

HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG
KHOA QUẢN TRỊ KINH DOANH 1
BỘ MÔN KINH TẾ

BÀI GIẢNG
NGUYÊN LÝ THỐNG KÊ KINH TẾ

Người biên soạn : TS. Vũ Trọng Phong

Hà nội - 2017

MỤC LỤC	Trang
Lời nói đầu	4
Chương 1: Những vấn đề chung về thống kê học	5
1.1 Đối tượng của thống kê học	5
1.2 Một số khái niệm trong thống kê	6
1.3 Thang đo trong thống kê	9
1.4 Hoạt động thống kê và quá trình nghiên cứu thống kê	11
Chương 2: Điều tra thống kê kinh tế	15
2.1 Khái niệm ý nghĩa của điều tra thống kê kinh tế	15
2.2 Các loại điều tra thống kê kinh tế	17
2.3 Các phương pháp thu thập thông tin trong điều tra thống kê kinh tế	19
2.4 Xây dựng phương án trong điều tra thống kê kinh tế	21
2.5 Xây dựng bảng hỏi trong điều tra thống kê	25
2.6 Sai số trong điều tra thống kê kinh tế	31
Chương 3: Phân tổ tổng hợp tài liệu điều tra thống kê kinh tế	33
3.1. Khái niệm chung về phân tổ thống kê	33
3.2. Tiêu thức phân tổ thống kê	38
3.3. Kỹ thuật trình bày kết quả tổng hợp tài liệu điều tra qua phân tổ thống kê	41
Chương 4. Nghiên cứu thống kê các mức độ của hiện tượng kinh tế xã hội	47
4.1 Số tuyệt đối trong thống kê	47
4.2 Số tương đối trong thống kê	47
4.3 Số bình quân trong thống kê	50
4.4 Các chỉ tiêu đo độ biến thiên của tiêu thức	59
4.5 Phân phối trong thống kê	64
4.6 Các chỉ tiêu biểu thị hình dáng của phân phối	69
Chương 5: Điều tra chọn mẫu trong thống kê	73
5.1 Khái niệm về điều tra chọn mẫu	73
5.2 Điều tra chọn mẫu ngẫu nhiên	74
5.3. Các phương pháp điều tra chọn mẫu ngẫu nhiên	82
5.4 Quy trình một cuộc điều tra chọn mẫu ngẫu nhiên	85
5.5 Điều tra chọn mẫu phi ngẫu nhiên	86
Chương 6: Phân tích tương quan và hồi quy	90
6.1 Nhiệm vụ của phân tích hồi quy và tương quan	90
6.2 Hồi quy tương quan tuyến tính giữa hai tiêu thức số lượng	91
6.3 Hồi quy tương quan phi tuyến tính giữa hai tiêu thức số lượng	93
6.4 Hồi quy tương quan tuyến tính bội	96
6.5. Tương quan hạng và tương quan giữa hai tiêu thức thuộc tính	99

Chương 7: Phân tích dãy số thời gian	101
7.1 Khái niệm, phân loại và ý nghĩa dãy số thời gian	101
7.2 Phân tích đặc điểm biến động của hiện tượng qua thời gian	102
7.3 Phân tích các thành phần của dãy số thời gian	107
Chương 8: Chỉ số phát triển và chỉ số kế hoạch kinh tế	110
8.1 Khái niệm, ý nghĩa và phân loại chỉ số	110
8.2 Chỉ số phát triển	111
8.3 Chỉ số kế hoạch	117
8.4 Hệ thống chỉ số	118
Chương 9: Thống kê xu hướng phát triển và dự báo thống kê kinh tế	124
9.1. Thống kê xu hướng phát triển kinh tế	124
9.2. Dự báo thống kê kinh tế	128

PTIT

LỜI MỞ ĐẦU

Nguyên lý thống kê kinh tế là môn học cơ sở của sinh viên chuyên ngành khối kinh tế và tài chính kế toán. Để đáp ứng yêu cầu giảng dạy và học tập của giáo

viên và sinh viên, chúng tôi tổ chức biên soạn bài giảng "Nguyên lý thống kê kinh tế". Với kinh nghiệm giảng, cộng với sự nỗ lực nghiên cứu từ các nguồn tài liệu khác nhau, bài giảng có nhiều thay đổi bổ sung để đáp ứng yêu cầu đào tạo trong giai đoạn mới.

Bài giảng gồm 9 chương đề cập đến toàn bộ những kiến thức về nguyên lý thống kê kinh tế.

Tác giả chân thành cảm ơn các đồng nghiệp đã giúp đỡ trong quá trình biên soạn bài giảng này. Tác giả mong muốn nhận được góp ý của đồng nghiệp, anh chị em sinh viên và bạn đọc để tiếp tục hoàn thiện nhằm nâng cao hơn nữa chất lượng bài giảng.

Hà Nội tháng 11 năm 2017

TÁC GIẢ

PTIT

CHƯƠNG 1

NHỮNG VẤN ĐỀ CHUNG VỀ THỐNG KÊ HỌC

1.1 Đối tượng nghiên cứu của thống kê học

Thống kê học là một môn khoa học xã hội. Tuy nhiên, khác với các môn học xã hội khác, thống kê học không trực tiếp nghiên cứu mặt chất của hiện tượng mà nó chỉ phản ánh bản chất, tính quy luật của hiện tượng thông qua các con số, các biểu hiện về lượng của hiện

tượng. Điều đó có nghĩa là thống kê học phải sử dụng các con số về quy mô, kết cấu quan hệ tỉ lệ, quan hệ so sánh, trình độ phát triển, trình độ phổ biến,... của hiện tượng để phản ánh, biểu thị bản chất, tính quy luật của hiện tượng nghiên cứu trong những điều kiện, hoàn cảnh cụ thể. Như vậy, các con số thống kê không phải chung chung, trừu tượng mà bao giờ cũng chứa đựng một nội dung kinh tế, chính trị, xã hội nhất định, giúp chúng ta nhận thức được bản chất và quy luật của hiện tượng nghiên cứu.

Theo quan điểm triết học, chất và lượng là hai mặt không thể tách rời của mọi sự vật, hiện tượng, giữa chúng luôn tồn tại mối liên hệ biện chứng với nhau. Trong mối quan hệ đó, sự thay đổi về lượng quyết định sự biến đổi về chất. Quy luật lượng - chất của triết học đã chỉ rõ: mỗi lượng cụ thể đều gắn với một chất nhất định, khi lượng thay đổi và tích lũy đến một chừng mực nhất định thì chất thay đổi theo. Vì vậy, nghiên cứu mặt lượng của hiện tượng sẽ giúp cho việc nhận thức bản chất của hiện tượng. Có thể đánh giá thành tích sản xuất của một doanh nghiệp qua các con số thống kê về tổng số sản phẩm làm ra, giá trị sản xuất đạt được, tỉ lệ hoàn thành kế hoạch sản xuất, giá thành đơn vị sản phẩm, năng suất lao động và thu nhập của công nhân.

Tuy nhiên, để có thể phản ánh được bản chất và quy luật phát triển của hiện tượng, các con số thống kê phải được tập hợp, thu thập trên một số lớn các hiện tượng cá biệt. Thống kê học coi tổng thể các hiện tượng cá biệt như một thể hoàn chỉnh và lấy đó làm đối tượng nghiên cứu. Mặt lượng của hiện tượng cá biệt thường chịu tác động của nhiều nhân tố, trong đó có những nhân tố tất nhiên và ngẫu nhiên. Mức độ và chiều hướng tác động của từng nhân tố này trên mỗi hiện tượng cá biệt là rất khác nhau. Nếu chỉ thu thập số liệu trên một số ít hiện tượng thì khó có thể rút ra bản chất chung của hiện tượng, mà nhiều khi người ta chỉ tìm thấy những yếu tố ngẫu nhiên, không bản chất. Ngược lại, khi nghiên cứu trên một số lớn các hiện tượng cá biệt, các yếu tố ngẫu nhiên sẽ bù trừ, triệt tiêu nhau và khi đó, bản chất, quy luật phát triển của hiện tượng mới được bộc lộ rõ.

Hiện tượng số lớn trong thống kê được hiểu là một tập hợp các hiện tượng cá biệt đủ bù trừ, triệt tiêu tác động của các yếu tố ngẫu nhiên. Giữa hiện tượng số lớn (tổng thể) và các hiện tượng cá biệt (đơn vị tổng thể) luôn tồn tại mối quan hệ biện chứng. Muốn nghiên cứu tổng thể, phải dựa trên cơ sở nghiên cứu từng đơn vị tổng thể. Mặt khác, trong quá trình phát triển không ngừng của xã hội, luôn nảy sinh những hiện tượng cá biệt mới, những điển hình tiên tiến hoặc lạc hậu. Sự nghiên cứu các hiện tượng cá biệt này sẽ giúp cho sự nhận thức bản chất của hiện tượng đầy đủ, toàn diện và sâu sắc hơn. Vì vậy trong thống kê, nhất là thống kê kinh tế, người ta thường kết hợp nghiên cứu hiện tượng số lớn với việc nghiên cứu hiện tượng cá biệt.

Đối tượng nghiên cứu của thống kê học bao giờ cũng tồn tại trong điều kiện thời gian và không gian cụ thể. Trong điều kiện lịch sử khác nhau, các đặc điểm về chất và biểu hiện về lượng của hiện tượng cũng khác nhau nhất là đối với các hiện tượng kinh tế - xã hội. Chẳng hạn, trình độ hiện đại hóa, một trong những nhân tố ảnh hưởng trực tiếp đến năng suất lao động của người công nhân, lại rất khác nhau giữa các doanh nghiệp. Ngay trong cùng một đơn vị, cũng lại có thể khác nhau giữa các giai đoạn, các thời kỳ,... Thậm chí, giữa các bộ phận trong cùng một đơn vị, nhiều khi cũng tồn tại những khác biệt đáng kể. Vì vậy các con số về năng suất lao động của người công nhân trong từng doanh nghiệp, từng thời kỳ khác nhau cũng có ý nghĩa khác nhau. Như vậy, khi sử dụng các số liệu thống kê phải luôn gắn nó với điều kiện thời gian và không gian cụ thể của hiện tượng mà số liệu phản ánh.

1.2 Một số khái niệm thống kê

1.2.1 Khái niệm thống kê

Thuật ngữ thống kê có thể hiểu theo hai nghĩa:

- **Thứ nhất**, thống kê là các số liệu được thu thập để phản ánh các hiện tượng tự nhiên, kỹ thuật, kinh tế và xã hội. Chẳng hạn như sản lượng sản phẩm được sản xuất ra của một doanh nghiệp nào đó, mức nước trên một dòng sông ở các tháng trong năm...

- **Thứ hai**, thống kê được hiểu là hệ thống các phương pháp ghi chép, thu thập và phân tích các con số về những hiện tượng tự nhiên, kỹ thuật, kinh tế và xã hội để tìm hiểu bản chất và tìm quy luật vốn có của những hiện tượng ấy.

Công việc của một nhà thống kê bao gồm các hoạt động trên một phạm vi rộng có thể tóm tắt thành những mục lớn như sau:

- Thu thập và xử lý dữ liệu
- Điều tra chọn mẫu
- Nghiên cứu mối liên hệ giữa các hiện tượng.
- Dự đoán
- Nghiên cứu các hiện tượng trong các hoàn cảnh không chắc chắn
- Ra quyết định trong điều kiện không chắc chắn.

Một cách tổng quát thống kê học chính là *khoa học nghiên cứu hệ thống các phương pháp thu thập, xử lý và phân tích các con số (mặt lượng) của những hiện tượng số lớn để tìm hiểu bản chất và tính quy luật vốn có của chúng (mặt chất) trong những điều kiện, địa điểm và thời gian cụ thể.*

Thống kê được chia thành hai lĩnh vực:

- Thống kê mô tả: bao gồm các phương pháp thu thập số liệu, mô tả và trình bày số liệu, tính toán các đặc trưng đo lường.
- Thống kê suy diễn: bao gồm các phương pháp ước lượng, kiểm định phân tích mối quan hệ, dự đoán trên cơ sở các thông tin thu thập từ mẫu.

1.2.2 Tổng thể thống kê và đơn vị tổng thể thống kê

Thống kê nghiên cứu một lượng của hiện tượng kinh tế xã hội số lớn phải xác định phạm vi hiện tượng được nghiên cứu cụ thể. Để chỉ đối tượng nghiên cứu cụ thể, người ta dùng khái niệm tổng thể. Tổng thể thống kê là hiện tượng kinh tế xã hội số lớn, gồm những đơn vị (hoặc phân tử, hiện tượng) cá biệt cần được quan sát, phân tích mặt lượng của chúng để tìm hiểu bản chất và tính quy luật vốn có của chúng (mặt chất) trong những điều kiện, địa điểm và thời gian cụ thể.

Có trường hợp các đơn vị cấu thành tổng thể, có thể thấy được bằng trực quan. Tổng thể bao gồm các đơn vị như vậy được gọi là tổng thể bộc lộ như các tổng thể nêu trên. Tổng thể các đơn vị cấu thành nó, không thể nhận biết được bằng trực quan là tổng thể tiềm ẩn như tổng thể những cán bộ công nhân viên ưa chuộng nghệ thuật sân khấu, tổng thể những người mê tín dị đoan, tổng thể những người trung thành với Tổ quốc ...

Các đơn vị tổng thể có thể giống nhau trên một số đặc điểm, các đặc điểm còn lại khác nhau. Do đó, tùy theo mục đích nghiên cứu mà phân biệt tổng thể đồng chất hay không đồng chất. Tổng thể đồng chất bao gồm các đơn vị giống nhau về một số đặc điểm chủ yếu có liên quan tới mục đích nghiên cứu tổng thể không đồng chất bao gồm các đơn vị khác nhau về các đặc điểm, các loại hình. Tổng thể bao gồm tất cả các đơn vị thuộc phạm vi nghiên cứu tổng thể chung, chỉ bao gồm một bộ phận đơn vị trong đó là tổng thể bộ phận.

Xác định tổng thể để đáp ứng mục đích nghiên cứu thống kê. Phải trên cơ sở phân tích lý luận kinh tế, chính trị hoặc xã hội, định nghĩa rõ tổng thể. Định nghĩa tổng thể không những phải giới hạn về thực thể (tổng thể là gì) mà còn phải giới hạn về thời gian và không gian (tổng thể tồn tại vào thời gian nào, ở đâu). Xác định tổng thể chính xác không dễ dàng. Vì có những hiện tượng có thể tương tự về hình thức, nhưng lại khác hẳn về nội dung. Chính vì vậy phải phân tích lý luận để thấy rõ nội dung của hiện tượng. Xác định tổng thể thống kê không chính xác sẽ lãng phí sức người và tiền của trong nghiên cứu, không đủ cơ sở để hiểu đúng bản chất cụ thể của hiện tượng.

Đúng nghĩa tổng thể làm rõ đặc trưng cơ bản chung của hiện tượng kinh tế xã hội số lớn phù hợp với mục đích nghiên cứu. Thông qua việc phân tích lý luận và thực tế phải làm rõ tổng thể gồm những hiện tượng (phần tử) cá biệt nào. Hiện tượng cá biệt này là đơn vị tổng thể. Tất cả các đơn vị tổng thể chỉ giống nhau trên một số mặt, còn các mặt khác không giống nhau. Cho nên trong thực tế phải nêu rõ ràng những hiện tượng cá biệt nào được kể là đơn vị tổng thể. Trong những trường hợp khó khăn cho việc giới hạn, người ta phải lập một danh mục các đơn vị hoặc trong giải thích cần xác định rõ phạm vi nào của các đơn vị thuộc tổng thể.

Đơn vị tổng thể bao giờ cũng có đơn vị tính toán phù hợp. Xác định đơn vị tổng thể là việc cụ thể hóa tổng thể. Đơn vị tổng thể là xuất phát điểm của quá trình nghiên cứu thống kê. Vì nó có mặt lượng mà ta cần nghiên cứu. Cho nên xác định đơn vị tổng thể cũng quan trọng như xác định tổng thể.

1.2.3 Tổng thể mẫu và quan sát

Tổng thể mẫu là tổng thể bao gồm một số đơn vị được chọn ra từ tổng thể chung theo một phương pháp lấy mẫu nào đó. Các đặc trưng mẫu được suy rộng ra các đặc trưng của tổng thể chung.

Quan sát là cơ sở để thu thập số liệu và thông tin cần nghiên cứu. Chẳng hạn trong điều tra chọn mẫu, mỗi đơn vị mẫu sẽ được tiến hành ghi chép, thu thập thông tin và được gọi là một quan sát.

1.2.4 Tiêu thức thống kê

Nghiên cứu thống kê phải dựa vào các đặc điểm của đơn vị tổng thể. Đơn vị tổng thể có nhiều đặc điểm. Tùy theo mục đích nghiên cứu, một số đặc điểm của đơn vị tổng thể được chọn ra để nghiên cứu. Các đặc điểm này được gọi là các tiêu thức. Ví dụ mỗi người lao động của doanh nghiệp có các tiêu thức: tên, tuổi, giới tính, trình độ văn hóa, nghề nghiệp, nơi ở ... Mỗi doanh nghiệp trong tổng thể có tiêu thức: tên doanh nghiệp, địa chỉ, số lượng cán bộ công nhân viên... Đơn vị tổng thể được làm rõ đặc trưng của nó qua các tiêu thức: thực thể, thời gian và không gian.

1. Tiêu thức thực thể

Nêu lên bản chất của đơn vị tổng thể, nó biến đổi trong bản chất này. Các tiêu thức: giới tính, tuổi, trình độ văn hóa, nghề nghiệp, số lượng cán bộ công nhân viên chức là các tiêu thức thực thể. Theo nội dung của nó, tiêu thức thực thể gồm hai loại: thuộc tính và số lượng.

Tiêu thức thuộc tính không có biểu hiện trực tiếp là các con số, như tiêu thức giới tính, trình độ văn hóa, nghề nghiệp ... tiêu thức thuộc tính có biểu hiện trực tiếp và gián tiếp, như giới tính có biểu hiện trực tiếp là nam và nữ. Tiêu thức đời sống vật chất có biểu hiện gián tiếp là lương tiêu dùng lương thực, thực phẩm theo đầu người, diện tích nhà ở theo đầu người. Các biểu hiện gián tiếp của tiêu thức thuộc tính còn được gọi là các chỉ báo thống kê.

Tiêu thức thuộc tính không có biểu hiện trực tiếp là con số, nên còn được gọi là tiêu thức phi lượng hóa.

Tiêu thức số lượng có biểu hiện trực tiếp là con số (gọi là lượng biến). Nó là kết quả của quá trình quan sát (cân, đo, đong, đếm) như tuổi đời, tuổi nghề, số lượng điện thoại, số bưu cục ... Tiêu thức số lượng còn gọi là tiêu thức lượng hóa vì nó có biểu hiện trực tiếp là con số.

Tiêu thức thực thể khi chỉ có hai biểu hiện không trùng nhau trên một đơn vị tổng thể, được gọi là tiêu thức thay phiên, như giới tính (nam và nữ), chất lượng sản phẩm dịch vụ (tốt và xấu). Tiêu thức thực thể có ba loại biểu hiện trở lên có thể trở thành tiêu thức thay phiên, như số lượng cán bộ công nhân viên nêu trên có nhiều biểu hiện nhưng rút gọn thành hai biểu hiện, hoặc các biểu hiện của tiêu thức trình độ văn hóa có thể rút gọn thành hai biểu hiện: chưa tốt nghiệp phổ thông trung học và tốt nghiệp phổ thông trung học trở lên. Những trường hợp này được tiến hành khi người ta chỉ quan tâm đến một biểu hiện nào đó xuất hiện hay không xuất hiện trên đơn vị tổng thể. Tiêu thức thực thể có phù hợp nhiều hay ít với việc đáp ứng mục đích nghiên cứu là tùy thuộc vào việc chọn những tiêu thức nào cho nghiên cứu.

2. Tiêu thức thời gian

Nêu hiện tượng kinh tế xã hội theo sự xuất hiện của nó vào thời gian nào. Những biểu hiện của tiêu thức thời gian là phút, giờ, ngày, tháng, năm. Thời hạn có giá trị của các chỉ dẫn về đối tượng nghiên cứu và những đơn vị tổng thể, về sự phân phối chúng trong một thời gian cũng như về sự thay đổi từ thời kỳ này tới thời kỳ khác được khẳng định qua tiêu thức thời gian. Ví dụ tổng số tài sản cố định của doanh nghiệp A tại thời điểm 1/1/2010 là 20.

3. Tiêu thức không gian:

Nêu phạm vi lãnh thổ bao trùm của đối tượng nghiên cứu và sự xuất hiện theo địa điểm của các đơn vị tổng thể. Những biểu hiện của nó được chỉ ra nhờ sự phân định về mặt quản lý hành chính hoặc theo điều kiện tự nhiên, phân vùng kinh tế ... Nghiên cứu thống kê theo tiêu thức không gian có ý nghĩa quan trọng trước hết là gắn với tiêu thức thực thể để quan sát sự phân phối về mặt lãnh thổ của các đơn vị tổng thể.

Các tiêu thức góp phần vào việc khẳng định đơn vị tổng thể cũng như tổng thể. Vì chúng nêu rõ các mặt và tính chất nhất định của đơn vị tổng thể cũng như của tổng thể. Nhờ đó có thể phân biệt đơn vị này với đơn vị khác cũng như tổng thể này với tổng thể kia.

1.2.5 Chỉ tiêu thống kê

Nghiên cứu thống kê không chỉ phản ánh lượng và chất của hiện tượng kinh tế xã hội cá biệt mà còn phản ánh và chất của hiện tượng kinh tế xã hội số lớn trong điều kiện thời gian và địa điểm cụ thể. Tính chất của các hiện tượng cá biệt được khái quát hóa trong chỉ tiêu thống kê. Do đó chỉ tiêu chỉ ra những mối quan hệ cần thiết, cái chung của tất cả các đơn vị hoặc của nhóm đơn vị. Ngoài ra chỉ tiêu còn phản ánh các mối quan hệ tồn tại khách quan, nhưng cũng không tự bộc lộ ra để hiểu trực tiếp là mối quan hệ. Phải điều tra mặt lượng của đơn vị cá biệt và từ đó phát hiện ý nghĩa theo số lượng của mối quan hệ bằng chỉ tiêu.

Chỉ tiêu thống kê có hai mặt: khái niệm và con số. Khái niệm của chỉ tiêu bao gồm định nghĩa và giới hạn về thực thể, thời gian và không gian của hiện tượng kinh tế xã hội. Mặt này chỉ rõ nội dung của chỉ tiêu thống kê. Con số của chỉ tiêu là trị số được phát hiện với đơn vị tính toán phù hợp. Nó nêu lên mức độ của chỉ tiêu. Theo nội dung, chỉ tiêu biểu hiện quy mô, cơ cấu, sự phát triển và mối quan hệ của hiện tượng số lớn trong điều kiện thời gian và địa điểm cụ thể.

Căn cứ vào nội dung có thể chia các chỉ tiêu thống kê thành hai loại: khối lượng và chất lượng. Chỉ tiêu khối lượng biểu hiện quy mô của tổng thể như số cán bộ công nhân viên, số máy điện thoại, khối lượng sản phẩm dịch vụ. Chỉ tiêu chất lượng biểu hiện trình độ phổ biến, mối quan hệ của tổng thể như giá thành sản phẩm dịch vụ. Việc phân loại này nhằm đáp ứng yêu cầu của một số phương pháp phân tích thống kê.

1.3 Thang đo trong thống kê

1.3.1 Thang đo định danh

Thang đo định danh (hay là đặt tên) là đánh số các biểu hiện cùng loại của một tiêu thức. Như giới tính biểu hiện “nam” được đánh số 1 và “nữ” đánh số 2. Giữa các con số ở đây không có quan hệ hơn, kém. Cho nên các phép tính với chúng đều vô nghĩa. Loại thang đo này dùng để đếm tần số của biểu hiện tiêu thức.

Ví dụ: Thang đo định danh trong câu hỏi phỏng vấn:

Tình trạng hôn nhân của ông (bà):

1. Có gia đình
2. Độc thân
3. Ly dị
4. Trường hợp khác

1.3.2 Thang đo thứ bậc

Thang đo thứ bậc cũng là thang đo định danh, nhưng giữa các biểu hiện tiêu thức có quan hệ thứ bậc, hơn, kém. Sự chênh lệch giữa các biểu hiện không nhất thiết phải bằng nhau, như huân chương có ba hạng: một, hai và ba. Hạng một hơn hạng hai, hạng hai hơn hạng ba. Trình độ văn hoá phổ thông có ba cấp: một, hai và ba. Cấp ba hơn cấp hai, cấp hai hơn cấp một. Con số có trị số lớn hơn không có nghĩa ở bậc cao hơn và ngược lại, mà do sự quy định. Thang đo loại này được sử dụng để tính toán đặc trưng chung của tổng thể một cách tương đối, trong một số trường hợp như tính cấp bậc bình quân của một doanh nghiệp, một đơn vị, bộ phận.

Ví dụ: Trong câu hỏi phỏng vấn:

Thu nhập của ông (bà) hàng tháng:

1. < 4 triệu đồng 2. Từ 4 đến 7 triệu đồng 3. > 7 triệu đồng

1.3.3 Thang đo khoảng

Thang đo khoảng là thang đo thứ bậc có các khoảng cách đều nhau. Có thể đánh giá sự khác biệt giữa các biểu hiện bằng thang đo loại này. Việc cộng trừ các con số có ý nghĩa, có thể tính các đặc trưng chung như số bình quân, phương sai. Yêu cầu có khoảng cách đều là đặt ra đối với thang đo, còn đối với biểu hiện của tiêu thức được đo không nhất thiết phải bằng nhau.

Ví dụ: Trong câu hỏi phỏng vấn:

Đề nghị ông (bà) cho ý kiến về tầm quan trọng của mục tiêu đào tạo cho sinh viên đại học sau đây bằng cách khoanh tròn các số tương ứng trên thang đánh giá chỉ mức độ từ 1 đến 5 (1 = không quan trọng; 5 = rất quan trọng).

	Không quan trọng	Bình thường			Rất quan trọng
Đạo đức	1	2	3	4	5
Khả năng tư duy	1	2	3	4	5
Năng lực giải quyết vấn đề	1	2	3	4	5
Kiến thức chuyên môn	1	2	3	4	5
Sức khỏe	1	2	3	4	5
Làm việc nhóm	1	2	3	4	5
Làm việc độc lập	1	2	3	4	5

1.3.4 Thang đo tỷ lệ

Thang đo tỷ lệ là thang đo khoảng với một điểm không (0) tuyệt đối (điểm gốc) để có thể so sánh được tỷ lệ giữa các trị số đo. Với thang đo loại này, có thể đo lường các biểu hiện của tiêu thức như các đơn vị đo lường vật lý thông thường (kg, m,...) và thực hiện được tất cả các phép tính với trị số đo.

1.4 Hoạt động thống kê và quá trình nghiên cứu thống kê

Mục đích cuối cùng của một cuộc nghiên cứu thống kê là thu thập những thông tin định lượng về hiện tượng nghiên cứu trong điều kiện lịch sử cụ thể, trên cơ sở đó phát hiện, bản chất quy luật phát triển của hiện tượng, giải quyết được một vấn đề lý thuyết hoặc một yêu cầu nhất định của một thực tiễn.

Ở Việt Nam, “hoạt động thống kê nhà nước là điều tra, báo cáo, tổng hợp phân tích và công bố các thông tin phản ánh bản chất và tính quy luật của các hiện tượng kinh tế - xã hội trong điều kiện không gian và thời gian cụ thể do cơ quan nhà nước tiến hành” và được điều chỉnh bởi Luật thống kê.

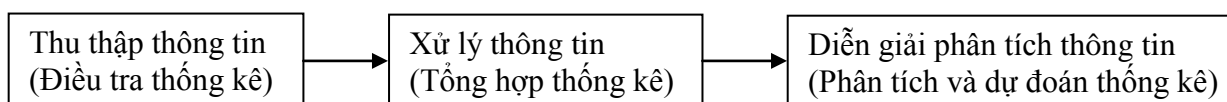
Hoạt động thống kê nhà nước Việt Nam phải theo 7 nguyên tắc cơ bản sau:

- 1) Đảm bảo tính trung thực, khách quan, chính xác, đầy đủ và kịp thời.
- 2) Đảm bảo tính độc lập về chuyên môn, nghiệp vụ thống kê.
- 3) Thống nhất về chỉ tiêu, biểu mẫu, phương pháp tính, bảng phân loại, đơn vị đo lường, niên độ thống kê và tính so sánh quốc tế.
- 4) Không trùng lặp, chồng chéo giữa các cuộc điều tra thống kê, các chế độ báo cáo thống kê.
- 5) Công khai về phương pháp thống kê, công bố thông tin thống kê.
- 6) Đảm bảo quyền bình đẳng trong việc tiếp cận và sử dụng thông tin thống kê nhà nước đã được công bố công khai.
- 7) Những thông tin thống kê về từng tổ chức, cá nhân chỉ được sử dụng cho mục đích tổng hợp thống kê.

Trong các nguyên tắc trên, nguyên tắc 1 là quan trọng nhất. Nó liên quan đến tất cả các hoạt động thống kê, từ điều tra, thực hiện chế độ báo cáo thống kê cơ sở, chế độ báo cáo thống kê tổng hợp, đến phân tích và công bố thông tin thống kê.

Ngoài ra, Luật Thống kê còn quy định 5 hành vi bị nghiêm cấm trong hoạt động thống kê, như: Không thực hiện hoặc cản trở việc thực hiện chế độ báo cáo, điều tra thống kê; tự thực hiện hoặc ép buộc người khác khai man, thông báo sai thông tin sai quy định, thực hiện các hoạt động thống kê trái pháp luật...

Các hoạt động thống kê đều phải trải qua một quá trình gồm nhiều giai đoạn, nhiều bước công việc kế tiếp nhau, có liên quan chặt chẽ với nhau. Có thể khái quát quá trình này bằng một sơ đồ hết sức đơn giản như sau:



Điều tra thống kê là việc tổ chức một cách khoa học với một kế hoạch thống nhất việc thu nhập, ghi chép nguồn tài liệu ban đầu về hiện tượng nghiên cứu trong điều kiện cụ thể về thời gian, không gian.

Như vậy, nhiệm vụ chủ yếu của giai đoạn này là thu thập được đầy đủ thông tin cần thiết cho việc nghiên cứu. Người ta thường gọi đây là những thông tin sơ cấp, hay nguồn tài liệu ban đầu. Đây là giai đoạn khởi đầu rất quan trọng. Có làm tốt giai đoạn khởi đầu rất quan trọng. Có làm tốt giai đoạn này các thông tin, số liệu mới đầy đủ và kịp thời, tạo điều kiện để

thực hiện tốt các giai đoạn tiếp theo. Để thực hiện nhiệm vụ của giai đoạn này, người ta thường vận dụng nhiều hình thức tổ chức, nhiều phương pháp điều tra khác nhau, tùy thuộc mục đích nghiên cứu, nhu cầu thông tin, điều kiện cụ thể của người nghiên cứu và đặc điểm cơ bản của hiện tượng nghiên cứu.

Tổng hợp thống kê

Tổng hợp thống kê được hiểu là quá trình tập trung chỉnh lý, hệ thống hoá các tài liệu thu được trong điều tra thống kê để làm cho các đặc trưng riêng biệt của từng đơn vị điều tra bước đầu chuyển thành những thông tin chung của toàn bộ hiện tượng nghiên cứu, làm cơ sở cho việc phân tích tiếp theo...

Do đối tượng nghiên cứu của thống kê là các hiện tượng số lớn, phức tạp, bao gồm nhiều đơn vị có những đặc điểm khác nhau, nên việc tổng hợp thường được thực hiện đến từng tổ, từng nhóm đại diện cho các loại hình khác nhau. Vì vậy, phương pháp phân tổ thống kê được coi là quan trọng nhất của tổng hợp thống kê.

Ngày nay, sự phát triển mạnh mẽ của công nghệ thông tin đã làm cho tổ hợp thống kê càng có vai trò quan trọng, phục vụ được nhiều yêu cầu của phân tích thống kê. Việc tổng hợp bằng máy thay thế cho tổng hợp thủ công chẳng những rút ngắn được thời gian tổng hợp, nâng cao độ chính xác của việc xử lý thông tin, mà còn đáp ứng được các yêu cầu khác nhau về thông tin. Tuy nhiên, muốn thực hiện được việc tổng hợp bằng máy, cần phải xây dựng được chương trình phần mềm nhập số liệu và xác định rõ phần mềm phân tích sử dụng.

Phân tích và dự báo thống kê

Phân tích và dự báo thống kê được hiểu là việc nêu lên một cách tổng hợp bản chất và tính quy luật của các hiện tượng và quá trình kinh tế - xã hội số lớn trong điều kiện lịch sử cụ thể qua các biểu hiện về lượng và tính toán và các mức độ trong tương lai nhằm đưa ra những căn cứ cho quyết định quản lý.

Như vậy, phân tích và dự báo thống kê là biểu hiện tập trung kết quả của toàn bộ quá trình nghiên cứu thống kê. Phân tích và dự báo thống kê giúp ta thấy rõ bản chất, quy luật phát triển của hiện tượng trong quá khứ, hiện tại, giúp tiên đoán được các mức độ của hiện tượng trong tương lai. Đồng thời, nó còn giúp chỉ rõ mối liên hệ nội tại giữa các bộ phận của tổng thể, mối liên hệ, tác động quay lại giữa hiện tượng nghiên cứu với các hiện tượng có liên quan. Trên cơ sở đó, giúp ta có nhận thức đúng đắn về hiện tượng, tìm các biện pháp thích hợp thúc đẩy hiện tượng phát triển theo hướng tốt nhất, phù hợp với điều kiện, hoàn cảnh kinh tế - xã hội hiện đại.

Tuy nhiên, để phân tích và dự báo thống kê thực hiện được nhiệm vụ của mình, mang lại những kết quả nghiên cứu chính xác, cần phải chú ý một số yêu cầu cơ bản sau:

- *Thứ nhất:* Phân tích và dự báo thống kê phải được tiến hành trên cơ sở phân tích lý luận kinh tế - xã hội.
- *Thứ hai:* Phân tích và dự báo thống kê phải căn cứ vào toàn bộ sự kiện thực tế và phải luôn đặt chúng trong mối liên hệ ràng buộc, tác động qua lại lẫn nhau. Không được bỏ qua hay xem thường bất kỳ một sự kiện nào, dù là nhỏ nhất.

Đối với các hiện tượng có tính chất và hình thức khác nhau, phải áp dụng phương pháp phân tích và dự báo khác nhau, phù hợp với bản thân chúng, phù hợp với điều kiện, hoàn cảnh cụ thể. Không có phương pháp phân tích và dự báo vạn năng, áp dụng cho mọi trường hợp.

TÀI LIỆU THAM KHẢO CHƯƠNG 1

1. TS. Nguyễn Công Nhựt - **Giáo trình Thống kê công nghiệp**. NXB Thống kê, 2013
2. PGS. TS. Trần Ngọc Phác, TS Trần Thị Kim Thu – **Giáo trình Lý thuyết thống kê**. Nhà xuất bản Thống kê, 2013
3. GS.TS Bùi Xuân Phong - **Thống kê và ứng dụng trong BCVT**. NXB Bưu điện, 2005
4. Hà Văn Sơn - **Giáo trình Lý thuyết thống kê ứng dụng trong quản trị và kinh tế**. Nhà xuất bản Thống kê, 2004
5. PTS. Phan Công Nghĩa - **Giáo trình Thống kê Lao động**. NXB Thống kê, 2009

CÂU HỎI ÔN TẬP CHƯƠNG 1

1. Phân tích đối tượng nghiên cứu của thống kê học.
2. Trình bày các khái niệm thường dùng trong thống kê, ý nghĩa của các khái niệm này? Cho ví dụ?
3. Trình bày các loại thang đo trong thống kê. Cho ví dụ về việc sử dụng các loại thang đo này trong thực tế?

CHƯƠNG 2

ĐIỀU TRA THỐNG KÊ KINH TẾ

2.1 Khái niệm ý nghĩa của điều tra thống kê kinh tế

2.1.1. Khái niệm, ý nghĩa của điều tra thống kê

Nhiệm vụ chủ yếu của điều tra thống kê là thu nhập được đầy đủ thông tin cần thiết về hiện tượng nghiên cứu, làm cơ sở cho việc tổng hợp và phân tích. Tuy nhiên đối tượng của thống kê thường là những hiện tượng số lớn, phức tạp bao gồm nhiều đơn vị, phần tử khác nhau. Mặt khác, các hiện tượng này lại luôn biến động theo thời gian và không gian. Do sự đa dạng, phong phú và phức tạp của đối tượng nghiên cứu, nên muốn đáp ứng được mục đích nghiên cứu, muốn giải quyết được một vấn đề lý thuyết hoặc thực tế đã được định trước đòi hỏi các cuộc điều tra thống kê phải được tổ chức một cách khoa học, có kế hoạch tập trung, thống nhất, có chuẩn bị chu đáo theo những nguyên tắc khoa học nhất định.

Điều tra thống kê là việc tổ chức một cách khoa học theo một kế hoạch thống nhất việc thu thập, ghi chép nguồn tài liệu ban đầu về hiện tượng nghiên cứu trong điều kiện cụ thể về thời gian, không gian.

Điều 3, Luật Thống kê của nước Cộng Hoà Xã Hội Chủ Nghĩa Việt Nam cũng định nghĩa: “Điều tra thống kê là hình thức thu thập thông tin thống kê theo phương án điều tra”. Định nghĩa này hoàn toàn phù hợp với khái niệm nêu trên, bởi lẽ phương án điều tra thống kê sẽ quy định rõ về mục đích, ý nghĩa, toàn bộ quá trình tổ chức điều kiện thời gian, không gian... của cuộc điều tra. Tính khoa học, tính kế hoạch của cuộc điều tra được thể hiện rõ trong phương án này.

Điều tra thống kê, nếu được tổ chức theo những nguyên tắc khoa học, chặt chẽ, sẽ đáp ứng được nhiều yêu cầu khác nhau cả về lý thuyết cũng như thực tế đặt ra. Trước hết, tài liệu do điều tra thống kê thu được là căn cứ tin cậy để kiểm tra, đánh giá thực trạng hiện tượng nghiên cứu, đánh giá tình hình thực hiện kế hoạch phát triển kinh tế, văn hoá, xã hội của từng đơn vị, từng địa phương và của toàn bộ nền kinh tế quốc dân. Thứ hai, điều tra thống kê cung cấp những luận cứ xác đáng cho việc phân tích, phát hiện, tìm ra những yếu tố tác động, những yếu tố quyết định sự biến đổi của hiện tượng nghiên cứu. Trên cơ sở đó, tìm biện pháp thích hợp thúc đẩy hiện tượng phát triển theo hướng có lợi nhất. Thứ ba, những tài liệu điều tra thống kê cung cấp một cách có hệ thống còn là căn cứ vững chắc cho việc phát hiện, xác định xu hướng, quy luật biến động của hiện tượng và dự đoán xu hướng biến động của hiện tượng trong tương lai. Trong quá trình điều hành, quản lý kinh tế - xã hội, các tài liệu này giúp cho việc xây dựng các định hướng, các kế hoạch phát triển kinh tế - xã hội trong tương lai, quản lý quá trình thực hiện các kế hoạch đó.

Theo cách thức tổ chức các hoạt động thống kê nhà nước của nước Cộng hòa XHCN Việt Nam hiện nay, điều tra thống kê được tổ chức thành hai cấp độ: Tổng điều tra thống kê và điều tra thống kê.

Tổng điều tra thống kê để thu thập những thông tin thống kê cơ bản, trên phạm vi cả nước theo chu kỳ dài, quy mô lớn, phạm vi rộng liên quan nhiều ngành, nhiều cấp, có sử dụng lực lượng và kinh phí rất lớn. Các cuộc Tổng điều tra dân số được tổ chức theo chu kỳ 10 năm một lần là một ví dụ điển hình thuộc loại này.

Điều tra thống kê được thực hiện đối với các tổ chức, hộ gia đình và cá nhân trong các trường hợp:

- Thu thập những thông tin thống kê từ các tổ chức không phải thực hiện chế độ báo cáo thống kê.
- Khi cần bổ sung thông tin từ các tổ chức có thực hiện chế độ báo cáo thống kê.
- Thu thập thông tin từ các hộ cơ sở kinh doanh cá thể, hộ gia đình, cá nhân.
- Thu thập thông tin thống kê khi có nhu cầu đột xuất.

2.1.2. Các yêu cầu cơ bản của điều tra thống kê

Muốn thực hiện được nhiệm vụ và các mục đích nghiên cứu nói trên, điều tra thống kê cần đảm bảo được các yêu cầu cơ bản của một hoạt động thống kê nói chung là: Trung thực, khách quan, chính xác, đầy đủ và kịp thời. Đây cũng chính là những vấn đề được quy định rõ trong nguyên tắc thứ nhất - nguyên tắc quan trọng nhất của hoạt động thống kê. Chúng không chỉ là những yêu cầu cơ bản của điều tra thống kê, mà nó còn liên quan đến tất cả các đối tượng áp dụng Luật Thống kê, liên quan đến mọi hoạt động thống kê, từ điều tra, thực hiện các chế độ báo cáo thống kê cơ sở, chế độ báo cáo thống kê tổng hợp đến xử lý số liệu, phân tích và công bố các thông tin thống kê.

Yêu cầu trung thực được đặt ra cho cả người tổ chức điều tra và người cung cấp thông tin. Yêu cầu này đòi hỏi người thu thập thông tin phải tuyệt đối trung thực, ghi chép đúng những điều đã được nghe, được thấy. Ngay trong việc đặt câu hỏi cũng phải hết sức khách quan, không áp đặt ý muốn chủ quan, thậm chí không được đưa ra những gợi ý có thể gây ảnh hưởng đối với người trả lời... nhằm giúp thu được những thông tin trung thực. Đối với người cung cấp thông tin, yêu cầu này đòi hỏi họ phải cung cấp những thông tin xác thực, không được che giấu và đặc biệt nghiêm cấm việc khai man thông tin...

Chính xác - khách quan trong điều tra thống kê nghĩa là các tài liệu thu thập phản ánh đúng đắn tình hình thực tế khách quan của hiện tượng nghiên cứu. Điều này đòi hỏi việc ghi chép phải được thực hiện một cách trung thực, không được tùy tiện thêm bớt, không được sao chép một cách tùy hứng, không được suy luận, “sáng tạo” ra các con số theo ý muốn chủ quan của bất kỳ một tổ chức hay cá nhân nào, nhằm bất cứ mục đích nào. Đây được coi là yêu cầu cơ bản nhất của điều tra thống kê. Tài liệu được điều tra chính xác mới có thể dùng làm căn cứ tin cậy cho việc tổng hợp, phân tích và rút ra kết luận đúng đắn về hiện trạng, các yếu tố ảnh hưởng đến biến động của hiện tượng, đến quy luật xu hướng biến động của nó. Chỉ có trên cơ sở tài liệu được điều tra chính xác - khách quan mới có thể tính toán, lập kế hoạch và quản lý tốt các mặt, các lĩnh vực của nền kinh tế quốc dân. Chính yêu cầu này cũng lại trở thành một yếu tố cơ bản quyết định chất lượng của công tác thống kê.

Yêu cầu kịp thời của điều tra thống kê được hiểu theo hai nghĩa. Trước hết, các tài liệu của điều tra thống kê phải có tính nhạy bén, phản ánh được sự biến đổi của hiện tượng nghiên

cứu đúng lúc cần thiết, đúng lúc hiện tượng có sự thay đổi về chất và phản ánh đầy đủ những bước ngoặt quan trọng nhất trong sự biến đổi của hiện tượng mà ta cần theo dõi. Tức là các tài liệu ghi chép được phải mang tính thời sự. Thứ hai thống kê phải cung cấp tài liệu phục vụ các yêu cầu nghiên cứu đúng lúc cần thiết. Trong quản lý kinh tế, yêu cầu kịp thời của điều tra thống kê giúp cho nhà quản lý ra các quyết định, mệnh lệnh có tính chuẩn xác, mang lại lợi ích kinh tế cao. Yêu cầu này cũng đòi hỏi việc cung cấp thông tin phải đảm bảo đúng hạn theo quy định của phương án điều tra.

Đầy đủ có ý nghĩa là tài liệu điều tra phải được thu thập theo đúng nội dung cần thiết cho việc nghiên cứu hoặc đã được quy định trong phương án điều tra. Đầy đủ cũng còn có nghĩa là phải thu thập thông tin đối với tất cả số đơn vị của hiện tượng nghiên cứu, không được đếm trùng hay bỏ sót bất kỳ đơn vị nào. Trong các cuộc điều tra không toàn bộ thì phải thu thập thông tin trên tất cả các đơn vị đã được quy định. Tài liệu điều tra đầy đủ sẽ không những giúp cho việc tổng hợp, phân tích chính xác mà còn giúp cho việc đánh giá phân tích hiện tượng nghiên cứu một cách chính xác, tránh đưa ra những kết luận phiến diện chủ quan.

Ngoài ra, điều tra thống kê muốn phản ánh đúng bản chất, tính quy luật của hiện tượng nghiên cứu cần phải dựa trên cơ sở quan sát số lớn. Quan sát số lớn trong điều tra thống kê có nghĩa là phải đảm bảo thu thập số liệu trên nhiều đơn vị hoặc hiện tượng cá biệt. Chỉ khi đó, các yếu tố ngẫu nhiên mới bị bù trừ, triệt tiêu lẫn nhau, bản chất và quy luật của hiện tượng mới được bộc lộ rõ qua tổng hợp và phân tích thống kê. Tuy nhiên, trong một số trường hợp nhất định, nhất là với những điển hình tiên tiến, lạc hậu, điều tra thống kê cũng có thể chỉ tiến hành trên một số đơn vị cá biệt, nhưng các đơn vị này phải được lựa chọn và xem xét trong mối quan hệ với tổng thể nghiên cứu.

2.2 Các loại điều tra thống kê kinh tế

Điều tra thống kê có nhiều loại khác nhau, tùy theo mục đích nghiên cứu, đặc điểm của đối tượng điều tra và điều kiện thực tế mà người ta có thể sử dụng loại nào cho phù hợp. Sau đây là một số cách phân loại điều tra chủ yếu:

2.2.1. Điều tra thường xuyên và không thường xuyên

Căn cứ vào tính liên tục, tính hệ thống của các cuộc điều tra, ta có thể phân biệt hai loại điều tra thống kê: Điều tra thường xuyên và không thường xuyên.

Điều tra thường xuyên là việc tiến hành thu thập, ghi chép tài liệu ban đầu của hiện tượng nghiên cứu một cách liên tục, có hệ thống và thường là theo quan sát quá trình phát sinh, phát triển của hiện tượng. Ví dụ, việc tổ chức chấm công lao động, theo dõi số công nhân đi làm hàng ngày tại các doanh nghiệp, việc ghi chép số sản phẩm nhập, xuất kho hàng ngày tại các kho hàng...

Điều tra thường xuyên giúp ta thu thập được những số liệu theo dõi tỉ mỉ tình hình phát triển của hiện tượng theo thời gian, đánh giá được sự phát triển, tích lũy của hiện tượng. Tài liệu của điều tra thường xuyên là cơ sở chủ yếu để lập báo cáo thống kê định kỳ, là công cụ quan trọng để theo dõi tình hình thực hiện kế hoạch. Loại điều tra này phù hợp với những hiện tượng có quá trình phát triển liên tục cần phải theo dõi.

Hình thức tổ chức chủ yếu và quan trọng nhất của các cuộc điều tra thống kê thường xuyên là “Báo cáo thống kê định kỳ”. Đây là một hình thức thu thập số liệu dựa vào các biểu mẫu báo cáo thống kê được lập sẵn. Theo đó, các đơn vị báo cáo ghi số liệu vào biểu mẫu và gửi lên cấp trên. Các báo cáo này được thực hiện một cách thường xuyên, định kỳ, theo nội

dung, phương pháp, biểu mẫu và chế độ báo cáo đã được quy định sẵn. Trong nhiều năm trước kia, đây là hình thức thu thập thông tin thống kê chủ yếu ở nước ta và thường được áp dụng cho các đơn vị kinh tế, cơ quan, tổ chức của nhà nước. Ngày nay, trong nền kinh tế thị trường thì hình thức thu thập thông tin thống kê này ít được sử dụng hơn.

Điều tra không thường xuyên là tiến hành thu thập, ghi chép tài liệu ban đầu của hiện tượng một cách không liên tục, không gắn với quá trình phát sinh, phát triển của hiện tượng.

Điều tra không thường xuyên thường được tiến hành đối với những hiện tượng ít biến động, biến động chậm hoặc không cần theo dõi thường xuyên, liên tục. Chỉ khi nào cần nghiên cứu, người ta mới tổ chức điều tra. Do đó, các cuộc điều tra không thường xuyên được tiến hành với mục đích nội dung, phạm vi, đối tượng và phương pháp điều tra không giống nhau. Tuy nhiên, để tiện cho việc theo dõi, so sánh, phân tích sự biến động của hiện tượng theo thời gian, nhiều cuộc điều tra không thường xuyên cũng được tiến hành lặp đi lặp lại theo chu kỳ nhất định và người ta thường cố gắng kế thừa những gì đã thực hiện tại các cuộc điều tra trước, nay vẫn còn có ý nghĩa. Chẳng hạn, các cuộc Tổng điều tra dân số được tiến hành 10 năm một lần ở nước ta hiện nay thuộc loại điều tra này.

Hình thức chủ yếu của điều tra không thường xuyên là các cuộc điều tra chuyên môn. Khác với báo cáo thống kê định kỳ, điều tra chuyên môn chỉ được tổ chức khi có nhu cầu. Mỗi cuộc điều tra thường được tiến hành theo kế hoạch và phương pháp quy định riêng. Các cuộc điều tra chuyên môn ngày càng được sử dụng rộng rãi và phục vụ được nhiều thị trường, vai trò của các cuộc điều tra chuyên môn trong việc thu thập các số liệu thống kê ngày càng quan trọng. Do đó, nó càng được sử dụng. Tuy nhiên, muốn thực hiện tốt các cuộc điều tra chuyên môn, cần xây dựng được phương án điều tra tỉ mỉ, toàn diện và chi tiết.

2.2.2. Điều tra toàn bộ và không toàn bộ

Căn cứ vào phạm vi của đối tượng được điều tra thực tế, điều tra thống kê được phân thành điều tra toàn bộ và điều tra không toàn bộ.

1. Điều tra toàn bộ

Điều tra toàn bộ là tiến hành thu thập tài liệu ban đầu trên toàn thể các đơn vị thuộc đối tượng điều tra, không loại trừ bất kỳ đơn vị nào. Ví dụ: Các cuộc Tổng điều tra dân số được tiến hành vào ngày 1/10/1979, ngày 1/4/1989, ngày 1/4/1999 và 1/4/2009 ở nước ta.

Điều tra toàn bộ là nguồn cung cấp tài liệu đầy đủ nhất cho các nghiên cứu thống kê. Do tài liệu được thu thập trên toàn bộ các đối tượng nghiên cứu, nên nó vừa là cơ sở để tính được các chỉ tiêu tổng hợp cho cả tổng thể, lại vừa cung cấp số liệu chi tiết cho từng đơn vị. Có thể nói, điều tra toàn bộ là nguồn cung cấp thông tin thống kê đầy đủ, toàn diện và trực tiếp, nên nó có thể đáp ứng được nhiều yêu cầu nghiên cứu khác nhau, đặc biệt là trong điều tra nắm bắt tình hình cơ bản của hiện tượng. Tuy nhiên, với những hiện tượng lớn và phức tạp, điều tra toàn bộ thường đòi hỏi phải có nguồn tài chính lớn, số người tham gia đông, thời gian dài. Vì vậy, điều tra toàn bộ ít được tiến hành thường xuyên và thường được giới hạn ở một số nội dung chủ yếu.

2. Điều tra không toàn bộ

Điều tra không toàn bộ là tiến hành thu thập tài liệu ban đầu trên một số đơn vị được chọn trong toàn bộ các đơn vị của tổng thể chung.

Do chỉ tiến hành điều tra trên một số đơn vị của hiện tượng nghiên cứu, nên điều tra không toàn bộ có thể rút ngắn được thời gian, tiết kiệm công sức, giảm chi phí. Đặc biệt là

loại điều tra này vừa có điều kiện mở rộng nội dung điều tra, vừa có thể kiểm tra, đánh giá độ chính xác của số liệu thu được một cách thuận lợi. Do những ưu điểm trên, nên điều tra không toàn bộ được sử dụng ngày càng nhiều trong thực tế, kịp thời đáp ứng yêu cầu của công tác quản lý kinh tế - xã hội, quản lý nhà nước... Tuy nhiên, điều tra không toàn bộ cũng có những hạn chế nhất định. Hạn chế lớn nhất của nó là luôn phát sinh sai số do chỉ dựa trên cơ sở số liệu của một số ít đơn vị để nhận định, đánh giá cho toàn bộ hiện tượng nghiên cứu. Hạn chế này có thể được khắc phục bằng cách áp dụng những phương pháp khoa học, phù hợp với hiện tượng nghiên cứu trong quá trình tổ chức điều tra.

Căn cứ vào phương pháp lựa chọn các đơn vị để điều tra ta có thể phân chia điều tra không toàn bộ thành 3 loại khác nhau:

- **Điều tra chọn mẫu:** Là loại điều tra không toàn bộ, trong đó người ta chỉ chọn ra một số đơn vị để điều tra thực tế. Các đơn vị này được chọn theo những nguyên tắc khoa học nhất định để đảm bảo tính đại diện của chúng cho tổng thể chung. Kết quả điều tra được dùng để đánh giá, suy rộng cho toàn bộ hiện tượng.

- **Điều tra trọng điểm:** Trong điều tra trọng điểm, người ta chỉ tiến hành điều tra ở bộ phận chủ yếu nhất của tổng thể chung. Kết quả điều tra không được dùng để suy rộng thành các đặc điểm chung của hiện tượng. Loại điều tra này thích hợp với những đối tượng có những bộ phận tương đối tập trung, chiếm tỷ trọng lớn trong tổng thể. Ví dụ: nghiên cứu tình hình trồng chè Thái Nguyên, cà phê ở Tây Nguyên...

- **Điều tra chuyên đề:** chỉ được tiến hành trên một số rất ít, thậm chí một đơn vị của tổng thể, nhưng lại đi sâu nghiên cứu chi tiết khía cạnh khác nhau của đơn vị đó nhằm rút ra vấn đề cốt lõi, tìm những bài học kinh nghiệm chung. Tài liệu thu được trong điều tra chuyên đề không dùng để suy rộng hoặc làm căn cứ đánh giá tình hình cơ bản của hiện tượng nghiên cứu. Loại điều tra này thường được dùng để nghiên cứu những vấn đề mới phát sinh để nghiên cứu kinh nghiệm của các đơn vị tiên tiến hoặc phân tích tìm nguyên nhân yếu kém của các đơn vị lạc hậu...

2.3. Các phương pháp thu thập thông tin trong điều tra thống kê kinh tế

Để thu thập thông tin trong điều tra thống kê, người ta có thể sử dụng nhiều phương pháp khác nhau. Tùy theo điều kiện thực tế và đặc điểm của hiện tượng nghiên cứu, khả năng về tài chính, thời gian, kinh nghiệm, trình độ của người tổ chức và điều tra viên... để lựa chọn phương pháp điều tra thích hợp. Phần này sẽ trình bày những vấn đề chung của một số phương pháp chủ yếu trong điều tra thống kê.

2.3.1. Phương pháp đăng ký trực tuyến

Theo phương pháp này, nhân viên điều tra phải trực tiếp tiếp xúc với đối tượng điều tra, trực tiếp tiến hành hoặc giám sát việc, cân, đo, đóng, đếm và sau đó ghi chép những thông tin thu được vào phiếu điều tra. Như trong điều tra tồn kho vật tư, hàng hoá: Nhân viên điều tra trực tiếp cân, đo, đóng, đếm, phân loại số lượng vật tư, thiết bị, hàng hoá còn tồn trong kho, rồi ghi chép kết quả vào phiếu điều tra. Phương pháp đăng ký trực tiếp thường được gắn liền với quá trình phát sinh, phát triển của hiện tượng.

Tài liệu ghi chép ban đầu thu được qua đăng ký trực tiếp có độ chính xác khá cao, nhưng lại đòi hỏi nhiều nhân lực và thời gian. Mặt khác, trong thực tế có rất nhiều hiện tượng không cho phép quan sát, cân đo trực tiếp quá trình phát sinh, phát triển của nó được như

nghiên cứu thu chi, mức sống dân cư, những nội dung chủ yếu trong các cuộc Tổng điều tra dân số... Vì vậy, phạm vi áp dụng của phương pháp này rất hạn chế.

2.3.2 Phương pháp phỏng vấn

Phỏng vấn là phương pháp điều tra thống kê được sử dụng nhiều nhất, theo đó việc ghi chép, thu thập tài liệu ban đầu được thực hiện thông qua quá trình hỏi - đáp giữa nhân viên điều tra và người cung cấp thông tin.

Trong điều tra thống kê, phỏng vấn không phải là một cuộc nói chuyện, hỏi đáp thông thường, cũng không phải là cuộc phỏng vấn lấy tin của các nhà báo, càng không phải là cuộc phỏng vấn giữa nhân viên điều tra và người bị nghi vấn, can phạm... Phỏng vấn trong thống kê phải tuân thủ theo mục tiêu nghiên cứu, theo đối tượng, khách thể, nội dung nghiên cứu đã được xác định rõ trong chương trình, phương án điều tra. Điều tra viên bắt buộc phải tuân thủ phương án điều tra, đặc biệt là các nội dung điều tra được thể hiện cụ thể trong phiếu điều tra. Do đó, cần phải có sự chuẩn bị kỹ lưỡng về kỹ năng phỏng vấn, về năng lực chuyên môn, sự am hiểu về các nội dung, về đối tượng điều tra. Ngay việc ghi chép cũng phải được sự thực hiện nghiêm chỉnh theo tất cả các hướng dẫn, các quy định của phiếu điều tra để tạo điều kiện thuận lợi cho việc xử lý, tổng hợp thông tin sau này.

Trong thực tế, phỏng vấn có thể thích ứng với nhiều hoàn cảnh khác nhau mà không cần bám sát quá trình phát sinh, phát triển của hiện tượng. Mặt khác, thông tin thu được qua phỏng vấn thường có độ tin cậy cao, dễ tổng hợp, lại tập trung vào những nội dung chủ yếu nhờ có bảng hỏi hoặc phiếu điều tra. Do đó, phỏng vấn được sử dụng rộng rãi nhất trong điều tra thống kê nhằm thu nhận nguồn tài liệu ghi chép ban đầu.

Nếu căn cứ vào tính chất của sự tiếp xúc giữa người hỏi và người trả lời, ta phân biệt 2 loại: Phỏng vấn trực tiếp và phỏng vấn gián tiếp.

1. Phỏng vấn trực tiếp

Đây là phương pháp ghi chép, thu thập tài liệu ban đầu được thực hiện thông qua quá trình hỏi - đáp trực tiếp giữa nhân viên điều tra và người cung cấp thông tin. Tức là nhân viên điều tra trực tiếp đến địa bàn điều tra, tìm gặp đối tượng trả lời vào phiếu điều tra. Do việc tiếp xúc trực tiếp giữa người hỏi và người trả lời, nên phương pháp này tạo ra những điều kiện đặc biệt để hiểu đối tượng sâu sắc, giúp điều tra viên có thể kết hợp việc phỏng vấn với việc quan sát đối tượng từ dáng vẻ bề ngoài, đến những cử chỉ biểu lộ tình cảm, thái độ, nên có thể phát hiện ngay những sai sót và uốn nắn kịp thời. Mặt khác, nhân viên điều tra có thể giải thích kỹ các câu hỏi và rà soát tại chỗ các câu trả lời. Do đó, có thể đảm bảo chất lượng của các tài liệu thu được. Tuy nhiên, phỏng vấn trực tiếp cũng có những nhược điểm là: Tốn kém về thời gian, chi phí, số cán bộ điều tra; nó đòi hỏi phải có sự chuẩn bị càng kỹ về điều tra viên, địa điểm phỏng vấn và cần có sự tổ chức các cuộc gặp gỡ sao cho người trả lời không thấy bị gò bó, miễn cưỡng. Ngoài ra cần có sự lựa chọn, tập huấn thật tốt cho điều tra viên để thu được những số liệu thật sự khách quan.

2. Phỏng vấn gián tiếp

Đây là phương pháp thu thập tài liệu ban đầu được thực hiện bằng cách người được hỏi nhận phiếu điều tra, tự mình viết vào phiếu rồi gửi lại cho cơ quan điều tra.

Đặc điểm cơ bản của phương pháp này là người hỏi và người trả lời không trực tiếp gặp nhau. Quá trình hỏi - đáp diễn ra thông qua một vật trung gian là phiếu điều tra. Muốn nâng cao chất lượng các thông tin thu được cần chú ý đến một số điều kiện cơ bản:

- Người được hỏi phải có trình độ văn hoá cao, có ý thức trách nhiệm và tự giác.
- Phiếu điều tra phải ngắn gọn.
- Các câu hỏi phải rõ ràng, dễ hiểu, dễ trả lời.
- Phải thiết lập được một mạng lưới phân phát và thu hồi phiếu điều tra hợp lý, hoạt động có kết quả, không để bị thất lạc phiếu.

Ưu điểm cơ bản của phương pháp là dễ tổ chức, tiết kiệm chi phí và điều tra viên. Tuy nhiên, lại khó có thể kiểm tra, đánh giá được độ chuẩn xác của các câu trả lời, tỷ lệ hồi phiếu trong nhiều trường hợp là không cao, nội dung điều tra bị hạn chế. Phương pháp này cũng chỉ có thể sử dụng được trong điều kiện trình độ dân trí cao.

Ngoài ra, trong điều tra thống kê người ta còn sử dụng nhiều phương pháp khác để thu thập nguồn tài liệu ghi chép ban đầu như: phương pháp quan sát; phương pháp thu thập thông tin qua những nguồn sẵn có; phương pháp qua các chứng từ sổ sách...

2.4 Xây dựng phương án trong điều tra thống kê kinh tế

Để tổ chức tốt một cuộc điều tra chuyên môn, đòi hỏi phải xây dựng được phương án điều tra thật chi tiết, tỉ mỉ, cụ thể và toàn diện. Đây chính là văn kiện hướng dẫn thực hiện cuộc điều tra, trong đó xác định rõ những bước tiến hành, những vấn đề cần phải giải quyết, cần được hiểu thống nhất trong suốt quá trình thực hiện. Đối với các cuộc điều tra lớn, có liên quan đến nhiều cấp, nhiều ngành, như Tổng điều tra dân số, việc xây dựng phương án điều tra cần có sự phối hợp, bàn bạc thống nhất giữa cơ quan thống kê và các ngành có liên quan và phải được cấp trên có thẩm quyền phê duyệt. Trong điều kiện nền kinh tế thị trường, phương án điều tra thường được xây dựng phê duyệt hoặc là căn cứ để cơ quan chủ quản tiến hành xét chọn thầu theo quy định chung của nhà nước.

Phương án của mỗi cuộc điều tra có thể khác nhau, tùy thuộc điều kiện cụ thể của nó. Nhìn chung, mỗi phương án điều tra thường gồm những nội dung chủ yếu sau:

- Xác định mục đích điều tra.
- Xác định nội dung điều tra và thiết lập phiếu điều tra.
- Xác định phạm vi, đối tượng và đơn vị điều tra.
- Chọn thời điểm, thời kỳ và thời hạn điều tra.
- Lựa chọn phương pháp điều tra, tổng hợp số liệu và phương pháp tính cách chỉ tiêu điều tra.
- Xây dựng phương án tài chính cho cuộc điều tra.
- Lập kế hoạch tổ chức và tiến hành điều tra.

2.4.1 Xác định mục đích điều tra

Bất kỳ hiện tượng kinh tế - xã hội nào cũng đều có thể được quan sát, xem xét, nghiên cứu trên nhiều mặt, nhiều khía cạnh, nhiều góc độ khác nhau. Nghiên cứu trên mỗi mặt, mỗi khía cạnh khác nhau sẽ cho ta đưa ra những kết luận khác nhau về hiện tượng và phục vụ những yêu cầu khác nhau. Vì vậy, trước khi tiến hành điều tra, cần xác định rõ xem cuộc điều tra này nhằm tìm hiểu vấn đề gì, phục vụ cho yêu cầu nghiên cứu nào. Đó chính là mục đích của cuộc điều tra.

Mục đích điều tra còn là một trong những căn cứ quan trọng để xác định đối tượng, đơn vị điều tra, xây dựng kế hoạch và nội dung điều tra. Vì vậy, việc xác định đúng, rõ ràng

mục đích điều tra sẽ là cơ sở quan trọng cho việc thu thập số liệu ban đầu đầy đủ, hợp lý, đáp ứng yêu cầu nghiên cứu đặt ra.

Căn cứ để xác định mục đích điều tra thường là những nhu cầu thực tế cuộc sống, hoặc những nhu cầu hoàn chỉnh lý luận... những nhu cầu này được biểu hiện trực tiếp bằng các yêu cầu, đề nghị, mong muốn của cơ quan chủ quản. Đối với những cuộc điều tra lớn ở nước ta, có liên quan đến toàn bộ đất nước, tổng điều tra dân số, việc xác định mục đích điều tra cần phải căn cứ vào định hướng phát triển kinh tế, xã hội của đất nước, vào đường lối, chủ trương, chính sách của Đảng, Nhà nước trong từng thời kỳ cụ thể. Mục đích của tổng điều tra dân số năm 2009 nhằm phục vụ hai yêu cầu cơ bản:

- Làm căn cứ chính xác phục vụ công tác nghiên cứu, phân tích quá trình phát triển kinh tế - xã hội và dân số nước ta.

- Kiểm điểm tình hình thực hiện các nhiệm vụ liên quan đến dân số, lao động. Trên cơ sở đó, đáp ứng yêu cầu xây dựng kế hoạch phát triển kinh tế - xã hội thời kỳ 2010 - 2020.

2.4.2 Xác định phạm vi, đối tượng và đơn vị kiểm tra

Xác định đối tượng điều tra là xác định xem những đơn vị tổng thể nào thuộc phạm vi điều tra cần được thu thập thông tin. Như vậy, khi các đối tượng điều tra được chỉ rõ, cũng có nghĩa là phạm vi nghiên cứu đã được xác định, ranh giới giữa hiện tượng nghiên cứu với các tổng thể khác, hiện tượng khác cũng được phân biệt rõ ràng, tránh được tình trạng trùng lặp hay bỏ sót tiến hành điều tra.

Muốn xác định chính xác đối tượng điều tra, một mặt phải dựa vào sự phân tích lý luận, nêu lên những tiêu chuẩn cơ bản phân biệt hiện tượng, nghiên cứu với các hiện tượng liên quan, phân biệt đơn vị tổng thể này với các đơn vị tổng thể khác, đồng thời cũng còn phải căn cứ vào mục đích nghiên cứu. Trong cuộc tổng điều tra dân số ngày 1/4/2009 ở nước ta, đối tượng điều tra được xác định là “nhân khẩu thường trú”. Điều này, vừa giúp thực hiện tốt các mục đích điều tra đã được nêu rõ trong mục trên, vừa giúp thực hiện tốt các mục đích điều tra không bị trùng hay bỏ sót bất kỳ một nhân khẩu nào của nước ta. Tuy nhiên, trong phương án điều tra cũng cần phải có những quy định cụ thể về tiêu chuẩn xác định “nhân khẩu thường trú” để tránh nhầm lẫn.

Đơn vị điều tra là đơn vị thuộc đối tượng điều tra và được điều tra thực tế. Đơn vị điều tra chính là nơi phát sinh các tài liệu ban đầu, điều tra viên cần đến đó thu thập trong mỗi điều tra. Như vậy, nếu xác định đối tượng điều tra là trả lời câu hỏi “điều tra ở đâu?”. Trong một số trường hợp, đơn vị điều tra và đối tượng điều tra đều là các doanh nghiệp công nghiệp nhà nước của thành phố. Nhưng cũng có nhiều trường hợp, chúng lại khác nhau. Ví dụ, trong Tổng điều tra dân số ở nước ta ngày 1/4/2009, đối tượng điều tra là “nhân khẩu thường trú”, còn đơn vị điều tra lại được xác định là “hộ gia đình” và các “hộ tập thể”. Trong các cuộc điều tra chọn mẫu, đơn vị điều tra chỉ bao gồm những đối tượng được chọn vào mẫu.

Cần phân biệt đơn vị điều tra và đơn vị tổng thể. Đơn vị tổng thể là các phần tử, các đơn vị cấu thành hiện tượng, mà qua đó ta có thể xác định được quy mô tổng thể. Việc xác định số đơn vị tổng thể liên quan đến việc lập phương án điều tra, chọn phương pháp điều tra, ước lượng kinh phí điều tra... còn việc xác định số đơn vị điều tra lại liên quan đến việc tổ chức ghi chép, đăng ký tài liệu, phân bổ cán bộ...

2.4.3 Xác định nội dung điều tra và thiết lập phiếu điều tra

Xác định nội dung điều tra là việc câu hỏi “điều tra cái gì?”. Nội dung điều tra là toàn bộ các đặc điểm cơ bản của từng đối tượng, từng đơn vị điều tra, mà ta cần thu được thông tin. Trong thực tế, các đơn vị của hiện tượng nghiên cứu thường có rất nhiều đặc điểm khác nhau. Tuy nhiên, ta không thể và cũng không cần thiết phải thu thập thông tin của toàn bộ các tiêu thức đó, mà chỉ cần những tiêu thức có liên quan đến mục đích nghiên cứu, phục vụ được cho việc nghiên cứu. Vì vậy, bất kỳ cuộc điều tra nào cũng cần phải xác định rõ, thật cụ thể nội dung điều tra.

Việc xác định nội dung điều tra, cần căn cứ vào các yếu tố sau:

- Mục đích điều tra: Mục đích điều tra chỉ rõ cần thu thập những thông tin nào để đáp ứng yêu cầu của nó. Mục đích điều tra khác nhau, nhu cầu thông tin cũng khác nhau. Mục đích càng nhiều, nội dung điều tra càng phải rộng, càng phải phong phú.

- Đặc điểm của hiện tượng nghiên cứu: Tất cả những hiện tượng mà thống kê nghiên cứu đều tồn tại trong điều kiện cụ thể về thời gian và không gian. Khi điều kiện này thay đổi, đặc điểm của hiện tượng cũng thay đổi. Khi đó, các biểu hiện của chúng cũng khác nhau. Vì vậy, việc lựa chọn tiêu thức nghiên cứu cũng phải khác nhau.

- Năng lực, trình độ thực tế của đơn vị, của người tổ chức điều tra. Điều này biểu hiện ở khả năng về tài chính, về thời gian, về kinh tế và trình độ tổ chức điều tra. Nếu tất cả các yếu tố này được đảm bảo tốt, có thể mở rộng nội dung điều tra, nhưng vẫn đảm bảo chất lượng của các thông tin thu được. Trường hợp ngược lại, cần kiên quyết loại bỏ những nội dung chưa thực sự cần thiết.

Ngoài ra, nội dung điều tra cũng chỉ nên bao gồm những tiêu thức có liên hệ chặt chẽ với nhau, để có thể kiểm tra tính chính xác của những thông tin thu được.

Để có thể thu được những thông tin một cách chính xác và đầy đủ, nội dung của mỗi cuộc điều tra phải được diễn đạt thành những câu hỏi ngắn gọn, rõ ràng, dễ hiểu và mọi người đều hiểu theo một nghĩa thống nhất. Về mặt hình thức, các câu hỏi này có thể được diễn đạt theo hai cách: Câu hỏi đóng là các câu hỏi đã có sẵn các phương án trả lời có thể, người trả lời chỉ cần chọn 1 trong những phương án trả lời, người được diễn đạt câu trả lời. Các cuộc điều tra thống kê ít sử dụng loại câu hỏi thứ hai này. Trong cuộc Tổng điều tra được thực hiện trong 18 câu hỏi. Những nội dung chủ yếu của cuộc tổng điều tra này là: Họ tên, quan hệ với chủ hộ, giới tính, dân tộc, tôn giáo, nơi thực tế thường trú, tình trạng đi học và trình độ học vấn cao nhất đã đạt được trình độ chuyên môn kỹ thuật, tình trạng hôn nhân, tình trạng việc làm, lĩnh vực làm việc...

Phiếu điều tra là tập hợp các câu hỏi của nội dung điều tra, được sắp xếp theo một trật tự logic nhất định. Tùy theo yêu cầu, nội dung với đối tượng, mỗi cuộc điều tra có thể phải xây dựng nhiều loại phiếu khác nhau. Trong tổng điều tra dân số, người ta xây dựng hai loại phiếu: Phiếu hộ và phiếu cá nhân. Trong hai cuộc điều tra mức sống dân cư Việt Nam (1997 – 1998 và 2002 – 2003) do Tổng cục Thống kê thực hiện, ngoài việc thu thập những thông tin chủ yếu từ các hộ gia đình, còn cần những thông tin về cộng đồng, về giá cả thị trường có ảnh hưởng trực tiếp đến mức sống của người dân, nên người ta phải sử dụng đến 5 loại phiếu khác nhau: Bảng câu hỏi hộ gia đình, bảng câu hỏi xã/thôn, bảng câu hỏi trường học, bảng câu hỏi trạm y tế và câu hỏi về giá.

Thông thường, trong các văn kiện của cuộc điều tra, người ta còn ban hành bản giải thích cách ghi phiếu điều tra. Bản giải thích này thường đi kèm theo phiếu điều tra nhằm giúp

cho nhân viên điều tra và người trả lời nhận thức thống nhất các câu hỏi được đặt ra, cách thu thập và ghi chép số liệu. Đối với những câu hỏi phức tạp, khó trả lời người ta còn đặt ra những ví dụ cụ thể và những quy định về các trường hợp ngoại lệ...

2.4.4. Chọn thời điểm, thời kỳ và quyết định thời hạn điều tra

Các hiện tượng kinh tế - xã hội mà thống kê nghiên cứu luôn thay đổi theo thời gian và không gian. Muốn thu thập được chính xác các thông tin về chúng, cần có quy định thống nhất về thời điểm, thời kỳ và thời hạn điều tra.

Thời điểm điều tra: Là mốc thời gian quy định thống nhất mà cuộc điều tra phải thu thập thông tin về hiện tượng tồn tại đúng thời điểm tra đó. Nếu cuộc điều tra được tiến hành vào thời điểm sau đó, thì người trả lời phải hồi tưởng lại để “miêu tả trạng thái của hiện tượng” vào đúng thời điểm điều tra. Ví dụ, thời điểm của tổng điều tra dân số lần thứ 4 ở nước ta được xác định vào 0 giờ ngày 1/4/2009. Như vậy, số dân nước ta có tại cuộc điều tra này chỉ tính những người sinh trước 0 giờ 1/4/2009 và đến thời điểm này đang còn sống.

Thời kỳ điều tra: Là khoảng thời gian (tuần, tháng, năm...) được quy định để thu thập số liệu về lượng của hiện tượng được tích lũy trong cả thời kỳ đó.

Thời hạn điều tra: Là khoảng thời gian dành cho việc thực hiện nhiệm vụ thu thập số liệu. Thời hạn dài hay ngắn phụ thuộc vào quy mô, tính phức tạp của hiện tượng nghiên cứu và nội dung điều tra, vào khả năng, kinh nghiệm của điều tra viên. Nhìn chung, thời hạn điều tra không nên quá dài, cách quá xa thời điểm điều tra vì có thể làm mất thông tin do người trả lời không nhớ đầy đủ các sự kiện xảy ra.

2.4.5. Lập kế hoạch tổ chức và tiến hành điều tra

Lập kế hoạch tổ chức và tiến hành điều tra là một vấn đề trọng yếu thống kê. Kế hoạch này quy định cụ thể từng bước công việc phải tiến hành trong quá trình từ khâu tổ chức đến triển khai điều tra thực tế. Vì vậy, nó được xây dựng ngày càng chi tiết, tỷ mỉ, rõ ràng, cụ thể thì càng dễ thực thi, chất lượng của cuộc điều tra càng được nâng cao. Tuy nhiên, đây là một công việc phức tạp, đòi hỏi người lập kế hoạch phải có kinh nghiệm và am hiểu tình hình thực tế. Một kế hoạch tổ chức và tiến hành điều tra gồm rất nhiều khâu công việc. Thông thường, nó có thể gồm một số khâu chủ yếu là:

- Thành lập ban chỉ đạo điều tra và quy định nhiệm vụ cụ thể cho cơ quan điều tra các cấp.
- Chuẩn bị lực lượng cán bộ điều tra, phân công trách nhiệm, địa bàn cho từng cán bộ và tiến hành tập huấn nghiệp vụ cho họ.
- Lựa chọn phương pháp điều tra thích hợp.
- Xác định các bước tiến hành điều tra.
- Phân chia khu vực và địa bàn điều tra.
- Tổ chức các cuộc hội nghị chuẩn bị.
- Tiến hành điều tra thử để rút kinh nghiệm, nâng cao trình độ nghiệp vụ cho cán bộ điều tra và hoàn thiện phương án điều tra, phiếu điều tra.
- Xây dựng phương án tài chính và chuẩn bị các phương tiện vật chất khác.
- Tuyên truyền mục đích, ý nghĩa của cuộc điều tra.
- ...

2.5 Xây dựng bảng hỏi trong điều tra thông kê

2.5.1. Bảng hỏi và yêu cầu của việc xây dựng bảng hỏi trong điều tra thông kê

Bảng hỏi là hệ thống các câu hỏi được sắp xếp trên cơ sở các nguyên tắc, trình tự logic và theo nội dung nhất định nhằm giúp cho người điều tra có thể thu được thông tin về hiện tượng nghiên cứu một cách đầy đủ, đáp ứng mục tiêu nghiên cứu đã được thiết lập.

Trong điều tra nói chung, bảng hỏi chính là sự thể hiện cụ thể toàn bộ nội dung nghiên cứu. Nó là công cụ quan trọng giúp truyền tải thông tin từ người nghiên cứu đến đối tượng và thu nhận thông tin ngược lại. Việc đo lường, nhận thức hiện tượng có đầy đủ hay không phụ thuộc chặt chẽ vào chỗ các câu hỏi bao quát nội dung nghiên cứu đến mức độ nào. Chính vì vậy, nhìn vào bảng hỏi, ta có thể biết cuộc điều tra được tiến hành nhằm giải quyết vấn đề gì, chương trình nghiên cứu như thế nào.

Bảng hỏi là một công cụ đo lường quan trọng. Nhờ đó, ta có thể đo lường được các biến số nhất định có liên quan đến chủ đề cần nghiên cứu. Vì vậy, bảng hỏi được soạn thảo tốt sẽ cho ta những thông tin đầy đủ, tin cậy và việc đo lường sẽ đạt độ chính xác, khoa học. Ngược lại, nếu các câu hỏi không đáp ứng được các yêu cầu đề ra thì khả năng thu thập thông tin sẽ giảm, thậm chí ta có thể nhận được các thông tin mào mó, xuyên tạc sự thực.

Bảng hỏi còn giữ vai trò cầu nối giữa người nghiên cứu và trả lời. Một mặt, nó chịu tác động của người nghiên cứu khi chuyển các nội dung cần thiết thành các câu hỏi cụ thể với mong muốn thu được những thông tin chính xác, đầy đủ theo chủ đề đã định và mong muốn sử dụng bảng hỏi cũng phải phù hợp với trình độ, khả năng của người trả lời để họ có thể dễ dàng hiểu chính xác từng câu hỏi và đưa ra những câu hỏi trả lời thật sự khách quan. Muốn vậy, bảng các câu hỏi phải phản ánh đầy đủ nội dung điều tra. Việc thiết kế phải đảm bảo các yêu cầu về mỹ thuật, tiết kiệm và tiện dụng. Về mặt kỹ thuật, phiếu hỏi phải được thiết kế đẹp, dễ đọc, có khả năng lôi kéo, duy trì sự quan tâm của người trả lời. Việc sắp xếp các hàng, các cột, bố trí khổ giấy... sao cho phải đảm bảo yêu cầu tiết kiệm, nhưng lại thuận lợi cho việc ghi chép, mã hoá, nhập số liệu và kiểm tra lại sau này. Các câu hỏi cũng phải được sắp xếp theo trình tự logic nhất định. Từng câu hỏi phải rõ ràng, đơn giản, dễ hiểu và đảm bảo cho mọi người đều hiểu theo một nghĩa thống nhất.

2.5.2 Các loại câu hỏi và kỹ thuật đặt các loại câu hỏi

1. Câu hỏi theo nội dung

Cơ sở để phân loại câu hỏi theo nội dung là thực tế kinh tế - xã hội mà các câu hỏi này đề cập đến và có thể truyền tải được. Đó chính xác là các yếu tố, các khía cạnh, các quá trình kinh tế - xã hội ta muốn thu thập thông tin qua các câu trả lời của đối tượng. Do đó, theo nội dung, ta có thể chia chi tiết thành các câu hỏi về kinh tế, về văn hoá, về xã hội. Về kinh tế, có thể phân biệt các câu hỏi về từng ngành: Công nghiệp, nông nghiệp, về xây dựng cơ bản,... Tuy nhiên, theo cách hiểu chung hơn, ta có phân biệt hai nhóm câu hỏi theo nội dung như sau:

Nhóm thứ nhất: Câu hỏi về sự kiện

Là những câu hỏi về một sự kiện hiện thực nào đó đã và đang tồn tại trong không gian, thời gian. Những câu hỏi này được đặt ra nhằm để nắm tình hình hiện thực khách quan, bao gồm cả tình hình về đối tượng điều tra.

Ví dụ: Bạn có phải là sinh viên Học viện công nghệ BCVT không? Bạn đã xây dựng gia đình chưa? Hoặc những câu hỏi về đặc trưng của người trả lời như: giới tính, độ tuổi, trình độ văn hoá, nghề nghiệp, thành phần gia đình...

Loại câu hỏi này thường có mặt trong hầu hết các phiếu điều tra, các bảng hỏi điều tra thống kê cũng như điều tra xã hội học, nếu thiếu nó, cuộc nghiên cứu thực tế sẽ khó có ý nghĩa đầy đủ. Ngoại trừ những câu hỏi có tính chất riêng tư, câu hỏi về sự kiện, nói chung thường dễ trả lời. Vì vậy, người ta thường dùng câu hỏi về sự kiện để bắt đầu hỏi trong phỏng vấn nhằm giúp người trả lời quen dần với cuộc tọa đàm hoặc để chuyển ý giữa những câu hỏi về thái độ, quan điểm, động cơ... Thông tin thu được từ những câu hỏi này thường có độ tin cậy và độ xác thực so với các câu hỏi về nội dung khác.

Cần chú ý: Với câu hỏi về những sự kiện xảy ra trong quá khứ quá xa, cần đề phòng những sai lầm có thể xảy ra do đối tượng đã bị quên. Có thể giúp đỡ người trả lời bằng cách gợi ý lại bối cảnh lịch sử, địa lý mà sự kiện đó xảy ra để họ tái hiện thông tin cần thiết. Với những câu hỏi mang tính chất riêng tư, kín đáo hoặc có thể gợi cho đối tượng những ký ức không vui, những câu hỏi thuộc những chủ đề nhạy cảm... thì cần có cách diễn đạt tế nhị, hoặc sử dụng cách hỏi gián tiếp để tránh cho đối tượng phải gặp những tình huống khó xử, bối rối và làm ảnh hưởng đến độ chính xác của câu trả lời.

Nhóm thứ hai:

Nhóm thứ hai bao gồm những câu hỏi dùng để đo lường mức độ của vấn đề nghiên cứu như: Trạng thái của hiện tượng, trình độ nhận thức, mong muốn, thái độ, tình cảm, động cơ... Tuy nhiên, tính chuẩn xác của các câu trả lời đối với loại câu hỏi này phụ thuộc chặt chẽ vào trình độ, khả năng nhận thức, đánh giá và mong muốn của người trả lời. Ví dụ: “Bạn hiểu gì về chuyên ngành đào tạo mà bạn đang theo học?”, “Trình độ tiếng Anh của bạn ở mức độ nào?”...

Việc phân chia hai nhóm câu hỏi như trên có liên quan đến phương án thu thập thông tin. Đối với nhóm thứ nhất, các sự kiện được hỏi luôn tồn tại và gắn liền với hiện thực cuộc sống, chúng có tính chất khách quan, nên thường dễ trả lời, thông tin thu được thường có độ chính xác cao và ổn định. Đối với nhóm thứ hai, nó liên quan đến sự đánh giá chủ quan, vào nhận thức của mỗi cá nhân trả lời. Nó đòi hỏi người trả lời không chỉ “biết” mà còn phải “hiểu” vấn đề. Mặt khác, những mong muốn, đánh giá thường nằm trong ý thức của cá nhân con người, dễ bị thay đổi và đôi khi gặp phải những khó khăn nhất định trong việc diễn đạt. Hơn nữa, nhiều đánh giá, nhận định liên quan đến vấn đề riêng tư, những vấn đề nhạy cảm biểu hiện lập trường, quan điểm làm cho việc trả lời gặp nhiều khó khăn, không chính xác. Thậm chí có trường hợp ta không thể thu được thông tin. Vì vậy, khi sử dụng nhóm câu hỏi thứ hai này, cần đặc biệt quan tâm đến cách đặt câu hỏi, sử dụng từ ngữ, hình thức câu hỏi...

2. Câu hỏi chức năng

Trên thực tế, để chuyển tải những nội dung của thông tin trong điều tra, đặc biệt là khi sử dụng phương pháp phỏng vấn cần phải có những câu hỏi mang tính chất kỹ thuật, đó là các câu hỏi chức năng

Câu hỏi tâm lý:

Câu hỏi tâm lý có thể là câu hỏi tiếp xúc để gạt bỏ những nghi ngờ có thể nảy sinh, để giảm bớt sự căng thẳng, hoặc chuyển từ chủ đề này sang chủ đề khác... thường dùng trong phỏng vấn trực tiếp.

- Để tiếp xúc, làm quen, người ta thường đưa ra câu hỏi rất đơn giản, dễ trả lời, thậm chí có thể đã biết trước câu trả lời. Mục đích của việc đặt những câu hỏi này không nhất

thiết là để thu thập thông tin, mà chủ yếu là mang tính chất làm quen, tạo không khí thoải mái, gạt bỏ nghi ngờ để thúc đẩy người được hỏi và lôi cuốn họ vào vấn đề nghiên cứu.

- Để giảm bớt sự căng thẳng, có thể biểu thị sự quan tâm tới người được hỏi, người thường đặt chen vào giữa những câu hỏi nội dung một số câu hỏi về đời sống hàng ngày, về gia đình, về những nét đẹp trong truyền thống của địa phương... rồi sau đó tiếp tục trở lại vấn đề cần hỏi.

- Để chuyển từ nội dung này sang nội dung khác khá cách biệt nhau, có thể sử dụng những câu chuyển tiếp để làm cho người trả lời không cảm thấy bị hụt hẫng hoặc không bị đảo lộn về tư duy, không cảm thấy nghi ngờ, thắc mắc về tính hợp lý của các câu hỏi. Những chiếc “cầu nối” này làm trật tự các câu hỏi trở nên nhịp nhàng hơn, hấp dẫn người trả lời hơn. Chẳng hạn: “Bây giờ ta bắt đầu vào đề...”, hoặc “Còn bây giờ ta nói chuyện một chút về vấn đề...”. Việc chuyển tiếp cần nhịp nhàng, tế nhị và đặc biệt không gây nên sự thắc mắc về tính hợp lý của tiến trình câu chuyện.

Những câu hỏi tâm lý có thể không có liên quan rõ ràng đến nội dung nghiên cứu, vì vậy sự dụng nó phải khéo léo và có mức độ. Có thể vì nó như gia vị trong một món ăn, nếu sử dụng nhiều hoặc không đúng chỗ có thể gây phản tác dụng. Chẳng hạn, cần tránh những câu hỏi về nhân khẩu học, các câu hỏi gây phiền hà về nguyên tắc, những câu hỏi về quan hệ nội bộ trong tập thể, những xích mích trong gia đình, mất mát, tổn thương... mà họ không muốn nhắc đến.

Câu hỏi lọc:

Câu hỏi lọc có tác dụng tìm hiểu xem người được hỏi có thuộc nhóm người dành cho những câu hỏi tiếp sau có được trả lời cũng không có giá trị đích thực, thậm chí bị xuyên tạc. Câu hỏi lọc có thể dùng trước khi tiến hành phỏng vấn hoặc trước khi tiếp vào một nội dung nào đó. Như vậy, loại câu hỏi này có tác dụng phân chia những người tham gia trả lời thành các nhóm khác nhau, để sau đó sẽ có các câu hỏi dành riêng cho mỗi nhóm.

Khi xây dựng bảng hỏi, người ta thường sử dụng kỹ thuật “bước nhảy”: với ý nghĩa cho phép “chuyển đến” hoặc “được phép chuyển đến” vì nếu không chuyển ngay vào mà vẫn theo trình tự bình thường thì các câu hỏi trả lời sẽ không có giá trị mục đích thực hay nói khác đi là vô nghĩa đối với người đang hỏi, thậm chí có thể làm sai lệch kết quả nghiên cứu.

Câu hỏi kiểm tra:

Loại câu hỏi này có tác dụng kiểm tra độ chính xác của những thông tin thu thập được. Nó được sử dụng khi gặp một câu trả lời nghi ngờ về tính xác thực, độ chính xác, người ta có thể đặt một câu hỏi khác, theo cách diễn đạt khác có liên quan đến vấn đề bị nghi ngờ nhằm kiểm tra lại thông tin trên.

Câu hỏi kiểm tra có thể được sử dụng để kiểm tra những câu trả lời về một vấn đề, xác định mức tin cậy đối với từng câu hỏi hoặc toàn bộ bảng hỏi. Phương án thực hiện có thể khác nhau, rất linh hoạt. Chẳng hạn có thể nêu câu hỏi tiếp sau đó đưa ra phương án trả lời mà trong đó đã gài một câu biết chắc chắn đúng một trăm phần trăm (hoặc không có thật) để thử. Người ta cũng có thể đưa ra một tình tiết, một vấn đề hoàn toàn hư cấu, nhưng về hình thức lại rất gần gũi với thực tế để xem đối tượng có nhận ra được không, qua đó để kiểm tra tính xác thực của các thông tin đã được cung cấp trước đó...

Trong thực tế, có nhiều vấn đề có liên quan chặt chẽ với nhau, mà khi trả lời người ta ít để ý đến. Ví dụ, thâm niên công tác có liên quan đến tuổi, mức lương... của người trả lời.

Vì vậy, ta có thể sử dụng chính các câu hỏi đã có, kiểm tra chéo lẫn nhau để xác định tính chân thực của các thông tin đã được cung cấp.

Một số vấn đề cần chú ý khi sử dụng các câu hỏi kiểm tra:

- Các câu hỏi kiểm tra thường làm tăng số câu hỏi cho một cuộc nghiên cứu. Nếu sử dụng quá nhiều sẽ ảnh hưởng đến chất lượng của cuộc điều tra. Vì vậy, chỉ nên sử dụng câu hỏi kiểm tra đối với nội dung nghiên cứu chủ yếu và hay gặp sai sót. Đồng thời nó cũng phải được chuẩn bị, kiểm nghiệm trong thực tiễn là có thể dùng để kiểm tra được.

- Câu hỏi kiểm tra không bao giờ được đi liền với câu trả lời mà nó kiểm tra và thường đặt xa khoảng 3 đến 4 câu hỏi khác nếu có thể phát sinh sự nghi ngờ ở người được hỏi làm ảnh hưởng đến tiến trình phỏng vấn tiếp theo.

- Sử dụng câu hỏi kiểm tra cũng phải hết sức khéo léo, cần tránh cho người trả lời cảm giác tự ái, chạm lòng, cảm thấy bị xúc phạm khi phải trả lời câu hỏi này. Điều đó ảnh hưởng đến việc cung cấp thông tin ở các câu hỏi tiếp theo.

3. Câu hỏi theo cách biểu hiện

Có thể phân loại câu hỏi theo hình thức biểu hiện của câu trả lời cũng như cách hỏi:

Theo biểu hiện của câu trả lời

Mở rộng bảng hỏi ra, điều dễ nhận thấy nhất là hình thức biểu hiện câu trả lời. Tùy theo cách thức đưa ra câu trả lời, người ra chia thành loại câu hỏi đóng, câu hỏi mở và kết hợp giữa hai loại này.

Câu hỏi đóng

Là dạng câu hỏi đã có trước những phương án trả lời cụ thể mà người trả lời chỉ việc chọn một hoặc số phương án họ cho là phù hợp nhất.

Ta có thể phân biệt hai loại câu hỏi đóng khác nhau:

Câu hỏi đóng lựa chọn (câu hỏi loại trừ). Đặc điểm chính của loại câu hỏi là các phương án trả lời được đưa ra mang tính chất loại trừ nhau, tức là người trả lời chỉ có thể lựa chọn một số trong các phương án trả lời được nêu ra: Khi đã chọn phương án này, thì đương nhiên, các phương án khác đã bị loại trừ.

Trong loại này, có một dạng câu hỏi khá đặc biệt là chỉ có hai phương án loại trừ nhau: Có hoặc không, đúng hoặc sai, đã hoặc chưa... Chẳng hạn: “Bạn đã lập gia đình chưa?”, “Anh chị có thích sinh con trai không?”... Trong thực tế, do đặc điểm tâm lý của con người, nêu câu trả lời có trách nhiệm hơn và muốn trả lời dễ thiên theo hướng “tích cực” của câu hỏi. Để người trả lời có trách nhiệm hơn và muốn trả lời hơn có thể thay đổi cách diễn đạt nhẹ nhàng, tế nhị hơn. Hoặc người ta có thể chuyển thành câu hỏi có nhiều phương án lựa chọn hơn. Ví dụ, câu hỏi “Ông (bà) có hài lòng với mức lương hiện nay của mình không?” với hai lựa chọn là có hoặc không được thay thế bằng câu hỏi “Mức độ hài lòng của ông (bà) về mức lương hiện tại của mình?” nhiều phương án để trả lời, lựa chọn từ rất hài lòng đến không hài lòng.

Câu hỏi đóng tùy chọn (câu hỏi tuyển): đặc điểm của loại câu hỏi này là các phương án trả lời không loại trừ nhau, người được hỏi có thể chọn một số khả năng nào đó họ cho là thích hợp nhất.

Ưu điểm của câu hỏi đóng:

Loại câu hỏi này tạo điều kiện thuận lợi, dễ dàng cho người trả lời vì chỉ cần lựa chọn một hoặc một số phương án được nêu ra. Mặt khác còn giúp cho người được hỏi hiểu câu hỏi một cách thống nhất, theo cùng một nghĩa. Hơn nữa, nó cũng làm cho việc xử lý thông tin thống kê, đo lường hiện tượng nghiên cứu được thuận lợi, rõ ràng hơn. Vì vậy, loại câu hỏi này luôn chiếm vị trí chủ đạo trong bảng hỏi của các cuộc điều tra thống kê.

Hạn chế lớn nhất của câu hỏi loại này là bó hẹp tư duy, gò ép đối tượng theo cách lập luận chủ quan của người nghiên cứu với phương án trả lời được nêu ra. Điều đó, hạn chế khả năng sáng tạo và hướng tư duy, suy nghĩ của đối tượng. Nếu không cẩn thận sẽ vi phạm nguyên tắc khách quan, khoa học. Khi sử dụng phải chú ý:

- Phải đặt mình vào vị trí của người được hỏi sẽ hiểu vấn đề và trả lời như thế nào, từ đó có cách diễn đạt, sắp xếp câu trả lời.
- Phải lường trước các phương án trả lời có thể có. Tức là tất cả các khía cạnh, các yếu tố của vấn đề được hỏi phải được thể hiện đầy đủ trong các phương án trả lời.
- Các phương án trả lời phải có mức độ đồng nhất tương đối với nhau theo cách phân chia nào đó. Việc sắp xếp trật tự các phương án trả lời cũng phải tuân theo những nguyên tắc nhất định để giúp cho người trả lời hiểu vấn đề một cách dễ dàng hơn.
- Số lượng các phương án trả lời cũng nên giới hạn trong một chừng mực nhất định. Sử dụng quá ít hoặc quá nhiều phương án trả lời đều làm việc lựa chọn gặp khó khăn mà còn làm cho việc đo lường nghiên cứu không chuẩn xác.

Câu hỏi mở

Là loại câu hỏi không có phương án trả lời được nêu trước mà nó hoàn toàn do người trả lời tự nghĩ ra. Chúng cho phép người được hỏi tự thông tin một cách tốt nhất những suy nghĩ của họ.

Tác dụng của câu hỏi mở: Để tìm hiểu vấn đề, thu thập ý kiến, quan điểm một cách đầy đủ nhất theo chủ đề hoặc trong phạm vi vấn đề đã nêu ra. Chúng có thể làm tăng tính tích cực, chủ động của người được hỏi. Vì vậy, loại câu hỏi này rất thuận tiện cho những nghiên cứu đặc tính.

Khó khăn đối với câu hỏi mở là ở vấn đề xử lý số liệu. Nguyên tắc của việc xử lý là phải tách thành các nhóm mà theo đó có thể thu thập từ những câu trả lời. Các nhóm này do người trả lời hình thành nên vì thế nhà nghiên cứu trở nên bị động hơn. Chính vì vậy, loại câu hỏi này ít được sử dụng trong các cuộc điều tra thống kê thuần túy.

Câu hỏi hỗn hợp (câu hỏi nửa đóng)

Về hình thức, đây là loại câu hỏi kết hợp giữa câu hỏi đóng và câu hỏi mở. Xét về nội dung nó được sử dụng trong những trường hợp sau:

- Khi không tìm hết được phương án diễn đạt theo câu hỏi đóng, cần để người trả lời tự diễn đạt.
- Khi chỉ cần xử lý, tổng hợp theo những phương án trả lời nhưng không để người trả lời bị rơi vào thế bị, hụt hẫng.

Theo biểu hiện của câu hỏi

Tuỳ theo cách hỏi các câu hỏi được biểu hiện dưới dạng trực tiếp hoặc gián tiếp.

- Câu hỏi trực tiếp: Là cách hỏi thẳng ngay vào nội dung vấn đề, người được hỏi không bị câu nệ và có thể trả lời vào chính nội dung đó.

- Câu hỏi gián tiếp: Là cách hỏi khôn khéo, không đi trực tiếp vào vấn đề, mà có thể hỏi vòng vo, thông qua những vấn đề có liên quan để thu thập thông tin về vấn đề cần nghiên cứu. Cách hỏi này thường được dùng cho những vấn đề mà xã hội thường gán cho nó tính “tiêu cực”, những chủ đề nhạy cảm, riêng tư sâu kín.

2.6. Sai số trong điều tra thống kê kinh tế

Các cuộc điều tra thống kê, dù có cố gắng làm thật tốt vẫn thường gặp những trường hợp mà số liệu điều tra không trùng khớp với số liệu thực tế của hiện tượng nghiên cứu. Người ta gọi là sai số. Sai số trong điều tra thống kê là chênh lệch giữa vị trí số thực của hiện tượng nghiên cứu so với trị số của nó mà điều tra thống kê thu được. Sai số này làm giảm chất lượng của các cuộc điều tra, ảnh hưởng đến chất lượng của toàn bộ quá trình nghiên cứu thống kê. Trong các cuộc điều tra thống kê, người ta phải cố gắng áp dụng nhiều biện pháp để hạn chế sai số này.

Căn cứ vào tính chất của các sai số, ta có thể phân biệt hai loại: Sai số do đăng ký và sai số do tính đại diện.

Sai số do đăng ký xảy ra đối với mọi cuộc điều tra thống kê. Nó phát sinh do việc đăng ký số liệu ban đầu không chính xác. Nguyên nhân gây ra loại sai số này rất đa dạng, có thể do cân, đo, đếm sai, tính toán sai, ghi chép sai, do dụng cụ đo lường không chuẩn xác... Nếu phân chia chi tiết hơn, ta có thể chia loại sai số này thành sai số ngẫu nhiên và sai số có hệ thống, do cố ý, có chủ định của người điều tra và người trả lời. Sai số ngẫu nhiên là những sai số phát sinh một cách tình cờ, không có chủ định, không có bất kỳ một sự sắp đặt trước nào của người điều tra. Nó xảy ra hoàn toàn ngẫu nhiên sẽ có khả năng bù trừ, triệt tiêu nhau làm cho sai số một cách có hệ thống do lỗi của hệ thống đo lường, hệ thống thang đo được thiết kế không chuẩn xác... Loại sai số này không chịu sự chi phối của quy luật số lớn, nên điều tra càng nhiều, khả năng xảy ra sai số sẽ càng lớn.

Sau số do tính đại diện chỉ xảy ra trong điều tra chọn mẫu. Nguyên nhân là do trong các cuộc điều tra này, người ta chỉ chọn một số đơn vị để điều tra thực tế. Các đơn vị này không đủ đảm bảo đại diện cho toàn bộ tổng thể, nên phát sinh sai số, ngay cả trong trường hợp việc lựa chọn số đơn vị điều tra được thực hiện một cách hoàn toàn ngẫu nhiên.

Để đảm bảo các kết quả điều tra độ chính xác cao, cần áp dụng một số biện pháp để hạn chế sai số.

- Làm tốt công tác chuẩn bị điều tra: Thông thường trong các cuộc điều tra thống kê, công tác chuẩn bị chiếm vị trí rất quan trọng, nó đòi hỏi một sự đầu tư chất xám khá lớn. Công tác chuẩn bị càng chu đáo, tỷ mỉ, thận trọng và chi tiết, đặc biệt là trong việc thiết lập phương án điều tra, xây dựng phiếu điều tra, lựa chọn và tập huấn cán bộ điều tra càng làm tốt, sai số điều tra càng giảm.

- Tiến hành kiểm tra có hệ thống toàn bộ cuộc điều tra: Kiểm tra là biện pháp có hiệu quả để sửa chữa, uốn nắn kịp thời các sai lầm có thể mắc phải trong quá trình điều tra. Việc kiểm tra có thể được tiến hành theo nhiều giai đoạn khác nhau. Trước hết, cần tiến hành kiểm tra ngay từ giai đoạn chuẩn bị xem các khâu cần chuẩn bị đã được đầy đủ, chu đáo chưa. Việc kiểm tra trong giai đoạn thu thập thông tin, việc ghi chép số liệu ban đầu nhằm nâng cao ý thức trách nhiệm của nhân viên điều tra cũng có ý nghĩa quan trọng. Nghiệm thu phiếu điều tra là một khâu kiểm tra có ý nghĩa quyết định. Trong thời gian giai đoạn này, người ta cần kiểm

tra xem các phiếu điều tra có đầy đủ không, các câu trả lời, các con số ghi chép từng phiếu có được tính toán đúng, đủ không, có hợp logic không, có mâu thuẫn với nhau không...Nhìn chung, việc kiểm tra, nghiệm thu phiếu điều tra có tác dụng rất lớn, nhưng nó đòi hỏi người kiểm tra phải có trình độ, kinh nghiệm, hiểu biết thực tế và rất nhạy cảm. Tiếp theo, việc nhập số liệu vào máy cũng cần được kiểm tra thật kỹ lưỡng. Thực tế cho thấy đây cũng là một khâu dễ làm phát sinh sai số. Nhiều cuộc điều tra, người ta yêu cầu nhập hai lần độc lập nhau, để khắc phục những sai sót có thể xảy ra trong quá trình nhập số liệu.

Ngoài ra, trong các cuộc điều tra không toàn bộ, người ta còn tiến hành kiểm tra tính đại diện của các đơn vị được chọn để điều tra.

TÀI LIỆU THAM KHẢO CHƯƠNG 2

1. PGS. TS. Trần Ngọc Phác, TS Trần Thị Kim Thu – **Giáo trình Lý thuyết thống kê**. Nhà xuất bản Thống kê, 2013
2. GS.TS Bùi Xuân Phong - **Thống kê và ứng dụng trong BCVT**. NXB Bưu điện, 2005
3. Hà Văn Sơn - **Giáo trình Lý thuyết thống kê ứng dụng trong quản trị và kinh tế**. Nhà xuất bản Thống kê, 2004
4. PTS. Phan Công Nghĩa - **Giáo trình Thống kê Lao động**. NXB Thống kê, 2009
5. PGS.PTS. Nguyễn Thiệp, PTS Phan Công Nghĩa - **Giáo trình Thống kê chất lượng**. NXB Thống kê, 2011

CÂU HỎI ÔN TẬP CHƯƠNG 2

1. Phân tích các yêu cầu cơ bản của điều tra thống kê kinh tế. Cho ví dụ minh họa?
2. Phân biệt các loại điều tra thống kê. Cho ví dụ minh họa?
3. Trình bày những nội dung chủ yếu của một phương án điều tra thống kê.
4. Sử dụng các loại câu hỏi để lập bảng hỏi đơn giản cho một cuộc điều tra.

CHƯƠNG 3

PHÂN TỔ TỔNG HỢP TÀI LIỆU ĐIỀU TRA THỐNG KÊ KINH TẾ

3.1. Khái niệm chung về phân tổ thống kê

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

Ta đã biết, hiện tượng và quá trình kinh tế - xã hội mà thống kê học nghiên cứu thường rất phức tạp. Vì chúng tồn tại và phát triển dưới nhiều loại hình có quy mô và đặc điểm khác nhau. Trong kết cấu nội bộ của hiện tượng nghiên cứu thường bao gồm nhiều tổ, nhiều bộ phận có tính chất khác nhau. Muốn phản ánh được bản chất và quy luật phát triển của hiện tượng, nếu chỉ dựa vào những con số tổng cộng chung chung thì không thể nêu được vấn đề một cách sâu sắc. Phải tìm cách nêu lên được đặc trưng của loại hình, của từng bộ phận cấu thành hiện tượng phức tạp, đánh giá tầm quan trọng của mỗi bộ phận. Nêu lên mối liên hệ giữa các bộ phận rồi từ đó nhận thức được các đặc trưng chung của toàn bộ. Yêu cầu nói trên chỉ có thể giải quyết được bằng phương pháp phân tổ thống kê.

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

Chẳng hạn, khi nghiên cứu tình hình nhân khẩu, căn cứ vào tiêu thức “giới tính” để chia tổng số nhân khẩu thành hai tổ: Nam và Nữ; còn căn cứ vào tiêu thức “tuổi” để chia số nhân khẩu thành nhiều tổ có độ tuổi khác nhau. Hay khi nghiên cứu tình hình sản xuất của các doanh nghiệp công nghiệp, có thể chia tổng số doanh nghiệp thành các nhóm theo các tiêu thức như: thành phần kinh tế, ngành sản xuất, số lượng lao động, giá trị sản xuất công nghiệp...

Khi phân tổ thống kê, trước hết ta thực hiện được việc nghiên cứu cái chung và cái riêng một cách kết hợp. Các đơn vị tổng thể được tập hợp lại thành một số tổ (và tiểu tổ): giữa các tổ đều có sự khác nhau rõ rệt về tính chất, còn trong phạm vi mỗi tổ các đơn vị đều có sự giống nhau (hay gần giống nhau) về tính chất theo tiêu thức được dùng làm căn cứ phân tổ.

Phân tổ thống kê là phương pháp cơ bản để tiến hành tổng hợp thống kê, vì ta sẽ không thể tiến hành hệ thống hoá một cách khoa học các tài liệu điều tra, nếu không áp dụng phương pháp này. Tính chất phức tạp của hiện tượng nghiên cứu đòi hỏi phải tổng hợp theo từng tổ, từng bộ phận. Vì vậy, khi tổng hợp thống kê, trước hết, người ta thường sắp xếp các đơn vị vào từng tổ, từng bộ phận, tính toán các đặc điểm của mỗi tổ hoặc bộ phận, rồi sau đó mới tính các đặc điểm chung của cả tổng thể.

Phân tích thống kê là một trong các phương pháp quan trọng của phân tích thống kê, đồng thời là cơ sở để vận dụng các phương pháp phân tích thống kê khác. Chỉ sau khi đã phân chia tổng thể nghiên cứu thành các tổ có quy mô và đặc điểm khác nhau, việc tính các chỉ tiêu phản ánh mức độ, tình hình biến động, mối liên hệ giữa các hiện tượng mới có ý nghĩa đúng đắn. Nếu việc phân tổ không chính xác, tổng thể được chia thành những bộ phận không đúng với thực tế, thì mọi chỉ tiêu tính ra cũng không giúp ta rút ra được những kết luận đúng đắn. Phương pháp phân tổ được vận dụng phổ biến nhất trong mọi trường hợp nghiên cứu kinh tế, vì không những phương pháp này đơn giản, dễ hiểu mà lại có tác dụng phân tích

sâu sắc. Các phương pháp phân tích thống kê khác như: phương pháp số tương đối, phương pháp số bình quân, phương pháp chỉ số phương pháp bảng cân đối, phương pháp tương quan... thường cũng phải dựa trên các kết quả phân tổ thống kê chính xác.

Phân tổ thống kê còn được vận dụng ngay trong giai đoạn điều tra thống kê, nhằm phân tổ đối tượng điều tra thành những bộ phận có đặc điểm, tính chất khác nhau từ đó chọn các đơn vị điều tra sao cho có tính đại biểu cho tổng thể chung.

Phân tổ thống kê giải quyết những nhiệm vụ nghiên cứu cơ bản sau đây:

Thứ nhất, Phân bố thực hiện việc phân chia các loại hình kinh tế - xã hội của hiện tượng nghiên cứu. Hiện tượng kinh tế - xã hội mà thống kê học nghiên cứu thường không phải là tổng thể đồng chất, mà là tổng thể bao gồm nhiều đơn vị, thuộc các loại hình rất khác nhau, phát triển theo những xu hướng không giống nhau. Vì vậy phương pháp nghiên cứu khoa học là phải nêu lên các đặc trưng riêng biệt của từng loại hình và mối quan hệ giữa các loại hình đó với nhau. Muốn vậy, trước hết phải dựa trên lý luận kinh tế, chính trị, xã hội để phân biệt các bộ phận khác nhau về tính chất đang tồn tại khách quan trong nội bộ hiện tượng.

Thứ hai, Phân tổ có nhiệm vụ biểu hiện kết cấu của hiện tượng nghiên cứu. Ta biết rằng một hiện tượng kinh tế - xã hội do nhiều bộ phận, nhiều nhóm đơn vị có tính chất khác nhau hợp thành. Các bộ phận hay nhóm này chiếm những tỷ trọng khác nhau trong tổng thể và nói lên tầm quan trọng của mình trong tổng thể đó. Mặt khác, tỷ trọng của các bộ phận còn nói lên kết cấu của tổng thể theo một tiêu thức nào đó. Muốn nghiên cứu được kết cấu của tổng thể, phải dựa trên cơ sở phân tổ thống kê.

Thứ ba, Phân tổ được dùng để biểu hiện mối liên hệ giữa các mục của tiêu thức. Hiện tượng kinh tế - xã hội phát sinh và biến động không phải một cách ngẫu nhiên, tách rời với các hiện tượng xung quanh, mà chúng có liên hệ và phụ thuộc lẫn nhau theo những quy luật nhất định. Giữa các tiêu thức mà thống kê nghiên cứu cũng thường có mối liên hệ với nhau: sự thay đổi tiêu thức này sẽ đưa đến sự thay đổi của tiêu thức kia theo một quy luật nhất định. Tìm hiểu tính chất và trình độ của mối liên hệ giữa các hiện tượng nói chung và giữa các tiêu thức nói riêng là một trong các nhiệm vụ quan trọng của nghiên cứu thống kê. Phân tổ thống kê là một trong các phương pháp có thể giúp ta thực hiện nhiệm vụ nghiên cứu này.

3.1.2 Các loại phân tổ thống kê

Thống kê thường có các cách phân loại phân tổ thống kê như sau:

1. Căn cứ vào nhiệm vụ của phân tổ thống kê

Tương ứng với ba nhiệm vụ nói trên của phân tổ thống kê, có ba loại phân tổ: Phân tổ phân loại; phân tổ kết cấu và phân tổ liên hệ.

** Phân tổ phân loại:*

Phân tổ phân loại giúp ta nghiên cứu một cách có phân biệt loại hình kinh tế - xã hội, nêu lên đặc trưng và mối quan hệ giữa chúng với nhau. Từ việc nghiên cứu riêng biệt mỗi loại hình đó, tiến thêm một bước nghiên cứu các đặc trưng của toàn bộ hiện tượng phức tạp, giải thích một cách sâu sắc bản chất và xu hướng phát triển của hiện tượng trong điều kiện thời gian và địa điểm cụ thể.

Tuỳ theo mục đích nghiên cứu, có thể phân loại các đơn vị theo tiêu thức khác nhau. Chẳng hạn, các doanh nghiệp công nghiệp nước ta có thể được phân loại theo thành phần kinh tế, theo cấp quản lý, theo nhóm, theo ngành, theo quy mô...

** Phân tổ kết cấu:*

Trong công tác nghiên cứu thống kê, các bảng phân tổ kết cấu được sử dụng rất phổ biến, nhằm mục đích nêu lên bản chất của hiện tượng trong điều kiện nhất định và để nghiên cứu xu hướng phát triển của hiện tượng qua thời gian. Kết cấu của tổng thể phản ánh một trong các đặc trưng cơ bản của hiện tượng trong điều kiện thời gian và địa điểm cụ thể. Sự thay đổi kết cấu của tổng thể qua thời gian có thể giúp ta thấy được xu hướng phát triển của hiện tượng. Chẳng hạn sự thay đổi kết cấu về tổng sản phẩm trong nước phân theo nhóm ngành phản ánh sự chuyển dịch cơ cấu ngành kinh tế trong quá trình phát triển của một quốc gia như bảng 3.1.

Phân bổ kết cấu giúp ta có thể so sánh được bản chất của các hiện tượng cũng loại trong điều kiện không gian khác nhau. Ví dụ, có thể so sánh cơ cấu công nhân của hai nhà máy, cơ cấu giống lúa của hai hợp tác xã. Phân tổ kết cấu còn được vận dụng trong phân tích thực hiện kế hoạch để thấy rõ tỷ trọng các bộ phận chưa hoàn thành, hoàn thành và hoàn thành vượt mức kế hoạch. Từ đó có thể đánh giá việc thực hiện kế hoạch; xem xét lại đạt kế hoạch như vậy có hợp lý không và có thể tính được khả năng tiềm tàng vượt mức kế hoạch, trên cơ sở kết hợp với các giả thiết khác.

Trong nhiều trường hợp nghiên cứu, phân tổ kết cấu có thể được xác định ngay trên cơ sở phân tổ phân loại, như vậy là hai loại phân tổ này thường kết hợp chặt chẽ với nhau. Mặt khác, ngay cả đối với một tổng thể đồng chất cũng vẫn thường bao gồm các bộ phận khác nhau do nhiều nguyên nhân cụ thể, cho nên vẫn cần phân tổ kết cấu. Như tổng thể công nhân thuộc cùng một nghề trong cùng một doanh nghiệp, số công nhân này vẫn khác nhau về giới tính, về tuổi nghề, về bậc thợ và về nhiều đặc điểm khác nhau. Như vậy là phân tổ kết cấu cần thiết đối với bất kỳ công tác nghiên cứu thống kê nào.

** Phân tổ liên hệ*

Khi tiến hành phân tổ liên hệ, các tiêu thức có liên hệ với nhau được phân biệt thành hai loại: tiêu thức nguyên nhân và tiêu thức kết quả. Tiêu thức nguyên nhân là tiêu thức gây ảnh hưởng; sự biến động của tiêu thức này sẽ dẫn đến sự thay đổi (tăng hoặc giảm) của tiêu thức phụ thuộc mà ta gọi là tiêu thức kết quả - một cách có hệ thống. Như vậy, các đơn vị tổng thể trước hết được phân tổ theo một tiêu thức (thường là tiêu thức nguyên nhân), sau đó trong mỗi tổ tiếp tục tính các trị số bình quân của tiêu thức còn lại (thường là tiêu thức kết quả). Quan sát sự biến thiên của hai tiêu thức. Như trong nhiều doanh nghiệp công nghiệp, ta thường nhận thấy mối liên hệ giữa năng suất lao động càng tăng thì giá thành đơn vị sản phẩm càng có điều kiện giảm. Nếu ta phân tổ các doanh nghiệp trong cùng một ngành theo năng suất lao động, sau đó từ mỗi tổ tính ra giá thành bình quân đơn vị sản phẩm, thì các kết quả tính toán sẽ cho thấy rõ mối liên hệ giữa năng suất lao động và giá thành đơn vị sản phẩm.

Phân tổ liên hệ còn có thể được vận dụng để nghiên cứu mối liên hệ giữa nhiều tiêu thức. Có thể nghiên cứu mối liên hệ giữa năng suất lúa với lượng phân bón, lượng nước tưới, mật độ cấy...; hoặc nghiên cứu mối liên hệ giữa năng suất lao động của công nhân với tuổi nghề, bậc thợ, trình độ trang bị kỹ thuật...

Khi phân tổ liên hệ giữa nhiều tiêu thức trước hết tổng thể được chia thành các tiểu tổ theo tiêu thức nguyên nhân thứ hai, cuối cùng tính trị số tổng hoặc bình quân của tiêu thức kết quả cho từng tổ và tiểu tổ đó. Sau đây là ví dụ về mối liên hệ giữa năng suất lao động với

trình độ kỹ thuật và tuổi nghề của công nhân trong một doanh nghiệp, được trình bày thành bảng phân tổ kết hợp như sau:

2. Căn cứ vào số lượng tiêu thức của phân tổ

Theo định nghĩa phân tổ thống kê, chúng ta có thể căn cứ vào một hay một số tiêu thức để tiến hành phân tổ. Vì vậy, có thể phân thành hai loại: phân tổ một tiêu thức và phân tổ theo nhiều tiêu thức.

Phân tổ theo một tiêu thức:

Là tiến hành phân chia các đơn vị thuộc hiện tượng nghiên cứu thành các tổ có tính chất khác nhau trên cơ sở một tiêu thức thống kê hay còn gọi là phân tổ giản đơn. Chẳng hạn, theo tiêu thức giới tính, tổng thể dân số được chia thành 2 tổ: nam và nữ hoặc theo tiêu thức thành phần kinh tế, tổng sản phẩm được chia thành các tổ tương ứng với các thành phần kinh tế...

Phân tổ theo nhiều tiêu thức:

Là tiến hành phân chia các đơn vị thuộc hiện tượng nghiên cứu thành các tổ và các tiểu tổ có tính chất khác nhau trên cơ sở nhiều tiêu thức thống kê (từ hai tiêu thức trở lên). Tùy thuộc vào mục đích nghiên cứu, đặc điểm của hiện tượng và các tiêu thức phân tổ mà phân tổ theo nhiều tiêu thức được chia thành hai loại: Phân tổ kết hợp và phân tổ nhiều chiều.

- *Phân tổ kết hợp* là tiến hành phân tổ lần lượt theo từng tiêu thức một. Các tiêu thức được sắp xếp theo thứ tự phù hợp với mục đích nghiên và đặc điểm của hiện tượng. Thông thường người ta hay phân tổ theo tiêu thức liên quan trực tiếp đến mục đích nghiên cứu và có ít biểu hiện trước. Chẳng hạn, tổng thể dân số trước hết được phân tổ theo tiêu thức giới tính, sau đó theo tiêu thức độ tuổi và đó là cơ sở để xây dựng tháp dân số, hoặc phân tổ tổng thể một loại lao động nào đó của một doanh nghiệp theo mức lương và số năm kinh nghiệm. Tuy nhiên theo cách này số liệu tiêu thức phân tổ không nên quá nhiều vì nếu như vậy sẽ chia tổng thể thành quá nhiều bộ phận nhỏ có thể gây khó khăn cho việc phân tích.

- *Phân tổ nhiều chiều* là cũng một lúc phân tổ theo nhiều tiêu thức khác nhau nhưng vai trò như nhau trong việc đánh giá hiện tượng. Chẳng hạn, để phản ánh quy mô của một doanh nghiệp có thể biểu hiện qua các tiêu thức: doanh thu, số lượng lao động, tổng vốn... Các tiêu thức này khác nhau về số lượng và đơn vị tính nhưng đều biểu hiện quy mô của doanh nghiệp và việc sắp xếp thứ tự trước sau các tiêu thức này trong phân tổ các doanh nghiệp trong ngành là không có ý nghĩa. Vì vậy, phải cũng một lúc phân tổ theo tất cả các tiêu thức bằng cách đưa các tiêu thức này về một tiêu thức tổng hợp gọi chung là phân tổ nhiều chiều.

3.1.3 Tiêu thức phân tổ và chỉ tiêu giải thích

1. Tiêu thức phân tổ

Tiêu thức phân tổ là tiêu thức được chọn làm căn cứ để tiến hành phân tổ thống kê. Lựa chọn tiêu thức phân tổ là vấn đề quan trọng đầu tiên phải đề ra và giải quyết chính xác. Tuy các đơn vị tổng thể có rất nhiều tiêu thức khác nhau, nhưng chúng ta không thể tùy tiện chọn bất kỳ tiêu thức nào làm căn cứ phân tổ.

Tiêu thức phân tổ khác nhau sẽ nói lên những mặt khác nhau của hiện tượng. Các tiêu thức phân tổ nói rõ được bản chất của hiện tượng, nhưng cũng có tiêu thức, nếu được chọn làm căn cứ phân tổ, sẽ không đáp ứng mục đích nghiên cứu, thậm chí còn làm cho ta hiểu sai lệch bản chất của hiện tượng. Bởi vì cũng những tài liệu như nhau mà cách sắp xếp khác

nhau, lại đưa đến những kết luận trái ngược hẳn với nhau. Như vậy, việc phân tổ chính xác và khoa học trước hết phụ thuộc vào việc lựa chọn tiêu thức phân tổ. Có thể nêu ra những yêu cầu sau đây về lựa chọn tiêu thức phân tổ.

Thứ nhất, phải dựa trên cơ sở phân tích lý luận một cách sâu sắc 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, phản ánh đặc trưng cơ bản của hiện tượng trong điều kiện thời gian và địa điểm cụ thể. Bản chất của mỗi hiện tượng có thể được phản ánh qua nhiều tiêu thức khác nhau, cho nên phải tùy theo mục đích nghiên cứu mà dùng lý luận để chọn ra tiêu thức bản chất. Chẳng hạn, muốn phân tổ các doanh nghiệp sản xuất công nghiệp để biểu hiện quy mô lớn nhỏ, ta phải căn cứ vào thực tế của các doanh nghiệp đó, để xem xét tiêu thức nào có khả năng phản ánh quy mô của chúng như: số lượng lao động, giá trị sản xuất, giá trị thiết bị chủ yếu, diện tích doanh nghiệp... Đối với những doanh nghiệp trong quá trình sản xuất chủ yếu còn dựa vào sức lao động để tiến hành phân tổ, vì số lượng lao động nhiều hay ít sẽ nói lên quy mô của doanh nghiệp lớn hay nhỏ. Nhưng đối với doanh nghiệp mà quá trình sản xuất được cơ giới hoá hoặc tự động hoá cao, thì muốn biểu hiện quy mô của chúng phải tuân theo tổ theo các tiêu thức như: giá trị sản xuất, giá trị thiết bị sản xuất chủ yếu... Đó là các tiêu thức bản chất nhất, có thể đáp ứng được mục đích nghiên cứu.

Thứ hai, phải căn cứ vào điều kiện lịch sử cụ thể của hiện tượng nghiên cứu để chọn ra tiêu thức phân tổ thích 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. Nếu chỉ dùng một tiêu thức phân tổ chung cho mọi trường hợp, thì tiêu thức đó trong điều kiện này có thể giúp ta nghiên cứu chính xác, nhưng trong điều kiện khác lại không có tác dụng gì cả.

Thứ ba, phải tùy theo mục đích nghiên cứu và điều kiện tài liệu thực tế mà quyết định phân tổ hiện tượng theo một hay nhiều tiêu thức.

Nói chung, hiện tượng nghiên cứu thường phức tạp, cho nên việc phân tổ theo một tiêu thức, dù là tiêu thức bản chất nhất cũng chỉ phản ánh được mặt nào đó của hiện tượng. Nếu phân bố kết hợp theo nhiều tiêu thức, sẽ phản ánh được nhiều mặt khác nhau của hiện tượng, các mặt này có thể bổ sung cho nhau và giúp cho việc nghiên cứu thêm sâu sắc. Trong nhiều trường hợp phân tổ kết hợp giúp ta nghiên cứu mối liên hệ giữa các tiêu thức. Ví dụ: Có thể phân tổ nhân khẩu theo giới tính và theo độ tuổi, phân tổ doanh nghiệp theo nhóm, theo ngành, và theo thành phần kinh tế. Tuy nhiên, cũng cần chú ý là không nên phân tổ kết hợp theo quá nhiều tiêu thức vì làm như vậy số tổ và tiểu tổ sẽ tăng lên nhiều, tổng thể bị chia nhỏ nhiều quá sẽ trở ngại cho việc nghiên cứu. Thường người ta chỉ phân tổ kết hợp theo hai hay ba tiêu thức và nếu cần thiết, có thể lập nhiều bảng phân tổ kết hợp khác nhau.

2. Các chỉ tiêu giải thích

Trong phân tổ thống kê, sau khi đã lựa chọn được tiêu thức phân tổ, xác định các chỉ tiêu giải thích để nói rõ đặc trưng của các tổ cũng như của toàn bộ tổng thể. Chẳng hạn, sau khi phân tổ các doanh nghiệp công nghiệp theo khu vực và thành phần kinh tế, có thể đưa ra số chỉ tiêu giải thích như sau: số lao động, doanh thu thuần, doanh thu bình quân trên một lao động.

Mỗi chỉ tiêu giải thích có ý nghĩa quan trọng riêng giúp chúng ta thấy rõ các đặc trưng số lượng của từng tổ cũng như của toàn bộ tổng thể, làm căn cứ để so sánh các tổ với nhau và để tính ra hàng loạt chỉ tiêu phân tích khác. Tuy nhiên, cũng không nên đề ra quá nhiều chỉ tiêu, mà phải lựa chọn một số chỉ tiêu nào thích hợp nhất đối với mục đích nghiên cứu.

Muốn xác định các chỉ tiêu giải thích, chủ yếu phải căn cứ mục đích nghiên cứu và nhiệm vụ của phân tổ để chọn ra các chỉ tiêu có liên hệ với nhau và bổ sung cho nhau. Mục đích nghiên cứu có thể tiếp cận từ nhiều khía cạnh khác nhau nên chỉ tiêu giải thích chọn ra phải hợp lý mới thỏa mãn mục đích nghiên cứu. Phải chọn các chỉ tiêu chỉ có thể nói lên biểu hiện số lượng về một mặt nào đó của hiện tượng nghiên cứu, cho nên cần có các chỉ tiêu giải thích bổ sung cho nhau mới giúp cho việc nghiên cứu được sâu sắc.

Cũng cần chú ý tới mối quan hệ nhất định giữa tiêu thức phân tổ với các chỉ tiêu giải thích. Chẳng hạn, khi phân tổ các xí nghiệp theo quy mô, thì các chỉ tiêu giải thích như: số lượng lao động, giá trị tài sản cố định, giá trị sản xuất là những chỉ tiêu giúp ta hiểu rõ thêm về quy mô của xí nghiệp. Trái lại, nếu chọn các chỉ tiêu giải thích như: mức độ hoàn thành kế hoạch, tiền lương bình quân... thì các chỉ tiêu này thường không trực tiếp chịu ảnh hưởng bởi quy mô của xí nghiệp.

Các chỉ tiêu giải thích có ý nghĩa quan trọng trong việc so sánh với nhau cần được bố trí gần nhau. Chẳng hạn, nên bố trí chỉ tiêu thực hiện gần chỉ tiêu kế hoạch, chỉ tiêu tương đối gần chỉ tiêu tuyệt đối có liên quan...

3.2. Tiêu thức phân tổ thống kê

3.2.1 Lựa chọn tiêu thức phân tổ

Lựa chọn tiêu thức phân tổ là bước đầu tiên làm cơ sở để tiến hành phân tổ, lựa chọn tiêu thức chính xác, phù hợp với mục đích nghiên cứu thì kết quả phân tổ mới thực sự có ích cho việc phân tích đặc điểm và bản chất của hiện tượng. Việc lựa chọn phải đảm bảo các yêu cầu đã trình bày ở trên.

3.2.2 Xác định số tổ và khoảng cách tổ

Sau khi đã chọn được tiêu thức phân tổ thích hợp, vấn đề tiếp theo là xem xét cần phân chia hiện tượng nghiên cứu thành bao nhiêu tổ và căn cứ vào đâu để xác định số tổ cần thiết đó.

Số tổ cần thiết thường được xác định tùy theo tiêu thức phân tổ là tiêu thức thuộc tính hay tiêu thức số lượng. Đối với mỗi loại tiêu thức này, vấn đề xác định số tổ cần thiết được giải thích khác nhau.

1. Phân tổ theo tiêu thức thuộc tính

Khi phân tổ theo tiêu thức thuộc tính, các tổ được hình thành không phải do sự khác nhau về lượng biến của tiêu thức mà thường do các loại hình khác nhau, tuy nhiên không nhất thiết lúc nào mỗi loại hình cũng phải hình thành nên một tổ.

Trường hợp các loại hình tương đối ít thì mỗi loại hình có thể hình thành nên một tổ, như khi phân tổ tổng thể nhân khẩu theo giới tính thì sẽ chia tổng thể đó thành hai tổ là Nam và Nữ, hoặc phân tổ các doanh nghiệp theo thành phần kinh tế...

Trường hợp số loại hình thực tế nhiều, nếu coi mỗi loại hình là một tổ thì số tổ sẽ quá nhiều, không thể khái quát chung được và cũng không nêu rõ được sự khác nhau giữa các tổ, nên cần ghép những loại hình giống nhau hoặc gần giống nhau vào cùng một tổ. Chẳng hạn

khi phân tổ tổng thể nhân khẩu theo nghề nghiệp, phân tổ các loại sản phẩm công, nông nghiệp, phân tổ các mặt hàng theo giá trị sử dụng, phân tổ các ngành kinh tế quốc dân... Số tổ thực tế có thể rất nhiều, có khi tới hàng nghìn, hàng vạn, nếu cứ phân chia tổng thể theo số tổ thực tế đó thì việc phân tổ gặp nhiều khó khăn và có thể không giúp gì được cho phân tích thống kê. Trong những trường hợp này phải giải quyết bằng cách ghép nhiều tổ nhỏ lại thành tổ lớn, theo nguyên tắc các tổ nhỏ ghép lại với nhau phải giống nhau về tính chất, về giá trị sử dụng, về loại hình... Yêu cầu của việc ghép nhiều tổ nhỏ thành một số tổ lớn nhằm rút bớt số tổ thực tế quá nhiều, tạo điều kiện cho việc phân tổ được gọn và hợp lý. Trên thực tế, người ta thường tiến hành sắp xếp và trình bày trong những văn bản gọi là bảng phân loại hay danh mục do Nhà nước quy định thống nhất và cố định trong một thời gian tương đối dài, nhằm đảm bảo tính chất so sánh được của tài liệu thống kê.

2. Phân tổ theo tiêu thức số lượng

Khi phân tổ theo tiêu thức số lượng tùy theo lượng biến của tiêu thức thay đổi nhiều hay ít mà cách phân tổ được giải quyết khác nhau. Mặt khác, cũng cần chú ý đến số lượng đơn vị tổng thể nhiều hay ít mà xác định số tổ thích hợp.

- Trường hợp lượng biến của tiêu thức thay đổi ít, tức là sự biến thiên về mặt lượng giữa các đơn vị không chênh lệch nhiều lần, biến động rời rạc và số lượng các lượng biến ít như số người trong gia đình, số máy do một công nhân phụ trách,... thì ở đây, số tổ có một giới hạn nhất định và thường tổ công nhân trong một nhà máy dệt cứ mỗi lượng biến là cơ sở để hình thành một tổ. Ví dụ phân tổ công nhân của một nhà máy dệt theo số máy mỗi người phụ trách như sau:

- Trường hợp lượng biến của tiêu thức biến thiên rất lớn, ta không thể áp dụng cách phân tổ nói trên được, nghĩa là không thể căn cứ vào mỗi lượng biến lập nên một tổ vì làm như vậy số tổ sẽ nhiều và không rõ sự khác nhau về chất giữa các tổ. Trong trường hợp này ta cần chú ý mối liên hệ giữa lượng và chất trong phân tổ, xét cụ thể xem lượng biến tích lũy đến một mức độ nào đó thì chất của hiện tượng mới thay đổi và làm nảy sinh ra một tổ khác. Như vậy, mỗi tổ sẽ bao gồm một phạm vi lượng biến, với hai giới hạn: giới hạn dưới là lượng biến nhỏ nhất để làm cho tổ đó được hình thành và giới hạn trên là lượng biến lớn nhất của tổ đó, nếu vượt quá giới hạn đó thì chất của tổ thay đổi và chuyển thành tổ khác. Trị số chênh lệch giữa giới hạn trên và giới hạn dưới của mỗi tổ gọi là khoảng cách tổ. Việc phân tổ theo các giới hạn như vậy gọi là *phân tổ có khoảng cách tổ*. Các khoảng cách tổ có thể đều nhau hoặc không đều.

Như vậy, cần phân biệt khi nào phân tổ theo khoảng cách tổ đều nhau và khi nào dùng khoảng cách tổ không đều nhau? Nói chung, việc xác định khoảng cách tổ đều nhau hay không đều nhau là phải căn cứ vào đặc điểm của hiện tượng nghiên cứu. Phải đảm bảo các đơn vị được phân phối vào mỗi tổ đều có cũng một tính chất và sự khác nhau về chất giữa các tổ đó. Trong thực tế sự thay đổi về lượng của các bộ phận trong hiện tượng xã hội thường không biểu diễn một cách đều đặn, bởi vì sự khác nhau về chất của chúng cũng không đều nhau, do vậy có nhiều trường hợp nghiên cứu, phải phân tổ theo khoảng cách tổ không đều nhau.

Riêng đối với các hiện tượng tương đối đồng nhất về mặt loại hình kinh tế - xã hội và lượng biến trên các đơn vị thay đổi một cách tương đối đều đặn, có thể áp dụng việc phân tổ theo khoảng cách tổ đều nhau. Cách phân tổ này tạo điều kiện dễ dàng cho việc vận dụng các công thức toán học và để trình bày số liệu trên các đồ thị thống kê.

Việc phân tổ khoảng cách tổ đều nhau tương đối đơn giản và trị số khoảng cách tổ được xác định theo công thức:

$$h = \frac{X_{max} - X_{min}}{n} \quad (3.1)$$

Trong đó:

d : giá trị khoảng cách tổ

X_{max} : lượng biến lớn nhất của tiêu thức phân tổ

X_{min} : lượng biến nhỏ nhất của tiêu thức phân tổ

n : số tổ định chia

Trên đây là lý luận về xác định số tổ cần thiết về khoảng cách tổ đối với các trường hợp phân tổ. Nói chung, khi tiến hành phân tổ cần chú ý sắp xếp làm sao cho số tổ đặt ra không qua nhiều hay quá ít, gây khó khăn cho việc nghiên cứu. Nếu số tổ quá nhiều, tổng thể bị xé lẻ, số đơn vị tổng thể bị phân tán vào nhiều tổ có tính chất giống nhau hoặc gần giống nhau; ngược lại nếu số tổ quá ít thì các đơn vị có tính chất khác nhau sẽ được phân phối vào cùng một tổ, điều đó làm cho mọi kết luận rút ra sẽ kém chính xác. Mặt khác, cũng cần bảo đảm phân phối cho mỗi tổ một số lượng đơn vị cần thiết. Có như vậy, việc phân tích đặc trưng và mối liên hệ giữa các loại hình mới có ý nghĩa. Tuy nhiên, cũng không nên loại trừ những trường hợp đặc biệt, khi cần phân tổ để vạch rõ những đơn vị điển hình tiên tiến. Các đơn vị này khi mới phát sinh tuy chỉ chiếm một bộ phận nhỏ trong toàn bộ, nhưng lại có ý nghĩa rất lớn đối với việc vận động, thúc đẩy phong trào chung.

3.2.3 Phân phối các đơn vị vào từng tổ

Sau khi xác định số tổ và khoảng cách tổ, bước cuối cùng là phân phối các đơn vị vào từng tổ và tính toán giá trị số của chỉ tiêu giải thích (nếu có).

Việc phân phối các đơn vị vào từng tổ căn cứ vào lượng biến của từng đơn vị tổng thể, vào số tổ và khoảng cách tổ đã xác định ở trên. Số lượng đơn vị của từng tổ nhiều hay ít, phân phối theo dạng nào là cơ sở để biểu hiện và phân tích đặc điểm cơ bản của hiện tượng cũng như tính toán các chỉ tiêu giải thích có liên quan hoặc các chỉ tiêu phản ánh bản chất của hiện tượng.

Các chỉ tiêu giải thích được tính toán cho từng tổ và chung trên cơ sở số lượng các đơn vị trong từng tổ. Tùy theo các chỉ tiêu đó là chỉ tiêu tuyệt đối, tương đối hay bình quân mà xác định phương pháp tổng hợp hay tính toán cho phù hợp.

3.3. Kỹ thuật trình bày kết quả tổng hợp tài liệu điều tra qua phân tổ thống kê

3.3.1 Bảng thống kê

Sau khi tổng hợp các tài liệu điều tra thống kê, muốn phát huy tác dụng của nó đối với giai đoạn phân tích thống kê, cần thiết phải trình bày kết quả tổng hợp theo một hình thức thuận lợi nhất cho việc sử dụng sau này. Có thể trình bày các kết quả tổng hợp bằng các hình thức: Bảng thống kê, đồ thị thống kê, bài viết...

1. Ý nghĩa tác dụng của bảng thống kê

Bảng thống kê là một hình thức trình bày các tài liệu thống kê một cách có hệ thống, hợp lý và rõ ràng, nhằm nêu lên các đặc trưng về mặt lượng của hiện tượng nghiên cứu. Đặc điểm chung của tất cả các bảng thống kê là bao giờ cũng có những con số của từng bộ phận và

chung có liên hệ mật thiết với nhau.

Bảng thống kê có nhiều tác dụng quan trọng trong mọi công tác nghiên cứu kinh tế nói chung và phân tích thống kê nói riêng. Các tài liệu trong bảng thống kê đã được sắp xếp lại một cách khoa học, nên có thể giúp ta tiến hành mọi việc so sánh đối chiếu, phân tích theo các phương pháp khác nhau, nhằm nêu lên sâu sắc bản chất của hiện tượng nghiên cứu. Nếu biết trình bày và sử dụng thích đáng các bảng thống kê, thì việc chứng minh vấn đề sẽ trở nên rất sinh động, có sức thuyết phục hơn cả những bài văn dài.

2. Cấu thành bảng thống kê

- *Về hình thức*: Bảng thống kê bao gồm các hàng ngang, cột dọc, các tiêu đề, tiêu mục và các tài liệu con số.

Các hàng ngang, cột dọc phản ánh quy mô của bảng thống kê. Các hàng ngang cột dọc cắt nhau tạo thành các ô dùng để điền các số liệu.

Tiêu đề của bảng thống kê phản ánh nội dung, ý nghĩa của bảng và của từng chi tiết trong bảng. Các tài liệu số thu thập được do kết quả tổng hợp thống kê được ghi vào các ô của bảng thống kê, mỗi con số phản ánh một đặc trưng về mặt lượng của hiện tượng nghiên cứu.

- *Về nội dung*: Bảng thống kê gồm hai phần: phần chủ đề và phần giải thích.

Phần chủ đề nói lên hiện tượng được trình bày trong bảng thống kê, tổng thể này được phân thành những đơn vị nào, bộ phận nào? Nó giải đáp những vấn đề: đối tượng nghiên cứu của bảng thống kê là những đơn vị nào, những loại hình gì?...

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

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

Phần chủ đề	Các chỉ tiêu giải thích			
(a)	(1)	(2)	(3)	(4)
Tên chủ đề (tên hàng)				

3. Các loại bảng thống kê:

Căn cứ theo kết cấu của phần chủ đề, có thể chia làm ba loại bảng thống kê: Bảng giản đơn, bảng phân tổ và bảng kết hợp.

Bảng giản đơn:

Bảng giản đơn là loại bảng thống kê, trong đó phần chủ đề không phân tổ. Trong phần chủ đề của bảng giản đơn có liệt kê các đơn vị tổng thể, tên gọi các địa phương hoặc các thời gian khác nhau của quá trình nghiên cứu.

Bảng phân tổ:

Bảng phân tổ là loại bảng thống kê, trong đó đối tượng nghiên cứu ghi trong phần chủ đề được phân chia thành tổ theo một tiêu thức nào đó. Các bảng phân tổ là kết quả của việc áp dụng phương pháp phân tổ thống kê. Bảng phân tổ cho ta thấy rõ các loại hình kinh tế - xã hội tồn tại trong bản thân hiện tượng nghiên cứu, nêu lên kết cấu và biến động kết cấu của hiện

tượng; trong nhiều trường hợp còn giúp ta phân tích được mối liên hệ giữa các hiện tượng.

Bảng kết hợp:

Bảng kết hợp là loại bảng thống kê, trong đó đối tượng nghiên cứu ghi trong phần chủ đề được phân tổ theo hai, ba, ... tiêu thức kết hợp với nhau. Loại bảng kết hợp như trên giúp ta nghiên cứu được sâu sắc bản chất của hiện tượng, đi sâu vào kết cấu nội bộ của hiện tượng trong quá trình phát triển.

4. Những yêu cầu đối với việc xây dựng bảng thống kê

- Quy mô bảng thống kê không nên quá lớn (quá nhiều tổ hoặc quá nhiều chỉ tiêu giải thích). Khi có nhiều tiêu thức cần phân tổ có nhiều chỉ tiêu giải thích thì nên tách ra xây dựng một số bảng thống kê.

- Các tiêu đề, tiêu mục trong bảng thống kê phải được ghi chính xác, đầy đủ, ngắn gọn và dễ hiểu.

- Các chỉ tiêu giải thích trong bảng thống kê cần được sắp xếp theo thứ tự hợp lý, phù hợp với mục đích nghiên cứu. Các chỉ tiêu có liên hệ với nhau nên sắp xếp gần nhau.

- Các ô trong bảng thống kê dùng để ghi các con số thống kê. Nếu không có số liệu để ghi vào một hoặc một số ô nào đó thì dùng các ký hiệu theo quy ước.

Trong bảng thống kê phải dùng đơn vị tính cụ thể cho từng chỉ tiêu. Nếu tất cả các số trong bảng có cùng đơn vị thì đơn vị tính ghi ở đầu bảng. Nếu các chỉ tiêu có đơn vị tính khác nhau thì đơn vị tính ghi ngay dưới tiêu mục.

Dưới bảng thống kê cần ghi rõ nguồn tài liệu sử dụng và các chi tiết cần thiết mà không thể hiện được trong bảng thống kê.

3.3.2. Đồ thị thống kê

1. Ý nghĩa và tác dụng của đồ thị

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

Đồ thị thống kê có mấy đặc điểm sau;

- Đồ thị thống kê sử dụng con số kết hợp với hình vẽ, đường nét và màu sắc để trình bày và phân tích các đặc trưng số lượng của hiện tượng. Vì vậy người xem không mất nhiều công sức đọc con số mà vẫn nhận thức được vấn đề chủ yếu một cách dễ dàng, nhanh chóng.

- Đồ thị thống kê chỉ trình bày một cách khái quát các đặc điểm chủ yếu về bản chất và xu hướng phát triển của các hiện tượng.

Do các đặc điểm nêu trên, đồ thị thống kê có tính quần chúng, có sức hấp dẫn và sinh động, làm cho người hiểu biết ít về thống kê vẫn lĩnh hội được vấn đề chủ yếu một cách dễ dàng, đồng thời giữ được ấn tượng khá sâu đối với hiện tượng.

Phương pháp đồ thị thống kê được ứng dụng rộng rãi trong nghiên cứu, nhằm mục đích hình tượng hóa:

- Sự phát triển của hiện tượng qua thời gian
- Kết cấu và biến động của kết cấu qua hiện tượng
- Trình độ phổ biến của hiện tượng
- Sự so sánh giữa các mức độ của hiện tượng
- Mối liên hệ giữa các hiện tượng

- Tình hình thực hiện kế hoạch

Ngoài ra, đồ thị thống kê còn được coi là một phương tiện tuyên truyền, một công cụ dùng để biểu dương các kết quả sản xuất.

2. Các loại đồ thị thống kê

Căn cứ theo nội dung phản ánh, có thể phân chia đồ thị thống kê thành các loại sau:

- Đồ thị so sánh
- Đồ thị phát triển
- Đồ thị kết cấu
- Đồ thị hoàn thành kế hoạch
- Đồ thị liên hệ
- Đồ thị phân phối.

Căn cứ vào hình thức biểu hiện, có thể phân chia thành:

- Biểu đồ hình cột
- Biểu đồ tượng hình
- Biểu đồ diện tích
- Đồ thị đường gấp khúc

3. Yêu cầu đối với việc xây dựng đồ thị thống kê

Một đồ thị thống kê phải đảm bảo các yêu cầu: chính xác, dễ xem, dễ hiểu. Ngoài ra còn phải thể hiện tính thẩm mỹ của đồ thị. Cho nên khi xây dựng đồ thị thống kê cần thực hiện các yêu cầu sau:

- Lựa chọn loại đồ thị cho phù hợp với nội dung, tính chất các số liệu cần diễn đạt. Để đảm bảo những yêu cầu trên, cần chú ý đến các yếu tố của đồ thị, quy mô, các ký hiệu hình học hoặc các hình vẽ, hệ tọa độ, thang và tỷ lệ xích, phần giải thích.

- Xác định quy mô đồ thị cho thích hợp. Quy mô của đồ thị được quyết định bởi chiều dài, chiều cao và quan hệ tỷ lệ giữa 2 chiều đó. Quy mô của đồ thị to hay nhỏ còn phải căn cứ vào mục đích sử dụng. Quan hệ tỷ lệ giữa chiều cao và chiều dài của đồ thị, thông thường được dùng từ 1: 1,33 đến 1: 1,50.

- Các ký hiệu hình học hoặc hình vẽ quyết định dáng của đồ thị. Các ký hiệu hình học có nhiều loại như: các chấm, các đường thẳng hoặc cong, các hình cột, hình vuông, hình chữ nhật, hình tròn v.v... Các hình vẽ khác trên đồ thị cũng có thể thay đổi nhiều loại tùy tính chất của hiện tượng nghiên cứu. Việc lựa chọn các ký hiệu hình học hoặc hình vẽ của đồ thị là vấn đề quan trọng, vì mỗi hình có khả năng diễn tả riêng. Như khi cần biểu hiện kết cấu thành phần thời gian quay vòng toa xe có thể vẽ các hình cột hoặc các hình tròn (có chia thành các hình quạt).

- Hệ tọa độ giúp cho việc xác định chính xác vị trí các ký hiệu hình học trên đồ thị. Các đồ thị thống kê thường dùng hệ tọa độ vuông góc. Trên hệ tọa độ vuông góc, trục hoành thường được dùng để biểu thị thời gian, trục tung biểu thị trị số của chỉ tiêu. Trong trường hợp phân tích mối liên hệ giữa hai tiêu thức, thì tiêu thức nguyên nhân được để ở trục hoành, tiêu thức kết quả được ghi trên trục tung.

- Thang và tỷ lệ xích giúp cho việc tính chuyển các đại lượng lên đồ thị theo các khoảng cách thích hợp. Người ta thường dùng các thang đường thẳng, được phân theo các

trục tọa độ. Cũng có khi dùng thang đường cong như thang tròn (ở đồ thị hình tròn) được chia thành 360^0 . Các thang tỷ lệ có thể có khoảng cách bằng nhau hoặc không bằng nhau. Các thang tỷ lệ có các khoảng cách không bằng nhau chỉ dùng để biểu hiện các tốc độ khi khoảng biến thiên của các mức độ quá lớn mà người ta chỉ chú ý đến biến động tương đối của chúng.

- Phải giải thích tên đồ thị, các con số và ghi chú dọc theo thang tỷ lệ, các con số bên cạnh từng bộ phận của đồ thị, giải thích các ký hiệu quy ước ... cần được ghi rõ, gọn dễ hiểu.

TÀI LIỆU THAM KHẢO

1. TS. Nguyễn Công Nhựt - **Giáo trình Thống kê công nghiệp**. NXB Thống kê, 2013
2. PGS. TS. Trần Ngọc Phác, TS Trần Thị Kim Thu – **Giáo trình Lý thuyết thống kê**. Nhà xuất bản Thống kê, 2013
3. Hà Văn Sơn - **Giáo trình Lý thuyết thống kê ứng dụng trong quản trị và kinh tế**. Nhà xuất bản Thống kê, 2004
4. PTS. Phan Công Nghĩa - **Giáo trình Thống kê Lao động**. NXB Thống kê, 2009
5. PGS.PTS. Nguyễn Thiệp, PTS Phan Công Nghĩa - **Giáo trình Thống kê chất lượng**. NXB Thống kê, 2011

CÂU HỎI ÔN TẬP CHƯƠNG 3

1. Trình bày khái niệm, ý nghĩa, nhiệm vụ và các vấn đề chủ yếu của tổng hợp thống kê?
2. Trình bày khái niệm, ý nghĩa, nhiệm vụ của phân tổ thống kê?
3. Trình bày các loại phân tổ thống kê?
4. Khái niệm, tác dụng của phân tổ nhiều chiều?
5. Khái niệm, tác dụng của các loại bảng thống kê, các yêu cầu khi xây dựng bảng thống kê?
6. Khái niệm, tác dụng của các loại đồ thị thống kê, các yêu cầu khi xây dựng đồ thị thống kê?

CHƯƠNG 4

NGHIÊN CỨU CÁC MỨC ĐỘ CỦA HIỆN TƯỢNG KINH TẾ XÃ HỘI

Mọi hiện tượng kinh tế - xã hội đều tồn tại trong điều kiện không gian, thời gian nhất định và mặt lượng của hiện tượng được biểu hiện ở các mức độ khác nhau. Ta có thể sử dụng các chỉ tiêu đo lường tính toán các mức độ này như số tuyệt đối, số tương đối, các đặc trưng đo lường khuynh hướng tập trung, các đặc trưng đo lường độ phân tán nhằm nêu lên đặc trưng phân phối của dãy số.

4.1 Số tuyệt đối trong thống kê

4.1.1 Khái niệm và ý nghĩa của số tuyệt đối

Trong thống kê số tuyệt đối biểu hiện quy mô, khối lượng hiện tượng trong điều kiện thời gian và không gian cụ thể. Số tuyệt đối trong thống kê có thể biểu hiện số đơn vị của tổng thể hay bộ phận (số doanh nghiệp, số công nhân) hoặc các trị số của 1 tiêu thức hay chỉ tiêu thống kê nào đó (khối lượng sản phẩm dịch vụ, tổng chi phí sản xuất)...

Ví dụ năm 2016, số lao động của doanh nghiệp X là 400 người và doanh thu của doanh nghiệp này là 60 tỉ đồng, các con số trên đều là số tuyệt đối

Số tuyệt đối có ý nghĩa quan trọng đối với công tác quản lý kinh tế.

- Thông qua số tuyệt đối có thể nhận thức được cụ thể về quy mô, khối lượng của hiện tượng nghiên cứu, khả năng tiềm tàng và kết quả đạt được về sự phát triển kinh tế xã hội.
- Số tuyệt đối là cơ sở đầu tiên để phân tích thống kê và là cơ sở để tính các loại chỉ tiêu thống kê khác như số tương đối, số bình quân.
- Là căn cứ để xây dựng kế hoạch và chỉ đạo thực hiện kế hoạch.

4.1.2 Đặc điểm của số tuyệt đối

Số tuyệt đối trong thống kê bao giờ cũng phải mang một nội dung kinh tế nào đó trong thời gian và không gian cụ thể, có nghĩa là số tuyệt đối trong thống kê phải phản ánh được nội dung gì? của đơn vị nào? khi nào?

Số tuyệt đối trong thống kê phải được xác định qua điều tra và tổng hợp thống kê, nó không phải là con số lựa chọn tùy ý. Nó được xác định thông qua một hoặc một số phép tính toán.

4.1.3 Phân loại số tuyệt đối

1. Số tuyệt đối thời kỳ: Số tuyệt đối thời kỳ là số tuyệt đối phản ánh quy mô, khối lượng của hiện tượng trong một độ dài thời gian nhất định. Nó được hình thành thông qua sự tích lũy về lượng của hiện tượng trong suốt thời gian nghiên cứu. Các số tuyệt đối thời kỳ của cùng một chỉ tiêu có thể cộng lại với nhau để có trị số của thời kỳ dài hơn. Thời kỳ càng dài, trị số của chỉ tiêu càng lớn.

Ví dụ: doanh thu doanh nghiệp X năm 2009 là 120 tỉ đồng, là số tuyệt đối thời kỳ. Nhiều chỉ tiêu khác cũng là số tuyệt đối thời kỳ như chi phí sản xuất, lượng hàng hóa tiêu thụ... đều là số tuyệt đối thời kỳ vì đó là kết quả tổng hợp mặt lượng của hiện tượng trong một độ dài thời gian nhất định.

2. Số tuyệt đối thời điểm: Số tuyệt đối thời điểm là số tuyệt đối phản ánh mặt lượng của hiện tượng nghiên cứu tại một thời điểm nhất định. Với số tuyệt đối thời điểm thì không thể cộng lại với nhau. Vì khi cộng lại chúng không phản ánh được nội dung kinh tế nào. Số tuyệt đối thời điểm chỉ phản ánh tình hình của hiện tượng tại một thời điểm nào đó, trước và sau thời điểm trạng thái của hiện tượng có thể khác. Muốn có số tuyệt đối thời điểm chính xác phải quy định thời điểm hợp lý và phải tổ chức điều tra kịp thời.

Ví dụ: dân số thành phố A vào 0 giờ ngày 1/4/2009 là 2,5 triệu người, đó là số tuyệt đối thời điểm. Nhiều chỉ tiêu khác như số lao động của doanh nghiệp đầu tháng, cuối tháng, nguyên vật liệu tồn kho cuối kỳ,... đều là số tuyệt đối thời điểm.

4.1.4 Đơn vị tính số tuyệt đối

Trong thống kê, các số tuyệt đối đều có đơn vị tính cụ thể để biểu thị nội dung của hiện tượng nghiên cứu. Tùy theo tính chất của hiện tượng và mục đích nghiên cứu có thể sử dụng các đơn vị tính khác nhau.

- *Đơn vị tự nhiên:* Là đơn vị tính phù hợp với đặc điểm vật lý của hiện tượng, cái, con, chiếc, ... hoặc thời gian (phút, giờ, ngày,...). Trong nhiều trường hợp phải dùng đơn vị kép như mật độ điện thoại (máy / 100 dân).

- *Đơn vị thời gian:* Thường dùng để tính lượng lao động hao phí sản xuất ra những sản phẩm dịch vụ không thể tổng hợp hoặc so sánh với nhau bằng đơn vị tính toán khác, hoặc những sản phẩm dịch vụ phức tạp do nhiều người thực hiện qua nhiều giai đoạn khác nhau.

- *Đơn vị tiền tệ (VNĐ, USD, EURO,...):* được sử dụng rộng rãi để biểu thị giá trị của sản phẩm dịch vụ. Nó giúp cho việc tổng hợp và so sánh nhiều loại sản phẩm có giá trị sử dụng và đo lường khác nhau. Tuy nhiên, do giá cả luôn thay đổi nên đơn vị tiền tệ không có tính chất so sánh được qua thời gian. Để khắc phục nhược điểm này phải dùng “giá cố định” ở thời gian nào đó.

4.2 Số tương đối trong thống kê

4.2.1 Khái niệm và ý nghĩa số tương đối

1. Khái niệm: Số tương đối trong thống kê là một loại chỉ tiêu biểu hiện quan hệ so sánh giữa hai mức độ của hiện tượng nghiên cứu cùng loại nhưng khác nhau về thời gian hoặc không gian, hoặc biểu diễn quan hệ tỷ lệ giữa hai chỉ tiêu thống kê khác loại nhưng có liên quan đến nhau.

2. Ý nghĩa: Trong thống kê số tương đối có ý nghĩa quan trọng. Nó là một trong những chỉ tiêu phân tích thống kê. Nếu như số tuyệt đối chỉ khái quát được về quy mô khối lượng của hiện tượng nghiên cứu, thì số tương đối cho phép phân tích đặc điểm của hiện tượng, nghiên cứu các hiện tượng trong mối quan hệ so sánh với nhau.

Trong thống kê số tương đối phản ánh các mặt sau:

- Số tương đối cho biết kết cấu của hiện tượng;
- Biểu hiện mối quan hệ giữa hai chỉ tiêu;
- Biểu hiện trình độ phát triển của hiện tượng (so sánh thực hiện giữa các năm);
- Trình độ phổ biến của hiện tượng (mật độ dân số, mật độ điện thoại,...);
- Giúp ta kiểm tra đánh giá tình hình thực hiện kế hoạch;
- Đi sâu phân tích so sánh đặc điểm của hiện tượng và giữ bí mật số tuyệt đối

4.2.2 Đặc điểm của số tương đối

Trong thống kê số tương đối không phải là con số trực tiếp thu được qua điều tra mà là kết quả so sánh giữa hai số tuyệt đối đã có. Vì vậy mỗi số tương đối phải có gốc để so sánh (gốc so sánh gọi là chỉ tiêu gốc). Tùy theo mục đích nghiên cứu mà chọn gốc so sánh cho phù hợp.

Trong thống kê số tương đối có thể biểu thị bằng lần hoặc phần trăm (%), phần nghìn (o/oo)

4.2.3 Các loại số tương đối

1. Số tương đối động thái (tốc độ phát triển, chỉ số phát triển)

Số tương đối động thái biểu hiện sự biến động về mức độ của hiện tượng nghiên cứu qua thời gian. Nó được xác định bằng cách so sánh 2 mức độ của chỉ tiêu cùng loại nhưng khác nhau về thời gian, được biểu hiện bằng số lần hoặc số phần trăm.

$$t_{dt} = \frac{y_1}{y_0} \text{ (lần)}, \text{ hoặc } t_{dt} = \frac{y_1}{y_0} 100 \text{ (\%)} \quad (4.1)$$

Trong đó: y_1 - mức độ của hiện tượng kỳ nghiên cứu (kỳ báo cáo);

y_0 - mức độ của hiện tượng kỳ gốc;

t_{dt} - Số tương đối động thái.

Ví dụ: vốn đầu tư của một doanh nghiệp năm 2005 là 250 tỉ đồng và năm 2007 là 300 tỉ đồng. Nếu đem so sánh vốn đầu tư năm 2007 và năm 2005 ta có số tương đối động thái:

$$\frac{300}{250} = 1,2 \text{ lần (hay 120\%)}$$

Như vậy, vốn đầu tư năm 2007 so với 2005 bằng 1,2 lần hay 120%. Trong thực tế số tương đối động thái này được gọi là tốc độ phát triển hay chỉ số phát triển.

Muốn có số tương đối động thái chính xác cần đảm bảo tính chất so sánh được giữa các mức độ của kỳ báo cáo và kỳ gốc. Tức là phải đảm bảo giống nhau về nội dung kinh tế, về phương pháp tính và đơn vị tính, về phạm vi và độ dài thời gian mà mức độ phản ánh.

2. Số tương đối kế hoạch

Số tương đối kế hoạch được dùng để lập và kiểm tra tình hình thực hiện kế hoạch về một chỉ tiêu nào đó. Có hai loại số tương đối kế hoạch:

+ Số tương đối hoàn thành kế hoạch biểu hiện quan hệ tỷ lệ giữa mức độ thực tế đạt được trong kỳ so với mức độ kế hoạch đã đề ra của chỉ tiêu kinh tế nào đó.

$$t_{thkh} = \frac{y_1}{y_{kh}} \quad \text{hoặc} \quad t_{thkh} = \frac{y_1}{y_{kh}} \times 100 \quad (4.2)$$

Trong đó : y_{kh} - mức độ kế hoạch;

y_1 - mức độ thực tế của kỳ báo cáo;

t_{thkh} - Số tương đối kế hoạch.

+ Số tương đối nhiệm vụ kế hoạch biểu hiện mối quan hệ tỷ lệ giữa mức độ cần đạt tới của chỉ tiêu kinh tế nào đó trong kỳ kế hoạch so với mức độ thực tế của chỉ tiêu đó ở kỳ gốc. Như vậy số tương đối nhiệm vụ kế hoạch phản ánh mục tiêu cần đạt tới của đơn vị.

$$t_{nvkh} = \frac{y_{kh}}{y_o} \quad \text{hoặc} \quad t_{nvkh} = \frac{y_{kh}}{y_o} \times 100 \quad (4.3)$$

Trong đó: t_{nvkh} - số tương đối nhiệm vụ kế hoạch;

Giữa số tương đối động thái, số tương đối hoàn thành kế hoạch, số tương đối nhiệm vụ kế hoạch của cùng một chỉ tiêu, cùng một thời gian có quan hệ như sau:

$$t_{dt} = t_{nvkh} \cdot t_{thkh}$$

Quan hệ này được vận dụng để tính mức độ chưa biết khi đã biết các mức độ kia.

3. Số tương đối kết cấu

Số tương đối kết cấu phản ánh tỷ trọng của mỗi bộ phận cấu thành trong một tổng thể. Qua chỉ tiêu này có thể phân tích được đặc điểm cấu thành của hiện tượng.

Số tương đối kết cấu được tính bằng cách so sánh trị số tuyệt đối của chỉ tiêu của từng bộ phận so với trị số tuyệt đối của chỉ tiêu của cả tổng thể. Nó thường được biểu hiện bằng lần hoặc số phần trăm.

$$t_{kc} = \frac{y_{bp}}{y_{tt}} \quad \text{hoặc} \quad t_{kc} = \frac{y_{bp}}{y_{tt}} \times 100 \quad (4.4)$$

Trong đó : y_{bp} - trị số tuyệt đối của chỉ tiêu của bộ phận;

y_{tt} - trị số tuyệt đối của chỉ tiêu của tổng thể;

t_{kc} - Số tương đối kết cấu.

Ví dụ: tỉ trọng thuê bao trả sau của vinaphone trong tổng số thuê bao của mạng là:

$$\frac{2,6}{26} \cdot 100 = 10\%$$

Muốn tính số tương đối kết cấu chính xác phải phân biệt được các bộ phận có tính chất khác nhau trong tổng thể nghiên cứu, có nghĩa là phải phân tổ chính xác, tức là phải dựa vào phương pháp phân tổ thống kê.

4. Số tương đối so sánh

Số tương đối so sánh biểu hiện quan hệ so sánh giữa các hiện tượng cùng loại khác nhau về không gian. Số tương đối so sánh còn biểu hiện sự so sánh giữa các bộ phận trong cùng một tổng thể, khi so sánh người ta lấy một bộ phận nào đó làm gốc rồi đem các bộ phận khác so sánh với nó.

5. Số tương đối cường độ

Biểu hiện trình độ phổ biến của hiện tượng nghiên cứu trong điều kiện lịch sử nhất định, nó được xác định bằng cách so sánh chỉ tiêu của 2 hiện tượng khác nhau nhưng có liên quan với nhau. Mức độ của hiện tượng mà ta cần nghiên cứu trình độ phổ biến được đặt ở tử số, còn mức độ của hiện tượng có liên quan được đặt ở mẫu số. Hình thức biểu hiện của số tương đối cường độ là đơn vị kép do đơn vị tính của tử số và mẫu số hợp thành.

Ví dụ:

$$\text{Mật độ dân số} = \frac{\text{Tổng số dân(người)}}{\text{diện tích đất đai(Km}^2\text{)}} = \quad (\text{đơn vị: Người/Km}^2) \quad (4.5)$$

4.3 Số bình quân trong thống kê

4.3.1 Khái niệm số bình quân

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

Một tổng thể thống kê thường bao gồm nhiều đơn vị, các đơn vị này cơ bản là giống nhau, nhưng biểu hiện cụ thể về mặt lượng theo các tiêu thức nào đó thường chênh lệch nhau. Những chênh lệch này do nhiều yếu tố, ngoài các nguyên nhân chung quyết định đặc điểm cơ bản của hiện tượng, còn có những nguyên nhân riêng ảnh hưởng tới từng đơn vị. Khi nghiên cứu tổng thể thống kê không chỉ nêu các đặc điểm riêng biệt của từng đơn vị, mà cần tìm một chỉ tiêu có khả năng mô tả một cách khái quát đặc điểm chung, điển hình nhất của hiện tượng. Mức độ đó chính là số bình quân.

4.3.2 Ý nghĩa của số bình quân

Số bình quân có tính tổng hợp và khái quát cao, chỉ cần một trị số là số bình quân nó có thể nêu lên mức độ chung nhất, phổ biến nhất của tiêu thức nghiên cứu. Nó không kể đến chênh lệch thực tế giữa các đơn vị tổng thể. Số bình quân không biểu hiện mức độ cá biệt mà biểu hiện mức độ tính chung nhất của các đơn vị tổng thể.

Do số bình quân biểu hiện đặc điểm chung của cả tổng thể nghiên cứu cho nên những nét riêng biệt có tính chất ngẫu nhiên của từng đơn vị cá biệt bị loại trừ, tức là số bình quân đã san bằng mọi chênh lệch giữa các đơn vị tổng thể về lượng biến của tiêu thức nghiên cứu. Sự san bằng này chỉ có ý nghĩa lớn khi tính cho một số khá nhiều đơn vị.

Số bình quân có vị trí quan trọng trong lý luận và trong công tác thực tế. Nó được sử dụng:

- *Nêu lên đặc điểm điển hình của hiện tượng kinh tế xã hội số lớn trong điều kiện thời gian và địa điểm cụ thể;*
- *Dùng để so sánh các hiện tượng không cùng quy mô (có thể so sánh kết quả của hoạt động sản xuất kinh doanh của 2 đơn vị A,B, có thể 2 đơn vị này có số vốn khác nhau, có số lao động khác nhau nhưng có thể so sánh năng suất lao động bình quân của 2 đơn vị);*
- *Có thể dùng để nghiên cứu sự biến động của hiện tượng theo thời gian, nhằm thấy được xu hướng phát triển cơ bản của hiện tượng số lớn, nghĩa là của đại bộ phận các đơn vị.*
- *Số bình quân được dùng trong công tác lập kế hoạch, phân tích thống kê.*

4.3.3 Các loại số bình quân trong thống kê

Có rất nhiều loại số bình quân. Trong thống kê sử dụng loại nào cần căn cứ vào đặc điểm nguồn tài liệu sẵn có, mục đích nghiên cứu và ý nghĩa kinh tế xã hội của số bình quân. Số bình quân được phân thành hai nhóm:

- Nhóm 1: Các số bình quân được tính từ tất cả các lượng biến bằng cách chia đều tổng các lượng biến cho số đơn vị có lượng biến đó. Nhóm này bao gồm số trung bình cộng, trung bình nhân, trung bình điều hoà.
- Nhóm 2: Số bình quân chuyên dùng để mô tả đặc trưng phân phối các đơn vị của tổng thể theo trị số tiêu thức nghiên cứu. Nhóm này gồm có một và trung vị.

1. Số bình quân số học (Trung bình cộng)

Số bình quân số học còn gọi là số trung bình cộng được xác định bằng cách đem chia tổng lượng biến của tiêu thức cho tổng số đơn vị tổng thể.

** Số trung bình cộng giản đơn*

- Nó được sử dụng trong trường hợp mỗi lượng biến chỉ xuất hiện một lần. Công thức tính:

$$\bar{x} = \frac{x_1 + x_2 + x_3 + \dots + x_n}{n} = \frac{\sum_{i=1}^n x_i}{n} \quad (4.6)$$

Trong đó : x_i - là lượng biến của tiêu thức của đơn vị thứ i ;

\bar{x} - Số trung bình của tổng thể;

n - số đơn vị tổng thể.

Ví dụ: tính năng suất lao động bình quân của một tổ công nhân có 6 người trong đó những người này sản xuất được số sản phẩm tương ứng là: 50, 55, 60, 65, 70, 72.

Theo công thức trên:

$$\bar{x} = \frac{50 + 55 + 60 + 65 + 70 + 72}{6} = 62 \text{ (sản phẩm)}$$

* *Số trung bình cộng gia quyền (số trung bình số học có trọng số, số bình quân gia quyền)*

Khi mỗi lượng biến x_i xuất hiện nhiều lần, tức là có tần số (f_i) khác nhau, muốn xác định số trung bình cộng cần nhân mỗi lượng biến x_i với tần số f_i tương ứng, sau đó cộng lại và đem chia cho số đơn vị tổng thể:

$$\bar{x} = \frac{x_1 f_1 + x_2 f_2 + x_3 f_3 + \dots + x_m f_m}{f_1 + f_2 + \dots + f_m} = \frac{\sum_{i=1}^m x_i f_i}{\sum_{i=1}^m f_i} \quad (4.7)$$

Trong đó: $\sum_{i=1}^m f_i$ - Số đơn vị tổng thể

x_i - Là lượng biến i ;

f_i - tần số của tổ i ;

m - Là số tổ;

$x_i f_i$ - Gọi là gia quyền.

* *Các dạng đặc biệt của số trung bình cộng gia quyền*

- Tính số trung bình cộng gia quyền từ dãy số lượng biến có khoảng cách tổ.

Trong trường hợp mỗi tổ có một phạm vi lượng biến, cần có lượng biến đại diện để làm căn cứ tính toán, và lượng biến đại diện là các trị số giữa:

x'_i là trị số giữa của tổ thứ i

$$x'_i = \frac{x_{i \min} + x_{i \max}}{2} \quad (4.8)$$

$x_{i \min}$, $x_{i \max}$ là giới hạn dưới, giới hạn trên của tổ i

$$\bar{x} = \frac{\sum_{i=1}^m x'_i f_i}{\sum_{i=1}^m f_i} \quad (4.9)$$

- Tính số trung bình cộng chung của các số trung bình cộng tổ.

Trong trường hợp này số trung bình cộng chung sẽ là số trung bình cộng gia quyền của các số trung bình cộng tổ trong đó quyền số là số đơn vị của mỗi tổ:

$$\bar{x} = \frac{\sum_{i=1}^m \bar{x}_i f_i}{\sum_{i=1}^m f_i} \quad (4.10)$$

Trong đó: \bar{x}_i - là số trung bình cộng của tiêu thức của tổ thứ i;

f_i - là tần số của tổ thứ i;

m - là số tổ.

- Tính số trung bình cộng gia quyền khi biết tỷ trọng và số đơn vị tổng thể của mỗi tổ :

$$d_i - \text{tỷ trọng của tổ thứ } i \text{ trong tổng thể : } d_i = \frac{f_i}{\sum f_i}$$

$$\bar{x} = \frac{\sum_{i=1}^m x_i d_i}{1} \quad (4.11)$$

Trong đó: x_i - Là lượng biến của tiêu thức của tổ có tỷ trọng d_i

Để tính số bình quân cộng nhanh chóng cần biết các tính chất của chúng.

Tính chất 1: Nếu đem tất cả các lượng biến x_i cộng thêm (hoặc trừ đi) một số a bất kỳ thì số bình quân tính được cũng sẽ tăng thêm (hoặc giảm đi) một trị số bằng a .

$$\frac{\sum (x_i + a) f_i}{\sum f_i} = \bar{x} \pm a \quad (4.12)$$

Tính chất 2: Nếu đem nhân (hoặc chia) tất cả các lượng biến x_i với cùng một số d bất kỳ thì số bình quân tính được cũng sẽ lớn hơn (hoặc nhỏ hơn) d lần.

$$\frac{\sum (x_i d) f_i}{\sum f_i} = \bar{x} \cdot d \quad (4.13)$$

$$\frac{\sum (x_i / d) f_i}{\sum f_i} = \bar{x} / d \quad (4.14)$$

Tính chất 3: Nếu đem nhân (hoặc chia) tất cả các quyền số f_i với cùng một số m bất kỳ thì số bình quân tính được vẫn không thay đổi.

$$\frac{\sum x_i(m.f_i)}{\sum m f_i} = \bar{x} \quad (4.15)$$

$$\frac{\sum x_i(f_i/m)}{\sum f_i/m} = \bar{x} \quad (4.16)$$

Từ đây có thể suy ra:

- Nếu quyền số của tất cả các lượng biến bằng nhau thì số bình quân cộng gia quyền sẽ bằng số bình quân cộng giản đơn. $f_1 = f_2 = \dots = f_n = a$

- Khi tính số bình quân cộng, có thể dùng các quyền số là số tương đối tức là tỷ trọng mỗi tổ chiếm trong tổng thể

$$\begin{aligned} \frac{\sum x_i f_i}{\sum f_i} &= \frac{x_1 f_1 + x_2 f_2 + x_3 f_3 + \dots + x_m f_m}{\sum f_i} = \\ &= x_1 d_1 + x_2 d_2 + \dots + x_m d_m \end{aligned}$$

Tính chất 4: Tổng các độ lệch giữa các lượng biến xi với số bình quân cộng của chúng bao giờ cũng bằng 0

$$\sum (x_i - \bar{x}) f_i = 0$$

Tính chất 5: Trong một dãy số lượng biến mà tất cả các lượng biến đều có trị số như nhau thì số bình quân tính được chính là trị số đó

Nếu $x_i = c$ với mọi i thì $\bar{x} = c$

2. Số bình quân điều hoà

Được sử dụng trong trường hợp khi không có số liệu thống kê về số đơn vị tổng thể nhưng có số liệu thống kê về tổng các lượng biến của tiêu thức. Công thức tính:

$$\bar{x} = \frac{\sum_{i=1}^m M_i}{\sum_{i=1}^m \frac{M_i}{x_i}} \quad (4.17)$$

Trong đó: x_i - Là lượng biến thứ i ;

$M_i = x_i f_i$ - Là tổng lượng biến của tiêu thức của tổ thứ i ;

m - Là số tổ.

Số bình quân tính theo công thức trên gọi là số bình quân điều hoà gia quyền.

Các trường hợp đặc biệt:

- Nếu các tổng lượng biến của tiêu thức có giá trị bằng nhau, nghĩa là $M_1 = M_2 = M_3 = \dots = M$ thì:

$$\bar{x} = \frac{mM}{M \sum_{i=1}^m \frac{1}{x_i}} = \frac{m}{\sum_{i=1}^m \frac{1}{x_i}}$$

- Nếu các tổng lượng biến tiêu thức được xác định dưới tỷ lệ kết cấu, nghĩa là không có tài liệu về M_i về số tuyệt đối, mà chỉ biết các tỷ lệ $M_i / \sum M_i$ thì số bình quân điều hoà được xác định như sau:

$$\bar{x} = \frac{\sum M_i}{\sum \frac{M_i}{x_i}} = \frac{1}{\frac{M_1}{\sum M_i} * \frac{1}{x_1} + \frac{M_2}{\sum M_i} * \frac{1}{x_2} + \dots + \frac{M_m}{\sum M_i} * \frac{1}{x_m}} = \frac{1}{\sum_{i=1}^m \frac{d_i}{x_i}}$$

Trong đó: $d'_i = \frac{M_i}{\sum M_i}$

- Nếu di tính bằng % với $\sum d_i = 100$ thì:

$$\bar{x} = \frac{100}{\sum \frac{d_i}{x_i}}$$

3. Số bình quân nhân

Số bình quân nhân được sử dụng trong trường hợp khi các lượng biến có quan hệ tích số với nhau, thường được sử dụng để tính tốc độ phát triển trung bình của chỉ tiêu thống kê nào đó. Có hai cách tính :

- *Số bình quân nhân giản đơn:*

Được sử dụng trong trường hợp mỗi lượng biến x_i chỉ xuất hiện một lần:

$$\bar{x} = \sqrt[n]{x_1 \cdot x_2 \cdot x_3 \cdot \dots \cdot x_n} = \sqrt[n]{\prod_{i=1}^n x_i} \quad (4.19)$$

- *Số bình quân nhân gia quyền:*

Được dùng trong trường hợp mỗi lượng biến x_i xuất hiện nhiều lần, tức là có tần số f_i khác nhau:

$$\bar{x} = \sqrt[m]{\prod_{i=1}^m x_i^{f_i}} \quad (4.20)$$

Trong đó: x_i - là lượng biến thứ i ;

f_i - là tần số của tổ thứ i ;

m - là số tổ .

4. Một

a. *Khái niệm:* Một là biểu hiện của tiêu thức được gặp nhiều nhất trong tổng thể hay trong một dãy số phân phối. Như vậy, trong dãy số lượng biến Một là lượng biến có tần số

lớn nhất. Trị số của Một không phụ thuộc vào lượng biến của tiêu thức mà phụ thuộc vào sự phân phối các đơn vị trong tổng thể.

Theo định nghĩa trên có thể xác định Một là 27 lao động vì lượng biến này có tần số lớn nhất ($f_{\max}=11$)

b. Phương pháp xác định một:

Phân biệt các trường hợp sau:

* *Đối với dãy số lượng biến không có khoảng cách tổ:* Trường hợp này một là lượng biến có tần số lớn nhất.

* *Đối với dãy số có khoảng cách tổ đều nhau:* Muốn tìm Một trước hết phải tìm tổ chứa một, tổ chứa một là tổ có tần số lớn nhất, sau đó tính trị số gần đúng của một theo công thức sau:

$$M_0 = x_{M_0 \min} + h_{M_0} * \frac{f_{M_0} - f_{M_0-1}}{(f_{M_0} - f_{M_0-1}) + (f_{M_0} - f_{M_0+1})} \quad (4.21)$$

Trong đó: $x_{M_0 \min}$ - là giới hạn dưới của tổ chứa Một;

h_{M_0} - là khoảng cách của tổ có chứa Một;

$f_{M_0}, f_{M_0-1}, f_{M_0+1}$, là tần số của tổ chứa Một, tổ đứng trước tổ chứa Một, và tổ đứng sau tổ chứa Một.

Ví dụ có tài liệu khảo sát 100 doanh nghiệp về tình hình thực hiện doanh thu như sau:

* *Trường hợp dãy số lượng biến có khoảng cách tổ không đều nhau,* tổ chứa một là tổ có mật độ phân phối lớn nhất, tức là xung quanh lượng biến ấy chứa nhiều tần số nhất. Mật độ phân phối được xác định bằng cách đem chia tần số cho khoảng cách tổ tương ứng:

$$d_i = f_i / h_i$$

Trong đó d_i - Mật độ phân phối tổ thứ i ;

f_i - Tần số của tổ thứ i ;

h_i - Khoảng cách tổ i .

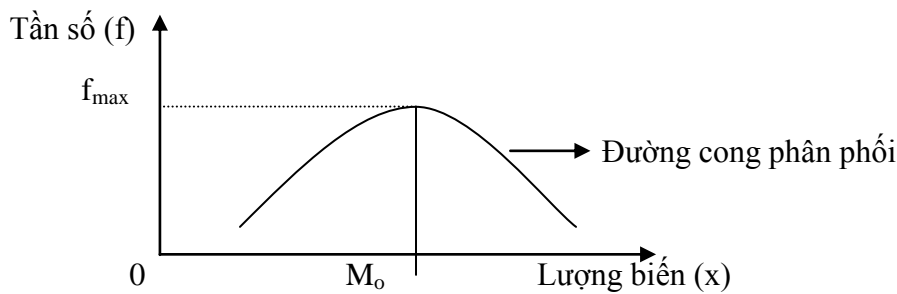
Khi đó Một được xác định theo công thức:

$$M_0 = x_{M_0 \min} + h_{M_0} * \frac{d_{M_0} - d_{M_0-1}}{(d_{M_0} - d_{M_0-1}) + (d_{M_0} - d_{M_0+1})} \quad (4.22)$$

c. Ý nghĩa của Một

Một biểu hiện mức độ phổ biến nhất của hiện tượng nghiên cứu, đồng thời bản thân nó lại không san bằng, bù trừ chênh lệch giữa các lượng biến, cho nên Một là chỉ tiêu có thể dùng bổ sung hoặc thay thế cho việc tính số trung bình khi cần thiết, trong trường hợp tính số trung bình gặp khó khăn.

Ngoài ra, một còn là một trong những mức độ để nêu lên đặc trưng phân phối của dãy số như nếu tính được số trung bình, môđ, trung vị có giá trị bằng nhau thì dãy số có phân phối chuẩn.



5. Số trung vị

a. Khái niệm

Số trung vị là lượng biến của tiêu thức của đơn vị đứng ở vị trí giữa trong dãy số lượng biến, chia dãy số lượng biến thành hai phần (phần trên và phần dưới số trung vị), mỗi phần có cùng có một số đơn vị tổng thể bằng nhau.

b. Phương pháp tính số trung vị:

+ Đối với dãy số lượng biến không có khoảng cách tổ.

Có hai trường hợp:

- Nếu số đơn vị của tổng thể lẻ: $n = 2k+1$, nghĩa là dãy số lượng biến:

$$x_1, x_2, \dots, x_k, x_{k+1}, \dots, x_{2k+1}$$

Thì số trung vị sẽ là lượng biến của đơn vị đứng ở vị trí thứ $k+1$ tức là lượng biến x_{k+1} . Ký hiệu số trung vị là Me . Ta có

$$Me = x_{k+1}$$

- Nếu số đơn vị tổng thể chẵn ($n = 2k$) nghĩa là dãy số lượng biến:

$$x_1, x_2, \dots, x_k, x_{k+1}, \dots, x_{2k}$$

Thì số trung vị được xác định căn cứ vào lượng biến của hai đơn vị đứng ở vị trí giữa nhất (đơn vị thứ k và $k+1$)

$$Me = \frac{x_k + x_{k+1}}{2}$$

+ Đối với dãy số lượng biến có khoảng cách tổ.

Muốn xác định số trung vị đầu tiên cần xác định tổ chứa trung vị, tức là tổ có chứa lượng biến của đơn vị ở vị trí giữa trong tổng số các đơn vị. Tổ có chứa số trung vị - đó là tổ đầu tiên có tần số tích lũy lớn hơn hoặc bằng nửa số đơn vị tổng thể, tức là tổ có tần số tích lũy bằng $(\sum f_i + 1)/2$. Sau đó tính giá trị gần đúng của số trung vị theo công thức sau:

$$Me = x_{Me \min} + h_{Me} * \frac{\sum \frac{f_i}{2} - S_{Me-1}}{f_{Me}} \quad (4.24)$$

Trong đó: $x_{Me \min}$: Là giới hạn dưới của tổ có số trung vị;

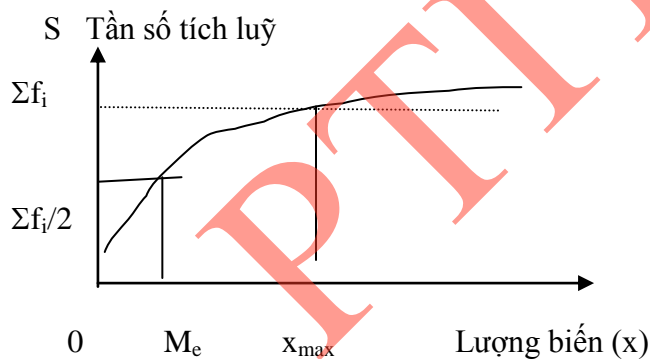
h_{Me} - Khoảng cách tổ có chứa số trung vị;

S_{Me-1} - Tần số tích lũy của tổ đứng trước tổ có số trung vị ;

f_{Me} - Tần số của tổ chứa số trung vị .

c. Ý nghĩa

Cũng như một, số trung vị có khả năng nêu lên mức độ điển hình của tổng thể mà không san bằng, bù trừ chênh lệch giữa các lượng biến. Vì vậy có thể bổ sung hoặc thay thế số trung bình cộng khi không có đủ các lượng biến để tính. Việc tính số trung vị cũng thường được dùng để nêu lên đặc trưng của dãy số phân phối.



4.4 Chỉ tiêu đánh giá độ biến thiên của tiêu thức

4.4.1 Ý nghĩa nghiên cứu

Khi xét số trung bình, số trung vị, một người ta mới chỉ xét đến tính chất điển hình của toàn bộ tổng thể nghiên cứu mà chưa xét đến từng đơn vị cá biệt và sự chênh lệch thực tế giữa các lượng biến (có khi chênh lệch này là đáng kể hoặc có khi là không đáng kể). Do vậy ngoài việc tính mức độ trung bình, cần phải đánh giá độ biến thiên của tiêu thức nhằm:

+ Đánh giá tính chất đại biểu của số trung bình. Nếu độ biến thiên của tiêu thức càng nhiều thì tính chất đại biểu của số trung bình càng ít và ngược lại.

+ Trong phân tích hoàn thành kế hoạch các chỉ tiêu đánh giá độ biến thiên của tiêu thức giúp ta thấy rõ chất lượng công tác và nhịp độ hoàn thành kế hoạch chung và từng bộ phận trong doanh nghiệp.

+ Các chỉ tiêu đánh giá độ biến thiên của tiêu thức giúp ta nghiên cứu độ tương quan và tính sai số trong điều tra chọn mẫu.

4.4.2 Các chỉ tiêu đánh giá độ biến thiên của tiêu thức

1. Khoảng biến thiên (R) hay còn gọi là toàn cự

Khoảng biến thiên là độ chênh lệch giữa lượng biến lớn nhất (x_{\max}) và lượng biến nhỏ nhất (x_{\min}) của tiêu thức nghiên cứu:

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

Khoảng biến thiên R càng nhỏ thì tính chất đồng đều của tổng thể nghiên cứu càng cao và ngược lại. Chỉ tiêu này giúp xác định nhanh chóng chênh lệch giữa đơn vị tiên tiến và đơn vị lạc hậu. Tuy nhiên nhược điểm của chỉ tiêu này là nó chỉ phụ thuộc vào lượng biến nhỏ nhất và lớn nhất của dãy số, không xét đến các lượng biến khác, cho nên nhiều khi dẫn đến những kết luận không hoàn toàn chính xác.

Ví dụ: Doanh thu tại 2 doanh nghiệp như sau (triệu đồng)

Doanh nghiệp A: 40; 50; 60; 70; 80

Doanh nghiệp B : 58; 59; 60; 61; 62

Doanh thu bình quân của mỗi doanh nghiệp là 60, nhưng thực ra 2 doanh nghiệp này không đồng đều về doanh thu. Doanh thu của doanh nghiệp A chênh lệch nhiều hơn so với doanh nghiệp B. Để đánh giá mức độ biến thiên của doanh thu, qua đó đánh giá tính chất đại biểu của số bình quân, ta tính khoảng biến thiên của 2 doanh nghiệp

$$R_A = 80 - 40 = 40 \text{ triệu đồng}$$

$$R_B = 62 - 58 = 4 \text{ triệu đồng}$$

Kết quả cho thấy R_A lớn hơn R_B , có nghĩa là độ biến thiên tiêu thức của doanh nghiệp A lớn hơn và vì thế tính chất đại biểu của số bình quân doanh nghiệp A thấp hơn.

2. Độ lệch tuyệt đối trung bình (\bar{d})

Độ lệch tuyệt đối trung bình là số trung bình cộng của các độ lệch tuyệt đối giữa các lượng biến và số trung bình cộng của các lượng biến đó.

Công thức tính:

- Khi lượng biến xuất hiện một lần:

$$\bar{d} = \frac{\sum_{i=1}^n |x_i - \bar{x}|}{n} \quad (4.25)$$

- Khi lượng biến xuất hiện nhiều lần:

$$\bar{d} = \frac{\sum_{i=1}^m |x_i - \bar{x}| f_i}{\sum_{i=1}^m f_i} \quad (4.26)$$

Trong đó: \bar{d} - Độ lệch tuyệt đối trung bình;

x_i - Các lượng biến của đơn vị thứ i (hoặc tổ i);

\bar{x} - Trung bình cộng của các lượng biến x_i .

f_i - Tần số của tổ thứ i.

Khi độ lệch tuyệt đối trung bình càng nhỏ càng nhỏ thì tính chất đồng đều của tổng thể càng lớn, tính chất đại biểu của số trung bình số học càng cao và ngược lại. Độ lệch tuyệt đối trung bình có ưu điểm hơn khoảng biến thiên vì số đo này lưu ý đến tất cả các lượng biến trong dãy số.

3. Phương sai

Phương sai là số trung bình cộng của bình phương độ lệch giữa các lượng biến x_i với số trung bình cộng của các lượng biến đó (\bar{x}).

Công thức tính:

$$\sigma^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n} \quad (4.27) \quad \text{hoặc} \quad \sigma^2 = \frac{\sum_{i=1}^m (x_i - \bar{x})^2 f_i}{\sum_{i=1}^m f_i}$$

Trong đó: σ^2 - là phương sai;

Phương sai càng nhỏ thì tính chất đồng đều của tổng thể càng cao, tính chất đại biểu của số trung bình số học càng cao và ngược lại.

Qua công thức tính phương sai cho thấy sự khác nhau về dấu giữa các độ lệch đã được khắc phục bằng cách tính bình phương. Song cũng chính vì thế mà trị số bị khuếch đại, đơn vị tính không phù hợp với thực tế. Để khắc phục nhược điểm này đưa ra chỉ tiêu độ lệch tiêu chuẩn.

4. Độ lệch tiêu chuẩn (σ)

Độ lệch tiêu chuẩn là căn bậc 2 của phương sai.

Công thức tính: $\sigma = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}}$ (4.28) hoặc

$$\sigma = \sqrt{\frac{\sum_{i=1}^m (x_i - \bar{x})^2 \cdot f_i}{\sum_{i=1}^m f_i}}$$

Chỉ tiêu độ lệch tiêu chuẩn là chỉ tiêu hoàn thiện nhất và thường dùng nhất trong nghiên cứu thống kê để đánh giá độ biến thiên của tiêu thức. Tuy nhiên việc tính toán là khá phức tạp.

5. Hệ số biến thiên (V)

Hệ số biến thiên là số tương đối so sánh giữa độ lệch tuyệt đối trung bình (\bar{d}) hoặc độ lệch tiêu chuẩn (σ) với số trung bình số học của các lượng biến đó.

$$V = \frac{\bar{d}}{\bar{x}} 100 \quad \text{và} \quad V = \frac{\sigma}{\bar{x}} 100$$

Chỉ tiêu hệ số biến thiên cho phép so sánh 2 tổng thể không cùng quy mô, hoặc so sánh giữa 2 chỉ tiêu thống kê khác nhau.

4.4.3 Các phương pháp tính phương sai

1. Các tính chất toán học của phương sai

Tính chất 1: Phương sai của một hằng số bằng 0

$$\sigma^2_C = 0$$

Tính chất 2: Nếu tất cả các lượng biến x_i được cộng (hoặc trừ) đi một số x_0 bất kỳ thì phương sai tính ra không đổi

$$\sigma^2_{(x_i \pm x_0)} = \sigma^2_{x_i}$$

Trong đó: $\sigma^2_{(x_i \pm x_0)}$ - Phương sai của các lượng biến $x_i \pm x_0$

$\sigma^2_{x_i}$ - Phương sai của các lượng biến x_i

Tính chất 3: Nếu tất cả các lượng biến x_i chia cho cùng một số d thì phương sai tính ra sẽ giảm đi d^2 lần.

$$\sigma^2_{x/d} = d^2 \cdot \sigma^2_{x_i}$$

Trong đó: $\sigma^2_{x/d}$ - Phương sai các lượng biến x_i đã chia cho cùng một số d .

Tính chất 4: Độ lệch bình phương bình quân giữa các lượng biến x_i với số bình

quân cộng x bao giờ cũng nhỏ hơn độ lệch bình phương bình quân giữa các lượng biến xi với một trị số A bất kỳ khác số bình quân cộng x. Nghĩa là phương sai có tính chất cực tiểu

$$\sigma^2_A = \sigma^2_x + (\bar{x} - A)^2$$

Trong đó: σ^2_A - Độ lệch bình phương bình quân giữa các lượng biến xi với trị số A

2. Các phương pháp tính phương sai

Như trên đã nêu công thức tính phương sai theo định nghĩa. Song khi dãy số lượng biến có các giá trị lớn thì tính toán sẽ không thích hợp. Do vậy có thể tính phương sai bằng các phương pháp khác.

a. *Tính phương sai theo công thức biến đổi*: Cách tính này được suy ra từ công thức định nghĩa.

$$\sigma^2 = \overline{x^2} - (\bar{x})^2$$

Trong đó: $\overline{x^2}$ - Bình quân cộng của các bình phương x_i^2

$(\bar{x})^2$ - Bình phương của số bình quân cộng xi

$$\overline{x^2} = \left[\frac{\sum x_i f_i}{\sum f_i} \right]^2$$

b. *Tính phương sai bằng đổi biến*: Khi tài liệu phân tổ phức tạp, việc tính phương sai bằng công thức định nghĩa cũng như công thức biến đổi vẫn dài dòng và tốn kém thời gian. Để khắc phục có thể tính bằng phương pháp đổi biến. Phương pháp này dựa trên cơ sở một số tính chất toán học của phương sai.

Quá trình tính toán phương sai bằng phương pháp đổi biến tiến hành như sau:

+ Tạo biến mới

$$x' = \frac{x_i - x_0}{d}$$

Trong đó: x_0 và d là các hằng số tự chọn $\neq 0$

+ Tính phương sai theo biến mới (bằng công thức định nghĩa hay công thức biến đổi)

$$\sigma^2 = \frac{\sum (x'_i - \bar{x}')^2 f_i}{\sum f_i}$$

$$\sigma^2 = \overline{x'^2} - (\bar{x}')^2$$

+ Tính phương sai theo biến cũ

$$\sigma^2 = \frac{1}{d^2} \sigma^2_{x'}$$

do đó

$$\sigma^2 = d^2 \cdot \sigma^2_{x'}$$

c. *Phương sai tiêu thức chất lượng*: Trong số các tiêu thức chất lượng thường gặp loại tiêu thức mà biểu hiện cụ thể của nó chỉ có 2 trường hợp: có hoặc không, nghĩa là ở một số đơn vị tổng thể có biểu hiện đặc tính quan tâm, còn một số đơn vị còn lại thì không biểu hiện.

Để xây dựng công thức tính số bình quân và phương sai cho loại tiêu thức này ta quy ước một số ký hiệu:

$x_1 = 1$ khi đơn vị quan sát có biểu hiện đặc tính quan tâm.

$x_2 = 0$ khi đơn vị quan sát không biểu hiện

p - Tỷ trọng của bộ phận có biểu hiện

q - Tỷ trọng của bộ phận không có biểu hiện.

p và q là những số tương đối nhỏ hơn 1, giữa chúng có mối quan hệ $p + q = 1$. Xác định số bình quân và phương sai căn cứ vào công thức định nghĩa.

+ Số bình quân cộng

$$\bar{x} = \frac{\sum x_i f_i}{\sum f_i} = \frac{(1 \times p) + (0 \times q)}{p + q} = p$$

+ Phương sai

$$\sigma^2 = \frac{\sum (x_i - \bar{x})^2 f_i}{\sum f_i} = \frac{(1 - p)^2 + (0 - q)^2}{p + q} = pq$$

4.5 Phân phối trong thống kê

4.5.1 Một số phân phối lý thuyết

1. Phân phối nhị thức

a. Công thức

Giả sử thực hiện m phép thử độc lập. Trong mỗi phép thử biến cố A xuất hiện với xác suất $P(A) = p$. Khi đó xác suất để biến cố A xuất hiện đúng x lần trong m phép thử được xác định bằng công thức:

$$P(x) = C_m^x p^x (1 - p)^{m-x}$$

Trong đó: x - Các đại lượng ngẫu nhiên có thể nhận các giá trị 0, 1, 2, 3,... m .

$(1 - p)$ - Xác suất không xuất hiện biến cố A trong mỗi phép thử $1 - p = q$.

C_m^x - Các hệ số của phân phối nhị thức (tổ hợp chập x của tập hợp gồm m phần tử). Nó được xác định:

$$C_m^x = \frac{m!}{x!(m-x)!}$$

Như vậy:
$$P(x) = \frac{m!}{x!(m-x)!} p^x q^{m-x}$$

b. Tính chất (không chứng minh)

- Tổng các xác suất tính từ mọi giá trị có thể của x bằng 1

$$\sum_{x=0}^m P(x) = 1$$

- Nếu $p = q = 0,5$ có phân phối nhị thức đối xứng. Nếu $p \neq q$ có phân phối nhị thức không đối xứng.

- Các giá trị của hệ số nhị thức C_m^x luôn có tính chất đối xứng và có thể biểu diễn dưới dạng "tam giác" Pascal.

Các giá trị của C_m^x được tính sẵn và ghi trong bảng

c. Các tham số đặc trưng

- Số bình quân: $\bar{x} = mp$

- Phương sai: $\sigma_x^2 = mpq$

- Độ lệch chuẩn: $\sigma_x = \sqrt{mpq}$

2. Phân phối Poisson

Trong phân phối nhị thức khi m lớn, việc xác định các giá trị P(x) rất tốn kém thời gian. Để khắc phục tình trạng đó, có thể sử dụng phân phối Poisson thay thế. Điều kiện thay thế một cách hợp lý là xác suất p hoặc q khá bé và m tương đối lớn.

a. Công thức

Đặt $\lambda = mp$ và lúc đó xác suất để biến cố A nào đó xuất hiện đúng x lần có thể xác định bằng công thức:

$$P(x) = \frac{\lambda^x \cdot e^{-\lambda}}{x!}$$

Trong đó: $e = 2,718 \dots$ Cơ số của logarit

$x = 0, 1, 2 \dots$ Đại lượng ngẫu nhiên rời rạc

b. Các tham số

- Số bình quân: $\bar{x} = \lambda = mp$

- Phương sai: $\sigma_x^2 = \lambda$

- Độ lệch chuẩn: $\sigma_x = \sqrt{\lambda}$

c. Tính chất

Tổng các xác suất tính từ mọi giá trị có thể của x bằng 1

$$\sum_{x=0}^m P(x) = 1$$

Việc xác định các giá trị P (x) đã được ghi sẵn .

3. Phân phối chuẩn

Phân phối chuẩn sử dụng để xét các đại lượng ngẫu nhiên liên tục.

a. Phương trình đường cong phân phối chuẩn

Phân phối chuẩn là một phân phối liên tục. Trên đồ thị các tung độ của đường cong được xác định:

$$f(x) = \frac{1}{\sigma_x \sqrt{2\pi}} e^{-\frac{(x-\bar{x})^2}{2\sigma_x^2}}$$

Trong đó: $\pi = 3,14159$; $e = 2,718$;

x - Đại lượng ngẫu nhiên liên tục có thể lấy mọi giá trị từ $-\infty$ đến $+\infty$

\bar{x} - Số bình quân ; σ_x - Độ lệch chuẩn.

Đường cong f(x) có dạng hình chuông đối xứng. Tại giá trị $x = \bar{x}$ tung độ đạt giá trị cực đại.

$$f(x) = f(\bar{x}) = \frac{1}{\sigma_x \sqrt{2\pi}} e^0 = \frac{1}{\sigma_x \sqrt{2\pi}}$$

Đỉnh của hình chuông là P ứng với tung độ cực đại. Ở hai bên của P tung độ giảm dần ; lúc đầu giảm ít, sau giảm nhiều, đến 2 điểm đối xứng ứng với hoành độ $\bar{x} - \sigma_x$ và $\bar{x} + \sigma_x$ là 2 điểm uốn đường cong. Sau đó tung độ giảm chậm và cuối cùng bên trái cũng như bên phải đường cong tiệm cận với trục hoành (xem hình).

b. Các tham số đặc trưng

- Số bình quân:
$$\bar{x} = \int_{-\infty}^{+\infty} xf(x) dx$$

- Phương sai:
$$\sigma_x^2 = \int_{-\infty}^{+\infty} (x - \bar{x})^2 f(x) dx$$

- Độ lệch chuẩn:
$$\sigma_x = \sqrt{\sigma_x^2}$$

Số bình quân xác định vị trí đường cong trên trục x, còn độ lệch chuẩn xác định hình dáng nhọn hay dẹt của đường cong. (Độ lệch chuẩn bí thì đường cong nhọn, độ lệch chuẩn

lớn thì đường cong dẹt).

Vì đường cong đối xứng qua \bar{x} nên số bình quân vừa là số mốt vừa là số trung vị. $\bar{x} = M_0 = M_e$

Hàm $f(x)$ là hàm mật độ xác suất nên có thể chứng minh được:

$$\int_{-\infty}^{+\infty} f(x)dx = 1$$

c. Đường cong chuẩn thu gọn.

Nếu trong công thức đặt:

$$t = \frac{x - \bar{x}}{\sigma_x}$$

Khi đó số bình quân theo biến mới (t) sẽ là:

$$\bar{t} = \frac{\bar{x} - \bar{x}}{\sigma_x} = 0$$

Các phương sai theo biến t:

$$\sigma_t^2 = \frac{\sigma_{x-\bar{x}}^2}{\sigma_x^2} = \frac{1}{\sigma_x^2} \sigma_{(x)}^2 = 1$$

Như vậy từ cách biến đổi công thức có thể viết thành dạng:

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

Hàm $\varphi(t)$ được gọi là phương trình đường cong chuẩn thu gọn, còn t được gọi là biến chuẩn hóa. Ở đây t được phân phối theo luật chuẩn với số bình quân $\bar{t} = 0$ và độ lệch chuẩn $\sigma_t = 1$.

4.5.2 So sánh phân phối thực nghiệm với phân phối lý thuyết

1. So sánh bằng tiêu chuẩn λ^2

Khi so sánh phân phối thực nghiệm với một phân phối lý thuyết bằng tiêu chuẩn λ^2 tiến hành theo các bước sau:

Bước 1: Vẽ đồ thị tần số thực nghiệm qua đó phán đoán dạng phân phối lý thuyết.

Do có nhiều dạng phân phối lý thuyết, trong số đó thường chỉ chọn ra 1 dạng để đem so sánh. Vì vậy để đỡ tốn thời gian phải chọn nhiều lần, ngay từ đầu cần có sự phán đoán tương đối chính xác dạng phân phối lý thuyết. Mỗi dạng phân phối lý thuyết có một đồ thị biểu diễn nhất định. Thông qua đồ thị tần số thực nghiệm sẽ có sự liên hệ và căn cứ hợp lý để phán đoán. Tuy nhiên điều phán đoán được vẫn không thể thay thế được kết luận cuối cùng.

Bước 2: Tính các tham số cần thiết của phân phối thực nghiệm

- Xác suất p, q (đối với phân phối lý thuyết và phân phối nhị thức)
- Số bình quân
- Độ lệch chuẩn
- Xác suất P (x)
- Tần số lý thuyết f_i

Cách xác định các tham số này đã được trình bày cụ thể trong phần trên.

Bước 3: Tính giá trị tiêu chuẩn λ^2 và so sánh kết luận

Giá trị tiêu chuẩn λ^2 được tính theo công thức:

$$\lambda^2 = \sum_{i=1}^k \frac{(f_i - f_i')^2}{f_i'}$$

Trong đó: f_i - Tần số thực nghiệm

f_i' - Tần số lý thuyết

$i = 1, 2, \dots, k$ - Thứ tự các tổ.

Chú ý: Khi áp dụng công thức cần thỏa mãn điều kiện trong tất cả các tổ tần số lý thuyết cũng như tần số thực nghiệm đều phải lớn hơn hoặc bằng 5. Nếu chưa thỏa mãn cần phải tiến hành ghép tổ.

Sau khi tính được λ^2 đem so sánh với giá trị λ_{bang}^2 (tra được từ bảng tính sẵn với độ tự do T và hệ số sai lầm cho phép Ps)

- Nếu $\lambda^2 < \lambda_{bang}^2$ kết luận phân phối thực nghiệm phù hợp với phân phối lý thuyết với hệ số sai lầm Ps.

- Nếu $\lambda^2 \geq \lambda_{bang}^2$ không có cơ sở kết luận phù hợp của 2 phân phối. Lúc đó tạm thời coi phân phối thực nghiệm không phù hợp với phân phối lý thuyết với hệ số sai lầm Ps.

2. So sánh bằng tiêu chuẩn Romanosky

Ở đây các bước tiến hành cũng giống như khi sử dụng tiêu chuẩn λ^2 . Nhưng để đánh giá kết quả và rút ra kết luận cuối cùng không sử dụng λ_{bang}^2 mà dùng chỉ tiêu R.

$$R = \frac{|\lambda^2 - T|}{\sqrt{2T}}$$

Trong đó: λ^2 - Tính theo công thức

T - Độ tự do

Căn cứ vào giá trị của R để kết luận:

- Nếu $R < 3$ phân phối thực nghiệm phù hợp với phân phối lý thuyết
- Nếu $R \geq 3$ phân phối thực nghiệm không phù hợp với phân phối lý thuyết.

3. So sánh bằng tiêu chuẩn Kolmogorop

Việc so sánh cũng được tiến hành theo ba bước, trong đó bước 1 và bước 2 giống như khi so sánh bằng tiêu chuẩn χ^2 .

Bước 3: Tính giá trị tiêu chuẩn và so sánh. Tiêu chuẩn Kolmogorop D được xác định bằng công thức:

$$D = \frac{1}{n} \max |F_i - F'_i|$$

Trong đó: F_i - Tần số thực nghiệm cộng dồn

F'_i - Tần số lý thuyết cộng dồn

$$n - \text{Tổng các tần số} \quad n = \sum f_i = \sum f'_i$$

Giá trị của D tính theo công thức được so sánh với $D_{\text{bảng}}$

- Nếu $D < D_{\text{bảng}}$ kết luận với hệ số sai lầm P_s phân phối thực nghiệm phù hợp với phân phối lý thuyết.
- Nếu $D \geq D_{\text{bảng}}$ phân phối thực nghiệm không phù hợp với phân phối lý thuyết.

Khi sử dụng tiêu chuẩn Kolmogorop không cần phải ghép các tổ có tần số nhỏ hơn 5.

4.6 Các chỉ tiêu biểu thị hình dáng của phân phối

Các chỉ tiêu biểu thị hình dáng của phân phối chủ yếu sử dụng đối với phân phối chuẩn.

4.6.1 Các chỉ tiêu biểu hiện sự không đối xứng của phân phối

Để biểu hiện sự không đối xứng của phân phối có thể sử dụng hai cách phổ biến sau:

1. So sánh trung bình, Một và trung vị, cụ thể:

- Nếu: $\bar{x} = M_0 = Me$ dãy số có phân phối chuẩn đối xứng
- Nếu: $\bar{x} > M_0 = Me$ dãy số có phân phối chuẩn lệch phải
- Nếu: $\bar{x} < M_0 < Me$ dãy số có phân phối chuẩn lệch trái

2. Hệ số không đối xứng:

$$K_A = \frac{\bar{x} - M_0}{\sigma}$$

- $K_A > 0$ dãy số có phân phối chuẩn lệch phải
- $K_A < 0$ dãy số có phân phối chuẩn lệch trái

- KA càng lớn dãy số có phân phối càng không đối xứng
- KA = 0 dãy số có phân phối chuẩn đối xứng

4.6.2 Các chỉ tiêu biểu hiện chiều cao và độ dốc của phân phối

Thường sử dụng các Mômen phân phối:

Công thức tính tổng quát:

$$M_k = \overline{(x_i - A)^k} \text{ gọi là mômen bậc } k$$

Trong đó A là đại lượng không đổi.

- Nếu A = 0:

$$M_k = \overline{(x_i - A)^k} = \frac{\sum x_i^k}{n}$$

Mômen ban đầu bậc 1:

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

- Nếu A ≠ 0 và A = x₀ tùy ý thì:

$$M'_k = \overline{(x_i - x_0)^k} = \frac{\sum (x_i - x_0)^k}{n}$$

- Nếu A = \bar{x} ta có mômen trung tâm:

$$\mu_k = \overline{(x_i - \bar{x})^k} = \frac{\sum (x_i - \bar{x})^k}{n}$$

Khi k = 2 gọi là mômen trung tâm bậc 2:

$$\mu_2 = \overline{(x_i - \bar{x})^2} = \frac{\sum (x_i - \bar{x})^2}{n} = \sigma^2$$

Khi k = 3 gọi là mômen trung tâm bậc 3:

$$\mu_3 = \overline{(x_i - \bar{x})^3} = \frac{\sum (x_i - \bar{x})^3}{n}$$

Khi k = 4 gọi là mômen trung tâm bậc 4:

$$\mu_4 = \overline{(x_i - \bar{x})^4} = \frac{\sum (x_i - \bar{x})^4}{n}$$

Đánh giá tính chất đối xứng của phân phối:

$$A_3 = \frac{\mu_3}{\sigma^3}$$

- $A_3 = 0$ dãy số có phân phối chuẩn đối xứng
- $A_3 > 0$ dãy số có phân phối chuẩn lệch phải
- $A_3 < 0$ dãy số có phân phối chuẩn lệch trái

Biểu hiện độ dốc của phân phối:

$$A_4 = \frac{\mu_4}{\sigma^4}$$

- $A_4 = 0$ dãy số có phân phối có độ dốc gần giống với phân phối chuẩn
- $A_4 > 0$ dãy số có phân phối có độ dốc cao hơn phân phối chuẩn
- $A_4 < 0$ dãy số có phân phối có độ dốc thấp hơn phân phối chuẩn

TÀI LIỆU THAM KHẢO CHƯƠNG 4

1. PGS. TS. Trần Ngọc Phác, TS Trần Thị Kim Thu – **Giáo trình Lý thuyết thống kê**. Nhà xuất bản Thống kê, 2013
2. GS.TS Bùi Xuân Phong - **Thống kê và ứng dụng trong BCVT**. NXB Bưu điện, 2005
3. Hà Văn Sơn - **Giáo trình Lý thuyết thống kê ứng dụng trong quản trị và kinh tế**. Nhà xuất bản Thống kê, 2004
4. PGS.PTS. Nguyễn Thiệp, PTS Phan Công Nghĩa - **Giáo trình Thống kê chất lượng**. NXB Thống kê, 2011
5. PGS Nguyễn Thị Kim Thủy. **Nguyên lý Thống kê ứng dụng trong quản lý kinh tế và kinh doanh sản xuất dịch vụ**. Nhà Xuất bản thống kê - 2009

CÂU HỎI ÔN TẬP CHƯƠNG 4

1. Trình bày khái niệm, ý nghĩa đặc điểm của số tuyệt đối trong thống kê.
2. Trình bày khái niệm, ý nghĩa đặc điểm của số tương đối trong thống kê.
3. Khái niệm, ý nghĩa số bình quân trong thống kê.
4. Trình bày ý nghĩa và nội dung các chỉ tiêu đánh giá mức độ biến thiên của tiêu thức.
5. Tuổi của sinh viên khoa QTKD Học viện CNBCVT như sau:

Tuổi	18	19	20	21	22	Cộng
Số SV	134	127	109	112	90	572

Hãy tính:

- a. Khoảng biến thiên
- b. Độ lệch tuyệt đối trung bình
- c. Phương sai
- d. Độ lệch tiêu chuẩn
- e. Hệ số biến thiên

6. Có tài liệu về năng suất lao động của công nhân của một doanh nghiệp như sau:

NSLĐ (nghìn sản phẩm)	Số công nhân
50 - 60	10
60 - 70	23
70 - 80	38
80 - 90	45
90 - 100	12

Hãy tính:

- a. Năng suất lao động trung bình của công nhân trong doanh nghiệp.
- b. Một về năng suất lao động
- c. Trung vị về năng suất lao động
- d. Nhận xét về phân phối của công nhân theo NSLĐ.

CHƯƠNG 5

ĐIỀU TRA CHỌN MẪU TRONG THỐNG KÊ

5.1. Khái niệm điều tra chọn mẫu

Trong chương 2 đã trình bày các loại điều tra thống kê, nếu theo số lượng các đơn vị của hiện tượng điều tra thực tế thì có thể phân các cuộc điều tra thống kê thành hai loại là điều tra toàn bộ và điều tra không toàn bộ.

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

Ví dụ: Để đánh giá về đời sống nhân dân của một địa phương nào đó thì có thể chọn ra một số hộ để thu thập tài liệu về lao động, về nghề nghiệp, về tình hình thu chi... Dựa vào tài liệu điều tra để suy rộng về đời sống của nhân dân toàn bộ địa phương đó.

Tại sao chỉ điều tra 1 số đơn vị tổng thể mà suy ra kết quả cho cả tổng thể. Vì quy luật số lớn đã chỉ ra rằng nếu nghiên cứu 1 số tương đối lớn hiện tượng thì những biểu hiện ngẫu nhiên, những đặc thù của hiện tượng riêng lẻ sẽ bù trừ và triệt tiêu cho nhau, tính quy luật sẽ được biểu hiện rõ. Hơn nữa lý thuyết xác suất cũng chứng minh rằng sự sai khác giữa số bình quân của một số rất lớn đại lượng ngẫu nhiên với kỳ vọng toán của nó là một đại lượng nhỏ tùy ý.

Khi chọn đơn vị để điều tra chọn mẫu người ta có thể chọn theo cách ngẫu nhiên, nghĩa là các đơn vị tổng thể đều có khả năng được chọn như nhau không phụ thuộc vào ý kiến chủ quan người chọn mẫu, hoặc không ngẫu nhiên (bàn bạc, phân tích tập thể để lựa chọn ra những đơn vị đại biểu).

So với điều tra toàn bộ điều tra chọn mẫu có những ưu điểm chủ yếu sau:

Thứ nhất, điều tra chọn mẫu thường nhanh hơn rất nhiều so điều tra toàn bộ, vì điều tra ít đơn vị, nên công tác chuẩn bị nhanh gọn, số lượng tài liệu ghi chép ít, thời gian điều tra ghi chép, tổng hợp và phân tích sẽ được rút ngắn. Do đó điều tra chọn mẫu mang tính kịp thời cao.

Thứ hai, do điều tra ít đơn vị, số nhân viên điều tra và mọi chi phí điều tra giảm, cho nên điều tra chọn mẫu sẽ tiết kiệm sức người, vật tư, tiền của.

Thứ ba, do điều tra ít đơn vị nên có thể mở rộng được nội dung điều tra, đi sâu nghiên cứu nhiều mặt của hiện tượng.

Thứ tư, tài liệu thu thập được trong điều tra chọn mẫu sẽ có độ chính xác cao, bởi vì số nhân viên điều tra ít nên có thể chọn được những người có trình độ chuyên môn cao và nhiều kinh nghiệm, đồng thời việc kiểm tra số liệu có thể tiến hành tỷ mỉ và tập trung, do đó các sai sót do ghi chép sẽ giảm đi nhiều.

Thứ năm, điều tra chọn mẫu không đòi hỏi một tổ chức lớn như điều tra toàn bộ, một tổ chức nhỏ cũng có thể tiến hành điều tra chọn mẫu.

Tóm lại điều tra chọn mẫu có rất nhiều ưu điểm so với điều tra toàn bộ, nhưng do điều tra chọn mẫu tiến hành với phạm vi nhỏ nên sẽ có sai số nhất định so với kết quả điều tra toàn bộ nên không thể dùng điều tra chọn mẫu để hoàn toàn thay thế cho điều tra toàn bộ.

Điều tra chọn mẫu có thể sử dụng trong các trường hợp sau:

- Dùng để thay thế điều tra toàn bộ khi đối tượng nghiên cứu cho phép vừa có thể điều tra toàn bộ, vừa có thể điều tra chọn mẫu, thì tiến hành điều tra chọn mẫu để có kết quả nhanh và tiết kiệm hoặc dùng cho những trường hợp việc điều tra có liên quan tới phá hủy đơn vị điều tra.

- Kết hợp với điều tra toàn bộ để mở rộng nội dung điều tra và đánh giá kết quả điều tra toàn bộ.

- Dùng trong trường hợp muốn so sánh các hiện tượng với nhau mà chưa có tài liệu cụ thể, hoặc kiểm định giả thiết đặt ra.

- Dùng để tổng hợp nhanh tài liệu điều tra toàn bộ, có thông tin nhanh, phục vụ kịp thời cho công tác quản lý.

Phân loại điều tra chọn mẫu

Điều tra chọn mẫu ngẫu nhiên

Điều tra chọn mẫu ngẫu nhiên là phương pháp chọn mẫu trong đó các đơn vị tổng thể được chọn vào mẫu một cách hoàn toàn ngẫu nhiên, nghĩa là các đơn vị tổng thể đều có khả năng được chọn như nhau, không phụ thuộc vào ý muốn chủ quan người chọn mẫu.

Điều tra chọn mẫu phi ngẫu nhiên

Điều tra chọn mẫu phi ngẫu nhiên là sự lựa chọn các đơn vị vào mẫu điều tra dựa trên kinh nghiệm và hiểu biết của người nghiên cứu về tổng thể nghiên cứu, hoặc căn cứ vào những qui định nhất định khi lấy mẫu.

5.2. Điều tra chọn mẫu ngẫu nhiên

5.2.1 Tổng thể chung và tổng thể mẫu

Từ khái niệm về điều tra chọn mẫu cho thấy có hai tổng thể là tổng thể chung và tổng thể mẫu.

Tổng thể chung là tổng thể bao gồm toàn bộ các đơn vị thuộc đối tượng điều tra. Số đơn vị của tổng thể chung thường được ký hiệu là N và được biểu thị bằng dãy số lượng biến.

$$X_1, X_2, X_3, \dots, X_n, \dots, X_N$$

Tổng thể mẫu là tổng thể bao gồm một số đơn vị nhất định được chọn ra từ tổng thể chung để điều tra thực tế. Số đơn vị của tổng thể mẫu thường được ký hiệu là n . Tổng thể mẫu gồm n đơn vị tổng thể thường được biểu diễn dưới dạng bản phân phối.

Có hai cách chọn ngẫu nhiên số lượng n đơn vị của tổng thể mẫu:

Thứ nhất, chọn mẫu nhiều lần (chọn hoàn lại)

Chọn mẫu nhiều lần là khi mỗi đơn vị được chọn ra để đăng ký rồi lại trả về tổng thể chung. Như vậy số đơn vị tổng thể chung không thay đổi trong suốt quá trình lấy mẫu. Xác suất được chọn của mỗi đơn vị tổng thể là bằng nhau.

Gọi K là số khả năng thiết lập được tổng thể mẫu thì số khả năng đó trong chọn nhiều lần được xác định theo công thức sau:

$$K = N^n$$

Thứ hai, chọn mẫu 1 lần (chọn không hoàn lại)

Chọn mẫu một lần là khi mỗi đơn vị đã được chọn để đăng ký rồi sẽ được xếp riêng ra không trả lại về tổng thể chung nữa, do đó không có khả năng chọn lại, xác suất được chọn của mỗi đơn vị không bằng nhau.

Khi đó số khả năng thiết lập tổng thể mẫu được xác định theo công thức:

$$K = C_n^N = \frac{N!}{n!(N-n)!} \quad (5.1)$$

Như vậy với cả hai cách chọn mẫu ngẫu nhiên thì số lượng mẫu có thể hình thành là rất lớn. Mẫu được chọn ra để điều tra chỉ là một trong số rất lớn số lượng mẫu có thể được hình thành.

Nếu ký hiệu:

- \bar{X}, P, σ^2 - là bình quân, tỉ lệ, phương sai của tổng thể chung.

- $\bar{x}_i, w_i, \sigma_i^2$ - là bình quân, tỉ lệ, phương sai của tổng thể mẫu thứ i ($i=1,2,3,\dots,Q$) và thực chất đó là biến ngẫu nhiên và người ta đã chứng minh được rằng: Nếu $n \geq 30$ thì \bar{x}_i, w_i được xem như phân phối chuẩn với:

- Kỳ vọng các bình quân mẫu:

$$E(\bar{x}_i) = \bar{X}$$

- Phương sai các bình quân mẫu:

$$\sigma_{\bar{x}}^2 = \frac{\sigma^2}{n} \quad (\text{chọn hoàn lại})$$

$$\sigma_{\bar{x}}^2 = \frac{\sigma^2}{n} \left(1 - \frac{n}{N}\right) \quad (\text{chọn không hoàn lại})$$

- Kỳ vọng các tỉ lệ mẫu:

$$E(w_i) = p$$

- Phương sai các tỉ lệ mẫu:

$$\sigma_w^2 = \frac{p(1-p)}{n} \quad (\text{chọn hoàn lại}) \quad (5.2)$$

$$\sigma_w^2 = \frac{p(1-p)}{n} \left(1 - \frac{n}{N}\right) \quad (\text{chọn không hoàn lại}) \quad (5.3)$$

Nếu $n < 30$ thì \bar{x}_i được xem phân phối theo quy luật Student

5.2.2 Chọn mẫu với xác suất đều và xác suất không đều

Chọn mẫu với xác suất đều là đảm bảo mỗi đơn vị của hiện tượng nghiên cứu đều có cơ hội được chọn vào mẫu như nhau. Tính bình đẳng còn thể hiện trong việc ước lượng kết quả vì kết quả thu được trên mẫu không phân biệt đơn vị nào. Phương pháp chọn mẫu không đều không lưu ý đến sự khác biệt giữa các đơn vị tổng thể và thường được sử dụng khi các đơn vị của tổng thể tương đối đồng đều nhau theo tiêu thức nghiên cứu, hoặc không biết trước được sự khác biệt giữa các đơn vị điều tra.

Chọn mẫu với xác suất không đều nghĩa là không cần đảm bảo khả năng được chọn vào mẫu của các đơn vị phải bằng nhau. Các đơn vị có thể được chọn vào mẫu theo xác suất

tỷ lệ vai trò của từng đơn vị. Xác suất ấn định riêng cho mỗi đơn vị về khả năng được chọn vào mẫu là xác suất bao hàm.

Việc chọn mẫu với xác suất không đều có khó khăn, phức tạp hơn vì phải có số liệu tiêu nghiệm về qui mô, kích cỡ của từng đơn vị.

5.2.3. Sai số trong chọn mẫu

Sai số chọn mẫu là sự chênh lệch về trị số giữa các chỉ tiêu tính ra được trong điều tra chọn mẫu và các chỉ tiêu tương ứng của tổng thể chung.

Các loại sai số

a. Sai số chọn mẫu:

Như đã nói, ta không biết được một cách chính xác các đặc trưng của tổng thể như trung bình hoặc tỷ lệ, mà chỉ có thể ước lượng chúng dựa trên thông tin thu thập từ mẫu. Các ước lượng đó, tất nhiên là không thể tránh khỏi sai sót, gọi là sai số chọn mẫu. Như vậy sai số chọn mẫu là sai số do sử dụng thông tin thu thập được chỉ trên một bộ phận tổng thể, hoặc là do mẫu không thể hiện, đại diện được cho cả tổng thể. Nói cách khác sai số chọn mẫu do hai yếu tố: yếu tố thiết kế, thực hiện và yếu tố ngẫu nhiên. Có thể hạn chế sai số bằng cách hết sức chú ý tới việc thiết kế, thực hiện chọn mẫu, song sai số chọn mẫu hầu như là tất nhiên, chỉ có thể giảm bớt bằng cách tăng qui mô của mẫu.

b. Sai số phi chọn mẫu:

Sai số phi chọn mẫu là sai số không thuộc về phương pháp chọn mẫu được sử dụng. Các điều kiện để phát sinh sai số phi chọn mẫu có thể là:

- *Sai số do ghi chép:* do đơn vị điều tra chưa hiểu đúng nội dung câu hỏi nên trả lời sai, do đo lường, do vô tình ghi chép sai hoặc cố ý ghi chép sai do mục đích nào đó.

Nếu là sai số ngẫu nhiên thì các sai lệch trong khi lấy số liệu có thể bù trừ nhau. Loại sai số này ít gây nguy hiểm.

Sai số hệ thống là sai số nguy hiểm, càng nhiều đơn vị điều tra thì sai số càng nhiều. Nó xảy ra do dụng cụ đo lường sai hay cố ý ghi sai.

Để có thể giảm bớt được sai số do ghi chép cần phải chuẩn bị tốt cho công tác điều tra, trình độ dụng cụ máy móc và ý thức người điều tra.

- *Sai số do mẫu được lấy từ một tổng thể không thích hợp.*
- *Sai số do tỷ lệ không trả lời quá cao.* Việc không nhận được các câu trả lời là một vấn đề quan trọng phải giải quyết của phương pháp chọn mẫu. Nó làm cho kết quả điều tra bị sai lệch do thông tin không đầy đủ, tổng thể trả lời có thể rất khác xa với tổng thể thực sự muốn điều tra.

5.2.4. Sai số bình quân chọn mẫu và phạm vi sai số chọn mẫu

1. Sai số bình quân chọn mẫu

a. Khi điều tra chọn mẫu dùng để suy rộng số trung bình của một tiêu thức nào đó thì sai số được tính như sau:

+ Khi chọn mẫu nhiều lần

$$\mu_x = \sqrt{\frac{\sigma^2}{n}} \quad (5.4)$$

Trong đó: μ_x - Sai số trung bình chọn mẫu.

σ^2 - Phương sai chung.

n - Số đơn vị tổng thể mẫu.

Do phương sai tổng thể chung không tính được nên phải dùng phương sai mẫu điều chỉnh để thay thế:

$$\mu_x = \sqrt{\frac{\sigma_n^2}{n-1}}$$

+ Khi chọn mẫu 1 lần

$$\mu_x = \sqrt{\frac{\sigma^2}{n} \left(1 - \frac{n}{N}\right)} \quad (5.5)$$

$$\text{hay } \mu_x = \sqrt{\frac{\sigma_n^2}{n-1} \left(1 - \frac{n}{N}\right)} \quad (5.6)$$

Trong đó: σ_n^2 - là phương sai mẫu điều chỉnh:

N - Số đơn vị của tổng thể chung.

b. Khi điều tra chọn mẫu để suy rộng tỷ lệ theo một tiêu thức nào đó thì sai số được tính như sau:

- Trường hợp chọn mẫu nhiều lần:

$$\mu_p = \sqrt{\frac{pq}{n}} = \sqrt{\frac{p(1-p)}{n}} \quad (5.7)$$

Trong đó: p - Tỷ lệ cấu thành của tổng thể chung:

$$p = \frac{M}{N}$$

M - Là số đơn vị mang dấu hiệu A nào đó mà ta quan tâm, theo định nghĩa xác suất thì p chính là xác suất để lấy đơn vị mang dấu hiệu A trong tổng thể chung.

q - là tỷ trọng của bộ phận không mang dấu hiệu A của tổng thể chung.

Do tỷ lệ cấu thành chung không tính được nên phải dùng tỷ lệ cấu thành mẫu điều chỉnh để thay thế:

$$\mu_p = \sqrt{\frac{w(1-w)}{n-1}} \quad (5.8)$$

Trong đó: w - Tỷ lệ cấu thành của tổng thể mẫu $w = \frac{m}{n}$

m - Là số đơn vị mang dấu hiệu A nào đó mà ta quan tâm thuộc tổng thể mẫu.

n - Số đơn vị của tổng thể mẫu.

- Trong trường hợp chọn mẫu 1 lần

$$\mu_p = \sqrt{\frac{pq}{n} \left(1 - \frac{n}{N}\right)} = \sqrt{\frac{p(1-p)}{n} \left(1 - \frac{n}{N}\right)} \quad (5.9)$$

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

Ý nghĩa của μ_x và μ_p

Sai số μ_x và μ_p biểu hiện trị số của các sai số chọn mẫu gặp phải khi suy rộng tài liệu. Nhưng do chọn mẫu ngẫu nhiên nên sai số này không được xác định trước về dấu(+ hoặc -) mà nó phản ánh 1 phạm vi chênh lệch có thể nhiều hơn hay ít hơn so với tham số của tổng thể chung, do đó chênh lệch giữa số trung bình mẫu (tỷ lệ mẫu) và số trung bình chung (tỷ lệ chung) là $\pm \mu_x$ và μ_p , nghĩa là:

$$\bar{X} = \bar{x} \pm \mu_x$$

$$p = w \pm \mu_p$$

\bar{X} - là số trung bình chung;

\bar{x} - Số trung bình mẫu.

p - Tỷ lệ chung.

w - Tỷ lệ mẫu.

2. Phạm vi sai số chọn mẫu

Phạm vi sai số chọn mẫu có thể được xác định theo công thức:

$$\varepsilon = t\mu$$

Trong đó: ε - Phạm vi sai số chọn mẫu.

t - Hệ số tin cậy

μ - sai số trung bình chọn mẫu.

Hoặc

$$\varepsilon_x = t\mu_x$$

$$\varepsilon_p = t\mu_p$$

3. Suy rộng kết quả điều tra chọn mẫu

Tức là tính toán các tham số của tổng thể chung trên cơ sở thu thập được trong điều tra chọn mẫu. Có hai phương pháp suy rộng:

a. *Phương pháp tính đối trực tiếp:*

Được áp dụng khi dùng các số trung bình mẫu hay tỷ lệ mẫu của tổng thể mẫu để tính ra các tham số của tổng thể chung:

$$\begin{aligned}\bar{X} &= \bar{x} \pm \varepsilon_x \Rightarrow \\ \bar{x} - \varepsilon_x &\leq \bar{X} \leq \bar{x} + \varepsilon_x\end{aligned}\quad (5.11)$$

$$\begin{aligned}p &= w \pm \varepsilon_p \\ w - \varepsilon_p &\leq p \leq w + \varepsilon_p\end{aligned}\quad (5.12)$$

b. *Phương pháp hệ số điều chỉnh:*

Được dùng để xác định kết quả điều tra toàn bộ. Căn cứ vào kết quả điều tra toàn bộ và điều tra chọn mẫu, tính ra tỷ lệ chênh lệch rồi dùng tỷ lệ này để làm hệ số điều chỉnh kết quả điều tra toàn bộ.

4. Xác định số đơn vị tổng thể mẫu

Khi điều tra chọn mẫu cần xác định số đơn vị điều tra sao cho thoả mãn hai yêu cầu sau:

- Bảo đảm sai số chọn mẫu nhỏ nhất;
- Chi phí thấp nhất.

Hai yêu cầu này đối lập nhau. Muốn đảm bảo tài liệu điều tra chính xác (sai số nhỏ nhất) phải điều tra một lượng đơn vị khá lớn, tức là cần một lượng kinh phí lớn. Ngược lại với một lượng kinh phí có hạn, chỉ có thể điều tra một số ít đơn vị thì phải chấp nhận một sai số nhất định.

Trong thực tế thường căn cứ vào yêu cầu của độ chính xác (phạm vi sai số chọn mẫu) để tính số đơn vị mẫu cần điều tra.

+ *Trường hợp để ước lượng số trung bình:*

- Nếu dự định chọn đơn vị tổng thể mẫu theo cách chọn nhiều lần (chọn hoàn lại):

$$\varepsilon_x = t\mu_x = t\sqrt{\frac{\delta^2}{n}} \quad \text{suy ra } n = \frac{t^2\delta^2}{\varepsilon_x^2} \quad (5.13)$$

- Trong trường hợp chọn một lần (chọn không hoàn lại):

$$\varepsilon_x = t\sqrt{\frac{\delta_x^2}{n}\left(1 - \frac{n}{N}\right)} \quad \text{suy ra } n = \frac{Nt^2\delta^2}{N\varepsilon_x^2 + t^2\delta^2} \quad (5.14)$$

+ Trường hợp để ước lượng tỷ lệ:

- Nếu dự định chọn các đơn vị tổng thể theo cách chọn nhiều lần:

$$\varepsilon_p = t\mu_p = t\sqrt{\frac{pq}{n}} \quad \text{suy ra } n = \frac{t^2 pq}{\varepsilon_p^2} \quad (5.15)$$

- Nếu dự định chọn mẫu theo cách chọn một lần:

$$\varepsilon_p = t\mu_p = t\sqrt{\frac{pq}{n}\left(1 - \frac{n}{N}\right)} \quad \text{suy ra } n = \frac{Nt^2 pq}{N\varepsilon_p^2 + t^2 pq} \quad (5.16)$$

Qua các công thức cho thấy các nhân tố ảnh hưởng đến số đơn vị mẫu điều tra bao gồm:

- Phạm vi sai số chọn mẫu. Nếu yêu cầu phạm vi sai số càng nhỏ thì số đơn vị mẫu cần điều tra càng nhiều và ngược lại. Số đơn vị mẫu tỷ lệ nghịch với bình phương phạm vi sai số chọn mẫu.

- Hệ số tin cậy (t). Yêu cầu trình độ tin cậy của tài liệu suy rộng cao hay thấp thì hệ số tin cậy cũng phải lớn hay nhỏ một cách tương ứng và do đó số đơn vị mẫu cũng phải tăng giảm cho phù hợp. Số đơn vị mẫu tỷ lệ thuận với bình phương hệ số tin cậy.

- Tính chất đồng đều của hiện tượng nghiên cứu. Số đơn vị mẫu tỷ lệ thuận với phương sai của tổng thể chung σ^2 và pq. Nếu tiêu thức biến thiên ít, phương sai tính ra nhỏ, số đơn vị cần điều tra sẽ giảm đi. Ngược lại nếu tiêu thức biến thiên lớn, phương sai sẽ lớn làm cho số đơn vị cần điều tra sẽ tăng lên.

Trong thực tế, khi tính số đơn vị mẫu cần điều tra thường không có tài liệu để tính phương sai chung, vì vậy phải giải quyết bằng cách lấy phương sai lớn nhất của các lần điều tra trước hoặc lấy phương sai của hiện tượng tương tự, hoặc tổ chức điều tra thí điểm để lấy phương sai.

Sau khi xác định được số đơn vị mẫu điều tra, người ta thường kiểm tra tính chất đại biểu của tổng thể đó. Đây là việc làm cần thiết vì nếu tổng thể mẫu không đại biểu cho tổng thể chung thì kết quả suy rộng sẽ không chính xác. Việc kiểm tra tính chất đại biểu của tổng thể mẫu có thể tiến hành bằng nhiều cách:

So sánh một số chỉ tiêu trung bình hoặc chỉ tiêu tỷ lệ của tổng thể mẫu đã chọn với các chỉ tiêu tương ứng của tổng thể chung (\bar{X} với \bar{x} , w với p). Tất nhiên là giữa những chỉ tiêu này thường có sự chênh lệch, nhưng nếu mức độ chênh lệch không vượt quá phạm vi cho phép ($\pm 5\%$) thì tổng thể mẫu được coi như đủ tính chất đại biểu. Nếu mức độ chênh lệch vượt quá phạm vi $\pm 5\%$, tổng thể mẫu chưa đủ tính chất đại biểu, do đó phải thay thế một số đơn vị hoặc tăng thêm số đơn vị điều tra.

Kiểm tra và xử lý những số liệu nghi ngờ. Nếu trong quá trình chọn các đơn vị điều tra gặp phải một vài đơn vị có số liệu quá lớn hay quá bé thì cần kiểm tra và xử lý bằng cách:

+ Nếu số đơn vị có số liệu đột xuất không nhiều, thì nên loại bỏ thay thế bằng các đơn vị khác.

+ Nếu số đơn vị có số liệu đột xuất không phải là cá biệt, thì nên chọn phân loại để đảm bảo có một số mẫu nhất định đại diện cho tổng thể chung.

+ Có thể so sánh sự phân phối các đơn vị theo tiêu thức nghiên cứu của 2 tổng thể. Cách so sánh này phức tạp hơn nhưng đảm bảo tính đại biểu hơn.

5.3. Các phương pháp điều tra chọn mẫu ngẫu nhiên

5.3.1. Phương pháp chọn mẫu ngẫu nhiên đơn thuần

Đây là phương pháp tổ chức chọn các đơn vị mẫu trong tổng thể chung một cách hết sức ngẫu nhiên không qua một sự sắp xếp nào. Mỗi đơn vị của tổng thể chung có thể được chọn một lần hoặc nhiều lần.

Phương pháp này đơn giản nhưng nó chỉ cho kết quả tốt khi tổng thể nghiên cứu tương đối đồng đều. Nếu tổng thể nghiên cứu có kết cấu phức tạp thì không nên chọn phương pháp này vì chọn như vậy số đơn vị tổng thể mẫu sẽ không được phân phối đều trong tổng thể chung, tính chất đại biểu của tổng thể mẫu sẽ không cao. Nếu tổng thể chung có qui mô lớn thì việc đặt số liệu rút thăm cho tất cả các đơn vị cũng gặp khó khăn. Chính vì vậy trong thực tế phương pháp chọn ngẫu nhiên đơn thuần ít được sử dụng.

5.3.2. Phương pháp chọn máy móc

Đây là phương pháp tổ chức chọn mẫu trong đó mỗi đơn vị được chọn căn cứ vào những khoảng cách nhất định. Trước hết xếp các đơn vị tổng thể chung theo một thứ tự nào đó như sắp xếp theo vần A, B, C,... của tên gọi, theo thứ tự, theo qui mô từ nhỏ tới lớn, v, v,... Sau đó lần lượt chọn các đơn vị theo thứ tự một cách máy móc, tức là cứ sau một khoảng cách nhất định lại chọn ra một đơn vị. Khoảng cách này được xác định bằng cách lấy số đơn vị tổng thể chung chia cho số đơn vị tổng thể mẫu:

$$d = \frac{N}{n}$$

Trong thực tế có những hiện tượng mà các đơn vị của nó đã được sắp xếp sẵn và trong suốt thời gian điều tra vẫn không thay đổi vị trí. Với những hiện tượng như vậy có thể lợi dụng sự bố trí sẵn có để chọn theo khoảng cách đã tính.

Phương pháp chọn máy móc có thủ tục tiến hành đơn giản hơn chọn ngẫu nhiên đơn thuần. Do lựa chọn theo từng khoảng cách nhất định nên số đơn vị mẫu được phân đều trong tổng thể chung, tính chất đại biểu của tổng thể mẫu được nâng cao.

5.3.3. Phương pháp chọn phân loại

Đó là tiến hành chọn các đơn vị mẫu khi tổng thể chung đã được phân chia thành các tổ theo các tiêu thức liên quan trực tiếp đến mục đích nghiên cứu. Việc chọn các đơn vị từ các tổ được tiến hành theo phương pháp chọn mẫu ngẫu nhiên.

Khi phân tổ, các tổ thường có các quy mô khác nhau. Nếu chọn số đơn vị tổng thể mẫu trong mỗi tổ tương ứng với tỷ trọng của nó chiếm trong tổng thể thì gọi là chọn phân loại theo tỷ lệ. Nếu chọn số đơn vị mẫu trong các tổ đều bằng nhau thì gọi là chọn không theo tỷ lệ. Trong chọn tỷ lệ, số đơn vị chọn trong mỗi tổ được tính theo công thức:

$$n_i = n \frac{N_i}{N}$$

Đối với phân loại tỷ lệ sai số trung bình chọn mẫu không phụ thuộc vào phương sai chung mà phụ thuộc vào trung bình các phương sai tổ. Vì vậy có công thức tính sau:

* Khi nhiệm vụ là để suy rộng số trung bình:

- Trường hợp chọn nhiều lần:

$$\mu_x = \sqrt{\frac{\delta^2}{n}} \quad (5.17)$$

- Trường hợp chọn một lần:

$$\mu_x = \sqrt{\frac{\delta^2}{n} \left(1 - \frac{n}{N}\right)} \quad (5.18)$$

$$\text{Với: } \delta^2_x = \frac{\sum \delta_i N_i}{\sum N_i}$$

* Khi nhiệm vụ là để ước lượng tỷ lệ:

- Trường hợp chọn nhiều lần:

$$\mu_p = \sqrt{\frac{p(1-p)}{n}} \quad (5.19)$$

- Trường hợp chọn một lần:

$$\mu_p = \sqrt{\frac{p(1-p)}{n} \left(1 - \frac{n}{N}\right)} \quad (5.20)$$

Với:

$$p(1-p) = \frac{\sum p(1-p)N_i}{\sum N_i}$$

Sai số trung bình chọn mẫu trong chọn phân loại nhỏ hơn trong chọn mẫu ngẫu nhiên đơn thuần và máy móc.

Đối với chọn phân loại không tỷ lệ, sai số trung bình chọn mẫu được tính như sau:

$$\mu = \frac{1}{N} \sqrt{\mu_i^2 N_i}$$

Trong đó: μ_i - Sai số trung bình trong từng tổ;

N_i - Số đơn vị trong từng tổ của tổng thể chung.

5.3.4. Phương pháp chọn cả khối

Theo phương pháp này số mẫu được rút ra không phải là lẻ tẻ đơn vị mà từng nhóm (khối đơn vị). Mỗi nhóm đơn vị được rút ra được điều tra hết không bỏ sót một đơn vị nào. Có thể nói rằng đây là điều tra toàn bộ trong các khối được chọn ra.

Muốn chọn theo phương pháp này, trước hết tổng thể chung phải được chia thành các khối có qui mô bằng nhau hoặc khác nhau (để tiện cho việc tính toán thường chia các khối có qui mô bằng nhau). Việc chọn khối có thể bằng phương pháp chọn ngẫu nhiên đơn thuần hoặc chọn máy móc.

Trong chọn cả khối sai số trung bình chọn mẫu không phụ thuộc vào phương sai chung mà phụ thuộc vào sự chênh lệch giữa các số trung bình khối hay còn gọi là phương sai các số trung bình khối. Vì vậy sai số trung bình chọn mẫu được tính theo công thức sau:

- Khi suy rộng số trung bình:

$$\mu_x = \sqrt{\frac{\delta}{r} \left(\frac{R-r}{R-1} \right)} \quad (5.21)$$

- Khi suy rộng tỷ lệ:

$$\mu_p = \sqrt{\frac{w(1-w)}{r} \left(\frac{R-r}{R-1} \right)} \quad (5.22)$$

Trong đó: R - Số khối của tổng thể chung;

r - Số khối được chọn ra để điều tra.

$$\delta^2 = \frac{\sum (\bar{x}_i - \bar{x})^2 n_i}{\sum n_i}$$

Trong đó: \bar{x}_i - Số trung bình của khối thứ i được chọn

\bar{x} - Số trung bình của tất cả các khối được chọn.

$$w(1-w) = \frac{\sum (w_i - w_r)^2 n_i}{\sum n_i}$$

$$w_r = \frac{\sum w_i n_i}{\sum n_i}$$

Nếu số đơn vị các khối bằng nhau:

$$w_r = \frac{\sum w_i}{r}$$

Trong đó: w_i - Tỷ lệ từng khối được chọn;

w_r - Tỷ lệ trung bình các khối được chọn.

5.3.5. Phương pháp chọn kết hợp

Chọn kết hợp là sử dụng kết hợp một số phương pháp chọn với nhau. Nếu chọn kết hợp cả khối với với chọn ngẫu nhiên đơn thuần thì trước hết tổng thể chung được chia thành các khối, chọn một số khối cần thiết, sau đó chọn ngẫu nhiên các đơn vị trong các khối đã chọn để điều tra. Với cách chọn này sai số trung bình chọn mẫu được tính theo công thức sau:

Trường hợp chọn nhiều lần:

$$\mu^2_x = \frac{\delta^2}{n} + \frac{\delta^2}{r} \quad (5.23)$$

Trường hợp chọn một lần:

$$\mu^2_x = \frac{\delta^2}{n} \left(1 - \frac{n}{N}\right) + \frac{\delta^2}{r} \left(1 - \frac{r}{R}\right) \quad (5.24)$$

5.4. Quy trình một cuộc điều tra chọn mẫu ngẫu nhiên

Một cuộc điều tra chọn mẫu ngẫu nhiên thường được tiến hành qua các giai đoạn sau đây:

5.4.1. Xác định mục tiêu nghiên cứu

Xác định mục tiêu nghiên cứu tức là phải xác định một cách rõ ràng là cuộc điều tra đó nhằm tìm hiểu những vấn đề gì, phục vụ cho yêu cầu cụ thể nào?

Xác định mục tiêu nghiên cứu là bước khởi đầu rất quan trọng, là tiền đề cho giai đoạn sau.

5.4.2. Xác định tổng thể nghiên cứu

Tổng thể nghiên cứu là tổng thể chung bao gồm tất cả các đơn vị của hiện tượng nghiên cứu tức xác định N. Để xác định tổng thể nghiên cứu thì phải dựa vào mục đích nghiên cứu, mục đích nghiên cứu khác nhau thì tổng thể nghiên cứu khác nhau. Thực chất của việc xác định tổng thể nghiên cứu là đi xác định các đơn vị của nó. Muốn vậy, ngoài việc dựa vào mục đích nghiên cứu còn phải dựa vào lý luận kinh tế xã hội, tình hình thực tế để đưa ra định nghĩa, những tiêu chuẩn làm căn cứ để xem xét một cách cụ thể.

5.4.3. Xác định nội dung điều tra

Xác định nội dung điều tra là xác định danh mục các tiêu thức cần điều tra trên các đơn vị của tổng thể mẫu và được cụ thể hóa bằng phiếu điều tra. Để xác định nội dung điều tra thì phải dựa vào mục đích nghiên cứu. Mục đích nghiên cứu đòi hỏi phải giải quyết nhiều vấn đề thì nội dung điều tra phải bao gồm nhiều tiêu thức.

5.4.4. Xác định số đơn vị của tổng thể mẫu và phương pháp tổ chức chọn mẫu

Xác định số đơn vị của tổng thể mẫu tức là xác định n. Để xác định số đơn vị của tổng thể mẫu cần phải cho trước phạm vi sai số chọn mẫu và xác suất suy rộng tài liệu. Các công thức tính số lượng đơn vị của tổng thể mẫu đã được trình bày ở trên.

Sau khi đã xác định số đơn vị của tổng thể mẫu, dựa vào các đặc điểm của hiện tượng nghiên cứu và khả năng tổ chức điều tra để áp dụng các phương pháp tổ chức chọn mẫu đã được đề cập ở phần trên nhằm xác định các đơn vị cụ thể của tổng thể mẫu.

5.4.5 Tiến hành thu thập tài liệu ở các đơn vị của tổng thể mẫu

Dựa vào phiếu điều tra để tiến hành thu thập tài liệu ở các đơn vị của tổng thể mẫu. Có nhiều phương pháp thu thập tài liệu như: phương pháp đăng ký trực tiếp, phương pháp phỏng vấn trực tiếp... Tùy thuộc vào điều kiện và tính chất của cuộc điều tra để áp dụng phương pháp thu thập tài liệu cho phù hợp.

5.4.6. Suy rộng kết quả điều tra chọn mẫu

Sau khi đã thu thập được đầy đủ tài liệu ở các đơn vị của tổng thể mẫu, căn cứ vào đó để tiến hành tính toán và suy rộng ra các đặc điểm của tổng thể chung. Có hai phương pháp suy rộng là suy rộng trực tiếp và suy rộng khoảng.

Suy rộng trực tiếp là coi các mức độ của tổng thể mẫu cũng là các mức độ của tổng thể chung. Suy rộng khoảng cách là các mức độ của tổng thể chung được xác định nhận giá trị trong một khoảng nào đó với xác suất cho trước.

5.4.7. Đưa ra kết luận về tổng thể chung

Đây là giai đoạn cuối cùng thể hiện kết quả của quá trình nghiên cứu. Câu hỏi được đặt ra là: "có thể có kết luận gì về tổng thể chung?", các kết luận có đáp ứng được mục đích nghiên cứu đã được đặt ra hay không? Có phản ánh được đặc điểm, bản chất của hiện tượng hay không. Từ đó cần đưa ra những giải pháp kiến nghị cụ thể để thúc đẩy sự phát triển của hiện tượng.

Các kết quả nghiên cứu đạt được cần được trình bày thông qua các bảng thống kê, các đồ thị thống kê và báo cáo phân tích.

5.5. Điều tra chọn mẫu phi ngẫu nhiên

Điều tra chọn mẫu phi ngẫu nhiên là chọn có dụng ý trước, nghĩa là dựa trên sự hiểu biết về hiện tượng nghiên cứu, tiến hành bàn bạc, phân tích để lựa chọn những đơn vị điều hình có khả năng đại diện cho tổng thể nghiên cứu để điều tra. Kết quả điều tra thường được dùng để suy rộng cho toàn bộ tổng thể hoặc để đánh giá một cách tổng quát. Điều tra chọn mẫu phi ngẫu nhiên được sử dụng rộng rãi để nghiên cứu các hiện tượng kinh tế phức tạp, phân tán, không ổn định đòi hỏi phải quan sát, phân tích tỷ mỉ trước khi thu thập tài liệu. Điều tra chọn mẫu phi ngẫu nhiên không hoàn toàn dựa trên cơ sở toán học như chọn mẫu ngẫu nhiên mà chủ yếu đòi hỏi phải kết hợp chặt chẽ giữa phân tích lý luận và thực tế. Sự nhận xét chủ quan của người tổ chức có ảnh hưởng đến chất lượng điều tra. Muốn đảm bảo chất lượng tài liệu điều tra phải giải quyết các vấn đề sau:

5.5.1. Đảm bảo phân tổ chính xác đối tượng điều tra

Vì mỗi đơn vị được chọn ra dù có đầy đủ tính chất đại biểu đến mấy cũng chỉ có khả năng đại diện cho mỗi bộ phận, một loại hình nào đó trong tổng thể phức tạp. Nếu tập hợp được các điển hình của nhiều bộ phận thì chúng có khả năng đại diện cho cả tổng thể hiện tượng phức tạp. Mặt khác, việc phân tổ có tác dụng thu hẹp độ biến thiên tiêu thức trong mỗi

bộ phận làm cho việc suy rộng tài liệu càng tỷ mỉ và chính xác. Tổng thể càng phức tạp, việc phân tích càng phải thận trọng, nhiều khi phân tổ còn phải trải qua nhiều bước để có những tổ chi tiết hơn.

5.5.2. Chọn đơn vị điều tra

Trong điều tra chọn mẫu phi ngẫu nhiên, người ta chọn các đơn vị điển hình có khả năng đại diện cho từng bộ phận khác nhau của tổng thể nghiên cứu. Có nhiều cách chọn:

- + Chọn những đơn vị có mức độ tiêu thức gần với số trung bình của từng bộ phận nhất, đồng thời cũng là mức độ phổ biến nhất trong bộ phận đó. Khi chọn phải thông qua quan sát bàn bạc phân tích tập thể thì mới chọn được những đơn vị điều tra có tính đại biểu cao. Sau khi chọn được các đơn vị điều tra phải kiểm tra tính chất đại biểu của các đơn vị đó, nếu chấp nhận được thì mới tiến hành điều tra thực tế.

- + Chọn những đơn vị có kinh nghiệm về một mặt nào đó.

- + Chọn một số địa phương đại diện cho từng vùng kinh tế. Trong các vùng này lại chọn ngẫu nhiên một số đơn vị điều tra.

5.5.3. Xác định số đơn vị điều tra

Điều tra chọn mẫu phi ngẫu nhiên cũng phải dựa trên cơ sở định luật số lớn, nghĩa là cần chọn ra một số đơn vị điều tra nhiều tới mức đủ khả năng đại biểu cho cả tổng thể. Vì chọn phi ngẫu nhiên nên không thể dùng công thức toán học để tính. Muốn xác định số mẫu cần:

- + Căn cứ vào tính chất phức tạp của tổng thể điều tra. Tổng thể càng phức tạp càng cần điều tra nhiều đơn vị.

- + Căn cứ vào kinh nghiệm của các địa phương khác, nước khác, các lần điều tra trước để quyết định số đơn vị cần điều tra thực tế.

- + Căn cứ vào mức độ đòi hỏi của việc nghiên cứu, lực lượng cán bộ và khả năng vật chất để quyết định tăng thêm hay giảm bớt số đơn vị cần điều tra.

5.5.4. Sai số chọn mẫu

Trong điều tra chọn mẫu phi ngẫu nhiên, sai số chọn mẫu không thể tính được bằng công thức toán học, mà phải thông qua nhận xét, so sánh để ước lượng ra. Nếu thấy sai số không lớn lắm có thể chấp nhận được thì dùng kết quả của điều tra chọn mẫu suy rộng ra các tham số của tổng thể chung. Nếu thấy nghi ngờ thì có thể chọn lại và điều tra lại. Ngoài ra có thể dùng phương pháp kiểm định thống kê để xác định chất lượng kết quả điều tra chọn mẫu. Khi suy rộng thì suy rộng trực tiếp, chứ không suy rộng có phạm vi như trong điều tra chọn mẫu ngẫu nhiên. Các đơn vị điều tra được lựa chọn từ các bộ phận khác nhau nên khi suy rộng phải theo thứ tự từng bước và chú ý đến tỷ trọng của mỗi bộ phận trong tổng thể.

5.5.5. Lực lượng tham gia điều tra

Điều tra chọn mẫu phi ngẫu nhiên đòi hỏi phải giải quyết nhiều vấn đề phức tạp, mà cơ sở chủ yếu là dựa trên sự phân tích sâu sắc đối tượng nghiên cứu. Muốn làm tốt công tác điều tra, phải thành thạo về nghiệp vụ, am hiểu hiện tượng nghiên cứu, làm tốt công tác tổ

chức vận động để hiểu rõ mục đích nghiên cứu, giải quyết tốt những vấn đề ảnh hưởng tới chất lượng điều tra.

TÀI LIỆU THAM KHẢO CHƯƠNG 5

1. TS. Nguyễn Công Nhựt - **Giáo trình Thống kê công nghiệp**. NXB Thống kê, 2013
2. PGS. TS. Trần Ngọc Phác, TS Trần Thị Kim Thu – **Giáo trình Lý thuyết thống kê**. Nhà xuất bản Thống kê, 2013
3. GS.TS Bùi Xuân Phong - **Thống kê và ứng dụng trong BCVT**. NXB Bưu điện, 2005
4. Hà Văn Sơn - **Giáo trình Lý thuyết thống kê ứng dụng trong quản trị và kinh tế**. Nhà xuất bản Thống kê, 2004
5. PTS. Phan Công Nghĩa - **Giáo trình Thống kê Lao động**. NXB Thống kê, 2009
6. PGS.PTS. Nguyễn Thiệp, PTS Phan Công Nghĩa - **Giáo trình Thống kê chất lượng**. NXB Thống kê, 2011
7. PGS Nguyễn Thị Kim Thúy. **Nguyên lý Thống kê ứng dụng trong quản lý kinh tế và kinh doanh sản xuất dịch vụ**. Nhà Xuất bản thống kê - 2009

PTTH

CÂU HỎI ÔN TẬP CHƯƠNG 5

1. Trình bày đối tượng áp dụng điều tra chọn mẫu và những ưu nhược điểm của điều tra chọn mẫu.
2. Thế nào là điều tra chọn mẫu ngẫu nhiên, phi mẫu nhiên?
3. Phân tích các yếu tố ảnh hưởng đến sai số chọn mẫu ngẫu nhiên.
4. Tại sao nếu tăng độ tin cậy khi suy rộng kết quả điều tra thì giá trị suy rộng của kết quả lại giảm?
5. Một địa phương gieo cấy 1500 ha lúa. Sắp đến vụ thu hoạch người ta chọn ngẫu nhiên 100 điểm (mỗi điểm $1m^2$) để gặt thống kê thu được kết quả sau:

Năng suất (Kg)	Số điểm gặt
Dưới 0,40	15
0,40 - 0,45	25
0,45 – 0,50	30
0,50 - 0,55	20
Từ 0,55 trở lên	10

- Tính năng suất lao động bình quân 1ha của toàn địa phương với xác suất bằng 0,9545
- Từ câu a, xác định lượng lúa của toàn địa phương
- Tỉ lệ diện tích lúa đạt 5 tấn/ha trở lên với xác suất bằng 0,6826

PTTH

CHƯƠNG 6

PHÂN TÍCH TƯƠNG QUAN VÀ HỒI QUY

6.1 Nhiệm vụ của phân tích hồi quy và tương quan

6.1.1. Liên hệ hàm số và liên hệ tương quan

Giữa các hiện tượng kinh tế xã hội tồn tại những mối liên quan ràng buộc lẫn nhau. Ngay trong cùng một hiện tượng nghiên cứu bao gồm nhiều tiêu thức khác nhau, thì những tiêu thức này cũng có những mối liên hệ qua lại nhất định. Nếu xét về mức độ phụ thuộc giữa các tiêu thức thì người ta thường chia mối liên hệ giữa các hiện tượng thành các loại sau:

Liên hệ hàm số: là loại liên hệ hết sức chặt chẽ giữa hai đại lượng nghiên cứu và biểu hiện sự liên hệ này theo những tỷ lệ tương ứng chặt chẽ, tức là khi hiện tượng thay đổi thì nó hoàn toàn quyết định sự thay đổi của hiện tượng liên quan theo một tỷ lệ tương ứng. Trường hợp này ít gặp trong thực tế, thường gặp trong toán học và vật lý.

Liên hệ tương quan: là mối liên hệ không hoàn toàn chặt chẽ như liên hệ hàm số, tức là khi trị số của tiêu thức nguyên nhân thay đổi nó dẫn đến tiêu thức kết quả thay đổi theo. Nhưng sự thay đổi của tiêu thức kết quả không hoàn toàn phụ thuộc vào sự thay đổi của tiêu thức nguyên nhân nói trên mà nó còn phụ thuộc vào các tiêu thức nguyên nhân khác, thống kê gọi mối liên hệ đó là mối liên hệ tương quan. Để phân tích và đánh giá mối liên hệ tương quan giữa các tiêu thức người ta thường sử dụng phương pháp tương quan, trên cơ sở quan sát hiện tượng số lớn. Phương pháp tương quan thường được vận dụng trong trường hợp trị số của tiêu thức nào đó bị thay đổi do ảnh hưởng của nhiều tiêu thức khác, trong đó mức độ ảnh hưởng của các tiêu thức có thể khác nhau. Tùy theo mục đích nghiên cứu có thể chọn riêng ra 1, 2 hoặc 3, 4 tiêu thức có ý nghĩa quyết định để xác định mối liên hệ tương quan giữa chúng với nhau. Trong thống kê thường phân biệt:

- *Tiêu thức nguyên nhân*: Là loại tiêu thức mà sự thay đổi trị số của nó là nguyên nhân làm ảnh hưởng tới sự thay đổi trị số của tiêu thức khác (thường ký hiệu x - là tiêu thức nguyên nhân);

- *Tiêu thức kết quả*: Là loại tiêu thức chịu ảnh hưởng tác động của các tiêu thức nguyên nhân (trong thống kê ký hiệu y - tiêu thức kết quả).

Ứng với mỗi giá trị (trị số) x_i của tiêu thức x ta có một trị số y_j của tiêu thức y . Số liệu ban đầu của hai tiêu thức này được trình bày dưới dạng bảng.

6.1.2. Nhiệm vụ của phương pháp hồi quy và tương quan

1. Khái niệm phương pháp hồi quy và tương quan

Hồi quy và tương quan là phương pháp toán học được vận dụng để phân tích mối liên hệ tương quan giữa các hiện tượng nghiên cứu gồm nhiều yếu tố, giữa các yếu tố có ảnh hưởng lẫn nhau, liên hệ tương quan với nhau.

2. Nhiệm vụ của phương pháp hồi quy và tương quan

Xác định phương trình hồi quy nhằm biểu diễn mối liên hệ tương quan dưới dạng một hàm số bao gồm:

- + Phân tích bản chất của hiện tượng;
- + Chọn dạng hàm số phù hợp với hiện tượng số lớn đã quan sát và bản chất của hiện tượng;
- + Tính toán các tham số của phương trình hồi quy;

Đánh giá mức độ chặt chẽ của mối liên hệ tương quan bằng cách tính hệ số tương quan và tỷ số tương quan.

6.1.3. Ý nghĩa của phân tích hồi quy và tương quan

Phương pháp phân tích hồi quy và tương quan là phương pháp thường được sử dụng trong thống kê để nghiên cứu mối liên hệ giữa các hiện tượng, như mối liên hệ giữa các yếu tố đầu vào của quá trình sản xuất với kết quả sản xuất, mối liên hệ giữa thu nhập và tiêu dùng, mối liên hệ giữa phát triển kinh tế và phát triển xã hội,...

Phương pháp phân tích hồi quy và tương quan còn được vận dụng trong một số phương pháp nghiên cứu thống kê khác như phân tích dãy số thời gian, dự báo thống kê,...

6.2 Hồi quy tương quan tuyến tính giữa hai tiêu thức số lượng

Tương quan tuyến tính là tương quan biểu thị mối liên hệ giữa hai (hay nhiều) tiêu thức phù hợp với một phương trình đường thẳng nào đó. Trong trường hợp này các giá trị thực nghiệm của tiêu thức nguyên nhân (x) và tiêu thức kết quả (y) thường được trình bày dưới dạng bảng.

Trong mối liên hệ giữa số lao động và giá trị sản xuất thì số lượng lao động là tiêu thức nguyên nhân – ký hiệu là x , giá trị sản xuất là tiêu thức kết quả - ký hiệu là y

Tài liệu cho thấy: Nhìn chung, cùng với sự tăng lên của số lượng lao động thì giá trị sản xuất cũng tăng lên, nhưng cũng có trường hợp không hẳn như vậy – như doanh nghiệp thứ hai so với doanh nghiệp thứ nhất : số lao động nhiều hơn nhưng giá trị sản xuất lại thấp hơn. Điều này chứng tỏ giữa số lượng lao động và giá trị sản xuất có mối liên hệ không hoàn toàn chặt chẽ, tức là liên hệ tương quan.

Giả sử giữa hai tiêu thức này có mối liên hệ tương quan tuyến tính theo một phương trình đường thẳng:

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

Trong đó: x - giá trị thực nghiệm của tiêu thức nguyên nhân;

\widehat{y}_x - giá trị lý thuyết của tiêu thức kết quả tính theo phương trình hồi qui;

a, b - là các tham số qui định vị trí của phương trình hồi qui.

Quá trình xác định và đánh giá mối tương quan tuyến tính giữa hai tiêu thức số lượng được tiến hành theo trình tự sau:

Các hệ số a, b được xác định bằng phương pháp bình phương bé nhất

$$\sum (y - \widehat{y}_x)^2 = \text{Min} \quad (6.1)$$

Từ đó ta có hệ phương trình sau:

$$\text{Hay: } \begin{cases} na + b \sum x = \sum y \\ a \sum x + b \sum x^2 = \sum xy \end{cases} \quad (6.2)$$

Hệ phương trình trên gọi là hệ phương trình chuẩn tắc xác định các tham số của phương trình hồi qui.

Giải hệ phương trình trên tính được a, b theo công thức sau:

$$b = \frac{\overline{xy} - \bar{x}\bar{y}}{\delta_x^2} \quad (6.3)$$

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

Thay các giá trị a, b vào phương trình hồi quy, ta sẽ có phương trình hồi qui biểu diễn mối liên hệ tương quan giữa hai giữa 2 tiêu thức số lượng x,y. Trong đó:

+ a là hệ số tự do, phản ánh ảnh hưởng của các tiêu thức khác đến tiêu thức kết quả y ngoài tiêu thức x

+ b là hệ số hồi quy, phản ánh mức độ ảnh hưởng của tiêu thức nguyên nhân x đến tiêu thức kết quả y.

Để tìm a, b cần tính $\sum x, \sum y, \sum xy, \sum x^2$

- *Hệ số tương quan:*

Hệ số tương quan là một số tương đối (biểu hiện bằng số lần) dùng để đánh giá trình độ chặt chẽ của mối liên hệ tương quan tuyến tính. Hệ số này được tính từ các số liệu thực nghiệm của tiêu thức x và y. Căn cứ vào các giá trị tính được của nó có thể đánh giá được mối liên hệ tương quan tuyến tính có chặt chẽ hay không và chặt chẽ tới mức độ nào?

Hệ số tương quan có thể được tính theo công thức sau:

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}} \quad (6.4)$$

Biến đổi ta được:

$$r = \frac{\overline{xy} - \bar{x}\bar{y}}{\sigma_x \sigma_y} = b \cdot \frac{\sigma_x}{\sigma_y} \quad (6.5)$$

Trong đó: x_i, y_i - là các trị số thực nghiệm của tiêu thức nguyên nhân và tiêu thức kết quả;

\bar{x}, \bar{y} - là số trung bình cộng của các trị số x_i, y_i

σ_x, σ_y - độ lệch tiêu chuẩn theo các tiêu thức x và y.

Tính chất: Giá trị của hệ số tương quan r luôn nằm trong khoảng $-1 \leq r \leq 1$.

+ Khi $r = +1$ hoặc $r = -1$ thì liên hệ giữa tiêu thức x,y là liên hệ hàm số.

+ Khi $r = 0$ thì giữa hai tiêu thức x và y không có mối liên hệ tương quan.

+ Khi r càng gần giá trị ± 1 thì mối liên hệ giữa tiêu thức x và y càng chặt chẽ và ngược lại.

+ Khi $1 > r > 0$ giữa x và y có mối tương quan tỉ lệ thuận.

+ Khi $-1 < r < 0$ giữa x và y có mối liên hệ tương quan tỉ lệ nghịch.

Trong ví dụ trên, $r = 0,961$ cho thấy: mối liên hệ giữa số lượng lao động và giá trị sản xuất rất chặt chẽ và đây là mối quan hệ tỉ lệ thuận.

6.3 Hồi quy tương quan phi tuyến giữa hai tiêu thức số lượng

Tương quan phi tuyến là tương quan biểu thị mối quan hệ giữa các tiêu thức phù hợp với một đường cong nhất định (parabol, hypebol, hàm số mũ, loga...).

Khi giữa hai tiêu thức x và y tồn tại mối liên hệ tương quan phi tuyến thì phương trình đặc trưng của mối liên hệ tương quan này không phải là một đường thẳng mà là một dạng đường cong nào đó (parabol, hypebol, hàm số mũ, loga...). Cách thức chung để xác định dạng phương trình hồi quy dạng phi tuyến cũng tương tự như khi xác định mối liên hệ tương quan tuyến tính.

6.3.1 Các dạng phương trình hồi quy

1. Phương trình Parabol bậc 2

Phương trình Parabol bậc 2 thường được sử dụng khi các giá trị của tiêu thức nguyên nhân tăng lên thì các trị số của tiêu thức kết quả tăng (hoặc giảm), việc tăng (hoặc giảm) đạt đến trị số cực đại (hoặc cực tiểu) rồi sau đó giảm (hoặc tăng). Phương trình hồi quy có dạng:

$$y_x = a + bx + cx^2$$

Trong đó: a, b, c là các tham số, được xác định bằng phương pháp bình phương nhỏ nhất, bằng cách giải hệ phương trình:

$$\begin{cases} na + b \sum x + c \sum x^2 = \sum y \\ a \sum x + b \sum x^2 + c \sum x^3 = \sum xy \\ a \sum x^2 + b \sum x^3 + c \sum x^4 = \sum yx^2 \end{cases} \quad (6.6)$$

2. Phương trình Hypebol

Phương trình Hypebol được áp dụng trong trường hợp khi các trị số của tiêu thức nguyên nhân tăng lên thì các trị số của tiêu thức kết quả có thể giảm và đến một giới hạn nào đó ($y_x = a$) thì hầu như không giảm. Phương trình hồi quy có dạng: $y_x = a + \frac{b}{x}$

Các tham số a, b được xác định bằng phương pháp bình phương nhỏ nhất bằng cách giải hệ phương trình chuẩn tắc:

$$\begin{cases} na + b \sum \frac{1}{x} = \sum y \\ a \sum \frac{1}{x} + b \sum \frac{1}{x^2} = \sum \frac{y}{x} \end{cases} \quad (6.7)$$

3. Phương trình hàm mũ:

Phương trình hàm mũ được áp dụng trong trường hợp cùng với sự tăng lên của các trị số tiêu thức nguyên nhân thì các trị số của các tiêu thức kết quả thay đổi theo cấp số nhân, nghĩa là có tốc độ phát triển xấp xỉ nhau. Phương trình hồi quy có dạng: $yx = a.b^x$

Các tham số a, b được xác định bằng cách giải hệ phương trình chuẩn tắc:

$$\begin{cases} n \lg a + \lg b \sum x = \sum \lg y \\ \lg a \sum x + \lg b \sum x^2 = \sum (\lg y)x \end{cases} \quad (6.8)$$

Ngoài 3 dạng phương trình phi tuyến ở trên, còn có nhiều dạng khác như Parabol bậc 3, lũy thừa, compec...

6.3.2. Các chỉ tiêu đánh giá mối liên hệ tương quan phi tuyến tính

Sau khi tìm được phương trình hồi quy ta đi đánh giá trình độ chặt chẽ của mối liên hệ tương quan phi tuyến. Thông thường để đánh giá trình độ chặt chẽ của mối liên hệ tương quan phi tuyến bằng tỷ số tương quan.

Khi có mối liên hệ tương quan giữa tiêu thức nguyên nhân x và tiêu thức kết quả y thì có thể tính các loại phương sai sau đây:

+ Phương sai chung, phản ánh độ biến thiên của y do tất cả các tiêu thức nguyên nhân, trong đó có cả tiêu thức x :

$$\sigma_y^2 = \frac{\sum (y_i - \bar{y})^2}{n} \quad (6.9)$$

+ Phương sai phản ánh độ biến thiên của y do ảnh hưởng riêng của tiêu thức x :

$$\sigma_{y_x}^2 = \frac{\sum (y_x - \bar{y})^2}{n} \quad (6.10)$$

+ Phương sai phản ánh độ biến thiên của y do ảnh hưởng của các tiêu thức nguyên nhân khác trừ tiêu thức x :

$$\sigma_{y(x)}^2 = \frac{\sum (y_i - y_x)^2}{n} \quad (6.11)$$

Mối quan hệ giữa 3 phương sai:

$$\sigma_y^2 = \sigma_{y_x}^2 + \sigma_{y(x)}^2 \quad (6.12)$$

Như vậy ta thấy nếu tiêu thức nguyên nhân x ảnh hưởng đến tiêu thức kết quả y càng lớn thì tỷ trọng của phương sai $\sigma_{y_x}^2$, phản ánh độ biến thiên của tiêu thức y do ảnh hưởng của tiêu thức x chiếm trong phương sai chung σ_y^2 càng nhiều và ngược lại. Do đó tỷ số giữa phương sai $\sigma_{y_x}^2$ và phương sai chung σ_y^2 có thể dùng để đánh giá trình độ chặt chẽ của mối liên hệ tương quan.

$$\eta = \sqrt{\frac{\sigma_{y_x}^2}{\sigma_y^2}} = \sqrt{1 - \frac{\sigma_{y(x)}^2}{\sigma_y^2}} \quad (6.13)$$

Hay:

$$\eta = \sqrt{1 - \frac{\sum (y_i - \bar{y}_x)^2}{\sum (y_i - \bar{y})^2}} \quad (6.14)$$

Trong đó: η - Tỷ số tương quan

y_i - là trị số thực nghiệm của tiêu thức kết quả;

\bar{y}_x - là trị số lý thuyết của tiêu thức kết quả, tính theo phương trình hồi quy;

\bar{y} : là số trung bình số học của các trị số thực nghiệm.

Tính chất của tỷ số tương quan:

Tỷ số tương quan lấy giá trị trong khoảng $0 \leq \eta \leq 1$

+ Nếu $\eta = 0$ giữa x và y không có mối liên hệ tương quan;

+ Nếu $\eta = 1$ giữa x và y có liên hệ hàm số ;

+ Nếu η càng gần 1 mối liên hệ giữa x và y càng chặt chẽ và ngược lại.

Trong cùng một tiêu thức nghiên cứu x và y tỷ số tương quan bao giờ cũng lớn hơn giá trị tuyệt đối của hệ số tương quan tuyến tính. Nếu $\eta > |r|$ thì giữa x và y có mối liên hệ tương quan phi tuyến. Còn lại nếu $\eta = |r|$ thì giữa x và y có mối liên hệ tương quan tuyến tính.

6.4. Hồi quy tương quan tuyến tính bội

Trong thực tế một tiêu thức thường có nhiều mối liên hệ với các tiêu thức khác. Khi phân tích mối liên hệ giữa nhiều tiêu thức, cần căn cứ vào mục đích nghiên cứu để chọn các tiêu thức có liên hệ với nhau và chỉ chọn những tiêu thức có ý nghĩa nhất, ảnh hưởng lớn nhất đối với tiêu thức kết quả.

6.4.1. Mô hình tuyến tính bội

Để biểu diễn mối liên hệ tương quan giữa nhiều tiêu thức số lượng thường dùng phương trình tuyến tính. Phương trình tổng quát có dạng:

$$\overline{y_{x_1 x_2 \dots x_n}} = a_0 + a_1 x_1 + a_2 x_2 + \dots + a_n x_n$$

Trong đó: x_1, x_2, \dots, x_n - là tiêu thức nguyên nhân;

y - là tiêu thức kết quả;

$a_0, a_1, a_2, \dots, a_n$ - là các tham số của phương trình hồi quy, thường được xác định bằng phương pháp bình phương nhỏ nhất.

Để đánh giá trình độ chặt chẽ của mỗi liên hệ tương quan tuyến tính giữa nhiều tiêu thức số lượng người ta thường sử dụng hệ số tương quan bội.

$$R_{y_{x_1 x_2 \dots x_n}}^- = \sqrt{1 - \frac{\sum (y - \bar{y}_{x_1 x_2 \dots x_n})^2}{\sum (y - \bar{y})^2}} \quad (6.15)$$

Trong đó: $R_{y_{x_1 x_2 \dots x_n}}^-$ - Hệ số tương quan bội, hệ số tương quan bội lấy giá trị trong khoảng $[0, 1]$. Nếu $R = 1$ thì có mỗi liên hệ hàm số;

$\bar{y}_{x_1 x_2 \dots x_n}$ - Giá trị lý thuyết tính theo phương trình hồi quy;

y - Giá trị thực nghiệm;

\bar{y} - Số trung bình số học của các giá trị thực nghiệm.

Giả sử có phương trình biểu diễn mỗi liên hệ tương quan tuyến tính giữa 3 tiêu thức, trong đó có 2 tiêu thức nguyên nhân (x_1 và x_2) và tiêu thức kết quả (y):

$$\bar{y}_{x_1 x_2} = a_0 + a_1 x_1 + a_2 x_2$$

Các tham số a_0, a_1, a_2 được xác định bằng phương pháp bình phương nhỏ nhất bằng cách giải hệ phương trình chuẩn tắc:

$$\begin{cases} na_0 + a_1 \sum x_1 + a_2 \sum x_2 = \sum y \\ a_0 \sum x_1 + a_1 \sum x_1^2 + a_2 \sum x_1 x_2 = \sum x_1 y \\ a_0 \sum x_2 + a_1 \sum x_1 x_2 + a_2 \sum x_2^2 = \sum y x_2 \end{cases} \quad (6.16)$$

Hệ số hồi quy chuẩn hóa : ký hiệu beta, được sử dụng đánh giá mức độ ảnh hưởng của từng tiêu thức nguyên nhân đối với tiêu thức kết quả và được tính theo công thức :

$$beta_i = b_i \frac{\sigma_{x_i}}{\sigma_y}$$

Dấu của $beta_i$ là dấu của b_i phản ánh chiều hướng mỗi liên hệ là thuận hay nghịch giữa tiêu thức nguyên nhân với tiêu thức kết quả.

$|beta_i|$ phản ánh mức độ ảnh hưởng của từng tiêu thức nguyên nhân đối với tiêu thức kết quả.

Như vậy, vốn đầu tư phát triển công nghiệp và số lượng lao động đều quan hệ tỉ lệ thuận với giá trị sản xuất công nghiệp và ảnh hưởng của vốn đầu tư phát triển công nghiệp đối với giá trị sản xuất công nghiệp lớn hơn ảnh hưởng của số lượng lao động đối với giá trị sản xuất công nghiệp.

Hệ số tương quan bội và hệ số tương quan riêng phần: Để đánh giá trình độ chặt chẽ mối liên hệ tương quan sử dụng hệ số tương quan bội theo công thức:

$$R_{y_{x_1x_2}} = \sqrt{\frac{r_{yx_1}^2 + r_{yx_2}^2 - 2r_{yx_1}r_{yx_2}r_{x_1x_2}}{1 - r_{x_1x_2}^2}} \quad (6.17)$$

Trong đó

$$r_{yx_1} = \frac{\overline{x_1y} - \overline{x_1}\overline{y}}{\delta_{x_1}\delta_y}; \quad r_{yx_2} = \frac{\overline{x_2y} - \overline{x_2}\overline{y}}{\delta_{x_2}\delta_y}; \quad r_{x_1x_2} = \frac{\overline{x_1x_2} - \overline{x_1}\overline{x_2}}{\delta_{x_1}\delta_{x_2}}$$

Đánh giá trình độ chặt chẽ mối liên hệ tương quan riêng giữa từng tiêu thức nguyên nhân x_1, x_2 với tiêu thức kết quả y dùng hệ số tương quan riêng phần. Hệ số tương quan riêng phần được sử dụng để đánh giá mức độ chặt chẽ giữa một tiêu thức nguyên nhân nào đó với tiêu thức kết quả y trong khi các tiêu thức nguyên nhân khác không đổi

Hệ số tương quan riêng phần giữa tiêu thức x_1 và tiêu thức y :

$$r_{yx_1(x_2)} = \frac{r_{yx_1} - r_{yx_2}r_{x_1x_2}}{\sqrt{(1 - r_{yx_2}^2)(1 - r_{x_1x_2}^2)}} \quad (6.18)$$

Hệ số tương quan riêng phần giữa x_2 và y :

$$r_{yx_2(x_1)} = \frac{r_{yx_2} - r_{yx_1}r_{x_1x_2}}{\sqrt{(1 - r_{yx_1}^2)(1 - r_{x_1x_2}^2)}} \quad (6.19)$$

6.4.2. Đa cộng tuyến

Khi xây dựng mô hình hồi quy giữa nhiều tiêu thức, về phương diện lý thuyết phải đảm bảo các tiêu thức nguyên nhân x_i không tương quan với nhau. Nếu giữa các tiêu thức nguyên nhân x_i có tương quan tuyến tính với nhau thì được gọi là hiện tượng đa cộng tuyến.

Hậu quả của đa cộng tuyến là làm cho việc ước lượng các hệ số của mô hình hồi quy sẽ không chính xác, ảnh hưởng đến việc suy rộng các kết quả tính toán.

Để khắc phục hiện tượng đa cộng tuyến, trong một số chương trình về thống kê có một số phương pháp xây dựng mô hình hồi quy sau đây:

- Phương pháp đưa vào dần: tiêu thức nguyên nhân đầu tiên được xem xét để đưa vào mô hình hồi quy là tiêu thức nguyên nhân có hệ số tương quan lớn nhất (về giá trị tuyệt đối) với tiêu thức kết quả. Để xem xét tiêu thức nguyên nhân này (và những tiêu thức nguyên nhân khác) có được đưa vào mô hình hồi quy hay không thì sử dụng tiêu chuẩn vào là thống kê F (mặc định $F=3,84$). Nếu tiêu thức nguyên nhân đầu tiên được xem xét để đưa vào mô hình hồi quy thỏa mãn tiêu chuẩn vào thì phương pháp đưa vào dần để tiếp tục, nếu không, không có tiêu thức nguyên nhân nào đưa vào mô hình hồi quy.

Khi tiêu thức nguyên nhân đầu đã được thỏa mãn tiêu chuẩn vào mô hình hồi quy thì tiêu thức nguyên nhân thứ hai được xem xét có thỏa mãn tiêu chuẩn vào hay không là tiêu thức nguyên nhân có hệ số tương quan riêng phần lớn nhất (về giá trị tuyệt đối) với tiêu thức kết quả. Nếu tiêu thức này thỏa mãn tiêu chuẩn sẽ được đưa vào mô hình hồi quy. Thủ tục này sẽ tiếp tục cho đến khi không còn tiêu thức nguyên nhân nào thỏa mãn tiêu chuẩn vào.

- Phương pháp loại trừ dần: Tất cả các tiêu thức nguyên nhân được đưa vào mô hình hồi quy. Sau đó loại trừ dần chúng bằng tiêu chuẩn loại trừ. Tiêu chuẩn loại trừ là giá trị F tối thiểu (được mặc định $F=2,71$) mà tiêu thức nguyên nhân phải đạt được để được ở lại trong mô hình hồi quy. Nếu các tiêu thức nguyên nhân có giá trị F nhỏ hơn giá trị F tối thiểu thì chúng bị loại khỏi mô hình hồi quy.

- Phương pháp chọn từng bước: Là sự kết hợp giữa hai phương pháp trên và là phương pháp thường được sử dụng.

Tiêu thức nguyên nhân đầu tiên được chọn để đưa vào mô hình hồi quy giống như phương pháp đưa vào dần, nếu nó không thỏa mãn tiêu chuẩn vào thì thủ tục này chấm dứt và không có tiêu thức nguyên nhân nào được chọn. Nếu nó thỏa mãn tiêu chuẩn vào thì tiêu thức nguyên nhân thứ hai được lựa chọn dựa vào hệ số tương quan riêng phần lớn nhất về giá trị tuyệt đối. Nếu tiêu thức nguyên nhân thứ hai thỏa mãn tiêu chuẩn vào thì nó cũng đi vào mô hình hồi quy.

Sau đó, dựa vào tiêu chuẩn ra để xem xét tiêu thức nguyên nhân thứ nhất có phải loại bỏ khỏi mô hình hồi quy hay không. Trong bước kế tiếp, các tiêu thức nguyên nhân không ở trong mô hình hồi quy được xem xét để đưa vào. Sau mỗi bước, các tiêu thức nguyên nhân ở mô hình hồi quy được xem xét để loại trừ ra cho đến khi không còn tiêu thức nguyên nhân nào thỏa mãn tiêu chuẩn ra thì kết thúc.

Các mô hình hồi quy được xây dựng theo các phương pháp trên có thể khác nhau. Tùy thuộc vào mục đích và nhiệm vụ nghiên cứu cụ thể để lựa chọn mô hình thích hợp.

6.5. Tương quan hạng và tương quan giữa hai tiêu thức thuộc tính

6.5.1. Tương quan hạng

Tương quan hạng có thể được sử dụng trong trường hợp số lượng đơn vị không nhiều để nghiên cứu mối liên hệ tương quan giữa hai tiêu thức. Đối với mỗi tiêu thức, cần phải xếp hạng từ thấp đến cao phù hợp với biểu hiện của tiêu thức - tức là sử dụng thang đo thứ bậc. Nếu biểu hiện tiêu thức của một số đơn vị giống nhau thì lấy hạng bình quân của các đơn vị đó.

Hệ số tương quan hạng r_s của Spearman được sử dụng để đánh giá mức độ chặt chẽ và tính theo công thức sau đây:

$$r_s = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)} \quad (6.20)$$

Trong đó:

n – Số đơn vị nghiên cứu;

d_i – Hiệu của hai hạng của đơn vị i , với $i=1, n$

Tính chất của hệ số tương quan hạng giống với tính chất của hệ số tương quan tuyến tính giữa hai tiêu thức số lượng tức là $-1 \leq r_s \leq 1$

6.5.2. Tương quan giữa hai tiêu thức thuộc tính

Để nghiên cứu mối liên hệ tương quan giữa hai tiêu thức thuộc tính, trước hết phải phân tổ kết hợp theo hai tiêu thức đó. Dựa vào bảng phân tổ kết hợp để tính hệ số liên hợp. Hệ số liên hợp thường được sử dụng là hệ số liên hợp của Cramer:

$$k = \sqrt{\frac{\chi^2}{n \cdot \min\{(d-1)|(c-1)\}}} \quad (6.21)$$

Trong đó:

$$\chi^2 = \sum \frac{(n_{ij} - \bar{n}_{ij})^2}{\bar{n}_{ij}} - \text{với } n_{ij} \text{ là tần số thực tế,}$$

$$\bar{n}_{ij} = \frac{\text{tổng dòng} \times \text{tổng cột}}{n} \text{ là tần số lý thuyết.}$$

n - số đơn vị nghiên cứu;

d - số dòng của bảng phân tổ kết hợp;

c - số cột của bảng phân tổ kết hợp;

k có giá trị trong khoảng 0,1. Nếu $k=0$ cho biết không có mối quan hệ; Nếu $k=1$ cho biết mối quan hệ hoàn toàn chặt chẽ.

TÀI LIỆU THAM KHẢO CHƯƠNG 6

1. TS. Nguyễn Công Nhựt - **Giáo trình Thống kê công nghiệp**. NXB Thống kê, 2013
2. PGS. TS. Trần Ngọc Phác, TS Trần Thị Kim Thu – **Giáo trình Lý thuyết thống kê**. Nhà xuất bản Thống kê, 2013
3. GS.TS Bùi Xuân Phong - **Thống kê và ứng dụng trong BCVT**. NXB Bưu điện, 2005
4. Hà Văn Sơn - **Giáo trình Lý thuyết thống kê ứng dụng trong quản trị và kinh tế**. Nhà xuất bản Thống kê, 2004
5. PTS. Phan Công Nghĩa - **Giáo trình Thống kê Lao động**. NXB Thống kê, 2009
6. PGS.PTS. Nguyễn Thiệp, PTS Phan Công Nghĩa - **Giáo trình Thống kê chất lượng**. NXB Thống kê, 2011

CÂU HỎI ÔN TẬP CHƯƠNG 6

1. Đặc điểm của liên hệ hàm số và liên hệ tương quan? Tại sao khi nghiên cứu các hiện tượng kinh tế xã hội lại thường gặp liên hệ tương quan.
2. Phân tích hồi quy và tương quan giải quyết những nhiệm vụ nghiên cứu gì?
3. Nêu những tính chất của r ?
4. Ý nghĩa của hệ số hồi quy chuẩn hóa?
5. Có tài liệu về tỉ lệ sinh đặc trưng theo tuổi của một địa phương như sau:

Tuổi	Tỉ lệ sinh
15-19	0,035
20-24	0,197
25-29	0,209
30-34	0,155
35-39	0,100
40-44	0,049
45-49	0,014

Yêu cầu:

- a. Xác định tiêu thức nguyên nhân và tiêu thức kết quả.
- b. Biểu hiện mối liên hệ bằng đồ thị.
- c. Xác định mô hình hồi quy phản ánh mối liên hệ giữa tiêu thức nguyên nhân và tiêu thức kết quả.
- d. Đánh giá mức độ chặt chẽ của mối liên hệ.

CHƯƠNG 7

PHÂN TÍCH DÃY SỐ THỜI GIAN

7.1. Khái niệm, phân loại và ý nghĩa của dãy số thời gian

7.1.1. Khái niệm

Mặt lượng của các hiện tượng kinh tế không ngừng biến động theo thời gian. Để nghiên cứu sự biến động này, người ta thường sử dụng dãy số thời gian. Dãy số thời gian là *dãy các trị số của chỉ tiêu thống kê được sắp xếp theo thứ tự thời gian*.

- *Thành phần dãy số thời gian*: Mỗi dãy số thời gian có hai thành phần: thời gian và mức độ của chỉ tiêu về hiện tượng nghiên cứu, trong đó:

+ *Thời gian của dãy số*: tùy theo mục đích nghiên cứu và chỉ tiêu biến đổi nhiều hay ít mà có thể biểu thị bằng ngày, tháng, năm. Độ dài giữa hai khoảng thời gian kế tiếp nhau gọi là khoảng thời gian. Có trường hợp khoảng thời gian đều nhau, có trường hợp khoảng thời gian là không đều nhau.

+ *Mức độ của chỉ tiêu về hiện tượng*: được phản ánh bằng các trị số của chỉ tiêu gọi là các mức độ của dãy số và nó có thể biểu diễn bằng số tuyệt đối, số tương đối hay số trung bình.

7.1.2. Phân loại dãy số thời gian

Căn cứ vào đặc điểm của yếu tố thời gian trong dãy số thì dãy số thời gian được chia thành 2 loại:

+ *Dãy số thời kỳ*: là dãy số phản ánh mặt lượng của hiện tượng qua từng thời kỳ, tức là dãy số phản ánh quy mô, khối lượng của hiện tượng nghiên cứu trong từng khoảng thời gian nhất định (năm, tháng, quý). Khoảng thời gian trong dãy số càng dài thì trị số của chỉ tiêu càng lớn vì thế có thể cộng các trị này với nhau để phản ánh mặt lượng của hiện tượng trong thời kỳ dài hơn. Ví dụ số liệu ở bảng 7.1 là dãy số thời kỳ phản ánh doanh thu của doanh nghiệp trong các năm từ 2003 đến 2009.

+ *Dãy số thời điểm*: phản ánh mặt lượng của hiện tượng nghiên cứu tại các thời điểm nhất định. Mức độ của thời điểm sau thường bao gồm toàn bộ hoặc một bộ phận mức độ của thời điểm trước đó. Vì vậy việc cộng các trị số của chỉ tiêu không phản ánh quy mô của hiện tượng.

Các mức độ của dãy số trên chỉ phản ánh giá trị hàng hóa tồn kho vào ngày đầu tháng, các ngày khác trong tháng thì giá trị hàng tồn kho có thể thay đổi do việc xuất nhập hàng hóa thường xuyên diễn ra trong quá trình kinh doanh.

Các dãy số thời kỳ và dãy số thời điểm được gọi là dãy số tuyệt đối. trên cơ sở các dãy số tuyệt đối có thể xây dựng các dãy số tương đối hoặc dãy số bình quân, trong đó các mức độ của dãy số là các số tương đối hoặc các số bình quân.

7.1.3. Ý nghĩa của dãy số thời gian

Dãy số thời gian giúp cho thống kê nghiên cứu các đặc điểm về sự biến động của hiện tượng và tính qui luật của phát triển của hiện tượng theo thời gian, đồng thời để dự đoán mức độ của hiện tượng trong tương lai.

7.1.4 Yêu cầu đối với dãy số thời gian

Để phản ánh một cách chính xác sự phát triển của hiện tượng theo thời gian thì khi xây dựng một dãy số thời gian phải đảm bảo tính chất có thể so sánh được giữa các mức độ của trong dãy số. Cụ thể là nội dung và phương pháp tính chỉ tiêu qua các thời gian khác nhau phải thống nhất. Phạm vi của tổng thể nghiên cứu trước sau phải nhất trí. Các khoảng cách thời gian trong dãy số nên bằng nhau để tiện lợi cho việc tính toán.

7.2 Phân tích đặc điểm biến động của hiện tượng qua thời gian

7.2.1 Mức độ bình quân theo thời gian

Mức độ bình quân theo thời gian là chỉ tiêu tổng hợp phản ánh mức độ điển hình của hiện tượng trong toàn bộ khoảng thời gian nghiên cứu, hoặc từng giai đoạn nghiên cứu.

Ký hiệu: y_1, y_2, \dots, y_n là các mức độ của dãy số thời gian

\bar{y} - Mức độ trung bình theo thời gian

Tùy theo dãy số thời kỳ hay thời điểm mà có các công thức tính khác nhau:

- Đối với dãy số thời kỳ:

$$\bar{y} = \frac{y_1 + y_2 + y_3 + \dots + y_n}{n} = \frac{\sum_{i=1}^n y_n}{n} \quad (7.1)$$

Trong đó - y_i ($i = 1, 2, \dots, n$) - là mức độ thứ i trong dãy số;

n - là số các mức độ của dãy số.

Đối với dãy số thời điểm

Dãy số thời điểm phản ánh qui mô, khối lượng của hiện tượng trong từng thời điểm. Nếu muốn tính mức độ trung bình chính xác người ta phải xác định trị số chỉ tiêu từng ngày. Nhưng trên thực tế, chỉ có trị số chỉ tiêu vào một ngày nào đó trong tháng nên phải giả thiết rằng người giữa hai thời điểm điều tra mật độ của hiện tượng tăng giảm đều đặn.

+ Dãy số thời điểm có khoảng cách thời gian bằng nhau:

$$\bar{y} = \frac{\frac{y_1 + y_2}{2} + \frac{y_2 + y_3}{2} + \frac{y_3 + y_4}{2} + \dots + \frac{y_{n-1} + y_n}{2}}{n-1}$$

$$\Rightarrow \bar{y} = \frac{\frac{y_1}{2} + y_2 + y_3 + \dots + \frac{y_n}{2}}{n-1} \quad (7.2)$$

Trong đó: y_i - Mức độ thứ i của dãy số thời điểm có khoảng cách thời gian bằng nhau.

+ Dãy số thời điểm có khoảng cách thời gian không bằng nhau:

$$\bar{y} = \frac{\sum_{i=1}^n y_i t_i}{\sum_{i=1}^n t_i} \quad (7.3)$$

Trong đó: y - là mức độ thứ i trong dãy số;

t_i - là độ dài thời gian có mức độ y_i .

7.2.2. Lượng tăng (hoặc giảm) tuyệt đối

Là chỉ tiêu phản ánh sự thay đổi tuyệt đối về mức độ của hiện tượng theo thời gian. Nó được xác định bằng hiệu số giữa hai mức độ trong dãy số. Tùy theo mục đích nghiên cứu mà có các cách tính:

1. Lượng tăng (hoặc giảm) tuyệt đối liên hoàn (từng kỳ):

Là chênh lệch giữa mức độ kỳ nghiên cứu (y_i) với mức độ của kỳ đứng liền trước nó (y_{i-1}), nhằm phản ánh mức độ tăng (hoặc giảm) tuyệt đối giữa hai thời gian liền nhau:

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

Trong đó:

δ_i - lượng tăng (hoặc giảm) tuyệt đối liên hoàn (từng kỳ) ở thời gian i so với thời gian đứng liền trước đó là $i-1$

y_i - mức độ tuyệt đối ở thời gian i

y_{i-1} - mức độ tuyệt đối ở thời gian $i-1$

2. Lượng tăng (hoặc giảm) tuyệt đối định gốc:

Là hiệu số giữa mức độ của kỳ nghiên cứu (y_i) với mức độ của một kỳ được chọn làm gốc cố định - thường là mức độ đầu tiên của dãy số y_1 , nhằm để phản ánh mức độ tăng (giảm) của hiện tượng trong khoảng thời gian dài.

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

Trong đó:

y_i - mức độ tuyệt đối ở thời gian i

y_1 - mức độ tuyệt đối ở thời gian đầu.

Xét các trường hợp sau:

+ Nếu $\delta_i, \Delta_i > 0$ thì mức độ hiện tượng tăng;

+ Nếu $\delta_i, \Delta_i = 0$ thì mức độ hiện tượng không thay đổi;

+ Nếu $\delta_i, \Delta_i < 0$ thì mức độ hiện tượng giảm;

+ Tổng đại số của các lượng tăng (giảm) tuyệt đối từng kỳ bằng lượng tăng (giảm) tuyệt đối định gốc tương ứng.

$$(y_2 - y_1) + (y_3 - y_2) + \dots + (y_i - y_{i-1}) = y_i - y_1 \quad \Leftrightarrow \quad \sum \delta_i = \Delta_i$$

3. Lượng tăng (hoặc giảm) tuyệt đối bình quân

Là số trung bình cộng của các lượng tăng (hoặc giảm) tuyệt đối liên hoàn. Nó phản ánh mức độ tăng trung bình của hiện tượng nghiên cứu trong thời kỳ dài:

$$\bar{\delta} = \frac{\sum_{i=2}^n \delta_i}{n-1} = \frac{(y_2 - y_1) + (y_3 - y_2) + \dots + (y_i - y_{i-1})}{n-1} = \frac{y_n - y_1}{n-1} \quad (7.4)$$

Trong đó: $\bar{\delta}$ - nói lên trong một đơn vị thời gian mức độ của hiện tượng tăng (hoặc giảm) bao nhiêu.

7.2.3. Tốc độ phát triển

Là chỉ tiêu tương đối phản ánh xu hướng phát triển của hiện tượng nghiên cứu qua thời gian. Chỉ tiêu này được xác định bằng tỷ số giữa hai mức độ của hiện tượng ở hai thời kỳ hoặc hai thời điểm.

Tuỳ theo mục đích nghiên cứu, ta có các loại tốc độ phát triển sau đây:

1. Tốc độ phát triển liên hoàn (từng kỳ)

Là tỷ số giữa mức độ của kỳ nghiên cứu (y_i) với mức độ của kỳ đứng ngay trước đó (y_{i-1}). Chỉ tiêu này phản ánh sự phát triển của hiện tượng giữa hai thời gian liên nhau:

$$t_i = \frac{y_i}{y_{i-1}} \quad (i = 2, 3, \dots, n) \quad (7.5)$$

$$\text{hay} \quad t_i = \frac{y_i}{y_{i-1}} 100 \quad (\%)$$

trong đó:

t_i – tốc độ phát triển liên hoàn thời gian I so với thời gian i-1 và có biểu hiện bằng lần hoặc %.

2. Tốc độ phát triển định gốc

Là tỷ số giữa mức độ của kỳ nghiên cứu (y_i) với mức độ của một kỳ được chọn làm gốc cố định, thường là mức độ đầu tiên của dãy số (y_1). Chỉ tiêu này biểu hiện sự phát triển của hiện tượng trong các khoảng thời gian dài.

$$T_i = \frac{y_i}{y_1} \quad \text{hay} \quad T_i = \frac{y_i}{y_1} 100 \quad (i = 2, 3, \dots, n) \quad (7.6)$$

T_i – Tốc độ phát triển định gốc thời gian I so với thời gian đầu của dãy số và có biểu hiện bằng lần hoặc %.

Giữa tốc độ phát triển liên hoàn và tốc độ phát triển định gốc có mối quan hệ sau:

- Tích các tốc độ phát triển liên hoàn bằng tốc độ phát triển định gốc, nghĩa là:

$$\frac{y_2}{y_1} \cdot \frac{y_3}{y_2} \cdot \frac{y_4}{y_3} \dots \frac{y_i}{y_{i-1}} = \frac{y_i}{y_1} \Leftrightarrow \prod t_i = T_i \quad (7.7)$$

- Thương của 2 tốc độ phát triển định gốc liên nhau bằng tốc độ phát triển liên hoàn giữa 2 thời gian:

$$\frac{T_i}{T_{i-1}} = \frac{\frac{y_i}{y_1}}{\frac{y_{i-1}}{y_1}} = \frac{y_i}{y_{i-1}} = t_i \quad (i = 2, 3, \dots, n) \quad (7.8)$$

3. Tốc độ phát triển bình quân

Là số trung bình cộng của các tốc độ phát triển liên hoàn. Chỉ tiêu này biểu hiện tốc độ phát triển trung bình của hiện tượng trong suốt thời gian nghiên cứu.

$$\bar{t} = \sqrt[n-1]{t_2 t_3 \dots t_n} = \sqrt[n-1]{\prod_2^n t_i} = \sqrt[n-1]{\frac{y_n}{y_1}} \quad (7.9)$$

Từ công thức tính là tốc độ phát triển bình quân cho thấy: chỉ nên tính chỉ tiêu này đối với những hiện tượng biến động theo một xu hướng nhất định.

7.2.4. Tốc độ tăng (hoặc giảm)

Là chỉ tiêu tương đối phản ánh mức độ của hiện tượng nghiên cứu giữa 2 thời gian đã tăng hoặc giảm bao nhiêu lần (hoặc bao nhiêu phần trăm). Tùy theo mục đích nghiên cứu ta có:

1. Tốc độ tăng (hoặc giảm) liên hoàn

Là tỷ số giữa lượng tăng (hoặc giảm) tuyệt đối liên hoàn với mức độ kỳ gốc liên hoàn (kỳ gốc đứng ngay trước nó).

$$a_i = \frac{y_i - y_{i-1}}{y_{i-1}} = t_i - 1 \quad (i = 2, 3, \dots, n) \quad (7.10)$$

Hay $a_i = t_i - 100$ (%)

2. Tốc độ tăng (hoặc giảm) định gốc

Là tỷ số giữa lượng tăng (hoặc giảm) tuyệt đối định gốc và mức độ kỳ gốc cố định:

$$A_i = \frac{y_i - y_1}{y_1} = T_i - 1 \quad (i = 2, 3, \dots, n) \quad (7.11)$$

Hay $A_i = T_i - 100$ (%)

Vậy tốc độ tăng (giảm) liên hoàn hay định gốc bằng tốc độ phát triển liên hoàn hay định gốc trừ 1 (hay 100%)

3. Tốc độ tăng (hoặc giảm) bình quân

Là chỉ tiêu tương đối nói lên nhịp điệu tăng (hoặc giảm) diễn hình của hiện tượng nghiên cứu trong thời gian dài.

$$\bar{a} = \bar{t} - 1 \quad (7.12)$$

Nếu \bar{t} tính bằng % thì $\bar{a} = \bar{t} - 100$

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

Chỉ tiêu này phản ánh cứ 1% tăng (hoặc giảm) của tốc độ tăng (hoặc giảm) liên hoàn thì tương ứng với 1 trị số tuyệt đối là bao nhiêu?

$$g_i = \frac{\Delta_i}{a_i} = \frac{y_i - y_{i-1}}{\frac{y_i - y_{i-1}}{y_{i-1}}} = \frac{y_{i-1}}{100} \quad (i = 2, 3, \dots, n) \quad (7.13)$$

Với a_i tính bằng %.

7.3 Phân tích các thành phần của dãy số thời gian

Các mức độ của dãy số thời gian y có thể được chia thành ba thành phần sau đây:

- Xu thế, ký hiệu f , phản ánh xu hướng phát triển cơ bản của hiện tượng kéo dài theo thời gian.
- Thời vụ, ký hiệu s , sự biến động có tính chất lặp đi lặp lại trong những khoảng thời gian nhất định của năm.
- Ngẫu nhiên, ký hiệu z , sự biến động do các yếu tố ngẫu nhiên xảy ra ở những thời gian khác nhau.

Ba thành phần trên đây được kết hợp lại với nhau theo một trong hai dạng sau đây:

- Kết hợp cộng:

$$y = f + s + z$$

- Kết hợp nhân:

$$y = f.s.z$$

7.3.1 Phân tích các thành phần theo kết hợp cộng

Trường hợp đơn giản được giả thiết rằng hàm xu thế là hàm tuyến tính:

$$f_t = a_0 + a_1 t$$

Với $t = 1, 2, 3, \dots$ là thứ tự thời gian trong dãy số.

Khi đó việc xác định xu thế và biến động thời vụ được dựa vào bảng sau:

Quý/tháng (j) Năm (i)	1	...	Tổng năm T_i	iT_i
1				
2				
...				
N				
Tổng quý/tháng T_j			$T = \sum T_j$	$S = \sum iT_i$
BQ quý \bar{y}_j				

Trong đó: $i = 1, 2, 3, \dots, n$ là số năm.

$j = 1, 2, 3, \dots, m$ là các khoảng thời gian dưới một năm $m = 4$ đối với quý, $m = 12$ đối với tháng.

Sau đó. Các hệ số của hàm xu thế tuyến tính và thời vụ được tính theo các công thức sau đây:

$$a_1 = \frac{12}{m.n.(n^2-1)} \left(\frac{S}{m} - \frac{n+1}{2m} T \right) \quad (7.14)$$

$$a_0 = \frac{T}{m.n} - a_1 \frac{m.n+1}{2} \quad (7.15)$$

$$s_t \equiv s_i = \bar{y}_j - \bar{y}_0 - a_1 \left(j - \frac{m+1}{2} \right) \quad (7.16)$$

Sau khi đã xác định được thành phần xu thế và thời vụ thì thành phần ngẫu nhiên được xác định như sau:

$$z_t = y_t - f_t - s_t$$

7.3.2 Phân tích các thành phần theo kết hợp nhân

Để phân tích thành phần của dãy số thời gian y_t theo kết hợp nhân, trước hết cần phải loại trừ thành phần thời vụ và thành phần ngẫu nhiên bằng cách xây dựng dãy số bình quân trượt \bar{y}_t với số lượng mức độ bằng 4 đối với tài liệu quý và bằng 12 đối với tài liệu tháng.

Từ dãy số y_t và dãy số bình quân trượt \bar{y}_t , tính $\frac{y_t}{\bar{y}_t}$. Từ đó xác định thành phần thời vụ s_t bằng tính các số bình quân \bar{s}_j . Sau đó tính hệ số điều chỉnh H, với $m = 4$ đối với quý, $m = 12$ đối với tháng.

Từ đó tính:

$$s_t = \bar{s}_j \cdot H$$

Sau khi đã xác định được s_t thì xác định dãy số y'_t là dãy số đã loại bỏ thành phần thời vụ như sau:

$$y'_t = \frac{y_t}{s_t}$$

Từ dãy số y'_t đi xây dựng hàm xu thế.

Cuối cùng, thành phần ngẫu nhiên được xác định bởi công thức sau:

$$z_t = \frac{y_t}{f_t \cdot s_t}$$

Trên đây đã trình bày phân tích các thành phần dãy số thời gian theo kết hợp cộng và kết hợp nhân. Vấn đề đặt ra là: đối với dãy số thời gian nào thì phân tích theo kết hợp cộng hoặc kết hợp nhân?

Nhiều nghiên cứu cho thấy, nếu đồ thị biểu hiện sự biến động của hiện tượng qua thời gian có biên độ dao động ít thì có thể phân tích theo kết hợp cộng, nếu có biên độ dao động thay đổi nhiều thì có thể phân tích theo kết hợp nhân.

TÀI LIỆU THAM KHẢO CHƯƠNG 7

1. TS. Nguyễn Công Nhự - **Giáo trình Thống kê công nghiệp**. NXB Thống kê, 2013
2. Hà Văn Sơn - **Giáo trình Lý thuyết thống kê ứng dụng trong quản trị và kinh tế**. Nhà xuất bản Thống kê, 2004
3. PTS. Phan Công Nghĩa - **Giáo trình Thống kê Lao động**. NXB Thống kê, 2009
4. PGS.PTS. Nguyễn Thiệp, PTS Phan Công Nghĩa - **Giáo trình Thống kê chất lượng**. NXB Thống kê, 2011
5. PGS Nguyễn Thị Kim Thúy. **Nguyên lý Thống kê ứng dụng trong quản lý kinh tế và kinh doanh sản xuất dịch vụ**. Nhà Xuất bản thống kê - 2009

CÂU HỎI ÔN TẬP CHƯƠNG 7

1. Thế nào là dãy số thời gian? Có mấy loại dãy số thời gian?
2. Phân tích các yêu cầu khi xây dựng dãy số thời gian?
3. Ý nghĩa của việc nghiên cứu dãy số thời gian?
4. Phương pháp tính các chỉ tiêu phân tích dãy số thời gian?
5. Có tài liệu về tình hình sản xuất của doanh nghiệp trong ba tháng đầu năm 2010 như sau:

Chỉ tiêu	Tháng 1	Tháng 2	Tháng 3
GO thực tế (tỉ đồng)	3,8	3,4	4,2
Tỉ lệ % hoàn thành kế hoạch GO	105	102	104
Số công nhân ngày đầu tháng (ng)	204	200	206

Số công nhân ngày ¼ là 208 người. Hãy tính:

- a. Giá trị sản xuất thực tế bình quân một tháng của quý 1.
- b. Số công nhân bình quân mỗi tháng và cả quý 1.
- c. Năng suất lao động bình quân mỗi tháng của một công nhân.

- d. Năng suất lao động bình quân một tháng trong quý 1 của một công nhân.
- e. Tỷ lệ hoàn thành kế hoạch bình quân một tháng của quý 1.

CHƯƠNG 8

CHỈ SỐ PHÁT TRIỂN VÀ CHỈ SỐ KẾ HOẠCH KINH TẾ

8.1 Khái niệm, ý nghĩa, phân loại chỉ số

8.1.1 Khái niệm chỉ số

Chỉ số trong thống kê là một loại chỉ tiêu tương đối biểu hiện mối quan hệ so sánh giữa hai mức độ nào đó của một hiện tượng kinh tế.

Chỉ số được tính bằng cách so sánh 2 mức độ của hiện tượng ở 2 thời gian khác nhau nhằm nêu lên biến động của hiện tượng theo thời gian. Chỉ số này được biểu hiện bằng số tương đối giống như số tương đối động thái. Trong những mục đích so sánh khác, các chỉ số tính được cũng là số tương đối, có thể là số tương đối kế hoạch, số tương đối so sánh.

Như vậy trong thống kê khái niệm chỉ số tương đối rộng rãi; nó là phương pháp biểu hiện các quan hệ so sánh khác nhau. Trong thực tế, đối tượng chủ yếu của phương pháp phức tạp bao gồm nhiều đơn vị, nhiều phần tử có tính chất khác nhau về tên gọi, về giá trị sử dụng, đơn vị tính...v.v. Muốn so sánh các mức độ của hiện tượng kinh tế phức tạp, cần biến đổi các đơn vị, các phần tử có tính chất khác nhau trở thành dạng đồng nhất.

8.1.2 Đặc điểm của chỉ số

Để đáp ứng yêu cầu so sánh các mức độ của hiện tượng phức tạp, chỉ số có đặc điểm là khi muốn so sánh các mức độ của hiện tượng kinh tế phức tạp, cần phải chuyển các đơn vị hoặc phần tử có tính chất khác nhau thành dạng giống nhau, để có thể cộng trực tiếp với nhau. Như khi các đơn vị có các sản phẩm dịch vụ khác nhau không thể trực tiếp cộng với nhau do đó phải chuyển về dạng giống nhau (dạng giá trị). Như vậy dễ dàng so sánh hai chỉ tiêu giá trị để tính ra chỉ số giá trị

Ngoài ra, chỉ số còn có đặc điểm là khi có nhiều nhân tố cùng tham gia vào việc tính toán số, phải giả định chỉ có một nhân tố thay đổi, còn các nhân tố khác không thay đổi.

8.1.3 Tác dụng của chỉ số

Chỉ số là một phương pháp không những có khả năng nêu lên biến động tổng hợp của hiện tượng phức tạp, mà còn có thể phân tích sự biến động này. Trong thống kê chỉ số có tác dụng:

- Chỉ số biểu hiện biến động của hiện tượng theo thời gian. Các chỉ số tính theo mục đích này được gọi là chỉ số phát triển. Nó được tính bằng cách so sánh mức độ của hiện tượng ở hai thời gian khác nhau.
- Biểu hiện sự biến động của hiện tượng theo không gian khác nhau giữa hai đơn vị, doanh nghiệp, hai địa phương. Chỉ số này gọi là chỉ số không gian.
- Biểu hiện nhiệm vụ và thực hiện kế hoạch của chỉ tiêu kinh tế nào đó. Chỉ số này gọi là chỉ số kế hoạch.
- Phân tích vai trò và ảnh hưởng của chỉ tiêu nhân tố đến sự biến động của chỉ tiêu tổng hợp.

8.1.4 Phân loại chỉ số

1. Căn cứ vào phạm vi tính toán

- *Chỉ số cá thể*: Biểu hiện sự biến động của từng phần tử, đơn vị cá biệt trong tổng thể phức tạp.
- *Chỉ số tổng hợp*: Biểu hiện sự biến động của tất cả các phần tử, các đơn vị của tổng thể nghiên cứu.

Cơ sở để tính chỉ số tổng hợp là chỉ số cá thể. Nó được dùng nhiều nhất trong thực tế.

2. Căn cứ vào tính chất của chỉ tiêu nghiên cứu

- *Chỉ số của chỉ tiêu số lượng*: Biểu hiện sự biến động của chỉ tiêu số lượng như khối lượng sản phẩm dịch vụ, số lượng công nhân, số lượng thiết bị.
- *Chỉ số của chỉ tiêu chất lượng*: Biểu hiện sự biến động của chỉ tiêu chất lượng như chỉ số giá thành sản phẩm, năng suất lao động...

3. Căn cứ vào điều kiện thời gian và không gian

- Chỉ số thời gian như chỉ số phát triển, chỉ số kế hoạch, chỉ số định gốc, chỉ số liên hoàn.

- Chỉ số không gian biểu hiện quan hệ so sánh các mức độ của hiện tượng nghiên cứu giữa hai doanh nghiệp, hai địa phương..

Trong quá trình tính toán thường sử dụng các ký hiệu sau: i - Chỉ số cá thể; I - Chỉ số tổng hợp; p - Biểu hiện mặt lượng của chỉ tiêu chất lượng; q - Biểu hiện mặt lượng của chỉ tiêu số lượng.

Biểu hiện các mốc thời gian, sử dụng các ký hiệu: 0 – Kỳ gốc; 1 – Kỳ báo cáo hay kỳ thực hiện; KH – Kỳ kế hoạch

Biểu hiện các địa phương, các khu vực hoặc các đơn vị khác nhau, sử dụng các chữ cái A, B, C

8.2 Chỉ số phát triển

Chỉ số thống kê được vận dụng trong phân tích kinh tế đối với nhiều chỉ tiêu thuộc nhiều lĩnh vực, bao gồm chỉ số giá tiêu dùng, chỉ số giá xuất khẩu, chỉ số khối lượng sản phẩm...

Để minh họa phương pháp luận thiết lập và phân tích chỉ số thống kê sau đây đề cập đến chỉ số giá và chỉ số lượng hàng hóa tiêu thụ

8.2.1 Chỉ số đơn (cá thể)

Chỉ số đơn giá:

Biểu hiện quan hệ so sánh giữa mức giá từng mặt hàng qua thời gian. Công thức tính:

$$i_p = \frac{p_1}{p_0}$$

Trong đó: i_p - chỉ số đơn giá

p_1 – giá bán của mặt hàng kỳ nghiên cứu;

p_0 – giá bán của mặt hàng kỳ gốc.

Chỉ số đơn giá phản ánh biến động giá bán của từng mặt hàng ở kỳ nghiên cứu so với kỳ gốc.

Chỉ số đơn lượng hàng tiêu thụ:

Biểu hiện quan hệ so sánh giữa khối lượng tiêu thụ của từng mặt hàng ở hai thời gian:

$$i_q = \frac{q_1}{q_0}$$

Trong đó: i_q - chỉ số đơn lượng hàng tiêu thụ

q_1 – khối lượng tiêu thụ của mặt hàng kỳ nghiên cứu;

q_0 - khối lượng tiêu thụ của mặt hàng kỳ gốc.

Chỉ số đơn lượng hàng tiêu thụ phản ánh biến động khối lượng tiêu thụ của từng mặt hàng ở kỳ nghiên cứu so với kỳ gốc.

8.2.2 Chỉ số tổng hợp

1. Chỉ số tổng hợp giá

Chỉ số tổng hợp giá biểu hiện quan hệ so sánh giữa giá bán của một nhóm hay toàn bộ các mặt hàng ở kỳ nghiên cứu với kỳ gốc và qua đó phản ánh biến động chung giá bán của các mặt hàng. Như đã đề cập về các đặc điểm của phương pháp chỉ số, khi xây dựng chỉ số tổng hợp giá không thể tổng hợp đơn thuần, nghĩa là cộng giá đơn vị của các mặt hàng ở kỳ nghiên cứu và kỳ gốc. Bản thân việc cộng giá đơn vị của các mặt hàng là không có ý nghĩa và đồng thời bỏ qua tình hình tiêu thụ thực tế của mỗi mặt hàng có tầm quan trọng khác nhau. Để đưa về đại lượng có thể tổng hợp được, khi xây dựng chỉ số tổng hợp giá, phải nhân giá mỗi mặt hàng với lượng tiêu thụ tương ứng trên cơ sở đó thiết lập quan hệ so sánh. Bằng cách thiết lập như vậy, chỉ số tổng hợp giá được biểu hiện qua công thức:

$$I_p = \frac{\sum p_1 q}{\sum p_0 q}$$

Trong đó:

I_p – Chỉ số tổng hợp giá;

p_0 và p_1 – Giá bán mỗi mặt hàng kỳ gốc và kỳ nghiên cứu;

q – Lượng tiêu thụ của mỗi mặt hàng.

Trong công thức trên, lượng tiêu thụ mỗi mặt hàng (q) đã tham gia vào công thức tính chỉ số giá và giữ vai trò là quyền số phản ánh tầm quan trọng của từng mặt hàng trong sự biến động chung của giá.

Mặt khác muốn nghiên cứu biến động chỉ số của nhân tố giá thì giá bán các mặt hàng ở hai kỳ phải được tổng hợp theo cùng một lượng hàng hóa tiêu thụ, nghĩa là cố định ở một kỳ nào đó trong cả tử và mẫu số ở mỗi công thức. Tùy theo mục đích nghiên cứu và điều kiện tài liệu tổng hợp được thực tế, Chỉ số tổng hợp giá cả có thể được xác định theo các công thức sau:

Chỉ số tổng hợp giá cả Laspeyres: Là chỉ số tổng hợp giá với quyền số là khối lượng tiêu thụ của mỗi mặt hàng ở kỳ gốc.

Công thức:

$$I_p^L = \frac{\sum p_1 q_0}{\sum p_0 q_0}$$

Với quyền số kỳ gốc, chỉ số tổng hợp giá *Laspeyres* phản ánh biến động của giá bán các mặt hàng ở kỳ nghiên cứu so với kỳ gốc và ảnh hưởng biến động riêng của giá cả đối với mức tiêu thụ các mặt hàng. Chênh lệch giữa tử số và mẫu số phản ánh lượng tăng giảm mức tiêu thụ do ảnh hưởng biến động của giá bán các mặt hàng với giả định lượng tiêu thụ ở kỳ nghiên cứu cũng như kỳ gốc.

Chỉ số tổng hợp giá cả *Laspeyres* với quyền số là lượng hàng hóa tiêu thụ kỳ gốc có hạn chế là không phản ánh cập nhật được những thay đổi về khuynh hướng tiêu dùng đồng thời không cho phép xác định được lượng tăng hay giảm thực tế của mức tiêu thụ do ảnh

hưởng biến động giá bán các mặt hàng. Tuy nhiên về mặt tính toán, áp dụng công thức chỉ số Laspeyres thường gặp nhiều thuận lợi vì dữ liệu về khối lượng tiêu thụ của các mặt hàng ở kỳ gốc đã được tổng hợp.

Chỉ số tổng hợp giá cả Passche: Là chỉ số tổng hợp giá với quyền số là khối lượng tiêu thụ của mỗi mặt hàng ở nghiên cứu.

Công thức:

$$I_p^P = \frac{\sum p_1 q_1}{\sum p_0 q_1}$$

Chỉ số tổng hợp giá cả Passche sử dụng quyền số là lượng tiêu thụ các mặt hàng kỳ nghiên cứu, do đó trong trường hợp có sự thay đổi lớn về khối lượng và cơ cấu tiêu thụ các mặt hàng thì sẽ không phản ánh ảnh hưởng biến động riêng của giá đối với mức tiêu thụ các mặt hàng. Tuy nhiên với quyền số q_1 chênh lệch giữa tử số và mẫu số của chỉ số giá Passche phản ánh lượng tăng hay giảm thực tế của mức tiêu thụ do ảnh hưởng biến động giá bán các mặt hàng.

Kết quả tính chỉ số tổng hợp giá cả theo công thức chỉ số Laspeyres và Passche thường có sự chênh lệch. Nguyên nhân cơ bản đó là sự khác biệt về thời kỳ quyền số và cũng được hiểu là kết quả của sự thay đổi cơ cấu tiêu thụ của các mặt hàng giữa hai kỳ. Hơn nữa, bản chất chỉ số tổng hợp giá Laspeyres và Passche đều có thể được xác định từ các chỉ số đơn giá nhưng theo các công thức bình quân khác nhau. Trong điều kiện cơ cấu tiêu thụ các mặt hàng không thay đổi thì chỉ số tổng hợp giá cả Passche có kết quả thấp hơn so với chỉ số tổng hợp giá Laspeyres. Tuy nhiên, thực tế luôn có nhiều nhân tố ảnh hưởng đến sự thay đổi cơ cấu tiêu thụ của các mặt hàng đó là sự thay đổi thu nhập, thị hiếu dân cư,... Sự thay đổi lớn về cơ cấu tiêu thụ các mặt hàng có thể dẫn đến kết quả tính chỉ số Laspeyres và Passche có chênh lệch lớn. Do vậy để phản ánh biến động chung về giá bán các mặt hàng có thể sử dụng một chỉ số thay thế cho hai chỉ số trên đó là chỉ số tổng hợp giá Fisher.

Chỉ số tổng hợp giá cả Fisher

Chỉ số tổng hợp giá cả Fisher phản ánh biến động chung giá bán của các mặt hàng dựa trên cơ sở san bằng chênh lệch giữa các chỉ số Laspeyres và Passche theo công thức bình quân nhân như sau:

$$I_p^F = \sqrt{\frac{\sum p_1 q_0}{\sum p_0 q_0} \times \frac{\sum p_1 q_1}{\sum p_0 q_1}}$$

Có thể thấy rằng chỉ số Fisher sử dụng kết hợp cả quyền số kỳ gốc và kỳ nghiên cứu nên có thể khắc phục được những ảnh hưởng về sự khác biệt cơ cấu tiêu thụ các mặt hàng giữa hai kỳ và qua đó xác định được kết quả chung phản ánh biến động giá bán các mặt hàng.

2. Chỉ số tổng hợp lượng hàng hóa tiêu thụ

Chỉ số tổng hợp lượng hàng hóa tiêu thụ biểu hiện quan hệ so sánh giữa khối lượng tiêu thụ của một nhóm hay toàn bộ các mặt hàng thuộc phạm vi nghiên cứu giữa hai thời gian và qua đó phản ánh biến động chung về khối lượng tiêu thụ của các mặt hàng.

Trong công thức tính chỉ số tổng hợp lượng hàng hóa tiêu thụ, giá bán của các mặt hàng giữ vai trò là nhân tố thông ước chung chuyển các mặt hàng có khối lượng tiêu thụ tính theo đơn vị khác nhau về cùng đơn vị giá trị để tổng hợp và thiết lập mối quan hệ so sánh. Đồng thời để phản ánh biến động của khối lượng tiêu thụ các mặt hàng thì nhân tố giá được giữ cố định làm quyền số trong công thức chỉ số hợp lượng hàng hóa tiêu thụ như sau:

$$I_q = \frac{\sum pq_1}{\sum pq_0}$$

Trong đó:

I_q – Chỉ số tổng hợp lượng hàng hóa tiêu thụ;

q_0 và q_1 – Lượng hàng hóa tiêu thụ mỗi mặt hàng kỳ gốc và kỳ nghiên cứu;

p – Giá bán của mỗi mặt hàng.

Như vậy khối lượng tiêu thụ chung của mỗi nhóm hay toàn bộ các mặt hàng thuộc phạm vi nghiên cứu ở mỗi kỳ đều không tổng hợp được theo một đơn vị hiện vật mà theo đơn vị giá trị. Tùy theo điều kiện dữ liệu thực tế, khối lượng tiêu thụ của các mặt hàng có thể được tổng hợp theo giá bán các mặt hàng kỳ gốc hay kỳ nghiên cứu. Xét ở kỳ nghiên cứu, nếu khối lượng tiêu thụ các mặt hàng được tổng hợp theo giá kỳ nghiên cứu thì chỉ tiêu thể hiện theo công thức $\sum p_1 q_1$ mang ý nghĩa là tổng giá trị hàng hóa tiêu thụ kỳ nghiên cứu theo giá hiện hành. Nếu tổng hợp theo giá kỳ gốc thì chỉ tiêu $\sum p_0 q_1$ mang ý nghĩa là tổng giá trị hàng hóa tiêu thụ kỳ nghiên cứu theo giá so sánh kỳ gốc.

Chỉ số tổng hợp lượng hàng hóa tiêu thụ Laspeyres

Chỉ số tổng hợp lượng hàng hóa tiêu thụ Laspeyres phản ánh biến động chung của lượng hàng hóa tiêu thụ và ảnh hưởng biến động đó đối với mức tiêu thụ các mặt hàng. Với quyền số là giá bán các mặt hàng kỳ gốc, chỉ số tổng hợp lượng hàng hóa tiêu thụ Laspeyres thể hiện theo công thức sau:

$$I_q^L = \frac{\sum p_0 q_1}{\sum p_0 q_0}$$

Chỉ số tổng hợp lượng hàng hóa tiêu thụ Passche

Chỉ số tổng hợp lượng hàng hóa tiêu thụ Passche có quyền số là giá bán các mặt hàng kỳ nghiên cứu theo công thức như sau:

$$I_q^P = \frac{\sum p_1 q_1}{\sum p_1 q_0}$$

Chỉ số tổng hợp lượng hàng hóa tiêu thụ Fisher

Chỉ số tổng hợp lượng hàng hóa tiêu thụ Fisher sử dụng kết hợp quyền số là giá các mặt hàng kỳ gốc và kỳ nghiên cứu thông qua công thức sau:

$$I_q^F = \sqrt{\frac{\sum p_0 q_1}{\sum p_0 q_0} \times \frac{\sum p_1 q_1}{\sum p_1 q_0}}$$

3. Quyền số của chỉ số phát triển

Quyền số của chỉ số là những đại lượng được giữ cố định trong công thức chỉ số tổng hợp. Như trong công thức chỉ số tổng hợp giá, lượng tiêu thụ các mặt hàng được cố định ở cả tử số và mẫu số và giữ vai trò là quyền số.

Trong công thức chỉ số tổng hợp, quyền số có hai tác dụng:

- Làm cho các phần tử với đại lượng biểu hiện không thể trực tiếp cộng được với nhau chuyển về cùng một đại lượng đồng nhất và có thể tổng hợp.
- Biểu hiện vai trò và tầm quan trọng của mỗi phần tử hay bộ phận trong toàn bộ tổng thể.

Trong từng chỉ số cụ thể, quyền số có thể thực hiện được một hoặc cả hai chức năng nói trên. Ví dụ quyền số của chỉ số tổng hợp giá là lượng tiêu thụ của các mặt hàng giúp cho việc so sánh giá giữa hai kỳ của các mặt hàng có tầm quan trọng khác nhau trong cơ cấu các mặt hàng. Quyền số của chỉ số tổng hợp lượng hàng hóa tiêu thụ là giá bán các mặt hàng có tác dụng như một nhân tố thông ước chung chuyển khối lượng tiêu thụ các mặt hàng theo đơn vị tính khác nhau về cùng dạng giá trị để tổng hợp. Đồng thời quyền số của từng mặt hàng với giá khác nhau cũng thể hiện tầm quan trọng khác nhau trong tổng mức tiêu thụ các mặt hàng.

Vấn đề chọn quyền số có ý nghĩa quan trọng đối với việc tính chỉ số tổng hợp vì nó quyết định ý nghĩa của mỗi chỉ số.

8.2.3 Chỉ số không gian

1. Khái niệm chỉ số không gian

Là chỉ số biểu hiện quan hệ so sánh mức độ của một hiện tượng nghiên cứu trong điều kiện không gian khác nhau (các doanh nghiệp khác nhau hoặc trên các thị trường khác nhau).

2. Phương pháp tính chỉ số không gian

a. Tính chỉ số không gian của chỉ tiêu số lượng:

Chỉ số không gian của chỉ tiêu số lượng biểu hiện mối quan hệ so sánh mức độ của hiện tượng nghiên cứu giữa hai doanh nghiệp hoặc hai địa phương khác nhau. Khi đó nếu q_A và q_B là giá trị của chỉ tiêu số lượng của 2 doanh nghiệp A và B và p là giá trị của chỉ tiêu chất lượng có liên quan được chọn làm quyền số của chỉ số không gian của chỉ tiêu số lượng ta có:

$$I_q(A/B) = \frac{\sum q_i^A p}{\sum q_i^B p}$$

Trong đó: p - là chỉ tiêu chất lượng có liên quan và đóng vai trò là quyền số, vì vậy nó phải mang tính đặc trưng của cả hai doanh nghiệp. Có thể chọn p theo hai cách sau:

- Nếu giá trị của p đã được nhà nước hoặc các bộ, ngành qui định chung cho các doanh nghiệp thì lấy giá trị đó làm quyền số. Ví dụ như khi tính chỉ số khối lượng sản phẩm thì quyền số là giá cố định do nhà nước qui định.

$$I_q(A/B) = \frac{\sum q_i^A p_{cd}}{\sum q_i^B p_{cd}}$$

Trong đó; pcd - là giá do nhà nước ban hành.

- Nếu giá trị của p chưa thống nhất giữa các địa phương thì phải tính toán giá trị trung bình:

$$\overline{p_i} = \frac{p_i^A q_i^A + p_i^B q_i^B}{q_i^A + q_i^B}$$

$$I_q(A/B) = \frac{\sum q_i^A \overline{p_i}}{\sum q_i^B \overline{p_i}}$$

Trong đó: q_i^A, q_i^B - là khối lượng hàng loại i bán ra trên thị trường A, B.

$\overline{p_i}$ là giá cả trung bình của mặt hàng i trên cả 2 thị trường.

b. Chỉ số không gian của chỉ tiêu chất lượng.

Khi xây dựng chỉ số không gian của chỉ tiêu chất lượng thì quyền số được chọn là chỉ tiêu số lượng có liên quan. Công thức tính:

$$I_p(A/B) = \frac{\sum p_i^A (q_i^A + q_i^B)}{\sum p_i^B (q_i^A + q_i^B)}$$

Trong đó: p_i^A, p_i^B - là giá trị của chỉ tiêu chất lượng tại địa phương A, B. Ví dụ: giá cả loại hàng i trên thị trường A, B

q_i^A, q_i^B - là giá trị của chỉ tiêu số lượng có liên quan tại địa phương A, B.

8.3 Chỉ số kết hoạch

Chỉ số kế hoạch biểu hiện nhiệm vụ kế hoạch hoặc tình hình thực hiện kế hoạch đối với từng chỉ tiêu. Khi thiết lập và tính các chỉ số tổng hợp phân tích kế hoạch đối với chỉ tiêu nghiên cứu, việc lựa chọn quyền số cũng căn cứ vào đặc điểm dữ liệu và mục đích nghiên cứu.

Trong trường hợp phân tích kế hoạch giá thành các sản phẩm của một doanh nghiệp, quyền số có thể là sản lượng thực tế hoặc sản lượng kế hoạch.

Nếu căn cứ vào dữ liệu sản lượng thực tế của doanh nghiệp ở các kỳ, có thể thiết lập các chỉ số sau:

Chỉ số kế hoạch giá thành:

$$I_z = \frac{\sum z_k q_0}{\sum z_0 q_0}$$

Chỉ số thực hiện kế hoạch giá thành:

$$I_z = \frac{\sum z_1 q_0}{\sum z_k q_0}$$

Nếu căn cứ vào sản lượng kế hoạch của doanh nghiệp, có thể thiết lập các chỉ số như sau:

Chỉ số kế hoạch giá thành:

$$I_z = \frac{\sum z_k q_k}{\sum z_0 q_k}$$

Chỉ số thực hiện kế hoạch giá thành:

$$I_z = \frac{\sum z_1 q_k}{\sum z_k q_k}$$

Trong các công thức chỉ số kế hoạch trên, mỗi loại quyền số có một tác dụng nhất định, chẳng hạn, việc dùng quyền số là sản lượng thực tế kỳ nghiên cứu có thể phản ánh được đúng điều kiện sản xuất thực tế của doanh nghiệp trong kỳ nghiên cứu. Còn trường hợp sử dụng quyền số là sản lượng kế hoạch có thể cho phép phân tích tình hình thực hiện kế hoạch giá thành trong điều kiện giả định doanh nghiệp thực hiện đúng kế hoạch về sản lượng.

8.4 Hệ thống chỉ số

8.4.1 Khái niệm và cấu thành của hệ thống chỉ số

Hệ thống chỉ số là một dãy các chỉ số có liên quan đến nhau, hợp thành một phương trình cân bằng.

Hệ thống chỉ số thông thường được vận dụng để phân tích mối liên hệ giữa các chỉ tiêu trong quá trình biến động. Trong nghiên cứu kinh tế, nhiều chỉ tiêu tổng hợp có thể được cấu thành từ những nhân tố liên quan thể hiện dưới dạng các phương trình kinh tế và chính mối quan hệ đó là cơ sở để thiết lập các hệ thống chỉ số.

Chỉ số sản lượng = Chỉ số năng suất lao động × Chỉ số quy mô lao động

Chỉ số doanh thu = Chỉ số giá × Chỉ số lượng hàng hóa tiêu thụ

Như vậy, cấu thành của một hệ thống chỉ số thường bao gồm một chỉ số toàn bộ và các chỉ số nhân tố.

Chỉ số toàn bộ phản ánh sự biến động của hiện tượng phức tạp (được biểu hiện qua một chỉ tiêu nào đó) do ảnh hưởng của tất cả các nhân tố cấu thành. Theo các ví dụ trên, chỉ số sản lượng và chỉ số doanh thu là các chỉ tiêu toàn bộ.

Chỉ số nhân tố phản ánh ảnh hưởng sự biến động của từng nhân tố đối với sự biến động của hiện tượng phức tạp. Đối với hệ thống chỉ số phân tích sản lượng như ví dụ trên thì chỉ số năng suất lao động và chỉ số quy mô lao động là các chỉ số nhân tố.

8.4.2 Tác dụng của hệ thống chỉ số

Trong phân tích thống kê, hệ thống chỉ số chủ yếu vận dụng đối với các chỉ tiêu có mối quan hệ với nhau và có tác dụng sau:

Xác định vai trò và mức độ ảnh hưởng biến động của các nhân tố đối với sự biến động của hiện tượng được cấu thành từ nhiều nhân tố. Trong đó, ảnh hưởng của từng nhân tố được biểu hiện bằng số tương đối hoặc số tuyệt đối. Căn cứ vào so sánh ảnh hưởng của các nhân tố có thể đánh giá được nhân tố nào có tác dụng chủ yếu đối với biến động chung nhằm phân tích mối liên hệ giữa các hiện tượng trong quá trình biến động và giải thích được nguyên nhân cơ bản đối với sự biến động của một hiện tượng.

Dựa vào hệ thống chỉ số có thể nhanh chóng xác định được một chỉ số chưa biết khi đã biết các chỉ số khác trong hệ thống.

8.4.3 Phương pháp xây dựng hệ thống chỉ số

Phương pháp liên hoàn

Phương pháp này dựa trên cơ sở lý luận là các nhân tố cấu thành một hiện tượng phức tạp đều cùng biến động, do đó để nghiên cứu ảnh hưởng của từng nhân tố phải giả định các nhân tố lần lượt biến động. Thứ tự phân tích của các nhân tố trong hệ thống chỉ số được xác định chủ yếu thông qua việc phân biệt các nhân tố mang đặc tính chất lượng hay số lượng.

Trên cơ sở xác định các nhân tố, hệ thống chỉ số theo phương pháp liên hoàn mang những đặc điểm sau :

Một chỉ tiêu tổng hợp của hiện tượng phức tạp có bao nhiêu nhân tố thì hệ thống chỉ số có bấy nhiêu chỉ số nhân tố. Mỗi nhân tố là cơ sở để hình thành một chỉ số nhân tố.

Trong hệ thống chỉ số, chỉ số toàn bộ bằng tích của các chỉ số nhân tố và mẫu số của chỉ số đứng trước giống với tử số của nhân tố đứng sau. Do đó, sự kết hợp của các chỉ số nhân tố hình thành một dãy các chỉ số liên tục, khép kín và đảm bảo quan hệ cân bằng. Nhờ đặc điểm này mà phương pháp mang tên liên hoàn.

Chênh lệch tuyệt đối giữa tử số và mẫu số của chỉ số toàn bộ bằng tổng các chênh lệch tuyệt đối giữa tử số và mẫu số của các chỉ số nhân tố. Thực chất những kết quả tính này là để phân tích biến động tuyệt đối của chỉ tiêu nghiên cứu ra thành những phần biến động do ảnh hưởng của các nhân tố cấu thành.

Theo những đặc điểm trên, việc thiết lập một hệ thống chỉ số theo phương pháp liên hoàn được thực hiện theo các bước sau :

- Phân tích chỉ tiêu nghiên cứu ra các nhân tố cấu thành, đồng thời sắp xếp các nhân tố theo thứ tự tính chất lượng giảm dần và tính số lượng tăng dần.
- Viết chỉ số toàn bộ và chỉ số cho các nhân tố. Trong đó, đối với các chỉ số nhân tố chất lượng thông thường sử dụng quyền số là nhân tố số lượng kỳ nghiên cứu và với chỉ số nhân tố số lượng sử dụng quyền số là nhân tố chất lượng kỳ gốc.

Phương pháp biểu hiện ảnh hưởng biến động riêng biệt

Phương pháp này dựa trên quan điểm cho rằng tất cả các nhân tố cấu thành nên hiện tượng nghiên cứu có vai trò như nhau và cùng biến động. Do đó, tất cả các chỉ số nhân tố đều được xây dựng trên cùng một nguyên tắc là thời kỳ quyền số của tất cả các chỉ số nhân tố phải giống nhau và được chọn là kỳ gốc để sao cho mỗi chỉ số nhân tố biểu hiện được ảnh hưởng biến động riêng của nhân tố.

Vì tất cả các chỉ số nhân tố đều có quyền số kỳ gốc, nên tích của các chỉ số này không bằng chỉ số toàn bộ. Để đảm bảo quan hệ cân bằng của hệ thống chỉ số, theo phương pháp này người ta thêm vào hệ thống chỉ số một đại lượng bổ sung gọi là chỉ số liên hệ.

Chỉ số liên hệ biểu hiện ảnh hưởng chung của tất cả các nhân tố cùng biến động và cùng tác động lẫn nhau.

Mô hình chung thiết lập hệ thống chỉ số theo phương pháp biểu hiện biến động riêng biệt được minh họa như sau:

$$\text{Chỉ số toàn bộ} = \text{Chỉ số nhân tố 1} \times \text{Chỉ số nhân tố 2} \times \dots \times \text{Chỉ số nhân tố n} \times \text{Các chỉ số liên hệ}$$

Tùy theo số lượng các nhân tố cấu thành của hiện tượng nghiên cứu mà trong hệ thống chỉ số có thể bao gồm một hay nhiều chỉ số liên hệ. Mỗi chỉ số liên hệ phản ánh kết quả cùng biến động và cùng tác động lẫn nhau của hai hay nhiều nhân tố.

Có thể nhận thấy việc thiết lập hệ thống chỉ số theo phương pháp này là tương đối phức tạp. Hiện tượng nghiên cứu được phân tích thành càng nhiều nhân tố thì trong hệ thống càng có nhiều thành phần biểu hiện sự biến động và cùng tác động giữa các nhân tố gây khó khăn trong quá trình xây dựng, tính toán và phân tích. Như vậy, phương pháp này có thể được áp dụng một cách thuận lợi trong trường hợp hiện tượng nghiên cứu được phân tích thành hai nhân tố.

8.4.4 Hệ thống chỉ số phân tích biến động chỉ tiêu bình quân và tổng lượng biến thiên

1. Hệ thống chỉ số phân tích biến động chỉ tiêu bình quân

Chỉ tiêu bình quân chịu ảnh hưởng biến động của hai nhân tố: tiêu thức nghiên cứu và kết cấu của tổng thể.

Theo cách xác định các nhân tố cấu thành chỉ tiêu bình quân như vậy thì một hệ thống chỉ số phân tích biến động chỉ tiêu bình quân theo phương pháp liên hoàn sẽ bao gồm một chỉ số toàn bộ phản ánh biến động của chỉ tiêu bình quân và hai chỉ tiêu nhân tố phản ánh ảnh hưởng biến động của các nhân tố trên đối với chỉ tiêu bình quân. Nếu chúng ta sử dụng ký hiệu:

- x_0 và x_1 – Lượng biến của tiêu thức kỳ nghiên cứu và kỳ gốc;
- \bar{x}_0 và \bar{x}_1 – Số bình quân kỳ nghiên cứu và kỳ gốc;
- f_0 và f_1 – Số đơn vị trong tổng thể kỳ nghiên cứu và kỳ gốc.

Các chỉ số trong hệ thống chỉ số phân tích chỉ tiêu bình quân được biểu hiện như sau:

Chỉ số cấu thành khả biến, biểu hiện quan hệ so sánh giữa mức độ của chỉ tiêu bình quân ở kỳ nghiên cứu và kỳ gốc. Chỉ số này được xác định bằng công thức:

$$I_{\bar{x}} = \frac{\bar{x}_1}{\bar{x}_0} = \frac{\frac{\sum x_1 f_1}{\sum f_1}}{\frac{\sum x_0 f_0}{\sum f_0}}$$

Theo công thức trên, chỉ số này bao hàm biến động của cả hai nhân tố: tiêu thức nghiên cứu (biểu hiện bằng các lượng biến x_0 và x_1) và kết cấu của tổng thể (biểu hiện bằng các tỉ trọng $d_1^f = \frac{f_1}{\sum f_1}$ và $d_0^f = \frac{f_0}{\sum f_0}$).

Chỉ số cấu thành cố định, phản ánh biến động của chỉ tiêu bình quân do ảnh hưởng của của tiêu thức nghiên cứu, trong điều kiện kết cấu của tổng thể được coi như cố định. Trường hợp phân tích hệ thống chỉ số theo phương pháp liên hoàn thì kết cấu của tổng thể được giữ cố định ở kỳ nghiên cứu. Công thức biểu hiện như sau:

$$I_{\bar{x}} = \frac{\frac{\sum x_1 f_1}{\sum f_1}}{\frac{\sum x_0 f_1}{\sum f_1}} = \frac{\bar{x}_1}{\bar{x}_{01}}$$

Chỉ số ảnh hưởng kết cấu, phản ánh biến động của chỉ tiêu bình quân do ảnh hưởng biến động của riêng kết cấu tổng thể. Chỉ số này được thiết lập với giả định các lượng biến tiêu thức không thay đổi và được cố định ở kỳ gốc.

$$I_{\bar{x}} = \frac{\frac{\sum x_0 f_1}{\sum f_1}}{\frac{\sum x_0 f_0}{\sum f_0}} = \frac{\bar{x}_{01}}{\bar{x}_0}$$

Các chỉ số trên hợp thành hệ thống chỉ số phân tích chỉ tiêu bình quân như sau:

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

Biến động tuyệt đối của chỉ tiêu bình quân được phân tích như sau:

$$\frac{\sum x_1 f_1}{\sum f_1} - \frac{\sum x_0 f_0}{\sum f_0} = \left(\frac{\sum x_1 f_1}{\sum f_1} - \frac{\sum x_0 f_1}{\sum f_1} \right) + \left(\frac{\sum x_0 f_1}{\sum f_1} - \frac{\sum x_0 f_0}{\sum f_0} \right)$$

2. Hệ thống chỉ số phân tích biến động tổng lượng biến tiêu thức

Tổng lượng biến tiêu thức trong phân tích thống kê được biểu hiện ở nhiều chỉ tiêu như: tổng sản lượng, tổng chi phí sản xuất, tổng quỹ lương... Nhân tố ảnh hưởng đến tổng lượng biến của tiêu thức có thể được biểu hiện một cách tổng quát như sau:

$$\text{Tổng lượng biến tiêu thức (T)} = \sum x_i f_i = \bar{x} \sum f_i$$

Trong đó x_i là lượng biến tiêu thức với f_i là số đơn vị (tần số) tương ứng.

Như vậy có hai cách xác định nhân tố ảnh hưởng đến tổng lượng biến của tiêu thức. Theo cách thứ nhất, tổng lượng biến được cấu thành từ hai nhân tố là bản thân các lượng biến và số đơn vị (tần số) tương ứng $\sum x_i f_i$. Theo cách thứ hai, tổng lượng biến tiêu thức được cấu thành từ hai nhân tố là chỉ tiêu bình quân và tổng số đơn vị tổng thể ($\bar{x} - \sum f_i$).

Trong trường hợp phân tích theo cách thứ nhất thì sự kết hợp của các nhân tố cho phép thiết lập hệ thống chỉ số tổng hợp giống như phân tích chỉ tiêu doanh thu theo hai nhân tố giá cả và khối lượng hàng hóa tiêu thụ của các mặt hàng. Do vậy, ở đây chỉ đề cập cụ thể hệ thống chỉ số phân tích tổng lượng biến tiêu thức theo cách thứ hai với các nhân tố bao gồm chỉ tiêu bình quân và quy mô tổng thể.

Hệ thống chỉ số được biểu hiện theo công thức sau:

$$\frac{T_1}{T_0} = \frac{\bar{x}_1 \sum f_1}{\bar{x}_0 \sum f_0} = \frac{\bar{x}_1 \sum f_1}{\bar{x}_0 \sum f_1} \times \frac{\bar{x}_0 \sum f_1}{\bar{x}_0 \sum f_0}$$

Biến động tuyệt đối:

$$\Delta_T = T_1 - T_0 = (\bar{x}_1 - \bar{x}_0) \sum f_1 + (\sum f_1 - \sum f_0) \bar{x}_0$$

Để phân tích sâu hơn, có thể tách chỉ số chỉ tiêu bình quân thành hai chỉ số để thiết lập mô hình phân tích như sau:

$$\frac{T_1}{T_0} = \frac{\bar{x}_1 \sum f_1}{\bar{x}_0 \sum f_0} = \frac{\bar{x}_1 \sum f_1}{\bar{x}_{01} \sum f_1} \times \frac{\bar{x}_{01} \sum f_1}{\bar{x}_0 \sum f_1} \times \frac{\bar{x}_0 \sum f_1}{\bar{x}_0 \sum f_0}$$

(1)
(2)
(3)
(4)

Biến động tuyệt đối:

$$\Delta_T = T_1 - T_0 = (\bar{x}_1 - \bar{x}_{01}) \sum f_1 + (\bar{x}_{01} - \bar{x}_0) \sum f_1 + (\sum f_1 - \sum f_0) \bar{x}_0$$

Trong mô hình trên:

Chỉ số (1) phản ánh biến động của tổng lượng biến tiêu thức do ảnh hưởng của tất cả các nhân tố.

Chỉ số (2) phản ánh biến động của lượng biến tiêu thức nghiên cứu và ảnh hưởng biến động đối với tổng lượng biến tiêu thức.

Chỉ số (3) phản ánh biến động của kết cấu tổng thể ảnh hưởng đến biến động đối với tổng lượng biến tiêu thức.

Chỉ số (4) phản ánh biến động của quy mô tổng thể ảnh hưởng đến biến động của tổng biến tiêu thức.

TÀI LIỆU THAM KHẢO CHƯƠNG 8

1. TS. Nguyễn Công Nhựt - **Giáo trình Thống kê công nghiệp**. NXB Thống kê, 2013
2. Hà Văn Sơn - **Giáo trình Lý thuyết thống kê ứng dụng trong quản trị và kinh tế**. Nhà xuất bản Thống kê, 2004
3. PTS. Phan Công Nghĩa - **Giáo trình Thống kê Lao động**. NXB Thống kê, 2009

4. PGS.PTS. Nguyễn Thiệp, PTS Phan Công Nghĩa - **Giáo trình Thống kê chất lượng**. NXB Thống kê, 2011

5. PGS Nguyễn Thị Kim Thúy. **Nguyên lý Thống kê ứng dụng trong quản lý kinh tế và kinh doanh sản xuất dịch vụ**. Nhà Xuất bản thống kê - 2009

CÂU HỎI ÔN TẬP CHƯƠNG 8

1. Trình bày khái niệm chỉ số và nêu rõ những loại số tương đối nào thuộc khái niệm chỉ số?
2. So sánh đặc điểm của chỉ số tổng hợp giá Laspeyres và chỉ số tổng hợp giá Passche?
3. Khái niệm và cấu thành của hệ thống chỉ số. Cho ví dụ minh họa?
4. Nội dung thiết lập hệ thống chỉ số theo phương pháp liên hoàn?
5. Có một tài liệu thống kê 3 phân xưởng của một doanh nghiệp như sau:

Phân xưởng	Giá thành một đơn vị SP (nghìn đồng/SP)		Giá bán một đơn vị SP (nghìn đồng/SP)		Lượng sản phẩm tiêu thụ (SP)	
	Kỳ gốc	Kỳ báo cáo	Kỳ gốc	Kỳ báo cáo	Kỳ gốc	Kỳ báo cáo
1(SXSP A)	12	11	15	14	100	200
2(SXSP B)	20	21	21	23	200	150
3(SXSP C)	15	17	17	21	300	250

Sử dụng phương pháp chỉ số để phân tích ba nhân tố ảnh hưởng đến lợi nhuận kỳ báo cáo so với kỳ gốc.

CHƯƠNG 9

THỐNG KÊ XU HƯỚNG PHÁT TRIỂN VÀ DỰ BÁO THỐNG KÊ KINH TẾ

9.1 Thống kê xu hướng phát triển kinh tế

Mức độ phát triển kinh tế xã hội qua thời gian chịu sự tác động của nhiều nhân tố. Ngoài những nhân tố cơ bản quyết định xu hướng phát triển cơ bản của hiện tượng, còn có những nhân tố ngẫu nhiên, cá biệt tác động, làm cho hiện tượng kinh tế xã hội phát triển sai lệch khỏi xu hướng phát triển cơ bản. Bằng những phương pháp thống kê thích hợp giúp chúng ta loại bỏ được những ảnh hưởng tác động của các nhân tố ngẫu nhiên, cá biệt, để biểu hiện rõ xu hướng cơ bản phát triển của hiện tượng nghiên cứu. Khi nghiên cứu, phân tích xu hướng phát triển của hiện tượng kinh tế xã hội có thể sử dụng các phương pháp cơ bản sau:

9.1.1 Phương pháp mở rộng khoảng cách thời gian

Phương pháp này được áp dụng khi dãy số có khoảng cách thời gian tương đối ngắn và có quá nhiều mức độ nên không phản ánh được xu hướng phát triển của hiện tượng. Người ta có thể rút bớt số các mức độ trong dãy số bằng phương pháp mở rộng khoảng cách thời gian để có dãy số mới có các mức độ trong đó sự tác động của các nhân tố ngẫu nhiên bị loại trừ (bởi chúng tác động theo chiều hướng trái ngược nhau)

9.1.2 Phương pháp số bình quân trượt (di động)

Phương pháp này dùng để điều chỉnh các mức độ trong dãy số có biến động tăng giảm thất thường nhằm loại trừ ảnh hưởng của các nhân tố ngẫu nhiên, vạch rõ xu hướng phát triển cơ bản của hiện tượng.

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

Giả sử có dãy số thời gian:

Bảng 9.1

Thời gian	t1	t2	t3	...	tn
Mức độ y	y1	y2	y3	...	Yn

Nếu tính số trung bình trượt từ nhóm m = 3 mức độ. Ta có:

$$\overline{y_2} = \frac{y_1 + y_2 + y_3}{3}$$

$$\overline{y_3} = \frac{y_2 + y_3 + y_4}{3}$$

.....

$$\overline{y_{n-1}} = \frac{y_{n-2} + y_{n-1} + y_n}{3}$$

Từ đó có thể xây dựng dãy số thời gian gồm các số trung bình trượt:

$$\overline{y_2}, \overline{y_3}, \dots, \overline{y_{n-1}}$$

Vấn đề được đặt ra là nên tính số trung bình trượt từ một nhóm bao nhiêu mức độ?

Nếu tính số trung bình trượt từ một nhóm ít mức độ thì ảnh hưởng của các nhân tố ngẫu nhiên ít bị loại trừ. Tuy nhiên ta sẽ có nhiều số trung bình trượt và do đó sẽ dễ dàng đánh giá xu hướng biến động của hiện tượng. Ngược lại nếu số trung bình trượt được tính từ một nhóm nhiều mức độ thì khả năng hạn chế, loại bỏ ảnh hưởng ngẫu nhiên sẽ lớn. Tuy nhiên số lượng số trung bình trượt tính được sẽ ít hơn, và có thể gây khó khăn trong việc đánh giá xu hướng phát triển của hiện tượng.

Do vậy trên thực tế nghiên cứu người ta thường làm như sau:

- Đối với những hiện tượng biến động không lớn và số mức độ thực tế không nhiều lắm thì số trung bình trượt có thể tính từ một nhóm 3 mức độ. Nếu biến động của hiện tượng lớn thì nên tính số trung bình trượt từ một nhóm nhiều mức độ hơn (5, 7,... mức độ).

- Nếu hiện tượng biến động theo chu kỳ thì nên chọn thời kỳ tính số trung bình đi động bằng với độ dài thời gian(hoặc bội số) của chu kỳ. Ví dụ nếu chu kỳ biến động là 3, 5 năm thì có thể tính số trung bình đi động với 7 mức độ. Hoặc với dãy số biến động thời vụ theo tháng, hoặc quý thì có thể tính số trung bình đi động từ nhóm 12 hoặc 4 mức độ.

9.1.3 Phương pháp hồi quy

1. Bản chất của phương pháp hồi quy

Là căn cứ vào đặc điểm biến động của các mức độ trong dãy số thời gian người ta tìm một hàm số (gọi là phương trình hồi quy) nhằm phản ánh sự biến động của hiện tượng theo thời gian.

Việc lựa chọn dạng của phương trình hồi quy (đường thẳng hay đường cong) phụ thuộc vào số liệu thống kê thực tế và phân tích đặc điểm biến động của hiện tượng qua thời gian, đồng thời kết hợp với một số phương pháp đơn giản khác.

Với biến thời gian t dạng tổng quát của phương trình hồi quy được biểu diễn như sau:
 $\overline{y_t} = f(t, a_0, a_1, \dots, a_n)$.

Trong đó: $\overline{y_t}$ là các giá trị lý thuyết;

t - là thứ tự thời gian;

$a_0, a_1, a_2, \dots, a_n$ - là các tham số, được xác định bằng phương pháp tổng bình phương nhỏ nhất. Tức là:

$$S = \sum_{i=1}^n (y_i - \overline{y_{t_i}})^2 \Rightarrow \min$$

Trong đó: y_i - là mức độ thực tế thứ i ;

$\overline{y_{t_i}}$ - là mức độ lý thuyết thứ i .

2. Các dạng phương trình hồi quy

a. Phương trình đường thẳng:

$$\overline{y_t} = a_0 + a_1 t$$

Các tham số a_0, a_1 được xác định bằng phương pháp bình phương nhỏ nhất.

$$S = \sum (y - a_0 - a_1 t)^2 \rightarrow \min \rightarrow \begin{cases} \frac{\partial S}{\partial a_0} = 0 \rightarrow 2 \sum (y - a_0 - a_1 t)(-1) = 0 \\ \frac{\partial S}{\partial a_1} = 0 \rightarrow 2 \sum (y - a_0 - a_1 t)(-t) = 0 \end{cases}$$
$$\rightarrow \begin{cases} \sum y - na_0 - a_1 \sum t = 0 \\ \sum yt - a_0 \sum t - a_1 \sum t^2 = 0 \end{cases} \rightarrow \begin{cases} na_0 + a_1 \sum t = \sum y \\ a_0 \sum t + a_1 \sum t^2 = \sum yt \end{cases}$$

Đây là hệ phương trình chuẩn tắc để tìm các tham số của phương trình.

Trong trường hợp này ta thấy t là thứ tự thời gian, cho nên có thể thay thế t bằng t' với $\sum t' = 0$, sao cho vẫn đảm bảo tính thứ tự, lúc đó việc tính toán sẽ đơn giản hơn. Khi sử dụng phép biến đổi cần lưu ý:

- Nếu số lượng thời gian là một số lẻ thì lấy thời gian đứng giữa bằng 0, các thời gian đứng trước lần lượt là -1, -2, .. và thời gian đứng sau lần lượt là +1, +2,...

- Nếu số lượng thời gian là một số chẵn thì lấy 2 thời gian đứng giữa là -1 và +1, các thời gian đứng trước lần lượt là -3, -5, .. và thời gian đứng sau lần lượt là +3, +5,...

Với $\sum t' = 0$ hệ phương trình trên sẽ có dạng:

$$\begin{cases} na_0 = \sum y \\ a_1 \sum t'^2 = \sum yt' \end{cases}$$

Giải hệ phương trình này ta sẽ xác định được các tham số a_0, a_1 .

Tác dụng của phương pháp hồi quy : ngoài việc biểu diễn xu hướng của hiện tượng còn có khả năng dự báo hiện tượng trong tương lai. Muốn có dự đoán hiện tượng vào thời điểm nào ta thay t vào phương trình hồi quy.

b. Phương trình hypebol

Phương trình hồi quy có dạng: $\overline{y_t} = a_0 + \frac{a_1}{t}$

Các tham số a_0, a_1 được xác định bởi hệ phương trình:

$$\begin{cases} na_0 + a_1 \sum \frac{1}{t} = \sum y \\ a_0 \sum \frac{1}{t} + a_1 \sum \frac{1}{t^2} = \sum \frac{1}{t} y \end{cases}$$

c. Phương trình hàm mũ

Phương trình hồi quy có dạng:

$$\overline{y}_t = a_0 a_1^t$$

Logarit hoá 2 vế của phương trình ta được: $\lg y_t = \lg a_0 + t \lg a_1$

Các tham số của phương trình được xác định từ hệ phương trình chuẩn tắc:

$$\begin{cases} n \lg a_0 + \lg a_1 \sum t = \sum \lg y_t \\ \lg a_0 \sum t + \lg a_1 \sum t^2 = \sum t \lg y_t \end{cases}$$

d. Phương trình parabol bậc 2:

Phương trình hồi quy có dạng:

$$\overline{y}_t = a_0 + a_1 t + a_2 t^2$$

Các tham số a_0, a_1, a_2 được xác định bằng cách giải hệ phương trình chuẩn tắc:

$$\begin{cases} na_0 + a_1 \sum t + a_2 \sum t^2 = \sum y \\ a_0 \sum t + a_1 \sum t^2 + a_2 \sum t^3 = \sum yt \\ a_0 \sum t^2 + a_1 \sum t^3 + a_2 \sum t^4 = \sum yt^2 \end{cases}$$

9.1.4 Phương pháp biểu hiện biến động thời vụ

Sự biến động của một số hiện tượng kinh tế xã hội thường mang tính thời vụ, nghĩa là hàng năm, trong từng thời gian nhất định, sự biến động được lặp đi, lặp lại.

Biến động thời vụ thường gây ra tình trạng làm ảnh hưởng đến hoạt động sản xuất kinh doanh của ngành và các ngành có liên quan. Ví dụ như sự thay đổi nhu cầu đi lại của dân chúng theo thời gian trong năm (tăng lên nhiều vào các dịp nghỉ hè, tết, ngày lễ...) sẽ gây khó khăn cho ngành vận tải trong việc bố trí phương tiện và nhân lực để phục vụ hết nhu cầu của nhân dân. Nghiên cứu biến động thời vụ cho phép chủ động trong công tác quản lý kinh tế-xã hội, hạn chế những ảnh hưởng của biến động thời vụ đối với sản xuất kinh doanh và đời sống xã hội.

Để xác định tính chất và mức độ biến động thời vụ người ta thường dùng phương pháp tính các chỉ số thời vụ dựa trên số liệu của nhiều năm (ít nhất là 3 năm).

Đối với dãy số thời gian tương đối ổn định, tức là dãy số trong đó mức độ của hiện tượng từ năm này qua năm khác không có biểu hiện tăng giảm rõ rệt:

$$I_{tvi} = \frac{\overline{y_i}}{y_0} 100$$

Trong đó: $\overline{y_i}$ - Số trung bình của tất cả các mức độ các tháng cùng tên i;

$\overline{y_0}$ - Số trung bình của tất cả các mức độ trong dãy số.

I_{tvi} - Chỉ số thời vụ.

Trong trường hợp các mức độ của dãy số năm này qua năm khác biến động nhiều, thì chỉ số thời vụ được tính theo công thức sau:

$$\overline{I_{tvi}} = \frac{\sum \frac{y_i}{y_{lti}}}{n} 100$$

Trong đó: y_i - Mức độ thực tế thứ i ;

y_{lti} - Mức độ lý thuyết thứ i tính theo phương trình hồi qui hoặc phương pháp số trung bình di động.

9.2 Dự báo thống kê kinh tế

Dự báo thống kê là xác định các mức độ của hiện tượng kinh tế xã hội có thể xảy ra trong tương lai. Biết được tương lai của hiện tượng sẽ giúp các cấp quản lý, các nhà quản trị chủ động trong công tác lập kế hoạch sản xuất, đầu tư, quảng bá, chuẩn bị đầy đủ cơ sở vật chất cũng như có những quyết định đúng trong kinh doanh.

Xuất phát từ đối tượng và nhiệm vụ nghiên cứu, từ nguồn tài liệu thống kê thích hợp, thống kê thường thực hiện dự đoán ngắn hạn gọi là dự đoán thống kê ngắn hạn.

Dự báo thống kê ngắn hạn là công cụ quan trọng để tổ chức quản lý một cách thường xuyên các hoạt động sản xuất kinh doanh từ ngành đến các cấp cơ sở, nó cho phép phát hiện những nhân tố mới, những sự mất cân đối để từ đó có biện pháp phù hợp trong quá trình quản lý.

Có nhiều phương pháp dự báo khác nhau, phụ thuộc vào nguồn thông tin cũng như mục tiêu của dự đoán. Nhưng nội dung cơ bản của dự báo thống kê là dựa trên các giá trị đã biết y_1, y_2, \dots, y_n , phân tích các yếu tố ảnh hưởng đến sự biến động của hiện tượng, thừa nhận rằng những yếu tố đã và đang tác động sẽ vẫn còn tác động đến hiện tượng trong tương lai, xây dựng mô hình để dự đoán các giá trị tương lai chưa biết của hiện tượng.

Dự báo thống kê có những đặc điểm cơ bản sau đây:

- Dự báo thống kê chỉ thực hiện được trên từng mô hình cụ thể. Tức là nó chỉ thực hiện được sau khi đã phân tích thực trạng biến động theo thời gian hoặc không gian và phân tích đánh giá các nguyên nhân ảnh hưởng đến tiêu thức kết quả. Trong phân tích thống kê cần phân biệt rõ 2 mô hình cơ bản sau:

+ Mô hình dãy số thời gian: là tính quy luật biến động của hiện tượng qua thời gian được biểu hiện bằng hàm xu thế trên cơ sở phân tích sự biến động dãy số tiền sử trong quá khứ, hiện tại và tiến tới tương lai.

+ Mô hình nhân quả: là mối quan hệ nhân quả giữa các hiện tượng nghiên cứu qua thời gian hoặc không gian được biểu hiện bằng các hàm kinh tế, phương trình kinh tế, phương trình tương quan.

Do đó, dự báo thống kê không phải là sự phán đoán theo định tính hoặc “đoán mò” mà là sự định lượng cái sẽ xảy ra, khả năng sẽ xảy ra nhiều nhất hoặc định lượng mức độ phải xảy ra trên cơ sở khoa học của phân tích thực tiễn, cho nên kết quả dự báo thống kê vừa mang tính khách quan vừa mang tính chủ quan và nó phụ thuộc vào trình độ nhận thức khách quan, hay khả năng tư duy của người dự báo.

- Nguyên tắc cơ bản để xác định mô hình dự báo là tính kế thừa lịch sử, tính quy luật phát sinh phát triển của hiện tượng, mối quan hệ biện chứng nhân quả giữa các hiện tượng cho nên điều kiện để xác lập mô hình dự báo là:

+ Các nguyên nhân, các yếu tố, các điều kiện cơ bản ảnh hưởng đến quy luật biến động phải tương đối ổn định, bền vững trong quá khứ đến hiện tại và tiến tới tương lai.

+ Một khi có sự thay đổi các yếu tố, các nguyên nhân thì phải xác định lại mô hình để thích nghi với hiện thực.

+ Để dễ điều chỉnh mô hình và đảm bảo mức độ chính xác phù hợp với thực tiễn thì tầm xa dự báo (là khoảng cách thời gian từ hiện tại đến tương lai) không nên quá $1/3$ thời gian tiền sử.

- Tính khả thi của mức độ dự báo mang tính xác suất

- Dự báo thống kê là dự báo ngắn hạn và dự báo trung hạn vì mức độ chính xác của kết quả dự báo thống kê tỷ lệ nghịch với tầm xa dự báo.

- Dự báo thống kê mang tính nhiều phương án. Cần phải lựa chọn phương án hay mô hình để làm hàm dự báo bằng cách kiểm định mô hình

- Phương tiện để dự báo thống kê là các thuật toán, kỹ thuật tính toán phân tích, phương tiện tính toán, vì tính và trình độ nhận thức của người dự báo

Để dự báo thống kê phải thực hiện tuần tự các bước sau:

- Phân tích thực trạng biến động của hiện tượng nghiên cứu bằng nhiều phương pháp thống kê để đánh giá bản chất, mối quan hệ nội tại của đối tượng nghiên cứu.

- Xác định mô hình dự báo, tính toán các tham số để định lượng chiều hướng, đáng điệu biến động của tính quy luật.

- Kiểm định lựa chọn mô hình làm hàm dự báo

- Phân tích hậu dự báo, theo dõi các yếu tố, nguyên nhân, điều kiện đã, đang và sẽ xảy ra, tham khảo ý kiến chuyên gia để điều chỉnh lại mô hình một khi cần thiết.

9.2.1 Mô hình dự báo theo mức độ tăng (giảm) tuyệt đối bình quân

Phương pháp này được áp dụng trong trường hợp dãy số thời gian có các lượng tăng (hoặc giảm) tuyệt đối liên hoàn (từng kỳ) xấp xỉ nhau. Mức độ dự đoán ở thời kỳ thứ $n+L$ được xác định như sau:

$$\hat{y}_{n+L} = y_n + \overline{\Delta y} \cdot L$$

$$\overline{\Delta y} = \frac{y_n - y_1}{n - 1}$$

Trong đó: y_n : là mức độ cuối cùng của dãy số;

L : là tầm xa dự đoán;

$\overline{\Delta y}$: là lượng tăng (hoặc giảm) tuyệt đối trung bình.

9.2.2 Mô hình dự báo theo tốc độ phát triển bình quân

Phương pháp này được sử dụng dùng trong trường hợp hiện tượng biến động với một nhịp độ tương đối ổn định, nghĩa là có tốc độ phát triển liên hoàn xấp xỉ nhau.

Mức độ dự đoán ở thời kỳ thứ $n+L$ là:

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

Trong đó: \hat{y}_{n+L} - Mức độ dự đoán tại thời điểm $n+L$

\bar{t} - Tốc độ phát triển trung bình

$$\bar{t} = \sqrt[n-1]{\frac{y_n}{y_1}}$$

y_n - Mức độ cuối cùng của dãy số;

y_1 - Mức độ đầu tiên của dãy số.

L - Tầm xa dự đoán.

9.2.3 Mô hình dự báo theo phương trình hồi quy

Từ biến động thực tế của hiện tượng, xây dựng hàm hồi quy theo thời gian: $\overline{y_t} = f(t)$.
Căn cứ vào hàm hồi qui đã xây dựng để dự đoán mức độ tương lai của hiện tượng.

Mức độ dự đoán ở thời điểm $(n+L)$ là:

$$\hat{y}_{n+L} = f(n + L)$$

9.2.4 Mô hình dự báo theo phương pháp san bằng mũ

Ở phần trên đã đề cập đến một số phương pháp dự báo thống kê mà khi xây dựng các mô hình dự báo thì các mức độ của dãy số thời gian được xem là như nhau, nghĩa là có cùng quyền số trong khi xây dựng mô hình.

Nhưng ở những thời gian khác nhau thì hiện tượng nghiên cứu chịu sự tác động của những nhân tố khác nhau và cường độ không giống nhau. Có những nhân tố mất đi và có những nhân tố mới xuất hiện; có những nhân tố yếu đi và có những nhân tố mạnh lên.

Vì vậy, để phản ánh biến động này đòi hỏi khi xây dựng mô hình dự đoán thì các mức độ của dãy số thời gian phải được chú ý một cách khác nhau. Các mức độ càng mới (ở cuối dãy số thời gian) càng cần phải được chú ý nhiều hơn với các mức độ càng cũ (ở đầu dãy số). Như vậy, mô hình dự báo có khả năng thích nghi với sự biến động của hiện tượng qua thời gian. Một trong những phương pháp để xây dựng mô hình dự báo như vậy là phương pháp san bằng mũ.

1. Mô hình đơn giản

Mô hình đơn giản được sử dụng đối với dãy số thời gian không có xu thế và không có biến động thời vụ rõ rệt. Giả sử ở thời gian t có mức độ thực tế là y_t và mức độ dự báo là \widehat{y}_t .

Mức độ dự báo của hiện tượng có thể viết:

$$\widehat{y}_{t+1} = \alpha y_t + (1 - \alpha) \widehat{y}_t$$

Đặt $1 - \alpha = \beta$ ta có:

$$\widehat{y}_{t+1} = \alpha y_t + \beta \widehat{y}_t$$

α, β được gọi là các tham số san bằng với $\alpha + \beta = 1$ và nhận giá trị trong khoảng 0,1. Như vậy mức độ dự đoán \widehat{y}_{t+1} là trung bình cộng gia quyền của y_t và \widehat{y}_t .

Mức độ dự đoán của hiện tượng ở thời gian t là:

$$\widehat{y}_t = \alpha y_{t-1} + \beta \widehat{y}_{t-1}$$

Thay vào phương trình trên ta có:

$$\widehat{y}_{t+1} = \alpha y_t + \beta \widehat{y}_{t-1} + \beta^2 \widehat{y}_{t-1}$$

Bằng cách thay các mức độ dự báo $\widehat{y}_{t-1}, \widehat{y}_{t-2}, \dots$ vào công thức trên ta có:

$$\widehat{y}_{t+1} = \alpha \sum \beta^i y_{t-i} + \beta^{i+1} \widehat{y}_{t-i}$$

Vì $\beta < 1$ nên khi $i \rightarrow \infty$ thì $\beta^{i+1} \rightarrow 0$ và $\sum \beta^i \rightarrow 1$

Khi đó:

$$\widehat{y}_{t+1} = \alpha \sum \beta^i y_{t-i}$$

Như vậy: mức độ dự đoán \widehat{y}_{t+1} là tổng tất cả các mức độ của dãy số thời gian được tính theo quyền số mà trong đó các quyền số giảm dần theo dạng mũ tùy thuộc vào mức độ cũ của dãy số. Cũng chính vì vậy mà phương pháp này gọi là phương pháp san bằng mũ.

Công thức đầu tiên có thể viết:

$$\widehat{y}_{t+1} = \widehat{y}_t + (1 - \beta)(y_t - \widehat{y}_t)$$

Nếu đặt $e_t = y_t - \widehat{y}_t$ là sai số dự báo ở thời gian t thì:

$$\widehat{y}_{t+1} = \widehat{y}_t + \alpha e_t$$

Từ công thức trên cho ta thấy có hai vấn đề quan trọng trong phương pháp san bằng mũ:

Thứ nhất là việc chọn α , được ràng buộc bởi điều kiện $0 \leq \alpha \leq 1$ và $\alpha + \beta = 1$. Nếu α được chọn càng lớn thì các mức độ càng mới sẽ càng được chú ý, ngược lại nếu α được chọn càng nhỏ thì các mức độ cũ được chú ý một cách thỏa đáng. Do đó để lựa chọn α đòi hỏi

dựa vào phân tích đặc điểm biến động của hiện tượng qua thời gian và kinh nghiệm nghiên cứu. nói chung giá trị α tốt nhất là giá trị làm cho tổng bình phương sai số dự đoán nhỏ nhất.

Thứ hai là san mũ được thực hiện theo phép đệ quy tức là để tính $\widehat{y_{t+1}}$ thì phải có $\widehat{y_t}$, để có $\widehat{y_t}$ thì phải có $\widehat{y_{t-1}}, \dots$ do đó để tính toán cần phải xác định được giá trị ban đầu (điều kiện ban đầu) – ký hiệu là y_0 . Có nhiều phương pháp khác nhau để xác định giá trị ban đầu như có thể lấy mức độ đầu tiên của dãy số, hoặc số trung bình của một số các mức độ đầu tiên của dãy số,...

Trên đây đã trình bày nội dung của phương pháp dự báo bằng san mũ với mô hình đơn giản. Mô hình này có thể viết:

$$\widehat{y_{t+1}} = a_o(t)$$

$$\text{Với } a_o(t) = \alpha y_t + (1 - \alpha) \widehat{y_t}$$

2. Mô hình xu thế tuyến tính và không có biến động thời vụ

Trong trường hợp sự biến động của hiện tượng qua thời gian có xu thế là tuyến tính và không có biến động thời vụ để dự đoán, ta sử dụng mô hình sau:

$$\widehat{y_{t+1}} = a_o(t) + a_1(t)$$

Trong đó:

$$a_o(t) = \alpha y_t + (1 - \alpha)[a_o(t - 1) + a_1(t - 1)] = \alpha y_t + (1 - \alpha) \widehat{y_t}$$

$$a_1(t) = \gamma[a_o(t) - a_o(t - 1)] + (1 - \gamma)a_1(t - 1)$$

α, γ là các tham số san bằng và nhận giá trị trong khoảng 0,1. Giá trị α, γ được chọn tốt nhất là các giá trị làm cho tổng bình phương của sai số dự đoán là bé nhất. việc lựa chọn các giá trị ban đầu có thể được tiến hành như sau:

$a_o(0)$ có thể là mức độ đầu tiên của dãy số.

$a_o(1)$ có thể là lượng tăng (giảm) tuyệt đối trung bình.

3. Mô hình xu thế tuyến tính và biến động thời vụ

Mô hình này được sử dụng đối với dãy số thời gian mà các mức độ của nó là tài liệu tháng hoặc quý của một số năm, tức là sau một khoảng thời gian k ($k=4$ đối với quý và $k=12$ đối với năm) thì lặp lại. việc dự đoán có thể được thực hiện theo một trong hai mô hình sau đây:

+ Mô hình cộng:

$$\widehat{y_{t+1}} = [a_o(t) + a_1(t)] + S(t + 1)$$

Trong đó:

$$a_o(t) = \alpha[y_t - S(t - k)] + (1 - \alpha)[a_o(t - 1) + a_1(t - 1)]$$

$$a_1(t) = \gamma[a_o(t) - a_o(t - 1)] + (1 - \gamma)a_1(t - 1)$$

$$S(t) = \delta[y_t - a_o(t)] + (1 - \delta)S(t - k)$$

+ Mô hình nhân:

$$\widehat{y_{t+1}} = [a_o(t) + a_1(t)].S(t + 1)$$

Trong đó:

$$a_o(t) = \alpha \frac{y_t}{S(t - k)} + (1 - \alpha)[a_o(t - 1) + a_1(t - 1)]$$

$$a_1(t) = \gamma[a_o(t) - a_o(t - 1)] + (1 - \gamma)a_1(t - 1)$$

$$S(t) = \delta \frac{y_t}{a_o(t)} + (1 - \delta)S(t - k)$$

Với α, γ, δ là các tham số san bằng nhận các giá trị trong khoảng 0;1.

TÀI LIỆU THAM KHẢO CHƯƠNG 9

1. TS. Nguyễn Công Nhựt - **Giáo trình Thống kê công nghiệp**. NXB Thống kê, 2013
2. Hà Văn Sơn - **Giáo trình Lý thuyết thống kê ứng dụng trong quản trị và kinh tế**. Nhà xuất bản Thống kê, 2004
3. PTS. Phan Công Nghĩa - **Giáo trình Thống kê Lao động**. NXB Thống kê, 2009
4. PGS.PTS. Nguyễn Thiệp, PTS Phan Công Nghĩa - **Giáo trình Thống kê chất lượng**. NXB Thống kê, 2011
5. PGS Nguyễn Thị Kim Thúy. **Nguyên lý Thống kê ứng dụng trong quản lý kinh tế và kinh doanh sản xuất dịch vụ**. Nhà Xuất bản thống kê - 2009

CÂU HỎI ÔN TẬP CHƯƠNG 9

1. Dự báo thống kê là gì? Ý nghĩa thực tiễn của dự báo thống kê?
2. Tại sao nguồn tài liệu chủ yếu được sử dụng trong dự báo thống kê là dãy số thời gian?
3. Dự báo theo phương pháp san bằng mũ có ưu điểm gì?
4. Sản lượng khai thác của mỏ than A qua một số năm như sau:

Năm	2012	2013	2014	2015	2016
Sản lượng (vạn tấn)	52	56	62	67	72

Yêu cầu: dự báo sản lượng năm 2017 và 2018 bằng:

- a. Lượng tăng tuyệt đối bình quân.
- b. Tốc độ phát triển bình quân.
- c. Trong hai phương pháp trên phương pháp nào cho kết quả tốt hơn