức giới tính và trình độ văn hoá tuy có biểu hiện nhưng không đáng kể. Hệ số kết hợp $K = 0,063$ cho ta thấy mối quan hệ không chặt chẽ giữa hai tiêu thức nói trên.

Chương VII

DÃY SỐ BIẾN ĐỘNG THEO THỜI GIAN

I. KHÁI NIỆM, PHÂN LOẠI VÀ Ý NGHĨA CỦA DÃY SỐ BIẾN ĐỘNG THEO THỜI GIAN

1.1. Khái niệm

Dãy số biến động theo thời gian là dãy các trị số của chỉ tiêu thống kê được sắp xếp theo thứ tự thời gian.

Dãy số biến động theo thời gian gồm có hai thành phần: thời gian tính và chỉ tiêu. Thời gian tính có thể là thời kỳ hay thời điểm; chỉ tiêu có thể được biểu hiện bằng số tuyệt đối, số tương đối, số bình quân.

Một dãy số biến động theo thời gian có dạng chung sau:

t_i	t_1	t_2	$\dots$	t_n
y_i	y_1	y_2	$\dots$	y_n

trong đó:

t_i - thời gian thứ i .

y_i - mức độ thứ i tương ứng với thời gian i .

1.2. Phân loại

Nếu căn cứ vào đặc điểm thời gian có thể chia ra làm hai loại là: dãy số thời kỳ và dãy số thời điểm.

+ Dãy số thời kỳ: mỗi mức độ phản ánh mặt lượng của hiện tượng trong suốt một khoảng thời gian nhất định.

+ Dãy số thời điểm: mỗi mức độ phản ánh mặt lượng của hiện tượng chỉ trong từng thời điểm nhất định.

Ví dụ: dãy số sau đây là dãy số thời kỳ:

Bảng 7-1

Năm	1997	1998	1999	2000	2001	2002
Lợi nhuận (Tr.d)	100	105	112	122	127	131

Dãy số sau đây là dãy số thời điểm:

Bảng 7-2

0 giờ ngày	01/01/2002	01/02/2002	01/3/2002	01/4/2002
Giá trị hàng tồn kho (Tr.đ)	350	364	366	382

Nếu căn cứ vào biểu hiện các mức độ của dãy số có thể chia ra làm ba loại là: dãy số tuyệt đối, dãy số tương đối và dãy số bình quân.

1.3. Ý nghĩa

Dãy số biến động theo thời gian vạch rõ xu hướng, tính quy luật của sự phát triển của hiện tượng, làm cơ sở dự đoán các mức độ tương lai của hiện tượng nghiên cứu. Ngoài ra, dãy số còn giúp nghiên cứu các đặc điểm về sự biến động của hiện tượng theo thời gian: tốc độ phát triển, tốc độ tăng trưởng, mức tăng giảm tuyệt đối,...

II. CÁC CHỈ TIÊU PHÂN TÍCH DÃY SỐ BIẾN ĐỘNG THEO THỜI GIAN

2.1. Mức độ bình quân theo thời gian ($\bar{y}$)

Chỉ tiêu này phản ánh mức độ điển hình của hiện tượng nghiên cứu theo thời gian, được tính bằng cách bình quân hóa các mức độ cá biệt trong dãy số.

$$\text{Đối với dãy số thời kỳ: } \bar{y} = \frac{\sum y_i}{n}$$

trong đó: n- số các mức độ trong dãy số.

Ví dụ: theo số liệu bảng 7-1, ta tính được lợi nhuận bình quân năm là:

$$\bar{y} = \frac{100 + 105 + 112 + 122 + 127 + 131}{6} = 116,166 \text{ (Tr.đ)}$$

Đối với dãy số thời điểm:

+ Trường hợp khoảng cách thời gian bằng nhau và có số liệu của dãy số tại nhiều thời điểm.

$$\bar{y} = \frac{\frac{y_1}{2} + y_2 + \dots + y_{n-1} + \frac{y_n}{2}}{n-1}$$

Ví dụ: theo số liệu bảng 7-2 ta tính được giá trị hàng tồn kho bình quân trong quý I là:

$$\bar{y} = \frac{\frac{356}{2} + 364 + 376 + \frac{382}{2}}{4-1} = 369,666 \text{ (Tr.đ)}$$

+ Trường hợp khoảng cách thời gian giữa các thời điểm khác nhau và có số liệu tại nhiều thời điểm.

$$\bar{y} = \frac{\sum y_i t_i}{\sum t_i}$$

trong đó: t_i - khoảng cách thứ i giữa các thời điểm.

Ví dụ: tính số công nhân bình quân trong danh sách của một doanh nghiệp trong tháng 4/2002 theo tài liệu sau:

- Ngày 01/4 có 246 người.
- Ngày 12/4 bổ sung thêm 4 người.
- Ngày 24/4 cho thôi việc 2 người, từ đó đến hết tháng số công nhân không thay đổi.

Ta lập bảng tính toán sau:

Bảng 7-3

Thời gian	Số ngày (t_i)	Số công nhân (y_i)
Từ 01/4 đến 11/4	11	246
Từ 12/4 đến 23/4	12	250
Từ 24/4 đến 30/4	7	248
Cộng	30	-

Số công nhân bình quân trong tháng 4 là:

$$\bar{y} = \frac{246 \times 11 + 250 \times 12 + 248 \times 7}{30} = 248 \text{ người}$$

2.2. Lượng tăng (giảm) tuyệt đối

Chỉ tiêu này phản ánh chênh lệch giữa hai mức độ của dãy số theo thời gian. Căn cứ vào việc chọn kỳ gốc khác nhau chia ra:

- Lượng tăng (giảm) tuyệt đối từng kỳ: phản ánh chênh lệch giữa hai mức độ cạnh nhau.

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

- Lượng tăng (giảm) tuyệt đối định gốc: phản ánh chênh lệch giữa một mức độ nào đó với mức độ của kỳ được chọn làm gốc cố định. Nếu chọn mức độ đầu tiên của dãy số làm gốc cố định (y_1) ta có:

$$\Delta_i = y_i - y_1 \quad (i = 2, 3, \dots, n)$$

Mối quan hệ giữa lượng tăng (giảm) tuyệt đối từng kỳ và định gốc là tổng đại số các lượng tăng (giảm) tuyệt đối từng kỳ bằng lượng tăng (giảm) tuyệt đối định gốc.

$$\Delta_k = \sum \delta_i \quad (i = 2, 3, \dots, k \text{ và } k \leq n)$$

Ví dụ: tính lượng tăng (giảm) tuyệt đối của chỉ tiêu doanh số bán hàng của doanh nghiệp thương mại X qua các năm như sau:

Bảng 7- 4

(Đơn vị: Triệu đồng)

Năm	1998	1999	2000	2001	2002
Doanh số bán hàng	2112	2213,4	2304,1	2384,7	2449,6
Lượng tăng (giảm) từng kỳ	-	+101,4	+90,7	+80,6	+64,9
Lượng tăng (giảm) định gốc	-	+101,4	+192,1	+272,7	+337,6

Để phản ánh mức độ tăng (giảm) bình quân cho cả thời kỳ nghiên cứu, ta tính lượng tăng (giảm) tuyệt đối bình quân.

$$\bar{\delta} = \frac{\sum \delta_i}{n-1}$$

trong đó:

δ_i - lượng tăng (giảm) tuyệt đối từng kỳ ($i = 2, 3, \dots, n$).

n - số mức độ trong dãy số.

Ví dụ: theo số liệu bảng 7-4, ta tính được lượng tăng tuyệt đối bình quân hàng năm của doanh số bán hàng là:

$$\bar{\delta} = \frac{(+102,4) + (+90,7) + (+80,6) + (+64,9)}{4} = 84,4 \text{ (Tr,đ)}$$

2.3. Tốc độ phát triển

Chỉ tiêu này phản ánh mức độ kỳ báo cáo so với mức độ kỳ gốc bằng bao nhiêu lần (hay bao nhiêu phần trăm). Căn cứ vào việc chọn kỳ gốc khác nhau, chia ra:

- Tốc độ phát triển liên hoàn (t_i) phản ánh tốc độ phát triển giữa hai thời kỳ liên nhau.

$$t_i = \frac{y_i}{y_{i-1}} \quad (i = 2, 3, \dots, n)$$

- Tốc độ phát triển định gốc (T_i) phản ánh tốc độ phát triển của nhiều thời kỳ nếu so với một thời kỳ được chọn là gốc cố định cho mọi lần so sánh (thường là mức độ đầu tiên dãy số y_1).

$$T_i = \frac{y_i}{y_1} \quad (i = 2, 3, \dots, n)$$

Mối quan hệ giữa tốc độ phát triển liên hoàn và định gốc là:

$$T_k = t_2 \cdot t_3 \cdot \dots \cdot t_k = \prod t_i \quad (i = 2, 3, \dots, k \text{ và } k \leq n)$$

Ví dụ: tính tốc độ phát triển của chỉ tiêu doanh số bán hàng của doanh nghiệp thương mại X qua các năm:

Bảng 7-5

Năm	1998	1999	2000	2001	2002
Doanh số bán hàng (Tr.d)	2112	2213,4	2304,1	2384,7	2449,6
Tốc độ phát triển liên hoàn	-	1,0480	1,0409	1,0349	1,0272
Tốc độ phát triển định gốc	-	1,0480	1,0909	1,1291	1,1598

Để phản ánh nhịp điệu phát triển điển hình của hiện tượng trong cả thời kỳ nghiên cứu, ta tính tốc độ phát triển bình quân:

$$\bar{t} = \sqrt[n]{t_2 t_3 \dots t_n}$$

trong đó:

$t_2, t_3, \dots, t_n$ - các tốc độ phát triển liên hoàn.

$n-1$ - số tốc độ phát triển liên hoàn tham gia bình quân hóa .

Ví dụ: theo số liệu bảng 7-5 ta có:

$$\bar{t} = \sqrt[4]{1,0480 \times 1,0409 \times 1,0349 \times 1,0272} = 1,0377 \text{ hay } 103,77\%$$

2.4. Tốc độ tăng (giảm)

Là chỉ tiêu phản ánh mức độ kỳ báo cáo đã tăng lên (hay giảm) bao nhiêu lần (hay bao nhiêu phần trăm) so với kỳ gốc. Căn cứ vào việc chọn gốc so sánh, chia ra:

- Tốc độ tăng liên hoàn (từng kỳ): $a_i = \frac{\delta_i}{y_{i-1}} = \frac{y_i - y_{i-1}}{y_{i-1}} = t_i - 1$

- Tốc độ tăng định gốc: $b_i = \frac{\Delta_i}{y_1} = \frac{y_i - y_1}{y_1} = T_i - 1$

- Tốc độ tăng bình quân: $r = \bar{t} - 1$

Ví dụ: tính tốc độ tăng của doanh số bán hàng của một doanh nghiệp thương mại X qua các năm:

Bảng 7-6

Năm	1998	1999	2000	2001	2002
Doanh số bán hàng	2112	2213,4	2304,1	2384	2449,6
Tốc độ tăng liên hoàn (%)	-	+4,80	+4,09	+3,49	+2,72
Tốc độ tăng định gốc (%)	-	+4,80	+12,91	+12,91	+15,98

Tốc độ tăng bình quân của doanh số bán hàng hàng năm là:

$$r = \bar{t} - 1 = 1,0377 - 1 = +0,0377 \text{ hay } +3,77\%$$

2.5. Giá trị tuyệt đối của 1% tăng (giảm)

Chỉ tiêu này xác định nội dung thực tế của 1% tăng lên (hay giảm) có quy mô, khối lượng cụ thể là bao nhiêu. Nó được tính bằng cách lấy lượng tăng giảm tuyệt đối từng kỳ so sánh với tốc độ tăng giảm từng kỳ, tính bằng số phần trăm.

III. CÁC PHƯƠNG PHÁP BIỂU HIỆN XU HƯỚNG PHÁT TRIỂN CƠ BẢN CỦA HIỆN TƯỢNG

Các dãy số biến động theo thời gian thường bị ảnh hưởng của các nhân tố ngẫu nhiên, làm dãy số chưa phản ánh một cách thực sự khách quan, chính xác quy luật của hiện tượng nghiên cứu. Vì vậy cần loại trừ các nhân tố ngẫu nhiên đó. Có thể sử dụng các phương pháp sau:

- Phương pháp mở rộng khoảng cách thời gian.
- Phương pháp số bình quân trượt.
- Phương pháp hồi quy.
- Phương pháp biểu hiện quy luật biến động thời vụ.

3.1. Phương pháp mở rộng khoảng cách thời gian

Trên cơ sở dãy số đã có lập một dãy số mới với khoảng cách thời gian dài hơn. Việc mở rộng khoảng cách thời gian phải tùy tính chất của hiện tượng mà lựa chọn cho phù hợp với quy luật của chúng. Có như vậy dãy số mới có ý nghĩa.

3.2. Phương pháp số bình quân trượt

Theo phương pháp này các số bình quân trượt được tính từ một nhóm nhất định các mức độ trong dãy số, bắt đầu từ mức độ đầu tiên, tiếp theo bắt đầu từ mức độ thứ hai, tiếp theo nữa bắt đầu từ mức độ thứ ba,...

Giả sử ta có dãy số: $y_1, y_2, \dots, y_n$. Nếu tính số bình quân trượt theo từng nhóm 5 mức độ ta có:

$$\bar{y}_1 = \frac{y_1 + y_2 + y_3 + y_4 + y_5}{5}$$

$$\bar{y}_2 = \frac{y_2 + y_3 + y_4 + y_5 + y_6}{5}$$

$$\bar{y}_3 = \frac{y_3 + y_4 + y_5 + y_6 + y_7}{5}$$

v.v...

3.3. Phương pháp hồi quy

Nếu đưa lên hệ toạ độ vuông góc với trục hoành là thời gian, trục tung là giá trị của hiện tượng nghiên cứu, thì các mức độ của dãy số sẽ tạo thành một đường gấp khúc thực tế. Tuy nhiên đường gấp khúc thực tế bị các nhân tố ngẫu nhiên tác động nên không nêu bật được quy luật của hiện tượng. Vì vậy để rút ra tính quy luật, loại trừ ảnh hưởng ngẫu nhiên, thống kê học sử dụng phương trình toán học, phương trình hồi quy để thay thế sao cho miêu tả một cách sát nhất sự biến động thực tế của hiện tượng. Tuy nhiên việc chọn phương trình hồi quy (đường thẳng hay đường cong) không thể tùy tiện mà phải dựa vào đồ thị thực nghiệm và phân tích lý luận để xác định bản chất và tính quy luật của hiện tượng.

Ta hãy xét một trường hợp đơn giản: điều chỉnh mức độ của dãy số biến động theo thời gian bằng phương trình tuyến tính có dạng:

$$\bar{y}_t = a + bt$$

trong đó:

t - thời gian.

$\bar{y}_t$ - mức độ lý thuyết của dãy số.

a, b - các tham số.

Đây là đường hồi quy lý thuyết có dạng tuyến tính. Vì vậy, ta có hệ phương trình sau để xác định các tham số a, b :

$$\begin{cases} \Sigma y = na + b\Sigma t \\ \Sigma yt = a\Sigma t + b\Sigma t^2 \end{cases}$$

Giải hệ phương trình ta được.

$$a = \frac{\Sigma y}{n} - b \cdot \frac{\Sigma t}{n} = \bar{y} - b \cdot \bar{t}$$

$$b = \frac{n\Sigma ty - \Sigma t\Sigma y}{n\Sigma t^2 - (\Sigma t)^2}$$

Khi $\Sigma t = 0$ ta có: $a = \bar{y}$

$$b = \frac{\Sigma ty}{\Sigma t^2}$$

Muốn làm cho $\Sigma t = 0$ ta có thể quy ước lại thứ tự thời gian. Có 2 trường hợp:

- Nếu dãy số có lẻ các mức độ, chọn thời gian giữa làm mốc và cho $t = 0$, các thời gian đứng trước nó nhận các giá trị $-1, -2, -3, \dots$ các thời gian sau nó nhận các giá trị $+1, +2, +3, \dots$

- Nếu dãy số có chẵn các mức độ, chọn 2 thời gian giữa dãy số và cho chúng các giá trị -1, +1, các thời gian đứng trước nhận các giá trị -3, -5, -7... và đứng sau nhận +3, +5, +7, ...

Ví dụ: có tài liệu về sản lượng sản phẩm của một doanh nghiệp sản xuất như sau:

Bảng 7-7

Năm	Sản lượng sản phẩm (1000 cái) y	t	t ²	t.y	$\bar{y}_t$
1995	250	-3	9	-750	257,86
1996	300	-2	4	-600	272,86
1997	260	-1	1	-260	287,86
1998	320	0	0	0	302,86
1999	300	+1	1	300	317,86
2000	340	+2	4	680	332,86
2001	350	+3	9	1050	347,86
Cộng	2120	0	28	420	

$$a = y = \frac{\Sigma y}{n} = \frac{2120}{7} = 302,86$$

$$b = \frac{\Sigma ty}{\Sigma t^2} = \frac{420}{28} = 15$$

Vậy: $\bar{y}_t = 302,86 + 15t$

Nhờ phương trình này ta lập được dãy số mới (xem cột $\bar{y}_t$ của bảng 7-7)

- Khi $t = 0$ (1998), $\bar{y}_t = 302,86$ (1000 cái)
- Khi $t = -1$ (1997), $\bar{y}_t = 302,86 + (-15) = 287,86$ (1000 cái)
- Khi $t = -2$ (1996), $\bar{y}_t = 302,86 + (-15 \times 2) = 272,86$ (1000 cái)
- Khi $t = -3$ (1995), $\bar{y}_t = 302,86 + (-15 \times 3) = 257,86$ (1000 cái)
- Khi $t = +1$ (1999), $\bar{y}_t = 302,86 + 15 = 317,86$ (1000 cái)
- Khi $t = +2$ (2000), $\bar{y}_t = 302,86 + (15 \times 2) = 332,86$ (1000 cái)
- Khi $t = +3$ (2001), $\bar{y}_t = 302,86 + (15 \times 3) = 347,86$ (1000 cái)

Dãy số vừa điều chỉnh biểu hiện khá rõ xu hướng tăng lên không ngừng của sản lượng sản phẩm của doanh nghiệp.

3.4. Phương pháp biểu hiện quy luật biến động thời vụ

Biểu hiện của tính thời vụ thể hiện ở chỗ mức độ của hiện tượng tăng lên hoặc giảm đi rõ rệt vào một thời kỳ nhất định trong năm. Nguyên nhân biến động có thể do ảnh hưởng của điều kiện thiên nhiên hoặc tập quán tiêu dùng.

Để chỉ rõ mức độ của biến động thời vụ trong thống kê dùng chỉ tiêu “chỉ số thời vụ” tính theo công thức:

$$I_{iv} = \frac{\bar{y}_i}{\bar{y}_0} \times 100$$

trong đó:

I_{iv} - chỉ số thời vụ.

$\bar{y}_i$ - số bình quân của các mức độ các tháng cùng tên.

$\bar{y}_0$ - số bình quân chung của tất cả các mức độ trong dãy số.

Ví dụ: có tài liệu về mức tiêu thụ của một loại hàng tại một khu vực trong 3 năm từ 2000 - 2002:

Bảng 7-8

Tháng	Mức tiêu thụ (triệu đồng)			Cộng các tháng cùng tên Σy_i	Số bình quân các tháng cùng tên $\bar{y}_i$	Chỉ số thời vụ $\frac{\%}{y_i}$ $I_{iv} = \frac{y_i}{y_0} \times 100$
	2000	2001	2002			
1	4	5	6	15	5	8,57
2	11	10	24	45	15	25,71
3	25	23	27	75	25	42,85
4	62	61	72	195	65	111,43
5	79	73	88	240	80	137,15
6	120	135	135	390	130	222,87
7	140	123	157	420	140	240,01
8	134	82	144	360	120	205,72
9	62	64	84	210	70	120,00
10	38	40	42	120	40	68,57
11	2	5	8	15	5	8,57
12	2	5	8	15	5	8,57
Cộng	679	626	795	2100	58,33	100,00

Cách tính: $\bar{y}_1 (\text{tháng } 1) = \frac{\Sigma y_1}{3} = \frac{15}{3} = 5$ triệu đồng.

$$\bar{y}_2 (\text{tháng } 2) = \frac{\Sigma y_2}{3} = \frac{45}{3} = 15 \text{ triệu đồng.}$$

v.v...

$\bar{y}_0$ - số bình quân chung.

$$\bar{y}_0 = \frac{\Sigma y_i}{n} = \frac{2100}{36} = 58,33 \text{ triệu đồng.}$$

Số liệu trên cho thấy, tiêu thụ tháng 1 chỉ bằng 8,57% mức bình quân chung, tháng 2 bằng 25,71%, tháng 3 bằng 42,85% ...

Như vậy mặt hàng này tiêu thụ mạnh (trên mức bình quân chung) vào các tháng từ 4 đến 9, và tiêu thụ ít (dưới mức bình quân chung) vào các tháng mùa rét.

Chương VIII

CHỈ SỐ THỐNG KÊ

I. KHÁI NIỆM, ĐẶC ĐIỂM VÀ TÁC DỤNG CỦA CHỈ SỐ

1.1. Khái niệm chỉ số

Chỉ số là chỉ tiêu tương đối biểu hiện quan hệ so sánh giữa các mức độ của một hiện tượng kinh tế ở hai thời gian hoặc địa điểm khác nhau nhằm nêu lên sự biến động của hiện tượng qua thời gian, qua không gian, qua các kỳ kế hoạch.

Khái niệm trên giúp ta phân biệt chỉ số với các số tương đối. Theo đó các loại số tương đối động thái, số tương đối kế hoạch và số tương đối không gian là chỉ số. Còn các loại số tương đối khác như số tương đối kết cấu, số tương đối cường độ không phải là chỉ số vì nó không thể hiện mối quan hệ so sánh giữa hai mức độ của cùng một hiện tượng kinh tế.

Tuy nhiên, đối tượng nghiên cứu chủ yếu của chỉ số là các hiện tượng kinh tế phức tạp mà số tương đối không phản ánh được, đó là những hiện tượng kinh tế bao gồm nhiều thành phần, nhiều đơn vị hoặc phần tử có tính chất khác nhau (về tên gọi, giá trị sử dụng, đơn vị đo lường,...), mà các mức độ của chúng không thể trực tiếp cộng lại với nhau. Việc so sánh những hiện tượng kinh tế phức tạp trong những điều kiện thời gian và không gian khác nhau không phải dễ dàng mà phải thông qua phương pháp chỉ số và kết quả so sánh được gọi là chỉ số.

Vậy chỉ số là chỉ tiêu tương đối thể hiện sự biến động của hiện tượng kinh tế phức tạp bao gồm các phần tử không thể cộng trực tiếp với nhau.

1.2. Đặc điểm của phương pháp chỉ số

Khi nghiên cứu biến động của một tổng thể phức tạp bao gồm các phần tử không thể trực tiếp cộng được với nhau, phương pháp chỉ số biến đổi chúng thành những phần tử có thể trực tiếp cộng được với nhau, dựa trên cơ sở mối quan hệ giữa nhân tố nghiên cứu với các nhân tố khác.

Mặt khác khi nghiên cứu biến động của một nhân tố, phương pháp chỉ số giả định các nhân tố khác còn lại không thay đổi.

Ví dụ: khi nghiên cứu biến động về lượng của nhiều mặt hàng, chúng không thể trực tiếp cộng được với nhau, phương pháp chỉ số tính giá trị khối lượng hàng hóa có thể trực tiếp cộng được với nhau. Tuy nhiên, giá trị khối lượng hàng hóa là tích số của đơn giá hàng hóa với khối lượng hàng hóa. Do vậy, để so sánh giá trị khối lượng hàng hóa nhưng lại nghiên cứu sự biến động của lượng hàng hóa, phương pháp chỉ số phải cố định nhân tố đơn giá hàng hóa.

1.3. Tác dụng của chỉ số

Chỉ số có tác dụng lớn trong phân tích thống kê:

- Nghiên cứu sự biến động của hiện tượng kinh tế qua thời gian (chỉ số phát triển).
- Nghiên cứu sự biến động của hiện tượng kinh tế qua không gian (chỉ số không gian hay chỉ số địa phương).
- Nghiên cứu tình hình xây dựng và thực hiện kế hoạch các chỉ tiêu kinh tế (chỉ số kế hoạch).
- Phân tích vai trò và ảnh hưởng của từng nhân tố đến biến động của toàn bộ hiện tượng phức tạp (hệ thống chỉ số).

1.4. Các loại chỉ số

a) Căn cứ vào kỳ nghiên cứu biến động, có thể chia thành ba loại chỉ số là: chỉ số phát triển, chỉ số không gian, chỉ số kế hoạch.

b) Căn cứ vào phạm vi tính toán, có hai loại chỉ số là: chỉ số cá thể và chỉ số chung.

- *Chỉ số cá thể:* phản ánh sự biến động của từng phần tử, từng đơn vị cá biệt trong tổng thể. Chỉ số cá thể ký hiệu là i .

Ví dụ: chỉ số giá cả từng mặt hàng, chỉ số lượng hàng hóa tiêu thụ của từng mặt hàng,...

- *Chỉ số chung:* phản ánh sự biến động của tất cả các phần tử, các đơn vị thuộc tổng thể hiện tượng phức tạp. Chỉ số chung ký hiệu là I .

Ví dụ: chỉ số giá cả của tất cả các mặt hàng bán lẻ tại một thị trường, chỉ số năng suất lao động của toàn bộ công nhân trong một doanh nghiệp sản xuất,...

c) Căn cứ vào tính chất của chỉ tiêu nghiên cứu, có hai loại chỉ số là: chỉ số chỉ tiêu chất lượng và chỉ số chỉ tiêu khối lượng.

- *Chỉ số chỉ tiêu chất lượng:* phản ánh sự biến động của các chỉ tiêu như: giá cả, giá thành, tiền lương, năng suất lao động,...

- *Chỉ số chỉ tiêu số lượng:* phản ánh sự biến động của các chỉ tiêu như: lượng hàng hóa tiêu thụ, lượng sản phẩm sản xuất, số lượng công nhân,...

Trong các phần sau đây chủ yếu đề cập đến các chỉ số phát triển vì nó có vị trí quan trọng trong phân tích kinh tế của các đơn vị cơ sở.

II. PHƯƠNG PHÁP TÍNH CHỈ SỐ

Để nghiên cứu phương pháp tính chỉ số phát triển phải có tài liệu của hiện tượng nghiên cứu kỳ nghiên cứu và kỳ gốc.

Ví dụ: có tài liệu về giá cả và lượng hàng tiêu thụ của một doanh nghiệp thương mại như sau:

Bảng 8-1

Tên hàng	Đơn vị tính	Đơn giá bán (1000đ)		Lượng hàng hóa tiêu thụ	
		Kỳ gốc	Kỳ báo cáo	Kỳ gốc	Kỳ báo cáo
A	cái	5,0	4,5	2.000	2.500
B	kg	1,2	1,0	5.000	5.300
C	m	2,0	2,0	1.000	1.200

2.1. Phương pháp tính chỉ số cá thể

2.1.1. Chỉ số cá thể của chỉ tiêu chất lượng

Công thức:
$$i_p = \frac{p_1}{p_0}$$

Chênh lệch tuyệt đối: $(p_1 - p_0)$

Với số liệu của bảng 8-1, giá là chỉ tiêu chất lượng (p) và chỉ số cá thể về giá (i_p) là chỉ số giá của từng mặt hàng.

Chỉ số giá của mặt hàng A: $i_{pA} = \frac{4,5}{5} = 0,9$ (hay 90%)

Chênh lệch tuyệt đối: $4,5 - 5 = -0,5$ (nghìn đồng/cái).

Nhận xét: Giá bán mặt hàng A kỳ báo cáo so với kỳ gốc giảm 10% tương ứng với giảm 0,5 (nghìn đồng /cái).

Tương tự tính chỉ số giá của các mặt hàng B, C.

2.1.2. Chỉ số cá thể của chỉ tiêu khối lượng

Công thức:
$$i_q = \frac{q_1}{q_0}$$

Chênh lệch tuyệt đối: $(q_1 - q_0)$

Với số liệu của bảng 8-1, lượng hàng tiêu thụ là chỉ tiêu khối lượng (q) và chỉ số cá thể về lượng (i_q) là chỉ số lượng hàng tiêu thụ của từng mặt hàng.

Chỉ số lượng hàng tiêu thụ của mặt hàng A: $i_{qA} = \frac{2.500}{2.000} = 1,25$ (hay 125%)

Chênh lệch tuyệt đối: $2.500 - 2.000 = +500$ cái

Nhận xét: lượng hàng tiêu thụ của mặt hàng A kỳ báo cáo so với kỳ gốc tăng 25% tương ứng với tăng 500 cái.

Tương tự tính chỉ số lượng hàng tiêu thụ của các mặt hàng B, C.

2.2. Chỉ số chung

Chỉ số chung có thể biểu hiện dưới hai dạng là: Chỉ số tổng hợp và chỉ số bình quân.

2.2.1. Chỉ số tổng hợp: là dạng cơ bản của chỉ số chung

- *Chỉ số tổng hợp của chỉ tiêu chất lượng:*

Công thức:
$$I_p = \frac{\sum p_1 q_1}{\sum p_0 q_1}$$

Chênh lệch tuyệt đối: $\sum p_1 q_1 - \sum p_0 q_1$

Theo tài liệu của bảng 8-1 ta có:

$$I_p = \frac{4,5 \times 2.500 + 1,0 \times 5.300 + 2,0 \times 1.200}{5 \times 2.500 + 1,2 \times 5.300 + 2,0 \times 1.200} = \frac{18.950}{21.260} = 0,891 \text{ (hay } 89,1\%)$$

Chênh lệch tuyệt đối: $18.950 - 21.260 = -2.310$ (nghìn đồng)

Nhận xét: giá bán của các mặt hàng kỳ báo cáo so với kỳ gốc giảm 10,9% làm cho tổng mức tiêu thụ giảm 2310 nghìn đồng.

- *Chỉ số tổng hợp của chỉ tiêu khối lượng:*

Công thức:
$$I_q = \frac{\sum p_0 q_1}{\sum p_0 q_0}$$

Chênh lệch tuyệt đối: $\sum p_0 q_1 - \sum p_0 q_0$

Theo tài liệu của bảng 8-1 ta có:

$$I_q = \frac{5 \times 2.500 + 1,2 \times 5.300 + 2,0 \times 1.200}{5 \times 2.000 + 1,2 \times 5.000 + 2,0 \times 1.000} = \frac{21.260}{18.000} = 1,181 \text{ (hay } 118,1\%)$$

Chênh lệch tuyệt đối: $21.260 - 18.000 = +3.260$ (nghìn đồng)

Nhận xét: lượng tiêu thụ của các mặt hàng kỳ báo cáo so với kỳ gốc tăng 18,1% làm cho tổng mức tiêu thụ tăng lên là 3.260 nghìn đồng.

- *Chỉ số tổng hợp của chỉ tiêu tổng thể:*

Công thức:
$$I_{pq} = \frac{\sum p_1 q_1}{\sum p_0 q_0}$$

Chênh lệch tuyệt đối: $\sum p_1 q_1 - \sum p_0 q_0$

Theo tài liệu của bảng 8-1 ta có:

$$I_{pq} = \frac{4,5 \times 2.500 + 1,0 \times 5.300 + 2,0 \times 1.200}{5,0 \times 2.000 + 1,2 \times 5.000 + 2,0 \times 1.000} = \frac{18.950}{18.000} = 1,052 \text{ (hay } 105,2\%)$$

Chênh lệch tuyệt đối: $18.950 - 18.000 = +950$ nghìn đồng.

Nhận xét: tổng mức tiêu thụ hàng hóa kỳ báo cáo so với kỳ gốc tăng 5,2% tương ứng với mức tăng 950 nghìn đồng.

Qua việc xây dựng các chỉ số chung như trên, ta rút ra các kết luận sau:

Kết luận 1: Trong công thức tính chỉ số chung có 2 nhân tố cấu thành là: nhân tố chỉ số hóa và quyền số.

- Nhân tố mà ta cần nghiên cứu sự biến động của nó gọi là *nhân tố chỉ số hóa*.

Ví dụ: trong chỉ số chung về giá, nhân tố chỉ số hóa là giá cả của các loại hàng, hoặc trong chỉ số chung về khối lượng sản phẩm, nhân tố chỉ số hóa là khối lượng sản phẩm mỗi loại.

- Nhân tố kia có tác dụng quy định sự ảnh hưởng của nó đến sự biến động của chỉ số và được cố định ở một kỳ nào đó (kỳ gốc hay kỳ báo cáo) gọi là *quyền số*.

Ví dụ: trong chỉ số giá cả, quyền số là lượng sản phẩm tiêu thụ, trong chỉ số khối lượng sản phẩm, quyền số là giá cả các mặt hàng.

Kết luận 2: Chỉ số chung được tính bằng cách nhân trực tiếp nhân tố chỉ số hóa với quyền số như trên đã trình bày gọi là *chỉ số tổng hợp*, nó là hình thức cơ bản của chỉ số chung.

Kết luận 3: Về phương pháp tính chỉ số tổng hợp:

- Khi dùng chỉ số để nghiên cứu sự biến động của hiện tượng kinh tế phức tạp, trước hết cần chuyển tổng thể đó sang một tổng thể khác bao gồm các phần tử có thể trực tiếp cộng được với nhau. Để giải quyết vấn đề này, người ta xác định cho nhân tố chỉ số hóa một nhân tố khác làm nhân tố chuyển tổng thể, đồng thời giữ vai trò làm quyền số. Thông thường, căn cứ vào các phương trình kinh tế biểu hiện mối quan hệ giữa các nhân tố, từ đó khi nghiên cứu nhân tố này thì dùng nhân tố kia làm quyền số của chỉ số.

- Khi nghiên cứu sự biến động nhân tố nào đó, thì phải cố định nhân tố khác. Có như vậy thì mới nêu lên được sự biến động riêng biệt của nhân tố cần nghiên cứu, trong khi đó các nhân tố khác vẫn tham gia vào việc tính chỉ số.

Kết luận 4: Về việc lựa chọn thời kỳ của quyền số trong chỉ số tổng hợp.

Việc lựa chọn thời kỳ quyền số để cố định quyền số ở kỳ gốc hay kỳ báo cáo là do việc phân tích nội dung của chỉ số quyết định. Nhìn chung có thể thấy:

- Đối với chỉ số của chỉ tiêu chất lượng, quyền số thường là chỉ tiêu số lượng được cố định ở kỳ báo cáo.

- Đối với chỉ số của chỉ tiêu số lượng, quyền số thường là chỉ tiêu chất lượng được cố định ở kỳ gốc.

Tuy nhiên, tùy theo yêu cầu nghiên cứu, dựa vào việc phân tích nội dung kinh tế mà chỉ số phản ánh, việc cố định quyền số có thể khác với những trường hợp chung trên.

Ví dụ: để loại trừ ảnh hưởng biến động của nhân tố giá cả khi tính chỉ số sản lượng sản phẩm công nghiệp qua nhiều năm, người ta không dùng quyền số là giá kỳ gốc mà dùng hệ thống giá cố định do Nhà nước quy định.

2.2.2. Chỉ số bình quân

Là một dạng chỉ số chung, nó là số bình quân của các chỉ số cá thể. Thông thường chỉ số bình quân được sử dụng khi thiếu tài liệu tính toán chỉ số tổng hợp. Nhưng kết quả tính toán của chỉ số bình quân và chỉ số tổng hợp sẽ nhất trí với nhau nếu xuất phát từ một nguồn tài liệu. Có hai loại chỉ số bình quân là: chỉ số bình quân cộng và chỉ số bình quân điều hòa.

- *Chỉ số bình quân cộng:* là số bình quân cộng gia quyền của các chỉ số cá thể, tính theo công thức:

$$I_q = \frac{\sum i_q p_0 q_0}{\sum p_0 q_0}$$

Chênh lệch tuyệt đối: $\sum i_q p_0 q_0 - \sum p_0 q_0$

Trong trường hợp cùng một nguồn tài liệu như nhau, chỉ số bình quân của lượng hàng tiêu thụ như trên có kết quả tính toán giống với chỉ số tổng hợp về lượng hàng hóa tiêu thụ.

Thật vậy, thay $i_q = \frac{q_1}{q_0}$ vào công thức tính chỉ số bình quân ta có:

$$I_q = \frac{\sum i_q p_0 q_0}{\sum p_0 q_0} = \frac{\sum \frac{q_1}{q_0} p_0 q_0}{\sum p_0 q_0} = \frac{\sum p_0 q_1}{\sum p_0 q_0}$$

- *Chỉ số bình quân điều hòa:* là số bình quân điều hòa gia quyền của các chỉ số cá thể, tính theo công thức:

$$I_p = \frac{\sum p_1 q_1}{\sum \frac{1}{i_p} p_1 q_1}$$

Chênh lệch tuyệt đối: $\sum p_1 q_1 - \sum \frac{1}{i_p} p_1 q_1$

Trong trường hợp cùng một nguồn tài liệu như nhau, chỉ số bình quân của giá cả như trên có kết quả tính toán giống với chỉ số tổng hợp về giá cả. Thật vậy nếu thay

$i_p = \frac{p_1}{p_0}$ vào công thức tính chỉ số bình quân, ta có:

$$I_p = \frac{\sum p_1 q_1}{\sum \frac{1}{i_p} p_1 q_1} = \frac{\sum p_1 q_1}{\sum \frac{p_0}{p_1} p_1 q_1} = \frac{\sum p_1 q_1}{\sum p_0 q_1}$$

Ví dụ: có tài liệu về tình hình giá cả và tiêu thụ tại một doanh nghiệp thương mại như sau:

Bảng 8-2

Tên hàng	Mức tiêu thụ (1000đ)		Chỉ số cá thể	
	Kỳ gốc (p_0, q_0)	Kỳ báo cáo (p_1, q_1)	i_q	i_p
A	10.000	11.250	1,26	0,90
B	6.000	5.300	1,06	0,833
C	2.000	2.400	1,20	1,00

Theo tài liệu trên, muốn tính chỉ số chung lượng hàng hóa tiêu thụ phải áp dụng công thức chỉ số bình quân cộng:

$$I_q = \frac{1,25 \times 10.000 + 1,06 \times 6.000 + 1,2 \times 2.000}{10.000 + 6.000 + 2.000} = \frac{21.260}{18.000} = 1,181 \text{ (hay 118,1\%)}$$

Chênh lệch tuyệt đối: $21.260 - 18.000 = + 3.260$ nghìn đồng.

Nhận xét: lượng tiêu thụ của các mặt hàng kỳ báo cáo so với kỳ gốc tăng 18,1% làm cho tổng mức tiêu thụ tăng lên là 3.260 nghìn đồng.

Cũng nguồn tài liệu trên muốn tính chỉ số chung giá phải áp dụng công thức chỉ số bình quân điều hòa:

$$I_p = \frac{\frac{11.250}{0,9} + \frac{5.300}{0,833} + \frac{2.400}{1,0}}{\frac{11.250}{1,26} + \frac{5.300}{1,06} + \frac{2.400}{1,20}} = \frac{18.950}{21.260} = 0,891 \text{ (hay 89,1\%)}$$

Chênh lệch tuyệt đối: $18.950 - 21.260 = -2.310$ (nghìn đồng)

Nhận xét: giá bán của các mặt hàng kỳ báo cáo so với kỳ gốc giảm 10,9% làm cho tổng mức tiêu thụ giảm 2.310 nghìn đồng.

III. HỆ THỐNG CHỈ SỐ

Các hiện tượng kinh tế phức tạp thường bao gồm nhiều nhân tố cấu tạo nên nó, mà các nhân tố này có quan hệ tích số với nhau. Ví dụ: Mức tiêu thụ bao gồm hai nhân tố là giá cả đơn vị và lượng hàng hóa tiêu thụ. Mỗi quan hệ giữa mức tiêu thụ và các nhân tố đó được biểu hiện ở phương trình kinh tế.

$$\text{Mức tiêu thụ} = \text{Giá cả đơn vị} \times \text{Lượng hàng hóa tiêu thụ}$$

Cũng vậy ta còn có phương trình kinh tế khác:

$$\frac{\text{Tổng giá thành}}{\text{sản phẩm}} = \frac{\text{Giá thành đơn vị}}{\text{sản phẩm}} \times \frac{\text{Số sản phẩm}}{\text{sản xuất ra}}$$

$$\frac{\text{Tổng sản lượng}}{\text{sản phẩm}} = \frac{\text{Năng suất lao động}}{\text{l công nhân}} \times \frac{\text{Số lượng}}{\text{công nhân}}$$

Như vậy bản thân hiện tượng biến động là kết quả tổng hợp các sự biến động các nhân tố gây nên. Do đó khi nghiên cứu sự biến động của hiện tượng phức tạp nhiều nhân tố có quan hệ tích số với nhau, một vấn đề quan trọng đặt ra là phải xác định được vai trò và ảnh hưởng của sự biến động từng nhân tố đến sự biến động chung của toàn bộ hiện tượng như thế nào, qua đó giúp ta đánh giá được nhân tố nào có ảnh hưởng nhiều, ít, tích cực hay tiêu cực đến sự biến động của hiện tượng, từ đó giúp ta hiểu được đúng đắn nguyên nhân làm cho hiện tượng phát triển.

Để thể hiện được vai trò và ảnh hưởng của từng nhân tố khác nhau đến sự biến động của hiện tượng nghiên cứu, người ta sử dụng hệ thống chỉ số, hệ thống chỉ số được xây dựng trên cơ sở các phương trình kinh tế.

Ví dụ: từ các phương trình kinh tế đã nêu ở trên, ta có thể xây dựng các hệ thống chỉ số sau:

- Hệ thống chỉ số phân tích sự biến động của mức tiêu thụ theo hai nhân tố ảnh hưởng là : giá cả và lượng hàng hóa tiêu thụ.

Số tương đối: $I_{pq} = I_p \times I_q$

Hay:
$$\frac{\sum p_1 q_1}{\sum p_0 q_0} = \frac{\sum p_1 q_1}{\sum p_0 q_1} \times \frac{\sum p_0 q_1}{\sum p_0 q_0}$$

Chênh lệch tuyệt đối:

$$\sum p_1 q_1 - \sum p_0 q_0 = (\sum p_1 q_1 - \sum p_0 q_1) + (\sum p_0 q_1 - \sum p_0 q_0)$$

- Hệ thống chỉ số phân tích sự biến động của tổng giá thành sản phẩm theo hai nhân tố ảnh hưởng là: giá thành đơn vị và lượng sản phẩm sản xuất.

Số tương đối: $I_{Zq} = I_Z \times I_q$

Hay:
$$\frac{\sum Z_1 q_1}{\sum Z_0 q_0} = \frac{\sum Z_1 q_1}{\sum Z_0 q_1} \times \frac{\sum Z_0 q_1}{\sum Z_0 q_0}$$

Chênh lệch tuyệt đối:

$$\sum Z_1 q_1 - \sum Z_0 q_0 = (\sum Z_1 q_1 - \sum Z_0 q_1) + (\sum Z_0 q_1 - \sum Z_0 q_0)$$

- Hệ thống chỉ số phân tích sự biến động của tổng sản lượng theo hai nhân tố ảnh hưởng là: năng suất lao động và số công nhân.

Số tương đối: $I_{WT} = I_W \times I_T$

Hay:
$$\frac{\sum W_1 T_1}{\sum W_0 T_0} = \frac{\sum W_1 T_1}{\sum W_0 T_1} \times \frac{\sum W_0 T_1}{\sum W_0 T_0}$$

Chênh lệch tuyệt đối:

$$\sum W_1 T_1 - \sum W_0 T_0 = (\sum W_1 T_1 - \sum W_0 T_1) + (\sum W_0 T_1 - \sum W_0 T_0)$$

Ví dụ: lấy số liệu của bảng 8-1 thay vào hệ thống chỉ số phân tích mức tiêu thụ, ta có:

Số tương đối:

$$\frac{18.950}{18.000} = \frac{18.950}{21.260} \times \frac{21.260}{18.000}$$

$$1,053 = 0,891 \times 1,181$$

Hay $105,3\% = 89,1\% \times 118,1\%$

$$(+5,3\%) \quad (-10,9\%) \quad (+18,1\%)$$

Chênh lệch tuyệt đối:

$$18.950 - 18.000 = (18.950 - 21.260) + (21.260 - 18.000)$$

$$+950 = -2.310 + 3.260 \text{ (nghìn đồng)}$$

Kết quả trên cho thấy:

Mức tiêu thụ kỳ báo cáo so với kỳ gốc tăng 5,3% hay tăng 950 nghìn đồng do ảnh hưởng của hai nhân tố là:

- Do giá cả các loại hàng hóa giảm 10,9% nên đã làm cho mức tiêu thụ giảm 2.310 nghìn đồng.

- Do lượng hàng hóa tiêu thụ tăng 18,1% nên đã làm cho mức tiêu thụ tăng 3.260 nghìn đồng.

IV. VẬN DỤNG PHƯƠNG PHÁP CHỈ SỐ ĐỂ PHÂN TÍCH CHỈ TIÊU BÌNH QUÂN VÀ CHỈ TIÊU TỔNG LƯỢNG BIẾN TIÊU THỨC

4.1. Phân tích biến động của chỉ tiêu bình quân

Chỉ tiêu bình quân biến động do ảnh hưởng của hai nhân tố là: tiêu thức nghiên cứu $\bar{x}$ và kết cấu tổng thể ($f/\Sigma f$). Phân tích và đánh giá vai trò ảnh hưởng của từng nhân tố đến sự biến động chung của chỉ tiêu bình quân sẽ giúp ta đánh giá đúng đắn chất lượng công tác của đơn vị. Để giải quyết nhiệm vụ này chúng ta sử dụng phương pháp chỉ số.

Số tương đối: $I_{\bar{x}} = I_x \times I_{F/\Sigma F}$

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

Viết gọn lại:
$$\frac{\bar{x}_1}{\bar{x}_0} = \frac{\bar{x}_1}{\bar{x}_{01}} \times \frac{\bar{x}_{01}}{\bar{x}_0}$$

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

Chênh lệch tuyệt đối: $\bar{x}_1 - \bar{x}_0 = (\bar{x}_1 - \bar{x}_{01}) + (\bar{x}_{01} - \bar{x}_0)$

Trong hệ thống chỉ số trên:

I_x - chỉ số cấu thành khả biến, nêu lên sự biến động của chỉ tiêu bình quân giữa kỳ báo cáo với kỳ gốc.

I_x - chỉ số cấu thành cố định, nêu lên sự biến động của chỉ tiêu bình quân do ảnh hưởng của sự biến động của tiêu thức nghiên cứu (tiêu thức được bình quân hóa).

$I_{f/\Sigma f}$ - chỉ số ảnh hưởng kết cấu, nêu lên sự biến động của chỉ tiêu bình quân do ảnh hưởng của sự thay đổi của kết cấu tổng thể.

Ví dụ: có tài liệu về lao động và tiền lương của công nhân tại hai phân xưởng sản xuất tại một xí nghiệp như sau:

Bảng 8-3

Phân xưởng	Kỳ gốc		Kỳ báo cáo	
	Tiền lương 1 CN (1000đ)	Số công nhân (người)	Tiền lương 1 CN (1000đ)	Số công nhân (người)
I	520	140	640	120
II	400	110	480	80

Từ số liệu bảng 8-3 ta tính được:

- Tiền lương bình quân 1 công nhân trong doanh nghiệp ở kỳ báo cáo:

$$\bar{x}_1 = \frac{640 \times 120 + 480 \times 80}{120 + 80} = 576 \text{ (nghìn đồng)}$$

- Tiền lương bình quân 1 công nhân trong doanh nghiệp ở kỳ gốc:

$$\bar{x}_0 = \frac{520 \times 140 + 400 \times 110}{140 + 110} = 467,2 \text{ (nghìn đồng)}$$

- Tiền lương bình quân 1 công nhân trong doanh nghiệp ở kỳ gốc tính theo kết cấu công nhân kỳ báo cáo

$$\bar{x}_{(1)} = \frac{520 \times 120 + 400 \times 80}{120 + 80} = 472 \text{ (nghìn đồng)}$$

Thay số liệu vào hệ thống chỉ số ta có:

Số tương đối:

$$\frac{576}{467,2} = \frac{576}{472} \times \frac{472}{467,2}$$

$$1,232 = 1,220 \times 1,010$$

hay: $123,2\% = 122\% \times 101\%$
 $(+23,2\%) (+22\%) (+1\%)$

Chênh lệch tuyệt đối:

$$576 - 467,2 = (576 - 472) + (472 - 467,2)$$

$$+108,8 = +104 + 4,8 \text{ (nghìn đồng)}$$

Kết quả trên cho thấy:

Tiền lương bình quân 1 công nhân trong doanh nghiệp kỳ báo cáo so với kỳ gốc tăng 23,2% hay tăng 108,8 nghìn đồng, do ảnh hưởng của hai nhân tố:

- Bản thân tiền lương của công nhân ở các phân xưởng tăng làm cho tiền lương bình quân 1 công nhân của doanh nghiệp tăng 22% tương ứng tăng 104 nghìn đồng.

- Kết cấu số lượng công nhân thay đổi làm cho tiền lương bình quân 1 công nhân của doanh nghiệp tăng 1% tương ứng tăng 4,8 nghìn đồng.

Như vậy, nếu vẫn giữ nguyên kết cấu công nhân như ở kỳ gốc, thì tiền lương bình quân 1 công nhân trong doanh nghiệp thực chất chỉ tăng 22% hay tăng 104 nghìn đồng.

4.2. Phân tích sự biến động của chỉ tiêu tổng lượng biến tiêu thức

Sau khi phân tích sự biến động của chỉ tiêu bình quân, ta có thể phân tích sự biến động của chỉ tiêu tổng lượng mà trong đó chỉ tiêu bình quân là một nhân tố ảnh hưởng đến sự biến động của chỉ tiêu tổng lượng.

Chỉ tiêu tổng lượng M được xác định bởi phương trình kinh tế:

$$M = \bar{x} \times n$$

Ví dụ:

$$\begin{array}{l} \text{Tổng sản lượng} \\ \text{sản phẩm} \end{array} = \begin{array}{l} \text{Năng suất lao động} \\ \text{bình quân 1 công nhân} \end{array} \times \begin{array}{l} \text{Số lượng} \\ \text{công nhân} \end{array}$$

$$\begin{array}{l} \text{Tổng giá thành} \\ \text{sản phẩm} \end{array} = \begin{array}{l} \text{Giá thành bình quân} \\ \text{1 đơn vị sản phẩm} \end{array} \times \begin{array}{l} \text{Số lượng} \\ \text{sản phẩm sản xuất} \end{array}$$

$$\begin{array}{ccccc} \text{Tổng quỹ lương} & = & \text{Tiền lương bình quân} & \times & \text{Số} \\ \text{công nhân} & & \text{l công nhân} & & \text{công nhân} \\ \text{v.v...} & & & & \end{array}$$

Từ các phương trình kinh tế trên, ta có thể dùng hệ thống chỉ số để phân tích ảnh hưởng của các nhân tố đến chỉ tiêu tổng lượng:

Số tương đối: $I_M = I_{\bar{x}} \times I_n$

hay: $\frac{\Sigma M_1}{\Sigma M_0} = \frac{\bar{x}_1}{\bar{x}_0} \times \frac{\Sigma n_1}{\Sigma n_0}$

Chênh lệch tuyệt đối: $\Sigma M_1 - \Sigma M_0 = (\bar{x}_1 - \bar{x}_0) \Sigma n_1 + (\Sigma n_1 - \Sigma n_0) \bar{x}_0$

Ví dụ: theo số liệu của bảng 8-3, ta tiến hành phân tích sự biến động của chỉ tiêu tổng lượng là tổng quỹ lương của công nhân theo hai nhân tố ảnh hưởng là: tiền lương bình quân l công nhân và tổng số công nhân.

$$\Sigma M_1 = 576 \times 200 = 115.200 \text{ (nghìn đồng)}$$

$$\Sigma M_0 = 467,2 \times 250 = 116.800 \text{ (nghìn đồng)}$$

Thay số liệu vào hệ thống chỉ số ta có:

Số tương đối:

$$\frac{115.200}{116.800} = \frac{576}{467,2} \times \frac{200}{250}$$

$$0,986 = 1,232 \times 0,8$$

hay: $98,6\% = 123,2\% \times 80\%$
 $(-1,4\%) \quad (+23,2\%) \quad (-20\%)$

Chênh lệch tuyệt đối:

$$\begin{array}{rcl} 115.200 - 116.800 & = & (576 - 467,2) \times 200 + (200 - 250) \times 467,2 \\ -1.600 & = & +21.760 + (-23.360) \text{ (nghìn đồng)} \end{array}$$

Kết quả cho thấy:

Tổng quỹ lương kỳ báo cáo so với kỳ gốc giảm 1,4% tương ứng giảm 1.600 nghìn đồng, do ảnh hưởng của hai nhân tố là:

- Do tăng tiền lương bình quân l công nhân, nên đã làm cho tổng quỹ lương tăng 23,2% tương ứng tăng 21.760 nghìn đồng.

- Do giảm số công nhân nên đã làm cho tổng quỹ lương giảm 20% tương ứng giảm 23.360 nghìn đồng.

TÀI LIỆU THAM KHẢO

1. *Giáo trình Lý thuyết thống kê*. Nguyễn Hữu Hòe. NXB Thống kê. 1984.
2. *Giáo trình Lý thuyết thống kê*. Trường Đại học Tài chính kế toán Hà Nội. 1992.
3. *Giáo trình Lý thuyết thống kê*. Lê Hồng (chủ biên). Trường Đại học Tài chính kế toán Thành phố Hồ Chí Minh. 1993.
4. *Lý thuyết thống kê*. Trần Bá Nhẫn. Trường Đại học Kinh tế Thành phố Hồ Chí Minh. 1998.
5. *Giáo trình Thống kê doanh nghiệp*. Trường Cao đẳng Xây dựng số 1. NXB Xây dựng. 2001.

MỤC LỤC

	Trang
Lời nói đầu	3
Chương I: Đối tượng nghiên cứu của thống kê học	
I. Đối tượng nghiên cứu của thống kê học	5
II. Một số khái niệm thường dùng trong thống kê học	7
Chương II: Quá trình nghiên cứu thống kê	
I. Xác định hệ thống chỉ tiêu thống kê	9
II. Điều tra thống kê	10
III. Tổng hợp thống kê	16
IV. Phân tích và dự đoán thống kê	18
Chương III: Điều tra chọn mẫu	
I. Khái niệm, ý nghĩa của điều tra chọn mẫu	22
II. Những vấn đề lý luận về điều tra chọn mẫu	22
III. Suy rộng kết quả điều tra chọn mẫu	28
Chương IV: Phân tổ thống kê	
I. Khái niệm, ý nghĩa, nhiệm vụ phân tổ thống kê	30
II. Tiêu thức phân tổ	32
III. Phân tổ thống kê	32
IV. Chỉ tiêu giải thích	37
V. Dây số phân phối	37
Chương V: Các mức độ của hiện tượng kinh tế-xã hội	
I. Số tuyệt đối trong thống kê	39
II. Số tương đối trong thống kê	41
III. Số bình quân trong thống kê	44
IV. Các chỉ tiêu đánh giá độ biến thiên của tiêu thức	53
Chương VI. Hồi quy và tương quan	
I. Phương pháp hồi quy và tương quan	57
II. Liên hệ tương quan tuyến tính giữa hai tiêu thức số lượng	58
III. Liên hệ tương quan phi tuyến giữa hai tiêu thức số lượng	61
IV. Liên hệ tương quan giữa hai tiêu thức "thay phiên"	66
	91

Chương VII. Dãy số biến động theo thời gian

- | | |
|--|----|
| I. Khái niệm, phân loại và ý nghĩa của dãy số biến động theo thời gian | 68 |
| II. Các chỉ tiêu phân tích dãy số biến động theo thời gian | 69 |
| III. Các phương pháp biểu hiện xu hướng phát triển cơ bản của hiện tượng | 73 |

Chương VIII: Chỉ số thống kê

- | | |
|---|----|
| I. Khái niệm, đặc điểm và tác dụng của chỉ số | 78 |
| II. Phương pháp tính chỉ số | 80 |
| III. Hệ thống chỉ số | 84 |
| VI. Vận dụng phương pháp chỉ số để phân tích chỉ tiêu bình quân và chỉ tiêu tổng lượng biến tiêu thức | 86 |

- | | |
|---------------------------|----|
| Tài liệu tham khảo | 90 |
|---------------------------|----|

GIÁO TRÌNH LÝ THUYẾT THỐNG KÊ

Chịu trách nhiệm xuất bản:

BÙI HỮU HẠNH

<i>Biên tập:</i>	NGUYỄN THỊ MINH KHÔI
<i>Chế bản:</i>	ĐINH THỊ PHƯỢNG
<i>Sửa bản in:</i>	NGUYỄN THỊ MINH KHÔI
<i>Trình bày bìa:</i>	Hs. NGUYỄN HỮU TÙNG

In 800 cuốn khổ 19×27 cm tại Xưởng in Nhà Xuất bản Xây dựng. Giấy chấp nhận đăng ký kế hoạch xuất bản số 810/XB-QLXB-9, ngày 15/6/2004. In xong nộp lưu chiểu tháng 9/2004.

#253 283

3-31	810 - 2004
XD- 2004	

Giá : 15.000^d