**ĐẠI HỌC KINH TẾ TP. HỒ CHÍ MINH**
**KHOA KINH TẾ PHÁT TRIỂN**

**----------//----------**

## TÀI LIỆU GIẢNG DẠY

# PHƯƠNG PHÁP NGHIÊN CỨU KINH TẾ

**Nhóm biên soạn:**
**TS. Trần Tiến Khai**
**ThS. Trương Đăng Thụy**
**Ths. Lương Vinh Quốc Duy**
**ThS. Nguyễn Thị Song An**
**ThS. Nguyễn Hoàng Lê**

**8/2009**

# MỤC LỤC


| Mục | Trang |
|---|---|
| Chương 1. Giới thiệu phương pháp nghiên cứu | 1 |
| 1. Vai trò của nghiên cứu khoa học và phương pháp nghiên cứu | 1 |
| 2. Các loại hình nghiên cứu khoa học | 1 |
| 2.1 Nghiên cứu thực nghiệm | 1 |
| 2.2 Nghiên cứu lý thuyết | 1 |
| 3. Các phương pháp tư duy khoa học | 2 |
| 3.1 Phương pháp diễn dịch | 3 |
| 3.2 Phương pháp quy nạp | 4 |
| 4. Quy trình nghiên cứu | 4 |
| 4.1 Bước 1 - Xác định vấn đề | 5 |
| 4.2 Bước 2 - Tìm hiểu các khái niệm, lý thuyết và các nghiên cứu liên quan | 5 |
| 4.3 Bước 3 - Hình thành giả thiết | 5 |
| 4.4 Bước 4 - Xây dựng đề cương nghiên cứu | 7 |
| 4.5 Bước 5 - Thu thập dữ liệu | 8 |
| 4.6 Bước 6 - Phân tích dữ liệu | 9 |
| 4.7 Bước 7 - Giải thích kết quả và viết báo cáo cuối cùng | 9 |
| Chương 2. Mô tả vấn đề nghiên cứu | 10 |
| 1. Xác định vấn đề nghiên cứu | 10 |
| 2. Xác định câu hỏi nghiên cứu | 11 |
| 3. Tiên đề | 12 |
| 4. Giả thiết | 13 |
| 5. Nghiên cứu tìm hiểu vấn đề | 13 |
| 6. Đánh giá vấn đề nghiên cứu | 13 |
| Chương 3. Xây dựng tổng quan tài liệu và cơ sở lý thuyết | 15 |
| 1. Giới thiệu về tổng quan tài liệu và cơ sở lý thuyết | 15 |
| 1.1 Khái niệm | 15 |
| 1.2 Mục đích của Tổng quan tài liệu và cơ sở lý thuyết | 15 |
| 1.3 Một số lưu ý | 15 |

**Mục** **Trang**

2. Vai trò của tổng quan tài liệu và cơ sở lý thuyết 16

3. Thế nào là một tổng quan tài liệu và cơ sở lý thuyết tốt? 16

4. Chiến lược khai thác thông tin, dữ liệu 16

5. Các bước xây dựng tổng quan tài liệu và cơ sở lý thuyết 17

5.1 Các cấp độ của thông tin dữ liệu 18

5.2 Các dạng nguồn thông tin 18

5.3 Các bước xây dựng tổng quan tài liệu và cơ sở lý thuyết 19

6. Cách viết trích dẫn và ghi tài liệu tham khảo 21

6.1 Các hình thức trích dẫn 21

6.2 Cách ghi tài liệu tham khảo (theo ISO 690 và thông lệ quốc tế) 21

Chương 4. Thu thập dữ liệu 25

1. Nguồn dữ liệu 25

1.1 Dữ liệu thứ cấp 25

1.2 Dữ liệu sơ cấp 26

2. Phương pháp thu thập dữ liệu sơ cấp 26

2.1 Phân biệt nghiên cứu định lượng và định tính 26

2.2 Phương pháp thu thập dữ liệu định tính và định lượng 27

3. Bảng hỏi 30

3.1 Các cách khác nhau trong việc áp dụng bảng hỏi trong thu thập số liệu, thông tin 30

3.2 Các dạng câu hỏi 30

3.3 Ưu nhược điểm của câu hỏi mở 32

3.4 Ưu nhược điểm của câu hỏi đóng 32

3.5 Một số chú ý khi đặt câu hỏi 32

3.6 Bốn bước cơ bản để đặt câu hỏi đúng 33

3.7 Trật tự của các câu hỏi 35

3.8 Kiểm tra và điều chỉnh bảng câu hỏi 35

3.9 Lựa chọn giữa phỏng vấn và bảng hỏi 35

4. Tổ chức điều tra khảo sát 36

4.1 Tập huấn phỏng vấn viên 36

4.2 Tổ chức khảo sát 37

**Mục** **Trang**

4.3 Các công cụ khảo sát 37

Chương 5. Bản chất, dạng và cách đo lường dữ liệu

1. Bản chất của việc đo lường 39
2. Thang đo 40
   2.1 Thang đo danh nghĩa 41
   2.2 Thang đo thứ bậc 41
   2.3 Thang đo khoảng cách 42
   2.4 Thang đo tỷ số 42
3. Sai số trong đo lường và nguồn sai số 42
   3.1 Nguồn sai số 43
4. Các đặc điểm của một đo lường tốt 43
   4.1 Tính hợp lệ 44
   4.2 Tính tin cậy 45
   4.3 Tính thực tế 46
5. Bản chất của thái độ 47
   5.1 Quan hệ giữa thái độ và hành vi 47
   5.2 Lập thang đo thái độ 48
6. Lựa chọn một thang đo 48
   6.1 Mục tiêu nghiên cứu 48
   6.2 Các kiểu trả lời 49
   6.3 Tính chất của dữ liệu 49
   6.4 Số lượng chiều kích 49
   6.5 Cân xứng hoặc bất cân xứng 50
   6.6 Bắt buộc hay không bắt buộc 50
   6.7 Số lượng điểm đo 50
   6.8 Sai số do người đánh giá gây ra 51
7. Thang đo cho điểm 51
   7.1 Thang đo cho điểm giản đơn 51
   7.2 Thang đo Likert 54
   7.3 Thang đo trắc biệt 55
   7.4 Thang đo số / Thang đo danh sách cho điểm 55

**Mục** **Trang**

7.5 Thang đo Stapel 56

7.6 Thang đo Tổng - Hằng số 56

7.7 Thang đo cho điểm đồ thị 57

8. Thang đo xếp hạng 57

8.1 Thang đo so sánh cặp 57

8.2 Thang đo xếp hạng bắt buộc 57

8.3 Thang đo so sánh 58

Chương 6. Phương pháp chọn mẫu và xác định cở mẫu 61

1. Bản chất của việc chọn mẫu 61

1.1 Tại sao phải lấy mẫu? 61

1.2 Thế nào là một mẫu tốt? 62

1.3 Các kiểu thiết kế mẫu 63

2. Các bước thiết kế chọn mẫu 65

2.1 Dân số mục tiêu là gì? 66

2.2 Các chỉ tiêu cần quan tâm là gì? 66

2.3 Khung mẫu là gì? 67

2.4 Phương pháp chọn mẫu phù hợp là gì? 67

2.5 Cỡ mẫu cần bao nhiêu là vừa? 67

3. Chọn mẫu xác suất 68

3.1 Chọn mẫu xác suất ngẫu nhiên đơn giản 68

3.2 Chọn mẫu xác suất phức tạp 68

4. Chọn mẫu phi xác suất 76

4.1 Các vấn đề thực tiễn 76

4.2 Các phương pháp chọn mẫu phi xác suất 77

5. Xác định cỡ mẫu 78

5.1 Các khái niệm căn bản liên quan đến chọn mẫu và xác định cỡ mẫu 78

5.2 Xác định cỡ mẫu theo trung bình 80

5.3 Xác định cỡ mẫu theo tỷ lệ 82

Chương 7. Nhập và xử lý dữ liệu 86

1. Phân tích khám phá dữ liệu 86

2. Nhập số liệu 87

**Mục** **Trang**

2.1 Cách bố trí dữ liệu trên máy tính 87

2.2 Cách nhập liệu 88

3. Thanh lọc dữ liệu 89

3.1 Phát hiện giá trị dị biệt trong dữ liệu 89

3.2 Phát hiện và xử lý dữ liệu bị khuyết 95

4. Phân tích thống kê mô tả 96

4.1 Phân tích thống kê mô tả định lượng 96

4.2 Phân tích thống kê mô tả định tính 101

5. Phân tích trắc nghiệm giả thiết 102

5.1 Trắc nghiệm giả thiết 102

5.2 Quy trình trắc nghiệm thống kê 103

5.3 Phân tích dữ liệu 103

Chương 8. Viết báo cáo nghiên cứu 105

1. Giới thiệu 105

2. Xây dựng thông điệp 106

2.1 Xác định mục tiêu 106

2.2 Độc giả 106

2.3 Trình bày ý tưởng chủ đạo 107

2.4 Chỉnh sửa 107

3. Sắp xếp ý tưởng 109

4. Viết bản thảo đầu tiên 109

4.1 Lời văn 110

4.2 Các kỹ thuật giải thích 110

4.3 Tóm tắt và giới thiệu 110

4.4 Trình bày bài viết 111

4.5 Tài liệu tham khảo và các nội dung khác 112

5. Chỉnh sửa 113

5.1 Cách viết một đoạn văn hiệu quả 114

5.2 Chỉnh sửa câu văn 114

5.3 Lựa chọn từ ngữ 114

# Chương 1. Giới thiệu phương pháp nghiên cứu

**Mục tiêu giảng dạy**

Chương này nhằm mục tiêu giới thiệu các vấn đề cơ bản của môn học Phương pháp nghiên cứu và khả năng ứng dụng của các kiến thức liên quan vào việc học tập, nghiên cứu khoa học, thực hiện các khóa luận, luận văn tốt nghiệp của sinh viên ở bậc đại học cũng như cao học.

## 1. VAI TRÒ CỦA PHƯƠNG PHÁP NGHIÊN CỨU

Nhiều người cho rằng viết báo cáo/nghiên cứu là để truyền tải thông tin. Tuy nhiên một bài nghiên cứu hiệu quả phải:

- Làm thay đổi cách nhìn nhận vấn đề của người đọc
- Thuyết phục người đọc tin vào một điều gì đó
- Đưa người đọc đến quyết định và hành động
- Dẫn dắt người đọc theo một quy trình nào đó

## 2. CÁC LOẠI HÌNH NGHIÊN CỨU KHOA HỌC

Có nhiều cách phân loại. Có thể chia làm 2 loại:

- Nghiên cứu thực nghiệm: liên quan đến các hoạt động của đời sống thực tế
- Nghiên cứu lý thuyết: thông qua sách vở, tài liệu, các học thuyết và tư tưởng

Thông thường một nghiên cứu sẽ liên quan đến cả 2 khía cạnh lý thuyết và thực nghiệm.

### 2.1 Nghiên cứu thực nghiệm

Có 2 loại:

- Nghiên cứu hiện tượng thực tế (thông qua khảo sát thực tế)
- Nghiên cứu hiện tượng trong điều kiện có kiểm soát (thông qua thí nghiệm)

### 2.2 Nghiên cứu lý thuyết

Có 2 loại:

- Nghiên cứu lý thuyết thuần túy: nghiên cứu để bác bỏ, ủng hộ, hay làm rõ một quan điểm/lập luận lý thuyết nào đó.

- Nghiên cứu lý thuyết ứng dụng. Thông thường lý thuyết là cơ sở cho hành động. Nghiên cứu loại này sẽ giúp tìm hiểu các lý thuyết được áp dụng như thế nào trong thực tế, các lý thuyết có ích như thế nào...

Cách phân loại nghiên cứu khác:

- Nghiên cứu quá trình: tìm hiểu lịch sử của một sự vật hiện tượng hoặc con người
- Nghiên cứu mô tả: tìm hiểu bản chất của sự vật hiện tượng
- Nghiên cứu so sánh: tìm hiểu điểm tương đồng và khác biệt, ví dụ giữa các doanh nghiệp, thể chế, phương pháp, hành vi và thái độ...
- Nghiên cứu tìm hiểu mối quan hệ: giữa các sự vật hiện tượng. Công cụ thông thường là các phương pháp thống kê
- Nghiên cứu đánh giá: tim hiểu và đánh giá theo một hệ thống các tiêu chí
- Nghiên cứu chuẩn tắc: đánh giá/dự đoán những việc sẽ xảy ra nếu thực hiện một sự thay đổi nào đó
- Nghiên cứu mô phỏng: đây là kỹ thuật tạo ra một môi trường có kiểm soát để mô phỏng hành vi/sự vật hiện tượng trong thực tế

## 3. CÁC PHƯƠNG PHÁP TƯ DUY KHOA HỌC

Có nhiều phương pháp khoa học, trong đó, hai phương pháp (cách tiếp cận) chủ yếu là phương pháp quy nạp (inductive method) và phương pháp diễn dịch (deductive method).

• Phương pháp diễn dịch liên quan đến các bước tư duy sau:

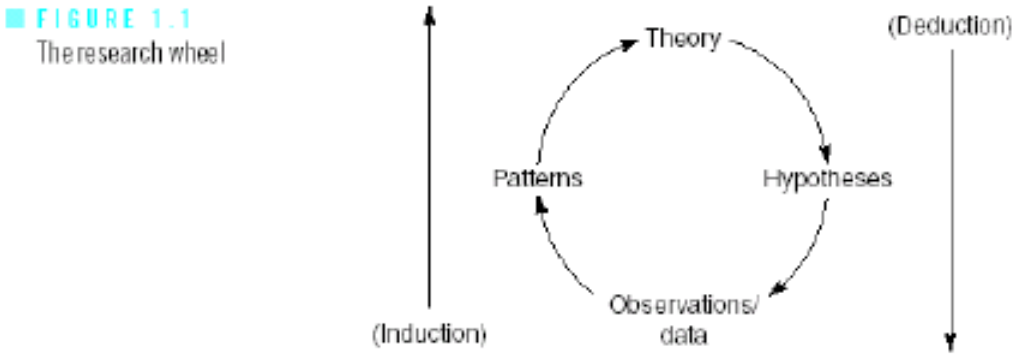
1. Phát biểu một giả thiết (dựa trên lý thuyết hay tổng quan nghiên cứu).
2. Thu thập dữ liệu để kiểm định giả thiết.
3. Ra quyết định chấp nhận hay bác bỏ giả thiết.

• Phương pháp quy nạp có ba bước tư duy:

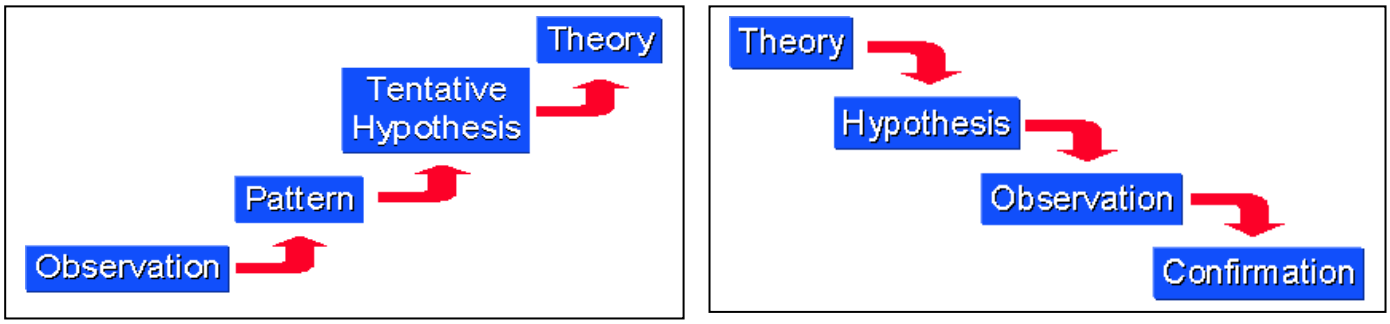
1. Quan sát thế giới thực.
2. Tìm kiếm một mẫu hình để quan sát.
3. Tổng quát hóa về những vấn đề đang xảy ra.

Trên thực tế, ứng dụng khoa học bao gồm cả hai cách tiếp cận quy nạp và diễn dịch (Hình 1.1) Phương pháp quy nạp đi theo hướng từ dưới lên (bottom up) rất phù hợp để

xây dựng các lý thuyết và giả thiết; trong khi phương pháp diễn dịch đi theo hướng từ trên xuống (top down) rất hữu ích để kiểm định các lý thuyết và giả thiết (Hình 1.2).




**Hình 1.1 Vòng tròn nghiên cứu**




**Hình 1.2 Lôgic của tư duy khoa học**

### 3.1 Phương pháp diễn dịch

Phương pháp diễn dịch là một hình thức tranh luận mà mục đích của nó là đi đến kết luận - kết luận nhất thiết phải đi theo các lý do cho trước. Các lý do này dẫn đến kết luận và thể hiện qua các minh chứng cụ thể. Để một suy luận mang tính diễn dịch là đúng, nó phải *đúng* và *hợp lệ*:

- Tiền đề (lý do) cho trước đối với một kết luận phải đúng với thế giới thực (đúng).
- Kết luận nhất thiết phải đi theo tiền đề (hợp lệ).

***Ví dụ 1***:

| | |
|---|---|
| Việc phỏng vấn các hộ gia đình trong khu phố cổ là khó khăn và tốn kém | (Tiền đề 1) |
| Cuộc điều tra này liên quan đến nhiều hộ gia đình trong khu phố cổ | (Tiền đề 2) |
| Việc phỏng vấn trong cuộc điều tra này là khó khăn và tốn kém | (Kết luận) |

### 3.2 Phương pháp quy nạp

Phương pháp quy nạp hoàn toàn khác với diễn dịch. Trong quy nạp, không có các mối quan hệ chặt chẽ giữa các lý do và kết quả. Trong quy nạp, ta rút ra một kết luận từ một hoặc hơn các chứng cứ cụ thể. Các kết luận này giải thích thực tế, và thực tế ủng hộ các kết luận này.

***Ví dụ 2***:

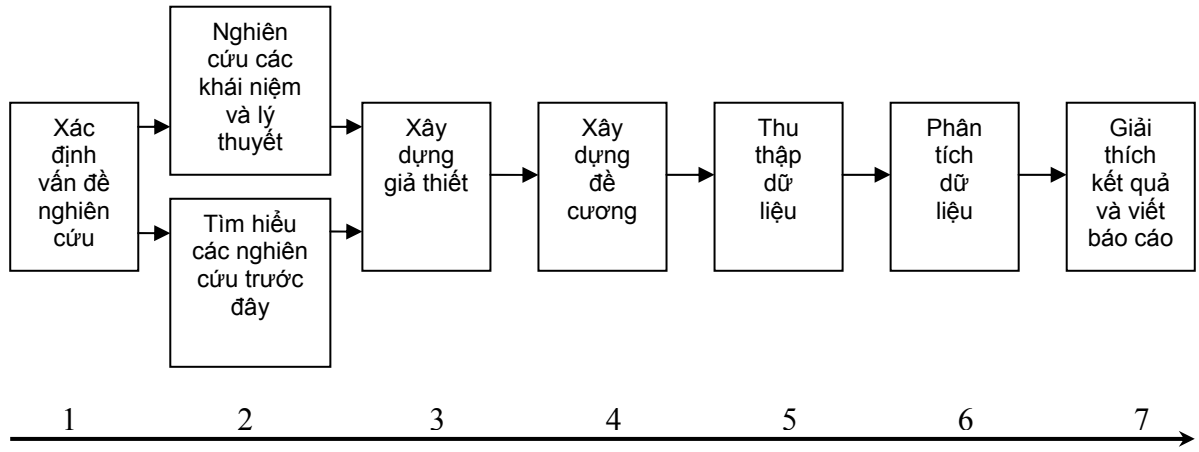
Một công ty tăng khoản tiền dành cho chiến dịch khuyến mãi nhưng doanh thu vẫn không tăng (thực tế). Tại sao doanh thu không tăng? Kết luận là chiến dịch khuyến mãi được thực hiện một cách tệ hại.

Các giải thích có thể là:

- Các nhà bán lẻ không có đủ kho trữ hàng để thỏa mãn nhu cầu của người tiêu dùng trong suốt thời gian diễn ra chiến dịch khuyến mãi.
- Một cuộc đình công của nhân viên các công ty vận tải xảy ra trong thời gian chiến dịch khuyến mãi làm cho xe tải không thể đưa hàng đến kho trữ hàng được.
- Một cơn bão cấp 8 xảy ra làm cho tát cả các cửa hàng bán lẻ phải đóng cửa trong vòng 10 ngày trùng với chiến dịch khuyến mãi.

## 4. QUY TRÌNH NGHIÊN CỨU

Quy trình nghiên cứu bao gồm một loạt các bước cần thiết để thực hiện một nghiên cứu (xem sơ đồ 1.1).




**Hình 1.3 Quy trình nghiên cứu**

Nói chung các bước trong quy trình nghiên cứu phải tuân theo một trình tự nhất định. Tuy nhiên quá trình nghiên cứu không phải đơn giản bắt đầu ở bước 1 và kết thúc ở bước 7 mà là một quá trình lặp đi lặp lại quy trình trên. Ví dụ: việc tìm hiểu khái niệm,

lý thuyết và những nghiên cứu trước đây sẽ giúp làm rõ hơn vấn đề nghiên cứu và đôi khi bắt buộc chúng ta phải xem xét lại vấn đề nghiên cứu. Ngoài ra, các bước trong quy trình có mối liên hệ chặt chẽ với nhau. Ví dụ: giả thiết và đề cương của chúng ta sẽ quyết định cần phải thu thập những dữ liệu gì, thu thập nhu thế nào và cả phương pháp phân tích dữ liệu. Sau đây là mô tả khái quát về các bước trong quy trình nghiên cứu.

### 4.1 Bước 1 - Xác định vấn đề

Có 2 loại vấn đề nghiên cứu: nghiên cứu một tình trạng thực tế nào đó hay nghiên cứu mối liên hệ giữa các biến số. Đầu tiên, người nghiên cứu phải xác định được lĩnh vực nghiên cứu mà anh ta quan tâm, từ đó thu hẹp lại thành một vấn đề nghiên cứu cụ thể. Đây là một bước hết sức quan trọng, đòi hỏi nhà nghiên cứu phải am hiểu vấn đề nghiên cứu và những khái niệm liên quan. Do vậy nhà nghiên cứu cũng phải đồng thời thực hiện bước thứ 2: tìm hiểu các khái niệm, lý thuyết và những nghiên cứu trước đây về những vấn đề tương tự để làm rõ thêm vấn đề nghiên cứu. Công việc này có thể phải lặp đi lặp lại nhiều lần, sau mỗi lần thì vấn đề nghiên cứu trở nên cụ thể hơn. Và kết thúc giai đoạn này chúng ta sẽ có được một vấn đề nghiên cứu rõ ràng, cụ thể và khả thi.

Cần lưu ý rằng việc xác định vấn đề nghiên cứu sẽ quyết định loại số liệu cần thu thập, những mối liên hệ cần phân tích, loại kỹ thuật phân tích dữ liệu thích hợp và hình thức của báo cáo cuối cùng.

### 4.2 Bước 2 - Tìm hiểu các khái niệm, lý thuyết và các nghiên cứu liên quan

Bước này đòi hỏi chúng ta phải tóm tắt lại tất cả những lý thuyết và nghiên cứu trước đây có liên quan có thể giúp chúng ta giải quyết vấn đề nghiên cứu. Cần nhớ rằng ở bước 1, chúng ta có thể phải đọc rất nhiều lý thuyết để làm rõ vấn đề nghiên cứu. Nhưng ở bước thứ 2, khi vấn đề nghiên cứu đã được xác định rõ, chúng ta chỉ sử dụng những lý thuyết thật sự liên quan và phù hợp có thể giúp giải quyết vấn đề nghiên cứu. Thư viện Khoa Kinh tế Phát triển có rất nhiều sách hay về mọi lĩnh vực của kinh tế phát triển, nhưng hầu hết là sách tiếng Anh. Do vậy, chúng ta cần chuẩn bị một vốn tiếng Anh đủ để đọc các loại sách này.

### 4.3 Bước 3 - Hình thành giả thiết

Sau khi xác định vấn đề nghiên cứu và tìm hiểu lý thuyết, chúng ta phải xây dựng một giả thiết nghiên cứu. Giả thiết nghiên cứu là một giả định của chúng ta, được xây dựng trên cơ sở của vấn đề nghiên cứu và những lý thuyết liên quan, để thông qua nghiên cứu có thể kiểm định tính hợp lý hoặc những hệ quả của nó. Đây là một bước quan trọng vì nó sẽ giúp chúng ta xác định tiêu điểm của vấn đề nghiên cứu. Nghĩa là mọi công việc trong quá trình nghiên cứu tiếp theo sẽ xoay quanh vấn đề này. Và mục đích

của cả quá trình nghiên cứu sẽ là kiểm định tính hợp lý của giả thiết. Chúng ta có thể thực hiện những công việc sau đây để xây dựng giả thiết:

- Thảo luận với bạn bè, đồng nghiệp và các chuyên gia trong lĩnh vực nghiên cứu về vấn đề nghiên cứu, nguồn gốc của nó và mục tiêu cụ thể của việc tìm ra lời giải đáp.
- Khảo sát những thông tin, dữ liệu sẵn có về vấn đề nghiên cứu.
- Khảo sát những nghiên cứu trước đây về những vấn đề liên quan, hoặc những nghiên cứu tương tự đã được thực hiện ở những địa phương/quốc gia khác.
- Thông qua quan sát và phán đoán của riêng chúng ta về vấn đề nghiên cứu, hoặc qua việc lấy ý kiến của các chuyên gia trong lĩnh vực nghiên cứu.

**(a) Tiên đề (Propositions) và Giả thiết (Hypotheses)**

Ta định nghĩa ***Tiên đề*** là một phát biểu về một hiện tượng quan sát được mà hiện tượng này có thể được phán xét là dung hay sai. Khi một định đề được viết lại nhằm mục tiêu kiểm định, ta gọi đó là ***giả thiết***.

**(b) Các loại giả thiết**

***Giả thiết mô tả (Descriptive Hypotheses)*** phát biểu về sự tồn tại, kích thước, dạng hình, hoặc phân phối của một biến nào đó. Thường các giả thiết mô tả được chuyển thành dạng câu hỏi nghiên cứu (research question). Ví dụ:

| Dạng Giả thiết mô tả | Dạng Câu hỏi nghiên cứu |
|---|---|
| Ở TP. HCM, bánh quy Kinh Đô chiếm 20% thị phần (biến). | Kinh Đô chiếm bao nhiêu thị phần bánh quy ở TP. HCM? |
| Các đô thị Việt Nam đang trải qua thời kỳ thâm hụt ngân sách (biến). | Có phải là các đô thị Việt Nam đang trải qua thời kỳ thâm hụt ngân sách hay không? |
| 80% cổ đông của Công ty A muốn Công ty tăng mức chia cổ tức (biến). | Có phải cổ đông của Công ty A muốn Công ty tăng mức chia cổ tức hay không? |

***Giả thiết quan hệ (Relational Hypotheses)*** là các phát biểu mô tả quan hệ giữa hai biến ở một số trường hợp.

*Giả thiết tương quan (Correlational hypotheses)* phát biểu rằng một số biến xuất hiệhn cùng với nhau theo một cách nào đó nhưng không có nghĩa là biến này là nguyên nhân của biến kia. Ví dụ:

- Phụ nữ trẻ (dưới 35 tuổi) mua sản phẩm của Công ty chúng ta ít hơn là phụ nữ ở độ tuổi 35.

- Số lượng bộ trang phục bán ra thay đổi theo chu kỳ kinh doanh.

*Giả thiết giải thích (nguyên nhân) (Explanatory causal hypotheses)* cho phép ám chỉ rằng sự hiện diện hoặc thay đổi của một biến gây ra hoặc dẫn đến sự thay đổi của một biến khác. Biến nguyên nhân được gọi là biến độc lập (independent variable - IV) và biến còn lại gọi là biến phụ thuộc (dependent variable - DV). Ví dụ:

- Một sự gia tăng về thu nhập của hộ gia đình (IV) dẫn đến một sự gia tăng về tỷ lệ tiền thu nhập tiết kiệm được (DV).
- Tính minh bạch của chính sách của một địa phương (IV) sẽ tạo ra niềm tin cho cộng đồng doanh nghiệp (DV) đối với địa phương đó.

**Vai trò của Giả thiết**

Trong nghiên cứu, một giả thiết đóng một số vai trò quan trọng:

- Hướng dẫn, định hướng nghiên cứu.
- Xác minh các sự kiện nào là phù hợp, và không phù hợp với nghiên cứu.
- Đề xuất các dạng nghiên cứu thích hợp nhất.
- Cung cấp khung sườn để định ra các kết luận về kết quả nghiên cứu.

***Như thế nào là một Giả thiết mạnh?*** Một giả thiết mạnh thỏa mãn đầy đủ ba điều kiện:

- Phù hợp với mục tiêu của nó
- Có thể kiểm định được
- Tốt hơn các giả thiết cạnh tranh khác

### 4.4 Bước 4 - Xây dựng đề cương nghiên cứu

Đây không đơn giản chỉ là những chương mục sẽ có trong báo cáo cuối cùng, mà là một "nghiên cứu khả thi" của dự án nghiên cứu của chúng ta. Đề cương nghiên cứu sẽ trình bày kết quả các bước chúng ta đã đạt được – bao gồm trình bày vấn đề nghiên cứu, các lý thuyết liên quan và giả thiết nghiên cứu, đồng thời trình bày kế hoạch tiếp theo để giải quyết vấn đề nghiên cứu. Một đề cương nghiên cứu thông thường bao gồm:

- Đặt vấn đề.
- Trình bày những khái niệm, lý thuyết và nghiên cứu liên quan.
- Giả thiết nghiên cứu.
- Khung phân tích: từ các khái niệm và lý thuyết liên quan, tìm ra các biến số thực tế tương ứng để kiểm định giả thiết.

- Phương pháp nghiên cứu
- Kỹ thuật thu thập và phân tích số liệu: cách thức để thu thập số liệu về các biến số đã xác định, chú ý đơn vị thu thập số liệu (cá nhân, hộ gia đình hay doanh nghiệp) và phạm vi thu thập số liệu (địa phương, tỉnh, vùng hay quốc gia). Nếu là số liệu sơ cấp thì cần thiết kế bảng câu hỏi và cách chọn mẫu. Từ tính chất của số liệu thu thập, người nghiên cứu phải xác định kỹ thuật phân tích số liệu thích hợp để kiểm định được giả thiết nghiên cứu hoặc tìm ra câu trả lời cho vấn đề nghiên cứu, và những kiểm định nhằm bảo đảm độ tin cậy của kết quả nghiên cứu.
- Trình bày cấu trúc dự kiến của báo cáo cuối cùng, bao gồm các chương mục. Đây chỉ là dự kiến và có thể thay đổi trong quá trình thực hiện.
- Lịch trình dự kiến: trình bày các bước tiếp theo cần phải thực hiện để hoàn thành nghiên cứu và thời gian cần thiết để thực hiện.
- Giới thiệu người tiến hành nghiên cứu, có thể là cá nhân hoặc nhóm, tóm tắt tiểu sử và quá trình học tập nghiên cứu. Trong luận văn tốt nghiệp ở bậc Đại học, có thể bỏ qua khâu này.
- Tài liệu tham khảo, bao gồm những tài liệu đã sử dụng để xây dựng đề cương nghiên cứu và những tài liệu đề nghị tham khảo tiếp theo cho quá trình nghiên cứu.
- Phụ lục (nếu có).

Sau khi đề cương nghiên cứu được chấp thuận (điều này có thể đòi hỏi người nghiên cứu phải sửa đi sửa lại đề cương nghiên cứu nhiều lần), bước tiếp theo là tiến hành nghiên cứu theo kế hoạch đã được vạch ra trong đề cương nghiên cứu. Bước tiếp theo sẽ là thu thập số liệu và phân tích số liệu. Lưu ý rằng trong quá trình này, người nghiên cứu vẫn phải tiếp tục tham khảo thêm các tài liệu liên quan để tiếp tục điều chỉnh các bước tiếp theo và nhằm chuẩn bị cho việc viết báo cáo cuối cùng.

### 4.5 Bước 5 - Thu thập dữ liệu

Tùy vào vấn đề nghiên cứu mà chúng ta sẽ phải thu thập loại dữ liệu thích hợp. Nói chung có 2 loại dữ liệu: thứ cấp và sơ cấp. Số liệu sơ cấp là số liệu được thu thập trực tiếp từ đối tượng nghiên cứu. Số liệu thứ cấp là số liệu tổng hợp từ số liệu sơ cấp.

Đối với số liệu thứ cấp, tùy vào lĩnh vực nghiên cứu mà chúng ta sẽ phải tìm nguồn cung cấp thích hợp. Thông thường là các niên giám thống kê, số liệu tổng hợp của các ngành và số liệu tổng hợp của các cơ quan chức năng.

Số liệu sơ cấp phải được thu thập trực tiếp từ đối tượng nghiên cứu thông qua bảng câu hỏi. Có thể thu thập bằng cách:

- Tự quan sát các hiện tượng.
- Thông qua phỏng vấn lấy ý kiến cá nhân.
- Phỏng vấn theo bảng câu hỏi. Có nhiều hình thức: phỏng vấn qua điện thoại, qua thư hoặc phỏng vấn trực tiếp. Đây là một quy trình phức tạp và tốn kém đòi hỏi phải có sự chuẩn bị cẩn thận.

Cần lưu ý rằng khi xác định vấn đề nghiên cứu, chúng ta phải cân nhắc trước về khả năng thu thập được số liệu cần thiết. Vấn đề nghiên cứu có thể rất hay và có ý nghĩa, nhưng nếu chúng ta không có khả năng thu thập được số liệu cần thiết thì nghiên cứu của chúng ta sẽ không khả thi. Và điều này là khá phổ biến trong điều kiện hiện nay.

### 4.6 Bước 6 - Phân tích dữ liệu

Tùy vào loại dữ liệu và giả thiết nghiên cứu mà chúng ta phải lựa chọn kỹ thuật phân tích dữ liệu thích hợp. Có thể là phân tích mô tả hoặc phân tích định lượng. Thông thường công việc này sẽ đòi hỏi chúng ta phải có những kỹ năng về thống kê và kinh tế lượng.

### 4.7 Bước 7 - Giải thích kết quả và viết báo cáo cuối cùng

Từ kết quả phân tích số liệu, chúng ta phải giải thích ý nghĩa của nó về mặt kinh tế. những câu hỏi cần phải trả lời là: Kết quả phân tích kết luận như thế nào về giả thiết nghiên cứu? Ý nghĩa của nó đối với vấn đề nghiên cứu? Ở đây chúng ta cần phải làm rõ ý nghĩa về mặt học thuật và ý nghĩa thực tiễn. Chúng ta phải tự hỏi: nghiên cứu của chúng ta có giá trị gì đối với những người nghiên cứu tiếp theo không? Nó có giúp những nhà hoạt động thực tiễn cải thiện được gì về vấn đề mà chúng ta nghiên cứu không?

Viết báo cáo cuối cùng là một công việc không quá khó khăn, nhưng đòi hỏi sự kiên trì. Nói chung báo cáo cuối cùng sẽ theo cấu trúc mà chúng ta đã đề nghị trong đề cương nghiên cứu. Tuy nhiên cũng không nhất thiết phải tuân thủ nghiêm ngặt. Chúng ta có thể thay đổi miễn sao nêu bật được:

- Vấn đề nghiên cứu
- Cơ sở khái niệm và lý thuyết của vấn đề
- Khung phân tích
- Phương pháp nghiên cứu
- Kết quả phân tích và giải thích kết quả phân tích số liệu
- Kết luận, đề xuất, ý nghĩa thực tiễn của nghiên cứu

# Chương 2. Mô tả vấn đề nghiên cứu

**Mục tiêu giảng dạy**

Chương này nhằm mục tiêu thảo luận về tầm quan trọng của việc xác định vấn đề nghiên cứu. Sinh viên cũng sẽ được giới thiệu về sự khác biệt và trật tự thang bậc giữa các dạng câu hỏi nghiên cứu, câu hỏi điều tra, câu hỏi đo lường và cách ứng dụng chúng vào trong một đề cương nghiên cứu.

Chương này cũng giới thiệu cách thức đánh giá thế nào là một vấn đề nghiên cứu tốt và phù hợp.

## 1. XÁC ĐỊNH VẤN ĐỀ NGHIÊN CỨU (Research Problems)

Để tiến hành một đề tài nghiên cứu, cần phải xác định một vấn đề cụ thể mà nghiên cứu của chúng ta tập trung vào. Các bước tiếp theo trong nghiên cứu sẽ thay đổi tùy thuộc vào vấn đề mà chúng ta lựa chọn.

Xác định vấn đề nghiên cứu là một công việc khó khăn. Tuy nhiên cuộc sống xung quanh chúng ta đầy rẫy các vấn đề. Có thể tìm vấn đề thông qua đọc tài liệu hoặc quan sát.

Lưu ý khi lựa chọn vấn đề nghiên cứu:

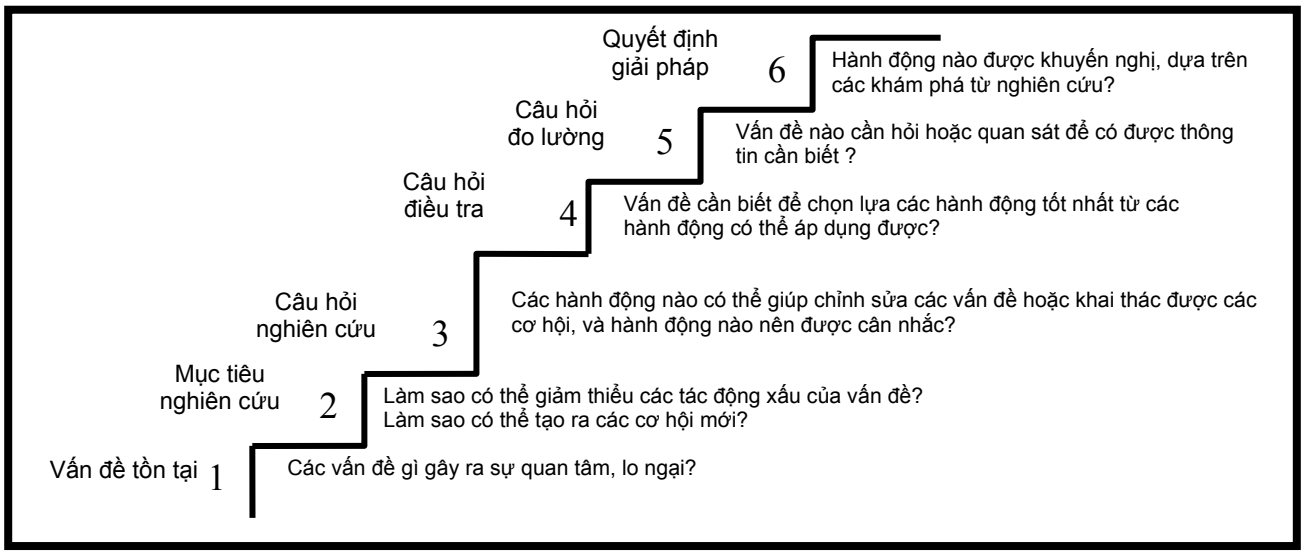
- Chúng ta cần phải thích thú với vấn đề. Chúng ta sẽ phải mất nhiều thời gian để thực hiện đề tài nên việc chọn một vấn đề chúng ta quan tâm sẽ giúp chúng ta có động cơ theo đuổi đến cùng.
- Vấn đề phải có ý nghĩa thực tiễn và phải có đóng góp. Chúng ta sẽ phí thời gian nếu thực hiện một đề tài mà người khác đã làm, hoặc sẽ chẳng có ai đọc. Ít nhất đề tài của chúng ta phải đóng góp điều gì đó về lý thuyết hoặc chính sách, hoặc đem lại những hiểu biết nhất định cho người đọc.
- Vấn đề của chúng ta phải cụ thể, không quá rộng vì chúng ta sẽ không có nhiều thời gian.
- Chúng ta cần phải bảo đảm là có thể thu thập được những thông tin/dữ liệu cần thiết để tiến hành đề tài.
- Chúng ta phải bảo đảm là có thể rút ra kết luận/bài học từ nghiên cứu của mình.
- Chúng ta phải trình bày vấn đề một cách rõ ràng, chính xác và ngắn gọn.

Khi đã chọn vấn đề nghiên cứu, chúng ta cần phải trình bày và triển khai để có thể thực hiện nghiên cứu. Sau đây là các cách thức liên quan.

## 2. XÁC ĐỊNH CÂU HỎI NGHIÊN CỨU

Nghiên cứu một vấn đề là tìm ra câu trả lời cho vấn đề đó. Do vậy đặt câu hỏi là cách tốt nhất để xác định vấn đề nghiên cứu.

Câu hỏi có thể rút ra trực tiếp từ vấn đề nghiên cứu. Có thể có nhiều câu hỏi cho một vấn đề. Có thể có câu hỏi chính và câu hỏi phụ. Từ câu hỏi nghiên cứu, ta sẽ cụ thể hóa và diễn giải chi tiết thành câu hỏi điều tra. Và để có được các thông tin, số liệu cụ thể, ta cần có các câu hỏi đo lường. Sau đây là hình minh họa sự phân chia thang bậc của vấn đề nghiên cứu và câu hỏi nghiên cứu.




**Hình 2.1 Các thang bậc vấn đề - câu hỏi nghiên cứu**

Ví dụ 2.1

| | |
|---|---|
| Vấn đề nghiên cứu | - Sự nghèo đói của đồng bào dân tộc ít người ở Tây Nguyên |
| Mục tiêu nghiên cứu | - Tìm các yếu tố ảnh hưởng đến sự nghèo đói và mức độ<br>- Đề xuất các giải pháp cải thiện tình hình |
| Câu hỏi nghiên cứu | - Đồng bào dân tộc ít người ở Tây Nguyên có nghèo đói hay không?<br>- Các yếu tố chính trị - kinh tế - xã hội, nhân chủng, dân tộc nào ảnh hưởng đến sự nghèo đói của đồng bào dân tộc ít người ở Tây Nguyên?<br>- Các yếu tố này ảnh hưởng như thế nào đến sự nghèo đói? |

Ví dụ 2.2

| | |
|---|---|
| Vấn đề nghiên cứu | - Đời sống của hộ gia đình ở các vùng quy hoạch phát triển khu công nghiệp |
| Mục tiêu nghiên cứu | - Tìm hiểu quá trình công nghiệp hóa và đô thị hóa ở vùng nghiên cứu<br>- Tình hiểu quá trình áp dụng các chính sách thu hồi đất, giải tỏa, đền bù và tái định cư ở vùng nghiên cứu.<br>- Mô tả, phân tích, so sánh tình hình đời sống của hộ gia đình trước và sau quá trình tái định cư.<br>- Đánh giá tác động của việc áp dụng các chính sách thu hồi đất, giải tỏa, đền bù và tái định cư đến đời sống hộ gia đình ở vùng nghiên cứu. |
| Câu hỏi nghiên cứu | - Các chính sách thu hồi đất, giải tỏa, đền bù và tái định cư nào được áp dụng ở vùng nghiên cứu?<br>- Đời sống của hộ gia đình trước và sau quá trình tái định cư có thay đổi hay không? Thay đổi như thế nào?<br>- Các chính sách thu hồi đất, giải tỏa, đền bù và tái định cư được áp dụng ở vùng nghiên cứu tác động như thế nào đến đời sống hộ gia đình? |

## 3. TIÊN ĐỀ

Nghiên cứu có thể tập trung vào một tiên đề thay vì giả thiết. Ví dụ:

Nghiên cứu tập trung vào vấn đề cung cấp nhà ở cho người có thu nhập thấp. Vấn đề nghiên cứu có thể được hình thành qua tiên đề:

- Chính phủ đưa ra các chương trình cung cấp nhà ở cho những người có thu nhập thấp
- Nhà ở do các chương trình này cung cấp không có những tính chất cần thiết phù hợp với người có thu nhập thấp
- Do vậy, có một sự không phù hợp giữa chương trình cung cấp nhà ở và nhu cầu nhà ở của người có thu nhập thấp

Nghiên cứu của chúng ta có thể tìm ra những điểm không phù hợp và đưa ra những kiến nghị để giải quyết vấn đề.

## 4. GIẢ THIẾT

Đây là các tiếp cận nghiên cứu theo phương pháp diễn dịch. Đầu tiên chúng ta đưa ra một giả thiết. Sau đó dùng các thông tin, dữ liệu để kiểm tra giả thiết (bác bỏ hay chấp nhận). Giả thiết có thể rút ra từ câu hỏi. Giả thiết nên:

- Là một câu khẳng định
- Phạm vi giới hạn
- Là một câu phát biểu về mối quan hệ giữa các biến số
- Có ý nghĩa rõ ràng
- Phù hợp với lý thuyết
- Được diễn tả một cách thích hợp với các thuật ngữ chính xác

## 5. NGHIÊN CỨU TÌM HIỂU VẤN ĐỀ

Thay vì nghiên cứu để trả lời câu hỏi hay kiểm định giả thiết, nghiên cứu của chúng ta có thể tìm hiểu một vấn đề nào đó. Ví dụ:

- *Mục đích của nghiên cứu này là tìm hiểu hệ thống thuế thu nhập cá nhân ở Nhật Bản và đánh giá theo hướng tìm ra những khía cạnh có thể áp dụng để cải thiện hệ thống thuế thu nhập cá nhân ở Việt Nam*
- *Mục đích của nghiên cứu này là xem xét và đánh giá hệ thống tiêu chuẩn nước thải ở Đức có thể áp dụng ở Việt Nam hay không*

## 6. ĐÁNH GIÁ VẤN ĐỀ NGHIÊN CỨU

Phần này giúp chúng ta đánh giá vấn đề nghiên cứu. Có thể xem xét có nên theo đuổi một vấn đề bằng cách trả lời những câu hỏi sau đây:

*Về tầm quan trọng của đề tài*

- Có phải là một vấn đề quan trọng không?
- Có cụ thể không?
- Có ý nghĩa về chính sách không?
- Có ý nghĩa về lý thuyết không?
- Có ý nghĩa về phương pháp không?
- Có phù hợp với chuyên ngành mà chúng ta theo học không?

*Về sở thích cá nhân*

- Chúng ta có quan tâm và hứng thú với vấn đề này không?
- Có giúp chúng ta thăng tiến trong học tập/nghề nghiệp không?
- Có thu hút sự quan tâm của người đọc không?
- Có được chấp nhận trong lĩnh vực mà chúng ta đang học tập/làm việc không?

*Về tính khả thi của đề tài*

- Có phù hợp với kiến thức của chúng ta không?
- Có phù hợp với nguồn tài liệu/dữ liệu mà chúng ta có thể có hoặc thu thập không?
- Có thể được xây dựng dựa trên lý thuyết, kiến thức và kinh nghiệm mà chúng ta có không?
- Có thể tiến hành trong điều kiện những hạn chế về thời gian, nguồn lực và tiền bạc của chúng ta không?

# Chương 3. Xây dựng tổng quan tài liệu và cơ sở lý thuyết

**Mục tiêu giảng dạy**

Sau khi học chương này, sinh viên có khả năng hiểu được:

1. Hiểu được các khái niệm về cơ sở lý thuyết và tầm quan trọng của cơ sở lý thuyết dối với nghiên cứu.
2. Các cách thức tìm kiếm tài liệu liên quan đến nội dung nghiên cứu.
3. Mục tiêu và quá trình nghiên cứu tài liệu.
4. Hai phương thức và ba mức độ của nguồn dữ liệu thứ cấp.
5. Năm kiểu thông tin bên ngoài và năm yếu tố quan trọng dùng để đánh giá giá trị của nguồn thông tin và các nội dung của nó.
6. Quá trình thực hiện tìm kiếm và nghiên cứu văn bản từ các nguồn tài liệu in và điện tử.
7. Cách thức ghi tài liệu tham khảo

## 1. GIỚI THIỆU VỀ TỔNG QUAN TÀI LIỆU VÀ CƠ SỞ LÝ THUYẾT (Literature review)

### 1.1 Khái niệm

Là một bản miêu tả chi tiết để chỉ ra rằng những lý thuyết nào sẽ được nói đến và sử dụng trong đề tài nghiên cứu của mình. Tổng quan tài liệu và cơ sở lý thuyết có vai trò rất quan trọng, cho phép ta kết luận vấn đề mình đang nghiên cứu có đáng để thực hiện và có khả năng thực hiện hay không?"

### 1.2 Mục đích của Tổng quan tài liệu và cơ sở lý thuyết

- Trình bày kiến thức và sự hiểu biết về vấn đề đang hoặc sẽ nghiên cứu.
- Đánh giá ưu - khuyết điểm của các lý thuyết sẽ áp dụng.

### 1.3 Một số lưu ý

- Tổng quan tài liệu và cơ sở lý thuyết không phải là một "bản danh sách" miêu tả những tài liệu, lý thuyết có sẵn hoặc tập hợp các kết luận.
- Tổng quan tài liệu và cơ sở lý thuyết phải là sự đánh giá có mục đích của những thông tin có tính chất tham khảo. Sự đánh giá này có thể dựa trên mục tiêu nghiên cứu hoặc những vấn đề gây tranh cãi trong đề tài nghiên cứu.

- Tổng quan tài liệu và cơ sở lý thuyết sẽ thể hiện kỹ năng của người làm nghiên cứu ở 2 lĩnh vực:
    - khả năng tìm kiếm thông tin, dữ liệu.
    - khả năng đánh giá vấn đề một cách sâu sắc và khách quan.

## 2. VAI TRÒ CỦA TỔNG QUAN TÀI LIỆU VÀ CƠ SỞ LÝ THUYẾT

- Cung cấp nền tảng lý thuyết cho việc nghiên cứu cũng như định hướng cho nghiên cứu của mình.
- Làm rõ ý nghĩa của việc liên kết những gì ta đề xuất khi nghiên cứu với những gì đã được nghiên cứu trước đó, từ đó giúp ta chọn lọc được phương pháp nghiên cứu phù hợp.
- Giúp tập trung và làm rõ ràng hơn vấn đề nghiên cứu, tránh sự tản mạn, lan man.
- Tăng cường khả năng phương pháp luận.
- Mở rộng tầm hiểu biết trong lĩnh vực ta đang nghiên cứu.
- Giảm thiểu các sai lầm, đặc biệt là những sai lầm mang tính "ngây thơ".
- Là bước quan trọng để định hướng việc tìm số liệu và thiết lập bảng câu hỏi về sau.

## 3. THẾ NÀO LÀ TỔNG QUAN TÀI LIỆU VÀ CƠ SỞ LÝ THUYẾT TỐT?

(1) Phải được sắp xếp hợp lý, bao quát từ tổng thể đến chi tiết từng câu hỏi nghiên cứu.

(2) Phải tổng hợp được các kết quả thành một kết luận, đồng thời chỉ rõ ra những ưu điểm cũng như mặt hạn chế của từng lý thuyết, nêu rõ cái gì đã biết và chưa biết.

(3) Nhận diện được những tranh luận nảy sinh giữa các lý thuyết.

(4) Thiết lập được những câu hỏi cần thiết để phục vụ cho các nghiên cứu về sau.

## 4. CHIẾN LƯỢC KHAI THÁC THÔNG TIN DỮ LIỆU

Việc khai thác các nguồn thông tin dữ liệu thứ cấp có thể được thực hiện ở tất cả các giai đoạn nghiên cứu. Nhưng hầu hết tập trung ở các giai đoạn đầu tiên của quá trình nghiên cứu, nhằm có cơ sở chuyển từ vấn đề nghiên cứu đến các câu hỏi nghiên cứu

cụ thể. Việc xây dựng cơ sở lý thuyết cho nghiên cứu cũng là một bộ phận quan trọng trong khai thác thông tin dữ liệu thứ cấp cho nghiên cứu.

Trong giai đoạn này, mục tiêu cần hoàn thành là:

- Mở rộng sự hiểu biết và nhận thức về vấn đề nghiên cứu.
- Tìm kiếm các cách thức đã được sử dụng để giải quyết vấn đề nghiên cứu hoặc câu hỏi nghiên cứu tương tự.
- Tập hợp các thông tin nền về chủ đề nghiên cứu để tinh lọc lại các câu hỏi nghiên cứu.
- Xác định các thông tin có thể được tập hợp để hình thành các câu hỏi điều tra.
- Xác định các dạng câu hỏi có thể sử dụng để thu thập dữ liệu theo các thang đo khác nhau.
- Xác định nguồn và các khung sườn có thể ứng dụng được để xác định phương thức lấy mẫu.

## 5. CÁC BƯỚC XÂY DỰNG TỔNG QUAN TÀI LIỆU VÀ CƠ SỞ LÝ THUYẾT

Trong hầu hết các trường hợp, nghiên cứu bắt đầu với việc tìm kiếm các tài liệu liên quan. Nhìn chung, việc tìm kiếm tài liệu thường được thực hiện theo 5 bước:

(1) Xác định vấn đề nghiên cứu hoặc/và câu hỏi nghiên cứu

(2) Tham khảo các bách khoa toàn thư, tự điển, sổ tay, sách và các tài liệu liên quan đến các thuật ngữ chủ yếu, con người, sự kiện liên quan đến vấn đề hoặc câu hỏi nghiên cứu.

(3) Áp dụng các thuật ngữ chủ yếu, con người, sự kiện vào việc tìm kiếm các chỉ mục (indexes), danh mục tài liệu tham khảo, và Internet để xác định các nguồn dữ liệu thứ cấp.

(4) Định vị và tổng quan các nguồn dữ liệu thứ cấp phù hợp.

(5) Đánh giá giá trị các nguồn và nội dung của dữ liệu thứ cấp.

Sau khi tổng quan tài liệu, ta có thể tìm thấy giải pháp sẵn có để trả lời cho vấn đề nghiên cứu, và khi đó, việc thực hiện nghiên cứu là không cần thiết. Tuy nhiên, có thể chưa có các giải pháp được các nghiên cứu trước chỉ ra, và ta quyết định thực hiện quá trình nghiên cứu.

### 5.1 Các cấp độ của thông tin dữ liệu

Các nguồn thông tin thường được chia theo ba cấp độ: (1) sơ cấp; (2) thứ cấp và (3) tam cấp.

**Dữ liệu sơ cấp (primary data)** là các kết quả nguyên thủy của các nghiên cứu hoặc các dữ liệu thô chưa được giải thích hoặc phát biểu đại diện cho một quan điểm hoặc vị trí chính thức nào đó.

Dữ liệu sơ cấp hầu hết có căn cứ đích xác vì các thông tin này chưa được lọc hoặc diễn giải bởi một người thứ hai. Nguồn dữ liệu sơ cấp thường là các số liệu ghi nhận trong nghiên cứu, các số liệu cá nhân, các bảng số liệu thô được mua, các bảng, biểu đồ số liệu thống kê.

**Dữ liệu thứ cấp (secondary data)** là các thông tin diễn dịch, giải thích của các dữ liệu sơ cấp. Trên thực tế, hầu hết các dữ liệu tham khảo đều thuộc nhóm này.

**Dữ liệu tam cấp (tertiary sources)** cũng có thể là các thông tin diễn dịch, giải thích của các dữ liệu thứ cấp, nhưng thông thường là các chỉ mục (indexes), danh mục tài liệu tham khảo (bibliographies), và các nguồn trợ giúp tìm kiếm thông tin khác, ví dụ các trang Web tìm kiếm thông tin Internet (Internet search engine).

### 5.2 Các dạng nguồn thông tin

Có 5 dạng nguồn thông tin quan trọng thường được các nhà nghiên cứu sử dụng.

**(1) Các Chỉ mục (Indexes) và Danh mục Tài liệu tham khảo (Bibliographies)** là nguồn tìm kiếm thông tin thư viện chủ yếu vì chúng có thể giúp ta xác định được một quyển sách hoặc một bài báo đơn lẻ có liên quan trong hàng triệu tài liệu ấn bản. Danh mục tài liệu tham khảo đơn lẻ quan trọng nhất là catalog trực tuyến (online catalog). Danh mục này rất cần thiết giúp tìm kiếm tác giả, tựa sách theo chủ đề quan tâm.

**(2) Tự điển chuyên ngành (Dictionaries)** Ta sử dụng tự điển chuyên ngành để thẩm định các thuật ngữ hoặc định nghĩa của các thuật ngữ chuyên môn. Ngoài các tự điển chuyên ngành được in ấn, hiện nay có rất nhiều các tự điển và các chú giải thuật ngữ (glossaries) trực tuyến trên Internet.

Ví dụ: *Glossaries of Financial Terms* của Federal Reserve Bank of Chicago (http://www.chicagofed.org/publications/glossary/index.cfm); Dictionary of Business and Management.

**(3) Tự điển Bách Khoa Toàn thư (Encyclopedias)**

Nên sử dụng Tự điển Bách Khoa toàn thư để tìm kiếm thông tin nền (background) hoặc thông tin lịch sử của một chủ đề nào đó, hoặc của một thuật ngữ, tên gọi nào đó. Tự điển Bách Khoa toàn thư còn cho phép ta mở rộng tìm kiếm khi chỉ ra các nguồn thông tin khác có liên quan.

**(4) Sách và Sổ tay (Handbooks)**

Sổ tay là một tập hợp của các thông tin liên quan đến một chủ đề nào đó. Các sổ tay thường bao gồm các số liệu thống kê, các thông tin niên giám điện thoại, các chú giải thuật ngữ, các dữ liệu khác ví dụ như luật lệ, quy định có liên quan để chủ đề.

**5.3 Các bước xây dựng Tổng quan tài liệu và cơ sở lý thuyết**

Bao gồm có 4 bước.

Bước 1: ***Tìm các tài liệu có liên quan đến đề tài đang và sẽ nghiên cứu.***

*Tham khảo các bách khoa toàn thư, tự điển, sổ tay, sách và các tài liệu liên quan đến các thuật ngữ chủ yếu, con người, sự kiện liên quan đến vấn đề hoặc câu hỏi nghiên cứu.*

Các nguồn để tìm:

- Internet
- Sách, báo, tạp chí
- Thư viện
- Từ điển kinh tế
- Phần "Index" của các sách và giáo trình nước ngoài
- Hỏi chuyên gia hoặc giáo viên hướng dẫn

Bước 2: ***Chọn lọc và giữ lại những tài liệu có độ tin cậy cao cũng như các lý thuyết phù hợp.***

*Áp dụng các thuật ngữ chủ yếu, con người, sự kiện vào việc tìm kiếm các chỉ mục (indexes), danh mục tài liệu tham khảo, và Internet để xác định các nguồn dữ liệu thứ cấp.*

*Đánh giá giá trị các nguồn và nội dung của dữ liệu thứ cấp.*

Việc chọn lọc sẽ căn cứ vào:

- Mức độ uy tín của tác giả, website
- Quan điểm
- Ý kiến của chuyên gia hoặc giáo viên hướng dẫn

**Các căn cứ để đánh giá giá trị của các nguồn và nội dung của dữ liệu thứ cấp.**

Có 5 yếu tố được dùng để đánh giá giá trị của các nguồn và nội dung của dữ liệu. Các yếu tố này có thể áp dụng được cho bất kỳ dạng nguồn dữ liệu nào, kể cả dữ liệu in ấn hoặc điện tử.

- Mục tiêu – Purpose (là gì?)
- Giới hạn phạm vi - Scope (như thế nào?)
- Tác giả - Authority (là ai?)
- Người đọc – Audience (là ai?)
- Định dạng - Format (như thế nào?)

**Mục tiêu**

Mục tiêu của nguồn dữ liệu là điều mà tác giả muốn hoàn thành. Một khi ta có thể xác định được mục tiêu của nguồn, ta có thể biết được nguồn thông tin này có bị thiên lệch hay không, và thiên lệch như thế nào. Chúng ta hy vọng rằng các nguồn thông tin từ các tổ chức độc lập là cân bằng hơn, thể hiện cả các thông tin có lợi (tốt) và thông tin bất lợi (xấu) về các chủ đề nghiên cứu, không thiên vị.

**Giới hạn phạm vi**

Gắn chặt với mục tiêu là giới hạn phạm vi.

- Ngày xuất bản, công bố;
- Độ sâu của chủ đề;
- Tầm bao quát của chủ đề (địa phương, quốc gia, quốc tế);
- Mức độ toàn diện;

Nếu chúng ta không biết giới hạn phạm vi của nguồn thông tin, chúng ta có thể mất thông tin vì dựa trên các nguồn không hoàn hảo.

**Tác giả**

Một trong những vấn đề quan trọng của bất kỳ người sử dụng thông tin nào là tác giả của nguồn thông tin. Tác giả và nhà xuất bản là những chỉ tiêu thể hiện cho tác giả.

**Người đọc**

Người đọc mà các tài liệu, nguồn thông tin đó hướng tới là ai. Điều này rất quan trọng và có ràng buộc chặt chẽ với mục tiêu của nguồn dữ liệu.

**Định dạng**

Yếu tố định dạng khác biệt nhau tùy theo nguồn thông tin. Vấn đề cần quan tâm là cách thức trình bày thông tin và việc tìm kiếm các mảnh thông tin đặc thù có dễ dàng hay không.

Bước 3: ***Tóm tắt và rút ra các nhân tố, các biến cần tìm và thang đo của nó.***

*(phần này dựa vào chủ đề của từng nhóm đã chọn để phân tích mẫu)*

Bước 4: ***Chắt lọc, tổng hợp và hoàn thành khung lý thuyết dựa vào tính khả thi của dữ liệu.***

- Chọn ra những lý thuyết tổng quát (key concepts).
- Tóm tắt ý chính của những lý thuyết có liên quan, trình bày ưu-nhược điểm của những lý thuyết đó.
- Trình bày kết quả nghiên cứu thực tiễn từ sách, báo, tạp chí, ... trong và ngoài nước mà ủng hộ vấn đề đang nghiên cứu để tăng sức thuyết phục cho lý thuyết mà ta đã chọn.

## 6. CÁCH VIẾT TRÍCH DẪN VÀ GHI TÀI LIỆU THAM KHẢO

**** Vai trò***: Đây là bước khá quan trọng và không thể thiếu trong đề tài nghiên cứu, nó thể hiện sự trung thực của người làm nghiên cứu, sự tôn trọng đối với các tác giả khác, cũng như giúp cho đề tài nghiên cứu mang tính thuyết phục cao hơn.

### 6.1 Các hình thức trích dẫn

- Trích dẫn nguyên văn
- Diễn đạt gián tiếp theo sự hiểu biết của mình
- Trích dẫn bảng biểu, hình vẽ minh họa

### 6.2 Cách ghi tài liệu tham khảo (theo ISO 690 và thông lệ quốc tế)

**(1) Đối với sách**

Các thành phần bắt buộc phải ghi đối với một tài liệu là sách:

- Tên tác giả
  + nước ngoài: họ, chữ viết tắt của tên và chữ lót
  + trong nước: tên, chữ viết tắt của họ và chữ lót
- Năm xuất bản (nằm trong ngoặc đơn)
- Tên sách (in nghiêng hoặc gạch dưới)
- Tên nhà xuất bản
- Tên địa danh nơi nhà xuất bản tọa lạc
- Lần tái bản (nếu có)

Ví dụ:

***Theo ISO 690***

Gall, J-C. Paléoécologie. Paysages et environnements disparus. 2e éd. Paris: Masson, 1998. 239p. ISBN 2-225-83084-3

***Theo cách tổng quan khoa học***

Gall, J-C. (1998). Paléoécologie. Paysages et environnements disparus. 2e éd. Paris: Masson

***Theo các cách phổ biến khác***

Aigner, D. J: *Basic Econometrics*, Prentice Hall, Englewood Cliffs, N.J., 1971.

American National Standards Institute, Inc. 1969. American national standard for the abbreviation of titles of periodicals. ANSI Z39.5-1969. American National Standards Institute, Inc., New York.

**(2) Đối với bài báo đăng trong tạp chí khoa học**

Các thành phần bắt buộc phải ghi đối với một tài liệu là bài báo đăng trong tạp chí khoa học:

- Tên tác giả
- Năm xuất bản (nằm trong ngoặc đơn)
- Tên bài viết (nằm trong dấu nháy đơn)
- Tên tờ báo hoặc tạp chí (in nghiêng)
- Số, kỳ
- Ngày tháng xuất bản và số trang

Ví dụ:

***Theo ISO 690***

Deleu, M *et al*. Apercu des techniques d'analyse conformationelle des macromolecules biologiques. Biotechnologie, Agronomie, Societé et Environnement, 1998, vol 2, no 4, p.234-247

***Theo cách tổng quan khoa học***

Deleu M., Watheler B., Brasseur R., Paquot M. (1998). Apercu des techniques d'analyse conformationelle des macromolecules biologiques. *Biotechnol. Agron. Soc. Environ.* **2(4)**, 234-247

***Theo các cách phổ biến khác***

McGirr, C. J. 1973. Guidelines for abstracting. Tech. Commun. **25**(2):2-5.

Rosner, J. L. 1990. Reflections on science as a product. Nature **345**:108.

Kaplinsky, R. (1999). "Globalisation and Unequalization: What Can Be Learned from Value Chain Analysis." Journal of Development Studies 37(2): 117-146.

**(3) Đối với bài đăng trong tuyển tập bài viết khoa học**

***Theo ISO 690***

Troxler, W.L. Thermal desorption. In Kearney, P. and Roberts, T. (eds), Pesticide remediation in soils and water. Chichester, UK: Wiley, 1998, p.105-128

***Theo cách tổng quan khoa học***

Troxler, W.L. (1998). Thermal desorption. In Kearney, P. and Roberts, T., eds. *Pesticide remediation in soils and water*. Chichester, UK: Wiley, p.105-128

***Theo các cách phổ biến khác***

Hugon, P., 1985. "Le miroir sans tain. Dépendance alimentaire et urbanisation en Afrique: un essai d'analyse mésodynamique en termes de filières", in Altersial, CERED & M.S.A. (eds.), Nourrir les villes, L'Harmattan, pp. 9 46.

Suhariyanto, K., Lusigi, A., Thirtle, C., 2001. Productivity growth and convergence in Asian and African agriculture. In: Lawrence, P., Thirtle, C. (Eds.), Africa and Asia in Comparative Economic Perspective. Palgrave, New York.

**(4) Đối với nguồn từ Internet**

- Tên tác giả
- Năm xuất bản (nằm trong ngoặc đơn)
- Tên bài viết (nằm trong dấu nháy đơn)
- Tên trang web chính (in nghiêng)
- Địa chỉ chi tiết của bài viết (nếu có)
- Ngày tháng năm truy cập

Ví dụ:

***Theo ISO 690-2***

Ashby J.A *et al*. Investing in Farmers as Researchers. Ciat publication $n^0$ 318 [online]. Cali, Colombia: CIAT, 2000 [ref. on Jan 20$^{th}$ 2002). Available on World Wide Web: <http://www.ciat.cigiar.org/downloads/pdf/Investing_farmers.pdf>

***Theo cách tổng quan khoa học***

Ashby J.A., Braun A.R., Gracia T., Del Pilar Guerrero L., Hernandez L.A., Quiros C.A., Roa J.I. (2000). *Investing in Farmers as Researchers. Ciat publication $n^0$ 318*

[online]. Cali, Colombia: CIAT, 2000. Available on World Wide Web: <http://www.ciat.cigiar.org/downloads/pdf/Investing_farmers.pdf>, Consulted Jan 20th 2002

# Chương 4. Thu thập dữ liệu

**Mục đích giảng dạy**

Mục đích của chương này là hướng dẫn sinh viên các phương pháp thu thập dữ liệu định tính và định lượng. cụ thể là giúp sinh viên hiểu được có mấy loại dữ liệu, tìm ở đâu ra dữ liệu đúng, làm thế nào để thu thập được dữ liệu phù hợp cho các loại nghiên cứu và các vấn đề nghiên cứu khác nhau.

## 1. NGUỒN DỮ LIỆU

Có hai loại nguồn dữ liệu cơ bản là (1) dữ liệu thứ cấp và (2) dữ liệu sơ cấp.

### 1.1 Dữ liệu thứ cấp

*Dữ liệu thứ cấp: là dữ liệu do người khác thu thập*, sử dụng cho các mục đích có thể là khác với mục đích nghiên cứu của chúng ta. Dữ liệu thứ cấp có thể là dữ liệu chưa xử lý (còn gọi là dữ liệu thô)  hoặc dữ liệu đã xử lý. Như vậy dữ liệu thứ cấp không phải do người nghiên cứu trực tiếp thu thập.

Có nhiều nhà nghiên cứu, sinh viên đánh giá thấp nguồn dữ liệu thứ cấp có sẵn. vì vậy chúng ta bắt đầu xem xét sự hợp lý của nguồn dữ liệu thứ cấp đối với vấn đề nghiên cứu của chúng ta trước khi tiến hành thu thập dữ liệu của chính mình. Các cuộc tổng điều tra về dân số, nhà ở, điều tra doanh nghiệp, điều tra mức sống dân cư, điều tra kinh tế xã hội hộ gia đình (đa mục tiêu)….do chính phủ yêu cầu là những nguồn dữ liệu rất quan trọng cho các nghiên cứu kinh tế xã hội.

Ngoài ra, một số nguồn dữ liệu dưới đây có thể là quan trọng cho các nghiên cứu của chúng ta bao gồm:

- các báo cáo của chính phủ, bộ ngành, số liệu của các cơ quan thống kê về tình hình kinh tế xã hội, ngân sách quốc gia, xuất nhập khẩu, đầu tư nước ngoài, dữ liệu của các công ty về báo cáo kết quả tình hình hoạt động kinh doanh, nghiên cứu thị trường….
- các báo cáo nghiên cứu của cơ quan, viện, trường đại học;
- các bài viết đăng trên báo hoặc các tạp chí khoa học chuyên ngành và tạp chí mang tính hàn lâm có liên quan;
- tài liệu giáo trình hoặc các xuất bản khoa học liên quan đến vấn đề nghiên cứu;
- cuối cùng nhưng không kém phần quan trọng là các bài báo cáo hay luận văn của các sinh viên khác (khóa trước) trong trường hoặc ở các trường khác.

Ưu điểm của việc sử dụng dữ liệu thứ cấp là tiết kiệm tiền, thời gian.

Nhược điểm trong sử dụng nguồn dữ liệu thứ cấp là:

(1) số liệu thứ cấp này đã được thu thập cho các nghiên với các mục đích khác và có thể hoàn toàn không hợp với vấn đề của chúng ta. khó phân loại dữ liệu, các biến số và đơn vị đo lường có thể khác nhau …

(2) dữ liệu thứ cấp thường đã qua xử lý nên khó đánh giá được mức độ chính xác, mức độ tin cậy của nguồn dữ liệu.

Vì vậy, trách nhiệm của người nghiên cứu là phải đảm bảo tính chính xác của dữ liệu, phải kiểm tra xem các kết quả nghiên cứu của người khác là dựa vào dữ liệu thứ cấp hay sơ cấp. vì vậy điều quan trọng là phải kiểm tra dữ liệu gốc.

### 1.2 Dữ liệu sơ cấp

Khi dữ liệu thứ cấp không có sẵn hoặc không thể giúp trả lời các câu hỏi nghiên cứu của chúng ta, chúng ta cần phải tự mình thu thập dữ liệu cho phù hợp với vấn đề nghiên cứu đặt ra. Các dữ liệu tự thu thập này được gọi là dữ liệu sơ cấp. Hay nói cách khác *dữ liệu sơ cấp là dữ liệu do chính người nghiên cứu thu thập*.

## 2. PHƯƠNG PHÁP THU THẬP DỮ LIỆU SƠ CẤP

### 2.1 Phân biệt nghiên cứu định lượng và định tính

**Bảng 4.1 Phân biệt nghiên cứu định lượng và định tính**

| Stt | Tính chất | Định lượng | Định tính |
|---|---|---|---|
| 1 | Mục đích | Mô tả sự kiện bằng những con số | Xác định ý nghĩ, quan điểm, cảm xúc, xu hướng bằng lời |
| 2 | Trình bày | Quan điểm, ngôn ngữ của nhà nghiên cứu | Quan điểm, ngôn ngữ của người được nghiên cứu |
| 3 | Chọn mẫu | Ngẫu nhiên hoặc ngẫu nhiên có phân tầng | Có mục đích |
| 4 | Câu hỏi | Đóng, trắc nghiệm, câu trả lời định sẵn. | Mở, câu trả lời tự do không định sẵn |
| 5 | Phỏng vấn | Cấu trúc. Bảng hỏi được soạn sẵn theo một cấu trúc cố định, không được thay đổi | Bán cấu trúc. Bảng hỏi chỉ mang tính chất gợi ý. Các câu hỏi được phát triển từ trả lời của người được phỏng vấn |

## 2.2 Phương pháp thu thập dữ liệu định tính và định lượng

Có 3 phương pháp phổ biến: (1) quan sát; (2) phỏng vấn và (3) điều tra qua bảng hỏi.

### (a) Phương pháp quan sát

(1) *Quan sát có tham dự* (nhập vai) ví dụ đóng vai là hành khách đi xe buýt công cộng để tìm hiểu chất lượng phục vụ hoặc đánh giá mức độ hài lòng của người sử dụng.

(2) *Quan sát không có tham dự* (không nhập vai). Ví dụ quan sát và đếm các loại phương tiện qua cầu, qua chốt giao thông; quan sát người công nhân trong dây chuyền sản xuất để làm định mức lao động; quan sát địa bàn nơi sẽ tiến hành khảo sát(nhà cửa, đường sá, các csvckt, chợ búa trường học, cách đi lại giao tiếp của người dân trong cộng đồng)...

*Những trở ngại khi sử dụng phương pháp quan sát là:*

(1) Đối tượng thay đổi hành vi khi cảm thấy bị quan sát theo hướng tích cực hoặc tiêu cực.

(2) Thiên lệch chủ quan của người quan sát.

(3) Diễn giải khác nhau cho cùng một quan sát giữa những người quan sát khác nhau.

(4) Quan sát phiến diện hoặc ghi chép thiếu. Quan sát kỹ, ghi chép thiếu hoặc quan sát thiếu nhưng ghi chép kỹ chú tâm quan sát quên ghi chép và ngược lại.

**Lưu ý:**

*Phương pháp quan sát thường được vận dụng trong nghiên cứu marketing (quan sát hành vi người tiêu dùng), hoặc quan sát bấm giờ trong nghiên cứu tổ chức sản xuất, tổ chức lao động, tính định mức lao động hoặc một số nghiên cứu tổ chức hệ thống giao thông vận tải, đếm lượng xe lưu thông qua cầu, phà…*

### (b) Phương pháp phỏng vấn

Phỏng vấn là phương pháp thu thập thông tin dữ liệu rất thông dụng. Trong cuộc sống đời thường chúng ta thu thập thông tin thông qua các dạng khác nhau của việc giao tiếp với người khác. Bất kỳ giao tiếp nào giữa 2 hay nhiều người với mục đích định trước gọi là *phỏng vấn.* Một mặt, phỏng vấn có thể rất linh hoạt, uyển chuyển khi phỏng vấn viên tự do đặt câu hỏi xung quanh vấn đề cần khảo sát, mặt khác, phỏng vấn có thể không linh hoạt khi phỏng vấn viên bám sát theo các câu hỏi đã được chuẩn bị sẵn. Do đó phỏng vấn được phân loại tùy vào mức độ linh hoạt như trình bày trong sơ đồ dưới đây:

Các dạng phỏng vấn: (1) cấu trúc; (2) không cấu trúc và (3) bán cấu trúc.

Trong phỏng vấn cấu trúc trình tự (trật tự, cấu trúc) phỏng vấn, nội dung phỏng vấn và câu hỏi phỏng vấn cùng với những câu từ trong đó đều được định sẵn. Ngược lại, trong phỏng vấn không cấu trúc thì trật tự phỏng vấn, nội dung phỏng vấn cũng như các câu hỏi phỏng vấn đều linh hoạt, có thể thay đổi tùy hoàn cảnh, tình huống cụ thể. Phỏng vấn bán cấu trúc là sự kết hợp của hai loại phỏng vấn trên.

**i. Phỏng vấn không cấu trúc**

Một số dạng phỏng vấn không cấu trúc thường được sử dụng là: phỏng vấn sâu, phỏng vấn chuyên gia, thảo luận nhóm mục tiêu, tường thuật và truyền miệng.

Phổ biến nhất hiện nay là phương pháp kết hợp **khảo sát định lượng** (phỏng vấn cấu trúc thường bằng các câu hỏi đóng, questionnaire) với **khảo sát định tính** bằng phương pháp đánh giá nhanh có sự tham dự PRA (participatory rapid appraisal). Phương pháp này áp dụng phỏng vấn bán cấu trúc bằng câu hỏi mở, thảo luận nhóm mục tiêu, phỏng vấn chuyên gia, phỏng vấn cá nhân với sự tham dự của các đối tượng nghiên cứu và phỏng vấn sâu.

***Phỏng vấn sâu*** là phỏng vấn trực tiếp trên cơ sở hiểu biết tin tưởng lẫn nhau (thường là phỏng vấn lặp lại). Người được phỏng vấn trình bày những nhận thức, hiểu biết, kinh nghiệm và hòan cảnh sống của họ bằng ngôn ngữ của chính họ. Phỏng vấn sâu thường sử dụng trong các nghiên cứu tình huống, nghiên cứu điển hình.

***Phỏng vấn nhóm mục tiêu*** (thảo luận nhóm mục tiêu): tương tự như phỏng vấn sâu nhưng người phỏng vấn trao đổi với 1 nhóm, còn người được phỏng vấn làm việc với cá nhân. Chủ đề phỏng vấn được phát triển rộng bởi người phỏng vấn hoặc nhóm. Những vấn đề chính sẽ được phát hiện quan thảo luận nhóm và các thành viên, chia xẻ nhận thức, quan điểm của họ về cùng nhhững vấn đề quan tâm. Người phỏng vấn cần ghi chép lại một cách trung thực ý kiến của nhóm. Tốt nhất nhờ thư ký hoặc ghi âm, ghi hình vì người phỏng vấn cần tập trung làm tốt vai trò người hướng dẫn thảo luận, sau đó cần kiểm tra lại những thông tin đã ghi chép. Chú ý khi bắt đầu vào thảo luận cần phải có thời gian để các thành viên trong nhóm tự giới thiệu về mình. Thư ký nên đánh số thứ tự cho các thành viên và khi họ phát biểu chỉ cần ghi lại số thứ tự đó (vừa ghi chép nhanh vừa đảm bảo khách quan hoặc bảo mật thông tin cá nhân cho người tham gia thảo luận).

***Phỏng vấn chuyên gia và những người chủ chốt*** **(key persons):** tương tự như phỏng vấn sâu nhưng đối tượng phỏng vấn là những chuyên gia trong lĩnh vực nghiên cứu và những người am hiểu cộng đồng, am hiểu địa bàn nơi tiến hành khảo sát (kỹ sư, giám đốc các cơ quan, cán bộ địa phương các cấp, lão nông tri điền, ...).

***Tường thuật*****:** nghe người trong cuộc tường thuật lại những gì đã xảy ra trong cuộc sống của họ, nhà nghiên cứu chỉ lắng nghe, thỉnh thỏang sử dụng các kỹ thuật để

khuyến khích người nói hứng khởi hơn; ví dụ dùng những tiếng đệm như: "à há"; "ừm ừm"; "yah"; "đúng rồi", v.v, vào những thời điểm thích hợp. Cơ bản là để cho người nói nói một cách tự nhiên, không được cắt ngang câu chuyện làm họ mất hứng. Tường thuật là một phương pháp thu thập thông tin dữ liệu rất hữu hiệu đặc biệt với những vấn đề nhạy cảm, ví dụ nghiên cứu vấn đề xâm hại tình dục trẻ em đối với những người đã trải qua chuyện đó. Là nhà nghiên cứu chúng ta đề nghị họ tường thuật lại kinh nghiệm đã qua và họ bị tác động như thế nào. Giống như phỏng vấn nhóm, cần chọn cách ghi chép thật thích hợp. Sau khi nghe tường thuật câu chuyện xong chúng ta phải ghi chép lại một cách tỉ mỉ, trung thực và phải đưa lại cho người tường thuật xem để kiểm tra tính chính xác của thông tin.

***Truyền miệng*:** giống như phương pháp tường thuật, phương pháp truyền miệng sử dụng cả 2 cách lắng nghe thụ động và chủ động. Phương pháp này thường áp dụng để nắm bắt những sự kiện lịch sử đã xảy ra trong quá khứ hay để hiểu biết văn hóa, phong tục tập quán hoặc những câu chuyện đã xáy ra trong quá khứ từ thế hệ này qau thế hệ khác. Nếu như tường thuật là kể lại câu chuyện của bản thân người đó thì truyền miệng là kể lại sự kiện lịch sử, xã hội hoặc văn hóa.

Thu thập thông tin dữ liệu bằng phỏng vấn không cấu trúc cực kỳ hữu ích trong trường hợp cần những thông tin sâu hoặc chưa hiểu biết nhiều về vùng nghiên cứu, lĩnh vực nghiên cứu. Sự linh hoạt giúp người phỏng vấn khai thác được nhiều thông tin phong phú trước khi tiến hành phỏng vấn cấu trúc. Tuy nhiên phỏng vấn không cấu trúc hạn chế khả năng so sánh và dễ bị thiên lệch trong quá trình thu thập thông tin. Do đó cần thiết phải có hướng dẫn phỏng vấn như là một phương tiện để thu thập dữ liệu. Phương pháp này cũng đòi hỏi người phỏng vấn phải có kỹ năng rất cao, cao hơn so với sử dụng phương pháp phỏng vấn cấu trúc.

**ii. Phỏng vấn cấu trúc**

Trong phỏng vấn cấu trúc, nhà nghiên cứu hỏi một loạt các câu hỏi xác định trước theo một trật tự nhất định trong bảng câu hỏi. Bảng câu hỏi gồm các câu hỏi đóng hoặc mở được chuẩn bị sẵn cho phỏng vấn viên. Thường là dùng hình thức trắc nghiệm, đưa ra các phương án trả lời khác nhau để người được phỏng vấn lựa chọn. Tuy nhiên thông thường bao giờ cũng có câu trả lời khác. Bảng câu hỏi là phương tiện còn việc phỏng vấn là phương pháp thu thập thông tin dữ liệu. Ưu điểm chính của phỏng vấn cấu trúc là cung cấp thông tin có khả năng so sánh. Phỏng vấn cấu trúc không đòi hỏi kỹ năng phỏng vấn cao như trong phỏng vấn bán cấu trúc.

## 3. BẢNG HỎI (Questionnaire)

Bảng hỏi là bảng liệt kê các câu hỏi mà người được phỏng vấn tự trả lời bằng cách tự viết vào. Khác nhau giữa phỏng vấn và bảng hỏi là người phỏng vấn có thể hỏi, giải thích nếu cần và ghi lại câu trả lời còn bảng câu hỏi là do chính người trả lời ghi vào. Bảng hỏi cần phải có câu hỏi rõ ràng, dễ đọc, dễ theo dõi, dùng ngôn ngữ phổ biến như văn nói giao tiếp thông thường mà người được phỏng vấn cảm thấy quen thuộc. Những câu hỏi nhạy cảm thường kèm theo sự giải thích rõ ràng. Tốt nhất là dùng font chữ khác để phân biệt với câu hỏi khác.

### 3.1 Các cách khác nhau trong việc áp dụng bảng hỏi trong thu thập số liệu, thông tin

(1) Gởi bưu điện là phổ biến nhất. Tuy nhiên cần phải có địa chỉ của người được phỏng vấn. Cần gởi kèm theo bì thư ghi địa chỉ phản hồi và dán sẵn tem để họ gởi lại sau khi điền câu trả lời. Cần có **thư ngỏ** đính kèm với bảng hỏi.

(2) Thu thập tại nơi hội họp, học tập hoặc nơi công cộng như sinh viên, học viên các chương trình, trung tâm mua sắm, y tế, bệnh viện, trường học, quán ăn, câu lạc bộ giải trí....

***Nội dung của thư ngỏ:***

Thư ngỏ đặc biệt quan trọng khi sử dụng bảng hỏi. Nội dung chủ yếu của thư ngỏ gồm:

(1) giới thiệu cơ quan tổ chức mà chúng ta đại diện;

(2) mô tả mục tiêu chính của nghiên cứu (2-3 câu);

(3) giải thích tầm quan trọng của nghiên cứu;

(4) những hướng dẫn chung;

(5) xác nhận rằng việc tham gia trả lời bảng hỏi là tự nguyện nếu người được hỏi không muốn trả lời họ có quyền trả lời;

(6) bảo đảm nguồn thông tin là do chính họ cung cấp;

(7) cung cấp cho họ số điện thoại, địa chỉ liên lạc trong trường hợp họ cần trao đổi thắc mắc hay hỏi lại những điều chưa rõ;

(8) địa chỉ gởi lại bảng trả lời và thời gian;

(9) cảm ơn vì sự hợp tác.

### 3.2 Các dạng câu hỏi

**Có 2 dạng câu hỏi chính**: ***câu hỏi đóng*** và ***câu hỏi mở***. Trong câu hỏi mở câu trả lời không được đưa ra trước để lựa chọn mà đối tượng phải tự trả lời theo cách của họ.

Trong câu hỏi đóng thường có sẵn các phương án trả lời cho lựa chọn và thường có câu trả lời khác kèm theo đề nghị giải thích. Ví dụ:

| **Câu hỏi đóng** | **Câu hỏi mở** |
|---|---|
| 1. Vui lòng đánh dấu vào ô thích hợp tuổi của chúng ta<br>Dưới 15 □ từ 15-19 □ từ 20-24 □ | 1. Hiện nay chúng ta bao nhiêu tuổi? |
| 2. Tình trạng hôn nhân hiện tại của chúng ta<br>Đã lập gia đình □ độc thân □<br>đã ly thân □ đã li dị □<br>Mới đính hôn □ | 2. Tình trạng hôn nhân hiện nay của chúng ta là gì? |
| 3. Thu nhập trung bình hàng năm của chúng ta là bao nhiêu?<br>Dưới 10 tr □ từ 10 - dưới 20 tr □<br>Từ 20–dưới 30tr □ từ 30–dưới 40tr □ từ 40 tr trở lên □<br>Hoặc tự xếp loại thu nhập hàng năm của chúng ta<br>Trên tb □ trung bình □ dưới tb □ | 3. Thu nhập trung bình hàng năm của chúng ta là bao nhiêu? |
| 4. Theo chúng ta một nhà quản trị giỏi có những đặc tính nào?<br>Có khả năng ra quyết định □<br>Ra quyết định nhanh □<br>Biết lắng nghe □<br>Kỹ năng giao tiếp tốt □<br>Công bằng, không thiên vị □<br>Đặc tính khác (ghi rõ)..... | 4. Theo chúng ta tiêu chí nào đánh giá một nhà quản trị giỏi? |

### 3.3 Ưu nhược điểm của câu hỏi mở

- Cung cấp thông tin sâu (nhất là người phỏng vấn có kinh nghiệm), phong phú nhưng xử lý thông tin và phân tích dữ liệu khó hơn. Nhà nghiên cứu cần phân tích nội dung thông tin để làm rõ dữ liệu.
- Tạo cho người trả lời sự tự do diễn đạt ý tưởng của họ chứ không trả lời theo một khuôn mẫu định sẵn. Tuy nhiên một số người không có khả năng trả lời một số câu hỏi sẽ dẫn đến kết quả thiếu thông tin.
- Tránh được thiên lệch từ phía người  người trả lời nhưng có thể bị thiên lệch từ người hỏi.

### 3.4 Ưu nhược điểm của câu hỏi đóng

- Thiếu thông tin sâu và ít có sự khác biệt.
- Thiên lệch do các câu trả lời định sẵn (thiên lệch từ ý tưởng của người đặt câu hỏi).
- Do câu trả lời định sẵn nên có thể không phản ánh đúng ý kiến của người được hỏi, trả lời thiếu động não.
- Ưu điểm lớn nhất là thông tin dữ liệu thu thập được dễ dàng phân tích và xử lý.

### 3.5 Một số chú ý khi đặt câu hỏi

- Câu hỏi phải đơn giản, ngắn gọn, dễ hiểu, sử dụng ngôn từ bình thường hàng ngày.
- Cần xem xét trình độ, kiến thức của đối tượng được hỏi liệu họ có trả lời được câu hỏi đặt ra không
- Cần phải chắc chắn rằng bất cứ người nào cũng hiểu được câu hỏi với cùng một kiểu tức là mỗi người đều hiểu ý nghĩa như nhau cho cùng một câu hỏi. Ví dụ: “chúng ta thấy căn-tin trường mình được không?” ; “công việc của chúng ta có gặp trở ngại vì chúng ta có con nhỏ không?”  Có? Không?. Có 2 trường hợp trả lời không: chúng ta không có con nhỏ hoặc chúng ta có con nhỏ nhưng không ảnh hưởng tới công việc.
- Mỗi câu hỏi chỉ liên quan đến 1 khía cạnh, không đặt câu hỏi ghép. Đừng hỏi những câu có 2 ý cùng một lúc. “chúng ta có thường đến thư viện không và mỗi lần đến khoảng bao lâu? Để làm gi?”
- Các câu hỏi phải được hình thành theo phương cách để tránh cho người trả lời mà không có lối thoát như trả lời “không biết” hay “không bình luận”…
- Các câu hỏi được hình thành cần sử dụng ngôn ngữ lịch sự và mềm dẻo như: xin ông/bà vui lòng cho biết…

- Đừng hỏi những câu có định hướng trả lời: "ở TP.HCM, tình trạng thất nghiệp đang tăng lên đúng không?" "chúng ta có nghĩ hút thuốc là có hại cho sức khỏe không?"
- Đừng hỏi những câu dựa trên giả định. "Một ngày chúng ta hút bao nhiêu điếu thuốc?"
- Các câu hỏi phải được sắp xếp từ câu hỏi tổng quan đến cụ thể.
- Cần ý kiến đóng góp của các chuyên gia, đồng nghiệp cho bảng hỏi.
- Tổ chức điều tra thử để xem xét, chỉnh sửa câu hỏi, bảng hỏi trước khi hoàn tất bảng hỏi và tiến hành điều tra chính thức.

### 3.6 Bốn bước cơ bản để đặt câu hỏi đúng

Nguyên tắc chung là "câu hỏi của chúng ta đặt ra phải gắn với mục tiêu nghiên cứu của chúng ta". Do đó xác định rõ mục tiêu đóng vai trò cực kỳ quan trọng. Mỗi câu hỏi đặt ra đều xuất phát từ mục tiêu nghiên cứu, câu hỏi nghiên cứu hoặc những giả thiết của nghiên cứu.

- ***Bước 1:*** xác định thật rõ mục tiêu, liệt kê ra tất cả mục tiêu cụ thể, các câu hỏi nghiên cứu hoặc các giả thiết đã được kiểm chứng (nếu có).

**Ví dụ 4.1 Nhận dạng chủ đề nghiên cứu vấn đề nghiện rượu trong xã hội**

Liệt kê các khía cạnh của chủ đề nghiên cứu

- Nguyên nhân gây nghiện rượu.
- Tác động của nghiện rượu đối với gia đình.
- Thái độ của cộng đồng đối với vấn đề nghiện rượu.
- ảnh hưởng của nghiện rượu đến năng suất lao động.

Lựa chọn

Tác động của nghiện rượu đối với gia đình

Xác định mục tiêu chung: tìm ra những tác động của nghiện rượu đối với gia đình.

Mục tiêu cụ thể:

1. Tìm ra những tác động của nghiện rượu đến quan hệ hôn nhân.
2. Xác định được những ảnh hưởng của nghiện rượu đến đời sống của con cái trong gia đình.

Tìm ra những ảnh hưởng của nghiện rượu đến tình hình tài chính của gia đình.

- ***bước 2:*** với mỗi mục tiêu/câu hỏi nghiên cứu, liệt kê tất cả những câu hỏi có liên quan mà chúng ta muốn trả lời thông qua nghiên cứu của chúng ta

Đặt câu hỏi

1. Nghiện rượu có tác động gì đến quan hệ hôn nhân ?
2. Nghiện rượu ảnh hưởng như thế nào đến đời sống của con cái trong gia đình ?
3. Nghiện rượu ảnh hưởng như thế nào đến tình hình tài chính của gia đình ?

- ***bước 3:*** với mỗi câu hỏi liệt kê ở bước 2, liệt kê các yêu cầu thông tin, chỉ số đo lường, đánh giá để trả lời câu hỏi đó

Thông tin yêu cầu/chỉ số đo lường:

1. **Các thông tin liên quan đến tình hình và kết quả học tập của con cái (điểm số trung bình, số ngày nghỉ học, bỏ học, thời gian tự học ở nhà....)**
2. Những thay đổi trong thái độ, quan hệ trong gia đình, quan hệ chúng ta bè (gọi điện thoại, vui buồn...)
3. Những thay đổi trong nếp sống, giờ giấc sinh hoạt (thời gian ba mẹ dành cho con, đưa đón con...)

- ***bước 4:*** thiết lập câu hỏi để đạt được thông tin yêu cầu

Điểm trung bình năm hoc.....?

Kết quả học tập năm.....đạt loại (giỏi, khá, trung bình, yếu)?

Số buổi nghỉ học có lý do?

Số buổi nghỉ học không lý do?

Số lần đi học trễ?

Thời gian tự học ở nhà trung bình mỗi ngày là bao nhiêu giờ?

So với năm trước tăng hay giảm? Lý do tăng, lý do giảm?

- **chú ý:** trong trường hợp một thông tin yêu cầu hoặc một chỉ số có thể có nhiều cách đặt câu hỏi khác nhau nhưng chỉ nên chọn 1.

### 3.7 Trật tự của các câu hỏi

Trật tự của các câu hỏi trong bảng questionnaire hay trong phỏng vấn là rất quan trọng vì nó ảnh hưởng đến mức độ tham gia cũng như thái độ của người trả lời và do đó ảnh hưởng đến chất lượng thông tin, dữ liệu thu thập được. Có hai quan điểm trong việc sắp xếp trật tự các câu hỏi (1) theo trật tự ngẫu nhiên (2) sắp xếp có hệ thống dựa trên mục tiêu nghiên cứu. Quan điểm thứ 2 cho rằng việc sắp xếp có hệ thống các câu hỏi sẽ dần dần giúp người trả lời đi vào vấn đề cần nghiên cứu, bắt đầu từ các câu hỏi dễ, đơn giản đến các câu hỏi khó và phức tạp hơn. Bằng cách hỏi như vậy sẽ tạo hứng thú cho người trả lời và họ không cảm thấy quá khó hoặc quá phức tạp. Tuy nhiên quan điểm sắp xếp câu hỏi ngẫu nhiên rất phù hợp trong các nghiên cứu mà người nghiên cứu muốn người trả lời trình bày sự đồng thuận hoặc không đồng thuận của họ với những khía cạnh khác nhau của vấn đề nghiên cứu. Trong trường hợp đó việc sắp xếp câu hỏi có hệ thống có thể tạo điều kiện cho người trả lời xuôi theo ý tưởng chủ quan của người đặt câu hỏi.

### 3.8 Kiểm tra và điều chỉnh bảng câu hỏi

Đây là một khâu rất quan trọng trong thu thập dữ liệu. Sau khi soạn bảng câu hỏi hòan chỉnh cần thiết phải điều tra thử để kiểm tra lại tính hợp lý của câu hỏi, sử dụng ngôn từ có đơn giản dể hiểu không? Người trả lời có hiểu sai câu hỏi không, có trả lời được không? Độ dài của bảng câu hỏi đã phù hợp chưa? Sắp xếp các phần nội dung có hợp lý không?

### 3.9 Lựa chọn giữa phỏng vấn và bảng hỏi

Lựa chọn giữa phỏng vấn và bảng hỏi là rất quan trọng. Cần cân nhắc kỹ ưu điểm của 2 phương pháp có thể ảnh hưởng đến tính chính xác, độ tin cậy của các kết quả.

| | **Phỏng vấn** | **Bảng hỏi** |
|---|---|---|
| Ưu điểm | - thích hợp cho các nghiên cứu tình huống phức tạp<br>- hữu ích trong thu thập thông tin sâu, chi tiết<br>- có nhiều thông tin bổ trợ thông qua quan sát<br>- có thể giải thích câu hỏi<br>- áp dụng rộng rãi phổ biến cho mọi đối tượng. | - ít tốn kém<br>- thông tin chính xác, người trả lời không e ngại. |

| | Phỏng vấn | Bảng hỏi |
|---|---|---|
| Nhược điểm | - tốn kém (thời gian và chi phí)<br>- chất lượng dữ liệu phụ thuộc vào quá trình phỏng vấn, quan hệ giữa người hỏi và người trả lời, kỹ năng của người hỏi<br>- chất lượng dữ liệu có thể khác nhau khi có nhiều người cùng thực hiện phỏng vấn<br>- có thể thiên lệch. | - hạn chế áp dụng<br>- tỉ lệ thu hồi thấp<br>- có thiên lệch, có thể trả lời theo tư vấn của người khác<br>- thiếu cơ hội làm rõ vấn đề, có thể không hiểu đúng câu hỏi<br>- thiếu các thông tin bổ trợ (quan sát). |

**Có 3 tiêu chí để lựa chọn**

***(1) Bản chất của điều tra.***

Nếu vấn đề nghiên cứu làm cho đối tượng cảm thấy lưỡng lự, do dự khi trả lời trực tiếp với phỏng vấn viên thì sử dụng bảng hỏi là thích hợp hơn. Ví dụ nghiên cứu vấn đề nghiện ngập, uống rượu, vấn đề tình dục, hành động tội phạm hoặc vấn đề tài chính cá nhân. Tuy nhiên trường hợp sử dụng phỏng vấn để điều tra các vấn đề nhạy cảm lại thu thập được thông tin tốt hơn. Điều đó tùy thuộc vào cộng đồng và kỹ năng của người phỏng vấn.

***(2) Độ phân tán của đối tượng được nghiên cứu.***

Nếu đối tượng sống quqá phân tán, rải rác thì không có cách lựa chọn nào khác hơn là dùng bảng hỏi vì phỏng vấn là vô cùng tốn kém.

***(3) Loại đối tượng nghiên cứu.***

Nếu là mù chữ hoặc còn quá nhỏ, quá già, bị tàn tật, người dân tộc thiểu số...thì bắt buộc phải phỏng vấn.

## 4. TỔ CHỨC ĐIỀU TRA KHẢO SÁT

### 4.1 Tập huấn phỏng vấn viên

Trong hầu hết các nghiên cứu có điều tra thu thập dữ liệu sơ cấp, bản thân nhà nghiên cứu không thể một mình thực hiện hết tất cả việc phỏng vấn mà bắt buộc phải nhờ vào đội ngũ phỏng vấn viên (có thể là sinh viên, cán bộ địa phương, các cộng tác viên...).

Do đó việc tập huấn phỏng vấn viên là rất cần thiết để đảm bảo tính chính xác, trung thực của thông tin dữ liệu.

### 4.2 Tổ chức khảo sát

Đòi hỏi phải có sự chuẩn bị kỹ với các kịch bản, kế hoạch rõ ràng. Thời gian, địa điểm, quan hệ với địa phương, thông báo đến tận các đối tượng điều tra. Đặc biệt cần làm tốt công tác tiền trạm, chọn địa bàn khảo sát. Cần phải phối hợp chặt chẽ giữa nhóm nghiên cứu và địa phương nơi tiến hành khảo sát. Chuẩn bị tài chính, hậu cần, văn phòng phẩm, ăn ở, phương tiện đi lại...tổ chức phối hợp các nhóm điều tra, kiểm tra lại thông tin dữ liệu thu thập được sau mỗi ngày điều tra để kịp thời bổ sung chỉnh sửa rút kinh nghiệm.

### 4.3 Các công cụ khảo sát

Có rất nhiều công cụ khảo sát. Tùy điều kiện và mục tiêu điều tra cũng như những yêu cầu thông tin mà chọn các công cụ cho thích hợp. Thông thường có các công cụ sau (1) vẽ sơ đồ/bản đồ theo không gian, theo mặt cắt (2) lịch thời vụ/ quá trình diễn biến sự việc theo thời gian (3) xếp hạng (4).

## TÓM TẮT CHƯƠNG

Chương này giúp chúng ta hiểu được các phương pháp thu thập dữ liệu khác nhau. Nguồn thông tin dữ liệu về tình huống, hiện tượng, vấn đề hay nhóm người có thể phân thành 2 loại chính (1) nguồn sơ cấp và (2) nguồn thứ cấp.

Phỏng vấn, quan sát và sử dụng bảng hỏi là 3 phương pháp chủ yếu dùng để thu thập dữ liệu sơ cấp. Tất cả những thông tin, dữ liệu có sẵn như tài liệu của chính phủ, các báo cáo, nghiên cứu trước…đều là nguồn thứ cấp.

Việc chọn phương pháp cụ thể nào để thu thập dữ liệu phụ thuộc vào mục tiêu nghiên cứu, loại thông tin dữ liệu cần thu thập, nguồn thông tin dữ liệu sẵn có, kỹ năng của chúng ta trong việc sử dụng phương pháp thu thập dữ liệu cụ thể cũng như đặc trưng kinh tế, xã hội, nhân chủng học của đối tượng nghiên cứu. Mỗi phương pháp đều có những ưu nhược điểm của nó và mỗi phương pháp chỉ thích hợp cho những tình huống nhất định. Việc chọn phương pháp thu thập dữ liệu cụ thể nào là rất quan trọng đảm bảo chất lượng và độ tin cậy của thông tin, dữ liệu. Không một phương pháp thu thập dữ liệu nào được cho là chính xác 100%. Chất lượng thông tin, dữ liệu chúng ta thu thập được phụ thuộc vào phương pháp luận, những yếu tố liên quan đến đối tượng nghiên cứu (người trả lời câu hỏi, người cung cấp thông tin), đến hoàn cảnh và khả năng của người nghiên cứu trong việc kiểm soát hay giảm thiểu được những yếu tố ảnh hưởng trong quá trình thu thập dữ liệu.

Việc dùng câu hỏi đóng hay câu hỏi mở tùy theo từng tình huống, hoàn cảnh. Cả hai loại câu hỏi đó đều có những ưu nhược điểm của nó, chúng ta cần cân nhắc xem xét việc áp dụng chúng cho những tình huống thích hợp. Các câu hỏi trong bảng hỏi hay trong phỏng vấn thường có nhiều vấn đề cần phải lưu ý vì chúng trực tiếp liên quan đến mục tiêu nghiên cứu của chúng ta và trực tiếp ảnh hưởng đến chất lượng cũng như kết quả nghiên cứu.

# Chương 5. Bản chất, dạng và cách đo lường dữ liệu

**Mục tiêu giảng dạy**

Sau khi học chương này, sinh viên có thể hiểu được:

1. Sự khác biệt giữa việc đo lường đối tượng nghiên cứu, các đặc điểm và các chỉ số đại diện cho các đặc điểm của đối tượng nghiên cứu.
2. Sự tương đồng và khác biệt giữa 4 dạng thang đo và cách thức đo lường tương ứng.
3. Các tiêu chí đánh giá thang đo tốt.
4. Các yếu tố ảnh hưởng đến việc chọn lựa một thang đo phù hợp.
5. Các đặc tính và cách sử dụng các thang đo cho điểm, thang đo xếp hạng, thang đo xếp trật tự (sorting) và các thang đo thường sử dụng khác.

Chương này cũng trình bày các quy trình giúp chúng ta hiểu được các thang đo để áp dụng vào việc chọn lựa hoặc thiết kế các cách thức đo lường nhằm thu thập dữ liệu một cách tốt nhất cho nghiên cứu kinh tế.

## 1. BẢN CHẤT CỦA VIỆC ĐO LƯỜNG

Việc đo lường gắn kết với nghiên cứu có nghĩa là gán các con số cho các sự kiện thực nghiệm, các đối tượng nghiên cứu hoặc các tính chất, hoặc các hành động theo các nguyên tắc nhất định. Định nghĩa này hàm ý rằng việc đo lường là một quá trình 3 bước:

1. Chọn lựa các sự kiện thực nghiệm[1] có thể quan sát được.
2. Phát triển các nguyên tắc để gán các con số hoặc biểu tượng để thể hiện các khía cạnh khác nhau của sự kiện được đo lường.
3. Áp dụng các nguyên tắc trên cho các quan sát tương ứng với từng sự kiện.

Mục tiêu của đo lường là cung cấp các dữ liệu, thông tin có chất lượng tốt nhất, ít sai sót nhất để trắc nghiệm giả thiết, để phỏng định, tiên lượng hoặc mô tả.

**Chúng ta có thể đo lường cái gì?**

Thông thường, trong nghiên cứu chúng ta sẽ đo lường các biến số (variables) và có thể phân loại chúng thành đối tượng (objects) hoặc là tính chất (properties).

[1] Nghiên cứu thực nghiệm (empirical study) là cách thức các nhà nghiên cứu áp dụng để mô tả, giải thích và tiên lượng dựa trên các thông tin có được từ các quan sát.

**Đối tượng (Objects)** nghiên cứu là một khái niệm rộng, ám chỉ tới chủ thể má chúng ta đang tiến hành nghiên cứu. Trong khoa học kinh tế, đối tượng nghiên cứu có thể là cá nhân, hộ gia đình, nhóm người, các chủ thể kinh tế, khu vực kinh tế, v.v. Thông thường, ta không trực tiếp đo lường được đối tượng nghiên cứu mà chúng ta diễn giải đối tượng nghiên cứu thông qua các tính chất, đặc điểm của đối tượng nghiên cứu.

**Tính chất (Properties)** là các đặc tính của đối tượng, ví dụ như:

- Các tính chất thực thể (Physical properties): chiều cao, cân nặng, tuổi tác, v.v.
- Các tính chất tâm lý (Psychological properties): thái độ, sự thông minh, tình cảm, v.v.
- Các tính chất xã hội (Social properties): khả năng lãnh đạo, quan hệ cộng đồng, v.v.

Mặc dù vậy, trên thực tế, chúng ta thường không đo lường các đối tượng nghiên cứu cũng như các tính chất, mà chúng ta đo lường các chỉ số đại diện (indicants) cho đối tượng hoặc tính chất. Ngoài ra, chúng ta cũng rất khó đo lường các tính chất tâm lý hoặc xã hội, trong khi lại có thể đo lường được các tính chất thực thể.

## 2. THANG ĐO (MEASUREMENT SCALES)

Khi đo lường, chúng ta đưa ra các nguyên tắc đo, và sau đó, diễn giải các quan sát của chúng ta về các chỉ số đại diện cho các tính chất của đối tượng nghiên cứu theo các nguyên tắc đo này. Có nhiều thang đo có thể áp dụng được, và chúng ta sẽ phải chọn các thang đo phù hợp nhất tuỳ theo chúng ta xây dựng các nguyên tắc như thế nào.

Các nguyên tắc đo có 4 đặc tính:

1. *Phân loại (Classification)*. Sử dụng các con số để chia nhóm hoặc sắp xếp các câu trả lời. Không sắp xếp theo trật tự thứ bậc.
2. *Thứ bậc (Order)*. Các con số được sắp xếp theo trật tự thứ bậc. Một con số phải lớn hơn, nhỏ hơn hoặc ngang bằng với một con số khác.
3. *Khoảng cách (Distance)*. Sự chênh lệch, sai biệt giữa các con số được xếp theo thứ bậc. Sai biệt giữa bất kỳ một cặp số liệu nào đều phải lớn hơn, nhỏ hơn hoặc ngang bằng với một con số khác với sự sai biệt của một cặp số bất kỳ khác.
4. *Số gốc (Origin)*. Các dãy số liệu có một số gốc duy nhất là số không.

Sự kết hợp của các đặc tính phân loại, thứ bậc, khoảng cách và gốc sẽ cho chúng ta 4 thang đo được sử dụng phổ biến là (1) thang đo danh nghĩa (nominal scale); (2) thang

đo thứ bậc (ordinal scale); (3) thang đo khoảng cách (interval scale) và (4) thang đo tỷ số (ratio scale).

### 2.1 Thang đo danh nghĩa (Nominal Scales)

Thang đo này đôi khi còn được gọi là thang đo định danh. Trong nghiên cứu kinh tế và quản trị, thang đo danh nghĩa được áp dụng rất phổ biến. Với thang đo này, chúng ta thu thập thông tin của một biến nào đó mà theo một cách thức tự nhiên hoặc được thiết kế, biến đó được chia thành hai loại hoặc nhiều hơn.

Với thang đo này, chúng ta chỉ có thể có được thông tin duy nhất là số đếm hoặc tỷ lệ các thành viên có trong mỗi nhóm. Nếu chúng ta sử dụng các biểu tượng dạng số theo nguyên tắc mà chúng ta đặt ra để xác định và biểu thị các nhóm, thì các con số này chỉ được hiểu là các nhãn (labels) một cách thuần túy, và không hề có giá trị định lượng nào hết.

Bởi vì ta chỉ có thể lượng hóa con số đếm của các thành viên trường hợp có trong mỗi nhóm (phân phối tần suất), ta không thể sử dụng chỉ số *mode và giá trị trung bình* như là một chỉ tiêu đo lường xu hướng trung tâm. Tương tự như vậy, ta cũng không thể đo lường độ phân tán của dữ liệu khi áp dụng thang đo danh nghĩa.

Mặc dù thang đo danh nghĩa không mạnh về khả năng thống kê, nhưng chúng vẫn rất hữu ích. Nếu không thể áp dụng thang đo nào khác, chúng ta vẫn có thể phân loại một bộ các tính chất thành một bộ các nhóm tương đương. Thang đo danh nghĩa rất có giá trị cho các nghiên cứu khai phá khi mà mục tiêu nghiên cứu là tìm hiểu các quan hệ hơn là bảo đảm các con số đo lường chính xác. Thang đo danh nghĩa củng được dùng phổ biến trong điều tra và các loại hình nghiên cứu khác khi cần phải phân loại dữ liệu theo các nhóm phụ của dân số điều tra. Thông thường, trong nghiên cứu kinh tế hoặc nghiên cứu khoa học xã hội, thang đo danh nghĩa được áp dụng để phân loại các đặc điểm cá nhân của một nhóm người nào đó, ví dụ như giới tính, tình trạng gia đình, học vấn, thái độ chính trị, v.v.

### 2.2 Thang đo thứ bậc (Ordinal Scales)

Thang đo thứ bậc, đầu tiên có các tính chất của thang đo danh nghĩa, cộng với tính chất chỉ thị thứ bậc. Ý nghĩa của thang đo thứ bậc là có thể chỉ ra khái niệm "lớn hơn", "nhỏ hơn" mà không cần phải nói chính xác lớn hơn bao nhiêu, nhỏ hơn bao nhiêu. Thang đo thứ bậc cũng có thể biểu thị được các trạng thái "cao hơn", "tốt hơn", "tệ hơn", "kém hơn", "quan trọng hơn" hoặc "kém quan trọng hơn".

Thang đo thứ bậc cũng được ứng dụng khi chúng ta cần đo lường nhiều hơn một tính chất mà chúng ta quan tâm. Chúng ta có thể xếp hạng tổng hợp bằng cách hoặc là xếp hạng dựa trên sự tổng hợp các tính chất của một đối tượng nào đó, hoặc bằng cách xây

dựng một bảng xếp hạng tổng hợp dựa trên các xếp hạng riêng lẻ dựa trên từng tính chất đơn lẻ.

Bởi vì các con số sử dụng trong thang đo thứ bậc chỉ có ý nghĩa xếp hạng, giá trị trung vị (median) là con số phù hợp để đo lường xu hướng trung tâm của dãy số biểu thị.

Khi thống kê các dữ liệu có được từ thang đo danh nghĩa hay thang đo thứ bậc, chúng ta có thể dùng các loại thống kê phi tham số (nonparametric tests) vì có nhiều loại trắc nghiệm thống kê mạnh, áp dụng một cách đơn giản, dễ tính toán và không đòi hỏi nhiều giả định như là thống kê tham số (parametric tests).

### 2.3 Thang đo khoảng cách (Interval Scales)

Thang đo khoảng cách có các đặc tính của thang đo danh nghĩa và thang đo thứ bậc, cộng thêm khả năng so sánh các khoảng cách giữa các cặp số (ví dụ khoảng cách sia biệt về thang đo giữa cặp số 1 và 2 tương đương với sai biệt giữa cặp số 2 và 3).

Có rất nhiều ứng dụng cụ thể đối với thang đo khoảng cách mà chúng ta sẽ thấy ở phần tiếp theo. Khi thang đo có tính chất khoảng cách và dữ liệu tương đối cân đối với một giá trị *mode*, ta có thể sử dụng giá trị *trung bình toán học* như là giá trị đo lường xu hướng trung tâm. Và vì vậy, ta có thể dùng giá trị độ lệch chuẩn (*standard deviation)* để đo lường sự phân tán của dữ liệu.

Khi phân phối các điểm số rút ra được từ thang đo khoảng cách bị thiên lệch về một hướng (lệch trái hoặc phải), chúng ta có thể sử dụng giá trị trung vị median để đo lường xu hướng trung tâm, và khoảng cách phân vị (*interquartile ran*ge) để đo lường độ phân tán.

### 2.4 Thang đo tỷ số (Ratio Scales)

Thang đo tỷ số **Ratio scales** có tất cả các đặc diểm của ba thang đo trên, cộng với đặc điểm có giá trị gốc là số không. Thang đo tỷ số thể hiện giá trị thực của một biến.

Trong nghiên cứu kinh tế, chúng ta áp dụng thang đo tỷ số cho rất nhiều loại dữ liệu dạng số thực, ví dụ như giá trị tiền, dân số, khoảng cách, thu nhập (bằng tiền), năng suất, sản lượng, v.v.

## 3. SAI SỐ TRONG ĐO LƯỜNG VÀ NGUỒN SAI SỐ

Một nghiên cứu lý tưởng nên được thiết kế và kiểm soát tốt để đạt được độ chính xác khi đo lường các biến. Nhưng rõ ràng, không nghiên cứu nào có thể đạt được sự kiểm soát hoàn hảo, nên sai số luôn luôn xảy ra. Hầu hết sai số có tính hệ thống (là kết quả từ sự sai lệch (bias), và phần còn lại là sai số ngẫu nhiên (xảy ra một cách bất thường). Có bốn nguồn sai số chủ yếu gây ảnh hưởng xấu đến kết quả nghiên cứu, là từ: (1)

người trả lời, (2) yếu tố tình huống, (3) người phỏng vấn, quan sát, đo lường, và (4) công cụ thu thập dữ liệu.

### 3.1 Nguồn sai số

**Người trả lời**

Sự sai biệt về ý kiến ảnh hưởng tới kết quả đo lường xuất phát từ các đặc điểm tính cách của người trả lời. Các đặc điểm này thể hiện các đặc trưng của tình trạng nghề nghiệp, dân tộc, tầng lớp xã hội, v.v. Người trả lời có thể:

- do dự, không muốn bày tỏ cảm nghĩ tích cực hoặc tiêu cực;
- bày tỏ thái độ một cách có chủ đích là họ cảm nhận sự việc khác với người khác;
- có ít hiểu biết về vấn đề được hỏi;
- bị tác động bởi các các yếu tố tạm thời như mệt mỏi, chán chường, giẫn dữ, bực tức, mất kiên nhẫn, các trạng thái tình cảm khác.

**Yếu tố tình huống**

Bất kỳ điều kiện nào gây ra sự căng thẳng cho cuộc phỏng vấn hay là quá trình đo lường đều có các ảnh hưởng nghiêm trọng đối với mối liên hệ giữa người phỏng vấn và người trả lời.

**Người phỏng vấn, quan sát, đo lường**

Người phỏng vấn có thể làm méo mó các trả lời bằng cách sửa từ, diễn giải dài dòng, hay là ghi nhận sai thông tin. Từ đó, sinh ra sai số. Ngoài ra, các trường hợp xử lý kỹ thuật thiếu cẩn thận như mã hóa sai, lập bảng sai, tính toán sai cũng gây ra nhiều sai số khác.

**Công cụ ghi nhận dữ liệu**

Công cụ ghi nhận dữ liệu bị thiếu sót cũng có thể gây ra sự bóp méo dữ liệu theo hai cách. Thứ nhất là gây ra sự lẫn lộn. Nghiệm trọng hơn nữa là chọn lựa các thông tin một cách nghèo nàn, đơn giản so với tổng thể các vấn đề cần quan tâm.

## 4. CÁC ĐẶC ĐIỂM CỦA MỘT ĐO LƯỜNG TỐT

Thế nào là một đo lường tốt? Làm sao có thể biết được các đặc tính thể hiện một đo lường tốt là gì? Có ba tiêu chuẩn để đánh giá một công cụ đo lường là: tính hợp lệ của dữ liệu, độ tin cậy, và tính thực tế.

## 4.1 Tính hợp lệ (Validity)

Tính hợp lệ của dữ liệu phản ảnh liệu chúng ta có đo lường đúng cái mà chúng ta muốn đo hay không. Tính hợp lệ phía ngoài (external validity) của kết quả nghiên cừu có nghĩa là dữ liệu  có khả năng tổng quát hóa được cho vấn đề nghiên cứu. Tính hợp lệ nội tại (internal validity) là là khả năng đo lường đúng cái mà chúng ta muốn đo lường.

Người ta thường chia tính hợp lệ thành ba kiểu chủ yếu: (1) hợp lệ về nội dung; (2) hợp lệ về tiêu chí và (3) hợp lệ về khái niệm.

### a. Hợp lệ về nội dung (Content validity)

Ta đạt được tính hợp lệ về nội dung của một công cụ đo lường khi nó cung cấp được một sự bao quát thích hợp đối với các câu hỏi điều tra dẫn dắt việc nghiên cứu. Nếu công cụ đo lường chứa đựng một mẫu đại diện của toàn thể về vấn đề cần quan tâm, thì sẽ có thể đạt được tính hợp lệ về nội dung.

Nhà nghiên cứu có thể đạt được tính hợp lệ về nội dung bằng các cách sau. Thứ nhất, thông qua việc xác định cẩn thận chủ đề nghiên cứu, soạn thảo các hạng mục cần đo lường, và các thang đo cần sử dụng. Cách thứ hai là tham vấn các chuyên gia xem liệu các công cụ đã thỏa mãn các tiêu chuẩn hay chưa. Và cũng quan trọng là đừng xác định các nội dung nghiên cứu quá hẹp.

### b. Hợp lệ về tiêu chí (Criterion-related validity)

Tính hợp lệ về tiêu chí phản ảnh sự thành công của việc đo lường nhằm mục tiêu phỏng định hoặc tiên lượng.

Nhà nghiên cứu phải bảo đảm là các tiêu chí sử dụng phái "hợp lệ". Bất kỳ tiêu chí nào cũng phải được xem xét theo bốn phẩm chất: (1) sự thích đáng - relevance, (2) không thiên lệch – freedom from bias, (3) sự tin cậy – reliability và (4) tính sẵn có - availability.

- Một tiêu chí là *thích đáng* khi nó đượcxác định và ghi nhận theo cách phù hợp, đúng đắn.
- *Không thiên lệch khi* mỗi người tham dự cung cấp dữ liệu thông tin đều có cơ hội ngang bằng nhau để cho điểm đúng nhất.
- Tiêu chí là *tin cậy* khi nó ổn định và có thể lặp lại được.
- Cuối cùng, thông tin được tiêu chí thể hiện có tính sẵn có. Nếu chưa sẵn có, thì liệu chi phí thu thập thông tin đó là bao nhiều, và lmức độ khó thu thập như thế nào.

**c. Hợp lệ về khái niệm (Construct validity)**

Đối với tính hợp lệ này, chúng ta phải xem xét cả hai khía cạnh về lý thuyết và công cụ đo lường được sử dụng.

**4.2 Tính tin cậy (Reliability)**

Ta sẽ đạt được sự tin cậy của một đo lường khi nó cung cấp được các kết quả nhất quán. Tính tin cậy là điều kiện cần, nhưng không phải là điều kiện đủ cho tính hợp lệ. Tinh tin cậy là mức độ mà dữ liệu không bị thiên lệch hoặc sai số quá mức. Các công cụ tin cậy là chúng không bị các yếu tố tình huống gây nhiễu. Các công cụ phải chắc chắn, và làm việc tốt ở mọi thời gian dưới mọi điều kiện. Thời gian và điều kiện là là những yếu tố cơ bản để đánh giá tính tin cậy, thông qua các chỉ tiêu – tính ổn định, tính tương đương và mực độ nhất quán nội tại.

**Tính ổn định (Stability)**

Một đo lường được coi là có tính ổn định khi chúng ta có thể bảo đảm kết quả nhất quán khi lặp lại trên cùng một người với cùng một công cụ. Một quá trình quan sát coi là ổn định khi nó cho cùng kết quả trên cùng một người khi lặp lại một hay nhiều lần.

Trong trường hợp điều tra phỏng vấn, ta khó đạt được tính bền vững của đo lường hơn so với nghiên cứu quan sát. Trong khi chúng ta quan sát lặp lại một vài hành động, chúng ta thường chỉ có thể điều tra lại một lần. Một vài khó khăn có thể xảy ra và gây ra sự thiên lệch như:

- *Thời gian giữa các lần đo lường quá lâu* dẫn đến sự thay đổi về tình huống.
- *Thời gian giữa các lần đo lường quá ngắn* làm cho người trả lời còn nhớ các câu trả lời trước đây, và lặp lại, làm cho độ tin cậy bị thiên lệch đi.
- *Người trả lời hiểu rõ mục tiêu thực của một nghiên cứu được ngụy trang* có thể gây ra thiên lệch nếu họ giữ riêng ý kiến liên quan tới mục tiêu nghiên cứu nhưng không được người nghiên cứu phát hiện.

**Tính tương đương (Equivalence)**

Vấn đề thứ hai ảnh hưởng đến tính tin cậy là liệu các điều tra viên khác nhau (trong nghiên cứu quan sát) hoặc các mẫu nghiên cứu khác nhau (trong các câu hỏi hoặc thang đo) có dẫn đến kết quả là sai số có khác nhau hay không. Vì vậy, tính tương đương được quan tâm ở khía cạnh sự biến thiên giữa các điều tra viên hoặc các mẫu nghiên cứu khác nhau. Có một phương thức tốt có thể được áp dụng để kiểm tra tính tương đương về kết quả ghi nhận được của các người quan sát - điều tra viên khác nhau là so sánh điểm mà họ cho trên cùng một sư kiện.

**Sự nhất quán nội tại (Internal Consistency)**

Cách tiếp cận thứ ba để đánh giá tính tin cậy là áp dụng một công cụ kiểm tra sự nhất quán nội tại, hoặc là tính đồng nhất (homogeneity) giữa các hạng mục nghiên cứu. Có thể dùng kỹ thuật *chia hai* (*split-half* technique) như là một công cụ để đánh giá khi trong bảng câu hỏi phỏng vấn có nhiều câu hỏi hay phát biểu tương tự nhau mà người được phỏng vấn có thể trả lời.

Các thực hiện là chia các hạng mục câu hỏi theo các số chẵn, lẻ hoặc chia làm hai phần một cách ngẫu nhiên. Khi xét sự tương quan giữa hai nửa này với nhau, nếu có tương quan chặt, thì ta có thể coi là kết quả nghiên cứu có tính tin cậy cao, theo ý nghĩa của tính nhất quán nội tại.

**4.3 Tính thực tế (Practicality)**

Các yêu cầu khoa học của một dự án nghiên cứu bao giờ cũng đòi hỏi quá trình đo lường phải có tính hợp lệ và tinh cậy, trong khi các yêu cầu hoạt động bao giờ cũng đòi hỏi quá trình đo lường phải có tính thực tế. Tính thực tế được xác định qua các tính chất *kinh tế (economy)*, *thuận tiện* (convienience) và *có khả năng diễn dịch được (interpretability)*.

**Kinh tế**

- Sự đánh đổi giữa sự lý tưởng của một nghiên cứu và ngân sách nghiên cứu.
- Nhiều hạng mục đo lường sẽ làm tăng độ tin cậy, nhưng cũng làm tăng chi phí.
- Việc lựa chọn phương thức thu thập dữ liệu có thể bị điều kiện kinh tế khống chế (ví dụ, phỏng vấn cá nhân và phỏng vấn qua điện thoại).

**Sự thuận tiện**

- Các phương tiện đo lường nên được quản lý một cách dễ dàng.
- Các phiếu điều tra, bảng hỏi hoặc các thang đo với các hướng dẫn cụ thể sẽ được trả lời đúng một cách dễ dàng hơn.
- Hiển nhiên là, nếu các khái niệm trong nghiên cứu càng phức tạp thì càng cần có các hướng dẫn cụ thể, rõ ràng.

**Khả năng diễn dịch được**

Các người thiết kế các công cụ thu thập dữ liệu phải cung cấp các thông tin mang tính hướng dẫn để làm cho việc diễn dịch câu hỏi trở nên dễ dàng hơn:

- Phát biểu về các chức năng của các trắc nghiệm được thiết kế để đo lường và các thủ tục liên quan.
- Các hướng dẫn chi tiết về quản lý.

- Các hướng dẫn về cách cho điểm.
- Các quy tắc.
- Các quan hệ nội bộ của các điểm phụ (sub-scores).
- Các quan hệ giữa kiểm định và các cách đo lường khác.
- Hướng dẫn sử dụng kiểm định.

## 5. BẢN CHẤT CỦA THÁI ĐỘ

Trong nghiên cứu kinh tế, nhiều khi đối tượng nghiên cứu chính là các cá nhân con người, hoặc là người đại diện cho một hộ, một gia đình, hoặc một nhóm người nào đó. Vì vậy, việc hiểu biết về thái độ và cách thức đo lường thái độ áp dụng các thang đo phù hợp là hết sức quan trọng. Chúng ta cũng cần phải hiểu các khía cạnh quan trọng khác nhau của thái độ, như là tính lâu bền tương đối của chúng, và sự liên hệ giữa thái độ với các sự kiện và các đối tượng xã hội.

Thái độ có thể được chia làm các dạng như sau:

*Thái độ dựa trên nhận thức (Cognitively based attitude)* thể hiện ký ức, dự đánh giá, và niềm tin về các tính chất của đối tượng nghiên cứu.

*Thái độ dựa trên xúc cảm (Affectively based attitude)* thể hiện các cảm giác, trực giác, giá trị và sự xúc cảm về đối tượng nào đó.

*Thái độ dự trên hành vi (Behaviorally based attitude)* phản ảnh các mong đợi và dự định mang tính hành vi đối với đối tượng nào đó.

### 5.1 Quan hệ giữa Thái độ và Hành vi

Quan hệ giữa thái độ và hành vi không có tính trực tiếp, mặc dù chúng có sự liên kết gần gũi với nhau. Thái độ và hành vi dự định không phải luôn luôn nhất quán với nhau. Ngoài ra, hành vi có thể ảnh hưởng đến thái độ.

Chúng ta cần xem xét thái độ một cách cẩn thận, vì tính phức tạp của nó, và mặc dù là ta có thể suy luận ra thái độ dựa trên các dữ liệu đo lường được, nhưng thực tế là ta không thực sự quan sát được thái độ.

Ngoài ra, nhà nghiên cứu cũng phải chú ý đến những khía cạnh mà dựa trên thái độ có thể tiên đoán được hành vi. Một số yếu tố tác động đến khả năng áp dụng nghiên cứu thái độ là:

- Khi tiên đoán hành vi, các thái độ đặc thù là các chỉ số tiên đoán tốt hơn là các thái độ chung chung.

- Khi tiên đoán hành vi, các thái độ mạnh mẽ là chỉ số tiên đoán tốt hơn là các thái độ yếu ớt.
- Các kinh nghiệm trực tiếp về một đối tượng nào đó sinh ra các hành vi tin cậy hơn.
- Thái độ dựa trên nhận thức ảnh hưởng đến hành vi nhiều hơn là thái độ dựa trên xúc cảm.
- Thái độ dựa trên xúc cảm thường là các chỉ số tiên đoán tốt hơn đối vớ ciác hành vi tiêu dùng.
- Sử dụng nhiều loại đo lường thái độ hoặc đánh giá hành vi nhiều lần theo thời gian giúp cải thiện các chỉ số tiên đoán.

### 5.2 Lập thang đo thái độ (Attitude Scaling)

Lập thang đo thái độ là một quá trình đánh giá một sự biểu hiện thái độ nào đó mà nhà nghiên cứu sử dụng một con số để thể hiện một điểm số đo lường thái độ dịch chuyển trong một khoảng từ thái độ biểu thị rất hài lòng đến rất không hài lòng.

Lập thang đo là "quá trình gán những điểm số cho một tính chất của một đối tượng nhằm sử dụng các đặc điểm của các điểm số để biểu thị cho các tính chất". Thông thường, chúng ta gán các điểm số cho các chỉ tiêu thể hiện các tính chất của đối tượng.

## 6. CHỌN LỰA MỘT THANG ĐO

Việc chọn lựa và xây dựng một thang đo đòi hỏi phải xem xét đến các yếu tố ảnh hưởng đến tính hợp lệ, tính tin cậy, và tính thực tế của thang đo. Các yếu tố này là:

- Mục tiêu nghiên cứu
- Các kiểu trả lời
- Tính chất của dữ liệu
- Số lượng chiều kích (dimensions)
- Cân xứng hay bất cân xứng
- Chọn lựa bắt buộc / không bắt buộc
- Số lượng điểm đo
- Sai số do người đánh giá gây ra

### 6.1 Mục tiêu nghiên cứu

Nhà nghiên cứu phải đối mặt với hai dạng mục tiêu nghiên cứu liên quan đến việc lập thang đo:

- Mục tiêu nhằm đo lường các đặc điểm của người tham gia vào nghiên cứu dưới góc độ là người trả lời phỏng vấn.
- Mục tiêu nhằm yêu cầu người tham gia để đánh giá một đối tượng nào đó.

Thông thường, trong nghiên cứu kinh tế, chúng ta thường kết hợp cả hai mục tiêu này vào trong một cuộc nghiên cứu. Nhà nghiên cứu thường đặt ra các thang đo để ghi nhận các đặc điểm cá nhân hoặc của gia đình, tầng lớp xã hội, v.v của người tham gia vào nghiên cứu, và cũng đồng thời, hỏi họ các ý kiến đánh giá về các đối tượng nghiên cứu mà nhà nghiên cứu quan tâm.

### 6.2 Các kiểu trả lời

Các thang đo thường rơi vào một trong bốn kiểu phổ biến sau: cho diểm, xếp hạng, phân loại và sắp xếp thứ tự.

**Thang đo cho điểm** (**rating scale)** là loại thang đo mà người tham gia sẽ cho điểm một đối tượng hoặc một chỉ tiêu nào đó mà không cần phải so sánh trực tiếp với một đối tượng khác.

**Thang đo xếp hạng (Ranking scale)** đòi hỏi người tham gia phải so sánh và quyết định trật tự thứ bậc giữa hai hoặc nhiều hơn các tính chất (hoặc các chỉ tiêu) hoặc các đối tượng.

**Thang đo phân loại (Categorization)** yêu cầu người tham gia phải tự phân loại chính họ hoặc các chỉ tiêu, các tính chất vào các nhóm, các loại khác nhau.

**Thang đo sắp xếp thứ tự (Sorting)** yêu cầu người tham gia sắp xếp các thẻ (đại diện cho các khái niệm nào đó) thành những nhóm khác nhau áp dụng các nguyên tắc phân loại do nhà nghiên cứu đưa ra.

### 6.3 Tính chất của dữ liệu

Khi quyết định sử dụng thang đo nào, nhà nghiên cứu thường xem xét đến các tính chất của dữ liệu có thể có được theo từng loại thang đo: danh nghĩa, thứ bậc, khoảng cách và tỷ số.

### 6.4 Số lượng chiều kích

Thang đo có thể hoặc là đơn chiều (*uni-dimensional)* hoặc đa chiều (*multi-dimensional).*

- Với một thang đo đơn chiều, người ta tìm cách đo lường chỉ một thuộc tính nào đó của người tham gia hoặc đối tượng nghiên cứu.
- Với một thang đo đa chiều, người ta mong muốn mô tả kỹ lưỡng hơn một đối tượng nào đó với nhiều chiều khác nhau hơn là chỉ một chiều duy nhất.

### 6.5 Cân xứng hoặc bất cân xứng

Một thang đo cho điểm cân xứng có số lượng các hạng mục phân hạng bằng nhau về phía trên hoặc phía dưới của điểm giữa. Nói chung, thang đo cho điểm nên cân xứng, một một số lượng cân bằng các cơ hội chọn lựa mang tính ưa thích hay không ưa thích.

Ví dụ: rất tốt – tốt – trung bình – tệ – rất tệ

Một thang đo cho diểm bất cân xứng có số lượng các cơ hội chọn lựa không cân bằng với nhau.

Ví dụ: Tệ – khá – tốt – rất tốt – tuyệt vời

Thông thường, chúng ta chỉ áp dụng thang đo cho điểm bất cân xứng khi chúng ta biết trước là hầu hết người đánh giá sẽ thiên về một hướng này hoặc hướng khác.

### 6.6 Bắt buộc hay không bắt buộc

Một thang đo cho điểm không bắt buộc luôn cho người trả lời một cơ hội để bày tỏ là họ không có ý kiến khi mà họ không thể quyết định chọn lựa bất kỳ một mục trả lời nào.

Ví dụ "không ý kiến", "không quyết định được", "không biết", "không chắc chắn"

Một thang đo cho điểm bắt buộc đòi hỏi người trả lời phải chọn một trong những mục chọn đề nghị.

### 6.7 Số lượng điểm đo

Ta cần bao nhiêu điểm đo cho một thang đo là vừa? Bao nhiêu điểm đo là lý tưởng?

Câu trả lời mang tính thực tế: một thang đo nên phù hợp với mục tiêu của nó. Để cho một thang đo mang tính hữu ích, thang đo đó nên phải phù hợp với các thông tin có thể rút ra được tỷ lệ với mức độ phức tạp của tính chất, đối tượng hoặc khái niệm mà ta đang nghiên cứu. Do đó, số lượng điểm phản ảnh tính chất phức tạp của đối tượng hoặc khái niệm mà chúng ta có thể hiểu được. Nếu số lượng điểm đo càng nhiều, mức độ biểu thị về chi tiết càng tăng, và ta có thể diễn giải sâu hơn các tính chất phức tạp của khái niệm. Tuy nhiên, khi số lượng điểm quá nhiều, ta có thể không phân biệt được một cách rõ rành ranh giới, hoặc sự khác biệt giữa các mức độ của tính chất, khái niệm. Ngược lại, nếu số lượng điểm đo quá ít, thì ta không thể phản ảnh đầy đủ bản chất phức tạp của tính chất, đối tượng hoặc khái niệm nghiên cứu.

Số lượng điểm có thể là 3, 5 hoặc nhiều hơn. Thông thường, người ta thường áp dụng thang đo 5 điểm hoặc 7 điểm.

### 6.8 Sai số do người đánh giá gây ra

Giá trị của các điểm số đo lường còn dựa trên giả định là người đánh giá sẽ có thể đưa ra phán quyết tốt. Tuy nhiên, chúng ta cũng nên xem xét lại liệu người đánh giá có xu hướng cho điểm lệch về một phía nào hay không.

Một dạng là người đánh giá ngại đưa ra các phán quyết mang tính đối cực, nên có xu hướng cho điểm xoay quanh giá trị điểm giữa. Lỗi này được gọi là lỗi theo xu hướng trung tâm (error of central tendency).

Người đánh giá cũng có thể là người cho điểm quá khó, hoặc cho điểm quá dễ, và lội này được gọi là "lỗi khoan dung" (error of leniency).

## 7. THANG ĐO CHO ĐIỂM

### 7.1 Thang đo thái độ giản đơn (Simple Attitude Scales)

Thang đo thái độ giản đơn được thiết lập nhằm ghi nhận sự đánh giá hoặc chọn lựa của người tham gia về một tính chất hay đối tượng nào đó. Thang đo này bao gồm một số loại phụ như sau:

- Thang đo thái độ giản đơn (**simple category scale** - *dichotomous scale*) có hai lựa chọn đơn giản (ví dụ, có/không; đồng ý/không đồng ý; quan trọng/không quan trọng).
- Thang đo nhiều lựa chọn, một trả lời (**multiple choice, single-response scale)**: nhiều mục lựa chọn; chỉ có một trả lời
- Thang đo nhiều lựa chọn, nhiều trả lời (**multiple-choice, multiple-response scale** - checklist): cho phép người trả lời chọn nhiều lựa chọn.

Thang đo thái độ giản đơn có tính chất là dễ thiết lập, có tính chuyên biệt cao, cung cấp thông tin hữu ích và phù hợp nếu có kỹ năng thiết lập.

## Các ví dụ về thang đo cho điểm

**Thang đo phân loại giản đơn**

(lưỡng phân)

Dữ liệu: danh nghĩa

Bạn có kế hoạch mua máy tính trong vòng 12 tháng tới hay không?

☐ Có

☐ Không

**Thang đo nhiều lựa chọn, một trả lời**

Dữ liệu: danh nghĩa

Bạn đọc tin tài chính chủ yếu ở tờ báo nào?

☐ Thanh Niên

☐ Tuổi Trẻ

☐ Thời Báo Kinh Tế Sài Gòn

☐ Người Lao Động

☐ Khác, (ghi rõ: …….)

**Thang đo nhiều lựa chọn, một trả lời**

**(checklist)**

Dữ liệu: danh nghĩa

Đánh dấu các nguồn thông tin mà bạn tham vấn khi vẽ kiểu nhà?

☐ Dich vụ lập bản vẽ trực tuyến

☐ Tại chí kiến trúc

☐ Các nhà xây dựng độc lập

☐ Các kiểu mẫu, bản vẽ của các nhà phát triển

☐ Nhà thiết kế

☐ Kiến trúc sư

☐ Khác (ghi rõ:……)

**Thang đo Likert**

**Cho điểm có thể cộng hợp được**

Dữ liệu: khoảng cách

Internet tốt hơn thư viện truyền thống khi tìm tài liệu chuyên sâu

| Rất đồng ý | Đồng ý | Không đồng ý Không phản đối | Phản đối | Rất phản đối |
|---|---|---|---|---|
| (5) | (4) | (3) | (2) | (1) |

**Thang đo trắc biệt Semantic Differential Scale**

Dữ liệu: khoảng cách

Khả năng giao hàng của dịch vụ phát chuyển nhanh TNT

NHANH ------: -----: -----: -----: -------: -------:------ CHẬM

CHẤT LƯỢNG CAO ---- : ----: ----: ----: ----:----: ---- : CHẤT LƯỢNG THẤP

**Thang đo số**

**Numerical Scale**

Dữ liệu: thứ bậc hoặc khoảng cách

Rất hài lòng 5 4 3 2 1 Rất không hài lòng

Sự hợp tác của nhân viên: …….

Kiến thức thực thi nhiệm vụ của nhân viên: …….

Hiệu quả lập kế hoạch của nhân viên: ……..

**Thang đo danh sách cho điểm**

**Multiple Rating List Scale**

Dữ liệu: khoảng cách

Vui lòng chỉ ra các tính chất dịch vụ sau đây quan trọng hoặc không quan trọng như thế nào

| | QUAN TRỌNG | | | | | | KHÔNG QUAN TRỌNG |
|---|---|---|---|---|---|---|---|
| Sửa chữa nhanh và tin cậy | 7 | 6 | 5 | 4 | 3 | 2 | 1 |
| Cung cấp dịch vụ tại nhà | 7 | 6 | 5 | 4 | 3 | 2 | 1 |
| Nhà chế tạo bảo trì | 7 | 6 | 5 | 4 | 3 | 2 | 1 |
| Kỹ thuật viên có kiến thức | 7 | 6 | 5 | 4 | 3 | 2 | 1 |
| Nhắc nhở nâng cấp | 7 | 6 | 5 | 4 | 3 | 2 | 1 |
| Hợp đồng dịch vụ sau bảo hành | 7 | 6 | 5 | 4 | 3 | 2 | 1 |

**Thang đo tổng hằng số**

**Constant-Sum Scale**

Dữ liệu: tỷ số

Hãy cho điểm các đặc tính của nhà cung cấp và xem xét tới chi phí. Loại nào (đặc tính và chi phí) là tương đối quan trọng với bạn? (tổng điểm là 100)

| | |
|---|---|
| Nhà cung cấp có chi phí thấp | |
| Các tính chất khác của nhà cung cấp | |
| Tổng | 100 |

**Thang đo Stapel**

**Stapel Scale**

Dữ liệu: thứ bậc hoặc danh nghĩa

**(Tên công ty)**

| | | |
|---|---|---|
| +5 | +5 | +5 |
| +4 | +4 | +4 |
| +3 | +3 | +3 |
| +2 | +2 | +2 |
| +1 | +1 | +1 |
| Dẫn đầu về công nghệ | Sản phẩm thú vị | Danh tiếng thế giới |
| -1 | -1 | -1 |
| -2 | -2 | -2 |
| -3 | -3 | -3 |
| -4 | -4 | -4 |
| -5 | -5 | -5 |

**Thang đo cho điểm đồ thị -Graphic Rating Scale**

Dữ liệu: thứ bậc, khoảng cách hay tỷ số

**Rất có thể** | | **Rất không có thể**

### 7.2 Thang đo Likert (Likert Scales)

Thang đo Likert, được ông Rensis Likert phát triển, là loại thang đo được sử dụng rất nhiều như là thang đo cho điểm có thể cộng điểm được (summated rating). Thang đo này bao gồm một phát biểu thể hiện một thái độ ưa thích hay không ưu thích, tốt hay xấu về một đối tượng nào đó.

Người tham dự được hỏi để trả lời đồng ý hay không với từng câu phát biểu. Mỗi trả lời được cho 1 điểm số phản ảnh mức độ ưa thích, và các điểm số có thể tổng hợp được để đo lường thái độ chung của người tham dự.

Thang đo Likert có thể là 5, 7 hoặc 9 điểm.

Lợi thế của thang đo Likert:

- Thiết lập dễ dàng và nhanh chóng.
- Tin cậy nhiều hơn và cung cấp nhiều lượng thông tin hơn nhiều loại thang đo khác.
- Dữ liệu đạt được là dữ liệu khoảng cách.

Cách thiết lập thang đo Likert

- Chọn một số lượng lớn phát biểu có hai tính chất: (1) phù hợp với thái độ được nghiên cứu; (2) phản ảnh vị trí của thái độ ưa thích hay không ưa thích.
- Người tham dự đọc từng phát biểu và cho điểm, sử dụng thang đo 5 điểm. Giá trị (1) có nghĩa thái độ rất không ưa thích. Giá trị (5) có nghĩa rất ưa thích.
- Các trả lời của mỗi người được cộng dồn để có một điểm tổng.
- Xếp dãy các điểm tổng để chọn các phần có điểm tổng cao nhất và thấp nhất (10 - 25% số có điểm cao nhất và thấp nhất).
- Hai nhóm có điểm tổng cao nhất và thấp nhất được đánh giá theo từng câu trả lời riêng lẻ.
- Tính các giá trị trung bình của từng nhóm có điểm cao nhất và thấp nhất, rồi kiểm định sự khác biệt dùng *t* test.
- Sau khi kiểm định *t* cho từng phát biểu, xếp hạng các giá trị trung bình, rồi chọn các phát biểu có giá trị *t* cao nhất.
- Chọn 20 - 25 mục có giá trị *t* cao nhất để gộp vào điểm cuối cùng.

### 7.3 Thang đo trắc biệt (Semantic Differential Scales - SD)

Thang đo trắc biệt nhằm đo lường ý nghĩa tâm lý của một đánh giá về đối tượng nghiên cứu sử dụng 2 tính từ đối cực. Thang đo này thường được dùng để đánh giá hình ảnh thương hiệu.

Phương pháp này bao gồm một bộ các thang đo cho điểm 2 cực, thường sử dụng thang đo 7 điểm.

Thang đo trắc biệt dựa trên giả định là một đối tượng có thể có nhiều chiều để đo lường ý nghĩa. Các ý nghĩa được định vị trong một không gian đa chiều, gọi là "không gian ý nghĩa" (semantic space).

Lợi thế của thang đo SD:

- Có hiệu quả và dễ dàng để đo lường thái độ từ một mẫu lớn.
- Có thể đo lường theo cả hướng (direction) và độ tập trung (intensity).
- Bộ tổng của các trả lời cung cấp một bức tranh sâu sắc về ý nghĩa của một đối tượng và sự đo lường của người đánh giá, cho điểm.
- Là một kỹ thuật chuẩn hóa, dễ lặp lại và không bị bóp méo.
- Cho dữ liệu dạng khoảng cách.

**Các bước xây dựng thang đo trắc biệt**

1. Chọn khái niệm: danh từ, nhóm danh từ, hoặc các phác họa hình ảnh. Các khái niệm được chọn sau khi xem xét, đánh giá và bằng khả năng phản ảnh bản chất của câu hỏi điều tra.
2. Chọn các cặp từ hoặc cụm từ đối cực phù hợp theo nhu cầu.
3. Tạo ra hệ thống tính điểm có trọng số. Hầu hết thang đo SD có 7 điểm: 7, 6, 5, 4 3, 2, và 1.
4. Tương tự như thang đo Likert, khoảng ½ các tính từ được lưu giữ một cách ngẫu nhiên để tối thiểu hóa hiệu ứng "halo".

### 7.4 Thang đo số/Thang đo danh sách cho điểm (Numerical/Multiple Rating List Scales)

Các thang đo số có các khoảng cách tương đương chia theo 5 hoặc 7 hoặc 10 điểm.

Người tham gia cho điểm (viết kế bên mục chọn). Nếu các câu hỏi cùng thể hiện tính chất của một sản phẩm nào đó, thì thang đo có thể cung cấp cả kết quả đo lường tuyệt đối về mức quan trọng và kết quả đo lường tương đối (xếp hạng) của các mục chọn khác nhau phản ảnh tính chất đó.

Lợi thế của thang đo số/danh sách cho điểm:

- Có tính tuyến tính
- Có tính giản đơn
- Tạo ra dữ liệu thứ bậc hoặc khoảng cách

Thang đo danh sách cho điểm cũng tương tự như thang đo số, nhưng có hai điểm khác biệt:

(1) Cho phép người đánh giá tự khoanh tròn mục số chọn lựa

(2) Hình thức thể hiện của thang đo này cho phép chúng ta hình tượng hóa kết quả. Dựa vào đó, chúng ta có thể xây dựng một bản đồ trí tuệ về sự đánh giá của các người tham gia để nghiên cứu sâu hơn.

### 7.5 Thang đo Stapel

Thang đo Stapel được sử dụng như là một phương pháp thay thế cho thang đo trắc biệt, nhất là khi chúng ta không thể tìm được một cặp tính từ đối cực phù hợp hợp câu hỏi điều tra.

Với thang đo này, chúng ta thường dùng thang đo số 5 điểm. Đôi khi thang đo ít điêm trả lời hơn cũng được áp dụng. Người tham gia sẽ chọn một con số để mô tả đắc điểm của đối tượng. Nếu sự mô tả càng chính xác, thì lựa chọn con số dương càng có giá trị lớn. Ngược lại, nếu sự mô tả càng kém chính xác, thì lựa chọn con số âm có giá trị tuyệt dối lớn. Thông thường, điểm số sẽ dao động trong khoảng từ +5 đến -5, và người tham gia sẽ lựa chọn một con số nào đó để mô tả các trạng thái từ rất chính xác đên không chính xác.

Giống như thang đo Likert, trắc biệt và số, thang đo Stapel cho dữ liệu khoảng cách.

### 7.6 Thang đo tổng-hằng số (Constant-Sum Scales)

Đây là loại thang đo cho phép nhà nghiên cứu phát hiện các tỷ lệ của các thuộc tính khác nhau trong đánh giá một đối tượng nào đó. Thang đo này yêu cầu người cho điểm phải phân phối các điểm số cho nhiều thuộc tính khác nhau và tổng của các điểm số này phải là một hằng số, ví dụ 100 hoặc 10.

Ưu điểm của thang đo này là:

- Tương thích với hệ thống tỷ lệ phần trăm, và có thể đo lường mức độ khác biệt theo tầm quan trọng của các thuộc tính so với nhau.
- Thường được sử dụng để ghi nhận thái độ, hành vi và các dự định hành vi.
- Tạo ra dữ liệu khoảng cách.

### 7.7 Thang đo cho điểm đồ thị (Graphic Rating Scales)

Thang đo này được tạo ra nhằm tạo điều kiện cho nhà nghiên cứu khám phá các sự khác biệt cực nhỏ. Về lý thuyết, có thể áp dụng một con số không xác định các điểm số để cho điểm nếu người tham gia đủ nhạy cảm để phát hiện sự khác biệt và ghi nhận chúng.

Người tham gia sẽ đánh dấu điểm trả lời của họ ở bất kỳ điểm nào dọc theo một cột liên tục. Thông thường, điểm số sẽ được xác định bằng cách đo độ dài của nó (tính bằng mi-li-mét) tính từ điểm khởi đầu. Kết quả cho dữ liệu khoảng cách.

Khó khăn trong việc áp dụng thang đo này là mã hóa và phân tích.

## 8. THANG ĐO XẾP HẠNG (RANKING SCALES)

Đối với thang đo xếp hạng, người tham gia so sánh trực tiếp hai đối tượng hoặc nhiều hơn với nhau và phải chọn lựa giữa chúng với nhau. Thông thường, người tham gia được yêu cầu chọn lấy một đối tượng được coi là "tốt nhất" hoặc "được ưa thích nhất". Khi chỉ có hai đối tượng được so sánh thì cách tiếp cận này rất dễ thực hiện, nhưng khi có nhiều hơn hai đối tượng so sánh thì rất khó có được kết quả chính xác.

### 8.1 Thang đo so sánh cặp (Paired-Comparison Scales)

Sử dụng thang đo so sánh cặp, người tham gia có thể bảy tỏ thái độ rõ ràng bằng cách chọn lựa một trong hai đối tượng.

Số lượng cặp so sánh được tính theo công thức $[(n)(n\text{-}1)/2]$, với $n$ là số lượng các đối tượng phải so sánh với nhau.

Sop sánh cặp có rủi ro là người tham gia sẽ quá mệt mỏi và không thể cho trả lời đúng, hoặc từ chối tiếp tục đánh giá. Người ta cho rằng nếu người tham gia còn phải trả lời các câu hỏi khác thì 5 – 6 đối tượng so sánh là quá nhiều rồi. Ngược lại, nếu dữ liệu cần thu thập chỉ bao gồm các so sánh cặp, thì có thể thiết kế câu hỏi với 10 đối tượng so sánh.

Thang đo so sánh cặp cho dữ liệu thứ bậc.

### 8.2 Thang đo xếp hạng bắt buộc (Forced-Ranking Scales)

Thang đo xếp hạng bắt buộc liệt kê danh sách các thuộc tính mà chúng có thể được xếp hạng một cách tương đối so với nhau. Phương pháp này cho kết quả nhanh hơn, dễ sử dụng hơn và khuyến khích người tham gia nhiều hơn trong việc đánh giá.

Với 5 đối tượng hoặc hạng mục, ta sẽ có đến 10 cặp so sánh, trong khi nếu xếp hạng 5 đối tượng, công việc sẽ dễ dàng hơn nhiều.

Tương tự như trên, luôn có câu hỏi là số lượng đối tượng hoặc hạng mục so sánh bao nhiêu là vừa. Nếu xếp hạng 5 đối tượng thì dễ dàng, nhưng nếu có 10 đối tượng trở lên, kết quả có thể không chính xác và người tham gia có thể xếp hạng không cẩn thận.

Ngoài ra, thang đo xếp hạng cho dữ liệu thứ bậc, vì ta không rõ khoảng cách giữa các khác biệt là bao nhiêu, và có bằng nhau hay không.

### 8.3 Thang đo so sánh (Comparative Scale)

Sử dụng một điểm, một đối tượng quy ước, thang đo so sánh sẽ là lý tưởng nếu người tham gia đã quen thuộc với một tiêu chuẩn nào đó. Thang do này đòi hỏi người tham gia so sánh một đối tượng hoặc các thuộc tính của đối tượng với một đối tượng được coi là chuẩn mực.

Về dạng dữ liệu, một vài nhà nghiên cứu cho là thang đo so sánh có thể cho dữ liệu khoảng cách khi điểm số phản ảnh sự khác biệt có tính khoảng cách giữa tiêu chuẩn và đối tượng quan sát. Chúng ta cũng có thể coi dữ liệu có tính thứ bậc trừ phi chứng minh được có sự tuyến tính giữa các biến được hỏi.

### Thang đo xếp hạng - Ranking Scales

**Thang đo so sánh cặp**

**Paired-Comparison Scale**

Dữ liệu: thứ bậc

Xin cho biết trong hai nhãn hiệu xe dưới đây, anh chị thích nhãn hiệu nào hơn?

------ Yamaha Nouvo LX

------ Honda Air Blade FI

**Thang đo xếp hạng bắt buộc**

**Forced Ranking Scale**

Dữ liệu: thứ bậc

Xếp hạng các đặc tính của bộ thu tín hiệu như trình bày dưới đây:

Số 1: thích nhất; số 2: kế tiếp…

----- Người sử dụng có thể lập trình

----- Tính năng không dây

----- Kích thước nhỏ gọn

----- Chức năng cảnh báo từ xa

----- Chức năng giảm thiểu báo động sai

**Thang đo so sánh**

**Comparative Scale**

Dữ liệu: thứ bậc

Hãy so sánh máy sấy tóc nhãn hiệu X này với máy sấy mà bạn đang dùng. Máy X thì:

| TỐT HƠN | | TƯƠNG ĐƯƠNG | | XẤU HƠN |
|---|---|---|---|---|
| ------ | ------ | ------ | ------ | ------ |
| 1 | 2 | 3 | 4 | 5 |

**Từ khóa**

| | |
|---|---|
| Đo lường | Measurement |
| Đối tượng | Objects |
| Thang đo khoảng cách | Interval scale |
| Thang đo danh nghĩa | Nominal scale |
| Thang đo thứ bậc | Ordinal scale |
| Thang đo tỷ số | Ratio scale |
| Tính chất | Properties |
| Tính hợp lệ nội bộ | Internal validity |
| Tính thực tế | Practicality |
| Tính tin cậy | Reliability |
| Tính hợp lệ | Validity |
| Thái độ | Attitude |
| Thang đo | Scale |
| Thang đo cho điểm | Rating scale |
| Thang đo xếp hạng | Rating scale |
| Thang đo so sánh | Comparative scale |
| Thang đo Likert | Likert scale |
| Thang đo xếp hạng cân xứng | Balanced rating scale |
| Thang đo nhiều lựa chọn, nhiều trả lời | Multiple-choice, multiple-response scale |
| Thang đo nhiều lựa chọn, một trả lời | Multiple-choice, single-response scale |
| Thang đo trắc biệt | Semantic differential (SD) scale |
| Thang đo tổng - hằng số | Constant-sum scale |
| Lập thang đo | Scaling |
| Phân loại | Categorization |
| Thang đo đa chiều | Multidimensional scale |
| Thang đo đơn hướng | Unidimensional scale |
| Thang đo phân loại giản đơn | Simple category scale |
| Thang đo Stapel | Stapel scale |

| | |
|---|---|
| Thang đo cho điểm tổng hợp | Summated rating scale |
| Thang đo cho điểm bất cân xứng | Unbalanced rating scale |
| Thang đo cho điểm không bắt buộc | Unforced-choice rating scale |
| Thang đo cho điểm danh sách | Multiple rating list scale |
| Thang đo số | Numerical scale |
| Thang đo so sánh cặp | Paired-comparison scale |
| Thang đo xếp hạng | Ranking scale |

# Chương 6. Phương pháp chọn mẫu và xác định cỡ mẫu

**Mục tiêu giảng dạy**

Sau khi học chương này, sinh viên có thể hiểu được:

1. Bản chất của việc chọn mẫu.
2. Các đặc điểm về tính đúng đắn và tính chính xác để đo lường mức độ hợp lệ của mẫu.
3. Các nguyên tắc cần thiết khi phát triển một kế hoạch chọn mẫu.
4. Hai nhóm kỹ thuật chọn mẫu và các phương pháp cụ thể.
5. Các kỹ thuật chọn mẫu và điều kiện áp dụng.

## 1. BẢN CHẤT CỦA VIỆC CHỌN MẪU

Chọn mẫu (**sampling)** là việc chọn lấy một số thành phần của một dân số (population), và từ đó, có thể rút ra các kết luận về chính dân số đó.

Một thành phần của dân số (**population element)** là một cá thể của đối tượng nghiên cứu hoặc một cá nhân người tham gia nghiên cứu mà nhà nghiên cứu sẽ tiến hành các đo lường. Đây chính là đơn vị nghiên cứu (unit of study).

Một dân số bao gồm tất cả các thành phần của dân số mà chúng ta muốn nghiên cứu.

Một điều tra tổng thể (census) là một nghiên cứu thực hiện trên tất cả mọi thành phần của dân số.

Danh sách của tất cả các thành phần của dân số mà dựa vào đó chúng ta rút ra mẫu được gọi là "khung mẫu" (**sample frame)**.

### 1.1 Tại sao phải lấy mẫu?

Khi thực hiện nghiên cứu, chúng ta rất hiếm khi điều tra tổng thể, vì lý do cơ bản là hết sức tốn kém và tốn rất nhiều thời gian, công sức. Trong khi đó, nếu chúng ta chỉ điều tra mẫu, thì có nhiều lợi thế như: (1) chi phí thấp, (2) vẫn đạt được tốt hơn độ chính xác cần có của kết quả, (3) đạt tốc độ thu thập dữ liệu nhanh; và (4) có được sự sẵn có của các thành phần dân số.

**Chi phí thấp**

Rõ ràng là điều tra nghiên cứu trên một mẫu nào đó của dân số sẽ có lợi thế về chi phí nhiều hơn là điều tra tổng thể.

**Đạt được tốt hơn độ chính xác cần có của kết quả**

Chất lượng của một nghiên cứu thực hiện điều tra chọn mẫu hoặc nghiên cứu trên mẫu thường đạt kết quả tốt hơn so với thực hiện điều tra tổng thể hoặc nghiên cứu tổng thể vì:

- Phỏng vấn tốt hơn.
- Điều tra nhiều hơn, sâu hơn về các thông tin nghi ngờ, sai, hoặc sót.
- Xử lý thông tin tốt hơn.

Chỉ khi nào dân số nghiên cứu quá nhỏ, dễ tiếp cận, và biến động nhiều thì điều tra tổng thể mới có thể đạt độ chính xác cao hơn điều tra mẫu.

**Tốc độ thu thập dữ liệu cao hơn**

Tốc độ thực hiện nhanh giúp làm giảm thời gian giữa giai đoạn chuẩn bị các thông tin cần thiết và giai đoạn thu thập thông tin.

**Tính sẵn có của các thành phần dân số**

Thông thường, một số thành phần dân số luôn có sẵn, và chúng ta có thể chọn lựa để thực hiện lấy mẫu để điều tra, nghiên cứu.

**Mẫu và Tổng thể**

Lợi thế của điều tra mẫu so với điều tra tổng thể sẽ mất đi nếu dân số nhỏ và có tính biến động cao. Có hai điều kiện làm cho việc nghiên cứu tổng thể phù hợp hơn: (1) có tính khả thi khi dân số nhỏ và (2) cần thiết khi mà mỗi cá thể đều rất khác biệt nhau.

## 1.2 Thế nào là một mẫu tốt?

Một mẫu được coi là tốt khi nó có thể đại diện cho các tính chất của dân số mà nó được rút ra. Nói theo thuật ngữ đo lường, nó phải có tính hiệu lực (validity). Tính hiệu lực của mẫu tùy thuộc vào hai tính chất: tính đúng đắn (accuracy) và tính chính xác (precision).

**Tính đúng đắn (Accuracy)**

Tính đúng đắn là mức độ mà mẫu tránh được các thiên lệch (bias). Khi mẫu được rút ra đúng cách, thì các các tính chất của một số thành phần nào đó của dân số sẽ được thể hiện ít hơn mức độ thực có của chúng. Ngược lại, sẽ có một số thành phần khác sẽ được thể hiện nhiều hơn mức độ thực có của chúng. Kết quả là, các biến số này của các thành phần sẽ bù trù lẫn nhau, và dẫn đến việc giá trị của mẫu sẽ gần với giá trị của dân số.

Tuy nhiên, để hiệu quả bù trừ này xảy ra, mẫu của chúng ta phải có đủ số lượng các thành phần, và chúng phải được rút ra từ dân số một cách đúng đắn để không gây ra sự thiên lệch.

Một mẫu đúng (không thiên lệch) là một mẫu mà các sai số được đánh giá quá cao hay quá thấp bù trừ lẫn nhau. Và do đó, phương sai hệ thống (**Systematic variance)** được định nghĩa như là "biến động trong đo lường do các ảnh hưởng biết được hay không biết được gây ra làm cho các điểm số bị thiên lệch về một phía nào đó".

Tăng cỡ mẫu (sample size) có thể làm giảm được phương sai hệ thống như là một nguồn sai số. Tuy nhiên, dù có tăng cỡ mẫu thì phương sai hệ thống vẫn có thể xảy ra nếu khung mẫu mà ta dựa vào để rút mẫu đã bị thiên lệch.

**Tính chính xác (Precision)**

Tiêu chuẩn thứ hai để thiết kế một mẫu tốt là tính chính xác của các ước lượng. Các nhà nghiên cứu đồng ý với nhau là không có mẫu nào có thể đại diện một cách đầy đủ dân số của nó ở mọi phương diện, mọi khía cạnh. Tuy nhiên, để diễn giải các phát hiện của nghiên cứu, chúng ta cần phải đo lường coi mẫu thể hiện được dân số chính xác tới mức nào. Các chỉ tiêu dạng số mô tả mẫu có thể khác với dân số do sai số ngẫu nhiên sinh ra trong quá trình chọn mẫu. Sai số này được gọi là sai số chọn mẫu (sampling error) hay là sai số chọn mẫu ngẫu nhiên (random sampling error), và nó phản ảnh ảnh hưởng của cơ hội rút ra các thành viên của mẫu.

Tính chính xác được đo lường bằng chỉ số sai số chuẩn của ước lượng, một dạng độ lệch chuẩn. Sai số chuẩn càng nhỏ có nghĩa là độ chính xác càng cao, và ngược lại. Một thiết kế chọn mẫu được coi là lý tưởng khi nó tạo ra sai số chuẩn của ước lượng nhỏ. Tuy nhiên, không phải là tất cả các kiểu thiết kế mẫu đều tạo ra các ước lượng cho mức độ chính xác, và các mẫu có cỡ mẫu bằng nhau có thể sinh ra các mức độ sai số khác nhau.

## 1.3 Các kiểu thiết kế mẫu (Types of Sample Design)

Khi thiết kế mẫu (hay là chọn lựa các chọn mẫu), các nhà nghiên cứu phải trả lời nhiều vấn đề. (Hình 6.1). Quá trình ra quyết định chọn mẫu phụ thuộc vào nhiều yếu tố. Có thể kế đến như bản chất của câu hỏi quản lý và các câu hỏi điều tra cụ thể được rút ra từ các câu hỏi nghiên cứu. Ngoài ra, các yếu tố khác ảnh hưởng đến thiết kế mẫu còn là các yêu cầu của dự án nghiên cứu và mục tiêu của nó, mức độ rủi ro mà các nhà nghiên cứu chấp nhận, ngân sách nghiên cứu, quỹ thời gian, các nguồn lực có thể có và văn hóa.

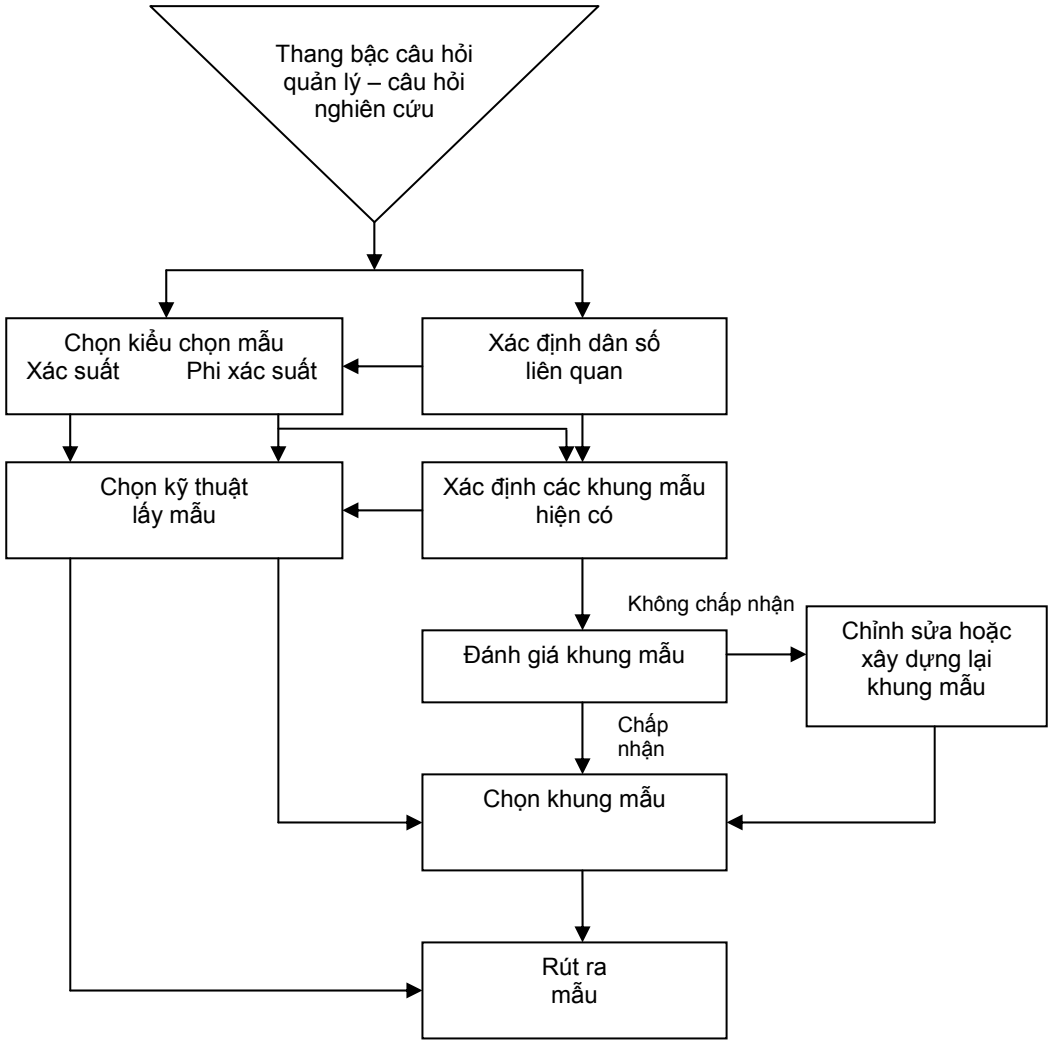
**Tính đại diện (Representation)**

Các thành phần trong một mẫu được chọn ra theo một trong hai qui trình: xác suất hay phi xác suất.

**Chọn mẫu phi xác suất (Nonprobability sampling)** có tính chất là tùy ý và có mục tiêu. Khi chúng ta chọn mẫu có mục tiêu, chúng ta thường chọn mẫu theo một kế hoạch định trước, và mỗi đơn vị nghiên cứu được rút ra từ dân số không có cơ hội được chọn ngang bằng nhau.

Sự khác biệt căn bản giữa chọn mẫu phi xác suất và chọn mẫu xác suất là tính chất "ngẫu nhiên". Chọn mẫu xác suất (**Probability sampling)** dựa trên các thành phần được chọn với cơ hội lựa chọn cho trước khác không. Chọn mẫu xác suất cho phép chúng ta xác định được các ước lượng về mức chính xác, và cho chúng ta cơ hội để tổng quát hóa các phát hiện cho các dân số nghiên cứu dựa trên dân số mẫu. Trong khi các nghiên cứu khám phá không đòi hỏi nhiều về việc này, nhưng các nghiên cứu giải thích, mô tả và nhân quả lại đòi hỏi điều này.

**Hình 6.1 Thiết kế chọn mẫu trong phạm vi quá trình nghiên cứu**



**Bảng 6.1 Các kiểu thiết kế mẫu**

| Chọn thành phần | Tính đại diện | |
|---|---|---|
| | Chọn mẫu xác suất | Chọn mẫu phi xác suất |
| Không hạn chế | Ngẫu nhiên đơn giản (Simple Random) | Thuận tiện (Convenience) |
| Hạn chế | Ngẫu nhiên phức tạp (Complex Random) | Có mục đích (Purposive) |
| | • Hệ thống (Systematic) | • Theo kinh nghiệm (Judgment) |
| | • Theo nhóm (Clustering) | • Hạn ngạch (Quota) |
| | • Phân tầng (Stratified) | Quả cầu tuyết (Snowball) |
| | Nhiều giai đoạn (Double) | |

**Chọn thành phần**

Các thành phần của mẫu được chọn theo từng cá thể và trực tiếp từ dân số. Cách chọn thành phần có hai loại cụ thể là không hạn chế và hạn chế. Chọn thành phần không hạn chế là cách thức mà các thành phần được rút ra theo từng cá thể từ dân số lớn. Cách chọn thành phần có hạn chế là các hình thức chọn mẫu còn lại (Hình 6.2).

## 2. CÁC BƯỚC THIẾT KẾ CHỌN MẪU

Khi lựa chọn cách chọn mẫu phù hợp nhất cho nghiên cứu, chúng ta phải trả lời một số câu hỏi đặt ra. Các câu hỏi này cũng chính là các nguyên tắc, hay là các bước mà chúng ta phải theo. Mỗi một câu hỏi đòi hỏi một thông tin duy nhất.

Các câu hỏi đi theo một trình tự nhất định. Tuy nhiên, để trả lời tốt một câu hỏi, ta phải xem xét lại câu hỏi và câu trả lời trước đó.

1. Dân số mục tiêu là gì?
2. Các chỉ tiêu (parameters) cần quan tâm là gì?
3. Khung mẫu là gì ?
4. Phương pháp chọn mẫu phù hợp là gì?
5. Cần cỡ mẫu bao nhiêu?

### 2.1 Dân số mục tiêu là gì?

Thông thường, khi chúng ta xác định vấn đề nghiên cứu và đặt ra câu hỏi nghiên cứu thì chúng ta đã dân số mục tiêu là gì rồi. Tuy nhiên, cũng có khi chúng ta vẫn chưa rõ ràng về dân số mục tiêu. Nếu chúng ta không biết rõ mục tiêu thì rất khó chọn mẫu phù hợp.

Ngoài ra, chúng ta vẫn có thể nhầm lẫn hoặc không biết chắc chắn là dân số bao gồm các cá nhân, hộ gia đình, gia đình hoặc là kết hợp các loại này.

Rõ ràng là đối với một nghiên cứu kinh tế thì việc xác định thành phần là một cá nhân hay là một hộ gia đình sẽ đưa đến các kết quả hoàn toàn khác nhau. Vì vậy, cần chú ý là ta ph ải xác định rõ khung phân tích, và khung hành động để chọn lựa đúng dân số liên quan.

### 2.2 Các chỉ tiêu (parameters) cần quan tâm là gì?

**Các chỉ số thể hiện cho dân số** là các chỉ tiêu mô tả tổng hợp (ví dụ giá trị trung bình, phương sai, v.v.) các biến số của dân số mà chúng ta quan tâm.

**Các chỉ số thống kê mẫu (Sample statistics)** là các chỉ tiêu mô tả cùng các biến số trên, nhưng không phải của dân số mà là của mẫu. Các chỉ số thống kê mẫu được dùng để ước lượng các chỉ số thống kê của dân số. Các chỉ số thống kê mẫu chính là cơ sở để chúng ta tham chiếu cho các chỉ số thống kê của dân số.

Tùy thuộc vào cách mà chúng ta đặt ra câu hỏi đo lường như thế nào (xem lại Hình 2.1, Chương 2), mỗi câu lại có thể thu thập dữ liệu ở các mức độ khác nhau. Mỗi mức độ khác biệt của dữ liệu lại sinh ra sự khác biệt về thống kê mẫu. Vì vậy, việc chọn lựa các chỉ tiêu cần quan sát sẽ thực tế quyết định kiểu chọn mẫu và cỡ mẫu.

Khi các biến số được đo lường với kiểu dữ liệu khoảng cách hay tỷ số (xem Chương 5), chúng ta sẽ sử dụng giá trị trung bình mẫu để ước lượng trung bình dân số, và độ lệch chuẩn của mẫu để ước lượng độ lệch chuẩn của dân số.

Khi các biến số được đo lường ở dạng thang đo danh nghĩa hoặc thứ bậc, chúng ta sẽ sử dụng các tỷ lệ của mẫu để ước lượng các tỷ lệ của dân số, và dùng chỉ số *pq* để ước lượng phương sai của dân số.

Trong trường hợp này, tỷ lệ của dân số sẽ bằng số lượng thành phần có trong dân số thuộc về một loại nào đó chia cho tổng số thành phần của dân số. Các đo lường tỷ lệ như thế này rất cần thiết cho dữ liệu danh nghĩa và được sử dụng rộng rãi cho các đo lường khác nữa.

Chỉ tiêu tỷ lệ thường dùng nhất chính là tỷ lệ phần trăm.

### 2.3 Khung mẫu là gì?

Khung mẫu có liên quan rất gần với dân số. Đó chính là danh sách của tất cả các thành phần có trong dân số mà từ đó chúng ta sẽ rút mẫu ra. Một khung mẫu lý tưởng chính là một danh sách hoàn thiện, đầy đủ và đúng tất cả các thành viên của dân số.

Tuy nhiên, trên thực tế, khung mẫu thường rất khác biệt với dân số lý thuyết.

Thường là chúng ta chấp nhận một khung mẫu bao gồm cả các người hoặc các trường hợp mà chúng ta không quan tâm. Nhưng chúng ta có thể giải quyết vấn đề này dễ dàng bằng cách rút một mẫu từ một dân số lớn hơn, và rồi sử dụng một quy trình lọc để loại bỏ các trường hợp mà chúng ta không quan tâm, hoặc không phải là thành viên của nhóm mà chúng ta muốn nghiên cứu.

### 2.4 Phương pháp chọn mẫu phù hợp là gì?

Nhà nghiên cứu phải đối mặt với một lựa chọn căn bản: chọn mẫu xác xuất hay phi xác suất. Với cách chọn mẫu xác suất, nhà nghiên cứu có thể đạt được các ước lượng cho nhiều chỉ tiêu nghiên cứu khác nhau dựa trên sự tin cậy về xác suất. Trong khi đó, chọn mẫu phi xác suất không cho được điều này.

Tuy nhiên, chọn mẫu xác suất có một vài hệ quả. Nhà nghiên cứu buộc phải theo các quy trình phù hợp mà:

- Phỏng vấn viên, điều tra viên không thể chỉnh sửa sự chọn lựa đã có.
- Chỉ có các thành phần được chọn từ khung mẫu gốc mới được tính tới.
- Không được thay thế thành phần này bằng thành phần khác ngoại trừ khi có các chỉ dẫn cụ thể theo các nguyên tắc định trước.

### 2.5 Cần cỡ mẫu bao nhiêu là vừa?

Cỡ mẫu chính là số đơn vị nghiên cứu mà ta cần có trong một mẫu khi rút ra từ dân số mục tiêu. Có nhiều quan niệm không chính xác về cỡ mẫu. Thứ nhất là một mẫu phải đủ lớn, nếu không nó sẽ không đại diện cho dân số. Thứ hai là một mẫu phải tương ứng với một tỷ lệ nào đó so với kích cỡ của dân số mà nó được rút ra.

Trên thực tế, cả hai câu chuyện này đều không chính xác.

Với mẫu phi xác suất, các nhà nghiên cứu khẳng định là số lượng nhóm phụ, các nguyên tắc lựa chọn và hạn chế về ngân sách là các yếu tố quyết định cỡ mẫu. Với cách chọn mẫu xác suất, cỡ mẫu phụ thuộc vào sự biến thiên của các chỉ số thống kê của dân số và mức độ chính xác của kết quả mà ta muốn có.

Một số nguyên tắc ảnh hưởng đến việc xác định cỡ mẫu là:

- Dân số càng biến thiên nhiều thì cỡ mẫu phải càng lớn để đạt tính chính xác;

- Độ chính xác mong muốn càng tăng thì cỡ mẫu phải càng lớn.
- Phạm vi sai số càng nhỏ thì cỡ mẫu phải càng lớn.
- Mức độ tin cậy của ước lượng càng cao thì cỡ mẫu càng phải lớn.
- Khi dân số có nhiều nhóm phụ, thì cỡ mẫu phải lớn để cỡ mẫu của từng nhóm phụ phải đạt yêu cầu tối thiểu

Các hạn chế về ngân sách cũng ảnh hưởng đến cỡ mẫu, cách chọn mẫu và phương pháp thu thập dữ liệu. Hầu hết các nghiên cứu đều bị giới hạn ngân sách, và điều này thúc đẩy các nhà nghiên cứu áp dụng các phương pháp chọn mẫu phi xác suất.

## 3. CHỌN MẪU XÁC SUẤT

### 3.1 Chọn mẫu xác suất ngẫu nhiên đơn giản (Simple Random Sampling)

Là một phương pháp chọn mẫu không hạn chế, phương pháp chọn mẫu xác suất ngẫu nhiên đơn giản là hình thức đơn giản nhất, thuần nhất của cách chọn mẫu xác suất. Khi mà tất cả các mẫu xác suất đều phải chọn lựa từng cá thể (đơn vị nghiên cứu) với một xác suất khác không cho trước thì phương pháp chọn mẫu ngẫu nhiên đơn giản được coi là một trường hợp đặc biệt vì mỗi một cá thể đều được lựa chọn với một xác suất biết trước và hoàn toàn ngang bằng nhau.

Xác suất chọn lựa = cỡ mẫu ÷ kích cỡ của dân số (%)

Để thực hiện chọn mẫu ngẫu nhiên đơn giản, việc đầu tiên là chúng ta phải có khung mẫu, hay chính là danh sách tất cả các cá thể (thành viên) của dân số mục tiêu. Dựa trên danh sách này, chúng ta sẽ đánh số và sử dụng bảng ngẫu nhiên để chọn lựa ra các cá thể (rút mẫu) để bảo đảm mọi cá thể đều có xác suất được chọn như nhau.

### 3.2 Chọn mẫu xác suất phức tạp (Complex Probability Sampling)

Một mẫu được coi là có hiệu quả hơn về phương diện thống kê là một mẫu mà nó có thể cho kích cỡ mẫu nhỏ hơn với một mức độ chính xác cho trước (sai số chuẩn của trung bình hoặc của tỷ lệ). Một mẫu được coi là có hiệu quả về phương diện kinh tế là một mẫu có thể đạt được một mức độ chính xác cho trước với chi phí thấp.

Ở các phần dưới đây, chúng ta sẽ thảo luận bốn cách thức chọn mẫu xác suất có khả năng thay thế nhau là: (1) chọn mẫu hệ thống (systematic sampling); (2) chọn mẫu phân tầng (stratified sampling); (3) chọn mẫu theo nhóm hoặc phân tổ (cluster sampling); và (4) chọn mẫu nhiều giai đoạn.

**a. Chọn mẫu hệ thống (Systematic sampling)**

Theo cách tiếp cận này, ta chọn lấy các thành phần thứ $k^{th}$ trong dân số, bắt đầu với một con số khởi điểm ngẫu nhiên trong phạm vi từ 1 đến k. Thành phần thứ $k^{th}$ , còn gọi là bước nhảy (***skip interval***), được tính bằng cách chia cỡ mẫu cho kích cỡ của dân số.

K = bước nhảy = dân số ÷ cỡ mẫu

Chúng ta cũng phải có khung mẫu chính xác và hoàn thiện.

Thủ tục để tiến hành chọn mẫu hệ thống theo các bước sau:

- Xác định, lập danh sách và đánh số các cá thể của dân số
- Xác định bước nhảy (k)
- Xác định con số khởi đầu một cách ngẫu nhiên
- Rút mẫu bằng cách chọn tất cả các cá thể theo các bước nhảy $k^{th}$.

Phương pháp chọn mẫu hệ thống có ưu điểm là đơn giản và mềm dẻo. Tuy vậy, phương pháp này cũng có thể sinh ra các thiên lệch khó thấy. Đầu tiên là tính chất chu kỳ của dân số có thể xảy ra song song với tỷ lệ mẫu (bước nhảy). Ngoài ra, các cá thể của dân số có thể đã được sắp xếp theo một trật tự đơn chiều nào đó. Trong nghiên cứu kinh tế, các dân số thường được sắp xếp theo trật tự sẵn có. Ví dụ, chúng ta có danh sách các cá nhân, hoặc hộ gia đình sắp xếp từ nghèo đến giàu, hoặc ngược lại; hoặc danh sách các hộ nông nghiệp sắp xếp theo quy mô tăng dần về diện tích đất canh tác, v.v. Chính vì vậy, khi chọn cá thể, chúng ta có thể bị thiên lệch về một phía nào đó của dãy số liệu.

Để tránh tình trạng thiên lệch như vậy, chúng ta nên:

- Sắp xếp ngẫu nhiên dân số trước khi chọn mẫu
- Chọn con số khởi điểm một cách ngẫu nhiên vài lần khi bắt đầu chọn mẫu
- Lặp lại cách chọn mẫu như vậy cho các mẫu khác.

Nếu thực hiện tốt, phương pháp này cho hiệu quả thống kê cao hơn phương pháp ngẫu nhiên đơn giản.

**b. Chọn mẫu phân tầng (Stratified Sampling)**

Hầu hết các dân số đều bao gồm các nhóm cá thể khác nhau. Các nhóm như vậy chính là các nhóm dân số phụ (subpopulation), hay là các tầng (strata). Quá trình chọn mẫu mà các cá thể được chọn lựa theo từng nhóm như vậy được gọi là chọn mẫu ngẫu nhiên phân tầng (**stratified random sampling)**. Phương pháp chọn mẫu phân tầng có hiệu quả thống kê cao hơn phương pháp chọn mẫu ngẫu nhiên đơn giản.

Tại sao chúng ta chọn phương pháp chọn mẫu phân tầng này? Phương pháp này cho chúng ta nhiều lợi ích như:

(1) Tăng hiệu quả thống kê của mẫu;

(2) Cung cấp dữ liệu phù hợp để phân tích từng nhóm dân số phụ hay từng tầng, và

(3) Cho phép sử dụng các phương pháp nghiên cứu và phân tích khác nhau cho cá nhóm dân số phụ khác nhau.

Nếu phân tầng một cách lý tưởng, ta sẽ có sự đồng nhất trong nội bộ từng nhóm và có sự dị biệt giữa các nhóm. Nếu phân tầng càng nhiều thì ta càng có thể tối đa hóa sự khác biệt giữa các nhóm và tối thiểu hóa sự biến thiên trong nội bộ từng nhóm.

Tuy nhiên, chi phí cũng là một yếu tố đáng quan tâm. Nếu tăng số nhóm nghiên cứu lên (số tầng) thì chi phí cũng tăng theo vì chi phí đi đôi với mức độ chọn mẫu chi tiết.

Ngoài ra, cũng phải chú ý đến các yếu tố sau: (1) kích cỡ tổng mẫu cần có và (2) tổng mẫu được phân bổ như thế nào giữa các tầng. Hai vấn đề này quan trọng vì chúng quyết định số cá thể cần có ở từng tầng.

Ví dụ, ta có hạn chế ngân sách nên chỉ có thể chọn cỡ mẫu tối đa là 250. Nếu ta chọn cách chia dân số làm 5 nhóm dân số phụ khác nhau, với tỷ lệ tương đương nhau, thì số lượng cá thể cần quan sát của mỗi mẫu phụ (tương ứng với mỗi nhóm dân số phụ, hay là từng tầng) là 50, tương đương với tỷ lệ 20% tổng mẫu. Số lượng 50 cá thể này có thể bảo đảm mức độ tin cậy về phân tích thống kê. Nhưng nếu chúng ta muốn chia dân số làm 10 nhóm dân số phụ, thì kích cỡ của mẫu phụ chỉ là 25. Số lượng đơn vị nghiên cứu có trong 1 mẫu phụ này có thể không bảo đảm tin cậy về phân tích thống kê.

Đối với cách phân bố mẫu cho các nhóm phụ (tầng) khác nhau, có hai cách là theo tỷ lệ (proportionate) và không theo tỷ lệ (disproportionate).

Đối với cách chọn mẫu phân tầng theo tỷ lệ (**proportionate stratified sampling**), cỡ mẫu của mỗi mẫu phụ (tầng) theo đúng tỷ lệ của các thành phần có trong từng dân số phụ so với tổng dân số. Cách chọn mẫu phân tầng theo tỷ lệ phổ biến nhiều hơn bất kỳ cách chọn mẫu phân tầng nào khác, bởi vì:

- có hiệu quả thống kê cao hơn phương pháp ngẫu nhiên đơn giản
- dễ thực hiện hơn các phương pháp phân tầng khác
- cung cấp một mẫu tự định trọng số (self-weighting sample); giá trị trung bình tổng thể hoặc tỷ lệ tổng thể có thể được ước lượng một cách dễ dàng.

Quy trình chọn mẫu phân tầng bao gồm các bước sau đây:

- *Quyết định các biến số dùng để phân tầng*. Trong nghiên cứu kinh tế - xã hội, các biến danh nghĩa thường được dùng để phân chia dân số thành các dân số

phụ. Thông thường là các biến nhân khẩu học (ví dụ độ tuổi, giới tính, nghề nghiệp, học vấn, v.v) hoặc các biến thể hiện sự khác biệt về vị thế kinh tế (ví dụ nghèo, cận nghèo, trung bình, khá, giàu).

- *Xác định tỷ lệ của từng nhóm dân số phụ so với dân số chung*. Để làm được việc này, rõ ràng là chúng ta phải có được khung mẫu của dân số tổng thể, và các khung mẫu của các dân số phụ dựa trên các biến danh nghĩa mà chúng ta dùng để phân chia.
- *Chọn lựa cách phân tầng theo tỷ lệ hoặc không theo tỷ lệ* tùy theo nhu cầu thông tin nghiên cứu và các rủi ro có thể xảy ra.
- *Thiết lập các khung mẫu của các dân số phụ*. Mỗi khung mẫu (phụ) thể hiện một tầng (nhóm dân số phụ).
- *Trộn ngẫu nhiên các thành phần* (cá thể, đơn vị nghiên cứu) trong từng khung mẫu của từng tầng.
- *Rút mẫu* cho các tầng bằng cách rút mẫu ngẫu nhiên hoặc hệ thống.

**c. Chọn mẫu theo nhóm (Cluster Sampling)**

Trong một mẫu ngẫu nhiên, mỗi thành phần của dân số được chọn lựa theo từng cá thể. Dân số cũng có thể được chia thành nhiều nhóm chứa đựng các thành phần cá thể mà có thể, một số nhóm như vậy được chọn ngẫu nhiên cho nghiên cứu. Đó chính là nguyên tắc của phương pháp chọn mẫu theo nhóm.

Ta có thể hình dung sự khác biệt giữa chọn mẫu phân tầng và chọn mẫu theo nhóm như sau. Giả sử hai dân số 1 và 2 đều chứa đựng các cá thể khác biệt, nhưng có thể chia làm ba nhóm chính, thể hiện bằng các ký tự x, Δ và o (Hình 6.3).

| | | |
|---|---|---|
| x x x | Δ Δ Δ | o o o |
| x x x | Δ Δ Δ | o o o |
| x x x | Δ Δ Δ | o o o |
| x x x | Δ Δ Δ | o o o |
| x x x | Δ Δ Δ | o o o |
| x x x | Δ Δ Δ | o o o |
| x x x | Δ Δ Δ | o o o |

Dân số 1 được chia thành các nhóm dân số phụ (tầng) dị biệt nhau dựa trên các đặc tính riêng biệt của các cá thể (chia nhóm x, Δ và o riêng biệt).

| | | |
|---|---|---|
| x x x | Δ Δ Δ | o o o |
| Δ Δ Δ | x x x | x x x |
| o o o | o o o | Δ Δ Δ |
| x x x | Δ Δ Δ | o o o |
| Δ Δ Δ | x x x | x x x |
| o o o | o o o | Δ Δ Δ |
| x x x | Δ Δ Δ | o o o |

Dân số 2 được chia thành các nhóm dân số phụ (nhóm) bao gồm các cá thể có tính đa dạng như nhau (x, Δ và o đều có mặt trong từng nhóm).

**Hình 6.2 Minh họa sự khác biệt giữa chọn mẫu phân tầng và theo nhóm**

Dựa trên tính chất khác biệt này, chúng ta có thể chọn mẫu theo hai cách khác biệt nhau. Cách thứ nhất là chúng ta chia dân số thành 3 nhóm dân số phụ theo các đặc tính x, Δ và o. Điều này cho phép chúng ta có 3 nhóm dân số phụ (tầng - stratum) bảo đảm sự đồng nhất trong nội bộ từng nhóm và có sự dị biệt giữa các nhóm. Ngược lại, chúng ta cũng có thể chia dân số thành 3 nhóm dân số phụ mà mỗi nhóm đều có các thành phần cá thể đa dạng với các đặc tính x, Δ và o. Kết quả là, ta có 3 nhóm dân số phụ (clusters) và có thể bảo đảm sự đa dạng hay dị biệt trong nội bộ từng nhóm và có sự đồng nhất giữa các nhóm.

Cách thứ nhất chính là chọn mẫu phân tầng. Cách thứ hai là chọn mẫu theo nhóm.

Hiệu quả thống kê của chọn mẫu theo nhóm thường thấp hơn chọn mẫu ngẫu nhiên đơn giản vì thông thường, các nhóm lại không có sự khác biệt cần thiết, mà lại có sự đồng nhất.

**Bảng 6.2 So sánh hai phương pháp chọn mẫu phân tầng và theo nhóm**

| **Chọn mẫu phân tầng - Stratified Sampling** | **Chọn mẫu theo nhóm - Cluster Sampling** |
|---|---|
| 1. Ta chia dân số thành *một số ít* nhóm phụ<br>- Mỗi nhóm phụ chứa *rất nhiều* thành phần.<br>- Các nhóm phụ được chọn lựa theo các tiêu chí liên quan đến các biến số nghiên cứu. | 1. Ta chia dân số thành *nhiều* nhóm phụ<br>- Mỗi nhóm phụ chứa *rất ít* thành phần.<br>- Các nhóm phụ được chọn lựa theo các tiêu chí để dàng hoặc có tính sẵn có để thu thập dữ liệu dễ hơn. |
| 2. Ta cố gắng bảo đảm *tính đồng nhất* (*homogeneity)* trong nội bộ từng nhóm phụ. | 2. Ta cố gắng bảo đảm *tính dị biệt* (*heterogeneity)* trong nội bộ từng nhóm phụ. |
| 3. Ta cố gắng bảo đảm *tính dị biệt* (*heterogeneity)* giữa các nhóm phụ. | 3. Ta cố gắng bảo đảm *tính đồng nhất* (*homogeneity)* giữa các nhóm phụ. |
| 4. Ta chọn lựa ngẫu nhiên các thành phần trong từng nhóm phụ. | 4. Ta chọn lựa ngẫu nhiên một số nhóm phụ để chúng ta nghiên cứu sâu. |

## Chọn mẫu theo vùng (Area Sampling)

Hầu hết các nghiên cứu kinh tế đều liên quan đến các dân số mà chúng có thể chia theo các vùng địa lý. Ví dụ khi nghiên cứu tình trạng nghèo đói, ta có thể thấy ở bất kỳ quốc gia nào (trên thế giới) hoặc ở bất kỳ vùng, miền, tỉnh nào (trong phạm vi một quốc gia) đều có người nghèo, giàu khác nhau. Như vậy, khi nghiên cứu, ta có thể chọn lựa một vài vùng miền nào đó thuận tiện cho nghiên cứu, và khi nghiên cứu ở các vùng như trên, ta vẫn bảo đảm có được các cá thể giàu, nghèo khác biệt nhau.

Khi ta có thể chia dân số theo vùng địa lý như vậy thì rõ ràng ta có thể sử dụng phương pháp chọn mẫu theo nhóm. Cách thức chọn mẫu như vậy còn được gọi là chọn

mẫu theo vùng, và có thể áp dụng ở mức độ quốc gia, vùng miền, thậm chí các đơn vị theo địa giới hành chính ở quy mô nhỏ hơn.

**Thiết kế chọn mẫu theo nhóm**

Khi chọn mẫu theo nhóm, kể cả chọn mẫu theo vùng, chúng ta cần trả lời các câu hỏi sau đây:

1. Các nhóm đồng nhất với nhau như thế nào?
2. Chúng ta tìm các nhóm có kích cỡ bằng nhau hay khác nhau?
3. Chúng ta sẽ chọn nhóm có kích cỡ bao nhiêu?
4. Chúng ta sẽ áp dụng phân nhóm một giai đoạn (single-stage cluster) hay nhiều giai đoạn (multistage cluster)?
5. Kích cỡ của mẫu bao nhiêu là vừa?

**d. Chọn mẫu nhiều giai đoạn (Double Sampling - sequential sampling - multiphase sampling)**

Trong nghiên cứu thực tế, người ta thường áp dụng phương chọn mẫu nhiều giai đoạn. Phương pháp này cho phép chúng ta sử dụng các thông tin có được từ các cuộc nghiên cứu ban đầu để làm cơ sở cho việc chọn mẫu ở các bước tiếp theo.

Trong nghiên cứu kinh tế, đôi khi chúng ta tiến hành nghiên cứu theo nhiều giai đoạn. Giai đoạn đầu tiên là nghiên cứu khám phá, là giai đoạn mà ta cần tìm hiểu các thông tin cơ bản của dân số mục tiêu thông qua mẫu. Dựa trên các thông tin cơ bản này, ta có thể hiểu về cấu trúc của dân số, và có thể phát hiện sự dị biệt cũng như tương đồng trong nội bộ dân số thông qua các chỉ tiêu thống kê ghi nhận được. Từ đó, chúng ta có thể tiếp tục rút ra các mẫu phụ từ mẫu mà chúng ta đã có để tiếp tục nghiên cứu ở các giai đoạn sau (nghiên cứu sâu).

Loại hình chọn mẫu nhiều giai đoạn thường được áp dụng trong nghiên cứu kinh tế - xã hội. Ở giai đoạn đầu, người ta thường chọn mẫu có cỡ mẫu lớn, thiết kế nội dung nghiên cứu đơn giản nhằm tìm hiểu các thông tin cơ bản của dân số mục tiêu. Sau đó, tùy theo mục tiêu nghiên cứu, người ta thiết kế các nghiên cứu sâu với các nội dung rất chi tiết, nhưng cần số đơn vị nghiên cứu ít hơn. Kết quả nghiên cứu trước cho phép rút ra các tiêu chí phân nhóm phù hợp cũng như bảo đảm khả năng rút các mẫu phụ chứa đựng các đơn vị nghiên cứu phù hợp từ mẫu đã nghiên cứu.

Thông thường, phương pháp chọn mẫu nhiều giai đoạn kết hợp nhiều phương pháp chọn mẫu khác nhau, ví dụ như chọn mẫu phân tầng, chọn mẫu theo nhóm, chọn mẫu hệ thống. Ví dụ sau đây sẽ minh họa rõ hơn.

| | |
|---|---|
| Nghiên cứu: | Thái độ và hành vi của người tiêu dùng đối với các nhãn hiệu xe máy Yamaha, Honda và Suzuki |
| Phạm vi nghiên cứu: | Việt Nam |
| Nghiên cứu sơ khởi: | Áp dụng phương pháp chọn mẫu theo vùng (Area sampling, một dạng của cluster sampling):<br>• chọn ra 1-2 hai thành phố đại diện cho các thành phố lớn ở Việt Nam<br>• chọn ra 7 tỉnh đại diện cho 7 vùng miền kinh tế ở Việt Nam<br>Ở mỗi tỉnh, thành phố, chọn người sử dụng xe máy để phỏng vấn sơ khởi. Có thể áp dụng các phương pháp phi xác suất như chọn mẫu thuận tiện (convenience sampling) hoặc xác suất như chọn mẫu ngẫu nhiên đơn giản hay chọn mẫu hệ thống dựa trên danh sách (khung mẫu) do cơ quan Công An hoặc hệ thống đại lý xe cung cấp. |
| Nghiên cứu sâu: | Dựa trên các thông tin ghi nhận được từ người tiêu dùng trong phỏng vấn sơ khởi và sự sẵn lòng của họ, tiến hành nghiên cứu sâu.<br>Áp dụng phương pháp chọn mẫu phân tầng (stratified sampling) theo tỷ lệ hoặc không theo tỷ lệ dựa trên các đặc điểm dị biệt về nhân khẩu học của người tiêu dùng, hoặc dựa trên nhãn hiệu xe mà họ sử dụng.<br>Rút mẫu từ mẫu nghiên cứu đã có. |

Với ví dụ trên, ta thấy nhà nghiên cứu có thể lựa chọn và áp dụng nhiều phương pháp chọn mẫu khác nhau cho các giai đoạn nghiên cứu khác nhau. Tất nhiên là các phương án chọn lựa còn tùy thuộc rất nhiều vào mục tiêu nghiên cứu, dân số mục tiêu, các chỉ tiêu cần thu thập, khả năng có được khung mẫu, sự dễ dàng, thuận tiện trong nghiên cứu, và khả năng tài chính đáp ứng cho nghiên cứu.

Như vậy, có nhiều phương pháp chọn mẫu xác suất khác nhau, với các ưu điểm và hạn chế của chúng. Bảng 6.3 giúp tóm tắt đặc điểm chính, ưu điểm và hạn chế của từng phương pháp.

**Bảng 6.3 So sánh các phương pháp chọn mẫu xác suất**

| Kiểu | Mô tả | Ưu điểm | Hạn chế |
|---|---|---|---|
| **Ngẫu nhiên đơn giản Simple Random**<br>Chi phí: Cao<br>Áp dụng: Trung bình | Mỗi thành phần của dân số đều có cơ hội được lự chọn ngang bằng nhau.<br>Mẫu được rút ra bằng cách sử dụng bảng số ngẫu nhiên hoặc phần mềm tạo bảng số ngẫu nhiên. | Dễ áp dụng, nhất là với cách phỏng vấn quan điện thoại do máy quay số ngẫu nhiên. Có thể áp dụng hệ thống trả lời tự động. | Đòi hỏi danh sách khung mẫu.<br>Tốn nhiều thời gian để thực hiện.<br>Cần cỡ mẫu lớn.<br>Tạo ra nhiều sai số. |
| **Hệ thống Systematic**<br>Chi phí: Trung bình<br>Áp dụng: Trung bình | Chọn ra một thành phần dân số khởi đầu một cách ngẫu nhiên, dùng bước nhảy $k^{th}$ để chọn các thành phần khác. | Thiết kế đơn giản.<br>Dễ áp dụng hơn chọn mẫu ngẫu nhiên đơn giản.<br>Dễ tính toán phân bố mẫu của giá trị trung bình hoặc tỷ lệ. | Tính chu kỳ của dân số có thể làm méo, sai lệch mẫu và kết quả.<br>Nếu dân số có xu hướng trật tự đơn chiều, có thể sinh ra kết quả thiên lệch. |
| **Phân tầng Stratified**<br>Chi phí: Cao<br>Áp dụng: Trung bình | Chia dân số thành các dân số phụ (tầng) và áp dụng chọn mẫu ngẫu nhiên đơn giản cho từng tầng. Kết quả có thể tính theo trọng số và kết hợp được. | Nhà nghiên cứu kiểm soát cỡ mẫu trong các tầng.<br>Tăng hiệu quả thống kê.<br>Cung cấp dữ liệu đại diện và phân tích nhóm phụ.<br>Cho phép sử dụng nhiều phương pháp phân tích khác nhau cho từng tầng. | Tăng sai số nếu các nhóm phụ được chọn ở các tỷ lệ khác nhau.<br>Đắt đỏ nếu phải tạo ra nhiều tầng khác nhau. |
| **Theo nhóm Cluster**<br>Chi phí: Trung bình<br>Áp dụng: Cao | Dân số được chia làm nhiều nhóm phụ dị biệt trong nội bộ. Chọn ngẫu nhiên một số nhóm để nghiên cứu sâu. | Cung cấp các ước lượng không thiên lệch nếu được thực hiện đúng cách.<br>Hiệu quả kinh tế cao hơn chọn mẫu ngẫu | Thường có hiệu quả thống kê thấp do các nhóm phụ có xu hướng đồng nhất hơn là dị biệt. |

| Kiểu | Mô tả | Ưu điểm | Hạn chế |
|---|---|---|---|
| | | nhiên đơn giản.<br>Chi phí thấp nhất, đặc biệt khi chia nhóm theo vùng địa lý.<br>Dễ làm, không cần danh sách khung mẫu. | |
| **Nhiều giai đoạn**<br>**(Double, sequential or multiphase)**<br>Chi phí: Trung bình<br>Áp dụng: Trung bình | Quá trình bao gồm việc thu thập dữ liệu từ một mẫu đã được xác định trước. Dựa trên các thông tin có được, chọn ra mẫu phụ cho các nghiên cứu tiếp theo. | Có thể làm giảm chi phí nếu kết quả giai đoạn đầu cho đầy đủ dữ liệu để phân tầng hoặc chia nhóm dân số. | Tăng chi phí nếu được áp dụng không phân biệt. |

## 4. CHỌN MẪU PHI XÁC SUẤT

Với cách tiếp cận có mục đích như chọn mẫu phi xác suất, ta sẽ không biết được xác suất lựa chọn các đơn vị nghiên cứu (thành phần của dân số). Có nhiều cách để chọn lựa các cá nhân hoặc các trường hợp cần có trong mẫu. Thường chúng ta hay cho phép các phỏng vấn viên lựa chọn người cần phỏng vấn. Khi điều này xảy ra, rõ ràng là các thiên lệch có thể phát sinh ra và làm méo mó kết quả nghiên cứu. Tuy nhiên, có những lý do thực tiễn mà người ta lựa chọn các phương pháp kém chính xác như vậy.

### 4.1 Các vấn đề thực tiễn

Chúng ta có thể sử dụng các thủ tục chọn mẫu phi xác suất vì các lý do sau:

- Chúng có thể thỏa yêu cầu chọn mẫu có mục tiêu.
- Nếu không có mong muốn hoặc không cần thiết phải tổng quát hóa các kết quả nghiên cứu cho dân số tổng thể thì ta không quan tâm lắm đến việc liệu là mẫu có đại diện đầy đủ cho dân số hay không. Điều này đúng với các nghiên cứu khám phá khi mà chúng ta có thể chỉ muốn gặp những cá nhân, những trường hợp không điển hình, không giống ai.
- Chọn mẫu phi xác suất ít tốn kém chi phí và thời gian so với chọn mẫu xác suất.

- Trong khi chọn mẫu xác suất có vẻ lý tưởng và rất tốt về lý thuyết, thì khi áp dụng vào thực tiễn, lại có nhiều thất bại. Ngay cả khi chúng ta áp dụng cẩn thận các bước chọn mẫu ngẫu nhiên đơn giản thì chất lượng nghiên cứu vẫn còn tùy thuộc vào mức độ áp dụng cẩn thận hay không cẩn thận của các người liên quan. Vì vậy, các phươjng pháp chọn mẫu xác suất lý tưởng lại chỉ có thể thành công một phần vì lỗi con người.
- Chọn mẫu phi xác suất có thể là cách thay thế duy nhất. Trong một số trường hợp, có thể ta không có dân số tổng thể cho nghiên cứu. Và vì vậy, ta không thể có khung mẫu hoặc có cơ sở để chọn mẫu xác suất.
- Theo một nghĩa khác, chính những người tham gia nghiên cứu (đối tượng nghiên cứu) có thể tự chọn chính mình để tham gia. Điều này cũng có nghĩa là nhà nghiên cứu không thể bảo đảm sự ngang bằng về cơ hội chọn lựa các dơn vị nghiên cứu.

### 4.2 Các phương pháp chọn mẫu phi xác suất

#### a. Chọn mẫu thuận tiện (Convenience)

Các mẫu phi xác suất không bị hạn chế được gọi là các mẫu "thuận tiện". Đây là các mẫu có mức tin cậy ít nhất, nhưng thường là rẻ nhất và dễ tiến hành nhất. Lý do chính là các nhà nghiên cứu hoặc các điều tra viên, có quyền tự do chọn lựa bất kỳ ai họ muốn, vì thế được gọi là "thuận tiện".

Trong khi chọn mẫu thuận tiện không có kiểm soát như thế có thể không bảo đảm tính chính xác, nhưng vẫn là một phương pháp hữu ích. Thường thì ta có thể áp dụng một mẫu như vậy để kiểm tra các ý tưởng hoặc để có được các ý tưởng về đối tượng nghiên cứu. Ở các giai đoạn đầu của nghiên cứu khám phá, khi ta tìm kiếm hướng đi, ta có thể áp dụng cách tiếp cận này. Các kết quả có thể rõ ràng đến mức không cần thiết phải áp dụng các phương pháp chọn mẫu phức tạp.

Các nghiên cứu thị trường thường sử dụng cách chọn mẫu thuận tiện này. Các cuộc thăm dò ý kiến khách hàng hầu hết được thực hiện một cách thuận tiện.

#### b. Chọn mẫu có mục đích (Purposive Sampling)

Chọn mẫu có mục đích là hình thức chọn mẫu phi xác suất mà nhà nghiên cứu muốn theo những tiêu chí nào đó. Có hai phương pháp chọn mẫu có mục đích là chọn mẫu theo kinh nghiệm (judgment sampling) và chọn mẫu theo hạn ngạch (quota sampling).

***Chọn mẫu theo kinh nghiệm*** xảy ra khi nhà nghiên cứu chọn các đơn vị nghiên cứu theo các tiêu chuẩn nào đó. Phương pháp này phù hợp khi được sử dụng vào các giai đoạn đầu của nghiên cứu khám phá. Khi ta muốn chọn một nhóm thiên lệch nào đó

nhằm mục tiêu thanh lọc dữ liệu thì chọn mẫu theo kinh nghiệm cũng là một phương pháp tốt.

Ví dụ, một công ty chọn nhân viên của chính họ để đánh giá những sản phẩm mới trước khi đưa ra thị trường. Nếu thất bại, thì các sản phẩm này khó có triển vọng đưa vào thị trường. Một trường hợp khác, ví dụ ta muốn nghiên cứu về thị trường xe ô tô gia đình ở Việt Nam. Dĩ nhiên là chúng ta phải chọn các đối tượng nghiên cứu là người ở tầng lớp trung lưu trở lên, và phải là người có kinh nghiệm sử dụng xe ô tô gia đình.

***Chọn mẫu theo hạn ngạch*** là kiểu chọn mẫu có mục đích thứ hai. Chúng ta áp dụng để cả thiện tính đại diện. Lý do chủ yếu là dân số có thể có vài chiều kích và chọn mẫu theo hạn ngạch có thể mô tả được các chiều kích này.

Trong chọn mẫu hạn ngạch, nhà nghiên cứu phải chỉ ra nhiều hơn một hướng kiểm soát. Mỗi hướng phải thỏa mãn hai điều kiện: (1) có một phân phối trong dân số để chúng ta có thể ước lượng và (2) thích hợp với chủ đề nghiên cứu. Để minh họa, ta quan sát các trường hợp sau:

*Giới tính: hai nhóm thuộc tính – nam, nữ.*

*Trình độ học vấn: hai nhóm thuộc tính: đại học – trung học.*

*Khoa ngành: sáu nhóm thuộc tính – nghệ thuật và khoa học, nông nghiệp, kiến trúc, kinh doanh, công nghệ, khác.*

*Tôn giáo: bốn nhóm thuộc tính - Phất giáo, Thiên chúa giáo, Tin lành, khác.*

*Thành viên hiệp hội: hai nhóm thuộc tính – thành viên, không phải thành viên.*

*Tầng lớp kinh tế - xã hội: ba nhóm thuộc tính: giàu, trung bình, nghèo.*

Tương tự như chọn mẫu phân tầng, chọn mẫu hạn ngạch có thể theo tỷ lệ hoặc không theo tỷ lệ.

Chọn mẫu hạn ngạch có vài hạn chế. Thứ nhất, không có gì bảo đảm mẫu sẽ đại diện cho các biến cần nghiên cứu. Thứ hai, việc chọn lựa đơn vị nghiên cứu tùy thuộc vào điều tra viên, và tùy thuộc vào kinh nghiệm của chính họ. Vì vậy, họ có thể chọn những người thân thiết, ban bè quen thuộc để dễ thực hiện công việc.

Tuy vậy, nhìn chung là chọn mẫu hạn ngạch có ít rủi ro về thiên lệch hệ thống, và thường thỏa mãn được các yêu cầu dự đoán nói chung.

**c. Chọn mẫu mở rộng (Snowball)**

Kiểu chọn mẫu này được áp dụng khi ta khó xác định các người trả lời và khó tiếp cận được. Cách này rất phù hợp cho cấc nghiên cứu định tính.

Ở giai đoạn đầu tiên, các cá nhân cần tìm hiểu sẽ được phát hiện bằng cách nào đó, có thể theo xác suất hoặc phi xác suất. Rồi sau đó các cá nhân này chỉ cho nhà nghiên cứu những người khác có các đặc điểm tương tự. Rồi cứ tiếp tục như thế, nhà nghiên cứu sẽ được các người trả lời chỉ cho những người khác và mở rộng mẫu nghiên cứu.

## 5. XÁC ĐỊNH CỠ MẪU

### 5.1 Các khái niện căn bản liên quan đến chọn mẫu và xác định cỡ mẫu

***Giá trị trung bình*** ($\bar{x}$) của mẫu rút ra từ một dân số cho trước là một giá trị ước lượng điểm và là thông số tốt nhất dùng để ước lượng giá trị trung bình chưa biết của dân số, $\mu$.

***Sai số chuẩn.*** Chúng ta không thể coi trung bình mẫu là trung bình dân số. Tuy nhiên, chúng ta có thể ước lượng khoảng tin cậy mà trung bình dân số $\mu$ rơi vào. Ta có thể áp dụng công thức tính sai số chuẩn (*standard error of the mean)* - $\sigma_{\bar{X}}$ hay là *se*.

$$\sigma X = \frac{\sigma}{\sqrt{n}}$$

với

$\sigma_{\bar{X}}$ = sai số chuẩn của giá trị trung bình hay là độ lệch chuẩn của tất cả giá trị trung bình $\bar{x}_s$ có thể có.

$\sigma$ = độ lệch chuẩn của dân số

n = cỡ mẫu

Độ lệch chuẩn của mẫu được sử dụng như là ước lượng không chệch cho độ lệch chuẩn của dân số.

$$\sigma x = \frac{s}{\sqrt{n}}$$

với

*s* = độ lệch chuẩn của mẫu *n*

*Giả sử ta có:*

$n_1$ = 10, $\bar{x}_1$ = 3,0 và $s_1$ = 1,15

$$\sigma x = \frac{s}{\sqrt{n}} = \frac{1.15}{\sqrt{10}} = 0{,}36$$

***Ước lượng giá trị trung bình của dân số***.

Giá trị trung bình của dân số, μ, có thể được ước lượng theo công thức sau:

$$\mu = \bar{x} \pm \sigma_{\bar{x}}$$

Bởi vì chúng ta không điều tra tổng thể nên ta chưa biết giá trị μ và σ. Tuy nhiên, ta có thể áp dụng công thức $\mu = \bar{x} \pm \sigma_{\bar{x}}$. Theo ví dụ trên, $\mu = \bar{x} \pm \sigma_{\bar{x}}. = 3,0 \pm 0,36$

Tuy nhiên, vì sai số chuẩn có tính chất như các thông số thống kê khác, ta chỉ có thể có mức tin cậy 68% về giá trị ước lượng này. Điều này có nghĩa là một sai số chuẩn chỉ chứa đựng ± 1Z hay là 68% diện tích dưới đường phân phối chuẩn.

Ta sẽ sử dụng chỉ số thống kê khoảng tin cậy (*confidence interval)*. Để tăng độ tin cậy lên 95%, ta phải nhân sai số chuẩn với ± 1,96 (Z), khi 1,96 (Z) bao phủ 95% diện tích dưới đường phân phối chuẩn. Tương tự như vậy, để nâng độ tin cậy lên 99%, ta phải nhân sai số chuẩn với ± 3,0 (Z), khi 3,0 (Z) bao phủ 99% diện tích dưới đường phân phối chuẩn.

Do đó, khoảng tin cậy của giá trị trung bình dân số, μ sẽ là:

Ở mức tin cậy 68%: 2,64 – 3,36 (μ = 3,0 ± 0,36)

Ở mức tin cậy 95%: 2,29 – 3,71 (μ = 3,0 ± 0,71)

Ở mức tin cậy 99%: 1,92 – 4,08 (μ = 3,0 ± 1,08)

**5.2 Xác định cỡ mẫu theo trung bình**

Trước khi tính cỡ mẫu mong muốn, chúng ta hãy coi lại các thông tin cần thiết:

1. Mức chính xác mong muốn và làm thế nào để lượng hóa nó:
    a. Mức tin cậy (*confidence level)* mà ta muốn đạt được.
    b. Độ lớn của khoảng tin cậy (*size of the interval estimate)*.
2. Độ biến thiên kỳ vọng của dân số.
3. Có cần thiết điều chỉnh dân số hữu hạn hay không.

***Mức chính xác***

Ta phải xác định rõ mức chính xác mong muốn. Thường thì mức chính xác 95% được áp dụng rộng rãi, tuy nhiên chúng ta vẫn có thể tăng hay giảm mức chính xác mong muốn tùy theo từng nghiên cứu cụ thể.

Tương tự như vậy, ta cũng cần xác định độ lớn của khoảng tin cậy nhằm tiên đoán các chỉ số của dân số dựa trên dữ liệu rút ra từ mẫu.

***Độ biến thiên của dân số***

Yếu tố kế tiếp ảnh hưởng đến cỡ mẫu với mức tin cậy cho trước là độ biến thiên của dân số. Độ biến thiên càng nhỏ thì cỡ mẫu ta cần cũng càng nhỏ. Ngược lại, độ biến thiên càng lớn thì cỡ mẫu sẽ phải càng lớn.

Tuy nhiên, không phải lúc nào ta cũng có các chỉ số thể hiện độ biến thiên của dân số (ví dụ phương sai, độ lệch chuẩn). Tuy nhiên, ta có thể biết được độ biến thiên của dân số nhờ vào:

- Sử dụng kết quả tính độ biến thiên từ các nghiên cứu trước đây trên cùng chủ đề.
- Tính phương sai dựa trên kết quả khảo sát thử nghiệm trên một mẫu rút ra từ dân số.
- Nguyên tắc: giả sử các quan sát tuân theo quy luật phân phối chuẩn, thì độ lêch chuẩn bằng khoảng 1/6 khoảng dao động của dữ liệu (tối thiểu - tối đa) với độ tin cậy 99.73%.

***Tính cỡ mẫu***

Cỡ mẫu được tính từ công thức:

$$\sigma x = \frac{s}{\sqrt{n}}$$

$$\sqrt{n} = \frac{s}{\sigma x}$$

$$n = \frac{s^2}{\sigma x^2}$$

Ví dụ: tính cỡ mẫu cho nghiên cứu thu nhập của sinh viên (đơn vị tính: triệu đồng/tháng)

- Chọn mức độ chính xác mong muốn:
    - Mức tin cậy (confidence level): 95% (Z=1,96)
    - Độ lớn của khoảng tin cậy ± 0,25 (tr.đồng/tháng) = Z*se, suy ra se = 0,25/Z
- Xác định độ biến thiên kỳ vọng trong dân số (expected dispersion in the population): dựa trên các kết quả nghiên cứu gần đây về thu nhập của sinh viên, ta có giá trị độ lệch chuẩn tham khảo = 0,7 (tr.đồng/tháng)
- Phỏng định sai số chuẩn: se = 0,25/Z = 0,25/1,96 = 0,127
- Xác định cỡ mẫu n = $s^2/\sigma x^2$ = $0,7^2/0,127^2$ = 30,38 = 30

Nếu ta muốn nâng mức độ chính xác mong muốn từ 95% lên 99%, thì Z thay đổi từ 1,96 đến 3,0. Áp dụng vào công thức tính ta có:

- Phỏng định sai số chuẩn: se = 0,25/3,0 = 0.083
- Cỡ mẫu n = $0,7^2/0,083^2$ = 71,02 = 71

Như vậy, khi tăng mức tin cậy từ 95% lên 99%, trong trường hợp này, chúng ta phải tăng cỡ mẫu lên 2,4 lần.

Nếu ta muốn giảm độ lớn của khoảng tin cậy xuống còn 0,1 triệu đồng/tháng thay vì 0,25 triệu đồng/tháng, và vẫn giữ mức tin cậy 95%. Áp dụng vào công thức tính, ta có:

- Phỏng định sai số chuẩn: se = 0,1/1,96 = 0.051
- Cỡ mẫu n = $0,7^2/0,051^2$ = 188,38 = 188

Như vậy, khi giảm độ lớn của khoảng tin cậy xuống 2,5 lần, cỡ mẫu phải tăng 6,3 lần trong trường hợp này.

**5.3 Xác định cỡ mẫu theo tỷ lệ**

Thay vì xác định giá trị trung bình của dân số, với cách xác định cỡ mẫu theo tỷ lệ, ta phải xác định tỷ lệ của dân số mà chúng có một thuộc tính cho trước, tỷ lệ này gọi là $p$. Và thay vì sử dụng độ lệch chuẩn, độ biến thiên của dân trong trường hợp này được xác định bằng $p$ x $q$, trong đó $q$ là tỷ lệ của dân số không có thuộc tính đó, tức là $q$ = (1 – $p$). Tương tự như vậy, sai số chuẩn của trung bình được thay thế bằng sai số chuẩn của tỷ lệ, $\sigma_p$.

- Giả sử rằng từ một khảo sát sơ khởi, ta biết $p$ = 30%.
- Ta quyết định ước lượng tỷ lệ thực đúng của dân số trong phạm vi sai số 10% ($p = 0.30 \pm 0.10$).
- Giả sử là ta muốn đạt mức tin cậy 95%.

Cách tính được thực hiện như sau:

| | | |
|---|---|---|
| $\pm 0.10$ | = | Khoảng tin cậy mong muốn mà ta kỳ vọng tỷ lệ dân số đạt được (quyết định mục tiêu) |
| $1.96\ \sigma_p$ | = | mức tin cậy 95% để ước lượng khoảng tin cậy mà ta kỳ vọng tỷ lệ dân số đạt được (quyết định mục tiêu) |
| $\sigma_p$ | = | 0.051 = sai số chuẩn của tỷ lệ (0.10/1.96) |
| $pq$ | = | Chỉ thị độ biến thiên của mẫu, được dùng như là một ước lượng của độ biến thiên của dân số |

$n$ = Cỡ mẫu

$$\sigma_p = \sqrt{\frac{pq}{n}}$$

$$n = \frac{pq}{\sigma^2_p}$$

Áp dụng công thức này với các giá trị trên, ta có:

$$n = \frac{0.3x0.7}{(0.051)^2} = 81$$

**Bảng 6.4 Tóm lược các bước xác định cỡ mẫu**

| | **Ví dụ** | |
|---|---|---|
| **Các bước xác định cỡ mẫu** | **Theo trung bình** | **Theo tỷ lệ** |
| 1. Độ chính xác mong muốn và làm sao để lượng hóa nó:<br>a. Mức tin cậy mong muốn<br>b. Độ lớn của khoảng tin cậy cần có | <br><br>95% (Z=1.96)<br>± 0.5 | <br><br>95% (Z=1.96)<br>± 0.10 (10 %) |
| 2. The accepted range in the population for the question used to measure precision: | 0 đến 30 | 0 đến 100% |
| **Các giá trị đo lường xu hướng trung tâm** | | |
| ▪ Trung bình mẫu | 10 | |
| ▪ Tỷ lệ mẫu của dân số có thuộc tính cho trước cần đo lường | | 30% |
| **Đo lường độ biến thiên** | | |
| ▪ Độ lệch chuẩn | 4.1 | |
| ▪ Độ biến thiên của dân số | | $pq$ = 0.30 x 0.70 = 0.21 |
| 3. Ước lượng sai số chuẩn của dân số | | |
| ▪ Sai số chuẩn của trung bình | 0.5/1.96 = 0.255 | |
| ▪ Sai số chuẩn của tỷ lệ | | 10/1.96 = 0.051 |
| 4. Cỡ mẫu | n = 259 | n = 81 |

**Từ khóa**

| | |
|---|---|
| Chọn mẫu | Sampling |
| Thành phần của dân số | Population element |
| Đơn vị nghiên cứu | Unit of study |
| Điều tra tổng thể | Census |
| Khung mẫu | Sample frame |
| Chọn mẫu phi xác suất | Nonprobability sampling |
| Chọn mẫu xác suất | Probability sampling |
| Chọn mẫu ngẫu nhiên đơn giản | Simple random sampling |
| Chọn mẫu ngẫu nhiên phức tạp | Complex random sampling |
| Chọn mẫu hệ thống | Systematic sampling |
| Chọn mẫu theo nhóm | Clustering sampling |
| Chọn mẫu phân tầng | Stratified sampling |
| Chọn mẫu nhiều giai đoạn | Double, sequential, multiphase sampling |
| Chọn mẫu thuận tiện | Convenience sampling |
| Chọn mẫu có mục đích | Purposive sampling |
| Chọn mẫu theo kinh nghiệm | Judgment sampling |
| Chọn mẫu hạn ngạch | Quota sampling |
| Chọn mẫu mở rộng | Snowball sampling |

# Chương 7. Nhập và xử lý dữ liệu

**Mục tiêu giảng dạy**

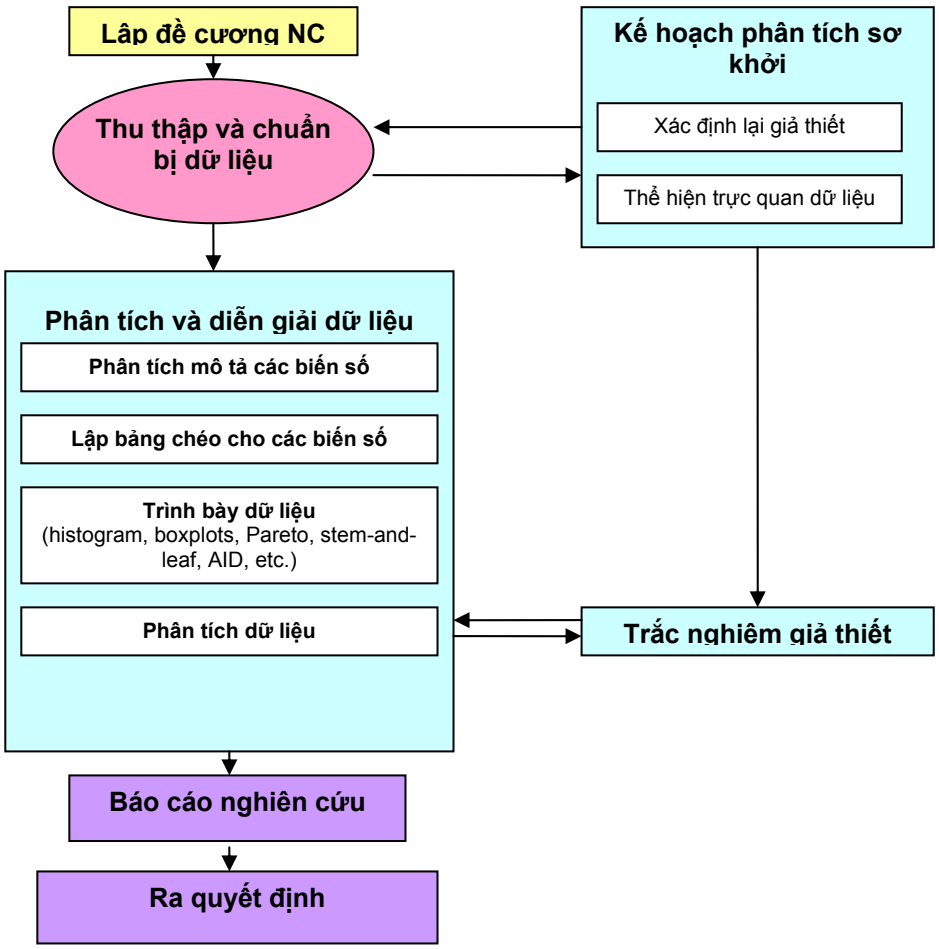
Nhằm hướng dẫn sinh viên cách:

1. Cách nhập liệu, xử lý và phân tích dữ liệu.
2. Các kỹ thuật phân tích dữ liệu mang tính khám phá (exploratory data analysis**)** cung cấp một sự thấu hiểu và chẩn đoán dữ liệu bằng cách nhấn mạnh việc trình bày trực quan các dữ liệu.
3. Cách sử dụng bảng chéo (cross-tabulation) để trắc nghiệm mối quan hệ giữa các biến phân loại (categorical variables), có vai trò như là một khung phân tích cho các trắc nghiệm thống kê sau này, và làm cho các phân tích dựa trên bảng số liệu sử dụng một hoặc nhiều biến khống chế (control variables) trở thành một công cụ thể hiện dữ liệu có hiệu quả.
4. Cách sử dụng các thống kê phân tích trắc nghiệm giả thiết.

## 1. PHÂN TÍCH KHÁM PHÁ DỮ LIỆU

Khi dùng phân tích khám phá dữ liệu - **exploratory data analysis (EDA)** ta có khả năng linh động đáp ứng lại các khuôn mẫu khác nhau của bước phân tích dữ liệu sơ khởi. Cách thức phân tích này cho phép xem xét và đánh giá lại kế hoạch phân tích dữ liệu. Tính mềm dẻo là một thuộc tính quan trọng của cách tiếp cận này.

**Phân tích xác nhận dữ liệu (Confirmatory data analysis)** là một quá trình phân tích theo hướng suy luận từ kết quả phân tích thống kê dựa trên trắc nghiệm ý nghĩa và độ tin cậy.




**Hình 5.1 Các bước khám phá, trắc nghiệm và phân tích trong quá trình nghiên cứu**

## 2. NHẬP SỐ LIỆU

### 2.1 Cách bố trí dữ liệu trên máy tính

a) Mục tiêu:

- Nhằm tạo điều kiện thuận tiện cho việc nhập liệu
- Nhằm tạo sự thuận lợi cho việc chỉnh sửa dữ liệu

b) Thực hiện:

- Nguyên tắc chung: đặt tên biến ngắn gọn, nên viết tắt (nên sử dụng tiếng Việt không dấu hoặc sử dụng tiếng Anh). Tên biến nên được đặt theo quy luật và trình tự của bảng câu hỏi hay nội dung khảo sát.

- Nếu lưu trữ bằng phần mềm Excel: ưu điểm là dễ thao tác và chỉnh sửa, nhược điểm là không gian lưu trữ hạn chế, công cụ thống kê và kinh tế lượng phát triển chưa đầy đủ cho nhu cầu phân tích.
- Nếu lưu trữ bằng phần mềm SPSS: ưu điểm là không gian lưu trữ gần như không hạn chế, công cụ thống kê và kinh tế lượng phát triển khá đầy đủ cho nhu cầu phân tích. Nhưng nhược điểm là đòi hỏi việc khai báo dữ liệu mất nhiều thời gian hơn.

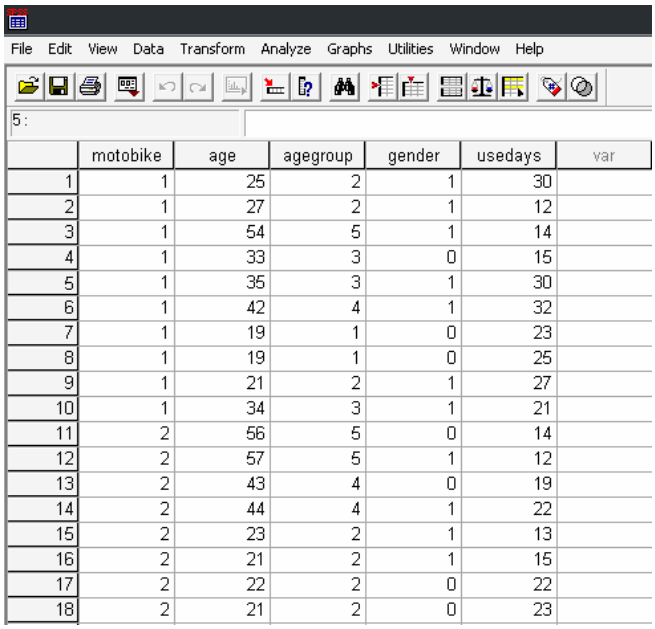
**2.2 Cách nhập liệu**

a) Đối với dữ liệu định lượng: nhập đúng giá trị trong bảng phỏng vấn, nên thao tác bằng các phím tại ô số trên bàn phím.

b) Đối với dữ liệu định tính

- Câu trả lời đóng:
    - Trường hợp câu hỏi có 1 câu trả lời hoặc chọn 1 trong 2 (ví dụ: có hoặc không, nam hay nữ): sử dụng giá trị 0 và 1 để lưu thông tin. Ví dụ: có là 1 , không là 0, nam là 1 , nữ là 0 hoặc ngược lại.
    - Trường hợp có từ 3 lựa chọn trở lên nhưng chỉ có 1 câu trả lời (ví dụ: không thích, thích và không ý kiến): sử dụng giá trị 1, 2 và 3 tương ứng theo câu trả lời.
    - Trường hợp có từ 3 lựa chọn trở lên và có ít nhất 2 câu trả lời (ví dụ: câu hỏi về sở thích: xem tivi, đọc báo và nghe radio): Tạo 3 biến, mỗi biến là một lựa chọn và sử dụng giá trị 0 và 1 để lưu thông tin, lựa chọn nào được đánh dấu trong bảng câu hỏi thì biến tương ứng sẽ có giá trị 1, nếu không được chọn thì đánh số 0.
- Câu trả lời mở: nhập chính xác câu trả lời ghi trong bảng câu hỏi, sau đó đọc và phân nhóm câu trả lời rồi mã hóa.

Chú ý: Cần phải tạo 1 file để chứa tên và giải thích ý nghĩa của các biến có trong dữ liệu để thuận tiện cho việc phân tích và kế thừa dữ liệu.

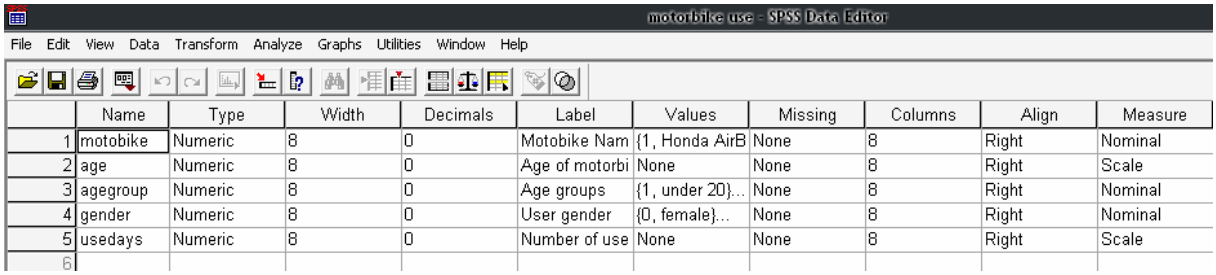


File Edit View Data Transform Analyze Graphs Utilities Window Help

5 :

| | motobike | age | agegroup | gender | usedays | var |
|---|---|---|---|---|---|---|
| 1 | 1 | 25 | 2 | 1 | 30 | |
| 2 | 1 | 27 | 2 | 1 | 12 | |
| 3 | 1 | 54 | 5 | 1 | 14 | |
| 4 | 1 | 33 | 3 | 0 | 15 | |
| 5 | 1 | 35 | 3 | 1 | 30 | |
| 6 | 1 | 42 | 4 | 1 | 32 | |
| 7 | 1 | 19 | 1 | 0 | 23 | |
| 8 | 1 | 19 | 1 | 0 | 25 | |
| 9 | 1 | 21 | 2 | 1 | 27 | |
| 10 | 1 | 34 | 3 | 1 | 21 | |
| 11 | 2 | 56 | 5 | 0 | 14 | |
| 12 | 2 | 57 | 5 | 1 | 12 | |
| 13 | 2 | 43 | 4 | 0 | 19 | |
| 14 | 2 | 44 | 4 | 1 | 22 | |
| 15 | 2 | 23 | 2 | 1 | 13 | |
| 16 | 2 | 21 | 2 | 1 | 15 | |
| 17 | 2 | 22 | 2 | 0 | 22 | |
| 18 | 2 | 21 | 2 | 0 | 23 | |

**Hình 5. 2 Cách nhập dữ liệu vào bảng tính SPSS**



motorbike use - SPSS Data Editor

File Edit View Data Transform Analyze Graphs Utilities Window Help

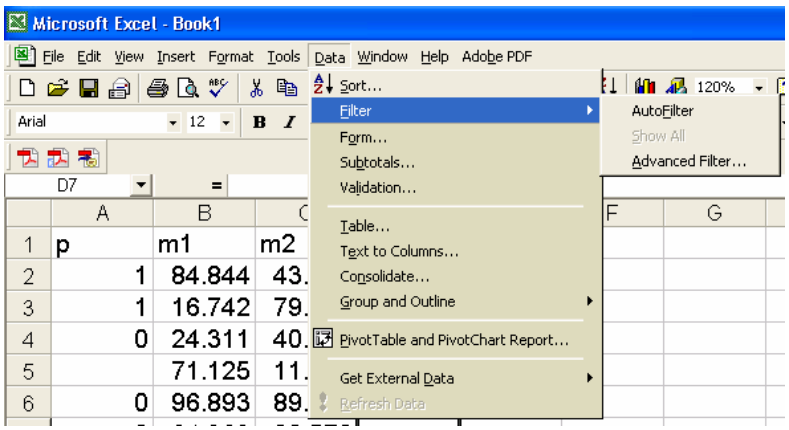
| | Name | Type | Width | Decimals | Label | Values | Missing | Columns | Align | Measure |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | motobike | Numeric | 8 | 0 | Motobike Nam | {1, Honda AirB | None | 8 | Right | Nominal |
| 2 | age | Numeric | 8 | 0 | Age of motorbi | None | None | 8 | Right | Scale |
| 3 | agegroup | Numeric | 8 | 0 | Age groups | {1, under 20}... | None | 8 | Right | Nominal |
| 4 | gender | Numeric | 8 | 0 | User gender | {0, female}... | None | 8 | Right | Nominal |
| 5 | usedays | Numeric | 8 | 0 | Number of use | None | None | 8 | Right | Scale |
| 6 | | | | | | | | | | |

**Hình 5.3 Cách định nghĩa các thuộc tính của các biến số định tính và định lượng**

## 3. THANH LỌC DỮ LIỆU (Data Screening)

### 3.1 Phát hiện giá trị dị biệt trong dữ liệu

**a. Sử dụng Excel: hàm Max và Min, công cụ Auto Filter, đồ thị Scatter.**



Microsoft Excel - Book1

File Edit View Insert Format Tools Data Window Help Adobe PDF

Sort...
Filter
Form...
Subtotals...
Validation...
Table...
Text to Columns...
Consolidate...
Group and Outline
PivotTable and PivotChart Report...
Get External Data
Refresh Data

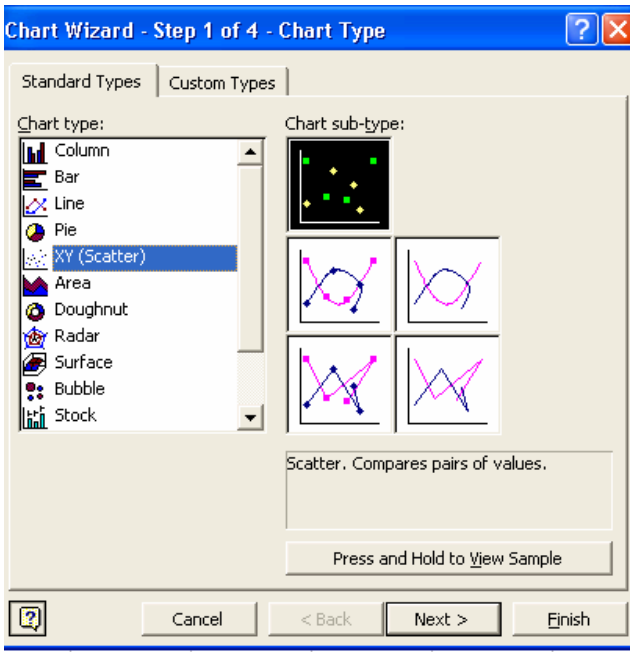
AutoFilter
Show All
Advanced Filter...

120%

Arial 12

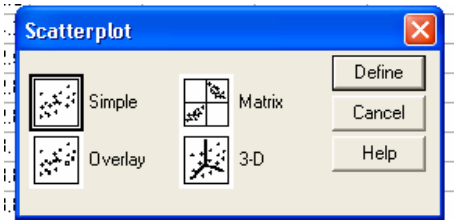
D7

| | A | B | C | F | G |
|---|---|---|---|---|---|
| 1 | p | m1 | m2 | | |
| 2 | 1 | 84.844 | 43. | | |
| 3 | 1 | 16.742 | 79. | | |
| 4 | 0 | 24.311 | 40. | | |
| 5 | | 71.125 | 11. | | |
| 6 | 0 | 96.893 | 89. | | |

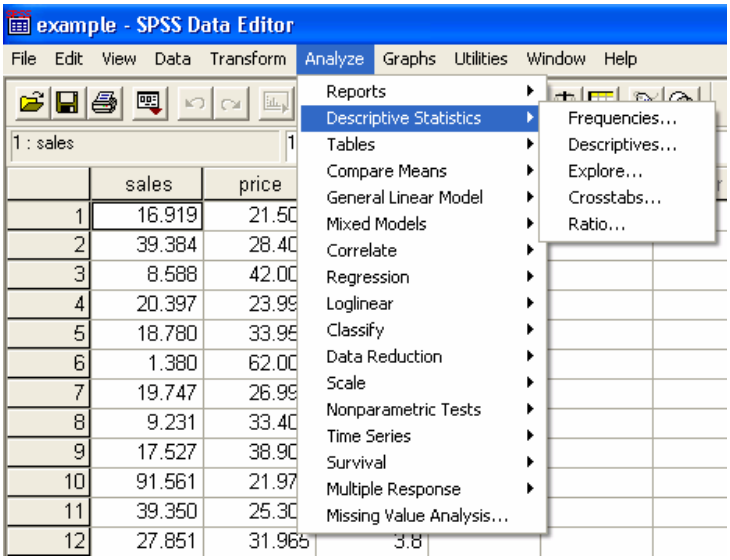



**Hình 5.4** Công cụ đồ thị Scatter trong Excel

**b. Sử dụng SPSS: đồ thị Scatter, công cụ Frequency, Bar Chart, Pie Chart, và Box Plot trong Explore**




**Hình 5.5 Đồ thị Scatter trong SPSS**




**Hình 5.6 Công cụ Frequency và Explore trong SPSS**

## i. Bảng tần suất (Frequency Tables)

Bảng tần suất là một công cụ đơn giản để sắp xếp dữ liệu. Nó giúp sắp xếp dữ liệu theo giá trị số, với các cột thể hiện các chỉ số phần trăm, phần trăm có hiệu lực (phần trăm sau khi điều chỉnh số liệu mất), và phần trăm cộng dồn.

### Ví dụ 5.1 Bảng tần suất của biến số Nhãn hiệu xe máy được sử dụng

**Statistics for motorbike use example**

Motobike Names

| | | |
|---|---|---|
| N | Valid | 100 |
| | Missing | 0 |

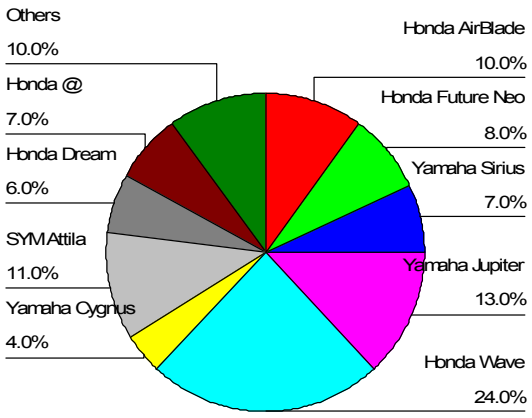
**Frequency table**

| | Frequency | Percent | Valid Percent | Cumulative Percent |
|---|---|---|---|---|
| Honda Air Blade | 10 | 10.0 | 10.0 | 10.0 |
| Honda Future Neo | 8 | 8.0 | 8.0 | 18.0 |
| Yamaha Sirius | 7 | 7.0 | 7.0 | 25.0 |
| Yamaha Jupiter | 13 | 13.0 | 13.0 | 38.0 |
| Honda Wave | 24 | 24.0 | 24.0 | 62.0 |
| Yamaha Cygnus | 4 | 4.0 | 4.0 | 66.0 |
| SYM Attila | 11 | 11.0 | 11.0 | 77.0 |
| Honda Dream | 6 | 6.0 | 6.0 | 83.0 |
| Honda @ | 7 | 7.0 | 7.0 | 90.0 |
| Others | 10 | 10.0 | 10.0 | 100.0 |
| **Total** | **100** | **100.0** | **100.0** | |

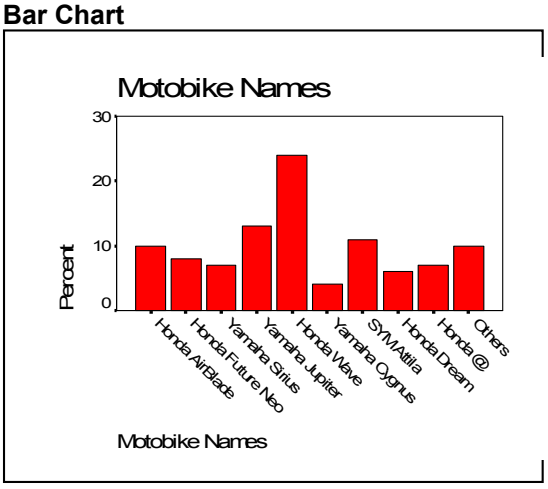
### ii. Biểu đồ thanh (Bar Charts) và Biểu đồ bánh (Pie Charts)

Các dữ liệu tương tự có thể được thể hiện dưới dạng biểu đồ thanh và biểu đồ bánh.

**Pie Chart**




**Bar Chart**




### iii. Biểu đồ Histograms

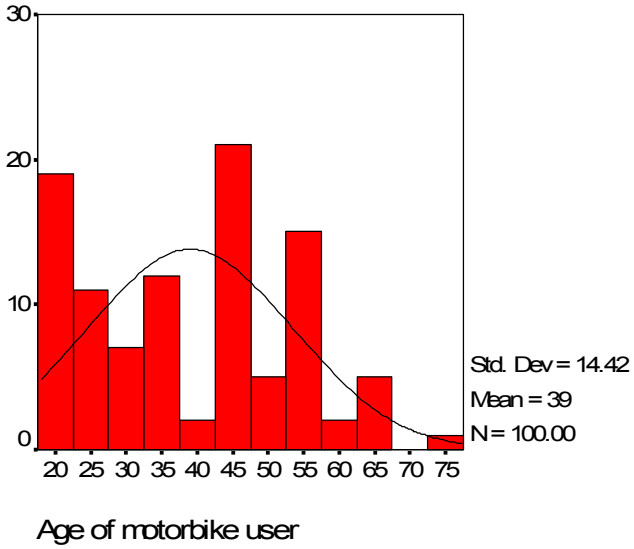
Biểu đồ histogram là một giải pháp quy ước dùng để thể hiện các dữ liệu tỷ lệ hoặc khoảng cách. Biểu đồ histogram được sử dụng để phân nhóm các giá trị dữ liệu của các biến số (variable) thành các khoảng cách. Biểu đồ histogram được xây dựng dưới dạng các thanh thể hiện giá trị dữ liệu.

Biểu đồ histogram rất hữu dụng cho việc: (1) thể hiện tất cả các khoảng cách trong một phân phối (distribution), và (2) trắc nghiệm dạng hình của phân phối như độ méo (skewness), độ nhọn (kurtosis).

***Ghi chú**: Biểu đồ histogram không dùng được cho các biến danh nghĩa.*

**Ví dụ 5.2 Phân phối biến số tuổi của người sử dụng xe máy**

**Histogram of user age**



### iv. Biểu đồ Thân-và-Lá (Stem-and-Leaf Displays)

Mỗi dòng của biểu đồ được gọi là một *thân*; và mỗi số liệu thể hiện trên một thân gọi là một *lá*.

Theo ví dụ 7.3, ý nghĩa của dòng (thân) thứ nhất là có 6 số liệu có chữ số đầu là 1 (hàng chục) là 18, 18, 19, 19, 19, 19.

Khi biểu đồ thân-và-lá được quay trái $90^0$ , nó sẽ có dạng hình tương tự như biểu đồ histogram.

**Ví dụ 5.3 Biểu đồ Thân-và Lá của biến số Tuổi của người sử dụng xe máy**

```
 Frequency    Stem &  Leaf

     6.00        1 .  889999
    18.00        2 .  000111122222233344
     8.00        2 .  55677788
    13.00        3 .  0012233334444
     4.00        3 .  5556
    12.00        4 .  123333334444
    13.00        4 .  5555566777789
    10.00        5 .  0123344444
     9.00        5 .  566667779
     2.00        6 .  03
     4.00        6 .  5567
      .00        7 .
     1.00        7 .  6

 Stem width:        10
 Each leaf:       1 case(s)
```

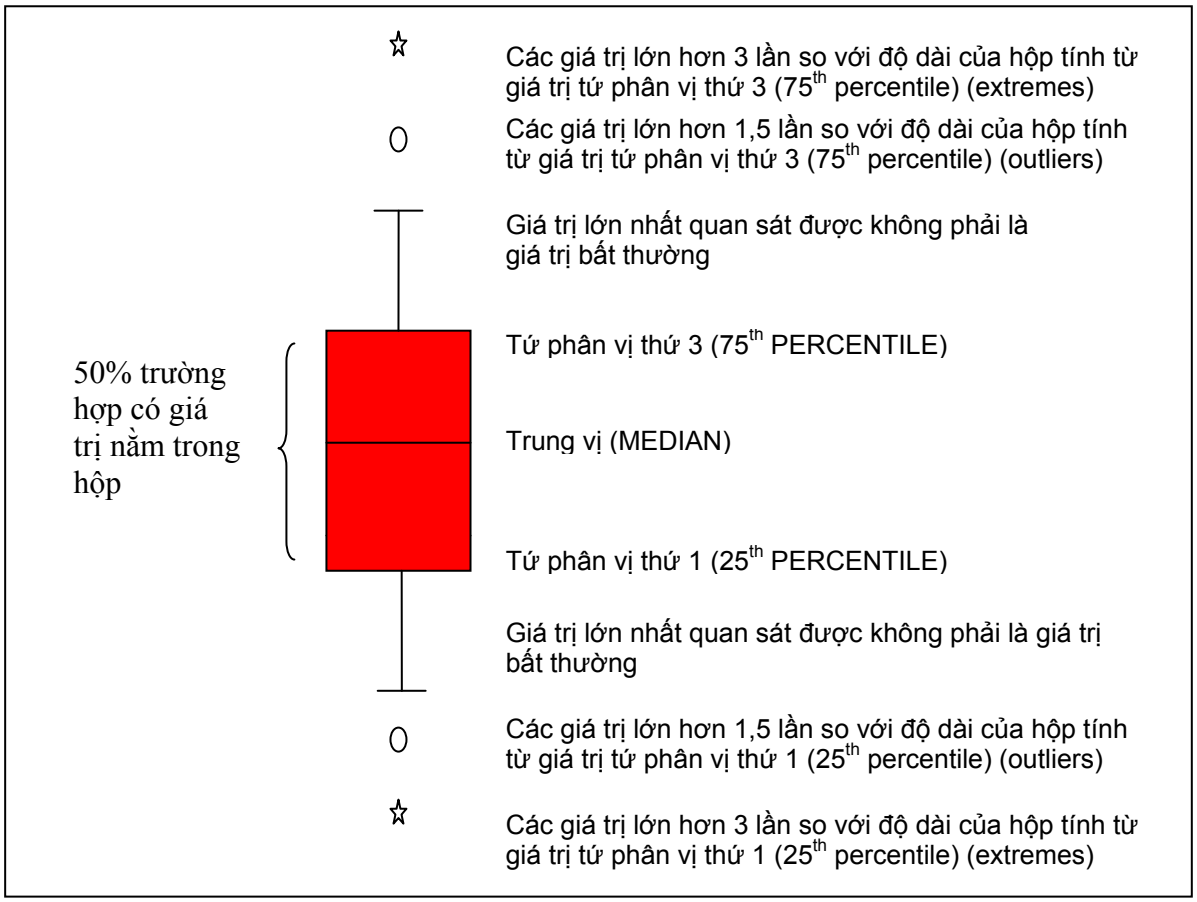
### v. Biểu đồ hộp (Box Plots)

Biểu đồ hộp, hay còn gọi là biểu đồ hộp-và-râu (*box-and-whisker plot)*, cho ta một hình ảnh trực quan khác về vị trí, độ phân tán, dạng hình, độ dài đuôi và các giá trị bất thường (outliers) của phân phối.

Biểu đồ hộp thể hiện tóm tắt 5 giá trị thống kê của một phân phối là trung vị (median), hai tứ phân vị trên và dưới (the upper and lower quartiles), và các giá trị quan sát lớn nhất và nhỏ nhất. Các thành phần chủ yếu của biểu đồ hộp là:

1. Hộp hình chữ nhật chứa đựng 50% các giá trị dữ liệu.
2. Đường thẳng ở trung tâm hộp là giá trị trung vị.
3. Hai lề của hộp thể hiện hai giá trị tứ phân vị thứ 1 và thứ 3 (tương ứng với giá trị thứ 25% ($25^{th}$ percentile) và giá trị thứ 75% ($75^{th}$ percentile) của dãy số liệu.
4. Các "râu" kéo dài từ lề phía trên và phía dưới của hộp thể hiện giá trị lớn nhất và nhỏ nhất. Các giá trị này nằm trong khoảng tối đa 1,5 lần khoảng cách giữa các tứ phân vị tính từ lề của hộp.

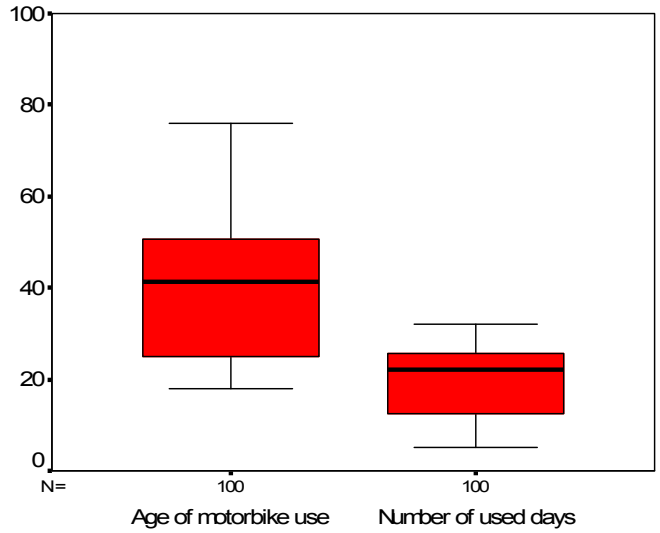
Khi trắc nghiệm dữ liệu, điều quan trọng là phải tách biệt các giá trị bất thường sinh ra từ các lỗi đo lường, hiệu đính, mã hóa và nhập dữ liệu. Các giá trị bất thường này vượt quá 1,5 lần khoảng cách tứ phân vị.




**Hình 5.7 Biểu đồ hộp và các chỉ số**

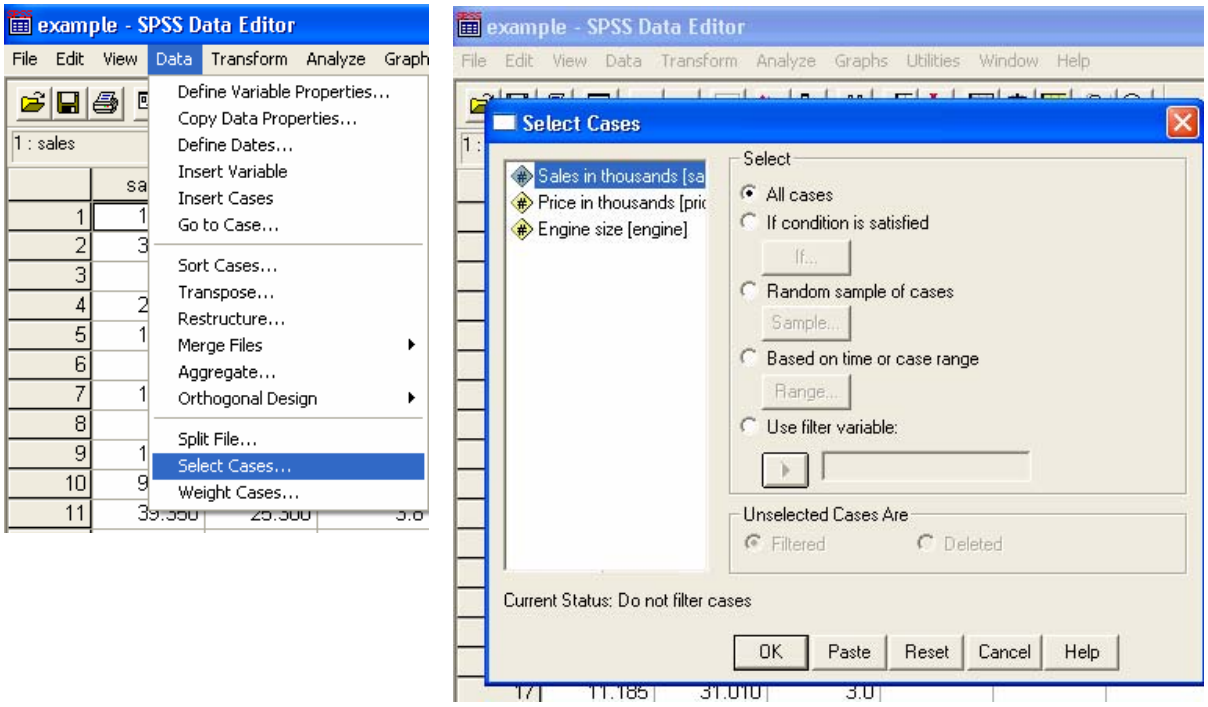
**Ví dụ 5.4 Biểu đồ hộp của biến số Tuổi của người sử dụng xe máy và số ngày sử dụng trong tháng**

**Boxplots: Age of motorbike user and number of used days**




## 3.2 Phát hiện và xử lý dữ liệu bị khuyết (Missing data)

- Sử dụng Excel: công cụ **Auto Filter**.
- Sử dụng SPSS: công cụ **Frequency** và **Select Cases**.




**Hình 5.8 và 5.9** Công cụ **Select Cases** trong SPSS

# 4. PHÂN TÍCH THỐNG KÊ MÔ TẢ

## 4.1 Phân tích thống kê mô tả định lượng

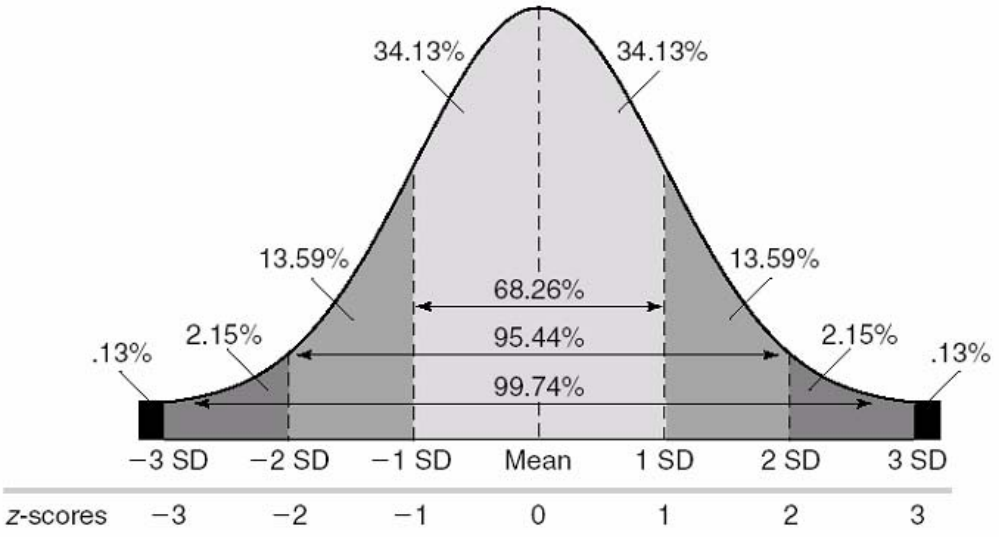
- Sử dụng Excel: công cụ **Descriptives Statistics** trong chức năng **Data Analysis**.
- Sử dụng SPSS: công cụ **Frequency, Descriptives, Explore** trong chức năng Descriptive Statistics của SPSS.

### 4.1.1 Ôn lại các khái niệm thống kê

#### (1) Phân phối bình thường

**Phân phối bình thường** là mô hình phân phối của một bộ dữ liệu theo dạng dường cong hình quả chuông. Đường phân phối chuẩn có các đặc tính sau:

- Đường cong tập trung ở phần trung tâm và giảm đều về hai bên. Điều này có nghĩa dữ liệu ít có xu hướng có các giá trị bất thường.
- Hình chuông cân đối, có nghĩa là xác suất lệch khỏi giá trị trung bình (mean) là bằng nhau kể cả về hai phía.




**Hình 5.10 Đường phân phối chuẩn và các đặc tính**

#### (2) Các chỉ tiêu thống kê mô tả

Phân tích thống kê mô tả chỉ ra các đặc điểm về xu hướng trung tâm, tính biến thiên và dạng hình phân phối của dữ liệu.

***Đo lường xu hướng trung tâm (Measures of Central Tendency)***

Các chỉ tiêu đo lường xu hướng trung tâm bao gồm giá trị trung bình (mean), trung vị (median) và mode.

- Giá trị ***trung bình*** (*mean)* là tổng tất cả giá trị của các dữ liệu chia cho số lượng của dữ liệu.

$$\overline{x} = \frac{x_1 + x_2 + \cdots + x_N}{N} = \frac{1}{N}\sum_{i=1}^{N} x_i \,.$$

- ***Trung vị*** (*median)* là giá trị của số liệu có vị trí nằm giữa bộ số liệu sắp xếp theo trật tự. Đây chính là điểm giữa của phân phối. Khi số quan sát là chẵn, trung vị là giá trị trung bình của hai quan sát ở vị trí trung tâm.
- ***mode*** là giá trị của quan sát có tần suất xuất hiện nhiều nhất trong bộ dữ liệu.
- ***Khoảng cách*** (*range)* là giá trị khác biệt giữa con số lớn nhất và nhỏ nhất trong bộ dữ liệu.

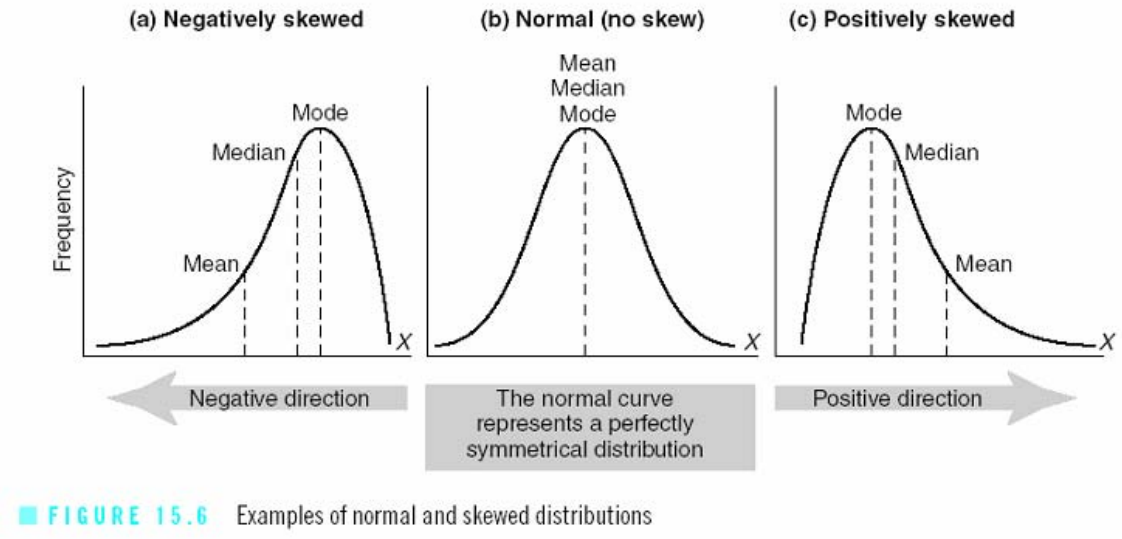



FIGURE 15.6 Examples of normal and skewed distributions

**Hình 5.11 Các dạng phân phối lệch trái và lệch phải so với phân phối bình thường**

***Đo lường tính biến thiên (Measures of Variability)***

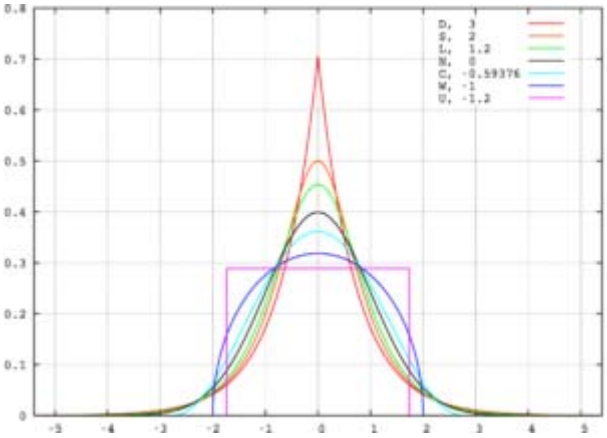
- ***Phương sai*** (Variance; $\boldsymbol{\sigma^2}$) là trung bình tổng các sai số bình phương giữa các giá trị của các quan sát và giá trị trung bình.
- ***Độ lệch chuẩn*** (Standard deviation; SD; $\boldsymbol{\sigma}$) đo lường mức độ phân tán của số liệu xung quanh giá trị trung bình.

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \overline{x})^2}\,,$$

- ***Sai số chuẩn*** của giá trị trung bình (Standard error of the mean; s.e.) đo lường phạm vi mà giá trị trung bình của quần thể (μ) có thể xuất hiện với một xác suất cho trước dựa trên giá trị trung bình của mẫu (mean).

***Đo lường dạng hình của phân phối (Measures of Shape)***

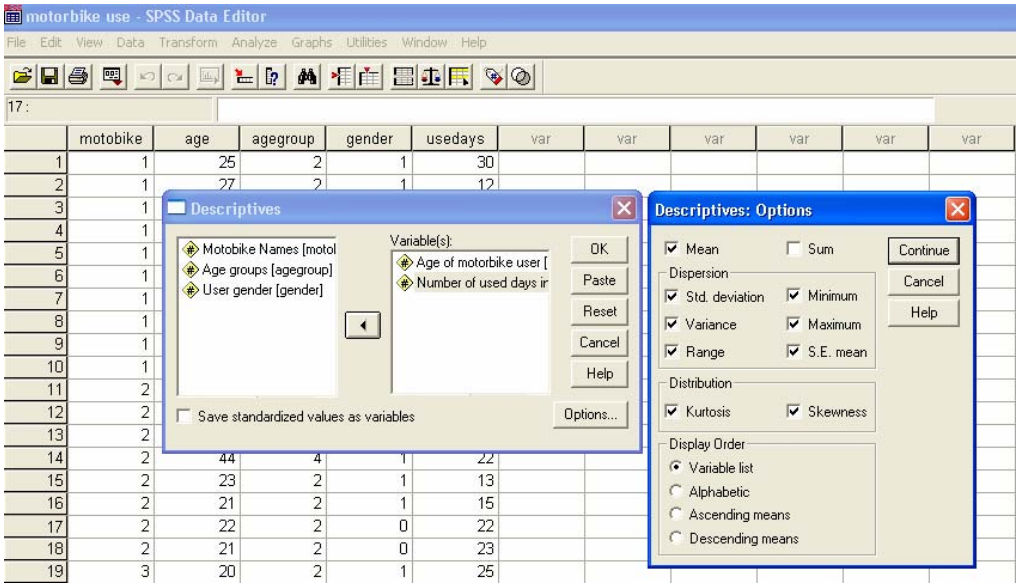
- Độ méo (skewness) đo lường độ lệch của phân phối về một trong hai phía. Phân phối méo trái (negative skew, left-skewed**)** khi đuôi phía trái dài hơn, và phần lớn số liệu tập trung ở phía phải của phân phối. Phân phối méo phải (positive sknew, right-skewed) khi đuôi phía phải dài hơn, và phần lớn số liệu tập trung ở phía trái của phân phối. Khi lệch phải, giá trị sknewness dương; khi lệch trái, giá trị skewness âm. Độ méo càng lớn thì giá trị sknewness càng lớn hơn 0.
- Độ nhọn (kurtosis) đo lường mức độ nhọn hay bẹt của phân phối so với phân phối bình thường (có độ nhọn bằng 0). Phân phối có dạng nhọn khi giá trị kurtosis dương và có dạng bẹt khi giá trị kurtosis âm.
- Với phân phối bình thường, giá trị của độ méo và độ nhọn bằng 0. Căn cứ trên tỷ số giữa giá trị skewness và kurtosis và sai số chuẩn của nó, ta có thể đánh giá phân phối có bình thường hay không (khi tỷ số này nhỏ hơn -2 và lớn hơn +2, phân phối là không bình thường).



**Hình 5.12 Các dạng phân phối nhọn và bẹt so với phân phối bình thường**

### 4.1.2 Thống kê mô tả

### i. Dùng công cụ Descriptives trong SPSS




**Hình 5.13 Các chức năng thống kê mô tả của công cụ Descriptives**

**Ví dụ 5.6 Thống kê mô tả các biến số Tuổi của người sử dụng xe máy và số ngày sử dụng trong tháng**

Descriptive Statistics

| | | Statistic | Std. Error |
|---|---|---|---|
| Age of motorbike user | N | 100 | |
| | Range | 58 | |
| | Minimum | 18 | |
| | Maximum | 76 | |
| | Mean | 39.01 | 1.44 |
| | Std. Deviation | 14.42 | |
| | Variance | 207.909 | |
| | Skewness | .242 | .241 |
| | Kurtosis | -.948 | .478 |
| Number of used days in a month | N | 100 | |
| | Range | 27 | |
| | Minimum | 5 | |
| | Maximum | 32 | |
| | Mean | 20.15 | .74 |
| | Std. Deviation | 7.40 | |
| | Variance | 54.715 | |
| | Skewness | -.302 | .241 |
| | Kurtosis | -1.138 | .478 |
| Valid N (listwise) | N | 100 | |

### ii. Dùng công cụ Explore trong SPSS

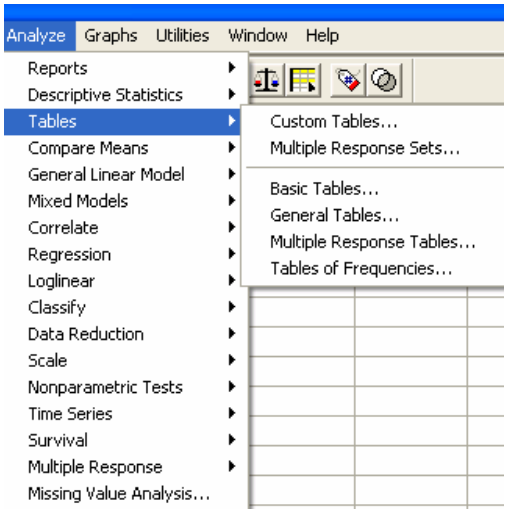
Công cụ Explore rất thích hợp để thống kê mô tả chi tiết các biến số phân nhóm theo một biến phân loại khác (factor variable).

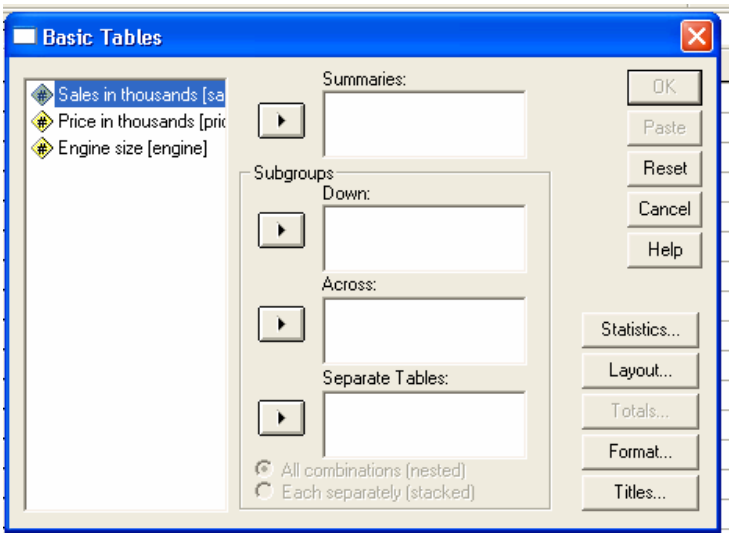
**Ví dụ 5.7 Thống kê mô tả các biến số Tuổi của người sử dụng xe máy và số ngày sử dụng trong tháng phân theo giới tính**

| | | | Age of motorbike user | | Number of used days in a month | |
|---|---|---|---|---|---|---|
| User gender | | | Statistic | Std. Error | Statistic | Std. Error |
| female | Mean | | 38.46 | 2.11 | 20.71 | 1.07 |
| | 95% Confidence Interval for Mean | Lower Bound | 34.19 | | 18.54 | |
| | | Upper Bound | 42.74 | | 22.88 | |
| | 5% Trimmed Mean | | 38.13 | | 20.95 | |
| | Median | | 41.00 | | 22.00 | |
| | Variance | | 183.205 | | 47.212 | |
| | Std. Deviation | | 13.54 | | 6.87 | |
| | Minimum | | 19 | | 7 | |
| | Maximum | | 65 | | 30 | |
| | Range | | 46 | | 23 | |
| | Interquartile Range | | 23.00 | | 11.00 | |
| | Skewness | | .118 | .369 | -.513 | .369 |
| | Kurtosis | | -1.089 | .724 | -.838 | .724 |
| male | Mean | | 39.39 | 1.97 | 19.76 | 1.01 |
| | 95% Confidence Interval for Mean | Lower Bound | 35.45 | | 17.74 | |
| | | Upper Bound | 43.33 | | 21.79 | |
| | 5% Trimmed Mean | | 38.87 | | 19.90 | |
| | Median | | 42.00 | | 21.00 | |
| | Variance | | 228.173 | | 60.460 | |
| | Std. Deviation | | 15.11 | | 7.78 | |
| | Minimum | | 18 | | 5 | |
| | Maximum | | 76 | | 32 | |
| | Range | | 58 | | 27 | |
| | Interquartile Range | | 28.00 | | 15.00 | |
| | Skewness | | .292 | .311 | -.175 | .311 |
| | Kurtosis | | -.932 | .613 | -1.271 | .613 |

## 4.2 Phân tích thống kê mô tả định tính

### 4.2.1 Sử dụng công cụ Basic Table trong SPSS




**Hình 5.14 và 5.15** Công cụ **Basic Table** trong SPSS

### 4.2.2 Sử dụng công cụ Bảng chéo (Cross-Tabulation)

Bảng chéo là một kỹ thuật dùng để so sánh dữ liệu từ hai hoặc nhiều hơn các biến phân loại hoặc danh nghĩa (categorical or nominal variables), ví dụ như là giới tính. Bảng chéo sử dụng các bảng có các cột và dòng thể hiện các mức độ hoặc các giá trị mã hóa của từng biến phân loại hoặc danh nghĩa.

Bảng chéo là bước đầu tiên để xác định các quan hệ giữa các biến. Khi bảng chéo được xây dựng để trắc nghiệm thống kê, ta gọi chúng là bảng contingency (contingency tables), và loại trắc nghiệm dùng để đánh giá liệu các biến phân loại có độc lập với nhau hay không là $\chi^2$ (Chi bình phương / chi-square).

**Ví dụ 5.5 Bảng chéo giữa hai biến số Nhãn hiệu xe máy và Giới tính của người sử dụng**

**Motobike Names * User gender Crosstabulation**

| | | User gender | | Total |
|---|---|---|---|---|
| | | female | male | |
| Motobike Names | Honda AirBlade | 3 | 7 | 10 |
| | Honda Future Neo | 4 | 4 | 8 |

| | | | | |
|---|---|---|---|---|
| | Yamaha Sirius | 3 | 4 | 7 |
| | Yamaha Jupiter | 6 | 7 | 13 |
| | Honda Wave | 9 | 15 | 24 |
| | Yamaha Cygnus | 2 | 2 | 4 |
| | SYM Attila | 5 | 6 | 11 |
| | Honda Dream | 2 | 4 | 6 |
| | Honda @ | 3 | 4 | 7 |
| | Others | 4 | 6 | 10 |
| Total | | 41 | 59 | 100 |

**Ví dụ 5.6 Bảng chéo giữa hai biến số Nhãn hiệu xe máy và Nhóm tuổi của người sử dụng**

**Motorbike Names * Age Group Crosstabulation**

| | | Age groups | | | | | | Total |
|---|---|---|---|---|---|---|---|---|
| | | under 20 | under 30 | under 40 | under 50 | under 60 | older than 60 | |
| Motobike Names | Honda AirBlade | 2 | 3 | 3 | 1 | 1 | | 10 |
| | Honda Future Neo | | 4 | | 2 | 2 | | 8 |
| | Yamaha Sirius | | 1 | | 1 | 2 | 3 | 7 |
| | Yamaha Jupiter | | 4 | 1 | 4 | 4 | | 13 |
| | Honda Wave | 1 | 2 | 8 | 7 | 5 | 1 | 24 |
| | Yamaha Cygnus | | 1 | 1 | 1 | 1 | | 4 |
| | SYM Attila | 3 | 4 | 1 | 2 | | 1 | 11 |
| | Honda Dream | | 3 | 1 | 1 | | 1 | 6 |
| | Honda @ | | 2 | | 1 | 4 | | 7 |
| | Others | | 2 | 2 | 5 | | 1 | 10 |
| Total | | 6 | 26 | 17 | 25 | 19 | 7 | 100 |

# 5. PHÂN TÍCH TRẮC NGHIỆM GIẢ THIẾT

## 5.1 Trắc nghiệm giả thiết

Mục tiêu của trắc nghiệm giả thiết là nhằm quyết định tính chính xác của giả thiết dựa trên các số liệu mẫu thu thập được. Chúng ta đánh giá tính chính xác của các giả thiết bằng cách áp dụng các kỹ thuật thống kê; và đánh giá tầm quan trọng của sự khác biệt có ý nghĩa thống kê.

Cách tiếp cận cổ điển hay là lý thuyết lấy mẫu thể hiện cách nhìn mục tiêu theo xác suất dựa trên phân tích dữ liệu mẫu. Một giả thiết được xây dựng, nó sẽ bị bác bỏ hoặc chấp nhận dựa trên mẫu dữ liệu thu thập.

**Bảng 5.2 Các kỹ thuật phân tích thống kê nên dùng theo loại dữ liệu và trắc nghiệm**

| **Measurement scale** | **One-sample Case** | **Two-Samples Tests** | | ***k*-Samples Tests** | |
|---|---|---|---|---|---|
| | | Related Samples | Independent Samples | Related Samples | Independent Samples |
| Nominal | - Binomial<br>- $\chi^2$ one-sample test | - McNemar | - Fisher exact test<br>- $\chi^2$ two-sample test | - Cochran *Q* | - $\chi^2$ for *k*-samples |
| Ordinal | - Kolmogorov-Smirnov one-sample test<br>- Runs test | - Sign test<br>- Wilcoxon matched-pairs test | - Median test<br>- Mann-Whitney U<br>- Kolmogorov-Smirnov<br>- Wald-Wolfowitz | - Friedman two-way ANOVA | - Median extension<br>- Kruskal-Wallis one-way ANOVA |
| Interval and Ratio | - T-test<br>- Z test | - T-test for paired samples | - T-test<br>- Z test | - Repeated-measured ANOVA | - One-way ANOVA<br>- N-way ANOVA |

## 5.2 Quy trình trắc nghiệm thống kê

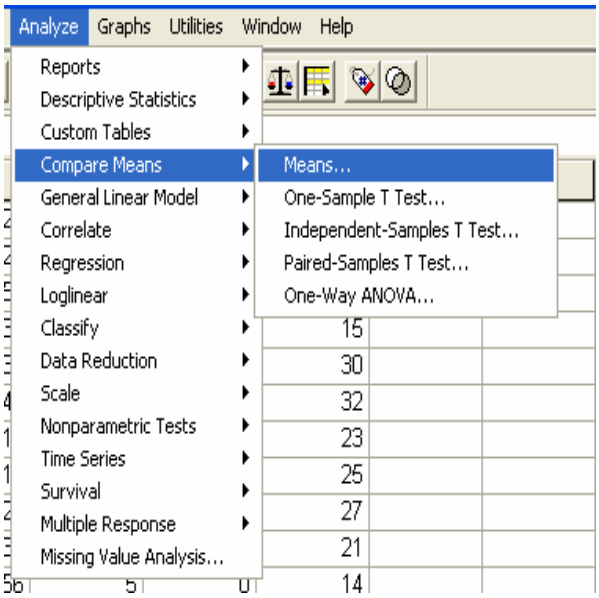
Trắc nghiệm ý nghĩa thống kê đi theo một trình tự 6 bước tương đối rõ ràng.

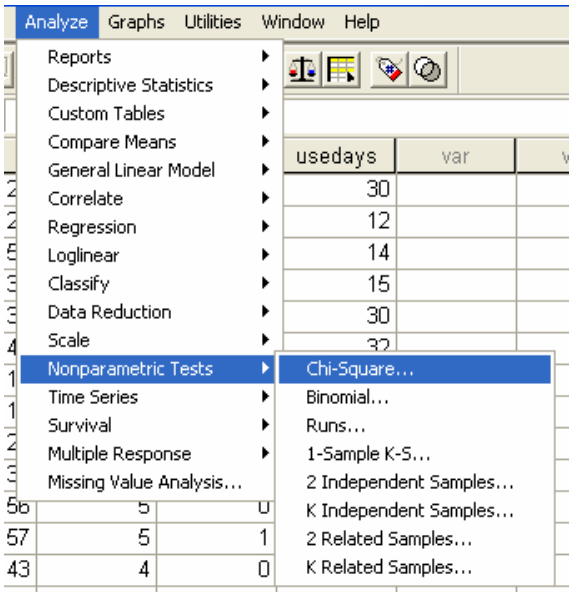
| |
|---|
| 1. Phát biểu giả thiết |
| 2. Chọn loại trắc nghiệm thống kê |
| 3. Chọn mức ý nghĩa mong muốn |
| 4. Tính giá trị khác biệt |
| 5. Có được giá trị trắc nghiệm |
| 6. Diễn giải kết quả trắc nghiệm |

## 5.3 Phân tích dữ liệu

a. Sử dụng Excel: công cụ **Correlation**, **Anova** và **Regression** trong chức năng **Data Analysis**.

b. Sử dụng SPSS: các công cụ Compare Means và Nonparametric Tests




**Từ khóa**

| | |
|---|---|
| Biểu đồ hộp | Boxplot |
| Phân tích dữ liệu xác nhận | Confirmatory data analysis |
| Biến kiểm soát, biến đối chứng | Control variable |
| Bảng chéo | Cross-tabulation |
| Phân tích dữ liệu khám phá | Exploratory data analysis |
| Bảng tần suất | Frequency table |
| Biểu đồ histogram | Histogram |
| Khoảng cách phân vị | Interquartile range |
| Điểm dị biệt | Outliers |
| Biểu đồ thân và lá | Stem-and-leaf display |

# Chương 8. Viết báo cáo nghiên cứu

Mục tiêu giảng dạy.

Các nội dung:

- Xác định thông điệp. Cần xác định đề tài nghiên cứu muốn truyền tải nội dung gì đến người đọc. Đây là nội dung quan trọng nhất trong báo cáo.
- Sắp xếp ý tưởng: truyền tải một cách hiệu quả các thông tin thành những ý tưởng thuyết phục, cấu trúc bài viết, đưa các ý tưởng vào cấu trúc bài viết một cách hiệu quả.
- Viết bản nháp: cách sử dụng các công cụ trong văn bản để trình bày ý tưởng một cách rõ ràng
- Sửa bản nháp: làm thế nào để lời văn rõ ràng hơn, dễ hiểu hơn, nhấn mạnh thông điệp của từng đoạn, cấu trúc câu và lựa chọn từ ngữ

## 1. GIỚI THIỆU

Nhiều người cho rằng viết báo cáo/nghiên cứu là để truyền tải thông tin. Tuy nhiên một bài viết hiệu quả còn hơn thế. Nó phải:

- Làm thay đổi cách nhìn nhận vấn đề của người đọc
- Thuyết phục người đọc tin vào một điều gì đó
- đưa người đọc đến quyết định và hành động
- dẫn dắt người đọc theo một quy trình nào đó

Để viết một cách hiệu quả, cần nhớ những nguyên tắc sau:

- bài viết phải có một ý tưởng chủ đạo
- viết nhằm vào một đối tượng độc giả cụ thể
- đưa ra các quan điểm và biện hộ cho nó
- các ý tưởng khác nên được dùng để biện luận cho ý tưởng chủ đạo và không nên quá nhiều

Bốn bước để xây dựng một bài viết hiệu quả

- xây dựng thông điệp
- sắp xếp ý tưởng

- viết bản nháp
- chỉnh sửa

## 2. XÂY DỰNG THÔNG ĐIỆP

Ý tưởng chủ đạo của bài viết là thông điệp chính... tất cả các ý tưởng khác đều xoay quanh và hỗ trợ cho ý tưởng này.

4 bước để xác định thông điệp/ý tưởng chủ đạo

- xác định mục tiêu nghiên cứu
- Xác định độc giả (viết cho ai?)
- Trnh bày ý tưởng chủ đạo
- Chỉnh sửa

### 2.1 Xác định mục tiêu

Cần xác định rõ nghiên cứu được thực hiện nhằm mục đích gì, tức là nghiên cứu sẽ có vai trò gì, đóng góp gì cho thực tế. Điều này khác với việc xác định nội dung nghiên cứu.

Để xác định mục tiêu cần phải:

- Người viết muốn đem lại sự thay đổi gì?
- Muốn người đọc làm gì?
- Bài viết sẽ đóng góp gì

### 2.2 Độc giả

Người đọc khác nhau sẽ có kiến thức, kỳ vọng và ưu tiên khác nhau. Sẽ dễ hơn nếu xác định được ai sẽ là người đọc nghiên cứu của mình. Việc xác định rõ người đọc sẽ có ích cho việc:

- Xác định hình thức của báo cáo
- Bản chất của kết quả báo cáo. Báo cáo c1 cần phải đưa ra một chương trình hành đõng cụ thể không?
- Mức độ chi tiết của báo cáo
- Ưu tiên và kỳ vọng của người đọc là gì?
- Ý tưởng của chúng ta có phù hợp với mục tiêu của người đọc không?

### 2.3 Trình bày ý tưởng chủ đạo

Một bài viết hiệu quả phải có một ý tưởng chủ đạo. Ý tưởng này sẽ quyết định toàn bào viết. tất cả những thứ khác, báo gồm những ý kiến ,lập luận... và cả cách trình bày đều do ý tưởng chủ đạo quyết định.

Hãy nghĩ về người đọc của chúng ta. Họ là ai? Họ mong muốn gì? Giả sử chúng ta đang trình bày và thảo luận với họ về nghiên cứu của chúng ta và học chưa biết gì về bài viết của chúng ta, chúng ta sẽ nói gì? Hãy thử phát biểu chỉ trong một câu...

Bây giờ hãy xem lại câu chúng ta vừa viết...

- Đó có phải là điều chúng ta muốn nói với người đọc không?
- Họ có hiểu ý chúng ta không?
- Ý tưởng và cách trình bày ý tưởng của chúng ta có phù hợp không?
- Họ có thích thú với ý tưởng của chúng ta không?

Ý tưởng chủ đạo cần phải:

- Diễn đạt được mục đích của chúng ta
- Có ý nghĩa
- Tập trung vào hành động
- Lôi cuốn độc giả
- Sử dụng ngôn từ thích hợp, dễ hiểu
- Cung cấp thông tin mới
- Gợi ra câu hỏi/vấn đề cho người đọc

### 2.4 Chỉnh sửa

Bao gồm 4 bước

- Tình huống
- Vấn đề
- Câu hỏi
- Trả lời/phản hồi

*Tình huống*

Một tình huống tốt phải rõ ràng, dễ hiểu và không gây tranh cãi

*Vấn đề*

Là điều xảy ra trong tình huống – là cái có thể được cải thiện nhờ bài viết của chúng ta... Vấn đề có thể là:

- Điều gì đó không đúng
- Điều gì đó có thể sẽ không đúng
- Xảy ra sự khác biệt giữa điều người ta mong đợi và thực tế xảy ra
- Tình hình đã thay đổi
- Chúng ta thấy trước sự thay đổi tình hình

*Câu hỏi*

Vấn đề đặt ra câu hỏi

| **Vấn đề** | **Câu hỏi** |
|---|---|
| Điều gì đó không đúng | Cái gì không đúng?<br>Làm sao để khắc phục?<br>Phải làm gì bây giờ? |
| Điều gì đó có thể sẽ không đúng | Cái gì có thể xảy ra?<br>Hậu quả của nó là gì? |
| Xảy ra sự khác biệt giữa điều người ta mong đợi và thực tế xảy ra | Khác biệt gì?<br>Điều gì gây ra sự khác biệt đó?<br>Phải khắc phục hoặc điều chỉnh kế hoạch tương lai như thế nào? |
| Tình hình đã thay đổi | Thay đổi như thế nào?<br>Hậu quả là gì?<br>Phải làm gì để khắc phục? |
| Chúng ta thấy trước sự thay đổi tình hình | Cái gì có thể thay đổi?<br>Khả năng xảy ra?<br>Có nên tính đến thay đổi này trong kế hoạch?<br>Phải đối phó thế nào đối với rủi ro? |

*Trả lời*

Câu trả lời cho vấn đề cũng chính là ý tưởng chủ đạo của bài viết.

## 3. SẮP XẾP Ý TƯỞNG

Để hiểu vấn đề một cách tường tận, cần "cắt các thông tin ra thành từng mảnh" và sắp xếp chúng lại một cách thích hợp, theo một cấu trúc hợp lý. Ba bước để sắp xếp ý tưởng

- Xây dựng cấu trúc ý tưởng
- Thu thập bằng chứng ủng hộ ý tưởng
- Viết đề cương trong đó trình bày các ý tưởng và mối quan hệ giữa chúng

Đề cương này sẽ giúp chúng ta chuyển cấu trúc ý tưởng thành một dàn ý có thể giúp chúng ta chuẩn bị bản nháp đầu tiên.

Việc viết đề cương giúp chúng ta hiểu rõ hơn vấn đề và trình bày một cách hợp lý.

## 4. VIẾT BẢN THẢO ĐẦU TIÊN (bản nháp)

Viết bản thảo đầu tiên có nghĩa là thêm vào đề cương: từ ngữ, tiêu đề, đánh số và hình, bảng biểu. Quy tắc chung:

- Viết nhanh, không suy nghĩ quá nhiều về từ ngữ, bỏ trống nếu cần thiết để giữ mạch suy nghĩ
- Viết bằng chính văn phong của chúng ta
- Viết không gián đoạn
- Viết không cần chỉnh sửa
- Giữ cấu trúc của đề cương. Dùng các câu trong đề cương để diễn đạt những gì chúng ta cần trình bày. Nếu cần suy nghĩ lại về một câu nào đó trong đề cương, hãy chuyển qua câu kế tiếp và tiếp tục viết

***Các vấn đề khi viết bản thảo***

- Lời văn
- Các kỹ thuật giải thích
- Tóm tắt và giới thiệu
- Trình bày bài viết (navigation aid và layout)

- Tài liệu tham khảo

### 4.1 Lời văn

Lời văn được dùng để:

- Mô tả
- Giải thích
- Lập luận
- Tường thuật

Hai chức năng chính là giải thích và lập luận.

### 4.2 Các kỹ thuật giải thích

- Ví dụ
- Phép so sánh (dùng những trường hợp tương tự)
- Định nghĩa
- Phân loại
- So sánh và tương phản
- Nguyên nhân và kết quả
- Phân tích quá trình

### 4.3 Tóm tắt và giới thiệu

**Tóm tắt** là một phiên bản của bài viết ở dạng ngắn gọn

Giới thiệu là một phần của bài viết, trình bày bài viết được viết như thế nào.

Phần tóm tắt thông thường không quá 400 từ, bao gồm những ý chính trong bài viết. Phần tóm tắt nằm ở đầu, thông thường theo thứ tự sau:

- Trang bìa
- Tóm tắt
- Mục lục
- Giới thiệu

Phần tóm tắt được viết cho những người:

- Không muốn đọc toàn bài viết
- đang suy nghĩ có nên đọc tiếp không
- muốn tìm những nội dung trong bài viết mà họ quan tâm

Lưu ý khi viết tóm tắt:

- trong 1 trang
- nhấn mạnh những điểm chính, thường đặt ở đầu mỗi đoạn
- nhớ trình bày cơ sở của bài viết, những thông tin mà người đọc cần biết để hiểu vấn đề

Tránh:

- lỗi lặp
- không liệt kê
- cung cấp thông tin chi tiết về các chỉ số thống kê
- ví dụ
- hình vẽ và đồ thị

**Mở đầu** hay **Giới thiệu** là phần trình bày tất cả những thông tin cần thiết để đi vào vấn đề chính. Phần này sẽ nhấn mạnh:

- vấn đề mà bài viết đề cập
- câu hỏi/vấn đề mà bài viết sẽ giải quyết

Phần mở đầu thường bao gồm:

- mục đích nghiên cứu
- phạm vi nghiên cứu
- phương pháp
- giải thích các thuật ngữ chính
- các khái niệm cơ bản
- lời cảm ơn
- trình bày ngắn gọn cấu trúc bài viết

### 4.4 Trình bày bài viết

- tiêu đề và định dạng tiêu đề
- tựa bài viết
- đánh số

*Tiêu đề*

Là phần quan trọng giúp người đọc nhận dạng nội dung của từng phần trong bài viết. Nó phản ánh cấu trúc của bài viết. Lưu ý:

- định dạng tiêu đề một cách hệ thống
- tiêu đề phải phản ánh nội dung chính
- giới thiệu các tiêu đề nhỏ và tóm tắt

*Tựa bài viết*

Số người đọc tựa bài viết sẽ nhiều hơn rất nhiều so với số người đọc bài viết của chúng ta. Do vậy, tựa bài viết càng hấp dẫn và gây ấn tượng càng tốt.

Lưu ý tựa bài viết phải:

- cụ thể
- ngắn gọn
- rõ ràng

*Đánh số*

- đánh số trang: nội dung chính nên đánh số 1, 2, 3...; các trang tựa nên đánh số i, ii, iii,...
- đánh số tiêu đề: các tiêu đề chính nên đánh số 1, 2, 3..; tiêu đề con nên đánh số 1.1, 1.2... và 1.1.1, 1.1.2. Không nên dùng quá 3 số.

**4.5 Tài liệu tham khảo và các nội dung khác**

Bao gồm:

- tài liệu tham khảo
- lời cảm ơn
- phụ lục

*Tài liệu tham khảo bao gồm:*

- sách
- tạp chí
- các báo cáo đã công bố
- website
- biên bản họp

- kỷ yếu hội nghị
- luận văn tốt nghiệp
- phỏng vấn cá nhân

Nguyên tắc ghi tài liệu tham khảo, theo thứ tự:

- tên tác giả
- năm xuất bản
- tên bài viết/sách
- nơi xuất bản
- nhà xuất bản
- số trang

*Lời cảm ơn*

Là phần ghi lời cảm ơn đến các cá nhân/tổ chức đã giúp đỡ chúng ta hoàn thành công trình nghiên cứu. Lưu ý:

- nếu có ít người thì ghi theo thứ tự mức độ đóng góp
- nếu nhiều thì ghi theo thứ tự chữ cái

*Phụ lục*

Là phần ghi những thông tin không trực tiếp hỗ trợ cho ý tưởng chính của đề tài nghiên cứu, thông thường là những thông tin rất chi tiết mà một số độc giả có thể muốn tìm hiểu thêm. Bao gồm:

- các chỉ số thống kê chi tiết
- các đồ thị/sơ đồ phức tạp
- kết quả trực tiếp từ máy tính
- trích đoạn từ tạp chí...

Nên đánh số. Ví dụ: Phụ lục A, Phụ lục B...

## 5. CHỈNH SỬA

Nhằm mục đích sửa lại bản thảo đầu tiên của chúng ta cho dễ đọc hơn.

Nguyên lý chung

- nên nghỉ một thời gian và đọc lại bản thảo như là ai đó chứ không phải chúng ta viết ra
- tham khảo ý kiến người khác
- nên sửa trên giấy chứ không nên sửa trên máy tính
- chọn ngôn từ thích hợp và dễ hiểu
- sửa một cách hệ thống, theo 3 cấp độ: đoạn văn, câu và từ.

### 5.1 Cách viết một đoạn văn hiệu quả

Các đoạn văn cấu tạo nên cấu trúc ý tưởng. Mỗi đoạn nên nhấn mạnh một ý tưởng hay lập luận nào đó. Lưu ý:

- sử dụng câu đầu tiên để tóm tắt ý chính của cả đoạn
- chọn độ dài thích hợp cho đoạn văn
- liên kết các đoạn một cách phù hợp
- sử dụng dấu gạch đầu dòng để liệt kê

### 5.2 Chỉnh sửa câu văn

Mỗi câu diễn đạt một ý. Các câu nên ngắn gọn và dễ hiểu để người đọc không phải đọc đi đọc lại. Lưu ý tránh:

- câu quá dài
- cú pháp phức tạp

### 5.3 Lựa chọn từ ngữ

Nên:

- sử dụng từ ngữ dễ hiểu. Mỗi câu chỉ nên có từ 15-20 từ.
- tránh dùng câu bị động. Câu chủ động sẽ ngắn gọn hơn và mạnh hơn.
- Sử dụng từ mạnh mẽ, rõ ràng và cụ thể
- Loại bỏ những từ không cần thiết (là những từ có thể bỏ mà không làm thay đổi ý của câu)

**Tài Liệu Tham Khảo**

Hoàng Trọng – Chu Nguyễn Mộng Ngọc. (2005). *Phân tích dữ liệu nghiên cứu với SPSS*, TP.HCM: NXB Thống Kê.

D. Cooper and P. Schindler. (2006). *Business Research Methods*. McGraw-Hill Irwin.

Kumar, R. (2005) *Research Methodology: A step-by-step guide for beginners*, London: Sage, 2nd Edition.