



**FPT POLYTECHNIC**

# **PHÂN TÍCH KINH DOANH ỨNG DỤNG**

**PHIÊN BẢN NỘI BỘ  
PB 0.1 - 30062020**



Arun Sukumar; Lucian Tipi;  
Jayne Revill

---

# PHÂN TÍCH KINH DOANH ỨNG DỤNG

# Mục lục

<b>Lời tựa.....</b>	<b>v</b>
---------------------	----------

## **Chương 1: Giới thiệu các khái niệm cơ bản về phân tích kinh doanh..... 1**

1.1 Mục tiêu của cuốn sách .....	1
1.2 Tiếp cận kiến thức cơ sở .....	1
1.3 Tìm hiểu kiến thức cơ sở.....	2
1.4 Phần ôn tập – bài tập .....	11
1.5 Tài liệu tham khảo chương .....	14

## **Chương 2: Ra quyết định, dữ liệu và thông tin..... 15**

2.1 Ra quyết định .....	15
2.2 Dữ liệu và thông tin .....	18
2.3 Nguồn dữ liệu .....	19
2.4 Thu thập dữ liệu .....	22
2.5 Chọn mẫu .....	23
2.6 Thông tin cần thiết để ra quyết định hiệu quả.....	29
2.7 Chất lượng thông tin .....	30
2.8 Kết luận.....	30
2.9 Tài liệu tham khảo chương .....	31

## **Chương 3: Phân tích dữ liệu; đưa ra thông tin..... 33**

3.1 Giới thiệu.....	33
3.2 Lưu trữ và cấu trúc dữ liệu .....	33
3.3 Tổ chức dữ liệu .....	35
3.4 Lập bảng biểu .....	35
3.5 Sắp xếp dữ liệu .....	36
3.6 Bảng phân phối tần số.....	39
3.7 Bảng phân phối tần số nhóm.....	41
3.8 Phân phối tần số tích lũy.....	44
3.9 Phân vị .....	45

3.10 Trình bày thông tin bằng hình ảnh .....	48
3.11 Kết luận .....	59
3.12 Tài liệu tham khảo chương .....	59

## **Chương 4: Tóm tắt dữ liệu ..... 61**

4.1 Giới thiệu.....	61
4.2 Một số khái niệm nền tảng về tóm tắt dữ liệu .....	61
4.3 Giá trị quy chuẩn hoặc trung bình.....	63
4.4 Tài liệu tham khảo chương .....	86

## **Chương 5: Đo lường và đánh giá sự biến động trong giá trị dữ liệu ..... 87**

5.1 Giới thiệu.....	87
5.2 Chênh lệch phần trăm .....	88
5.3 Chỉ số bình quân gia quyền .....	94
5.4 Giảm phát chuỗi thời gian – sử dụng chỉ số giá như chỉ số giảm phát ....	98
5.5 Tài liệu tham khảo chương .....	100

## **Chương 6: Liên kết giữa các biến số ..... 101**

6.1 Giới thiệu.....	101
6.2 Phân tích tương quan.....	102
6.3 Biểu đồ phân tán.....	104
6.4 Độ mạnh của liên hệ tuyến tính – hệ số tương quan $r$ .....	107
6.5 Phân tích hồi quy .....	109
6.6 Hệ số xác định .....	113
6.7 Ước tính sử dụng mô hình hồi quy tuyến tính.....	114
6.8 Tài liệu tham khảo chương .....	115

## **Chương 7: Chuỗi thời gian và dự báo..... 117**

7.1 Giới thiệu.....	117
7.2 Yếu tố chuỗi thời gian .....	118
7.3 Đánh giá và so sánh các mô hình.....	135
7.4 Tài liệu tham khảo chương .....	136

**Chương 8: Mô hình hóa tài chính ..... 137**

8.1 Giới thiệu.....	137
8.2 Chi phí, doanh thu, lợi nhuận và số dư đảm phí.....	137
8.3 Minh họa mô hình tài chính bằng đồ thị .....	140
8.4 Liên hệ phi tuyến tính .....	146
8.5 Các kiểu liên hệ hữu dụng khác .....	151
8.6 Tài liệu tham khảo chương .....	152

**Chương 9: Phân tích tài chính tiền tệ - lãi suất và giá trị thời gian ..... 153**

9.1 Giới thiệu.....	153
9.2 Lãi suất.....	154
9.3 Ứng dụng tính lãi suất .....	159
9.4 Giá trị hiện tại .....	164

**Chương 10: Kết luận và cân nhắc ..... 171**

# LỜI TỰA

Ngày nay, mỗi chuyên gia thực sự phải dựa vào một lượng dữ liệu có tổ chức để đưa ra những quyết định đúng đắn và hiệu quả. Chính lượng dữ liệu này khiến họ cần có các phương pháp phân tích kinh doanh để nắm vững và vận dụng nhiều tập dữ liệu lớn, nhằm rút ra thông tin hỗ trợ việc đưa ra quyết định. Hai khía cạnh phân tích kinh doanh trên giúp họ thu thập được một lượng thông tin đáng kể và hữu ích từ các tập dữ liệu dạng số:

- Kiến thức cập nhật về các kỹ thuật phân tích dữ liệu,
- Khả năng sử dụng phần mềm phân tích dữ liệu để thực hiện các kỹ thuật phân tích dữ liệu.

Là đối tượng nghiên cứu của cuốn sách, nên thông qua các chương dưới đây, quý độc giả sẽ gặp hàng loạt kỹ thuật phân tích dữ liệu, cùng vô số ví dụ về triển khai các phương pháp này trong phần mềm phân tích dữ liệu. Trong cuốn sách này, Microsoft Excel là phần mềm phân tích dữ liệu theo chuẩn chuyên ngành và được sử dụng trong rất nhiều tổ chức. Tuy nhiên, do tính năng phân tích dữ liệu dạng số khá tương đồng trong nhiều phần mềm phân tích dữ liệu (như Open Office, Google Apps), nên chúng ta có thể trực tiếp sử dụng các ví dụ thực hành mà cuốn sách đưa ra với nhiều phần mềm đa dạng khác nhau.

Cuốn sách trình bày nhiều ví dụ thực hành với các tập dữ liệu thực tế nhằm phục vụ quá trình phân tích kinh doanh. Sau khi được nhóm tác giả sàng lọc qua thời gian, các ví dụ này được sử dụng để hỗ trợ nhiều học viên muốn phát triển sự nghiệp trong lĩnh vực kinh doanh / marketing / nhân sự / kế toán và tài chính.

Cuốn sách hướng tới các mục tiêu:

- Khuyến khích một hướng tiếp cận có cấu trúc đối với quá trình phân tích kinh doanh trong các tổ chức,
- Phát triển nhận thức và năng lực sử dụng các khuôn mẫu và mô hình phân tích kinh doanh có cấu trúc, nhằm rút ra thông tin cần thiết từ các tập dữ liệu,
- Thể hiện rõ giá trị gia tăng và hiệu quả phân tích khi áp dụng các hướng tiếp cận và phần mềm phân tích dữ liệu quy củ trong việc xử lý các tập dữ liệu có tổ chức.

Học xong cuốn sách này, bạn sẽ có thể:

- Áp dụng các kỹ thuật và mô hình phân tích dữ liệu thích hợp để xử lý các tập dữ liệu có tổ chức,
- Đánh giá các tùy chọn để tìm ra kỹ thuật có thể áp dụng với tình hình thực tế của tổ chức,
- Sử dụng công cụ phần mềm thích hợp để phân tích dữ liệu từ các tập dữ liệu có tổ chức.

Cuốn sách được thiết kế nhằm giới thiệu các khái niệm và ví dụ về nghiên cứu kinh doanh với độ khó tăng dần, bắt đầu từ các khái niệm nhập môn và cao nhất là phân tích tài chính. Đây cũng là cách mà nhóm tác giả khuyên bạn nên tìm hiểu. Tuy nhiên, tùy vào đặc tính riêng của từng ví dụ cũng như khả năng của bản thân, hãy thoải mái làm theo cách mà bạn cảm thấy tự nhiên nhất để đạt được mục đích học tập.

Với tiêu chí đó, nhóm tác giả chúc bạn có một chuyến đồng hành thú vị cùng cuốn sách. Dù là người mới bắt đầu hay đã dày dặn kinh nghiệm trong phân tích kinh doanh, thì các ví dụ mà sách cung cấp cũng sẽ nâng cao năng lực phân tích dữ liệu của bạn một cách đáng kể.



## Chương 1

# GIỚI THIỆU CÁC KHÁI NIỆM CƠ BẢN VỀ PHÂN TÍCH KINH DOANH

## 1.1 MỤC TIÊU CỦA CUỐN SÁCH

Hiện nay, trong môi trường làm việc chuyên nghiệp tại các tổ chức luôn tồn tại một lượng dữ liệu và thông tin dồi dào mà họ cần giải quyết. Bởi thế, đưa ra quyết định có cơ sở là một kỹ năng có giá trị nhất trong gia tài công cụ quản lý của bất cứ người quản lý chuyên nghiệp nào. Vì vậy, việc hiểu rõ dữ liệu dạng số thể hiện các khía cạnh của một tổ chức là cần thiết, và đây là mục tiêu mà cuốn sách này sẽ giúp bạn đạt được.

## 1.2 TIẾP CẬN KIẾN THỨC CƠ SỞ

Để sử dụng cuốn sách một cách hiệu quả nhất nhằm nâng cao kiến thức và kỹ năng của học viên, chúng tôi giả định rằng bạn đã thông thuộc các khái niệm toán học phổ thông. Nếu đã học toán hay thống kê ở cấp độ trên phổ thông, thì bạn sẽ thấy một số khía cạnh của quyển sách dễ giải quyết hơn hẳn. Hãy xem điểm cộng này như một cách để phát triển bản thân và công việc, chứ đừng lấy đó là một lý do để tự tin thái quá. Lợi ích của việc ở trên cấp độ phổ thông nằm ở mặt tự học – bạn sẽ có thể nắm bắt nội dung trong cuốn sách nhanh chóng hơn. Tuy nhiên, bạn vẫn nên tìm hiểu kỹ, đặc biệt chú ý đến những khía cạnh mà bản thân cảm thấy cần phát triển sâu hơn.

Để nhớ lại và vận dụng được kiến thức, kỹ năng cấp phổ thông (hoặc tương đương), hãy tìm hiểu các khái niệm dưới đây và đảm bảo bạn nắm rõ, cũng như có thể sử dụng tất cả. Nếu có nội dung nào được đề cập mà bạn không hiểu, thì nên áp dụng ngay nhiều biện pháp để khắc phục vấn đề. Nếu cần tìm hiểu sâu và chi tiết hơn một số khái niệm được minh họa trong sách, bạn nên tham khảo thư mục mở rộng ở cuối chương này.

Hãy nhớ rằng chìa khóa để làm chủ các khái niệm được minh họa trong sách này là thực hành – càng thực hành nhiều, bạn càng dễ nhớ và sử dụng chúng. Nội dung cuốn sách được trình bày đan xen giữa lý thuyết và ví dụ thực hành (bao gồm nhiều ví dụ sử dụng Microsoft Excel). Chúng tôi đặc biệt khuyến khích học viên thực hành các ví dụ đi kèm trên Microsoft Excel, do cuốn sách không thể hiện được điều đó. Cuốn sách sẽ giới thiệu một loạt mô hình – được minh họa bằng các công thức / phương trình / ví dụ Microsoft Excel giúp các chuyên gia và quản lý xử lý có hiệu quả các tập dữ liệu kinh doanh.

### 1.3 TÌM HIỂU KIẾN THỨC CƠ SỞ

Trong các phần tiếp theo, chúng ta sẽ tìm hiểu một loạt khái niệm cần thiết để làm chủ việc phân tích kinh doanh và ứng dụng các khái niệm này trong tình hình thực tế của tổ chức.

#### 1.3.1 KÝ HIỆU

Bảng sau đây chứa một loạt ký hiệu toán học sẽ được sử dụng xuyên suốt và bất cứ ai đọc sách cũng cần nắm rõ.

+	cộng	-	trừ
=	bằng	≡	đồng dư
>	lớn hơn	±	cộng hoặc trừ
<	nhỏ hơn	√	căn bậc hai
≥	lớn hơn hoặc bằng	%	phần trăm
≤	nhỏ hơn hoặc bằng	a, b, x, y, ...	ký hiệu minh họa biến số hoặc khái niệm số khác

**Bảng 1.1** Ký hiệu toán học thường được sử dụng.

#### 1.3.2 THUẬT NGỮ CHÍNH

Bên cạnh ký hiệu toán học, cuốn sách cũng sử dụng một loạt từ khóa thể hiện các thuật ngữ cụ thể liên quan đến phân tích kinh doanh. Bảng dưới đây giới thiệu và đưa ra giải thích sơ lược về từng từ khóa này.

trung bình cộng	giá trị trung bình thu được khi cộng nhiều số với nhau, rồi lấy kết quả chia cho tổng lượng số được cộng vào. Ví dụ: trung bình cộng của 3, 4 và 5 là $(3 + 4 + 5) / 3 = 4$ .
trục	đường thẳng tham chiếu cố định, có thể được dùng khi giao với một hoặc nhiều đường tham chiếu tương tự để xác định vị trí của một điểm hoặc một tập hợp điểm tạo thành một đường thẳng, đường cong hoặc mặt phẳng. Ví dụ: trục x, trục y, trục z, ...
hệ số	số được đặt trước và nhân với một số khác (thường là biến số). Ví dụ: $5 * x$ , $7 * b$ , $23 * x * y$

hằng số	số (lượng) được giả định là không đổi trong một phép toán. Ví dụ: 3, 5.87, -3.2
tọa độ	sự biểu diễn bằng số nhằm xác định vị trí của một điểm, đường thẳng hoặc mặt phẳng trên một hệ trục. Ví dụ: (3,5) trong không gian hai chiều (X,Y)
thập phân	vị trí của một chữ số sau dấu thập phân. Ví dụ: trong số 0,0247, chữ số 4 nằm ở hàng thập phân phần nghìn (vị trí thập phân thứ ba)
phương trình	biểu thức giả định đẳng thức giữa hai đại lượng. Ví dụ: $2 + 3 * x = 26$
công thức	quy tắc được thể hiện bằng các ký hiệu đại số. Công thức được dùng để tính toán nhiều kết quả số khác nhau, tùy theo các giá trị đầu vào. Ví dụ: $x + 3 * y$ , trên thực tế x và y cần được gán các giá trị số nhất định để tính được giá trị số của công thức
tần số	số lần xảy ra hoặc xuất hiện trong một đơn vị thời gian cụ thể. Ví dụ: 34 lần, 4 lần
phân phối tần số	tập hợp dãy tần số được quan sát hoặc đo lường có liên hệ với nhiều thể loại đối tượng khác nhau. Ví dụ: danh sách chứa tầm giá của sản phẩm kem chống nắng kèm số lượng có sẵn
số nguyên	số không chứa giá trị thập phân. Ví dụ: 1, 15, 3467, ...
phương trình tuyến tính	phương trình đa thức bậc nhất (không chứa biến số nào có lũy thừa lớn hơn 1). Ví dụ: $x - y = 25$ , $x + 3 = 68 * y$
số trung vị	số nằm giữa trong một dãy số. Nếu dãy có lượng số là chẵn, thì số trung vị được tính là trung bình của hai số nằm giữa. Ví dụ: số trung vị của dãy số 1, 5, 6, 7, 9 là 6. Số trung vị của dãy số 1, 2, 5, 6, 7, 9 là $(5 + 6) / 2 = 5,5$

phần trăm	tỷ lệ thường được tính trên một trăm, thể hiện tương quan của một phần với tổng số cho trước. Ví dụ: một phần tư của một số được tính là 25 phần trên tổng số 100 phần của số đó, tức là $25/100$ số đó. Tỷ lệ này thường được biểu diễn là 25%. Đôi khi tỷ lệ có thể được biểu diễn dưới dạng giá trị nằm giữa 0 và 1, ví dụ 25% sẽ được biểu diễn là 0,25.
khoảng biến thiên	hiệu số giữa giá trị cao nhất và thấp nhất của một dãy số. Ví dụ: khoảng biến thiên của dãy số 1, 2, 5, 6, 7, 9 là $9 - 1 = 8$
chữ số có nghĩa	tất cả chữ số khác 0 trong một số, chữ số 0 nằm ở giữa và nằm cuối biểu thị chính xác một số cho trước. Ví dụ: với số 0,025600 thì chữ số có nghĩa là 256 và hai chữ số 0 ở cuối thể hiện chính xác số thập phân đến hàng triệu.
biến số	đại lượng có thể nhận bất kỳ giá trị nào từ một khoảng giá trị nhất định. Ví dụ: biến số nguyên $x$ có thể nhận bất kỳ giá trị nguyên nào, như 3, 76, 8, 2456, ...

**Bảng 1.2** Thuật ngữ thông dụng.**1.3.3 SỐ THẬP PHÂN VÀ SỐ CÓ NGHĨA**

Phần này giải thích cách thích hợp để viết số thập phân hoặc số có nghĩa.

Nếu đáp án của một phép tính là 53,025602, chắc hẳn bạn sẽ không viết lại toàn bộ con số này. Thay vào đó, bạn sẽ “làm tròn” số. Có hai cách để làm điều này. Bạn có thể làm tròn đến một số nhất định:

- số thập phân, hoặc
- số có nghĩa

Khi đó ta viết 53,025602 đến 1 chữ số thập phân là 53,0 – viết 1 chữ số sau dấu thập phân. Viết 53,025602 đến 4 chữ số thập phân là 53,0256 – viết 4 chữ số sau dấu thập phân.

Tuy nhiên, viết 53,025602 đến 1 chữ số có nghĩa là 50 – bắt đầu từ bên trái, ta viết 1 chữ số (bắt buộc viết cả số 0 để giữ nguyên “giá trị hàng” – chỉ rõ rằng số này là 50 chứ không phải 5). Viết 53,025602 đến 4 chữ số có nghĩa là 53,03 – bắt đầu từ bên trái, ta viết 4 chữ số thì được kết quả là 53,02; tuy nhiên chữ số tiếp đó là 5, nên ta giả định là số này gần với 53,03 hơn 53,02.

Quy tắc là: khi làm tròn số, nếu chữ số ngay sau vị trí dừng lại lớn hơn hoặc bằng 5, thì thêm 1 vào chữ số cuối cùng mà bạn viết.

Nếu kết quả của một phép tính là 0,00256023164, ta sẽ “làm tròn” kết quả đó. Số này được làm tròn đến 5 chữ số thập phân là 0,00256. Bạn viết 5 chữ số sau dấu thập phân.

Để làm tròn số này đến 5 chữ số có nghĩa, ta phải viết 5 số. Tuy nhiên, do số này nhỏ hơn 1 nên ta không tính đến các số 0 ở đầu. Như vậy, viết số này với 5 chữ số có nghĩa, ta sẽ được 0,0025602 (5 chữ số tính từ chữ số khác 0 đầu tiên).

Nếu làm tròn 4,909 đến 2 chữ số thập phân, ta sẽ được 4,91.

Làm tròn 3,486 đến 3 chữ số có nghĩa, kết quả sẽ là 3,49.

Làm tròn 0,0096 đến 3 chữ số thập phân là 0,010 (thêm 1 vào 9, thành 10). Khi làm tròn đến một số thập phân nhất định, ta luôn viết bất cứ chữ số 0 nào ở cuối. Nếu yêu cầu là làm tròn đến 3 chữ số thập phân, ta viết 3 chữ số thập phân, bất kể chữ số cuối cùng là số 0.

Khi cần, bạn có thể tham khảo thêm tài liệu về chủ đề ở cuối chương.

### 1.3.4 PHÂN TRĂM

Trong phần này, chúng ta sẽ xét đến việc đánh giá mức tăng/giảm tỷ lệ phần trăm và cách ứng dụng phần trăm trong nhiều ứng dụng.

Bạn còn nhớ khái niệm về phần trăm chứ? Phần trăm là:

“tỷ lệ hoặc phần của một trăm phần”

Ví dụ, để tính phần trăm của một giá trị số, ta cần nhân phần trăm với giá trị đó rồi chia cho 100. Dĩ nhiên phép tính phần trăm khá dễ thực hiện nhờ có máy tính và các phần mềm văn phòng khác (như Microsoft Excel).

Giả sử chúng ta muốn tính 3% của 225:

Kết quả là  $(3 * 225) / 100 = 675 / 100 = 6,75\%$

Một cách khác để tính phần trăm là chia giá trị số mà chúng ta muốn tính phần trăm cho 100 để được 1% của giá trị số này, rồi nhân kết quả đó với số phần trăm muốn tính.

Sử dụng phương pháp này, ta tính 3% của 225 như sau:

$225 / 100 = 2,25$  – tức 1% của 225. Do đó 3% của 225 là  $2.25 * 3 = 6,75\%$

Vẫn còn một cách khác để tính phần trăm: biểu thị 3% là  $3 / 100 = 0,03$  rồi nhân số này với giá trị chúng ta đang muốn tính phần trăm. Ví dụ trên sẽ cho kết quả  $0,03 * 225 = 6,75\%$ .

Trong môi trường kinh doanh, khi tính phần trăm, ta cần tính toán sự thay đổi phần trăm với mục đích so sánh hoặc ra quyết định.

**Ví dụ:** Mức lương hàng năm của nhân viên trong một công ty đã thay đổi từ 20.000£ (1£ xấp xỉ 30.000 VNĐ) lên 25.000£. Vậy mức lương của người này đã tăng lên bao nhiêu phần trăm?

**Lời giải:** Phần trăm thay đổi được tính bằng công thức sau:

$$[(\text{Giá trị mới} - \text{Giá trị cũ}) / \text{Giá trị cũ}] * 100$$

Trong ví dụ này, Giá trị mới = 25.000£, Giá trị cũ = 20.000£

Do đó phần trăm tăng lên sẽ là:

$$[(25.000 - 20.000) / 20.000] * 100 = (5.000 / 20.000) * 100 = 0,25 * 100 = 25\%$$

Như vậy, tỷ lệ phần trăm tăng lương của người này = 25%.

**Bài tập:** Một sản phẩm có giá 25,02£ tính cả 20% VAT (T/N: Thuế giá trị gia tăng). Giá của sản phẩm này không bao gồm VAT là bao nhiêu?

**Lời giải:**

Giá chưa VAT + VAT = Giá đã gồm VAT

Giá chưa VAT + Giá chưa VAT \* 0,2 = Giá đã gồm VAT

Giá chưa VAT \* (1 + 0,2) = Giá đã gồm VAT

Giá chưa VAT = Giá đã gồm VAT / (1 + 0,2)

Giá chưa VAT = 25,02 / 1,2

Giá chưa VAT = 20,85£

### 1.3.5 PHƯƠNG TRÌNH TUYẾN TÍNH VÀ ĐỒ THỊ

Phần này sẽ thảo luận về phương trình tuyến tính và đồ thị.

Trong phân tích kinh doanh, các kỹ năng sau rất quan trọng:

- nhận biết dạng phương trình tuyến tính, ví dụ phương trình một đường thẳng có dạng  $y = a + b * x$  hoặc  $y = m * x + n$ , trong đó a, b, m, n là hằng số. Ta gọi phương trình là tuyến tính vì khi được biểu diễn trên đồ thị thì các giá trị của phương trình sẽ tạo thành một đường thẳng
- vẽ đồ thị phương trình tuyến tính

Trong phân tích kinh doanh, việc nhận biết được phương trình tuyến tính rất quan trọng, ví dụ để mô tả một xu hướng; nội dung này sẽ được thảo luận trong phần sau của cuốn sách này.

Giờ, giả sử ta sử dụng công thức sau cho một phương trình tuyến tính:

$$y = 3 + 2 * x$$

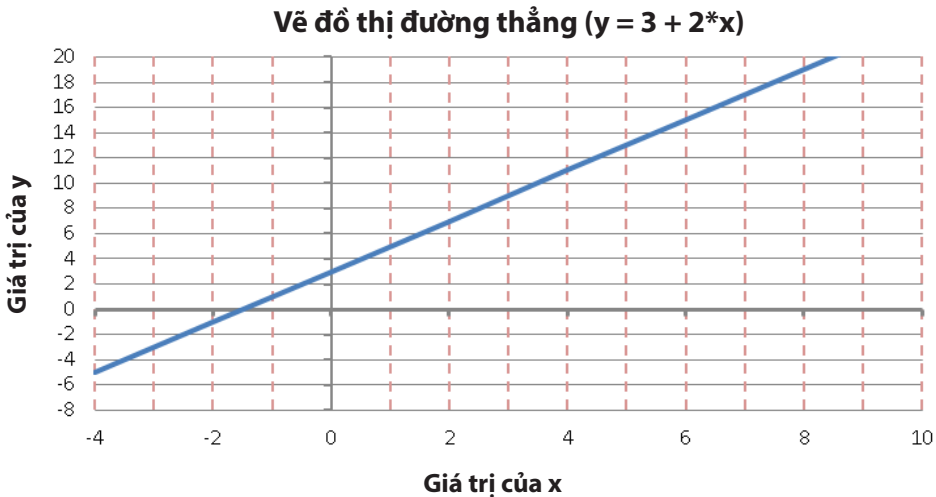
Để vẽ đồ thị cho phương trình này, ta cần chọn ít nhất hai giá trị cho  $x$  và tính giá trị  $y$  tương ứng (trên thực tế hai giá trị cho  $x$  là đủ, vì có thể xác định chính xác một đường thẳng trong không gian hai thông qua 2 tọa độ, tất cả các giá trị khác của cùng phương trình sẽ nằm trên chính đường thẳng đó). Do vậy chúng ta sẽ vẽ được đường thẳng biểu thị cho công thức trên.

Cách minh họa tốt nhất là tạo ra một bảng chứa các giá trị  $x$  và  $y$ . Giả sử  $x$  là một giá trị nguyên, ta có thể chọn hai giá trị nguyên cho  $x$  và thu được giá trị  $y$  tương ứng như sau:

Giá trị $x$	Giá trị $y$
2	$3 + 2 * 2 = 3 + 4 = 7$
4	$3 + 2 * 4 = 3 + 8 = 11$

**Bảng 1.3** Ví dụ về tính các giá trị trên một đường thẳng.

Kết quả của công thức tạo thành đường thẳng dưới đây, được biểu thị từ hai tọa độ (2, 7) và (4, 11) trong bảng:



**Hình 1.1** Vẽ đồ thị đường thẳng.

Chúng ta thấy đường thẳng này cắt trục  $x$  ở tọa độ -1,5; trong khi nếu  $x = 0$ , thì  $y = 3$ . Các tính chất này của đường thẳng sẽ được cuốn sách đề cập thêm, do chúng rất quan trọng trong việc xác định các đặc điểm cụ thể của đường thẳng.

Trong đường thẳng có dạng  $y = a + b \cdot x$ , thì  $a$  được gọi là hệ số chặn  $y$  vì đây là giá trị của  $y$  tại điểm mà đường thẳng cắt trục  $y$  và  $b$  được gọi là hệ số góc của đường thẳng do nó cho biết số đo về độ dốc của đường thẳng.

### 1.3.6 TÍNH GIÁ TRỊ TRUNG BÌNH

Giá trị trung bình là khái niệm quan trọng được sử dụng rộng rãi trong phân tích kinh doanh để đánh giá các chỉ số trong một tổ chức. Do đó, việc nắm rõ các khái niệm này rất quan trọng. Cách tốt nhất để làm được điều này là thực hành nhiều phương pháp tính giá trị trung bình khác nhau.

Dựa trên các khái niệm được đề cập trong mục 1.3.2, hãy làm các bài tập sau đây:

**Bài 1.1:** tìm trung vị của dãy số sau: 1, 2, 3, 6, 7

**Lời giải:** số trung vị của dãy số là 3

**Bài 1.2:** tìm trung vị của dãy số sau: 1, 2, 3, 5, 6, 7

**Lời giải:** số trung vị của dãy số là  $(3 + 5) / 2 = 8 / 2 = 4$

**Bài 1.3:** tìm trung bình cộng của dãy số sau: 23, 34, 27, 15, 19

**Lời giải:** trung bình cộng là  $(23 + 34 + 27 + 15 + 19) / 5 = 118 / 5 = 23,6$

**Bài 1.4:** tìm trung vị của bảng phân phối tần số sau (dữ liệu lấy từ một mẫu khảo sát số trẻ em trong hộ gia đình):

Số trẻ em trong hộ gia đình	Tần số
0	11
1	15
2	24
3	17
4	9
5	7

**Bảng 1.4** Phân phối tần số.

**Lời giải:** Ta cần tính tổng điểm dữ liệu, bằng cách cộng tổng các giá trị trong cột Tần số:  $11 + 15 + 24 + 17 + 9 + 7 = 83$ . Do đó, trung vị của bảng phân phối tần số này là giá trị số trẻ em trong hộ gia đình tương ứng với giá trị 42 của cột tần số.



Số trẻ em trong hộ gia đình	Tần số	Tổng giá trị tần số đếm được
0	11	11
1	15	26
2	24	50
3	17	67
4	9	76
5	7	83

**Bảng 1.5** Phân phối tần số – tính giá trị trung vị.

Vậy số trẻ em tương ứng với giá trị 42 là 2, do giá trị 42 xuất hiện ở dòng mô tả hộ gia đình có 2 trẻ em.

**Bài 1.5:** sử dụng máy tính bỏ túi để tìm giá trị trung bình của bảng phân phối tần số sau (dữ liệu lấy từ một mẫu khảo sát số trẻ em trong hộ gia đình):

Số trẻ em trong hộ gia đình	Tần số
0	11
1	15
2	24
3	17
4	9
5	7

**Bảng 1.6** Phân phối tần số – tính giá trị trung bình.

**Lời giải:** ta cần tính thêm cột thứ ba, đó là kết quả phép nhân giữa các giá trị trong hai cột đầu tương ứng với nhau:

Số trẻ em trong hộ gia đình (n)	Tần số (f)	$n * f$
0	11	$0 * 11 = 0$
1	15	$1 * 15 = 15$
2	24	$2 * 24 = 48$
3	17	$3 * 17 = 51$

Số trẻ em trong hộ gia đình (n)	Tần số (f)	$n * f$
4	9	$4 * 9 = 36$
5	7	$5 * 7 = 35$

**Bảng 1.7** Phân phối tần số – tính giá trị trung bình.

Bây giờ ta cần cộng tổng các giá trị  $n * f$  và tìm tổng điểm dữ liệu trong bảng phân phối tần số này:

Số trẻ em trong hộ gia đình (n)	Tần số (f)	$n * f$
0	11	0
1	15	15
2	24	48
3	17	51
4	9	36
5	7	35
Tổng	$11 + 15 + 24 + 17 + 9 + 7 = 83$	$0 + 15 + 48 + 51 + 36 + 35 = 185$

**Bảng 1.8** Phân phối tần số – tính giá trị trung bình.

Do đó giá trị trung bình của bảng phân phối tần số đã cho là  $185 / 83 = 2,2289$  hoặc làm tròn đến 2 chữ số thập phân là 2,23.

### 1.3.7 NÂNG LÊN LŨY THỪA VỚI SỐ MŨ NGUYÊN DƯƠNG

Để thực hiện phép tính và đơn giản hóa phép toán, thì biểu thức dạng  $a * a * a$  có thể viết thành  $a^3$  – có nghĩa là lũy thừa bậc 3 của a.

Trong các biểu thức toán học, phép lũy thừa có dạng  $a^n$  hoặc  $x^2$ . Thông thường, số mũ n là một giá trị nguyên dương.

Việc hiểu được ý nghĩa của biểu thức an rất quan trọng, với a là một số và n là số nguyên dương, như  $a^2$ ,  $a^3$ ,  $a^4$ , ...

Ví dụ ta có  $5^2$  tương đương  $5 \times 5 = 25$  hoặc  $2^3 = 2 \times 2 \times 2 = 8$  hoặc  $10^6$  tương đương  $10 \times 10 \times 10 \times 10 \times 10 \times 10 = 1.000.000$  hay một triệu.

Lưu ý rằng  $a^0 = 1$  bất kể giá trị của a là bao nhiêu, ví dụ  $10^0 = 1$ .

## 1.4 PHẦN ÔN TẬP – BÀI TẬP

1. Biết  $y = 1/x$ , hãy viết giá trị của  $y$ , làm tròn đến 2 chữ số thập phân, khi  $x = 8$

**Lời giải:**  $y = 1/8 = 0,125$ . Vậy làm tròn 0,125 đến hai chữ số thập phân, ta được 0,13.

2. Biết doanh số bán một sản phẩm tăng 5% trong năm đầu và 8% trong năm kế tiếp, vậy tổng doanh số đã tăng bao nhiêu phần trăm trong hai năm đó? (làm tròn đến một chữ số thập phân)

**Lời giải:**

Doanh số sau năm thứ nhất =  $1,05 \times$  doanh số ban đầu

Doanh số sau năm thứ hai =  $1,08 \times (1,05 \times \text{doanh số ban đầu}) = 1,134 \times$  doanh số ban đầu

Do đó tổng cộng doanh số tăng 13,4% ( $0,134 \times 100$ )

3. Tính giá trị trung bình của dãy bao gồm 11 số dưới đây. Làm tròn kết quả đến 2 chữ số thập phân.

7, 3, 2, 3, 8, 6, 7, 0, 9, 2, 4

**Lời giải:** Để tìm giá trị trung bình của dãy số, chia tổng các số cho số các số trong dãy.

Có tất cả 11 số và tổng của chúng là 51. Do đó giá trị trung bình bằng  $51 / 11 = 4,636363 = 4,64$  khi làm tròn đến 2 chữ số thập phân.

4. Tính trung vị của dãy 11 số trong bài tập 3.

**Lời giải:** Để tìm trung vị của một dãy số, trước hết ta sắp xếp chúng theo thứ tự từ số nhỏ nhất. Giá trị trung vị là số nằm giữa dãy số.

Dãy số sắp xếp theo thứ tự là: 0, 2, 2, 3, 3, 4, 6, 7, 7, 8, 9

Vậy giá trị nằm giữa dãy này là 4.

5. Hãy chia 25.000£ cho ba người theo tỷ lệ 5 : 3 : 2. Giá trị của phần lớn nhất là bao nhiêu?

**Lời giải:** Ta cần tìm tổng số lượng phần bằng nhau  $5 + 3 + 2 = 10$ .

Sau đó ta chia 25.000£ thành 10 phần bằng nhau:  $25.000£ / 10 = 2.500£$

Do đó phần lớn nhất sẽ là  $2.500£ \times 5 = 12.500£$

6. Tổng hóa đơn, tính cả VAT 20%, là 301,25£ – đã làm tròn đến chữ số thập phân hàng trăm. Nếu chưa tính VAT thì hóa đơn này là bao nhiêu, khi làm tròn đến chữ số thập phân hàng trăm?

**Lời giải:**

Hóa đơn chưa VAT =  $b$

Hóa đơn đã tính VAT = 301,25 =  $b * (1 + 0,2) = b * 1,2$

Nên  $1,2 * b = 301,25$  suy ra  $b = 301,25 / 1,2 = 251,041666 = 251,04£$  làm tròn đến chữ số thập phân hàng trăm

7. Một công ty vận tải hàng hóa đường dài đang mở rộng phạm vi hoạt động từ trong Liên minh Châu Âu (EU) đến Balkans (bán đảo nằm ở đông nam châu Âu). Một trong những tuyến đường mới chạy từ Frankfurt (Đức) đến Belgrade (Serbia) đi qua Ljubljana (Slovenia). Khoảng cách từ Frankfurt đến Ljubljana là 725km và từ Ljubljana đến Belgrade là 445km. Tính tổng khoảng cách từ Frankfurt đến Belgrade bằng đơn vị mile (dặm Anh) biết 1 mile = 1,609km. Làm tròn kết quả đến số nguyên gần nhất.

**Lời giải:** Tổng khoảng cách từ Frankfurt đến Belgrade tính theo km là  $725 + 445 = 1170$  km

1 mile = 1.609 km nên 1 km =  $1/1.609$  mile

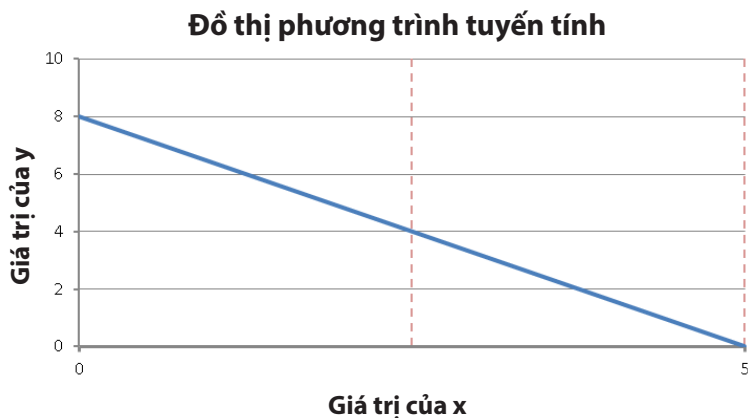
Do đó tổng khoảng cách này tính theo mile là  $1170 \text{ km} * 1 / 1,609 = 1170 / 1,609 = 727,159726 \text{ mile} = 727 \text{ mile}$  (đã làm tròn đến số nguyên gần nhất)

8. Tính khoảng cách từ Ljubljana đến Belgrade dưới dạng tỷ lệ phần trăm trên tổng khoảng cách của cả tuyến đường. Viết kết quả dưới dạng số có 3 chữ số có nghĩa. Sử dụng thông tin đã cung cấp ở Bài tập 7.

**Lời giải:** Khoảng cách từ Ljubljana đến Belgrade là 445 km. Tổng khoảng cách là 1170 km.

Tỷ lệ phần trăm của 445 so với 1170 là  $445 / 1170 \times 100 = 0,380341 \times 100 = 38,0341$ . Do đó câu trả lời là 38,0 với 3 chữ số có nghĩa.

9. Tính giá trị của hàm số chặn y trên đồ thị dưới đây.



**Hình 1.2** Đồ thị của một phương trình tuyến tính.

**Lời giải:** Hệ số chặn y là giá trị của y khi đồ thị cắt trục y (trục tung). Do đó hệ số này là 8.

10. Sử dụng đồ thị trong bài tập trên, hãy viết phương trình đường thẳng dưới dạng  $y = a + b \cdot x$

**Lời giải:** Phương trình đường thẳng có thể viết dưới dạng  $y = a + b \cdot x$  với a và b là hằng số.

Khi  $x = 0$  (tại điểm nằm trên trục y) thì  $y = 8$ , do đó  $a = 8$ . Đây là hệ số chặn của đường thẳng. Tại đây, ta biết phương trình này là  $y = 8 + b \cdot x$

Khi  $y = 0$  (tại điểm nằm trên trục x) thì  $x = 5$ , do đó  $0 = 8 + 5 \cdot b$

Như vậy  $5 \cdot b = -8$ , suy ra  $b = -8 / 5 = -1,6$ . Đây là hệ số góc của đường thẳng. Do đó phương trình đầy đủ của đường thẳng này là  $y = 8 - 1,6 \cdot x$ .

## 1.5 TÀI LIỆU THAM KHẢO CHƯƠNG

1. Morris, C, Quantitative Approaches in Business Studies (*tạm dịch*: Phương pháp định lượng trong nghiên cứu kinh doanh), tái bản lần thứ 8, FT/Prentice Hall, 2012 (eBook, tái bản lần thứ 7, 2008), Chương 1.
2. Oakshott, L, Essential Quantitative Methods for Business (*tạm dịch*: Phương pháp nghiên cứu định lượng thiết yếu trong kinh doanh, quản lý & tài chính), tái bản lần thứ 5, Palgrave Macmillan, 2012, Chương 1.
3. Waters, D, Quantitative Methods for Business (*tạm dịch*: Phương pháp định lượng trong kinh doanh), tái bản lần thứ 5, FT/Prentice Hall, 2011, (có bản eBook), Chương 1 và 2.
4. Wisniewski, M, Quantitative Methods for Decision Makers (*tạm dịch*: Phương pháp định lượng hỗ trợ người ra quyết định), tái bản lần thứ 5, FT/Prentice Hall, 2009 (có bản eBook), Chương 2 và 10.

## Chương 2

# RA QUYẾT ĐỊNH, DỮ LIỆU VÀ THÔNG TIN

## 2.1 RA QUYẾT ĐỊNH

Quá trình điều hành một doanh nghiệp hay tổ chức phức tạp nào cũng đều liên quan đến các lựa chọn. Bất cứ lúc nào, quyết định mới sẽ được đưa ra và có ảnh hưởng tới một vài khía cạnh của doanh nghiệp cũng như khách hàng, nhà cung ứng, nhân viên hay cổ đông của tổ chức. Các nhà quản lý thường đảm nhiệm một số vai trò, mà phần nhiều trong đó liên quan đến việc đưa ra quyết định.

Nhìn chung, *ra quyết định (decision making)* là việc đưa ra một lựa chọn trong các phương hướng hành động, dựa trên bộ tiêu chí hoặc quy tắc nhất định.

### 2.1.1 CHỨC NĂNG QUẢN LÝ

Các chức năng quản lý có thể được phân loại cụ thể như sau:

- lập kế hoạch
- tổ chức và điều phối
- dẫn dắt và thúc đẩy
- kiểm soát tiến trình

**Lập kế hoạch (planning)** là quá trình quyết định trước việc sẽ làm và cách thức thực hiện (ví dụ xây dựng nhà máy mới ở đâu, đầu tư bao nhiêu cho marketing). Quá trình này được thể hiện thành các kế hoạch (phương hướng hành động được định trước) phản ánh mục tiêu và mục đích hoạt động của tổ chức. Rõ ràng quá trình này có liên quan đến việc đưa ra quyết định.

**Tổ chức và điều phối (organising and co-ordinating)** nhân lực, tài nguyên, tài liệu để thực thi kế hoạch nhằm hoàn thành mục tiêu. Đôi khi chức năng này liên quan đến việc lựa chọn giữa các phương hướng hành động, nhân sự hay tài nguyên khác nhau để hoàn thành công việc.

**Dẫn dắt và thúc đẩy (leading and motivating)** nhân sự có thể liên quan đến việc đưa ra lựa chọn giữa nhiều cách thức quản lý và hoạt động trong những thời điểm khác nhau.

**Kiểm soát tiến trình (*controlling the process*)** giúp đảm bảo công việc được thực thi đúng theo kế hoạch. Chúng ta thường kiểm soát bằng cách so sánh hiệu suất thực tế với mục tiêu tương ứng và dựa vào chênh lệch để điều chỉnh cách thức hoạt động, nhờ đó mang lại kết quả mong muốn (ví dụ nếu nhiệt độ trong văn phòng quá thấp, thì tăng nhiệt hệ thống sưởi lên). Cứ khi nào cần lựa chọn một phương thức hoạt động, thì ta sẽ phải ra quyết định.

### 2.1.2 QUYẾT ĐỊNH XÁC ĐẮNG

Việc đưa ra một quyết định đúng đắn (hay ít nhất là phù hợp) không chỉ cần tới sự phán đoán, mà còn đòi hỏi cơ sở để đưa ra lựa chọn ấy. Nhiều quyết định được đưa ra dựa trên **thông tin (*information*)** có liên quan và ảnh hưởng đến nó. Quyết định về lượng sản phẩm đặt hàng tuần này phụ thuộc tương đối vào lượng sản phẩm bán được tuần trước hoặc cùng kỳ năm ngoái. Lựa chọn công việc của một cá nhân phải dựa vào chất lượng và thành tựu đã đạt được. Thông tin cần thiết cho quá trình ra quyết định sẽ được thảo luận chi tiết hơn ở các phần sau.

### 2.1.3 RA QUYẾT ĐỊNH TRONG PHẠM VI TỔ CHỨC

Trong phạm vi tổ chức, các quyết định được đưa ra ở nhiều cấp độ:

**quyết định vận hành hay giao dịch (*operational or transactional decision*)**

được thực hiện bởi quản lý cấp cơ sở hoặc nhân viên

ảnh hưởng ngay lập tức đến sự vận hành của tổ chức (hoặc khối)

Ở cấp độ này thường có những vấn đề định kỳ cần giải quyết. Ví dụ:

- lịch công việc hàng tuần đối với một dây chuyền sản xuất cụ thể
- lịch bảo dưỡng máy móc hàng tuần
- kiểm tra kho vật liệu thô hàng ngày

Thông tin phải chính xác, thường không thuộc phạm vi tài chính mà liên quan đến một chính sách đã được cấp quản lý cao hơn quy định.

**quyết định chiến thuật (*tactical decision*)**

được thực hiện bởi quản lý cấp trung

ảnh hưởng đến quá trình hoạt động trung hạn của tổ chức (hoặc bộ phận)

Quản lý cấp trung có trách nhiệm đưa ra các quyết định thường xuyên hoặc định kỳ (hàng năm, hàng quý, hàng tháng). Quyết định chiến thuật thường ngắn hạn, áp dụng cho các chu kỳ kế hoạch khoảng một năm. Các quyết định này chủ yếu đòi hỏi thông tin liên quan đến lịch sử doanh nghiệp (ví dụ hồ sơ doanh nghiệp) hoặc tình trạng tài chính trong phạm vi tổ chức. Ví dụ:



- ngân sách tuyển dụng nhân lực cho năm tài chính tiếp theo
- chi phí quảng cáo quý sau
- mục tiêu bán hàng quý sau

### ***quyết định chiến lược (strategic decision)***

được thực hiện bởi quản lý cấp cao

ảnh hưởng đến quá trình phát triển dài hạn của tổ chức

Đây là cấp quản lý điều hành cao nhất, chủ yếu chịu trách nhiệm với các vấn đề liên quan đến kế hoạch trường kỳ. Ví dụ:

- tầm vóc của tổ chức trong 10 năm tới?
- số lượng dây chuyền sản xuất trong 5 năm tới?
- nên áp dụng chính sách nghiên cứu và phát triển nào?
- cách phát triển mẫu sản phẩm trong 20 năm tới?

Đối với dạng ra quyết định này, ban quản lý cần tiếp cận với mọi nguồn thông tin nội bộ, cũng như thông tin bên ngoài liên quan. Thông tin này không được sử dụng thường xuyên, nghĩa là các quyết định như thế này không mang tính định kỳ.

Không chỉ phục vụ các cấp quản lý, thông tin còn phục vụ một nhóm đối tượng liên quan đến tổ chức khác, như:

- Nhân viên nói chung cần thông tin về lương, sự phát triển của công ty, cải thiện về cơ sở vật chất...
- Cổ đông yêu cầu thông tin về hiệu quả hoạt động hiện tại và mục tiêu trong thời gian tới của công ty.

## **2.1.4 PHÂN LOẠI QUYẾT ĐỊNH**

Một số quyết định được thực hiện định kỳ và số khác thì bất thường hơn.

***Quyết định được lập trình (programmed decision)*** thường lặp lại định kỳ, đi kèm các tùy chọn rõ ràng và quy tắc quyết định biết trước (ví dụ dừng lại khi đèn đỏ; luôn chất hàng hóa có kích cỡ lớn nhất lên xe tải trước). Loại quyết định này thường có xu hướng liên quan đến các cấp vận hành trong tổ chức.

***Quyết định không lập trình (non-programmed decision)*** thì bất thường hơn và phi cấu trúc, đi kèm các tùy chọn phức tạp và quy tắc quyết định không rõ ràng (ví dụ làm gì khi một thiết bị phát sinh lỗi lạ, cách ứng phó tốt nhất khi một đối thủ mới xuất hiện trên thị trường). Loại quyết định này có xu hướng được thực hiện ở cấp chiến thuật hoặc chiến lược.

## 2.2 DỮ LIỆU VÀ THÔNG TIN

Hãy xem những gì xảy ra ở một công ty lữ hành.

Công ty phải đưa ra một số **quyết định (decision)**, như:

- đặt bao nhiêu chỗ trên chuyến bay thuê bao cho dịp lễ này năm tới
- gửi sách quảng cáo mới nhất cho những ai
- quảng cáo vào các dịp lễ nào

**Thông tin (information)** hỗ trợ việc ra các quyết định này bao gồm:

- số người đã đặt chỗ trong một dịp lễ
- số nơi chưa được đặt chỗ
- bảng so sánh nhu cầu cho dịp lễ này trong ba năm gần đây
- lợi nhuận từ du lịch trọn gói
- số ngày lễ mỗi gia đình đi nghỉ

Thông tin này đến từ dữ liệu được thu thập thường xuyên trong quá trình hoạt động:

- chi tiết các gói du lịch trọn gói (giá cả, địa điểm, thời gian)
- chi tiết đặt chỗ du lịch trọn gói (số lượng người, ngày khởi hành, dịp lễ được chọn, số tiền chi trả)

Mặc dù được sử dụng thay thế lẫn nhau, song về cơ bản thì hai thuật ngữ **dữ liệu (data)** và **thông tin (information)** có nghĩa tách biệt rõ ràng.

**Dữ liệu (data)** là dữ kiện thô, chưa được sắp xếp và thường không liên quan đến nhau. Dữ liệu thường có dạng số (định lượng) nhưng không bắt buộc là vậy. Dữ liệu có thể không là dạng số (định tính).

Ví dụ về dữ liệu:

- một thiết bị cụ thể đã bị hỏng 4 lần vào tuần trước,
- có 13 nhân viên trong Phòng Kế toán,
- ngân sách cho Phòng Nhân sự năm ngoái là 157.000£,
- mức độ hài lòng của nhân viên về điều kiện làm việc trong nhà máy,
- trong tháng 12, đã có 1.500 xe cút kít được sản xuất,
- tháng trước đã có 385 đề nghị thanh toán công tác phí được trình nộp,

Trên đây là các ví dụ về dữ liệu *nội bộ (internal)*, phát sinh từ bên trong một tổ chức.

Dữ liệu *bên ngoài (external)* được sinh ra từ bên ngoài tổ chức. Ví dụ:

- tỷ lệ lạm phát tháng trước tăng tới 4% mỗi năm,
- Nghị viện vừa thông qua pháp chế mới về công tác kiểm soát ô nhiễm,
- năm ngoái, lượng ô tô Nhật Bản nhập khẩu tăng 5%,
- một khách hàng khen ngợi sản phẩm xe cút kít mà anh ta đã mua.

Chúng ta sẽ xét đến các nguồn dữ liệu bên ngoài chi tiết hơn ở phần sau.

**Thông tin (information)** là dữ kiện thu được từ quá trình xử lý dữ liệu nào đó. Thông tin là một tổ hợp dữ liệu có liên quan đến nhau. Ví dụ:

- 3 tuần trước, thiết bị đã hỏng 4 lần vào tuần trước đã được đại tu,
- 13 nhân viên Phòng Kế toán tương ứng với 10% toàn bộ nhân viên công ty,
- mức ngân sách 157.000£ cho Phòng Nhân sự năm ngoái đã tăng 12% so với năm trước đó,

Tổng thể các dữ liệu liên quan sẽ cung cấp cho các nhà quản lý thông tin quan trọng. Thông tin thể hiện bối cảnh sinh ra và giải thích ý nghĩa của dữ liệu.

Dữ liệu có thể được xử lý thành thông tin bằng cách:

- tập hợp các dữ liệu liên quan rồi lập bảng thống kê, kết hợp, chất lọc hoặc đơn giản là sắp xếp lại chúng;
- tóm tắt;
- lập bảng và sử dụng biểu đồ;
- phân tích thống kê;
- phân tích tài chính

## 2.3 NGUỒN DỮ LIỆU

Quá trình hoạt động của bất cứ tổ chức nào cũng thường sinh ra lượng dữ liệu lớn cần sắp xếp và xử lý bằng nhiều cách, nhằm thỏa mãn nhu cầu thông tin của từng bộ phận trong tổ chức. Hơn nữa, dữ liệu “nhà trồng” còn được bổ sung bởi lượng dữ liệu khổng lồ từ bên ngoài, tạo nên nền tảng thông tin hữu dụng cho tổ chức.

Vấn đề mà tổ chức phải đương đầu là quản lý nguồn dữ liệu này, sao cho thu được thông tin cần thiết vào đúng thời điểm và ở dạng thức phù hợp.

### 2.3.1 NGUỒN DỮ LIỆU NỘI BỘ

Chúng ta biết rằng các tổ chức luôn nắm giữ lượng lớn dữ liệu giao dịch trong cơ sở dữ liệu riêng. Nếu được xử lý thích hợp, các dữ liệu này có thể cung cấp thông tin hỗ trợ việc đưa ra quyết định vận hành, chiến lược và chiến thuật.

### 2.3.2 NGUỒN DỮ LIỆU BÊN NGOÀI

Các tổ chức thường sử dụng dữ liệu có được từ bên ngoài. Ví dụ:

- khảo sát nghiên cứu thị trường nhằm xác định mức độ hài lòng của khách hàng đối với một sản phẩm nhất định.
- thông tin từ các ấn phẩm chính thống về quy mô và đặc điểm dân cư có ích trong việc ước tính số lượng khách hàng tiềm năng mua sản phẩm mới.
- thông tin từ báo cáo doanh nghiệp về các hoạt động của bên đối thủ (bán hàng, đầu tư, tiếp quản...) là quan trọng để duy trì tính cạnh tranh trên thị trường.

### 2.3.3 DỮ LIỆU SƠ CẤP VÀ DỮ LIỆU THỨ CẤP

Dữ liệu được sử dụng hoàn toàn theo mục đích thu thập ban đầu gọi là *dữ liệu sơ cấp (primary data)*.

Dữ liệu được sử dụng khác với mục đích thu thập ban đầu gọi là *dữ liệu thứ cấp (secondary data)*.

Thuật ngữ “sơ cấp” và “thứ cấp” ám chỉ mục đích sử dụng dữ liệu.

### 2.3.4 VẤN ĐỀ KHI SỬ DỤNG DỮ LIỆU THỨ CẤP

Trong phần lớn trường hợp, người ta ưa dùng dữ liệu sơ cấp hơn vì dữ liệu được thu thập cho mục đích cụ thể thường tốt hơn, chính xác và đáng tin cậy hơn. Tuy nhiên, không phải lúc nào việc sử dụng dữ liệu sơ cấp cũng khả thi; do đó chúng ta cần lường trước những vấn đề khi sử dụng dữ liệu thứ cấp. Chúng ta hay gặp một số vấn đề sau:

- a) Dữ liệu này do người khác thu thập. Chúng ta không kiểm soát được cách thức thu thập dữ liệu đó. Nếu sử dụng bảng khảo sát, liệu:
  - bảng hỏi có nội dung thích hợp?
  - kích thước mẫu đủ lớn?
  - tổ chức thực hiện thu thập dữ liệu có uy tín?
  - dữ liệu thu được đủ độ chính xác cần thiết?
- b) Dữ liệu có cập nhật không? Dữ liệu thường nhanh chóng trở nên lỗi thời vì nhiều nguyên nhân, chẳng hạn thị hiếu người dùng thay đổi. Các mức tăng giá cũng có thể ảnh hưởng nghiêm trọng đến thị trường.
- c) Dữ liệu có thể chưa hoàn thiện. Đôi khi số liệu được công bố không tính đến một số nhóm nhất định, ví dụ, lượng người thất nghiệp không bao gồm người không có việc làm. (*Vậy nhóm đối tượng nào đã không được tính?*) Ví dụ, thống kê cụ thể do Hiệp hội kinh doanh xe máy (*Motor Traders Association*) công bố có thể không bao gồm số liệu về xe ba bánh, xe tải nhỏ (*van*) và nhà

di động (*motor-caravan*). Chúng ta phải nắm rõ các hạng mục nào có trong dữ liệu định sử dụng.

- d) Thông tin có thực không, có được điều chỉnh thời vụ, đánh giá hay chỉ là dự đoán?
- e) Số liệu công bố có thể chưa đủ chính xác và chúng ta không có quyền truy cập dữ liệu thô. Ví dụ, số liệu dân số đã công bố có thể được làm tròn đến hàng nghìn, mà chúng ta lại muốn biết con số chính xác.
- f) Lý do thu thập dữ liệu ban đầu có thể không được biết, nên khó mà đánh giá xem số liệu đã công bố có thích hợp để sử dụng trong hoàn cảnh hiện tại hay không.

Nếu buộc phải sử dụng dữ liệu thứ cấp, chúng ta phải trả lời các câu hỏi trên. Đôi khi câu trả lời có sẵn ngay trong chính dữ liệu, hoặc đôi khi ta cần liên hệ với người đã thu thập dữ liệu. Nếu không, chúng ta phải lường trước được những hạn chế khi đưa ra quyết định dựa trên thông tin đến từ nguồn dữ liệu thứ cấp.

### 2.3.5 NGUỒN DỮ LIỆU THỨ CẤP

Có rất nhiều nguồn dữ liệu thứ cấp, chủ yếu được phân loại thành hai nhóm sau đây:

- a) Dữ liệu đưa ra bởi mỗi công ty, chính quyền địa phương, công đoàn, nhóm lợi ích... Ví dụ:
  - i) Bản tin hàng quý của Ngân hàng Anh Quốc (*Bank of England Quarterly Bulletin*) – báo cáo tài chính và kinh tế.
  - ii) Báo cáo doanh nghiệp (chủ yếu là báo cáo thường niên) – thông tin hoạt động và báo cáo của mỗi công ty.
  - iii) Nghiên cứu lao động (hàng tháng) – bài viết về các ngành nghề, việc làm, công đoàn và đảng phái chính trị.
  - iv) Thời báo Tài chính (*Financial Times*) (hàng ngày) – chia sẻ giá cả và thông tin trong kinh doanh.
- b) Dữ liệu từ các Bộ trực thuộc Chính phủ. Đây là nguồn dữ liệu rộng, kèm theo các loại ấn phẩm tổng hợp như Nguyệt san Thống kê (*Monthly Digest of Statistics*), hay tài liệu cụ thể hơn như Khảo sát thu nhập (*New Earnings Survey*).

*Thống kê Chính phủ (Government Statistics)* – hướng dẫn đơn giản về các nguồn tài nguyên liệt kê tất cả các ấn phẩm và danh bạ bộ ngành.

*Hướng dẫn Thống kê Chính thống (Guide to Official Statistic)* là một danh sách toàn diện hơn.

Các bạn có thể truy cập dữ liệu trực tuyến của toàn bộ các ấn phẩm thống kê chính phủ tại trang <http://www.statistics.gov.uk>

## 2.4 THU THẬP DỮ LIỆU

Mọi tổ chức đều thu thập cái gọi là “dữ liệu thường xuyên”, bao gồm hồ sơ nhân viên, khách hàng, hóa đơn, bán hàng, rủi ro nghề nghiệp, hàng hóa đang có, hàng hóa đặt mua... Mỗi tổ chức sẽ thiết lập các hệ thống riêng để thu thập và lưu trữ loại dữ liệu này cho đến khi cần dùng.

Ngoài ra, trong nhiều trường hợp, tổ chức cần thu thập dữ liệu cho vài mục đích cụ thể. Ví dụ:

1. khảo sát phản ứng của người tiêu dùng với một sản phẩm mới ra mắt
2. đánh giá hiệu suất làm việc của dây chuyền sản xuất
3. điều tra các loại lỗi đã xuất hiện ở một sản phẩm nhất định
4. khảo sát phản ứng của nhân viên về những thay đổi tại nhà ăn giành cho họ.

Đối với các trường hợp này, phải áp dụng phương pháp thu thập dữ liệu không thường xuyên. Các phương pháp phổ biến nhất là:

- a) quan sát hoặc điều tra trực tiếp (ví dụ 2. ở trên)
- b) bảng hỏi dạng viết (tính cả gửi qua bưu điện và trực tuyến) (ví dụ 4. ở trên)
- c) phỏng vấn cá nhân (tính cả phỏng vấn qua điện thoại) (ví dụ 1. ở trên)
- d) trích rút từ hồ sơ hoặc dữ liệu thống kê đã được công bố (ví dụ 3. ở trên)

***Quan sát trực tiếp (direct observation)*** cho phép giám sát tình huống nghiên cứu một cách kín đáo. Đây là phương pháp lý tưởng dành cho điều tra viên, vì rủi ro với thông tin thiếu chính xác là rất nhỏ. Tuy nhiên, đây là cách làm khá tốn kém.

Việc quan sát mà không ảnh hưởng đến hành vi của đối tượng được quan sát là hết sức quan trọng. Ví dụ, nếu biết mình đang bị quan sát hành vi mua sắm, thì nhiều người sẽ thay đổi hành vi tiêu dùng. Thậm chí họ còn có thể rời khỏi đó luôn!

Phương pháp này chủ yếu được áp dụng cho các khảo sát khoa học, khảo sát giao thông và điều tra, ví dụ xác định hình thức dịch vụ khách hàng.

***Điều tra trực tiếp (direct inspection)*** sử dụng quy trình chuẩn hóa để xác định một số tính chất hoặc chất lượng của đối tượng hay chất liệu. Ví dụ, lấy ra 5 ổ bánh mỳ từ cùng một mẻ bánh trong tiệm, rồi cắt ra để kiểm tra độ quánh của hỗn hợp bột.

***Bảng hỏi dạng viết (written questionnaire)*** là một trong những cách thu thập dữ liệu hữu dụng nhất khi vấn đề cần nghiên cứu không phức tạp, nên dạng câu hỏi thường đơn giản, ngắn gọn. Nên thiết lập bảng hỏi với nhiều loại câu hỏi mà chỉ cần câu trả lời đơn giản, như là Đúng/Sai, đánh dấu vào ô, khoanh tròn lựa chọn đúng nhất...

Chi phí thực hiện khảo sát dạng bảng hỏi khá rẻ – thời gian (và chi phí) gửi đi 500 bảng hỏi sẽ ít hơn nhiều so với phỏng vấn 500 người. Vấn đề chính là tỷ lệ phản hồi đối với bảng hỏi gửi qua bưu điện/trực tuyến thường khá thấp, khoảng 10%.

Việc thiết kế bảng hỏi cũng rất quan trọng và không hề đơn giản.

**Phỏng vấn cá nhân (personal interviewing)** có thể được sử dụng để khảo sát thái độ đối với một vấn đề cụ thể, như thăm dò dư luận công chúng. Người phỏng vấn cần có kinh nghiệm và luôn giữ quan điểm công bằng trong suốt cuộc phỏng vấn. Câu hỏi có thể phức tạp hơn loại được sử dụng trong bảng hỏi, vì người phỏng vấn sẽ giúp giải đáp thắc mắc và ghi lại những phản hồi phức tạp hơn. Đôi lúc cuộc phỏng vấn được ghi âm. Rõ ràng là chi phí phỏng vấn cá nhân sẽ cao. Việc dùng điện thoại sẽ giảm được chi phí, song lại làm tăng sai số của mẫu vì đâu dễ tiếp cận mọi phần tử của tổng thể thông qua điện thoại.

**Trích rút từ hồ sơ hoặc thống kê đã được công bố (abstraction from records or published statistics)** là phương pháp thu thập dữ liệu vô cùng kinh tế và tiện lợi. Tuy nhiên, dữ liệu được sử dụng thường đã được thu thập để phục vụ mục đích khác và có thể không ở dạng cần dùng. Khi đó mọi vấn đề của việc sử dụng dữ liệu thứ cấp sẽ phát sinh.

## 2.5 CHỌN MẪU

### 2.5.1 TỔNG THỂ VÀ MẪU – TẠI SAO PHẢI CHỌN MẪU?

Thuật ngữ **tổng thể (population)** chỉ toàn bộ nhóm người hoặc vật liên quan tới phép điều tra thống kê. Thuật ngữ **mẫu (sample)** chỉ một nhóm nhỏ được chọn ra từ tổng thể đó. Tương tự như vậy, ta dùng thuật ngữ **tham số (parameter)** để chỉ số đo của tổng thể, và thuật ngữ **thống kê (statistic)** chỉ số đo của mẫu tương ứng. Ví dụ, nếu coi tổng thể là thành viên hiện tại của Institute of Directors (tạm dịch: Viện Thành viên Hội đồng quản trị), thì mức lương trung bình của tất cả thành viên là một tham số tổng thể. Tuy nhiên, nếu ta lấy mẫu gồm 100 thành viên, thì lương trung bình của nhóm này gọi là một thống kê mẫu.

Trong một khảo sát ý kiến sinh viên về dịch vụ ăn uống tại trường đại học, tổng thể đích là tất cả sinh viên đăng ký ăn ở trường đó. Tuy nhiên, nếu khảo sát về dịch vụ ăn uống nói chung ở các trường đại học, tổng thể đích sẽ là tất cả sinh viên tại tất cả các trường. Dù ở trong trường hợp nào thì cũng khó mà thu thập được dữ liệu từ toàn bộ tổng thể, nên nhóm khảo sát sẽ chọn ra một nhóm sinh viên nhỏ để đại diện cho tổng thể này. Khi đó nhóm khảo sát sẽ đưa ra kết luận thích hợp về tổng thể, dựa trên thông tin thu được từ mẫu. Việc định nghĩa cho tổng thể từ đầu quá trình điều tra là rất quan trọng, nhằm đảm bảo ý nghĩa của mọi kết luận được đưa ra.

Tổng thể bị giới hạn về kích cỡ được gọi là tổng thể hữu hạn (finite). Tổng thể không bị giới hạn kích cỡ được gọi là tổng thể vô hạn (infinite). Trên thực tế, nếu một tổng thể đủ lớn đến mức việc loại đi một phần tử không ảnh hưởng đáng kể đến khả năng lựa chọn phần tử khác, thì về mặt thống kê, tổng thể này được coi là vô hạn.

Ban đầu, chúng ta có thể cho rằng dùng toàn bộ tổng thể để điều tra thống kê là tốt hơn, nếu thực hiện được. Tuy nhiên, thực tế cho thấy việc sử dụng mẫu thì tốt hơn. Ưu điểm của việc dùng mẫu so với tổng thể là:

a) **Tính thực tiễn (practicality)**

Tổng thể có thể rất lớn, hoặc vô hạn. Do đó, khả năng thực tế là không thể thu thập dữ liệu từ toàn bộ tổng thể.

b) **Thời gian (time)**

Nếu cần đến dữ liệu gấp, có thể sẽ không có đủ thời gian để thực hiện khảo sát toàn bộ tổng thể. Ví dụ, nếu chúng ta muốn kiểm tra chất lượng hàng hóa đã được sản xuất hàng loạt, thì việc giao hàng trễ do phải kiểm tra từng món hàng một là không thể chấp nhận được.

c) **Chi phí (cost)**

Chi phí thu thập dữ liệu toàn bộ tổng thể có thể rất đắt. Như trong ví dụ trên, chi phí kiểm tra từng món hàng có thể khiến giá của sản phẩm được sản xuất hàng loạt trở nên quá mức đắt đỏ.

d) **Lỗi (error)**

Khi dữ liệu được thu thập từ một tổng thể lớn, thì công việc thu thập, quản lý và phân tích dữ liệu trên thực tế sẽ liên quan đến lượng nhân lực lớn và nguy cơ xảy lỗi tăng nhanh. Do đó, việc sử dụng một mẫu với tập dữ liệu nhỏ hơn sẽ phát sinh ít lỗi hơn.

e) **Kiểm tra phá hủy (destructive)**

Quá trình thu thập dữ liệu có thể đòi hỏi phương pháp kiểm tra phá hủy. Khi đó với một số bài kiểm tra, rõ ràng việc xử lý toàn bộ tổng thể là việc không mong muốn. Ví dụ, một nhà sản xuất muốn xác định độ bền của một loại pin cụ thể. Anh ta sẽ thực hiện kiểm tra trên một số pin cho đến khi chúng hỏng, để xác định nội dung tuyên bố phù hợp về tất cả số pin.

Khi nào nên sử dụng tổng thể thay vì một mẫu?

a) **Tổng thể nhỏ (small population)**

Nếu kích cỡ tổng thể nhỏ, mà bất cứ mẫu nào được lấy sẽ ảnh hưởng lớn đến kích cỡ tổng thể, thì thời gian, chi phí và độ chính xác khi sử dụng tổng thể sẽ không khác mấy so với khi sử dụng mẫu.



### b) Tính chính xác (accuracy)

Nếu thông tin rút ra từ dữ liệu phải thật chính xác, thì kết luận thống kê từ dữ liệu mẫu có thể chưa đủ độ tin cậy. Ví dụ, một cửa hàng cần biết chính xác số tiền giao dịch tại quầy trong một năm. Nếu chủ cửa hàng chỉ ghi lại số liệu từ một mẫu trong vài ngày của năm thì không đủ. Nguy cơ phát sinh lỗi vẫn còn đó, song sai sót trong dữ liệu thường nằm ở phép toán hơn là phép thống kê.

Mục đích điều tra thống kê có thể là đo lường một tham số của tổng thể, ví dụ, tuổi trung bình của kế toán viên chuyên nghiệp tại Anh, hoặc mức lương dao động của công nhân ngành thép ở Pháp. Ngoài ra, mục đích điều tra có thể là xác nhận một quan điểm nào đó về tổng thể, ví dụ, ý kiến cho rằng bạo lực trên truyền hình góp phần làm gia tăng bạo lực trong xã hội, hay việc chạy bộ giúp bạn giữ dáng. Thông thường không thể điều tra toàn bộ tổng thể, vì thế phải sử dụng một mẫu trong đó.

Việc sử dụng dữ liệu từ một mẫu thay vì tổng thể có ý nghĩa quan trọng với phép điều tra thống kê, dẫn dắt ta tới các suy luận mang tính thống kê; vì thế cần nắm vững kết luận rút ra từ mẫu trong tổng thể. Thống kê mẫu cho biết điều gì về tham số tổng thể, hay bằng chứng từ mẫu cho phép kết luận ra sao với quan điểm nào đó về tổng thể? Tuổi trung bình của *mẫu* trong nhóm kế toán viên chuyên nghiệp liên quan thế nào đến tuổi trung bình của *tổng thể* gồm tất cả kế toán viên chuyên nghiệp? Giả dụ, với *mẫu* là người trưởng thành, thì những người chạy bộ có cơ thể săn chắc hơn, thì liệu có thể nói rằng trên toàn *tổng thể* thì việc chạy bộ sẽ giúp bạn giữ dáng? Với một mẫu chọn thích hợp, có thể ước lượng tham số tổng thể từ thống kê mẫu, và sử dụng bằng chứng từ mẫu để kiểm tra quan điểm nào đó về cả tổng thể.

Kết luận có tính thống kê là một khía cạnh lớn và quan trọng của thống kê. Thông tin được tập hợp từ một mẫu và được sử dụng để suy luận về khía cạnh nào đó của tổng thể. Ví dụ, kiểm toán viên có thể kiểm tra mẫu là các giao dịch của một công ty; nếu mẫu này thỏa đáng, thì anh ta sẽ đưa ra kết luận rằng mọi giao dịch của công ty này đều thỏa đáng. Anh ta sẽ sử dụng mẫu, vì cách làm này rẻ hơn, nhanh hơn và thực tế hơn là kiểm tra tất cả giao dịch được thực hiện trong công ty đó.

### 2.5.2 CHỌN MẪU

Việc chọn lựa phần tử của mẫu là rất quan trọng, đảm bảo mẫu càng đại diện được cho tổng thể càng tốt, trong điều kiện thiếu hụt tài nguyên, thời gian và tiền bạc. Một mẫu thiên kiến sẽ cho thông tin sai lệch về tổng thể.

Có nhiều phương pháp chọn phần tử cho mẫu. Các phương pháp này có thể chia làm hai loại – thiết kế mẫu **ngẫu nhiên (random)** và thiết kế mẫu **không ngẫu nhiên (non-random)**.

### 2.5.3 CHỌN MẪU NGẪU NHIÊN

Có thể đưa ra kết luận thống kê chỉ nhờ phương pháp chọn mẫu này. Chọn mẫu ngẫu nhiên có nghĩa là mọi phần tử trong tổng thể đều có cơ hội được làm mẫu.

Nếu tổng thể chứa nhiều nhóm phần tử với tính chất khác nhau – điểm này rất quan trọng trong điều tra nghiên cứu – thì việc chọn mẫu ngẫu nhiên sẽ thu được một mẫu chứa các phần tử thuộc về từng nhóm đó. Nhóm càng lớn thì các phần tử trong nhóm càng có khả năng được chọn vào mẫu, khiến tính đại diện của nhóm trong mẫu cũng lớn hơn. Kích cỡ của nhóm liên quan đến xác suất được lựa chọn của phần tử trong nhóm, vì thế chọn mẫu ngẫu nhiên có xu hướng tạo ra một mẫu có tính đại diện cho tổng thể.

Bước đầu tiên trong chọn mẫu ngẫu nhiên từ một tổng thể hữu hạn là thiết lập một **khung mẫu (sampling frame)**. Đây là danh sách gồm tất cả phần tử của tổng thể, có thể ở bất cứ dạng nào, miễn là mỗi phần tử đều có thể xác định. Mỗi phần tử thuộc tổng thể được gán cho một số, và phương pháp ngẫu nhiên nào đó sẽ được sử dụng để chọn các số, từ đó xác định các phần tử cho mẫu. Tính đại biểu của mẫu phụ thuộc vào chất lượng khung mẫu. Khung mẫu cần có các tính chất quan trọng sau:

- a) **Tính đầy đủ (completeness)** – tất cả phần tử của tổng thể đều phải hiện diện trong khung mẫu. Tính không đầy đủ có thể dẫn đến thiếu sót trong mẫu, đặc biệt nếu các phần tử vắng mặt cùng thuộc về một nhóm trong tổng thể.
- b) **Tính chính xác (accuracy)** – thông tin cho mỗi phần tử phải chính xác và không được để xảy ra trường hợp trùng lặp phần tử.

Cần phân biệt danh sách các phần tử này với *tổng điều tra (census)*. Mục đích duy nhất của khung mẫu là xác định rõ các phần tử để chọn mẫu. Khi đó, ta thu thập dữ liệu cần thiết từ các phần tử đã được chọn của mẫu. Nếu dữ liệu được thu thập từ mọi phần tử của khung mẫu, hay là toàn bộ tổng thể, thì đây là tổng điều tra. Phương pháp lựa chọn ngẫu nhiên có thể được tiến hành thủ công như việc nhặt ngẫu nhiên những quả bóng được đánh số từ một cái túi. Ngoài ra, có thể dùng máy tính bỏ túi hoặc máy vi tính để chọn số ngẫu nhiên. Những số này sẽ cho ra mẫu ngẫu nhiên tương ứng.

### 2.5.4 THIẾT KẾ MẪU NGẪU NHIÊN

#### ***Thiết kế mẫu ngẫu nhiên đơn (simple random sample design)***

Mục trước đã mô tả nguyên tắc cơ bản và phác thảo quy trình lựa chọn mẫu ngẫu nhiên đơn. Một mẫu ngẫu nhiên đơn là thích hợp nhất trong trường hợp tất cả phần tử của tổng thể đều tương đồng đối với mục đích điều tra. Ví dụ, mẫu ngẫu nhiên đơn sẽ phù hợp khi chọn mẫu gồm 20 nhân viên trong một công ty để thực hiện khảo sát, nếu chúng ta chỉ quan tâm đến việc người được chọn là nhân viên trong công ty.

### ***Chọn mẫu ngẫu nhiên phân tầng (stratified random sampling)***

Nếu các phần tử tổng thể có tính chất khác nhau, mà những tính chất này thuộc phạm vi quan tâm, thì phương pháp chọn mẫu ngẫu nhiên đơn có thể không chọn ra được mẫu tiêu biểu nhất đối với kích thước mẫu cho trước. Ví dụ, khi khảo sát nhân viên trong một công ty, có thể cần phân biệt giữa nhân viên nam và nữ. Thiết kế mẫu ngẫu nhiên đơn có thể cho ra quá nhiều phần tử thuộc về một giới tính, trừ khi sử dụng một mẫu có kích thước lớn.

Phương pháp chọn mẫu ngẫu nhiên phân tầng cho phép tạo ra mẫu phản ánh chính xác hơn cấu phần của tổng thể. Quy trình đòi hỏi khung mẫu phải được chia nhỏ thành nhiều nhóm thuộc từng phạm vi quan tâm. Nhóm này được gọi là tầng (stratum). Trong ví dụ của chúng ta, nên xác định số nhân viên nam và nữ. Chúng ta cần biết số phần tử trong từng nhóm của tổng thể. Từ đó ta sử dụng phương pháp chọn mẫu ngẫu nhiên đơn để chọn phần tử mẫu từ mỗi tầng riêng, theo tỷ lệ tương ứng với số lượng của chúng trong tổng thể. Nếu cần chọn mẫu gồm 100 phần tử từ tổng thể bao gồm 60% nam và 40%, thì ta chọn 60 nhân viên nam từ tầng nam và 40 nhân viên nữ từ tầng nữ. Ưu điểm chính của phương pháp chọn mẫu ngẫu nhiên phân tầng là có thể sử dụng kích cỡ mẫu nhỏ hơn, để cho ra kết quả giống như khi chọn mẫu ngẫu nhiên đơn.

### ***Thiết kế mẫu hệ thống (systematic sample design)***

Việc chọn mẫu hệ thống có ưu điểm là đơn giản hơn nhiều so với các phương pháp ở trên. Nếu ta cần mẫu chứa 100 phần tử trong tổng thể bao gồm 1.000 phần tử, thì cứ lần lượt chọn (1000/100) phần tử thứ 10 từ khung mẫu. Yếu tố ngẫu nhiên nằm ở chỗ điểm xuất phát ban đầu được chọn là ngẫu nhiên. Giả dụ ta có một tập hóa đơn được đánh số từ 1 đến 1.000. Ta sử dụng số ngẫu nhiên để chọn số hóa đơn đầu tiên, và từ đó cứ cách 10 hóa đơn ta chọn ra 1. Giả sử số ngẫu nhiên là 3, thì ta chọn hóa đơn số 3, 13, 23, 33... để tạo thành mẫu theo yêu cầu. Như vậy ta chỉ cần chọn một số ngẫu nhiên, so với việc phải chọn 100 số như ở các phương pháp khác. Quy trình này thỏa đáng, với điều kiện số hóa đơn không có hiệu ứng chu kỳ hoặc quy luật gộp nhóm đặc biệt. Nếu vì nguyên nhân nào đó mà một khách hàng luôn có số hóa đơn kết thúc bằng số 3, thì mẫu thu được sẽ bị thiên lệch và phương pháp này không được chấp nhận.

### ***Thiết kế mẫu nhiều giai đoạn (multi-stage sample design)***

Như tên gọi, phương pháp chọn mẫu này được thực hiện qua nhiều giai đoạn. Ta thường sử dụng phương pháp này nhất trong trường hợp tổng thể được phân phối trên vùng địa lý rộng. Ví dụ, tổng thể có thể là tất cả thành viên trên thế giới của Hội cựu sinh viên Đại học Sheffield Hallam. Giai đoạn đầu tiên là chia tổng thể thành một số vùng được xác định cụ thể; trong ví dụ này có thể là các quốc gia. Tỷ lệ mẫu phân bổ trên các vùng này được xác định bằng tỷ lệ tổng thể trên mỗi vùng. Như vậy, nếu cần mẫu chứa 1.000 phần tử mà 80% thành viên trong Hội ở Anh và xứ Wales, thì ta phân bổ 800 phần tử

mẫu cho Anh và xứ Wales, tương tự như chọn mẫu phân tầng. Bước tiếp theo là xác định các khu vực nhỏ hơn, có thể là quận hành chính địa phương rồi đến các công ty trong quận. Ta lấy mẫu gồm các quận hành chính địa phương. Trong các quận đó, chọn mẫu gồm các công ty. Cuối cùng là chọn hội viên từ công ty thuộc các quận đã lựa chọn. Nhờ phương pháp này, việc lấy mẫu thực tế sẽ nhanh chóng và thuận tiện hơn.

### ***Thiết kế mẫu chùm (cluster sampling design)***

Với các loại hình thiết kế trước, chúng ta đã chọn từng phần tử một. Với phương pháp chọn mẫu chùm, ta sẽ thiết lập các chùm phần tử được coi là có tính đại diện phù hợp cho toàn bộ tổng thể. Khi đó các chùm phần tử được lựa chọn ngẫu nhiên và mọi phần tử thuộc chùm đều nằm trong mẫu. Ví dụ, một xưởng lớn lưu trữ hóa đơn thành các tập chứa 50 hóa đơn. Nếu trong một năm phát sinh 10.000 hóa đơn, thì sẽ có tổng cộng 200 tập. Tập hóa đơn như vậy có thể coi là chùm. Giả sử xưởng muốn lấy mẫu gồm 300 hóa đơn. Có thể thực hiện bằng cách chọn 300 hóa đơn tách biệt một cách ngẫu nhiên trong số 10.000. Ngoài ra, thiết kế mẫu chùm cho phép ta chọn 6 chùm ngẫu nhiên trong số 200 chùm. Đây là một phương pháp dễ dàng và nhanh chóng hơn nhiều, nhưng ta phải đảm bảo không có tình trạng thiên kiến trong tập hóa đơn.

Còn một số thiết kế chọn mẫu ngẫu nhiên khác, tuy nhiên các phương pháp được đề cập trên đây thường được sử dụng nhất.

### **2.5.5 THIẾT KẾ MẪU KHÔNG NGẪU NHIÊN**

Với nhiều khảo sát, đặc biệt trong lĩnh vực nghiên cứu thị trường, thường không tồn tại khung mẫu. Giả sử muốn khảo sát quan điểm của người nội trợ về một sản phẩm mới, thì ta sẽ khó mà phác họa trước khung mẫu về những người nội trợ.

**Chọn mẫu định ngạch (quota sampling)** là một phương pháp chọn mẫu thông dụng trong trường hợp như vậy. Đầu tiên, tương tự như chọn mẫu phân tầng, ta xác định các tính chất quan trọng nhất của tổng thể đích, ví dụ: nam hay nữ, độ tuổi, tầng lớp xã hội... Mẫu được chia càng chi tiết theo tỷ lệ vào các nhóm càng tốt; sau đó, cách chọn phần tử mẫu sẽ được bỏ ngỏ để nhóm khảo sát thực tế quyết định. Ta không cần quan tâm xem nên xác định cá thể trước hay chọn phần tử ngẫu nhiên.

Trong kiểm toán, người ta có thể sử dụng phương pháp gọi là **chọn mẫu phán đoán (judgement sampling)**, tức là kế toán viên sẽ kết hợp trực giác, kiến thức nền và phán đoán để chọn mẫu. Chúng ta không cần đến sự phân tầng hay chọn lựa ngẫu nhiên.

Bất cứ quy trình không ngẫu nhiên nào cũng đều khiến việc phân tích thống kê sau này mất hiệu lực, do đó nên tránh sử dụng phương pháp này nếu có thể.

### 2.5.6 KHẢO SÁT VÀ ĐIỀU TRA THỐNG KÊ

Khi thực hiện một điều tra thống kê thuộc loại được nhắc đến ở trên, cần đáp ứng các yêu cầu sau:

- Xác định mục đích khảo sát – cần thu thập thông tin gì, cho ai/cái gì.
- Xác định tổng thể đích
- Xác định phương pháp chọn mẫu
- Chọn phương pháp thu thập dữ liệu thích hợp – đảm bảo rằng phương pháp này sẽ mang lại thông tin cần thiết
- Tiến hành khảo sát thử – đây là quy trình “tổng duyệt” trước khi bắt đầu toàn bộ khảo sát, cho biết mức độ thích hợp của phương pháp thu thập dữ liệu; cụ thể là tính đầy đủ của bảng hỏi được sử dụng. Ta sẽ thực hiện bước này nếu cần thiết.
- Tiến hành khảo sát chính
- Phân tích và trình bày kết quả

## 2.6 THÔNG TIN CẦN THIẾT ĐỂ RA QUYẾT ĐỊNH HIỆU QUẢ

Để đưa ra được quyết định hiệu quả về hoạt động và chiến lược, người quản lý cần định vị công ty trong hiện tại và tương lai. Ví dụ:

<b>vị trí hiện tại</b>	<b>vị trí muốn đạt được</b>
<b>cấp độ vận hành</b>	
đặt hàng theo quy trình sản xuất	giao đơn hàng đúng hạn
kỹ sư đang phục vụ tại nhà khách hàng	tiếp tục phục vụ tại nhà theo yêu cầu
<b>cấp độ chiến thuật</b>	
cập nhật chi phí hoạt động của bộ phận	hoạt động trong phạm vi ngân sách bộ phận
đơn hàng đã giao trong tháng	tính toán hiệu quả giao hàng trong tháng
<b>cấp độ chiến lược</b>	
nhu cầu thường niên cho mỗi sản phẩm	xác định sản lượng chung
số liệu khai thác của mỗi nhà máy	đặt ra số liệu khai thác mục tiêu

**Bảng 2.1** Ví dụ về quyết định ở các cấp khác nhau.

Thông tin có cấu trúc phân tầng, do yêu cầu thông tin ở cấp độ cao được gộp lại từ các cấp thấp hơn. Ví dụ, thông tin tổng nhu cầu thường niên là tổng hợp tất cả nhu cầu sản phẩm từ các đơn hàng khác nhau.

## 2.7 CHẤT LƯỢNG THÔNG TIN

Để có quyết định đúng đắn, thông tin cần phải:

liên quan	làm tăng hiểu biết, giảm sự thiếu chắc chắn, có thể sử dụng cho mục đích định trước	ví dụ: độ tuổi, sở thích ăn uống của khách sử dụng gói nghỉ dưỡng
chính xác	phép đo/đếm là chính xác	khối lượng chính xác, có chỉ ra lỗi để ước tính
hoàn thiện	bao gồm tất cả khía cạnh chính liên quan	tài liệu không sót thông tin
đáng tin cậy	dữ liệu được thu thập bởi các phương thức chính xác và nhất quán, dựa vào bằng chứng, có thể kiểm tra chéo	ví dụ: khảo sát mức độ hài lòng của khách hàng
kịp thời	được đưa ra kịp thời để quyết định có hiệu quả	hàng giờ, hàng ngày, hàng tuần...
được truyền tải đến đối tượng thích hợp	trong cấu trúc phân tầng về cấp độ ra quyết định	được truyền tải đến đối tượng thích hợp nhất

**Bảng 2.2** Các tính chất của thông tin hữu dụng.

## 2.8 KẾT LUẬN

Nhìn chung, thông tin phải phù hợp với mục đích sử dụng ban đầu, thì mới có thể hỗ trợ việc đưa ra quyết định đúng đắn nhất ở cấp độ phù hợp trong phạm vi tổ chức.

## 2.9 TÀI LIỆU THAM KHẢO CHƯƠNG

1. Morris, C, Quantitative Approaches in Business Studies (*tạm dịch*: Phương pháp định lượng trong nghiên cứu kinh doanh), tái bản lần thứ 8, FT/Prentice Hall, 2012 (eBook, tái bản lần thứ 7, 2008), Chương 3.
2. Oakshott, L, Essential Quantitative Methods for Business (*tạm dịch*: Phương pháp nghiên cứu định lượng thiết yếu trong kinh doanh, quản lý & tài chính), tái bản lần thứ 5, Palgrave Macmillan, 2012, Chương 4.
3. Wisniewski, M, Quantitative Methods for Decision Makers (*tạm dịch*: Phương pháp nghiên cứu định lượng phục vụ ra quyết định), tái bản lần thứ 5, FT/Prentice Hall, 2009 (có bản eBook), Chương 3.





## Chương 3

# PHÂN TÍCH DỮ LIỆU; ĐƯA RA THÔNG TIN

### 3.1 GIỚI THIỆU

Trong chương trước, ta đã thảo luận về bản chất của dữ liệu và thông tin, xem xét cách thức mà tổ chức sử dụng để thu thập và quản lý dữ liệu. Bây giờ chúng ta sẽ chuyển sang thảo luận cách sắp xếp và trình bày dữ liệu, sao cho dễ dàng đưa ra thông tin liên quan. Nhiều ý tưởng và kỹ thuật được trình bày ở đây khá đơn giản, nhưng phải áp dụng đúng cách thì mới có thể quản lý thông tin một cách hữu dụng và kịp thời.

Ở các phần tiếp theo, giả định dữ liệu cần thiết đã được thu thập thông qua một phương pháp thích hợp từ tổng thể hoặc mẫu thích hợp. Giờ ta đã sẵn sàng để xử lý dữ liệu và tìm ra thông tin cần thiết.

### 3.2 LƯU TRỮ VÀ CẤU TRÚC DỮ LIỆU

Khi có khối lượng lớn dữ liệu được sinh ra từ trong tổ chức, thì người ta thường lưu trữ trong một cơ sở dữ liệu điện tử.

Cơ sở dữ liệu là cách lưu trữ dữ liệu có cấu trúc để có thể tìm thấy (khôi phục) chúng khi cần đến. Dữ liệu được lưu trữ bằng cấu trúc **bảng biểu (table)**. Mỗi bảng giống như một ngăn kéo được dán nhãn trong tủ hồ sơ, và mọi thứ trong đó đều có cấu trúc dữ liệu như vậy. Điều này giúp cho việc lưu trữ trở nên **tiết kiệm (economic)** (tránh tối đa việc bị nhân đôi dữ liệu) và **hiệu quả (efficient)** (dữ liệu có thể được khôi phục dễ dàng). Các gói cơ sở dữ liệu thương mại (như Microsoft Access) chia dữ liệu thành các hạng mục được gọi là **thực thể (entity)**. Cơ sở dữ liệu này có một bảng cho mỗi thực thể, trên đó chứa dữ liệu liên quan trong mục tương ứng. Hàng ngang trong bảng được gọi là **bảng ghi (record)** còn cột dọc được gọi là **trường (field)**. Nhìn chung, mỗi bảng ghi của cơ sở dữ liệu sẽ liên quan đến một tập hoặc mục dữ liệu khác nhau (một bệnh nhân, một đơn hàng, một khách hàng) và chỉ lưu trữ dữ liệu của duy nhất bảng ghi đó. Mỗi trường phải chỉ rõ **loại (type)** dữ liệu lưu trữ (dạng số, văn bản, ngày tháng...).

### Ví dụ 3.1 – Cấu trúc dữ liệu

Giả sử bạn muốn thiết lập một cơ sở dữ liệu cho bộ sưu tập đĩa CD của mình. Bạn có thể xác nhận một số thực thể; ví dụ, trong cơ sở dữ liệu đĩa CD, bạn có thể thiết lập một bảng lưu dữ liệu về album và một bảng lưu dữ liệu về nghệ sĩ. Cơ sở dữ liệu của bạn sẽ chứa nhiều bảng ghi, mỗi bảng ghi tương ứng với một đĩa CD trong bộ sưu tập. Bạn sẽ phải chỉ ra trường yêu cầu – tính chất đĩa CD mà bạn muốn tạo nhãn.

#### **bảng album**

tiêu đề album

tên nghệ sĩ

ngày phát hành

hãng sản xuất

danh sách bài hát

#### **bảng nhãn**

tên nhãn

địa chỉ

chi tiết khác

#### **bảng nghệ sĩ**

tên nghệ sĩ

tiểu sử (như ngày thành lập, xuất xứ)

#### **bảng danh sách bài hát**

số thứ tự bài hát

tiêu đề bài hát

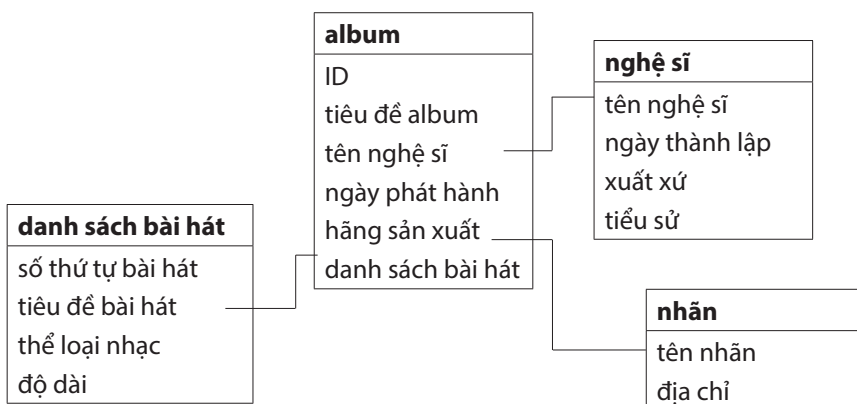
thể loại nhạc

độ dài

**Hình 3.1** Trình bày một cách cấu trúc dữ liệu thành các bảng khác nhau.

Sau khi xác lập, các bảng có thể được gán dữ liệu liên quan và liên kết với nhau, để tất cả dữ liệu thuộc về đĩa CD được khôi phục khi cần đến.

Dữ liệu trong một bảng có thể *liên quan (related)* đến dữ liệu trong một bảng khác khi có chung một trường – *cơ sở dữ liệu quan hệ (relational database)*. Hãy xem **Hình 3.2**.



**Hình 3.2** Ví dụ cơ sở dữ liệu quan hệ.

Sau khi được thiết lập và tiếp nhận dữ liệu, thì cơ sở dữ liệu có thể được khảo sát để khôi phục thông tin cần thiết; ví dụ, tất cả đĩa CD của một nghệ sĩ hay dòng nhạc cụ thể mà bạn có.

Tuy cơ sở dữ liệu thường được sử dụng, nhưng không phải tất cả dữ liệu đều được lưu trữ theo cách này. Thêm vào đó, ta sẽ không thể khảo sát cơ sở dữ liệu này nếu chưa rõ dữ liệu cần tìm. Ta cũng cần thêm cách thức mở để phân tích dữ liệu.

### 3.3 TỔ CHỨC DỮ LIỆU

Do dữ liệu thô (chưa có trong cơ sở dữ liệu) có thể là một mớ thông tin không liên quan, nên bước đầu tiên trong quá trình xử lý là tổ chức dữ liệu bằng cách nào đó. Ta có thể sử dụng nhiều dạng thức khác nhau; mà dạng thức phù hợp nhất phụ thuộc vào bản chất của dữ liệu, thông tin cần thiết cũng như các bước phân tích cần thực hiện về sau.

### 3.4 LẬP BẢNG BIỂU

Một trong những cách thông dụng nhất để tổ chức dữ liệu *dạng số (numerical)* là lập bảng. Đây là phương pháp trình bày thông tin dạng số thường được sử dụng nhất, tuy nhiên nhiều người gặp khó khăn với việc diễn dịch các loại bảng biểu. Khi thiết kế và sử dụng bảng để truyền tải thông tin đến người khác, bạn luôn cần lưu ý tới yếu tố này.

#### 3.4.1 DIỄN DỊCH BẢNG BIỂU

Khi sử dụng bảng do người khác lập sẵn, để hiểu đúng thông tin được đưa ra, cần thực hiện một số nguyên tắc đơn giản như sau.

1. Đầu tiên, nhìn vào tiêu đề của hàng và cột. Chọn và tự diễn giải ý nghĩa (càng rõ ràng càng tốt) bất kỳ số nào trong bảng. Tự hỏi xem cách giải thích này có nghĩa hay không.
2. Nếu trong bảng có xuất hiện phần trăm, hãy xác định nguồn tạo nên phần trăm đó.
3. Nếu có thể, hãy dự đoán số liệu trong bảng đại diện cho mục nào trong bảng, rồi so sánh số liệu có sẵn với dự đoán.
4. Khi diễn dịch ta thường sử dụng phép so sánh; ví dụ, xem xét các thay đổi theo thời gian, hoặc thay đổi giữa các vùng.

#### 3.4.2 THIẾT KẾ BẢNG

Cách thiết kế bảng phụ thuộc vào mục đích sử dụng, do người dùng dữ liệu có thể đòi hỏi các tính năng khác nhau. Việc đưa ra nhiều chi tiết hơn nhu cầu người dùng không dễ làm gì cả. Có nhiều giai đoạn để thiết kế bảng:

- a) Quyết định các mục sẽ sử dụng trong bảng. Đơn giản nhất là phân loại theo hai dấu hiệu, nhưng vẫn có thể đặt các mục con.

Phân loại có thể dựa trên:

- thuộc tính (như màu mắt – xanh lá, nâu, đen...)
- dao động đo lường (như nhỏ hơn 100£, từ 100£ – 200£...)
- khoảng thời gian

Ví dụ:

Mức độ quan tâm đến thể thao	Tuổi					
	Dưới 25		25-40		Trên 40	
	Nam	Nữ	Nam	Nữ	Nam	Nữ
<b>Chơi thể thao</b>						
<b>Xem thể thao</b>						
<b>Không quan tâm</b>						

**Bảng 3.1** Ví dụ về quyết định ở nhiều cấp độ khác nhau.

- b) Quyết định mức độ chính xác cần thiết và đơn vị đo lường (như £, nghìn £, m, làm tròn đến 10 dặm Anh, km...).
- c) Dự kiến khung bảng, đặt một số mục theo chiều ngang của trang và một số theo chiều dọc, với phân mục con phù hợp, nếu cần. Người ta thường cho rằng sẽ dễ đọc hơn nếu bảng có nhiều hàng hơn cột. Mỗi hàng và cột nên được đặt tiêu đề ngắn gọn nhưng rõ ràng.
- d) Có thể trình bày sẵn tổng phụ theo cột và hàng, thường với màu tối hơn.
- e) Dữ liệu phần trăm có thể hữu ích. Ta luôn cần biết rõ nguồn gốc của tỷ lệ phần trăm.
- f) Mỗi bảng nên có tiêu đề tổng quát và trích nguồn dữ liệu.

Việc lập các loại bảng kể trên không hẳn lúc nào cũng thích hợp. Tuy nhiên, ta sẽ khó thấy được tổng quan từ một tập dữ liệu lớn, nếu không sử dụng vài phương pháp sắp xếp hoặc phân loại.

Tiếp theo ta sẽ xét đến các phương pháp sắp xếp dữ liệu thành định dạng thích hợp để có thể khôi phục dữ liệu.

## 3.5 SẮP XẾP DỮ LIỆU

### 3.5.1 XẾP HẠNG (RANKING)

Nếu tập dữ liệu nhỏ, thì chỉ cần xếp hạng các giá trị để thu thập thông tin cần thiết, ví dụ sắp xếp dữ liệu theo thứ tự tăng hoặc giảm dần về kích thước.

### Ví dụ 3.2 – Xếp hạng dữ liệu

Mười người ứng tuyển vào một công việc văn phòng trong một công ty. Như một phần trong quy trình phỏng vấn, họ được yêu cầu thực hiện một bài kiểm tra tư duy tiêu chuẩn. Hãy xếp hạng theo kết quả kiểm tra được trình bày dưới đây.

Ứng viên	A	B	C	D	E	F	G	H	I	J
IQ	104	100	105	101	95	103	105	100	99	100

**Bảng 3.2** Dữ liệu IQ.

#### Lời giải:

Điểm số bài kiểm tra tư duy của 10 ứng viên theo thứ tự là:

95, 99, 100, 100, 100, 101, 103, 104, 105, 105

Giờ ta có thể thấy xu hướng từ dữ liệu, nếu có.

### 3.5.2 MÔ HÌNH THÂN-VÀ-LÁ

Một cách đơn giản để làm quen với một tập số là **mô hình thân-và-lá (stem-and-leaf)** của Tukey. Sơ đồ này sử dụng tập dữ liệu thô và trình bày giá trị cụ thể theo kiểu hình phân phối chung.

#### Ví dụ 3.3 – Mô hình thân-và-lá

Công ty trách nhiệm hữu hạn Multipress sản xuất sản phẩm vệ sinh nhà cửa, bao gồm cả bột giặt. Để có một sản phẩm đáng tin dùng, thì % độ ẩm của bột giặt phải được kiểm soát kỹ lưỡng. Số liệu dưới đây cho thấy % độ ẩm trong lô bột giặt được sản xuất trong các ca làm việc khác nhau tại nhà máy.

Lô	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
% độ ẩm	7,2	6,5	5,3	3,9	6,7	5,9	6,0	6,5	6,9	4,3	6,3	7,6	5,9	5,7	5,5

**Bảng 3.3** Dữ liệu ca ngày (15 lô).

Lô	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
% độ ẩm	5,9	5,6	3,3	4,4	6,9	6,1	5,6	2,9	6,4	8,8	8,1	7,1	7,1	4,8	4,0	2,5	6,4	6,5

**Bảng 3.4** Dữ liệu ca đêm (18 lô).

Lập một mô hình thân-và-lá cho ca ngày.

### Lời giải:

Phần “thân” của sơ đồ biểu diễn chữ số đầu tiên của giá trị. Phần “lá” biểu diễn chữ số thứ hai, **sắp xếp theo thứ tự xếp hạng**.

**% độ ẩm trong 15 lô bột giặt do công nhân ca ngày sản xuất**

Đơn vị, %	Phần mười, %					
3	9					
4	3					
5	3	5	7	9	9	
6	0	3	5	5	7	9
7	2	6				

**Bảng 3.5** Ví dụ về sơ đồ thân-và-lá.

Giờ ta dễ dàng thấy được các giá trị lớn nhất và nhỏ nhất, và vị trí tập trung của các giá trị. Khi đó có thể kiểm tra kiểu hình phân phối chung.

Khi lập sơ đồ thân-và-lá, việc chỉ sử dụng chữ số chính cho phần thân có thể làm mất thông tin giá trị. Có thể tránh trường hợp này bằng việc sử dụng mỗi giá trị phần thân hai lần; gán các đơn vị 0, 1, 2, 3, 4 với phần thứ nhất và đơn vị 5, 6, 7, 8, 9 với phần thứ hai. Các giá trị thân được đánh dấu phân biệt, ví dụ:

2\* cho các đơn vị 0, 1, 2, 3, 4 và

2. cho các đơn vị 5, 6, 7, 8, 9

### Ví dụ 3.4 – Mô hình thân-và-lá chi tiết hơn

Trong 30 tuần gần đây, doanh số tuần của một cửa hàng nhỏ như sau:

4300	4100	7400	6100	7900	6000	7100	6900	6300	7700
7000	6600	6400	7100	7100	7400	5600	7400	4100	7100
6300	5700	5700	6800	6400	6200	5900	5200	4000	7600

**Bảng 3.6** Doanh số mỗi tuần (£).

Lập mô hình thân-và-lá.

**Lời giải:**

4*	0	1	1	3				
4.								
5*	2							
5.	6	7	7	9				
6*	0	1	2	3	3	4	4	
6.	6	8	9					
7*	0	1	1	1	1	4	4	4
7.	6	7	9					

Đơn vị giá trị thân: nghìn £ ('000£)

Đơn vị giá trị lá: trăm £ ('00£)

**Bảng 3.7** Doanh số mỗi tuần (£) của một cửa hàng trong 30 tuần gần đây.

Giờ ta đã thấy rõ tính chất phân phối doanh số hàng tuần của cửa hàng này.

### 3.6 BẢNG PHÂN PHỐI TẦN SỐ

Trong tập dữ liệu lớn hơn, khả năng các giá trị được lặp lại là rất cao. Khi đó, kiểu hình chung của tập dữ liệu có thể được xác định thông qua **bảng phân phối tần số (frequency distribution)**.

#### Ví dụ 3.5 – Lập bảng phân phối tần số

50 hộ gia đình trong một vùng đã được khảo sát. Dưới đây là dữ liệu cư trú của mỗi hộ gia đình.

4	7	4	1	4	2	3	6	3	5
6	3	4	9	12	1	3	4	2	2
1	1	3	8	1	1	4	2	3	4
3	2	1	4	6	5	6	1	2	3
4	4	4	1	4	2	3	5	4	4

**Bảng 3.8** Dữ liệu cư trú của 50 hộ gia đình.

Lập bảng phân phối tần số cho dữ liệu này.

**Lời giải:** Biến số là “số người cư trú trong hộ gia đình”.

Giá trị nhỏ nhất là 1 và giá trị lớn nhất là 12, do đó, trong tập dữ liệu này, biến số nhận giá trị từ 1 đến 12. Liệt kê các giá trị có thể.

Số hộ gia đình có số người cư trú nhất định là “tần số”. Xác định tần số của mỗi giá trị bằng hệ thống kiểm đếm (tally system). Trong lúc đếm dữ liệu, bạn hãy gạch dấu tally theo nhóm giá trị thích hợp. Cuối cùng, cộng tổng số tally để xác định tổng tần số xuất hiện cho mỗi giá trị. Thông thường ta đánh dấu nhóm 5 giá trị một – 4 gạch thẳng và 1 gạch chéo – như vậy giúp cho việc cộng tổng số tally ở cuối trở nên dễ dàng hơn. Chúng ta cần (và nên) đọc qua tập giá trị này một lần để giảm thiểu rủi ro xảy ra lỗi. Xem **Bảng 3.9**.

Số người cư trú trong hộ gia đình	Kiểm đếm		Số hộ gia đình (tần số)
<b>1</b>	++++	1111	9
<b>2</b>	++++	11	7
<b>3</b>	++++	1111	9
<b>4</b>	++++	1111	14
<b>5</b>	111		3
<b>6</b>	1111		4
<b>7</b>	1		1
<b>8</b>	1		1
<b>9</b>	1		1
<b>10</b>			0
<b>11</b>			0
<b>12</b>	1		1
<b>Total</b>			50

**Bảng 3.9** Dữ liệu cư trú của các hộ gia đình.

Tổng số các giá trị nên được kiểm tra liên tục.

Khi đã có bảng tóm tắt dữ liệu ban đầu, ta có thể thấy kiểu hình chung trong phân phối các giá trị. Phần đuôi dài dòng, ít giá trị ở cuối không mang ý nghĩa thông tin quan trọng; do đó ta muốn rút ngắn bảng phân phối bằng việc gộp lại một số giá trị cao hơn, **Bảng 3.10**.



Số người cư trú trong hộ gia đình	Số hộ gia đình
1	9
2	7
3	9
4	14
5 hoặc 6	7
từ 7 trở lên	4
<b>Tổng</b>	50

**Bảng 3.10** Dữ liệu cư trú của các hộ gia đình.

### 3.7 BẢNG PHÂN PHỐI TẦN SỐ NHÓM

Trong ví dụ về số người cư trú trong hộ gia đình ở trên, đôi khi ta thấy việc nhóm các giá trị với nhau là có ích. Phương pháp này càng cần thiết khi biến số là liên tục.

#### Ví dụ 3.6 – Lập bảng phân phối tần số nhóm

Xem lại dữ liệu ở **Ví dụ 3.4** (Doanh số cửa hàng)

Giá trị nhỏ nhất là 4100£ và giá trị lớn nhất là 7900£. Chọn các nhóm bao quát được khoảng này sát nhất có thể. Ta chỉ nên đọc qua tập dữ này một lần duy nhất, và sử dụng hệ thống kiểm đếm để quy giá trị cho các nhóm. Xem **Bảng 3.11**.

Doanh số mỗi tuần (£)	Kiểm đếm		Số tuần (tần số)
4000 - nhỏ hơn 4500	1111		4
4500 - nhỏ hơn 5000			0
5000 - nhỏ hơn 5500	1		1
5500 - nhỏ hơn 6000	1111		4
6000 - nhỏ hơn 6500	1111	11	7
6500 - nhỏ hơn 7000	111		3
7000 - nhỏ hơn 7500	1111	111	8
7500 - nhỏ hơn 8000	111		3
<b>Tổng</b>			30

**Bảng 3.11** Doanh số hàng tuần của một cửa hàng.

Các nhóm thường được gọi là **lớp (class)** hoặc **khoảng gộp (class interval)**. Khoảng giá trị trong một lớp được gọi là **chiều rộng lớp (class width)**. Giá trị nhỏ nhất và lớn nhất trong nhóm là **phạm vi lớp (class boundary)**. Các số được sử dụng để xác định lớp là **giới hạn lớp (class limit)**.

Ví dụ: Trong **Ví dụ 3.6**, lớp đầu tiên là “4000 – nhỏ hơn 4500” (£)

chiều rộng của lớp là 500£,            tức (4500£ trừ 4000£)

phạm vi lớp là:                            4000£ và 4500£

giới hạn lớp cũng là:                    4000£ và 4500£

Ở một ví dụ khác (**Ví dụ 3.7**), chúng ta có thể chia thành các lớp:

Mức tiêu thụ nhiên liệu (mpg)	Số lượng xe
30–31	0
32–33	2
...	

**Bảng 3.12** Dữ liệu tiêu thụ nhiên liệu.

Trong trường hợp này, đối với lớp đầu tiên “30 – 31” mpg:

- giới hạn lớp là: 30 và 31
- phạm vi lớp là: 29,5 và 31,5; tất cả giá trị từ 29,5 trở lên nhưng chưa tới 31,5 sẽ được phân vào nhóm “30 – 31”
- chiều rộng của lớp là 2 mức, 31,5 – 29,5 mpg

### **Ví dụ 3.7 – Lập bảng phân phối tần số nhóm**

Mức tiêu thụ nhiên liệu của 35 xe ô tô cùng loại được đo thông qua một bài kiểm tra tiêu chuẩn. Kết quả được trình bày dưới đây.

29,8	39,2	38,3	34,6	37,0	35,9	40,6	35,8	35,3	39,7	38,3	35,5
38,5	37,7	36,4	35,3	38,8	34,7	38,6	32,6	36,9	36,7	38,0	33,3
36,1	37,8	37,3	35,4	38,6	37,1	36,4	38,5	37,9	36,2	37,5	

**Bảng 3.13** Dữ liệu tiêu thụ nhiên liệu (mpg).

Lập bảng phân phối tần số nhóm.

**Lời giải:** Giá trị nhỏ nhất là 29,8 mpg. Giá trị lớn nhất là 40,6 mpg.

Mức tiêu thụ nhiên liệu (mpg)	Kiểm đếm			Số lượng xe
dưới 30	1			1
30 – nhỏ hơn 32				0
32 – nhỏ hơn 34	11			2
34 – nhỏ hơn 36	1111	111		8
36 – nhỏ hơn 38	1111	1111	111	13
38 – nhỏ hơn 40	1111	1111		10
40 trở lên	1			1
Tổng				35

**Bảng 3.14** Dữ liệu tiêu thụ nhiên liệu (mpg).

Lưu ý việc sử dụng các lớp mờ. Đây là một cách hữu dụng để tránh tình trạng lớp có quá ít giá trị ở cuối chuỗi phân phối, nhưng phải nhớ là việc sử dụng lớp mờ như vậy sẽ dẫn đến mất thông tin.

### Hướng dẫn chung khi chọn khoảng gộp:

- khoảng gộp nên bao quát toàn bộ khoảng giá trị, nhưng sát nhất có thể;
- nên tránh trùng lặp, đơn giản và rõ ràng (sử dụng bội số của 2 hoặc 5);
- chiều rộng các khoảng gộp nên đều nhau, nếu có thể thấy rõ.

Nếu chiều rộng không đều nhau, thì các khoảng gộp nên có mối tương quan về số với nhau. Ví dụ:

Khoảng gộp	Chiều rộng lớp	
0 – nhỏ hơn 10	10	
10 – nhỏ hơn 30	20	gấp đôi chiều rộng lớp 1
30 – nhỏ hơn 50	20	gấp đôi chiều rộng lớp 1
50 – nhỏ hơn 100	50	gấp năm lần chiều rộng lớp 1
100 – nhỏ hơn 200	100	gấp mười lần chiều rộng lớp 1

**Bảng 3.15** Ví dụ về các khoảng gộp có chiều rộng không đều.

- Lớp nên thể hiện sự “phân nhóm tự nhiên” trong môi trường tồn tại; ví dụ: nếu biến số là tuổi người, thì “phân nhóm tự nhiên” có thể gồm học sinh, thiếu niên, thanh niên, người hưu trí. Với **Ví dụ 3.6**, hãy nhóm các giá trị cùng thể hiện mức doanh số tương tự nhau, ví dụ rất thấp, thấp, trung bình...

- e) Nên cẩn thận khi chọn số lớp. Nếu quá nhiều lớp, thì vài lớp sẽ có quá ít giá trị khiến kiểu hình phân phối chung trở nên không rõ ràng. Nếu quá ít lớp thì nhiều thông tin quan trọng có thể bị bỏ lỡ. Nếu bạn còn phân vân, hãy làm theo mẹo đơn giản là chọn số lớp gần tương đương với căn bậc hai của số các giá trị.

### 3.8 PHÂN PHỐI TẦN SỐ TÍCH LŨY

Trong nhiều trường hợp, thông tin rút ra từ một tập dữ liệu phải ở dạng số hoặc tỷ lệ quan sát nhỏ hơn (hoặc lớn hơn) một giá trị cho trước.

#### Ví dụ 3.8 – Lập bảng phân phối tần số tích lũy

Xem lại **Ví dụ 3.4** (Doanh số hàng tuần của một cửa hàng). Sử dụng phân phối tần số nhóm để lập bảng phân phối tần số tích lũy.

**Lời giải:** Xem **Bảng 3.16**.

Doanh số mỗi tuần (£)	Số tuần có doanh số thấp hơn giá trị cho trước (tần số tích lũy)
nhỏ hơn 4000	0
nhỏ hơn 4500	4
nhỏ hơn 5000	4
nhỏ hơn 5500	5
nhỏ hơn 6000	9
nhỏ hơn 6500	16
nhỏ hơn 7000	19
nhỏ hơn 7500	27
ít hơn 8000	30

Lưu ý: Tần số được tích lũy đến phạm vi trên của lớp.

**Bảng 3.16** Ví dụ về các khoảng gộp có chiều rộng không đều.

#### Ví dụ 3.9 – Lập bảng phân phối tần số tích lũy

Lập bảng phân phối tần số tích lũy “nhỏ hơn” từ bảng phân phối tần số nhóm đã lập ở **Ví dụ 3.7** (mức tiêu thụ nhiên liệu).

**Lời giải:** Xem **Bảng 3.17**.

Mức tiêu thụ nhiên liệu (mpg)	Số lượng xe tiêu thụ nhiên liệu ít hơn giá trị cho trước (tần số tích lũy)
nhỏ hơn 30	1
nhỏ hơn 32	1
nhỏ hơn 34	3
nhỏ hơn 36	11
nhỏ hơn 38	24
nhỏ hơn 40	34
Có 1 xe tiêu thụ nhiên liệu ở mức 40 mpg trở lên	

**Bảng 3.17** Ví dụ về các khoảng gộp có chiều rộng không đều.

Lưu ý rằng ta có thể gặp khó khăn với lớp mở cuối cùng.

### 3.9 PHÂN VỊ

Mức “y” phần trăm mà giá trị dữ liệu rơi vào dưới được gọi là phân vị y. Ví dụ, bách phân vị thứ 30 là mức có 30% giá trị dữ liệu rơi vào.

Ta thường sử dụng một số phân vị nhất định. Đó là:

Phân vị thứ 25 – điểm tứ phân vị đầu tiên, hoặc thấp hơn, ký hiệu là  $Q_1$

Phân vị thứ 50 – trung vị, ký hiệu là M

Phân vị thứ 75 – điểm tứ phân vị thứ 3, hay cao hơn, ký hiệu là  $Q_3$

#### Ví dụ 3.10 – Xác định phân vị

Xem lại **Ví dụ 3.5** (Tình trạng cư trú tại các hộ gia đình). Lập bảng phân phối tần số tích lũy và sử dụng để tính phân vị thứ 25, 50 và 75.

**Lời giải:** Xem **Bảng 3.18**.

Số người cư trú trong mỗi hộ gia đình	Số hộ gia đình có số người cư trú ít hơn giá trị cho trước (tần số tích lũy)
nhỏ hơn 1	0
nhỏ hơn 2	9
nhỏ hơn 3	16
nhỏ hơn 4	25
nhỏ hơn 5	39

Số người cư trú trong mỗi hộ gia đình	Số hộ gia đình có số người cư trú ít hơn giá trị cho trước (tần số tích lũy)
nhỏ hơn 6	42
nhỏ hơn 7	46
nhỏ hơn 8	47
nhỏ hơn 9	48
nhỏ hơn 10	49
nhỏ hơn 11	49
nhỏ hơn 12	49
nhỏ hơn 13	50

**Bảng 3.18** Dữ liệu cư trú của các hộ gia đình – tần số tích lũy.

Phân vị thứ 25 là điểm có 25% giá trị rơi vào bên dưới. Trong ví dụ này, 25% giá trị đầu tiên cho tổng lũy tích là 12,5 hộ gia đình. Tổng này nằm giữa hai giá trị tần số tích lũy là 9 và 16, do đó phân vị 25 thuộc quan sát có 2 người cư trú trên hộ gia đình.

Tương tự, phân vị thứ 50 là điểm có 50% giá trị rơi vào bên dưới. Quan sát bảng trên, ta thấy 50% giá trị nằm dưới 4 (giá trị của quan sát thứ 25). Do đó, trung vị là 3 người cư trú trên mỗi hộ gia đình.

Phân vị thứ 75 là giá trị quan sát 37,5. Tại đây có 4 người cư trú trên mỗi hộ gia đình.

Có thể ước tính phân vị dựa trên bảng phân phối tần số nhóm. Phương pháp này sẽ được mô tả trong phần sau.

### 3.9.1 TẦN SỐ TƯƠNG ĐỐI VÀ TẦN SỐ PHẦN TRĂM

Cần sử dụng tần số tương đối hoặc tần số phần trăm khi so sánh giữa các tập dữ liệu có **tổng** số lượng quan sát khác nhau.

#### Ví dụ 3.11 – Tần số phần trăm

Thời gian đáp ứng đơn hàng cho các loại sản phẩm tương tự nhau của hai công ty được đưa ra dưới đây.

Thời gian đáp ứng (tuần)	Số lượng đơn hàng	
	Công ty A	Công ty B
nhỏ hơn 1	1	4
1 - nhỏ hơn 2	1	12
2 - nhỏ hơn 3	3	25
3 - nhỏ hơn 4	7	4
4 - nhỏ hơn 5	5	2
từ 5 trở lên	3	3
Tổng	20	50

**Bảng 3.19** Số đơn hàng cho hai sản phẩm.

Do hai công ty có tổng số lượng đơn hàng khác nhau, nên không thể so sánh trực tiếp tần số. Ta phải tính tần số tương đối hoặc tần số phần trăm.

$$\text{Tần số tương đối} = \frac{\text{tần suất trong lớp}}{\text{tổng tần suất}}$$

$$\text{Tần số phần trăm} = \text{tần số tương đối} \times 100$$

Thời gian đáp ứng (tuần)	Phần trăm đơn hàng	
	Công ty A	Công ty B
nhỏ hơn 1	5	8
1 - nhỏ hơn 2	5	24
2 - nhỏ hơn 3	15	50
3 - nhỏ hơn 4	35	8
4 - nhỏ hơn 5	25	4
từ 5 trở lên	15	6
Tổng	100	100

**Bảng 3.20** Số đơn hàng cho hai sản phẩm.

Giờ ta đã có thể đánh giá sự khác biệt trong kiểu hình phân phối của số đơn hàng giữa hai công ty.

### 3.10 TRÌNH BÀY THÔNG TIN BẰNG HÌNH ẢNH

Sử dụng một trong các dạng thức nói trên, việc tổ chức một tập dữ liệu sẽ tự đưa ra được nhiều thông tin hữu ích về kiểu hình và xu hướng của dữ liệu. Tuy nhiên, nếu được trình bày bằng hình ảnh, thì ta sẽ dễ thấy được các tính chất của tập dữ liệu hơn.

Có rất nhiều hình thức trình bày hình ảnh. Việc lựa chọn hình thức trình bày hoàn toàn phụ thuộc vào loại dữ liệu, mục đích trình bày và mức độ chi tiết được yêu cầu.

Trong các phần tiếp theo, chúng ta sẽ thảo luận về một số hình thức trình bày khác nhau. Ta sẽ cân nhắc phương pháp xây dựng, thời điểm thích hợp để trình bày và cách sử dụng Microsoft Excel để xây dựng hình ảnh trình bày.

#### 3.10.1 BIỂU ĐỒ VÀ SƠ ĐỒ TRÌNH BÀY DỮ LIỆU PHÂN LOẠI

Các loại biểu đồ thông dụng nhất để trình bày dữ liệu phân loại là:

- a) Biểu đồ tần số
- b) Biểu đồ cột – biểu đồ đơn, biểu đồ nhiều cột và biểu đồ cột chồng
- c) Biểu đồ tròn

Còn nhiều lựa chọn khác, nhưng nên cẩn thận khi sử dụng các loại biểu đồ hiếm gặp.

Nếu chưa thông thạo các loại biểu đồ này, hãy đọc thêm các nội dung được đề cập trong phụ lục tham khảo. Phần dưới đây chỉ đưa ra tóm tắt sơ lược.

Bạn cần biết cách **sử dụng Microsoft Excel** để dựng biểu đồ cột và biểu đồ tròn. Trong phần tiếp theo, ta giả định các biểu đồ thực tế được dựng bằng Microsoft Excel.

#### 3.10.2 DỰNG BIỂU ĐỒ VÀ SƠ ĐỒ

- a) **Biểu đồ tần số (pictogram)** Tất cả ký hiệu trong sơ đồ phải cùng kích thước. Ta diễn tả sự thay đổi tần số bằng cách thay đổi số lượng ký hiệu, *chứ không* thay đổi kích thước ký hiệu.
- b) **Biểu đồ cột (bar chart)** Biểu đồ cột chỉ thích hợp với dữ liệu phân loại (category data). Thang biểu diễn trên trục tần số *luôn* phải gồm mốc 0.
- c) **Biểu đồ tròn (pie chart)** Sơ đồ này chỉ phù hợp khi ta có một biến duy nhất được chia thành nhiều phần. Khi đó biểu đồ tròn sẽ biểu diễn kích thước tương quan giữa các phần với nhau. Do đó, không nên sử dụng biểu đồ tròn để biểu diễn dữ liệu kiểu như một chuỗi thời gian, hay biến có giá trị thay đổi theo thời gian.



**Ví dụ 3.12 – Dựng biểu đồ tần số, biểu đồ cột và biểu đồ tròn**

Bảng dưới đây cho thấy Tổng sản phẩm quốc nội (GDP) của Wyeland, chia thành ba phân loại cho ba năm khác nhau.

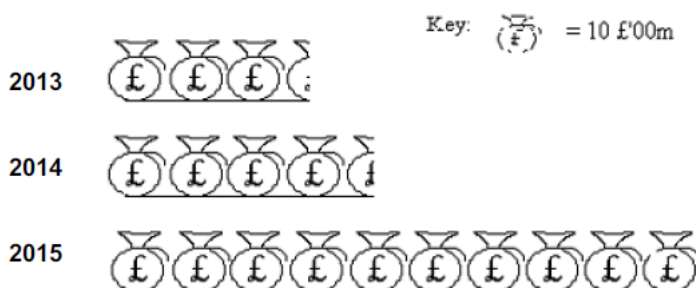
Phân loại	GDP (trăm triệu £)		
	2013	2014	2015
Chi tiêu công cộng	3	5	15
Sản xuất	12	15	30
Khác	18	25	55
Tổng GDP	33	45	100

**Bảng 3.21** Tổng sản phẩm quốc nội.

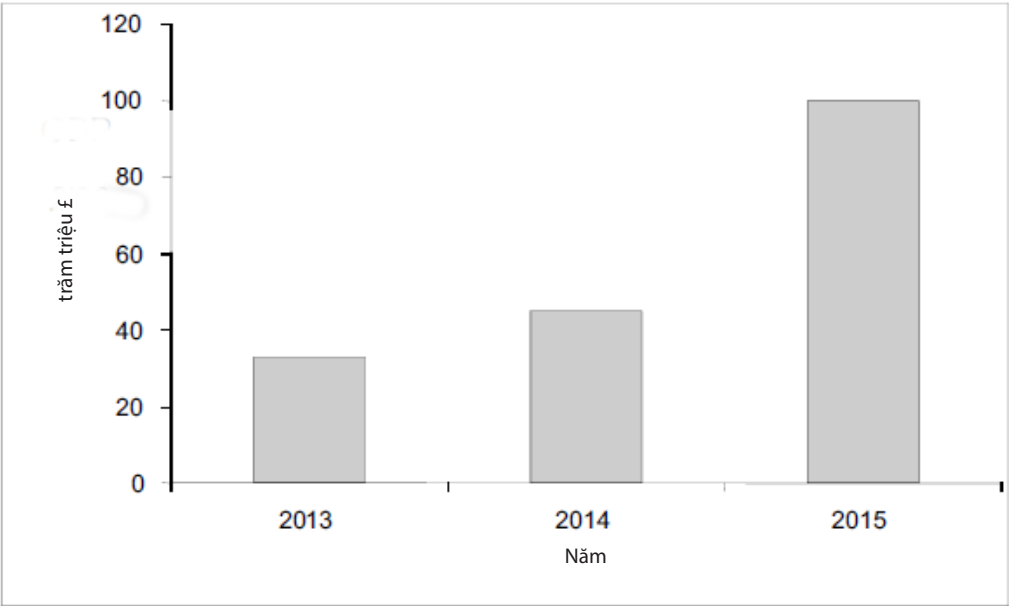
- Biểu diễn sự thay đổi trong tổng GDP bằng biểu đồ tần số.
- Biểu diễn sự thay đổi trong tổng GDP bằng biểu đồ cột.
- Biểu diễn sự thay đổi trong tổng GDP bằng:
  - biểu đồ nhiều cột
  - biểu đồ cột chồng
- Biểu diễn các thành phần GDP năm 2015 bằng biểu đồ tròn.

**Lời giải:**

- Biểu đồ tần số. Xem **Hình 3.3**.

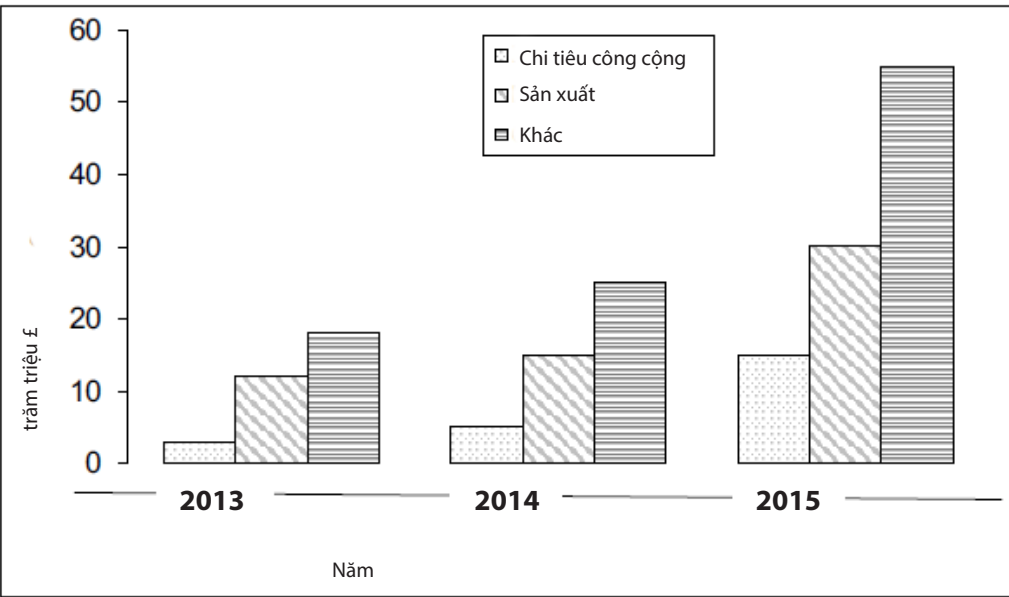
**Hình 3.3** Tổng GDP của Wyeland năm 2013, 2014 và 2015.

b) Biểu đồ cột. Xem **Hình 3.4**.



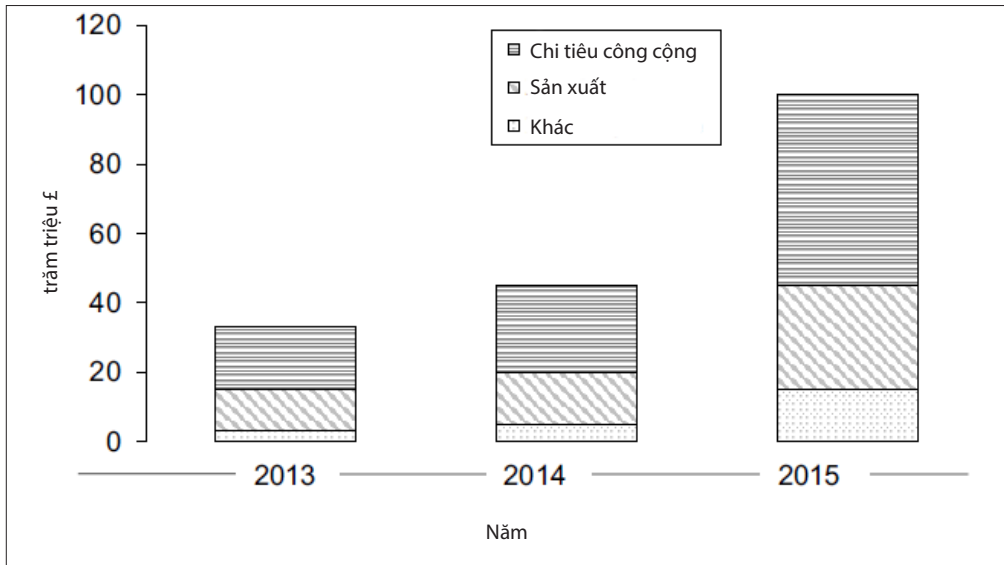
**Hình 3.4** Tổng GDP của Wyeland năm 2013, 2014 và 2015.

c) i) Biểu đồ nhiều cột. Xem **Hình 3.5**.



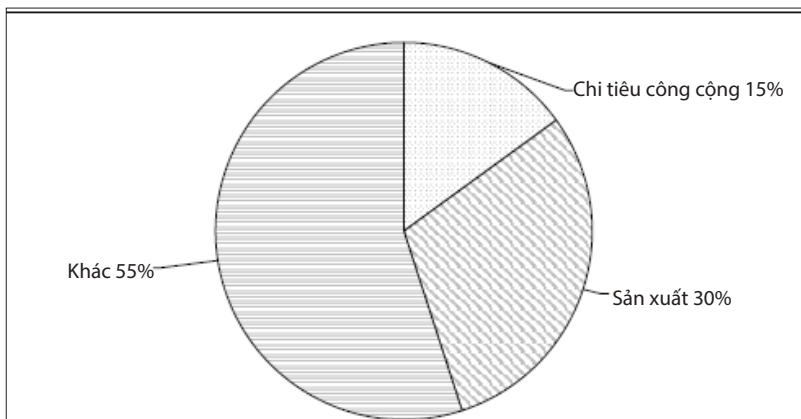
**Hình 3.5** Thành phần GDP của Wyeland năm 2013, 2014 và 2015.

iv) Biểu đồ cột chồng. Xem **Hình 3.6**.



**Hình 3.6** Thành phần GDP của Wyeland năm 2013, 2014 và 2015.

d) Biểu đồ tròn. Xem **Hình 3.7**.



**Hình 3.7** Thành phần GDP của Wyeland năm 2015.

**Bảng tính góc**

Phân loại	2015	Tính góc
Chi tiêu công cộng	15	$15/100 \times 360 = 54$
Sản xuất	30	$30/100 \times 360 = 108$
Khác	55	$55/100 \times 360 = 198$
Tổng GDP	100	360

**Bảng 3.22** Tính góc.**3.10.3 DIỄN DỊCH BIỂU ĐỒ VÀ SƠ ĐỒ**

Việc dựng sơ đồ rất dễ khiến người sử dụng hiểu nhầm. Chúng ta luôn cần thận trọng để đảm bảo rằng sơ đồ truyền tải được thông điệp chính xác của dữ liệu. Đồng thời, bạn cũng nên xem xét kỹ lưỡng sơ đồ người khác dựng, tránh trường hợp hiểu nhầm thông tin.

**3.10.4 SƠ ĐỒ BIỂU DIỄN DỮ LIỆU KHÔNG PHÂN LOẠI**

Có thể diễn tả phân phối tần số bằng nhiều cách; mà việc lựa chọn phương pháp sẽ phụ thuộc vào bản chất của thông tin yêu cầu. Không có loại biểu đồ hay sơ đồ nào đã được nhắc đến ở trên là phù hợp nếu biến được xét có tính chất liên tục, ví như thời gian hoặc độ dài.

Nếu dữ liệu được tổ chức thành bảng phân phối tần số hoặc phân phối tần số nhóm, ta có thể sử dụng biểu đồ tần số (*histogram*) hoặc đa giác tần số (*frequency polygon*) (hai loại này gần giống nhau) để mô tả sự phân phối. Xem **Hình 3.8 đến 3.10**.

Nếu dữ liệu được tổ chức thành bảng phân phối tần số tích lũy, ta có thể sử dụng đường vòm cung (*ogive*) hay đa giác tần số tích lũy (*cumulative frequency polygon*) (hai loại này là một) để mô tả sự phân phối. Xem **Hình 3.11**.

Một loại sơ đồ hữu dụng nữa là biểu đồ hộp (Box plot hoặc Box and Whisker plot), được dựng bằng các giá trị lớn nhất và nhỏ nhất của dữ liệu và ba phân vị chính. Xem **Hình 3.12**.

Trừ đa giác tần số tích lũy ra, thì không dễ dựng các sơ đồ kể trên thông qua Microsoft Excel, do đó các phần sau chỉ đưa ra thông tin đến mức độ này.

### 3.10.5 BIỂU ĐỒ TẦN SỐ

Biểu đồ tần số là một trong những sơ đồ thông dụng nhất dùng để diễn tả phân phối tần số. Nhìn qua thì tưởng giống hệt, song thực ra biểu đồ này khác với biểu đồ cột ở hai yếu tố quan trọng sau:

- Trục ngang là một đường **liên tục (continuous)**, **tuyến tính (linear)**, không phải một chuỗi các phân loại. Độ rộng của mỗi cột thể hiện độ rộng của lớp tương ứng trong bảng phân phối tần số.
- Diện tích (area)** của mỗi cột (chứ *không phải* chiều cao) tỷ lệ với tần số xảy ra trong lớp.

Những sự khác biệt đó dẫn đến hệ quả là:

- Biên trên trang luôn là **mật độ tần số (frequency density)** – tần số trên mỗi khoảng gộp đơn vị. Ví dụ, tần số trên mỗi £, tần số trên mỗi 10£, tần số mỗi 5 năm...

Nếu ta ký hiệu tần số của lớp là  $f$ , chiều cao cột là  $h$  và chiều rộng lớp là  $w$ , thì:

$$f \propto h \times w \text{ (}\propto\text{ nghĩa là “tỷ lệ với”)}$$

do đó chiều cao của cột (biên trên trang) sẽ:

$$h \propto \frac{f}{w}$$

- Không có khoảng cách giữa các cột.
- Khi phân phối tần số chứa các lớp có chiều rộng khác nhau, thì chiều cao của các cột phải được điều chỉnh để diện tích cột tỷ lệ với tần số.

**Lưu ý:** Muốn dựng biểu đồ tần số chính xác với Microsoft Excel, ta cần khéo léo một chút. Công cụ Histogram (“Biểu đồ tần số”) của Microsoft Excel trên thực tế không cho ra biểu đồ chính xác hoàn toàn.

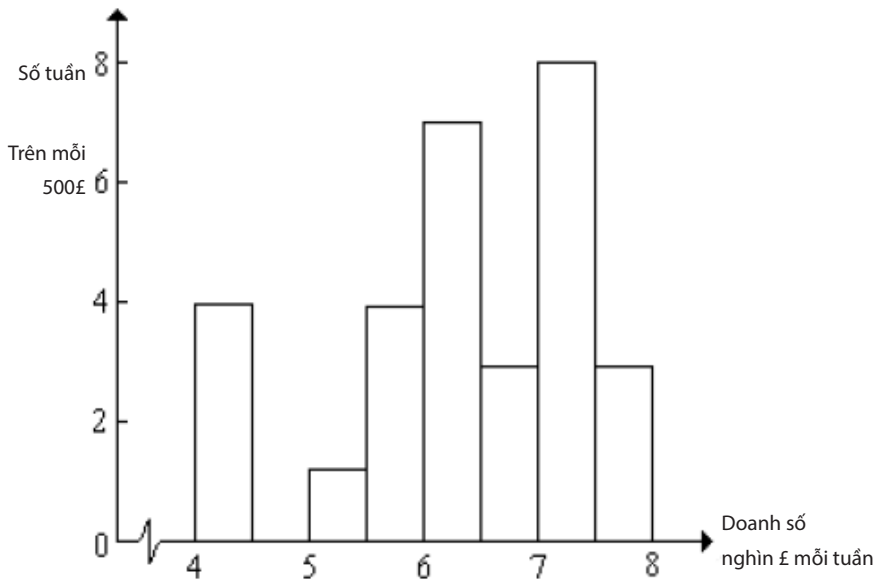
#### Ví dụ 3.13 – Dựng biểu đồ tần số

Xem lại **Ví dụ 3.6** (doanh số hàng tuần của một cửa hàng). Minh họa phân phối tần số nhóm bằng biểu đồ tần số.

**Lời giải:** Tất cả các lớp có chiều rộng như nhau (500£), nên ta chọn diễn tả “Tần số trên mỗi 500£” tương ứng với “doanh số/tuần”. Điều này có nghĩa là:

$$\text{tần số lớp} \propto \text{chiều cao cột}$$

Xem **Hình 3.8**.

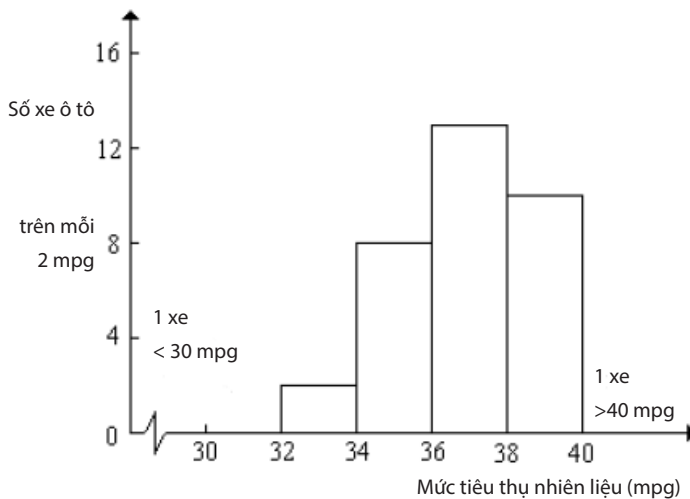


**Hình 3.8** Doanh số tuần của một cửa hàng trong 30 tuần gần đây.

### Ví dụ 3.14 – Dựng biểu đồ tần số

Xem lại **Ví dụ 3.7** (mức tiêu thụ nhiên liệu). Minh họa phân phối tần số nhóm bằng biểu đồ tần số.

Để diễn tả lớp mở, ta chỉ cần viết trên sơ đồ rằng có 1 xe với lượng tiêu thụ nhiên liệu dưới 30 mpg và 1 xe với lượng tiêu thụ trên 40 mpg.



**Hình 3.9** Mức tiêu thụ nhiên liệu của 35 xe ô tô cùng loại.

### 3.10.6 ĐA GIÁC TẦN SỐ

Đây là một biến thể có thể sử dụng thay thế cho biểu đồ tần số. Đa giác tần số được dựng bằng cách biểu diễn *mật độ tần số (frequency density)* tại trung điểm của lớp. Các điểm đã biểu diễn được nối lại, từng chấm với nhau bằng đường thẳng. Đa giác là hình kín, do đó điểm đầu và cuối sẽ được nối với trục (mật độ tần số không) ở trung điểm sát đó.

#### Ví dụ 3.15 – Dựng đa giác tần số

Xem lại **Ví dụ 3.6** (doanh số hàng tuần của một cửa hàng). Biểu diễn phân phối tần số nhóm bằng đa giác tần số.

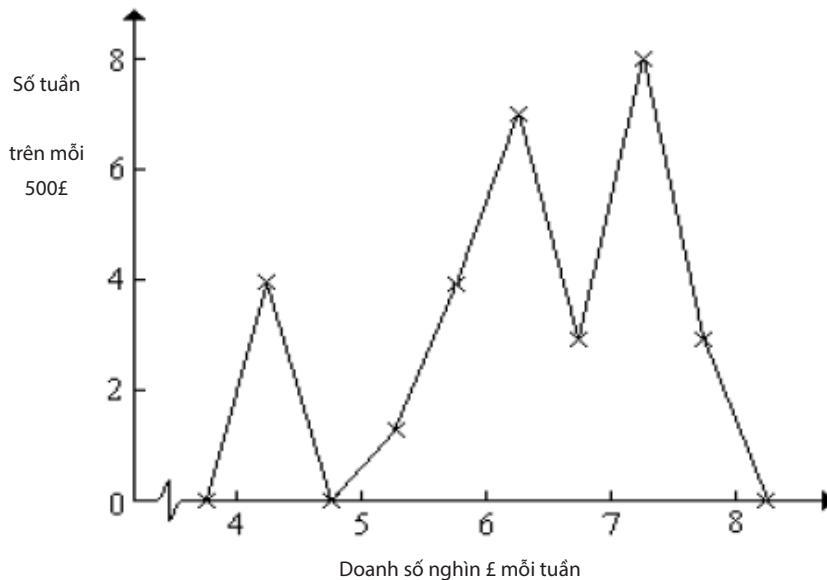
#### Lời giải:

Trong ví dụ này, các lớp đều có chiều rộng là 500£, do đó ta sẽ biểu diễn “Tần số trên mỗi 500£” ở trung điểm của lớp.

Doanh số hàng tuần (£)	Số tuần (tần số)	Trung điểm lớp, £ mỗi tuần
4000 nhưng < 4500	4	4250
4500 nhưng < 5000	0	4750
5000 nhưng < 5500	1	5250
5500 nhưng < 6000	4	5750
6000 nhưng < 6500	7	6250
6500 nhưng < 7000	3	6750
7000 nhưng < 7500	8	7250
7500 nhưng < 8000	3	7750

**Bảng 3.23 Dữ liệu doanh số.**

Đa giác tần số được đóng kín ở đầu dưới bằng cách nối điểm biểu diễn đầu tiên với trục ở 3750£; đóng kín ở đầu trên bằng cách nối điểm biểu diễn cuối cùng với trục ở 8250£.



**Hình 3.10** Doanh số tuần của một cửa hàng trong 30 tuần gần đây.

Đa giác tần số là một cách hữu hiệu để so sánh kiểu hình phân phối của hai hoặc nhiều phân phối. Khi đó ta nên đặt chồng các đa giác lên cùng trục, với điều kiện tổng tần số giống nhau hoặc tần số của các phân phối đều được chuyển thành tỷ lệ phần trăm.

### 3.10.7 ĐA GIÁC TẦN SỐ TÍCH LŨY HOẶC ĐƯỜNG VÒM CUNG

Sơ đồ này dùng để minh họa phân phối tần số tích lũy, cho phép chúng ta ước tính số lượng quan sát trong bảng phân phối rơi vào dưới một giá trị nhất định.

Để dựng đường vòm cung, ta biểu diễn tần số tích lũy với **phạm vi trên của lớp (upper class boundary)**. Từ đó nối liền các điểm biểu diễn với nhau bằng đường thẳng.

Đa giác được đóng kín ở đầu dưới bằng cách nối điểm biểu diễn đầu tiên với trục tại phạm vi dưới của lớp đầu tiên; đóng kín ở đầu trên bằng cách nối điểm biểu diễn cuối cùng trực tiếp với trục, cũng là phạm vi trên của lớp cuối cùng.

#### Ví dụ 3.16 – Dựng đa giác tần số tích lũy

Xem lại **Ví dụ 3.9** (doanh số hàng tuần của một cửa hàng). Biểu diễn phân phối tần số tích lũy bằng đường vòm cung. Ước tính xem trong bao nhiêu tuần thì cửa hàng đạt được:

- Doanh số nhỏ hơn 5250£ mỗi tuần;
- Doanh số lớn hơn 6750£ mỗi tuần;
- Doanh số nằm giữa 5250£ và 6750£ mỗi tuần.

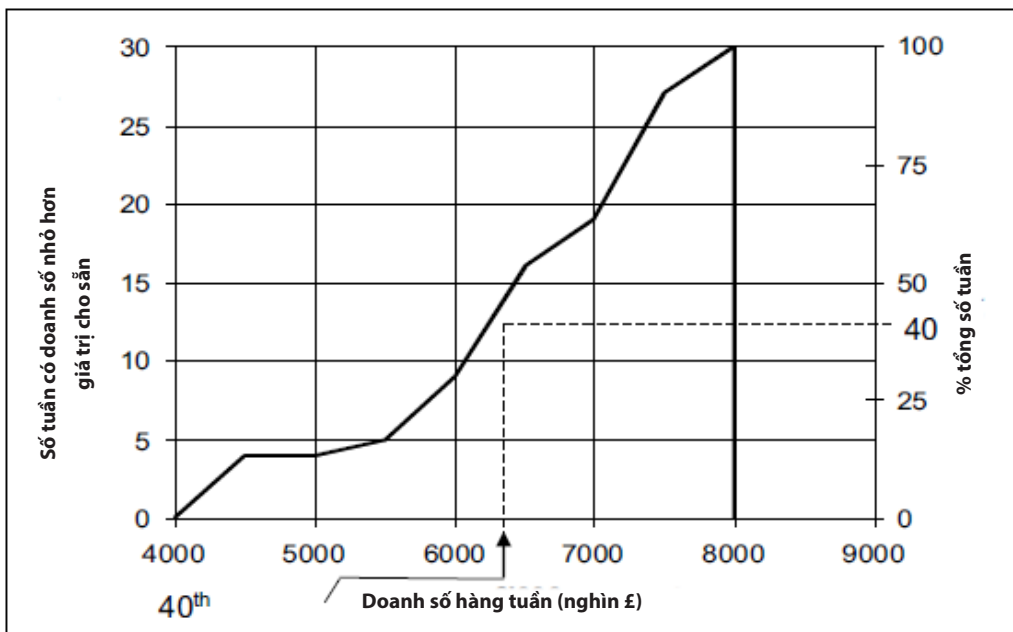


**Lời giải:** Bảng phân phối tần số tích lũy được lập lại ở đây:

Doanh số hàng tuần (£)	Số tuần có doanh số thấp hơn giá trị cho trước (tần số tích lũy)
nhỏ hơn 4000	0
nhỏ hơn 4500	4
nhỏ hơn 5000	4
nhỏ hơn 5500	5
nhỏ hơn 6000	9
nhỏ hơn 6500	16
nhỏ hơn 7000	19
nhỏ hơn 7500	27
nhỏ hơn 8000	30

**Bảng 3.24** Dữ liệu doanh số.

Ta dựng được đường vòm cung như ở **Hình 3.11**.



**Hình 3.11** Đường vòm cung doanh số của một cửa hàng trong 30 tuần gần đây.

- a) Theo đồ thị ta thấy doanh số nhỏ hơn 5250£ mỗi tuần trong tổng số 4,5 tuần.
- b) Theo đồ thị ta thấy doanh số nhỏ hơn 6750£ mỗi tuần trong tổng số 17,5 tuần.
- c) Từ a) và b), ta thấy doanh số nằm giữa 5250£ và 6750£ mỗi tuần trong tổng số  $(17,5 - 4,5) = 13$  tuần.

Tần số tích lũy dạng phần trăm thường có ích trong việc dựng đường vòm cung. Có thể dễ dàng ước tính phân vị từ đường vòm cung phần trăm.

Đồ thị ở **Hình 3.11** có thể chuyển thành đường vòm cung phần trăm bằng việc điều chỉnh tỷ lệ trên trang. Hãy quan sát phía bên phải hình. Ta có thể dựa vào tỷ lệ này để ước tính, ví như phân vị thứ 40.

Phân vị thứ 40 (giá trị mà 40% quan sát rơi phía dưới) là 6220£ (làm tròn đến 10£).

Với dạng đồ thị này, ta không cần điều chỉnh nếu chiều rộng các lớp không đều. Tuy nhiên, cần thận trọng để đảm bảo tỷ lệ trên trang là tuyến tính.

### 3.10.8 BIỂU ĐỒ HỘP

Biểu đồ hộp là mô hình đồ họa mô tả hành vi của giá trị dữ liệu trong khoảng giữa của phân phối, đồng thời chỉ ra sự tồn tại và vị trí của giá trị ngoại lệ (*outlier*) bất kỳ. Có thể thấy vị trí giá trị ngoại lệ tương quan với khoảng giữa của phân phối.

Biểu đồ hộp là một cách hữu hiệu để so sánh hai hoặc nhiều mô hình phân phối với nhau; cho phép đánh giá kiểu hình chung của phân phối; thể hiện rõ các điểm bất đối xứng. Tuy nhiên, với phân phối có hai yếu vị (bi-modal) thì ta sẽ không thấy được thông tin này.

Trong cuốn sách này, chúng ta sẽ dùng một phương pháp ước lượng để dựng đồ thị, dựa vào trung vị, hai tứ phân vị và khoảng giá trị trong dữ liệu. Với đa số tình huống, việc ước lượng này sẽ không làm phát sinh sai sót đáng kể nào.

Biểu đồ hộp được dựng bằng cách vẽ một khối hộp, có các đỉnh tương ứng với tứ phân vị (do đó hình hộp sẽ chiếm khoảng 50% tổng phân phối); vị trí của trung vị được đánh dấu trong khối hộp. Mỗi góc hộp có một đường thẳng lần lượt nối ra đến các giá trị nhỏ nhất và lớn nhất của phân phối.

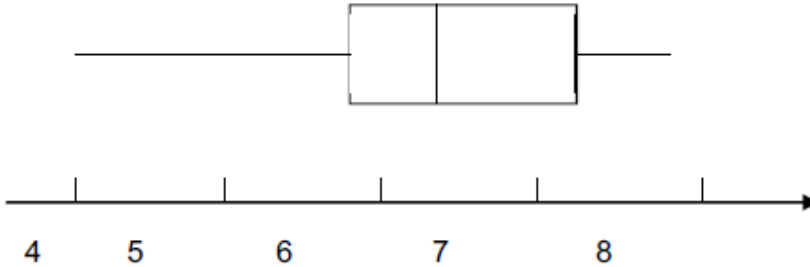
#### Ví dụ 3.17 – Dựng biểu đồ hộp

Dựng biểu đồ hộp cho dữ liệu ở **Ví dụ 3.4**. (Doanh số hàng tuần của một cửa hàng)

**Lời giải:** Từ tập dữ liệu ban đầu, doanh số nhỏ nhất là 4000£ mỗi tuần và doanh số lớn nhất là 7900£ mỗi tuần.

Từ đường vòm cung ở **Hình 3.11**, ước tính được phân vị thứ nhất  $Q_1$  là 5800£, phân vị thứ ba  $Q_3$  là 7200£ và trung vị là 6400£.

Giờ ta đã có thể dựng biểu đồ hộp. Xem **Hình 3.12**.



**Hình 3.12** Doanh số hàng tuần của một cửa hàng trong 30 tuần gần đây (nghìn £).

### 3.11 KẾT LUẬN

Việc lựa chọn biểu đồ phù hợp ảnh hưởng rất nhiều đến khả năng truyền tải thông tin chính xác. Đặc biệt, trong các tình huống cần đưa ra quyết định, chính việc lựa chọn sơ đồ và đồ thị đúng sẽ mang đến những quyết định có hiệu quả được đưa ra ở nhiều cấp độ quản lý. Do đó, đặc biệt cần trọng khi lựa chọn, tạo dựng và trình bày biểu đồ/sơ đồ thể hiện thông tin quản lý.

### 3.12 TÀI LIỆU THAM KHẢO CHƯƠNG

1. Morris, C, Quantitative Approaches in Business Studies (*tạm dịch*: Phương pháp định lượng trong nghiên cứu kinh doanh), tái bản lần thứ 8, FT/Prentice Hall, 2012 (eBook, tái bản lần thứ 7, 2008), Chương 5.
2. Oakshott, L, Essential Quantitative Methods for Business (*tạm dịch*: Phương pháp nghiên cứu định lượng thiết yếu trong kinh doanh, quản lý & tài chính), tái bản lần thứ 5, Palgrave Macmillan, 2012, Chương 5.



## Chương 4

# TÓM TẮT DỮ LIỆU

### 4.1 GIỚI THIỆU

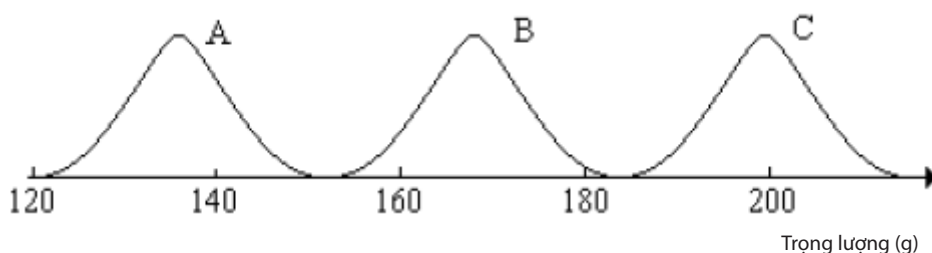
Trong chương trước, chúng ta đã xét đến các cách tổ chức và trình bày dữ liệu thô để đưa ra được thông tin hữu ích. Thông tin này chủ yếu mang tính chất mô tả. Ví dụ, ta có thể biết được kiểu phân phối dữ liệu đã đối xứng hay chưa, có giá trị cực biên hay không, khoảng dao động các giá trị là gì...

Trong chương này, ta sẽ thảo luận các phương pháp tóm tắt dữ liệu nhằm thu được nhiều thông tin hơn. Có thể xác định hai hoặc ba giá trị biểu trưng cho phân phối của toàn bộ tập dữ liệu. Các giá trị này được gọi là số đo. Số đo tóm tắt có thể giúp chúng ta so sánh các tập dữ liệu khác nhau, hoặc phân tích thống kê chuyên sâu hơn.

### 4.2 MỘT SỐ KHÁI NIỆM NỀN TẢNG VỀ TÓM TẮT DỮ LIỆU

Chúng ta sẽ xét đến một số khái niệm nền tảng, giúp xây dựng và củng cố kiến thức về tầm quan trọng của tóm tắt dữ liệu nói chung.

Nhìn vào sơ đồ phân phối trong **Hình 4.1**. Những phân phối này có thể đến từ một chiếc máy đóng gói trà với 3 loại cỡ túi trà khác nhau.

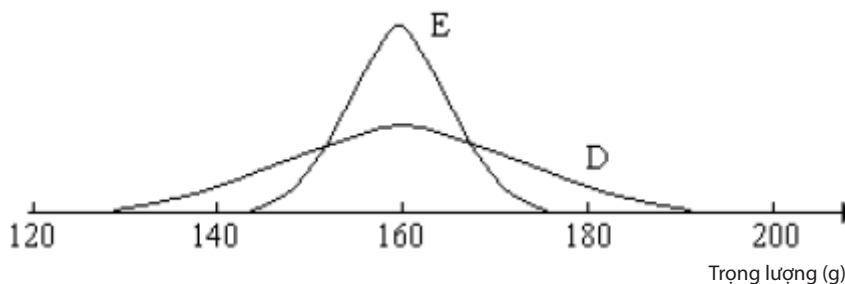


**Hình 4.1** Phân phối trọng lượng quy chuẩn của một máy đóng gói.

Các phân phối có kích cỡ và kiểu hình tương tự nhau, nhưng mỗi phân phối bao quát một dãy giá trị khác nhau. Do đó chúng ta có thể tóm tắt từng phân phối bằng cách mô tả kích cỡ của các giá trị trong phân phối đó – nói cách khác, ta cần một giá trị

**quy chuẩn (typical)**, có thể đại diện cho toàn bộ tập dữ liệu. Giá trị đơn này thường được gọi là **độ đo vị trí (measure of location)**.

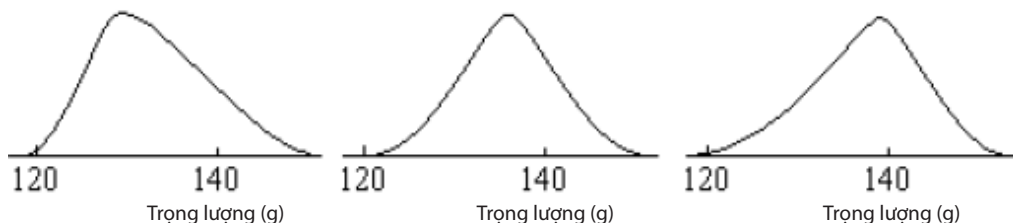
Giờ ta hãy xem hai phân phối trong **Hình 4.2**. Các phân phối này có thể đến từ hai máy đóng gói trà khác nhau, dùng để đóng các túi trà *cùng một kích cỡ*.



**Hình 4.2** Phân phối trọng lượng quy chuẩn của hai máy đóng gói.

Độ đo vị trí mà ta chọn cho phân phối D có thể khá tương đồng với độ đo vị trí mà ta chọn cho phân phối E, tuy rõ ràng hai phân phối này là khác nhau. Như vậy ngoài giá trị đại diện, ta cũng phải cân nhắc độ phân tán của tập dữ liệu tương ứng với giá trị đó. Nói cách khác, chúng ta cần đến **độ phân tán (measure of spread hoặc measure of dispersion)**, bên cạnh độ đo vị trí.

Khía cạnh quan trọng cuối cùng của phân phối là mức độ đối xứng. Hãy xem ba phân phối trong **Hình 4.3**:



**Hình 4.3** Phân phối có mức độ đối xứng khác nhau.

Mức độ đối xứng của một phân phối có thể tính toán được, tuy nhiên trong học phần này, ta sẽ chỉ mô tả độ đối xứng này bằng lời – “đối xứng” (*symmetrical*), “lệch trái” (*peak to the left*), “lệch phải” (*peak to the right*).

Giờ ta sẽ xét đến các tính chất này một cách chi tiết hơn.

### 4.3 GIÁ TRỊ QUY CHUẨN HOẶC TRUNG BÌNH

Có một số cách khác nhau để diễn tả giá trị “quy chuẩn” của một tập dữ liệu. Việc lựa chọn phương pháp phụ thuộc vào bản chất của dữ liệu, mục đích sử dụng giá trị đó cũng như khối lượng tính toán định thực hiện. Các giá trị đo thường được sử dụng nhất là:

**Yếu vị (mode)** – đây là (các) giá trị thường xuất hiện nhất trong tập dữ liệu, được ký hiệu là  $m$  trong chương này.

**Trung vị (median)** – đây là phân vị thứ 50, nghĩa là sau khi được sắp xếp theo thứ tự kích cỡ thì 50% tổng giá trị rơi dưới giá trị đó. Trong chương này, trung vị được ký hiệu là  $M$ .

**Trung bình cộng (arithmetic mean)** – là tổng của mọi giá trị trong tập dữ liệu chia cho số các giá trị, ký hiệu là  $\bar{x}$ . Số đo này thường được gọi là “số trung bình” (*average*).

#### 4.3.1 TÍNH YẾU VỊ, TRUNG VỊ VÀ TRUNG BÌNH CỘNG CỦA DỮ LIỆU KHÔNG THƯỜNG XUYỀN (NON-FREQUENCY DATA)

##### Ví dụ 4.1 – Tính yếu vị, trung vị và trung bình cộng

Công ty trách nhiệm hữu hạn đại chúng Shefftex có 90 nhân viên điều hành máy. Giám đốc Nhân sự công ty này đang xem xét tình hình nhân viên nghỉ phép, vắng mặt. Trong vòng 25 ngày làm việc liên tiếp, số lượng nhân viên vắng mặt từng ngày như sau:

4, 2, 4, 3, 5, 9, 5, 6, 6, 3, 3, 6, 4, 3, 3, 2, 3, 5, 4, 3, 3, 2, 4, 3, 5

Hãy xác định: a) yếu vị; b) trung vị; c) trung bình số lượt vắng mặt mỗi ngày.

**Lời giải:**

a) Yếu vị: Sắp xếp dữ liệu thành một phân phối tần số, ta có bảng sau:

Số lượt vắng mặt trong một ngày	Số ngày
2	3
3	9
4	5
5	4
6	3
7	-
8	-

Số lượt vắng mặt trong một ngày	Số ngày
9	1
Tổng	25

**Bảng 4.1** Dữ liệu vắng mặt, ví dụ 1.

Giá trị 3 xuất hiện 9 lần. Đây là giá trị xuất hiện thường xuyên nhất, như vậy giá trị yếu vị là 3 lượt vắng mặt mỗi ngày.

Lưu ý: Nếu có từ hai hoặc nhiều giá trị với tần số xuất hiện nhiều nhất, thì tập dữ liệu này có nhiều hơn một yếu vị, nghĩa là đa yếu vị (*multimodal*). Ví dụ:

Số lượt vắng mặt trong một ngày	Số ngày
2	3
3	6
4	3
5	6
6	2
7	1
Tổng	21

**Bảng 4.2** Dữ liệu vắng mặt, ví dụ 2.

Các giá trị 3 và 5 xuất hiện thường xuyên nhất, do đó chúng đều là yếu vị. Như vậy, tập dữ liệu này có **hai yếu vị (bi-modal)**.

Khi có nhiều hơn một giá trị yếu vị, thì yếu vị sẽ không còn được coi là giá trị “quy chuẩn” nữa.

- b) Trung vị: Sắp xếp dữ liệu ban đầu theo thứ tự kích cỡ, ta được dãy sau:  
2, 2, 2, 3, 3, 3, 3, 3, 3, 3, 3, 3, 4, 4, 4, 4, 4, 5, 5, 5, 5, 6, 6, 6, 9

Trung vị là giá trị nằm giữa tập dữ liệu. Trong tập này, giá trị giữa nằm ở vị trí thứ 13 – tức có giá trị 4 lượt vắng mặt trong một ngày.

Lưu ý: Tập dữ liệu trong ví dụ trên có số lượng giá trị lẻ, nên dễ dàng xác định giá trị nằm giữa.

$$\frac{n+1}{2} = \frac{25+1}{2} = \text{giá trị thứ 13}$$



Tuy nhiên, nếu tập dữ liệu có 30 ngày làm việc thì:

$$\frac{n+1}{2} = \frac{30+1}{2} = \text{giá trị thứ 15,5}$$

Nói cách khác, ta có hai giá trị nằm giữa – giá trị thứ 15 và thứ 16. Thông thường ta lấy trung vị là trung bình của hai giá trị này.

Ví dụ, nếu giá trị thứ 15 tương ứng với 3 lượt vắng mặt mỗi ngày và giá trị thứ 16 tương ứng với 4 lượt vắng mặt, thì ta lấy trung vị là:

$$\frac{3+4}{2} = 3,5 \text{ lượt vắng mặt mỗi ngày}$$

- c) Trung bình cộng – đây là tổng tất cả các giá trị trong tập dữ liệu chia cho số các giá trị:

$$\bar{x} = \frac{4+2+4+3+\dots+3+5}{25} = 4,0 \text{ lượt vắng mặt mỗi ngày}$$

Giá trị trung bình cộng cũng có thể được xác định thông qua các chức năng thống kê trên máy tính bỏ túi hoặc một gói phần mềm, như Microsoft Excel.

Nếu tập dữ liệu đủ nhỏ, ta có thể ước tính trung bình cộng khi nhìn vào kiểu hình phân phối tần số. Giá trị ước tính này có thể dùng để kiểm tra lại kết quả tính toán. Trong ví dụ trên, đa số các giá trị là 2, 3 hoặc 4, do đó ta đoán trung bình cộng nằm trong khoảng 3 hoặc 4, nhưng khả năng cao là gần 4 hơn 3 vì giá trị đơn lẻ 9 sẽ khiến trung bình cộng trở nên lớn hơn.

Trong thực tiễn, do các tập dữ liệu doanh nghiệp thường lớn và được lưu trữ điện tử nên hiển nhiên những công cụ bảng tính (như trong Microsoft Excel) là cách để tìm các giá trị tóm tắt dữ liệu kể trên. Xem thêm Mục 3.3 để tìm hiểu chi tiết hơn.

#### **Ví dụ 4.2 – Độ nhạy của yếu vị, trung vị và trung bình cộng**

Ta xét lại tình huống trong **Ví dụ 1**. Bây giờ Giám đốc Nhân sự phát hiện ra là dữ liệu 5 lượt vắng mặt của ngày 18 đúng ra là có 7 lượt vắng mặt. Sự thay đổi này sẽ ảnh hưởng đến giá trị của ba độ đo vị trí như thế nào?

#### **Lời giải:**

Xem lại quá trình xác định yếu vị và trung vị ở trên, bạn có thể thấy việc thay đổi một giá trị từ 5 thành 7 không khiến yếu vị hay trung vị khác đi.

Tuy nhiên, giá trị trung bình cộng mới là:

$$\bar{x} = \frac{102}{25} = 4,1 \text{ lượt vắng mặt mỗi ngày}$$

Trong trường hợp này, khi có sự thay đổi trong tập dữ liệu, trung bình cộng là độ đo vị trí duy nhất bị ảnh hưởng, giá trị đo đã khác đi một chút.

#### **4.3.1.1 Ưu điểm và nhược điểm của ba loại độ đo vị trí**

##### **Yếu vị**

##### ***Ưu điểm:***

- Giá trị xuất hiện thường xuyên nhất có thể là độ đo hữu dụng/chứa thông tin nhất trong nhiều trường hợp. Ví dụ, trong nghiên cứu thị trường, chúng ta quan tâm đến ý kiến số đông về một sản phẩm, sự kiện, chính sách...
- Việc tính yếu vị rất dễ dàng.
- Có thể áp dụng với dữ liệu không phải dạng số.
- Tính chất mở của một số phân phối không ảnh hưởng đến quá trình tìm yếu vị.

##### ***Nhược điểm:***

- Có khả năng không tồn tại giá trị thích hợp, ví dụ trong tình huống tất cả giá trị trong tập dữ liệu đều khác nhau.
- Nếu phân phối có hai yếu vị, thì sẽ có hai giá trị yếu vị, vì thế không còn được coi là giá trị “quy chuẩn” nữa.
- Trong một phân phối rất bất đối xứng, yếu vị có thể là một giá trị cực hạn nên không thích hợp để coi là giá trị quy chuẩn.
- Không thể dùng yếu vị để tính toán thống kê sâu hơn.
- Giá trị yếu vị không bao hàm mọi điểm dữ liệu.
- Thay đổi nhỏ trong tập dữ liệu không ảnh hưởng đến giá trị của yếu vị.
- Với các đặc điểm trên, yếu vị thường được dùng để mô tả hơn là phân tích phân phối.

**Trung vị:*****Ưu điểm:***

- Việc xác định được giá trị mà 50% phân phối rơi bên dưới có thể hữu ích, chứa nhiều thông tin.
- Giá trị của trung vị không bị ảnh hưởng bởi các giá trị dữ liệu ngoại lệ hoặc tính mở của một số phân phối.

***Nhược điểm:***

- Giá trị trung vị không bao hàm mọi điểm dữ liệu.
- Thay đổi nhỏ trong các giá trị dữ liệu không được phản ánh tương ứng trên giá trị trung vị.
- Không thể kết hợp trung vị của các phân phối khác nhau để tạo thành một trung vị chung.
- Không thể dùng trung vị để tính toán thống kê sâu hơn.
- Với các đặc điểm ở trên, trung vị thường được dùng để mô tả hơn là phân tích phân phối.

**Trung bình cộng*****Ưu điểm:***

- Tất cả giá trị trong tập dữ liệu đều góp phần tạo ra giá trị trung bình cộng.
- Có thể “gom” các giá trị từ hai hoặc nhiều tập dữ liệu để tính giá trị trung bình cộng chung cho tập dữ liệu kết hợp. Xem Mục 3.4 để tìm hiểu chi tiết hơn.
- Có thể dùng độ đo này cho phân tích thống kê sâu hơn; thay vì chỉ phục vụ mục đích mô tả đơn thuần.
- Độ đo này có thể phản ánh tương ứng với các thay đổi nhỏ trong tập dữ liệu.

***Nhược điểm:***

- Giá trị trung bình cộng bị ảnh hưởng bởi các giá trị ngoại lệ, do đó trong các phân phối rất bất đối xứng thì việc coi đây là giá trị quy chuẩn sẽ không thỏa đáng.
- Có thể khó diễn dịch giá trị trung bình cộng. Ví dụ, trong nghiên cứu về thành phần trong hộ gia đình tại Anh, trung bình số trẻ em trong mỗi gia đình là 2,47 – như vậy có nghĩa là gì? Số lượng trẻ em không phải là số nguyên thì có hợp lý không?

### 4.3.2 ĐỘ ĐO HỮU ÍCH KHÁC

Trong trường hợp này, tứ phân vị có thể hữu ích trong việc mô tả sự phân phối các giá trị của một biến. Khái niệm về tứ phân vị đã được nhắc đến ở chương trước.

Tứ phân vị thứ 1 (dưới),  $Q_1$ , là phân vị thứ 25 của một tập giá trị. 25% các giá trị của biến nằm dưới phân vị này.

Tứ phân vị thứ 3 (trên),  $Q_3$ , là phân vị thứ 75 của một tập giá trị. 75% các giá trị của biến nằm dưới phân vị này.

Thông thường, tứ phân vị được ước lượng dựa vào đường vòm cung hoặc tính bằng công cụ bảng tính.

### 4.3.3 SỬ DỤNG CÔNG CỤ MICROSOFT EXCEL TƯƠNG ƯNG

Microsoft Excel có một số công cụ tích hợp sẵn có thể dùng để tính các giá trị tóm tắt dữ liệu đã được thảo luận trong chương này. Các công cụ tương ứng là:

Công cụ Microsoft Excel	Chức năng
=MODE()	Tính yếu vị của tập dữ liệu
=MEDIAN()	Tính trung vị của tập dữ liệu
=AVERAGE()	Tính trung bình cộng của tập dữ liệu
=QUARTILE()	Tính tứ phân vị của tập dữ liệu

**Bảng 4.3** Các công cụ Microsoft Excel tính giá trị tóm tắt dữ liệu.

Các công cụ này phải được sử dụng với dữ liệu thô, nghĩa là không thể áp dụng với dữ liệu đã được sắp xếp thành bảng phân phối tần số. Phần ngoặc tròn () chứa khoảng dao động các giá trị dữ liệu. Với các tứ phân vị, chúng ta phải xác định cụ thể mình cần tính giá trị tứ phân vị nào. Hãy xem **Ví dụ 4.3**.

### Ví dụ 4.3 – Sử dụng các công cụ Microsoft Excel

	A	B	C
1	Sinh viên	Điểm SAT /10	
2	Adey	6	
3	Adrian	3	
4	Ashling	5	
5	Ben	8	
6	Catherine	5	
7	Daniel	6	
8	Dipen	6	
9	Emma	7	
10	Frances	2	
11	Hannah	7	
12	Jessica	7	
13	Jonathan	9	
14	Joseph	7	
15	Khalil	7	
16	Lauren	8	
17	Luke	5	
18	Matthew	5	
19	Max	5	
20	Michael	4	
21	Naomi	7	
22	Nicole	8	
23	Shahid	6	
24	Simon	5	

Dữ liệu cho thấy điểm kiểm tra tự đánh giá (SAT) của 25 sinh viên.

Các ô B28:B32 chứa hàm công cụ cần nhập.

Các ô C28:C32 chứa kết quả của ô B28 đến B32.

25	<b>Susan</b>	6	
26	<b>Thomas</b>	8	
27			
28	yếu vị	=mode(B2:B26)	5
29	trung vị	=median(B2:B26)	6
30	trung bình cộng	=average(B2:B26)	6.08
31	tứ phân vị $Q_1$	=quartile(B2:B26,1)	5
32	tứ phân vị $Q_3$	=quartile(B2:B26,3)	7

**Bảng 4.4** Ví dụ sử dụng Microsoft Excel để tính giá trị tóm tắt.

#### 4.3.4 TÍNH TRUNG BÌNH CỘNG CỦA CÁC GIÁ TRỊ KẾT HỢP

##### 4.3.4.1 Trung bình cộng chung của hai nhóm giá trị trở lên

Cho biết thông tin lương của nhân viên dưới cấp quản lý ở một công ty lớn như sau:

Nhóm nhân viên	Số lượng	Lương trung bình hàng năm (£)
Nam	74	18,274
Nữ	57	16,962

**Bảng 4.5** Dữ liệu lương.

(Chú ý: Trong phần này, “trung bình” tương đương với “trung bình cộng”).

Giả sử thông tin bạn cần là mức lương trung bình của toàn bộ nhân viên, không phân biệt giới tính. Bạn cũng không được phép truy cập dữ liệu thô, tức là mức lương của từng cá nhân.

Nhớ lại công thức tính giá trị trung bình, ta có thể xác định mức lương trung bình của nhóm kết hợp:

Trung bình cộng = tổng tất cả giá trị / tổng số giá trị

Trong bảng trên, lương trung bình của nhóm nhân viên nam được tính như sau:

Lương trung bình nhóm nam = tổng lương nhóm nam / 74 = 18.274£

Do đó, tổng lương nhóm nam =  $18.274\text{£} \times 74 = 1.352.276\text{£}$

Tương tự với nhóm nhân viên nữ, tổng lương nhóm nữ =  $16.962\text{£} \times 57 = 966.834\text{£}$

Suy ra tổng toàn bộ lương (của cả nhóm nam + nhóm nữ) =  $1.352.276\text{£} + 966.834\text{£} = 2.319.110\text{£}$

Do đó mức lương trung bình của toàn bộ là tổng toàn bộ lương / tổng số người:

$2.319.110\text{£} / (74 + 57) = 2.319.110\text{£} / 131 = \mathbf{17.703\text{£}}$  (làm tròn đến 1£).

**Quy trình chung (general procedure)** cho phép tính này được mô tả như sau:

Giả sử ta có một tập hợp giá trị trung bình cộng tương ứng với nhiều nhóm con của cùng một biến, như mức lương trung bình của từng giới trong ví dụ trên, hoặc lương theo nhóm tuổi, hoặc điểm thi học phần Phân tích Dữ liệu cho Quyết định Kinh doanh của từng khóa...

Muốn biết trung bình cộng của toàn bộ nhóm, ta tính như sau:

(Số giá trị nhóm 1 x trung bình nhóm 1 + Số giá trị nhóm 2 x trung bình nhóm 2 + ... + Số giá trị nhóm n x trung bình nhóm n) / tổng số giá trị trong tất cả các nhóm.

#### **4.3.4.2 Trung bình cộng có trọng số (hay bình quân gia quyền)**

Với phương pháp tính trung bình cộng ở trên, ta giả sử rằng tất cả dữ liệu đều có mức độ quan trọng tương đương nhau. Tuy nhiên không phải lúc nào điều này cũng đúng. Ví dụ, người ta đánh giá học phần qua cả phần nghiên cứu lẫn phần kiểm tra, nhưng trọng số của hai phần này không tương đương. Với học phần Phân tích Dữ liệu cho Quyết định Kinh doanh, tỷ trọng phần nghiên cứu / phần kiểm tra là 60 / 40. Nếu cần tính điểm trung bình của một sinh viên đang tham gia học phần này, ta phải cân nhắc đến yếu tố trọng số nữa.

Ta tính như sau:

$$(60 \times \text{điểm nghiên cứu} + 40 \times \text{điểm kiểm tra}) / (60 + 40)$$

Do đó, nếu đạt 70% điểm nghiên cứu và 55% điểm kiểm tra, thì điểm trung bình cộng có trọng số cho học phần này của sinh viên đó là:

$$(60 \times 70 + 40 \times 55) / (60 + 40) = (4200 + 2200) / 100 = 64\%$$

Tóm lại, giá trị trung bình cộng có trọng số bằng:

$$(\text{tỷ trọng}_1 \times \text{giá trị}_1 + \text{tỷ trọng}_2 \times \text{giá trị}_2 + \text{tỷ trọng}_3 \times \text{giá trị}_3 + \dots + \text{tỷ trọng}_n \times \text{giá trị}_n) / \text{tổng các tỷ trọng}$$

**Ví dụ 4.4 – Trung bình cộng có trọng số**

Bảng tính ở **Hình 4.4** cho thấy điểm của một sinh viên cho học phần Phân tích Dữ liệu cho Quyết định Kinh doanh.

Điểm được nhập vào các ô C5, E5 và F5; đồng thời ở các ô K5:M5 trong **Hình 4.5**.

Ô G5 chứa điểm trung bình có trọng số cho phần nghiên cứu.

Ô I5 chứa điểm trung bình có trọng số cho toàn bộ học phần.

Các giá trị ở cột C3:F3 là trọng số cho mỗi cấu phần của phần nghiên cứu, ví dụ: điểm SAT chiếm 12% tổng điểm nghiên cứu, điểm Đồ án 1 chiếm 25% tổng điểm nghiên cứu...

Ô G2 và H2 lần lượt chứa tỷ trọng điểm nghiên cứu và điểm kiểm tra.

Điểm ở ô D5 là tổng điểm có trọng số của ba bài tập Microsoft Excel. Các điểm lẻ và trọng số được trình bày ở **Hình 4.5**.

	A	B	C	D	E	F	G	H	I
1	<b>Phân tích Kinh doanh</b>		<b>Danh mục Nghiên cứu</b>						
2							0.60	0.40	
3		Trọng số	0.12	0.26	0.25	0.37	1.00		
4	<b>Tên sinh viên</b>		<b>SAT</b>	<b>Mô hình Excel</b>	<b>Đồ án 1</b>	<b>Đồ án 2</b>	<b>Tổng nghiên cứu</b>	<b>Kiểm tra</b>	<b>Tổng chung</b>
5	Bloggs	Jack	67		52	56		42	

**Hình 4.4** Điểm của một sinh viên học phần Phân tích Dữ liệu cho Quyết định Kinh doanh.

	<b>K</b>	<b>L</b>	<b>M</b>
1	0.07	0.12	0.07
2	<b>Bài tập Excel 1</b>	<b>Bài tập Excel 2</b>	<b>Bài tập Excel 3</b>
3	84	85	73

**Hình 4.5** Chi tiết điểm của một sinh viên học phần Phân tích Dữ liệu cho Quyết định Kinh doanh.



Đầu tiên, ta phải tính giá trị nhập cho ô D5 – trung bình có trọng số của ba loại điểm trong các ô K5:M5. Ta tính như sau:

Trung bình cộng:  $(0,07 \times 84 + 0,12 \times 85 + 0,07 \times 73) / (0,07 + 0,12 + 0,07) = 81,5\%$

Trong Microsoft Excel: Nhập hàm sau vào ô D5:

$= (K1 * K2 + L1 * L2 + M1 * M2) / \text{SUM}(K1:M1)$ . Ô D5 sẽ xuất hiện giá trị 82.

Việc tiếp theo là tính giá trị nhập cho ô G5 – trung bình có trọng số của tất cả các câu phần nghiên cứu. Ta tính như sau:

Trung bình cộng:  $(0,12 \times 67 + 0,26 \times 82 + 0,25 \times 52 + 0,37 \times 56) / (0,12 + 0,26 + 0,25 + 0,37) = 63,1\%$

Trong Microsoft Excel: Nhập hàm sau vào ô G5:

$= (C3 * C5 + D3 * D5 + E3 * E5 + F3 * F5) / \text{SUM}(C3:F3)$ . Ô G5 sẽ xuất hiện giá trị 63.

Việc cuối cùng là tính giá trị nhập cho cột I5 – trung bình có trọng số của tất cả điểm nghiên cứu và điểm kiểm tra. Ta tính như sau:

Trung bình cộng:  $(0,60 \times 63 + 0,40 \times 42) / (0,60 + 0,40) = 54,6\%$

Trong Microsoft Excel: Nhập hàm sau vào ô I5:

$= (G2 * G5 + H2 * H5) / \text{SUM}(G2:H2)$ . Ô I5 sẽ xuất hiện giá trị 55.

**Lưu ý:** Nếu các tỷ trọng có tổng bằng 1 như ở mục tính thứ nhất và thứ hai phía trên, thì ta không nhất thiết phải chia cho tổng tỷ trọng. Ở mục tính cuối cùng, nếu ta sử dụng tỷ trọng 60 và 40 (thay vì 0.60 và 0.40), thì ta sẽ cần chia cho tổng tỷ trọng, nghĩa là chia cho 100.

Trong ví dụ này, rõ ràng ta không nên chọn tính giá trị đối với từng sinh viên một. Do có sẵn một danh sách sinh viên, nên ta sẽ điền công thức tính tự động cho toàn bộ danh sách. Tuy nhiên, để công cụ này hoạt động thì ta cần tạo tham chiếu ô tuyệt đối cho các giá trị trọng lượng – tức là cố định ô liên quan trong công thức. Để tạo tham chiếu ô tuyệt đối, ta chèn dấu \$ vào trước phần tham chiếu muốn cố định – hàng, cột hoặc cả hai. Do đó, ba công thức trên cho ô D5, G5 và I5 nên được sửa lần lượt như sau:

$= (K\$1 * K2 + L\$1 * L2 + M\$1 * M2) / \text{SUM}(K\$1:M\$1)$

$= (C\$3 * C5 + D\$3 * D5 + E\$3 * E5 + F\$3 * F5) / \text{SUM}(C\$3:F\$3)$

$= (G\$2 * G5 + H\$2 * H5) / \text{SUM}(G\$2:H\$2)$

4.3.5 ẢNH HƯỞNG CỦA SỰ THAY ĐỔI TOÀN BỘ GIÁ TRỊ DỮ LIỆU ĐẾN ĐỘ ĐO VỊ TRÍ

Khi thường xuyên xử lý dữ liệu và số đo tóm tắt, ta thấy giá trị dữ liệu thay đổi khi thêm/bớt một số nhất định hoặc tăng/giảm một số phần trăm nào đó. Nếu tất cả các giá trị dữ liệu đều thay đổi giống nhau, thì ta có thể điều chỉnh các giá trị tóm tắt một cách vô cùng đơn giản mà không cần tính lại từ đầu (mặc dù việc tính toán không hề khó, nếu sử dụng Microsoft Excel).

Ví dụ 4.5 – Giá trị dữ liệu tăng/giảm một lượng cố định

Bảng sau cho thấy mẫu là danh sách nhà gỗ có 2 phòng ngủ ở khu vực Bakewell được Công ty Letting Agent cho thuê trong kỳ nghỉ, và giá thuê hàng tuần đến trước 30 tháng 6.

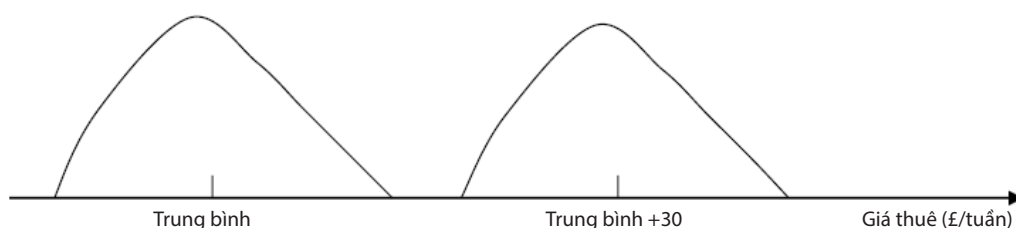
Tên nhà	Giá thuê (£/tuần)	Tên nhà	Giá thuê (£/tuần)
Rose	420	Woodside	440
Honeysuckle	475	1 Farm Road	450
Briar	500	3 Chimneys	450
Riverside	450	Bridge	500
3 Top Lane	550	5 Town Mews	450
Bluebell	575	May	600

Bảng 4.6 Dữ liệu nhà gỗ.

Hãy tự tính toán để xác nhận: giá thuê yếu vị là 450£/tuần; giá thuê trung vị là 463£/tuần và giá thuê trung bình là 488£/tuần (hai giá trị cuối được làm tròn đến 1£).

Giả sử từ ngày 01 tháng 7, toàn bộ giá thuê tăng 30£ mỗi tuần. Điều này sẽ ảnh hưởng thế nào đến ba độ đo vị trí?

Do tất cả giá trị dữ liệu đều thay đổi như nhau, ta không cần tính lại bất cứ số đo nào từ đầu; mà chỉ cần thêm phần tăng 30£ vào mỗi số đo. Điều duy nhất xảy ra là đường phân phối giá thuê đã chuyển dịch sang bên phải thêm 30£ trên trục giá thuê.



**Hình 4.6** Biểu diễn giá trị trung bình sau khi tăng 30£.

Do đó, giá thuê yếu vị mới là 480£/tuần; giá thuê trung vị mới là 493£/tuần và giá thuê trung bình mới là 518£/tuần (*hai giá trị cuối được làm tròn đến 1£ gần nhất*).

Giả sử giờ có đề xuất tăng giá thuê 10% cho thời điểm trước 30 tháng 6 năm 2016. Ba độ đo vị trí sẽ bị ảnh hưởng như thế nào?

Ta lập luận tương tự như trường hợp tăng giá cố định – mỗi số đo tóm tắt sẽ tăng 10% do tất cả giá trị dữ liệu đều tăng như vậy.

Do đó, giá thuê yếu vị mới là 495£/tuần; giá thuê trung vị mới là 509£/tuần và giá thuê trung bình mới là 537£/tuần (*hai giá trị cuối đã được làm tròn đến 1£; giá trị tăng đều từ số liệu ban đầu trong bảng trên*).

Nguyên tắc tương tự áp dụng cho trường hợp giảm giá trị hoặc phần trăm cố định. Các số đo tóm tắt giảm tương ứng với mức giảm giá trị dữ liệu.

### 4.3.6 SỐ ĐO TÍNH DAO ĐỘNG TRONG GIÁ TRỊ DỮ LIỆU (ĐỘ PHÂN TÁN)

#### 4.3.6.1 Giới thiệu

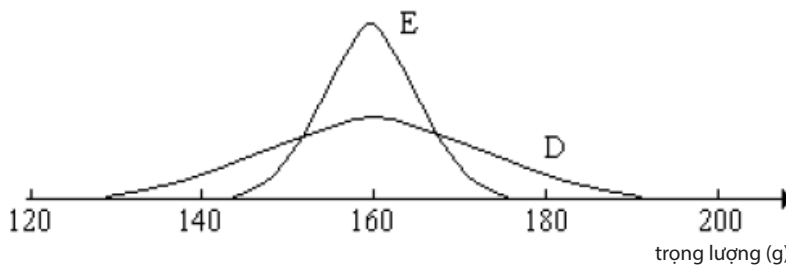
Khi tìm giá trị tóm tắt để đại diện hoặc mô tả một tập dữ liệu, việc chỉ tìm ra số đo trung bình “quy chuẩn” là không đủ, ta cũng phải cân nhắc đến mức độ dàn trải (độ phân tán) trong các giá trị dữ liệu.

Ví dụ, xét hai phân phối trong **Hình 4.7**. Các phân phối này có thể đến từ dữ liệu của hai máy đóng túi trà khác nhau.

Ta thấy có thể sử dụng cùng một độ đo vị trí để mô tả hai phân phối tương đối khác nhau.

Với phân phối E, các độ đo vị trí (trung bình cộng, trung vị hoặc yếu vị) có ý nghĩa tượng trưng cho toàn bộ phân phối. Tất cả các giá trị có tần số xảy ra cao đều tập trung trong một khoảng hẹp xoay quanh độ đo vị trí; như vậy máy này đang hoạt động ổn định.

Với phân phối D, số đo vị trí có ít ý nghĩa quy chuẩn cho toàn bộ phân phối hơn – do các giá trị đơn lẻ dàn trải hơn ở phân phối E. Trọng lượng trà được đóng gói trong mỗi túi dao động nhiều hơn – điều này không tốt cho công ty.



**Hình 4.7** Hai phân phối có chung giá trị trung bình, nhưng khác nhau về độ phân tán.

#### Ví dụ 4.6 – Minh họa nhu cầu cần đến độ phân tán

Ta xét hai phòng, trong đó mỗi phòng có 5 người. Trong phòng xanh, tuổi của 5 người là 54, 35, 24, 10 và 2 tuổi. Tuổi của 5 người trong phòng đỏ là 30, 27, 24, 23 và 21.

Độ tuổi trung bình ở mỗi phòng giống nhau (25 tuổi) và tuổi trung vị cũng vậy (24 tuổi). Tuy nhiên, mức độ dao động tuổi ở hai phòng này khác nhau.

Khi đó, ta cần đến **độ phân tán (measure of spread hoặc measure of dispersion)** bên cạnh độ đo vị trí. Kết hợp hai số đo này, chúng ta có thể phân biệt được hai phân phối kiểu như D và E trong ví dụ trên.

Độ phân tán là số đo thể hiện tính dàn trải, phân tán, dao động hoặc khó đoán biết của dãy giá trị dữ liệu tương ứng với số đo vị trí.

Vậy làm thế nào để tìm được số đo độ phân tán thích hợp cho giá trị trung bình hoặc trung vị của giá trị dữ liệu? (Từ đây ta không xét đến yếu vị nữa). Số đo độ phân tán được sử dụng phải thể hiện rõ tính đáng tin cậy của độ đo vị trí, cũng như đại diện cho toàn bộ phân phối.

#### 4.3.6.2 Khoảng biến thiên là độ phân tán

Để đánh giá mức độ dàn trải hoặc phân tán của tập dữ liệu, thì sử dụng khoảng biến thiên các giá trị là một cách vô cùng đơn giản. Theo định nghĩa, cách tính khoảng biến thiên của một tập giá trị là:

Khoảng biến thiên = Giá trị lớn nhất – Giá trị nhỏ nhất

Xét **Ví dụ 4.5**, khoảng biến thiên độ tuổi trong phòng xanh là  $(54 - 2) = 52$  tuổi, còn khoảng biến thiên trong phòng đỏ là  $(30 - 21) = 9$  tuổi.

Do đó, ta có thể nói trong phòng xanh độ tuổi trung bình là 25 tuổi với khoảng biến thiên 52 tuổi, còn trong phòng đỏ độ tuổi trung bình là 25 tuổi với khoảng biến thiên là 9 tuổi.

Các giá trị này biểu thị rõ ràng sự khác biệt của hai tập dữ liệu.

Tuy nhiên, việc sử dụng khoảng biến thiên làm độ đo phân tán có thể nảy sinh vấn đề trong trường hợp tập dữ liệu bao gồm các giá trị ngoại lệ. Ví dụ, giả sử tuổi của mọi người trong phòng đỏ là 30, 27, 24, 23 và 2 tuổi. Khi đó tuổi trung bình là 21 tuổi với khoảng biến thiên 29 tuổi.

Nếu được đưa ra không kèm với dữ liệu gốc, thì thông tin này có thể khiến người đọc hiểu sai về độ phân tán của các giá trị xoay quanh độ tuổi trung bình.

Không nên bỏ qua các giá trị ngoại lệ mà không suy tính kỹ; nhưng đồng thời, ta cũng không nên để các giá trị này bóp méo giá trị độ đo phân tán theo hướng bất hợp lý.

Tóm lại, việc sử dụng khoảng biến thiên như đơn vị đo độ phân tán của dữ liệu là tiện lợi, giúp ta dễ tính toán và dễ hiểu. Tuy nhiên, khoảng biến thiên có những hạn chế do chỉ sử dụng hai giá trị duy nhất từ tập dữ liệu và dễ dàng bị bóp méo do ảnh hưởng của các giá trị ngoại lệ.

#### 4.3.6.3 *Tứ phân vị là độ phân tán*

Vấn đề khoảng biến thiên bị bóp méo do giá trị ngoại lệ có thể khắc phục bằng cách sử dụng tứ phân vị để đo độ phân tán của dữ liệu. Chênh lệch giữa tứ phân vị  $Q_1$  và tứ phân vị  $Q_3$  cho ta biết khoảng biến thiên của nửa giữa các giá trị của dữ liệu. Ý tưởng này đã được sử dụng một lần trong phần trước, khi chúng ta thảo luận về biểu đồ hộp.

Có thể sử dụng **độ lệch tứ phân vị (quartile deviation, viết tắt: QD)** kết hợp tứ phân vị  $Q_1$  và tứ phân vị  $Q_3$  trong một số đo đơn lẻ, để mô tả độ phân tán của các giá trị dữ liệu quanh trung vị. Độ lệch tứ phân vị được tính như sau:

$$QD = \frac{Q_3 - Q_1}{2}$$

Việc mô tả độ phân tán của phần giữa tập dữ liệu bằng độ lệch tứ phân vị thường hữu ích, tuy nhiên phương pháp này không phù hợp để phân tích thống kê sâu hơn, mà chỉ phục vụ mục đích mô tả.

#### **Ví dụ 4.7 – Tính toán và sử dụng độ lệch tứ phân vị**

Ở chương trước, **Ví dụ 4.3**, chúng ta đã xét đến % độ ẩm của các lô bột giặt, lần lượt do công nhân ca ngày và ca đêm sản xuất. Dữ liệu này được nhắc lại trong bảng dưới đây.

Lô	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
% độ ẩm	7,2	6,5	5,3	3,9	6,7	5,9	6,0	6,5	6,9	4,3	6,3	7,6	5,9	5,7	5,5

**Bảng 4.7** Ca ngày (15 lô).

Lô	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
% độ ẩm	5,9	5,6	3,3	4,4	6,9	6,1	5,6	2,9	6,4	8,8	8,1	7,1	7,1	4,8	4,0	2,5	6,4	6,5

**Bảng 4.8** Ca đêm (18 lô).

Xác định trung vị, độ lệch tứ phân vị của % độ ẩm các lô bột giặt ca ngày và ca đêm. So sánh hai kiểu phân phối thu được.

**Lời giải:**

Sắp xếp các giá trị dữ liệu ca ngày theo thứ tự kích cỡ, ta có:

3,9 4,3 5,3 5,5 5,7 5,9 5,9 6,0 6,3 6,5 6,5 6,7 6,9 7,2 7,6

Có tất cả 15 giá trị, do đó trung vị là giá trị thứ 8 ( $= (n+1) / 2$ ). Trung vị ở đây là 6,0%.

Tứ phân vị  $Q_1$  là giá trị thứ  $(n+1) / 4 = 4$ , do đó tứ phân vị  $Q_1$  là 5,5%.

Có một số quan điểm khác nhau về vị trí chính xác của các tứ phân vị. Điều này là chấp nhận được, chỉ cần bạn sử dụng một cách thống nhất.

Tương tự vậy, ta lấy tứ phân vị  $Q_3$  là giá trị thứ  $3 \times (n+1) / 4 = 12$ , vậy  $Q_3$  là 6,7%.

Độ lệch tứ phân vị bằng:

$$QD = \frac{Q_3 - Q_1}{2} = \frac{6,7 - 5,5}{2} = 0,6\%$$

Đối với ca ngày, trung vị % độ ẩm là 6,0% với độ lệch tứ phân vị là 0,6%.

Ta sắp xếp các giá trị dữ liệu ca đêm theo thứ tự kích cỡ như sau:

2,5 2,9 3,3 4,0 4,4 4,8 5,6 5,6 5,9 6,1 6,4 6,4 6,5 6,9 7,1 7,1 8,1 8,8

Có tất cả 18 giá trị. Trung vị là giá trị trung bình của giá trị thứ 9 và thứ 10. Như vậy trung vị bằng:

$$\frac{5,9 + 6,1}{2} = 0,6\%$$

Tương tự như với dữ liệu ca ngày, ta lấy tứ phân vị  $Q_1$  là giá trị thứ 4,75; khoảng 4,4%.

Tứ phân vị  $Q_3$  là khoảng 6,9%. Độ lệch tứ phân vị là:

$$QD = \frac{Q_3 - Q_1}{2} = \frac{6,9 - 4,4}{2} = 1,25\%$$

Như vậy với ca đêm, trung vị % độ ẩm là 6,0% với độ lệch tứ phân vị bằng 1,25%.

Cả hai ca đều có trung vị % độ ẩm giống nhau, nhưng giá trị của các lô ca đêm phân tán rộng hơn các lô ca ngày – điều này thể hiện sự thiếu nhất quán hơn trong chất lượng làm việc buổi đêm.

#### 4.3.6.4 Độ lệch chuẩn là độ phân tán

Giờ ta cần một số đo độ phân tán có thể sử dụng với giá trị trung bình cộng và thích hợp để phân tích thống kê sâu hơn. Nói cách khác, ta phải thiết lập một độ phân tán mà sử dụng tất cả mọi giá trị dữ liệu. Độ phân tán thỏa mãn các điều kiện này gọi là độ lệch chuẩn (standard deviation).

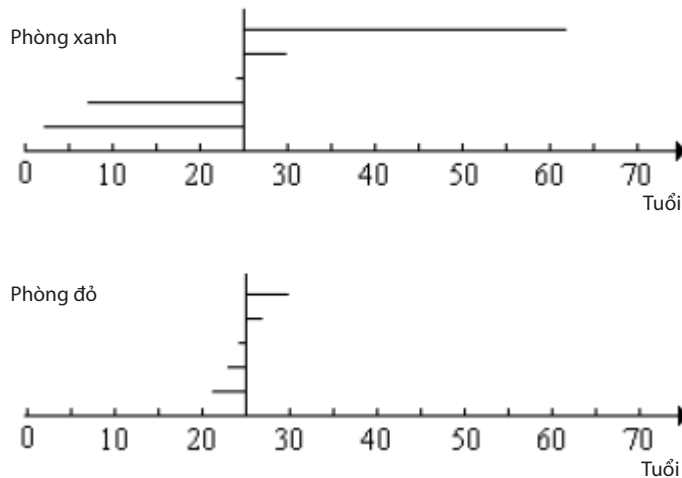
Xét lại **Ví dụ 4.5**. Ví dụ này nhắc đến độ tuổi của hai nhóm gồm 5 người: Trong phòng xanh, tuổi của 5 người là 54, 35, 24, 10 và 2. Trong phòng đỏ, tuổi của nhóm là 30, 27, 24, 23 và 21.

Ta đã xác định độ tuổi trung bình ở cả hai phòng đều là 25 tuổi. Giờ ta muốn xét độ phân tán của các giá trị trong mỗi nhóm. Hãy nhìn vào hiệu số chênh lệch giữa từng giá trị trong tập dữ liệu với giá trị trung bình. Hiệu số này được gọi là **độ lệch (deviation)** so với giá trị trung bình. Xem **Bảng 4.9**.

Phòng xanh		Phòng đỏ	
Tuổi	Tuổi – Tuổi trung bình	Tuổi	Tuổi – Tuổi trung bình
54	29	30	5
35	10	27	2
24	-1	24	-1
10	-15	23	-2
2	-23	21	-4
<b>Tổng</b>	<b>0</b>	<b>Tổng</b>	<b>0</b>

**Bảng 4.9** Độ lệch so với giá trị trung bình.

Độ lệch này được minh họa như trong **Hình 4.8**:



**Hình 4.8** Độ lệch tuổi của từng người so với tuổi trung bình.

Kiểu hình của độ lệch các giá trị so với giá trị trung bình là một chỉ số tốt để đánh giá độ phân tán trong phạm vi dữ liệu. Trong trường hợp lý tưởng, ta sẽ tính trung bình cộng của các độ lệch này và coi đó là độ phân tán dữ liệu. Tuy nhiên, như có thể thấy từ số liệu trong **Bảng 1**, các độ lệch so với giá trị trung bình ở cả hai phòng đều có tổng bằng không. Tình huống này sẽ luôn xảy ra, do giá trị trung bình cộng thể hiện giá trị của “các phần bằng nhau”, như trong ví dụ ở trên, tổng số tuổi được chia đều cho 5 người trong phòng, có nghĩa mỗi người 25 tuổi. Trong mỗi trường hợp, tổng độ lệch số âm sẽ cân bằng tương ứng với tổng độ lệch số dương.

Có một số cách để giải quyết vấn đề này. Ta có thể đơn giản là bỏ qua ký hiệu số âm và chỉ xét đến kích cỡ độ chênh lệch. Trung bình cộng của kích cỡ các độ lệch có thể dùng làm độ phân tán của tập dữ liệu. Độ phân tán này được gọi là **độ lệch trung bình (mean deviation)**.

Tuy nhiên, việc đơn thuần bỏ qua ký hiệu số âm là cách làm không hoàn toàn thỏa đáng, xét theo phương diện toán học. Vì thế, chúng ta sử dụng cách thay thế là bình phương các độ lệch. Khi đó, ta tính trung bình cộng của bình phương các độ lệch, gọi là **phương sai (variance)**. Ta sẽ sử dụng dữ liệu từ **Ví dụ 4.5** (Phòng đỏ và xanh) để minh họa cho quá trình này. Độ lệch tuổi mỗi người so với tuổi trung bình được ký hiệu là  $d$ .

Xem **Bảng 4.10**.



Phòng xanh			Phòng đỏ		
Tuổi	Tuổi – Tuổi trung bình (d)	d <sup>2</sup>	Tuổi	Tuổi – Tuổi trung bình (d)	d <sup>2</sup>
62	37	1369	30	5	25
30	5	25	27	2	4
24	-1	1	24	-1	1
7	-18	324	23	-2	4
2	-23	529	21	-4	16
<b>Tổng</b>	<b>0</b>	<b>2248</b>	<b>Tổng</b>	<b>0</b>	<b>50</b>

**Bảng 4.10** Tính phương sai.

Trung bình cộng của bình phương các độ lệch, hay phương sai = tổng tất cả các giá trị

$$d^2 / n = \frac{\sum d^2}{n}$$

$$\text{Phòng xanh: Phương sai} = \frac{2248}{5} = 449,6 \text{ (tuổi)}^2$$

$$\text{Phòng đỏ: Phương sai} = \frac{50}{5} = 10 \text{ (tuổi)}^2$$

Bản thân phương sai đã là một độ phân tán hữu dụng, nhưng trong ví dụ này ta sẽ lấy căn bậc 2 của phương sai. Số đo thu được gọi là **độ lệch chuẩn (standard deviation)**. So với phương sai, số đo này có lợi thế bởi có cùng đơn vị tính với các giá trị trong tập dữ liệu.

$$\text{Độ lệch chuẩn} = \sqrt{\text{phương sai}}$$

$$\text{Phòng xanh: Độ lệch chuẩn} = \sqrt{449,6} = 21,2 \text{ (tuổi)}$$

$$\text{Phòng đỏ: Độ lệch chuẩn} = \sqrt{10} = 3,2 \text{ (tuổi)}$$

Vì cả hai tập dữ liệu có cùng giá trị trung bình 25 tuổi, nên ta có thể nói ngay rằng độ tuổi trong phòng xanh có độ phân tán quanh giá trị trung bình cao hơn so với trong phòng đỏ.

Điểm quan trọng cần nhớ là độ lệch chuẩn có liên quan đến độ lệch trung bình tương ứng với giá trị trung bình của toàn bộ giá trị dữ liệu trong tập.

#### 4.3.7 SỬ DỤNG CÔNG CỤ MICROSOFT EXCEL TƯƠNG ỨNG

Microsoft Excel có chức năng tích hợp sẵn để tính độ lệch chuẩn và ta có thể dùng một công thức đơn giản để tính độ lệch tứ phân vị.

=STDEVP()    Tính độ lệch chuẩn.

Trong Excel, có nhiều tính năng tính độ lệch chuẩn với các ứng dụng cho vô số loại hình phân tích thống kê khác nhau. Với mục đích sử dụng cho môn Phân tích Kinh doanh, chúng ta luôn dùng hàm =STDEVP().

Với độ đo vị trí, chức năng này phải được sử dụng với tập dữ liệu thô. Sử dụng một phần dữ liệu ở **Ví dụ 4.3** như dưới đây, ta minh họa việc sử dụng tính năng này như sau.

#### Ví dụ 4.8 – Sử dụng các chức năng Microsoft Excel

	A	B	C	Dữ liệu cho thấy điểm 25 sinh viên đạt được trong bài kiểm tra tự đánh giá (SAT).
1	Sinh viên	Điểm SAT/10		
2	Adey	6		
3	Adrian	3		
4	Ashling	5		
...				Ô B33:B34 chứa hàm cần nhập.  Ô C33:C34 chứa các giá trị sẽ xuất hiện ở ô B33 và B34
24	Simon	5		
25	Susan	6		
26	Thomas	8		
27				
28	yếu vị	=mode(B2:B26)	5	
29	trung vị	=median(B2:B26)	6	
30	trung bình cộng	=average(B2:B26)	6.08	
31	tứ phân vị $Q_1$	=quartile(B2:B26,1)	5	
32	tứ phân vị $Q_3$	=quartile(B2:B26,3)	7	
33	độ lệch chuẩn	=STDEVP(B2:B26)	1.62	
34	độ lệch tứ phân vị	=(B32-B31)/2	1.0	

**Bảng 4.11** Ví dụ sử dụng Microsoft Excel để tính phương sai.

### 4.3.8 DIỄN DỊCH ĐỘ LỆCH CHUẨN

Giờ ta phải xem xét thông tin mà giá trị của độ lệch chuẩn mang đến.

#### Ví dụ 4.9 – Diễn dịch độ lệch chuẩn

Xem lại **Ví dụ 4.1**. Số lượt vắng mặt mỗi ngày ở Công ty Shefftex là:

4, 2, 4, 3, 5, 9, 5, 6, 6, 3, 3, 6, 4, 3, 3, 2, 3, 5, 4, 3, 3, 2, 4, 3, 5

Số lượt vắng mặt trung bình mỗi ngày là 4,0. Dùng hàm chức năng STDEVP() của Microsoft Excel, ta tìm được độ lệch chuẩn là 1,6 lượt vắng mặt mỗi ngày.

Do đó, tập dữ liệu với 25 thông tin thống kê vắng mặt có thể được đại diện bằng giá trị trung bình 4,0 lượt vắng mặt mỗi ngày. Độ phân tán các giá trị dữ liệu đơn lẻ xoay quanh giá trị trung bình này là 1,6 lượt mỗi ngày. Điều này thể hiện độ phân tán rộng hay hẹp?

Rất khó để trả lời một cách chắc chắn, vì nó còn phụ thuộc vào nhiều yếu tố. Tất cả mọi thứ ta có thể làm là phát triển, thông qua thực tiễn và kinh nghiệm, “cảm quan” về ý nghĩa các giá trị độ lệch chuẩn cụ thể.

Ở ví dụ này, 1,6 lượt vắng mặt mỗi ngày tương đương xấp xỉ 39% giá trị trung bình (4,0 lượt vắng mặt mỗi ngày). Điều đó thể hiện độ dàn trải tương đối lớn so với loại dữ liệu này. Ở đây giá trị trung bình có vẻ không phải là giá trị tiêu biểu cho toàn bộ tập dữ liệu. Tuy nhiên, cần lưu ý là tập có chứa một giá trị “ngoại lệ” – 9 lượt vắng mặt mỗi ngày. Số liệu này làm tăng giá trị độ phân tán của dữ liệu tương ứng với số trung bình – ngoài ra cũng làm tăng giá trị trung bình.

Do giá trị độ lệch chuẩn chịu ảnh hưởng bởi sự có mặt của các giá trị ngoại lệ, nên ta luôn phải kiểm tra để đảm bảo không phát sinh giá trị độ lệch chuẩn lớn do một hoặc hai giá trị ngoại lệ, và các giá trị còn lại của tập dữ liệu tập trung sát giá trị trung bình.

Khi cần so sánh hai hoặc nhiều phân phối, chúng ta càng dễ sử dụng độ lệch chuẩn hơn.

Giả sử dữ liệu ở **Ví dụ 4.9** là của tháng 8 và dữ liệu tóm tắt tương ứng của tháng 9 như sau:

số lượt vắng mặt trung bình mỗi ngày là 6,2

độ lệch chuẩn số lượt vắng mặt mỗi ngày là 0,8

Các số liệu này được tính từ thông tin vắng mặt tương đương cho các ngày làm việc trong tháng 9.

Ta có thể thấy ngay là số lượt vắng mặt trung bình đã tăng, tuy nhiên độ lệch chuẩn tháng 9 thấp hơn độ lệch chuẩn tháng 8, thể hiện rằng các lượt vắng mặt trong tháng 9

tập trung sát giá trị trung bình – nghĩa là không có giá trị ngoại lệ. Như vậy, các giá trị trong tập dữ liệu tháng 9 ổn định hơn so với tháng 8. (Việc xem xét dữ liệu trước khi đi đến kết luận cuối cùng sẽ rất có ích.)

#### 4.3.9 ĐỘ PHÂN TÁN TƯƠNG ĐỐI

Khi so sánh các giá trị trung bình và độ lệch chuẩn của hai hoặc nhiều phân phối, việc diễn đạt độ lệch chuẩn dưới dạng phần trăm của giá trị trung bình là một cách thức hữu hiệu. Số đo này được gọi là độ phân tán tương đối (**relative dispersion**) hoặc **hệ số biến thiên (coefficient of variation)**.

$$\text{Hệ số biến thiên} = \frac{s}{\bar{x}} \times 100$$

Sử dụng dữ liệu từ **Ví dụ 4.9**, ta có:

$$\text{Lượt vắng mặt tháng 8:} \quad \text{Độ phân tán tương đối} = \frac{1,6}{4,0} \times 100\% = 39,4\%$$

$$\text{Lượt vắng mặt tháng 9:} \quad \text{Độ phân tán tương đối} = \frac{0,8}{6,2} \times 100\% = 12,9\%$$

Điều này một lần nữa khẳng định rằng dữ liệu của tháng 9 có độ phân tán hẹp hơn quanh giá trị trung bình.

Việc sử dụng hệ số biến thiên là quan trọng khi so sánh các tập dữ liệu có giá trị biến khá khác biệt về kích cỡ.

#### 4.3.10 ẢNH HƯỞNG CỦA SỰ THAY ĐỔI TOÀN BỘ GIÁ TRỊ DỮ LIỆU ĐẾN ĐỘ LỆCH CHUẨN

Cũng như với các độ đo vị trí (ở Phần 4.3.5), nếu tất cả giá trị dữ liệu đều được thay đổi theo cách giống nhau, ta rất dễ điều chỉnh giá trị của độ lệch chuẩn mà không cần tính lại từ đầu (mặc dù việc tính lại là không khó, nếu sử dụng Microsoft Excel).

Xem lại **Ví dụ 4.5**. Hãy tự tính toán để kiểm chứng rằng độ lệch chuẩn của giá cho thuê hiện tại là 55,60£ mỗi tuần (*làm tròn đến chữ số thập phân hàng chục*).

Ta sẽ tìm độ lệch chuẩn trong các trường hợp:

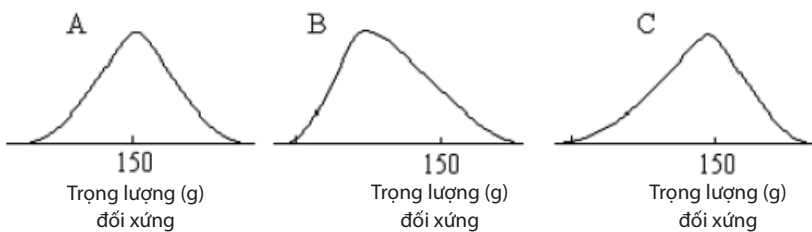
- Tất cả giá thuê tăng 30£ mỗi tuần từ ngày 01 tháng 7;
- Tất cả giá thuê tăng 10% cho năm 2008.
- Giá thuê tăng đều 30£ mỗi tuần sẽ làm tăng giá thuê trung bình thêm 30£, nhưng độ phân tán dữ liệu tương ứng với giá trị trung bình không đổi, do đó độ lệch chuẩn vẫn là 55,60£ mỗi tuần.

- d) Giá thuê tăng 10% khiến giá thuê trung bình tăng 10%, và lần này độ phân tán dữ liệu tương ứng với giá thuê trung bình có thay đổi. Có thể thấy độ lệch chuẩn cũng tăng 10%, thành 61,20£ mỗi tuần (làm tròn đến chữ số thập phân hàng chục).

#### 4.3.11 TÍNH ĐỐI XỨNG

Khía cạnh thứ ba của một tập dữ liệu mà chúng ta cần cân nhắc là tính đối xứng hoặc thiếu đối xứng, (còn được gọi là **tính bất đối xứng - skewness**) trong phân phối các giá trị. Ta sẽ không cố gắng tìm ra một phương pháp đo lường tính đối xứng của một phân phối, thay vào đó ta sẽ chỉ mô tả bằng lời.

Hãy xem ba phân phối được minh họa ở **Hình 4.9**.



**Hình 4.9** Phân phối bất đối xứng.

Phân phối A cho thấy sự đối xứng. Trong trường hợp này, trung bình cộng = trung vị = yếu vị. Đây là kiểu phân phối thường thấy khi bạn thu thập dữ liệu kiểu như:

- trọng lượng các gói của một loại bánh quy nhất định
- chiều dài các ống thép được cắt bởi một máy cắt
- chỉ số thông minh (IQ) của một mẫu người ngẫu nhiên

Phân phối B có đuôi lệch về bên phải – các giá trị cao hơn. Người ta nói phân phối **B lệch dương (positively skewed)**. Trong trường hợp này, trung bình cộng > trung vị > yếu vị. Đây là kiểu phân phối thường gặp khi bạn thu thập dữ liệu kiểu như:

- lương trả cho công nhân sản xuất
- độ tuổi của sinh viên học khóa văn bằng 1
- thời gian phục vụ trong gara ô tô

Phân phối C có đuôi lệch về bên trái – các giá trị thấp hơn. Người ta nói phân phối **C lệch âm (negatively skewed)**. Trong trường hợp này, trung bình cộng < trung vị < yếu vị. Đây là kiểu phân phối thường gặp khi bạn thu thập dữ liệu kiểu như:

- độ tuổi bắt đầu cần đeo kính
- thời gian chờ xếp hàng trong một số tình huống
- tuổi đời của một số loại thiết bị điện

#### 4.4 TÀI LIỆU THAM KHẢO CHƯƠNG

1. Morris, C, Quantitative Approaches in Business Studies (*tạm dịch*: Phương pháp định lượng trong nghiên cứu kinh doanh), tái bản lần thứ 8, FT/Prentice Hall, 2012 (eBook, tái bản lần thứ 7, 2008), Chương 6 (một phần).
2. Oakshott, L, Essential Quantitative Methods for Business (*tạm dịch*: Phương pháp nghiên cứu định lượng thiết yếu trong kinh doanh, quản lý & tài chính), tái bản lần thứ 5, Palgrave Macmillan, 2012, Chương 6.
3. Waters, D, Quantitative Methods for Business (*tạm dịch*: Phương pháp nghiên cứu định lượng trong kinh doanh), tái bản lần thứ 5, FT/Prentice Hall, 2011, (có bản eBook), Chương 5 và 6.
4. Wisniewski, M, Quantitative Methods for Decision Makers (*tạm dịch*: Phương pháp nghiên cứu định lượng phục vụ ra quyết định), tái bản lần thứ 5, FT/Prentice Hall, 2009 (có bản eBook), Chương 2 (một phần) và 3 (một phần).

## Chương 5

# ĐO LƯỜNG VÀ ĐÁNH GIÁ SỰ BIẾN ĐỘNG TRONG GIÁ TRỊ DỮ LIỆU

### 5.1 GIỚI THIỆU

Đương đầu với sự biến động giá trị trong các biến số chính như giá cả, chi phí, nhu cầu... là một trong những hoạt động chính trong doanh nghiệp. Biến đổi này phát sinh cả trong và ngoài phạm vi doanh nghiệp. Thay đổi giá cả các loại hàng hóa nhất định có thể ảnh hưởng nhiều đến nhu cầu sản phẩm trên thị trường; tình hình lạm phát hoặc kiểm soát tỷ giá có thể ảnh hưởng đến chi phí mua nguyên vật liệu thô...

Ngay kích cỡ và phương hướng biến động giá trị trong các biến số chính cũng có thể tự biến động, mang theo mức độ quan trọng tương ứng khác nhau. Ví dụ, việc tăng giá 10 xu mỗi kg nguyên liệu thô có thể không quan trọng, nếu giá tăng từ 20£ lên 20,1£ (1£ bằng 100 xu); nhưng tình hình sẽ nghiêm trọng hơn, nếu tăng giá từ 10 xu lên 20 xu (tăng gấp đôi). Người quản lý sẽ không chỉ phải điều tiết các thay đổi này khi lập kế hoạch và dự trù ngân sách, mà còn phải **đánh giá (evaluate)** mức độ quan trọng tương ứng với biến đổi.

Để đánh giá ảnh hưởng của sự thay đổi trong các giá trị biến, thông thường ta chỉ cần tính phần trăm là đủ. Ví dụ, nếu tăng từ 20£ lên 20,1£ thì đây là mức tăng 0,5%; nhưng nếu tăng từ 10 xu lên 20 xu thì sẽ là mức tăng 100%.

Theo dõi giá của một sản phẩm hàng tháng, ta có thể tính được phần trăm thay đổi từ tháng này đến tháng khác và hình thành cái nhìn về tình hình biến động giá cả.

Ngoài ra, ta có thể so sánh giá một tháng với giá của một số tháng cố định nào đó. Giá tham khảo được chọn là cơ sở so sánh giá các tháng khác.

Trên thực tế, ta không tính rõ ràng phần trăm **thay đổi (change)** trong giá tương ứng với giá cố định, mà chỉ đơn giản là tính giá trị giá như là một phần trăm nào đó của giá gốc. Số đo này được gọi là **chỉ số (index number)**. Chỉ số của giá tham khảo luôn là 100 và ta tính tất cả các giá khác tương ứng với giá trị đó. Quy trình này giúp việc tính toán trở nên dễ dàng hơn.

Ví dụ, nếu giá một sản phẩm ở tháng tham khảo là 5,65£ và tháng hiện tại là 6,45£, thì phần trăm thay đổi giá sẽ là:

$$\frac{6,45 - 5,65}{5,65} \times 100 = 14,2\%$$

trong khi chỉ số sẽ là:

$$\frac{6,45}{5,65} \times 100 = 114,2 \text{ nghĩa là vượt } 14,2\% \text{ so với mốc tham khảo } 100$$

Cả hai phương pháp đều cho thông tin giống nhau, nhưng phép tính chỉ số được “tổ chức hợp lý” hơn.

Chỉ số cho phép ta kiểm soát hướng thay đổi của giá trị biến bất kỳ qua thời gian hoặc không gian, đồng thời đánh giá kích thước cũng như tầm quan trọng tương ứng của sự thay đổi. Bằng cách này, có thể so sánh giữa các giá trị phát sinh ở nhiều thời điểm khác nhau. Ta có thể sử dụng bất cứ biến hoặc khoảng thời gian nào.

Điểm tham khảo được gọi là **giá trị gốc (the base)** và chỉ số của biến số tại gốc luôn là **100**.

Có nhiều cách để tính chỉ số, tùy thuộc vào đối tượng mà ta cần đánh giá:

- một sản phẩm đơn lẻ hay nhóm nhiều sản phẩm
- giá cả hay số lượng...
- giá trị đơn lẻ hay giá trị có trọng số

Tiếp theo ta sẽ xét đến một vài chỉ số thường được sử dụng nhất, bắt đầu bằng cách sử dụng phần trăm để đánh giá sự thay đổi.

## 5.2 CHÈNH LỆCH PHẦN TRĂM

Đôi khi sẽ rất khó để diễn dịch và đánh giá thông tin được đưa ra, dựa vào giá trị tuyệt đối. Ví dụ:

- Tổng doanh thu cả năm từ chi nhánh John Lewis ở Birmingham là 3.160.000£ giai đoạn 2014/2015. Liệu doanh thu như vậy là tốt?
- Giá của một ổ bánh mì lớn và 500g bí-tết đều tăng 10 xu so với tháng trước. Như vậy là tăng nhiều hay ít?

Trong ví dụ đầu tiên, doanh thu 3,2 triệu £ mỗi năm nghe có vẻ nhiều, nhưng cũng không biết chắc được – để đánh giá (evaluate), ta phải đặt số liệu này vào bối cảnh cụ thể.

Nếu giờ thông tin ta biết là doanh thu nhà hàng ở Birmingham chiếm 14% tổng doanh thu tất cả các nhà hàng của John Lewis; lúc này ta sẽ có cái nhìn rõ ràng hơn – số liệu này đạt trên mức trung bình.



Thêm vào đó, ta thấy rằng doanh số hàng năm của nhà hàng ở Birmingham đã tăng 11% so với năm ngoái, tăng 10% so với giai đoạn 2012/2013 – như vậy đã có mức tăng nhẹ trong tốc độ tăng trưởng.

Các số liệu phần trăm này giúp ta diễn giải được giá trị thực tế của doanh thu giai đoạn 2014/2015. Ta có thể đánh giá doanh thu theo thời gian và so sánh với các số liệu tương đương.

Ở ví dụ thứ hai, 10 xu tương đương với mức tăng 7,5% trên giá ổ bánh mỳ, nhưng chỉ là 1% trên giá nửa kg bít-tết. Như vậy, việc tính số liệu phần trăm lại giúp chúng ta đánh giá được giá trị tuyệt đối.

Quay lại ví dụ đầu tiên về nhà hàng John Lewis ở Birmingham, ta đã kết luận rằng doanh thu hàng năm đã tăng 10% từ giai đoạn 2012/2013 đến 2013/2014 và tăng tiếp 11% từ 2013/2014 đến 2014/2015.

Các số liệu phần trăm cung cấp thông tin hữu ích, nhưng đôi khi ta nên sử dụng một điểm cố định và đối chiếu mọi thay đổi với điểm cố định đó. Vì thế trong ví dụ này, có thể nói chênh lệch phần trăm từ giai đoạn 2012/2013 đến 2013/2014 là 10% và chênh lệch phần trăm từ 2012/2013 đến 2014/2015 là 22%.

Cả hai số liệu phần trăm đều liên quan đến giai đoạn đầu tiên là năm 2012/2013. Trong trường hợp này, người ta thường sử dụng chỉ số hơn là phần trăm, mặc dù cả hai loại đều khá giống nhau.

### 5.2.1. CHỈ SỐ ĐỐI VỚI SẢN PHẨM ĐƠN LẺ

Thuật ngữ **giá so sánh (price relative)** là cách gọi thay thế cho chỉ số đơn lẻ. Đó là giá của sản phẩm tại một thời điểm (như một năm nào đó) được thể hiện ở dạng phần trăm của giá sản phẩm đó tại thời điểm gốc.

$$\text{Giá so sánh} = \frac{\text{giá tại một thời điểm, } P_n}{\text{giá tại thời điểm gốc, } P_0} \times 100$$

#### Ví dụ 5.1 – Đánh giá sự thay đổi trong giá trị của sản phẩm đơn lẻ

**Bảng 5.1** cho thấy giá trị nhập khẩu một loại hàng hóa nhất định vào Anh trong các năm 2011 – 2015. Xác định chỉ số mô tả sự thay đổi giá trị nhập khẩu hàng hóa trong giai đoạn này.

Năm	Nhập khẩu (tỷ £)
2011	10,89
2012	11,41
2013	12,73
2014	14,46
2015	14,50

**Bảng 5.1** Nhập khẩu hàng hóa hàng năm.

**Lời giải:** Một cách để đo lường sự thay đổi trong giá trị nhập khẩu là sử dụng phần trăm.

**Bảng 5.2** cho thấy sự thay đổi đối với giá trị hàng hóa nhập khẩu tương ứng với năm 2011, dưới dạng số và dạng phần trăm.

Năm	Nhập khẩu (tỷ £)	Thay đổi so với năm 2011 (tỷ £)	Thay đổi tương ứng với năm 2011 (%)
2011	10,89	-	-
2012	11,41	0,52	4,78
2013	12,73	1,84	16,90
2014	14,46	3,57	32,78
2015	14,50	3,61	33,15

**Bảng 5.2** Chênh lệch phần trăm giá trị nhập khẩu tương ứng với năm 2011.

Ta nhận thấy giá trị nhập khẩu chỉ tăng nhẹ giữa năm 2011 và 2012, nhưng tăng rất nhanh giữa năm 2012 và 2014. Từ 2014 đến 2015 mức độ tăng đã chững lại.

Thay vào đó, biến động giá trị nhập khẩu tương ứng với năm 2011 có thể được thể hiện thông qua chỉ số. Trong **Ví dụ 5.1**, chỉ số của năm 2013 tương ứng với 2011 là:

Chỉ số nhập khẩu năm 2013 với giá trị gốc là

$$\text{năm 2011} = \frac{\text{giá trị nhập khẩu năm 2013} \times 100}{\text{giá trị nhập khẩu năm 2011}} = 116,90$$

Chỉ số từ năm 2011 đến 2015 được thể hiện trong **Bảng 5.3**.

Rõ ràng chỉ số ở giá trị gốc luôn là 100. Các chỉ số tiếp theo sẽ hơn hoặc kém 100, tùy thuộc vào việc đã có sự tăng hay giảm giá trị biến so với giá trị gốc.

Ví dụ, ta có thể nói chỉ số năm 2014 là 132,78, nên giá trị nhập khẩu đã tăng xấp xỉ 32,8% từ năm 2011.

Năm	Nhập khẩu (tỷ £)	Chỉ số với 2011 là giá trị gốc
2011	10,89	$\frac{10,89}{10,89} \times 100 = 100,00$
2012	11,41	$\frac{11,41}{10,89} \times 100 = 104,78$
2013	12,73	$\frac{12,73}{10,89} \times 100 = 116,90$
2014	14,46	$\frac{14,46}{10,89} \times 100 = 132,78$
2015	14,50	$\frac{14,50}{10,89} \times 100 = 133,15$

**Bảng 5.3** Chỉ số nhập khẩu giai đoạn 2011 – 2015 với giá trị gốc là 2011.

Cần làm rõ là việc xác định giá trị gốc khi lập chỉ số là rất quan trọng – chỉ số sẽ vô nghĩa nếu không đi kèm thông tin này.

Cách viết tắt để chỉ ra giá trị gốc là “**2011 = 100**” – như trong **Ví dụ 5.1**. Bất cứ năm nào cũng có thể được chọn làm giá trị gốc, không nhất thiết phải là giá trị đầu tiên trong chuỗi.

Ví dụ này có thể thực hành với Microsoft Excel như sau:

	<b>A</b>	<b>B</b>	<b>C</b>	<b>D</b>	Dữ liệu từ <b>Bảng 5.3.</b>  Ô C2:C6 chứa hàm cần nhập.  Ô D2:D6 chứa giá trị sẽ xuất hiện ở các ô C2:C6.
1	<b>Năm</b>	<b>Nhập khẩu (tỷ £)</b>	<b>Chỉ số với 2011 là giá trị gốc</b>		
2	<b>2011</b>	10,89	=B2/B2*100	100,00	
3	<b>2012</b>	11,41	=B3/B2*100	104,78	
4	<b>2013</b>	12,73	=B4/B2*100	116,90	
5	<b>2014</b>	14,46	=B5/B2*100	132,78	
6	<b>2015</b>	14,50	=B6/B2*100	133,15	

**Bảng 5.4** Ví dụ tính chỉ số thông qua Microsoft Excel.

**Bảng 5.5** cho thấy chỉ số nhập khẩu trong Ví dụ 5.1 với giá trị gốc là năm 2013.

<b>Năm</b>	<b>Nhập khẩu (tỷ £)</b>	<b>Chỉ số với 2013 là giá trị gốc</b>
2011	10,89	$\frac{11,41}{12,73} \times 100 = 85,55$
2012	11,41	$\frac{12,73}{12,73} \times 100 = 89,63$
2013	12,73	$\frac{12,73}{12,73} \times 100 = 100,00$
2014	14,46	$\frac{14,46}{12,73} \times 100 = 113,59$
2015	14,50	$\frac{14,50}{12,73} \times 100 = 113,90$

**Bảng 5.5** Chỉ số nhập khẩu năm 2011 – 2015, với 2013 = 100.

Chỉ số mới đo lường các thay đổi trong cùng một khoảng thời gian, nhưng giờ đây tất cả các thay đổi được đánh giá tương ứng với năm 2013, chứ không phải năm 2011.

Lưu ý rằng chỉ số năm 2011 giờ là 85.5. Điều này có nghĩa là gì? Con số này thấp hơn 100, cho thấy nhập khẩu năm này ở mức thấp hơn năm 2013. Trên thực tế, chỉ số thấp hơn:

$100 - 85,55 = 14,45\%$  so với mức năm 2013.

### 5.2.2 CHỈ SỐ ĐỐI VỚI NHÓM NHIỀU SẢN PHẨM

Chỉ số nhập khẩu đã được tính ở Ví dụ 1 đo lường sự thay đổi trong giá trị của một sản phẩm. Ta cũng cần đến chỉ số sẽ đo lường mức độ thay đổi trong giá trị của nhóm nhiều sản phẩm khác nhau, nên sẽ sử dụng *chỉ số giá nhóm hàng hóa/chỉ số chứng khoán (share price index)*.

Các trung tâm tài chính trên thế giới có chỉ số chứng khoán riêng – như FTSE100 (London), Dow Jones (phố Wall), Nikkei (Tokyo)... Các chỉ số này được công bố trên mục tài chính của các báo và là công cụ quan trọng chỉ thị biến động giá chứng khoán trên thị trường. Chúng đo lường sự thay đổi giá của không chỉ một loại cổ phiếu, mà là một mẫu nhiều loại cổ phiếu khác nhau.

Để tính loại chỉ số này, người thu thập đầu tiên phải chọn mẫu, hoặc “rổ” (*basket*) gồm các loại cổ phiếu. Mẫu được chọn đại diện cho các loại cổ phiếu thường được giao dịch nhất trên thị trường chứng khoán thế giới. Thay đổi giá trong rổ cổ phiếu được theo dõi hàng ngày và tính **chênh lệch bình quân (average change)**.

#### Ví dụ 5.2 – Tính chỉ số giá nhóm hàng hóa đơn giản

**Bảng 5.6** cho thấy biến động giá của ba loại cổ phiếu giữa tháng 3 và tháng 9 năm 2015. Hãy tính chỉ số giá mô tả sự thay đổi trong giá trị bình quân của các loại cổ phiếu trong giai đoạn này.

Cổ phiếu	Giá	Giá
	Tháng 3/2015	Tháng 9/2015
UK Enterprises plc	9,10£	12,15£
World Products Co.	3,40£	3,85£
Euroland Holdings	2,75£	3,55£

**Bảng 5.6** Biến động giá cổ phiếu.

**Lời giải:** Tính giá cổ phiếu bình quân tháng 3 và tháng 9 năm 2015.

$$\text{Giá cổ phiếu bình quân tháng 3/2015 là } \frac{9,10 + 3,40 + 2,75}{3} = 5,08£$$

$$\text{Giá cổ phiếu bình quân tháng 9/2015 là } \frac{12,15 + 3,85 + 3,55}{3} = 6,52£$$

Sử dụng hai giá trị bình quân này để tính chỉ số chứng khoán tháng 9/2015 (chỉ số tháng 3 = 100):

$$\frac{6,52}{5,08} \times 100 = 128$$

Do đó chỉ số giá chứng khoán là:

Tháng 3/2015 = 100

Tháng 9/2015 = 128

Có nghĩa là giá cổ phiếu\* trung bình đã tăng 28% trong giai đoạn này.

Ghi chú: \* chỉ giá các loại cổ phiếu đã được chọn trong “rổ”.

Đối với giá bán lẻ hàng hóa tiêu dùng, phương pháp này thường không phù hợp lắm. Ví dụ, vật giá các sản phẩm trong “giỏ mua sắm” thực phẩm thường không tương đương, do với mỗi sản phẩm thì ở các thời điểm khác nhau người mua sẽ mua chúng với khối lượng hay số lượng khác nhau. Một đơn vị bơ (khoảng 250g) sẽ đủ cho 3 đến 4 đơn vị (ô) bánh mỳ.

Khi xét một nhóm hàng hóa mà mỗi sản phẩm trong “rổ” không tương đương, ta phải cân nhắc đến **tính bình quân gia quyền (weighting)** của chúng đối với chỉ số.

### 5.3 CHỈ SỐ BÌNH QUÂN GIA QUYỀN

Nếu các sản phẩm trong “rổ” có tỷ trọng không tương đương, một cách để tính bình quân gia quyền chỉ số của từng sản phẩm trong phân phối là dựa vào phí tổn cho sản phẩm này, nghĩa là chi phí chi trả để mua sản phẩm tại một thời điểm nhất định. Phí tổn phụ thuộc trực tiếp vào **khối lượng** cũng như giá cả sản phẩm được mua tại thời điểm đó.

Đây là nguyên tắc tính Chỉ số Giá Bán lẻ (*Retail Price Index – RPI*), dùng để kiểm soát chi phí tiêu dùng tại Anh, trong số các loại giá khác.

Chỉ số RPI được xác định dựa trên giỏ mua sắm các loại hàng hóa và dịch vụ rõ ràng của một mẫu nghiên cứu lớn gồm nhiều hộ gia đình trong một thời điểm nhất định. Chỉ số này phản ánh chính xác giá trị tương đối giữa nhiều sản phẩm đã được mua trong thời điểm cho trước, dựa vào phí tổn trên mỗi sản phẩm.

$$\text{Chỉ số giá} = \frac{\text{tổng phí tổn chi trả cho "giỏ mua sắm" tại thời điểm hiện tại}}{\text{tổng phí tổn chi trả cho "giỏ mua sắm" tại năm mốc}}$$

Tổng phí tổn chi trả cho một sản phẩm được tính bằng khối lượng nhân với giá trên mỗi đơn vị của sản phẩm đó. Cộng tất cả các phí tổn với nhau, ta có tổng phí tổn chi trả cho toàn bộ “giỏ mua sắm”.

Nếu giá một sản phẩm thay đổi theo thời gian, thì khả năng cao là khối lượng mua hàng cũng thay đổi theo. Khi giá một mặt hàng tăng, chúng ta thường phản ứng lại theo hướng mua ít món đó hơn. Ngoài ra, có thể sau một thời gian, ta có nhiều tiền hơn và sẽ mua nhiều món đó hơn, bất kể giá có tăng hay không.

Do đó, để tính chỉ số giá, ta phải quyết định xem nên sử dụng khối lượng mua hàng ở năm gốc hay khối lượng ở thời điểm hiện tại để tính phí tổn trên mỗi sản phẩm.

Có hai hướng tiếp cận thường được sử dụng. Cả hai hướng này đều được thực hiện với điều kiện coi khối lượng mua hàng không đổi theo thời gian, ngay cả khi thay đổi giá.

### 5.3.1 TÍNH BÌNH QUÂN GIA QUYỀN DỰA TRÊN NĂM GỐC

Hướng tiếp cận thứ nhất là phương pháp của Laspeyres và có thể được gọi là chỉ số giá Laspeyres. Một cách rõ ràng hơn, do chỉ số này coi khối lượng mua hàng không đổi so với năm gốc, nên có thể gọi là **chỉ số giá dựa trên trọng số năm gốc (base-weighted price index)**.

Chỉ số giá dựa trên trọng số năm gốc

$$= \frac{\text{tổng phí tổn chi trả cho khối lượng mua năm gốc với giá hiện tại}}{\text{tổng phí tổn chi trả cho khối lượng mua năm gốc với giá năm gốc}} \times 100$$

Ta ký hiệu  $p_n$  là giá đơn vị hiện tại;  $p_0$  là giá đơn vị năm gốc;  $q_n$  là khối lượng mua ở hiện tại;  $q_0$  là khối lượng mua ở năm gốc. Công thức tính chỉ số giá dựa trên trọng số năm gốc có thể viết là:

$$\text{Chỉ số giá dựa trên trọng số năm gốc} = \frac{\sum p_n \times q_0}{\sum p_0 \times q_0} \times 100$$

(ký hiệu  $\sum$  có nghĩa là “tổng số”)

#### Ví dụ 5.3 – Xác định chỉ số giá dựa trên trọng số năm gốc

Giá trên mỗi đơn vị bánh mì, sữa và bơ của tháng 1 năm 2014 và 2015 được đưa ra trong **Bảng 5.7**. Giả sử ta đã tiến hành nghiên cứu và biết vào tháng 1/2014, trung bình người mua đã phân bổ cơ cấu mua hàng như sau: cứ mỗi 1£ mua bơ thì có 1£ mua sữa và 3£ mua bánh mì. Tính chỉ số giá tiêu dùng gia quyền của ba loại hàng này vào tháng 1/2015, với trọng số là tháng 1/2014.

**Lời giải:** Sử dụng các số 3, 1 và 1 lần lượt là khối lượng mua hàng bánh mì, sữa và bơ; do ta có thể thấy rõ là các số này tương ứng với khối lượng mỗi mặt hàng được mua.

Chỉ số giá gia quyền được tính trong **Bảng 5.7**.

Mặt hàng	Giá đơn vị (xu)		Khối lượng, $w_0$	$w_0 \times p_0$	$w_0 \times p_n$
	Tháng 1/2014	Tháng 1/2015			
Bánh mì	90	99	3	270	297
Sữa	60	61	1	60	61
Bơ	80	90	1	80	90
<b>Tổng</b>				<b>410</b>	<b>448</b>

**Bảng 5.7** Tính chỉ số giá gia quyền (Tháng 1/2014 = 100).

$$\text{Chỉ số giá gia quyền} = \frac{448}{410} \times 100 = 109$$

Số liệu này cho thấy giá đã tăng 9% trong năm.

### 5.3.2 TÍNH BÌNH QUÂN GIA QUYỀN DỰA TRÊN NĂM HIỆN TẠI

Trong phần trước, chúng ta đã coi khối lượng mua các mặt hàng ở thời điểm hiện tại không đổi so với năm gốc. Ngoài ra, thay vào đó ta có thể giả sử rằng trong năm gốc, người mua đã mua khối lượng giống như ở thời điểm hiện tại. Hướng tiếp cận này cho ta kết quả là **chỉ số giá dựa trên trọng số năm hiện tại (current year weighted price index)**, theo phương pháp của Paasche. Ta có:

Chỉ số giá dựa trên trọng số năm hiện tại

$$= \frac{\text{tổng phí tổn chi trả cho khối lượng mua hiện tại với giá hiện tại}}{\text{tổng phí tổn chi trả cho khối lượng mua hiện tại với giá hiện tại}} \times 100$$

Ký hiệu như phần trước, ta viết:

$$\text{Chỉ số giá sử dụng trọng số hiện tại} = \frac{\sum p_n \times q_0}{\sum p_0 \times q_n} \times 100 \text{ (ký hiệu } \sum \text{ có nghĩa là “tổng số”)}$$

### Ví dụ 5.4 – Tính chỉ số giá gia quyền dựa trên trọng số năm hiện tại

Xem lại **Ví dụ 5.3**. Nghiên cứu thị trường năm 2015 cho thấy cứ mỗi 5£ dùng để mua bánh mì, trung bình người ta mua 3£ sữa và 1£ bơ. Dựa vào trọng số này, hãy chỉ số giá cho tất cả sản phẩm năm 2015 (2014 = 100).



**Lời giải:** Phép tính chỉ số mới được thể hiện ở **Bảng 5.7**.

$$\text{Chỉ số giá gia quyền dựa trên trọng số năm hiện tại} = \frac{768}{710} \times 100 = 108$$

Số liệu này cho thấy giá đã tăng 8% trong năm, hơi thấp hơn giá trị đạt được với trọng số năm 2014.

Mặt hàng	Giá đơn vị (xu)		Khối lượng, $w_n$	$w_n \times p_0$	$w_n \times p_n$
	Tháng 1/ 2014	Tháng 1/2015			
Bánh mì	90	99	5	450	495
Sữa	60	61	3	180	183
Bơ	80	90	1	80	90
<b>Tổng</b>				<b>710</b>	<b>768</b>

**Bảng 5.8** Tính chỉ số giá gia quyền (tháng 1/2014 = 100).

Ví dụ này có thể được thực hành với phần mềm Microsoft Excel như sau:

	A	B	C	D	E	F	Dữ liệu từ <b>Bảng 5.8</b> .  Ô E3:F6 chứa hàm cần nhập và giá trị sẽ xuất hiện (trong ngoặc đơn)
1	Mặt hàng	Giá đơn vị (xu)		Khối lượng, $w_n$	$w_n \times p_0$	$w_n \times p_n$	
2		tháng 1/ 2014	tháng 1/ 2015				
3	Bánh mì	90	99	5	=D3*B3 (450)	=D3*C3 (495)	
4	Sữa	60	61	3	=D4*B4 (180)	=D4*C4 (183)	
5	Bơ	80	90	1	=D5*B5 (80)	=D5*C5 (90)	
6	Tổng				=sum(E3:E5) (710)	=sum(F3:F5) (768)	

**Bảng 5.9** Ví dụ sử dụng Microsoft Excel để tính chỉ số giá gia quyền.

### 5.3.3 TÍNH BÌNH QUÂN GIA QUYỀN DỰA TRÊN NĂM GỐC HAY NĂM HIỆN TẠI

Trong **Ví dụ 5.3** ta đã tính chỉ số dựa trên trọng số năm gốc (*base-weighted*). Với **Ví dụ 5.4** ta sử dụng trọng số hiện tại để cho ra kết quả là chỉ số giá gia quyền với trọng số năm hiện tại (*current-weighted*).

Chỉ số dựa trên trọng số năm gốc đo lường thay đổi giá cả trung bình trong một “rổ” các loại hàng hóa cố định thường được mua trong năm gốc.

Chỉ số dựa trên trọng số năm hiện tại cũng đo lường thay đổi giá cả trung bình, nhưng phạm vi là trong rổ hàng thường được mua ở năm hiện tại.

Như vậy, loại chỉ số ta nên lựa chọn phụ thuộc vào ý định sử dụng của chúng ta.

Ví dụ, trọng số năm gốc sẽ có ích trong trường hợp ta muốn xác định giá của rổ hàng nhất định được mua từ năm trước đã tăng bao nhiêu. Phương pháp này cũng có lợi thế do giá trị trọng số là cố định.

Trọng số hiện tại sẽ được lựa chọn khi ta cần biết sự thay đổi giá của rổ hàng hiện tại đã diễn ra như thế nào trong các năm qua. Tuy nhiên, lưu ý rằng phương pháp này sẽ tốn thêm phí để liên tục thu thập dữ liệu mới nhằm xây dựng tỷ trọng tương ứng với từng thời điểm.

Bất kể lựa chọn chỉ số nào, tỷ trọng cần được duy trì cố định xuyên suốt quy trình xác định chỉ số.

## 5.4 GIẢM PHÁT CHUỖI THỜI GIAN – SỬ DỤNG CHỈ SỐ GIÁ NHƯ CHỈ SỐ GIẢM PHÁT

Xét trường hợp dưới đây. Giữa năm 2013 và 2015, lương theo tuần trung bình trong một ngành công nghiệp thay đổi như sau:

Năm	2013	2014	2015
Lương (£/tuần)	420	460	485

**Bảng 5.10** Dữ liệu lương theo tuần trung bình.

Người sử dụng lao động trong ngành cho biết mức lương đã tăng thực sự đáng kể trong giai đoạn này, nhưng công đoàn thì không cho là như vậy. Để làm sáng tỏ vấn đề này, công đoàn đã tham khảo chỉ số giá bán lẻ trong cùng giai đoạn:

Năm (tháng 1)	2013	2014	2015
Chỉ số giá, (tháng 1/2007=100)	186,7	192,0	198,1

**Bảng 5.11** Dữ liệu chỉ số giá bán lẻ.

Biết rằng lạm phát giá ảnh hưởng đến mức lương và thu nhập, như vậy làm thế nào để ta xác định xem mức lương trong ngành này đã ứng phó được mức lạm phát hay chưa? Ta muốn biết mức lương sẽ thay đổi ra sao, nếu để nó tự điều chỉnh tương ứng với thay đổi giá tiêu dùng.

Ta có thể sử dụng chỉ số giá bán lẻ để loại bỏ ảnh hưởng của lạm phát giá trong số liệu lương. Quá trình này gọi là giảm phát (*deflating*) chuỗi lương, nghĩa là ta sử dụng chỉ số giá bán lẻ làm chỉ số giảm phát (*deflator*).

Trước khi phân tích tiếp, ta phải thay đổi chỉ số gốc của giá bán lẻ từ 2007 = 100 đến 2013 = 100. Xem **Bảng 5.12**.

Năm (tháng 1)	2013	2014	2015
Chỉ số giá, (tháng 1/2007=100)	186,7	192,0	198,1
Chỉ số giá, (tháng 1/2013=100)	100,0	$\frac{460}{102,8} \times 100 = 102,8$	$\frac{460}{102,8} \times 100 = 106,1$

**Bảng 5.12** Đổi chỉ số gốc.

Giữa năm 2013 và 2015, mức lương tăng  $\frac{485}{420} \times 100 = 115,5$  nghĩa là 15,5%

Cùng kỳ, giá tăng 6,1%.

Để điều chỉnh lương tương ứng với lạm phát, ta tính:

$$\text{Giá trị lương đã giảm phát (giá trị thực)} = \frac{\text{giá trị lương thực tế}}{\text{chỉ số giá bán lẻ}} \times 100$$

Diễn giải bằng lời, ta chia giá trị muốn giảm phát cho chỉ số thích hợp tại thời điểm đang xét và nhân với 100.

Giờ ta đã có chỉ số giá bán lẻ với gốc 2013 = 100. Ta có thể giảm phát giá trị chuỗi lương. Xem **Bảng 5.13**.

Năm	2013	2014	2015
Lương nhận (£/tuần)	420	460	485
Lương thực tế (£/tuần)	420	$460 \times 100 = 447$ 102,8	$485 \times 100 = 457$ 106,1

**Bảng 5.13** Giảm phát chuỗi lương.

Ta đã chuyển đổi các số liệu lương theo hướng điều chỉnh so với lạm phát giá. Giá trị đã giảm phát thu được thường được gọi là lương “thực tế” (*real wages*).

Ta có thể thấy lương thực tế đã tăng – phần giá trị tiền tăng có giảm đi, nhưng không bị lạm phát giá phủ nhận hoàn toàn.

## 5.5 TÀI LIỆU THAM KHẢO CHƯƠNG

1. Morris, C, Quantitative Approaches in Business Studies (*tạm dịch*: Phương pháp định lượng trong nghiên cứu kinh doanh), tái bản lần thứ 8, FT/Prentice Hall, 2012 (eBook, tái bản lần thứ 7, 2008), Chương 7.
2. Oakshott, L, Essential Quantitative Methods for Business (*tạm dịch*: Phương pháp nghiên cứu định lượng thiết yếu trong kinh doanh, quản lý & tài chính), tái bản lần thứ 5, Palgrave Macmillan, 2012, Chương 2.
3. Wisniewski, M, Quantitative Methods for Decision Makers (*tạm dịch*: Phương pháp nghiên cứu định lượng phục vụ ra quyết định), tái bản lần thứ 5, FT/Prentice Hall, 2009 (có bản eBook), Chương 7.

## Chương 6

# LIÊN KẾT GIỮA CÁC BIẾN SỐ

### 6.1 GIỚI THIỆU

Trong chương trước, chúng ta đã thảo luận về sự biến động giá trị các biến số trong kinh doanh. Ta thấy rất nhiều hoạt động trong kinh doanh có xu hướng dễ thay đổi. Trong cùng một hoạt động (bán hàng, đi công tác, sản xuất hàng hóa), thông thường hai lượt thực hiện khác nhau sẽ có kết quả hơi khác nhau, về thời gian thực hiện hoặc chất lượng kết quả đầu ra. Tình trạng này là khá tự nhiên, và ở mức độ nào đó thì sự biến động là không thể tránh khỏi.

Công ty có thể nhận thấy doanh thu bán hàng thay đổi giữa các tháng. Câu hỏi là, liệu công ty có thể giải thích nguyên do gây ra sự biến động này (và có thể kiểm soát biến động bằng cách nào đó) hay đây là thực tế bắt buộc phải chấp nhận? Đối với công ty, một số biến động trong doanh thu bán hàng có thể được giải thích là do sản phẩm được quảng cáo. Điều này ảnh hưởng đến đơn giá, cũng như số lượng sản phẩm bán được. Một yếu tố có thể ảnh hưởng khác là mức độ quảng cáo, cũng thay đổi từ tháng này qua tháng khác. Các công ty thường tin rằng chi nhiều tiền quảng cáo hơn sẽ giúp tăng doanh thu bán hàng cho tháng tiếp theo. Có những yếu tố khác cũng ảnh hưởng đến biến động doanh thu bán hàng. Ví dụ, hành vi mua sắm của người tiêu dùng có thể dễ bị tác động bởi thời tiết (một yếu tố dễ thay đổi), từ đó ảnh hưởng đến giá bán sản phẩm của đối thủ cũng như tình hình quảng cáo của nhà sản xuất.

Như vậy trong tình huống này, có nhiều yếu tố biến động đến mức nào đó, nhưng một biến động sẽ tương quan thế nào đến những biến động còn lại? Ví dụ, liệu thực sự có một mối liên kết giữa các biến động trong chi phí quảng cáo và biến động trong doanh thu bán hàng? Nếu xác định được mức độ của mối liên kết này (nếu tồn tại), thì công ty có thể quyết định chắc chắn hơn cách phân bổ quảng cáo hợp lý nhất để tăng doanh thu. Mặt khác, giữa doanh thu bán hàng và thời tiết – được đo lường bởi mức nhiệt hoặc lưu lượng mưa trung bình hàng tháng – có thể có mối liên kết lớn hơn nhiều. Trong trường hợp này, việc điều chỉnh chi phí quảng cáo có thể chỉ ảnh hưởng rất ít đến doanh thu hàng tháng và khi đó sẽ tốn kém rất nhiều tiền bạc.

Có thể đánh giá mức độ liên kết giữa các cặp yếu tố bằng phương pháp phân tích tương quan (correlation analysis). Phương pháp này đo lường mức độ liên kết giữa hai lượng biến đổi (mà ta gọi là “biến số”). Sự biến động có thể diễn ra theo thời gian (như trong ví dụ về doanh thu bán hàng) hoặc không gian từ nơi này đến nơi khác. Tuy nhiên, trước khi vội vàng sử dụng Microsoft Excel để thực hiện phân tích tương quan, ta phải đưa ra một số quyết định tiên quyết.

Giả sử ta quyết định phân tích tương quan giữa doanh thu hàng tháng của một sản phẩm với chi phí quảng cáo hàng tháng cho sản phẩm đó. Câu hỏi đầu tiên là:

*“Liệu có đúng khi giả sử có mối liên kết giữa hai yếu tố này?”*

Câu trả lời hần là “có” – nhiều năm kinh nghiệm giúp ta hiểu rằng việc quảng cáo sản phẩm có xu hướng ảnh hưởng đến tình hình bán hàng, vì thế tác động đến doanh thu.

Tuy nhiên, giả sử ta đang quan tâm đến số lượng vé máy bay từ Anh đến Tây Ban Nha được đặt mỗi tuần và nhiệt độ trung bình ở Anh trong tuần trước đó. Trong trường hợp này, ta thấy mối liên hệ ít rõ ràng hơn – do sẽ có nhiều yếu tố ảnh hưởng đến việc mua vé máy bay và nhiệt độ hiện tại ở Anh – nên dù có ảnh hưởng, thì cũng có thể không phải yếu tố ảnh hưởng chính.

Ta lại giả sử đang có trong tay dữ liệu doanh thu bán ô tô hàng tuần ở New York và lượng tiêu thụ xăng dầu hàng tuần ở Anh. Ta có thể thực hiện phân tích tương quan giữa hai loại số liệu này, nhưng liệu việc này có tác dụng gì không? Với cả ba ví dụ đã được nhắc đến, việc phân tích tương quan sẽ cho ra kết quả nào đó, nhưng chúng có thể không hoàn toàn có ích – như vậy nếu cân nhắc một chút từ trước, ta có thể tránh lãng phí rất nhiều thời gian và công sức.

## 6.2 PHÂN TÍCH TƯƠNG QUAN

Giả sử ta đã khẳng định được việc thực hiện phân tích tương quan là hữu ích, khi đó phép phân tích (thực hiện bởi Microsoft Excel) cho kết quả là một số (gọi là **hệ số tương quan - correlation coefficient**), dựa trên mức độ dao động chung của hai biến số. Giá trị hệ số tương quan thể hiện mối liên hệ tương quan dương (positive) hay âm (negative) và mức độ mạnh yếu của mối liên hệ đó. Hệ số tương quan được ký hiệu là **r** hoặc **R** và luôn có giá trị nằm trong khoảng giữa -1 và +1.

Giá trị gần với +1 (ví dụ 0,9) cho thấy mối liên hệ tương quan dương **tuyến tính (linear)** mạnh giữa hai yếu tố, khẳng định rằng một yếu tố thường nhận giá trị cao (trong khoảng biến động) khi yếu tố còn lại cũng nhận giá trị cao (trong khoảng biến động). Tương tự như vậy, cặp giá trị sẽ có xu hướng cùng nhận giá trị thấp như nhau.

Giá trị gần với -1 (ví như -0,9) cho thấy mối liên hệ tương quan âm **tuyến tính** mạnh và khi đó một yếu tố nhận giá trị cao nếu yếu tố còn lại nhận giá trị thấp, và ngược lại.

Trong cả hai trường hợp này, nếu giá trị của hệ số tương quan càng về gần 0 thì mức độ **tuyến tính** của liên hệ càng giảm. Tuy nhiên, trường hợp hệ số tương quan bằng 0,8 không có nghĩa là nó tốt gấp đôi hệ số tương quan bằng 0,4. Ta chỉ có thể nói là:

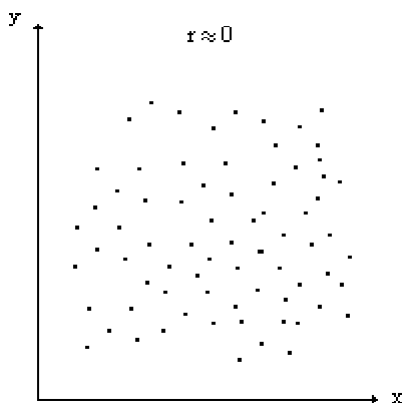
*Nếu  $r = 0,8$ ; giá trị này gần với 1. Do đó ta giả định rằng giữa các biến có mối liên hệ tuyến tính dương khá mạnh.*

*Nếu  $r = 0,4$ ; giá trị gần với 0 hơn 1. Do đó ta giả định giữa các biến có mối liên hệ tuyến tính hơi yếu một chút.*

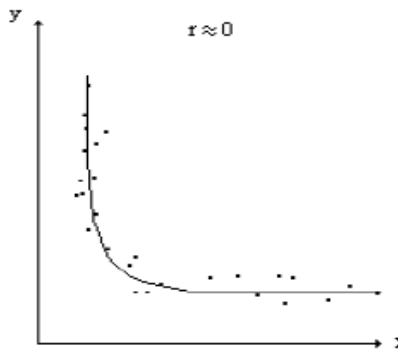
Liên hệ tuyến tính giữa hai biến càng mạnh thì giá trị của  $r$  càng gần 1. Mức độ tuyến tính của liên hệ càng giảm, thì giá trị  $r$  càng về gần 0. Khi  $r = 0$ , giữa các biến không còn liên hệ **tuyến tính**. Điều này không hoàn toàn có nghĩa là giữa các biến số không tồn tại bất cứ mối quan hệ nào.

**Hình 6.1** và **6.2** thể hiện các giá trị có hệ số tương quan gần bằng 0.

Cần hiểu rõ là hệ số tương quan đo lường độ mạnh yếu của liên hệ **tuyến tính** giữa các biến – chứ không phải bất cứ liên hệ nào. Do đó, để đánh giá bản chất của bất kỳ mối liên hệ nào giữa các biến thì ngoài phân tích tương quan, ta nên kèm theo **biểu đồ phân tán (scatter plot)** của dữ liệu.



**Hình 6.1** Dữ liệu cho thấy không có bất kỳ quan hệ nào giữa các biến.



**Hình 6.2** Dữ liệu cho thấy có mối liên hệ mạnh, phi tuyến tính giữa các biến.

Cuối cùng, sau khi thực hiện phân tích tương quan, ta phải cẩn thận trong quá trình diễn giải ý nghĩa số liệu. Đừng vội vàng kết luận rằng chỉ vì hai biến số có tương quan chặt chẽ nên giữa chúng bắt buộc phải có quan hệ nhân quả. Trường hợp hai biến A và B tương quan chặt chẽ với nhau có thể do một trong các trường hợp sau:

- A gây ra B
- B gây ra A
- A và B cùng biến động do chịu ảnh hưởng bởi một biến số C thứ ba
- Tương quan là trùng hợp ngẫu nhiên

Do đó, cao phân tích tương quan có khả năng cao chỉ là bước đầu của quá trình nghiên cứu đánh giá.

### 6.3 BIỂU ĐỒ PHÂN TÁN

Một khi đã khẳng định rằng nghiên cứu đánh giá mối liên hệ giữa hai biến là điều đáng làm, thì việc hữu ích đầu tiên là vẽ biểu đồ phân tán.

Mặc dù để thực hiện phân tích tương quan, ta có thể chọn bất kỳ cặp biến số nào, nhưng về sau ta sẽ thấy là nên cân nhắc khi chọn cặp biến số để biểu diễn trên trục x và y của biểu đồ phân tán. Khái niệm về sự phụ thuộc sẽ được xét đến ở đây. Ta nên biểu diễn biến số có xu hướng phụ thuộc trên trục y, nghĩa là biến số mà dễ bị ảnh hưởng hoặc tác động bởi biến số còn lại – được biểu diễn trên trục x. Như vậy, biểu đồ phân tán chỉ đơn giản là sơ đồ thể hiện các cặp giá trị tương ứng của hai biến số. Kiểu hình thu được từ các điểm đã biểu diễn cho ta thông tin về mối liên hệ giữa các biến, cũng như sự phù hợp của tập dữ liệu trước khi ta tốn thời gian để tính toán.

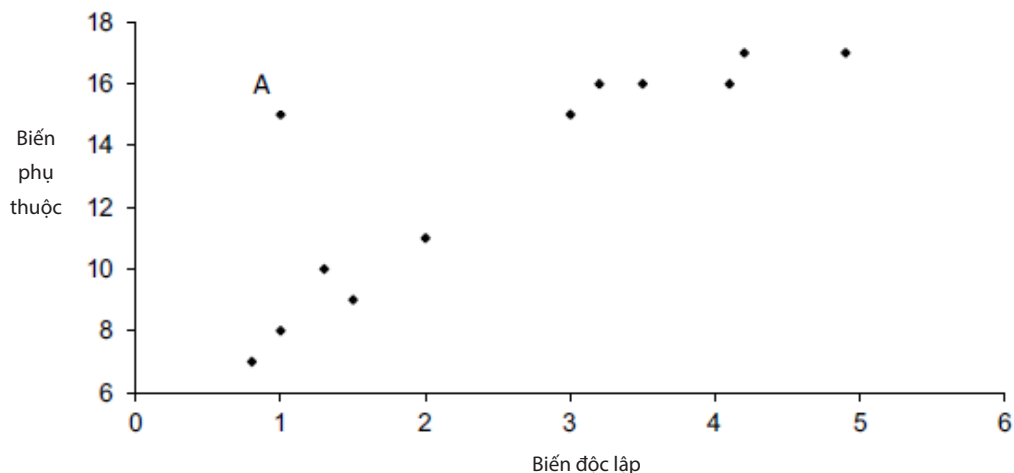
Khi khảo sát biểu đồ phân tán, ta nên cân nhắc nhiều yếu tố trước khi đưa ra kết luận về mối liên hệ giữa các biến. Yếu tố đầu tiên thật dễ thấy là liệu có tồn tại kiểu hình mẫu cố định giữa các điểm đã được biểu diễn không – một kiểu hình rõ ràng cho biết



một mối liên hệ tương quan giữa các biến. Nếu giả định là có dấu hiệu về một mối liên hệ như vậy, thì bước tiếp theo ta sẽ xác định bản chất của mối liên hệ là tuyến tính hay phi tuyến tính.

Cuối cùng, ta nên xác định xem có đặc tính bất thường nào trong các điểm đã được biểu diễn không. Đặc tính này có thể là các “giá trị ngoại lệ” (*outlier*), hoặc khoảng trống lớn trong kiểu hình biểu diễn, hoặc bằng chứng cho thấy tồn tại nhiều hơn một phân phối.

**Hình 6.3** sẽ minh họa cho các đặc tính này.



**Hình 6.3** Biểu đồ phân phối cho thấy các đặc tính bất thường trong tập dữ liệu.

Điểm A có thể được coi là một “giá trị ngoại lệ”, do nằm khá xa so với khoảng phân phối chính của các giá trị dữ liệu khác. Giờ đây xuất hiện câu hỏi tại sao lại xảy ra tình trạng này.

Đây là lỗi có thể được sửa lại sau khi ta kiểm tra và phát hiện. Đây có thể là giá trị chính xác nhưng không điển hình, nghĩa là dữ liệu này được ghi lại trong hoàn cảnh bất thường; vì lỗi này rất dễ bị bỏ qua nên cần thận trọng trước khi loại bỏ bất kỳ giá trị dữ liệu nào. Hoặc thay vào đó, điểm này có thể hoàn toàn hợp lý và việc coi lỗi là ngoại lệ chỉ vì kích cỡ nhỏ của mẫu, nên có thể khắc phục bằng việc sử dụng tập dữ liệu lớn hơn, cho thấy mối quan hệ giữa các biến số yếu hơn so với mối quan hệ đang được xét ở hiện tại.

Ta cũng phải xét xem lý do xuất hiện khoảng trống trong tập dữ liệu giữa 6 điểm bên trái (các giá trị thấp của biến số không phụ thuộc) và 6 điểm bên phải (các giá trị cao của biến số không phụ thuộc). Đó có thể là trùng hợp khi chọn mẫu ngẫu nhiên hoặc các giá trị giữa 2 và 3 của biến độc lập có vấn đề nào đó. Trong trường hợp này, có khả năng đây là hai phân phối khác nhau và ta không nên tính một hệ số tương quan chung cho toàn bộ tập dữ liệu.

### Ví dụ 6.1 – Đánh giá mối liên hệ giữa hai biến số

Sau đây ta sẽ phân tích ví dụ về thời gian để vận chuyển đơn hàng. Chúng ta đang cung cấp dịch vụ vận chuyển đặc biệt cho các đơn hàng với khoảng cách ngắn trong một thành phố. Để tính phí dịch vụ, ta cần đánh giá mối liên hệ giữa thời gian vận chuyển đơn hàng với khoảng cách vận chuyển.

Rõ ràng ngoài khoảng cách vận chuyển, còn có các yếu tố khác ảnh hưởng đến thời gian xử lý – tắc đường, khoảng thời gian trong ngày, sửa đường, thời tiết, hệ thống giao thông, tài xế, loại hình giao thông... Tuy nhiên, theo lối tư duy ban đầu, ta kết luận rằng khoảng cách vận chuyển có khả năng cao là yếu tố chính để xác định thời gian vận chuyển đơn hàng.

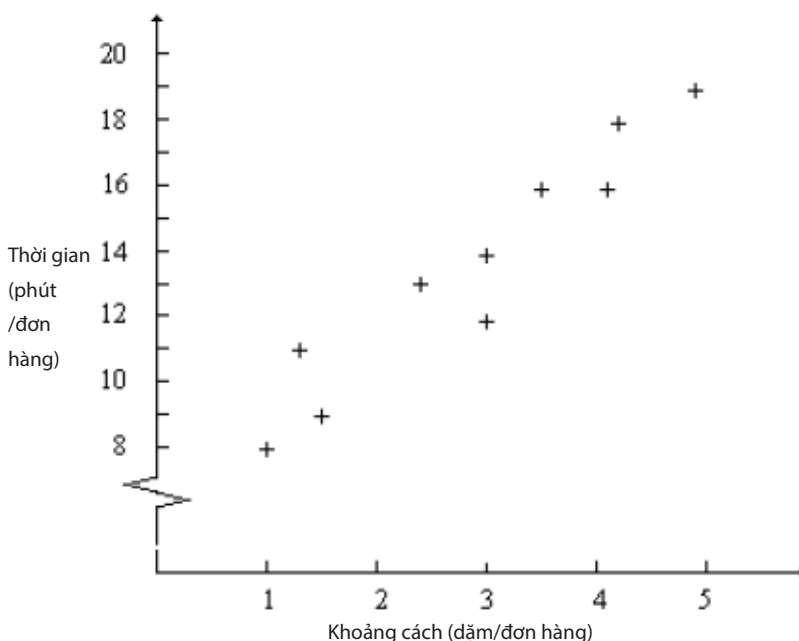
Để nghiên cứu đánh giá đơn giản nhất có thể, ta sẽ chỉ xét đến khoảng cách vận chuyển trong mối liên hệ với thời gian vận chuyển đơn hàng. Khoảng cách được đo lường là đường đi ngắn nhất trong thực tế theo đơn vị dặm Anh (*mile*; 1 mile xấp xỉ 1,6 km), và thời gian vận chuyển đơn hàng tính bằng phút. (Lưu ý: cần làm rõ thời gian vận chuyển bao gồm chính xác thời gian làm những gì, để đảm bảo mọi số liệu đo lường đều mô tả hoạt động tương đương, ví dụ thời gian vận chuyển có tính thời gian xếp và tháo dỡ hàng hay không).

Dữ liệu phân tích được thể hiện trong **Bảng 6.1**. Bảng này cho biết giá trị của hai biến số liên quan trong một số lượt vận chuyển khác nhau. Ta đã chọn ngẫu nhiên các lượt vận chuyển này từ một tập dữ liệu vận chuyển lớn hơn rất nhiều.

Khoảng cách (dặm)	3,5	2,4	4,9	4,2	3,0	1,3	1,0	3,0	1,5	4,1
Thời gian (phút)	16	13	19	18	12	11	8	14	9	16

**Bảng 6.1** Dữ liệu mẫu về khoảng cách và thời gian vận chuyển.

Bước đầu tiên để phân tích tương quan là vẽ biểu đồ phân tán. Trong trường hợp này, thời gian thường là yếu tố bị phụ thuộc vào khoảng cách. Do đó, ta biểu diễn thời gian trên trục y và gọi là **biến phụ thuộc (dependent variable)**. Khoảng cách được biểu diễn trên trục x và được gọi là **biến độc lập (independent variable)**. Xem **Hình 6.4**.



**Hình 6.4** Biểu đồ thời gian vận chuyển tương ứng với khoảng cách cho một mẫu đơn hàng ngẫu nhiên.

Biểu đồ phân tán ở **Hình 6.4** cho thấy xu hướng tăng thời gian tương ứng với khoảng cách. Ta cũng thấy có vẻ như các điểm được biểu diễn đều tập trung quanh một đường thẳng. Dữ liệu không cho thấy đặc tính bất thường nào cả.

Vì thế, có bằng chứng rõ ràng khẳng định mối liên hệ tuyến tính giữa hai biến số này. Tuy không hoàn toàn nằm trên một đường thẳng, nhưng các điểm tập trung trên một kiểu hình tương đối tuyến tính. Thực tế mà nói, sẽ thật ngạc nhiên nếu tất cả các điểm cùng nằm trên một đường thẳng tuyệt đối, nếu xét đến tất cả các yếu tố khác ta biết có thể ảnh hưởng đến thời gian vận chuyển đơn hàng.

Bây giờ, ta có thể phân tích tương quan thông qua một hàm của Microsoft Excel để xác định mức độ chặt chẽ trong tương quan giữa các biến số.

## 6.4 ĐỘ MẠNH CỦA LIÊN HỆ TUYẾN TÍNH – HỆ SỐ TƯƠNG QUAN R

Độ mạnh của liên hệ tuyến tính được đo lường bởi hệ số tương quan Pearson (**Pearson product moment correlation coefficient**), **r**. Có thể tính hệ số này bằng hàm công cụ `correl()` trong Microsoft Excel.

Sử dụng Microsoft Excel, nếu bạn gõ `=correl`, hàm công cụ này sẽ hiện lên trong ô và khi đó chỉ cần nhấn vào hàm cần dùng. Hoặc ta sử dụng phím chức năng `fx` để mở hộp thoại *Insert Function* (chèn hàm), sau đó gõ *Correl* vào ô *Search for a function* (tìm

kiểm hàm) và ấn *Go* (đi tới), hoặc chọn *Statistical* (thống kê) trong ô *Select a category* (chọn danh mục) và cuộn xuống chọn *correl*. Khi ô *Function Arguments* (đối số) hiện ra, nhấn vào Array 1 và đánh dấu tất cả các giá trị của một biến số (biến nào cũng được), rồi nhấn vào Array 2 và đánh dấu tất cả giá trị của biến số còn lại.

#### 6.4.1 TÍNH HỆ SỐ TƯƠNG QUAN R

Quay lại **Ví dụ 6.1** ở trên, ta đã đánh giá mối liên hệ giữa thời gian vận chuyển các đơn hàng có khoảng cách cho trước trong cùng một thành phố. Sử dụng hàm *correl()*, ta tìm được hệ số tương quan  $r = 0,958$  (làm tròn đến 3 chữ số thập phân).

Giá trị hệ số tương quan này rất gần với +1, thể hiện mối liên hệ tuyến tính rất mạnh giữa khoảng cách và thời gian giao hàng. Điều này đã khẳng định lại đánh giá chủ quan thu được từ biểu đồ phân tán.

#### 6.4.2 LIÊN HỆ TUYẾN TÍNH

Trong phần trên, ta đã nói biểu đồ phân tán của thời gian vận chuyển tương ứng với khoảng cách vận chuyển cho thấy có một liên hệ *tuyến tính* giữa hai biến số này – các điểm đã được biểu diễn chủ yếu nằm trên một kiểu hình tuyến tính. Do đó, ta có thể *mô tả* liên hệ giữa hai biến số này dưới dạng một đường thẳng. Đường thẳng này được gọi là **mô hình (model)** của mối liên hệ giữa các biến.

Rõ ràng mô hình này sẽ không thể mô tả *chính xác* mối liên hệ – do các điểm được biểu diễn không hoàn toàn nằm trên một đường thẳng tuyệt đối. Tuy nhiên, vì giá trị hệ số tương quan chỉ ra mối liên hệ tuyến tính mạnh giữa các biến, nên mô hình tuyến tính sẽ phù hợp với dữ liệu ở đây – độ lệch giữa các điểm dữ liệu thực tế và điểm tương ứng trên mô hình sẽ nhỏ thôi.

Mục đích của việc thiết lập mô hình có hai mặt:

- mô tả mối liên hệ giữa các biến trong khoảng dữ liệu đang được sử dụng;
- cho phép ta **ước tính (estimate)** các giá trị biến phụ thuộc (trong Ví dụ 1 là thời gian vận chuyển) tương ứng với bất kỳ giá trị biến độc lập cho trước (trong Ví dụ 1 là khoảng cách vận chuyển) trong khoảng dữ liệu được xét.

Phép phân tích được sử dụng để thiết lập mô hình tuyến tính thích hợp nhằm mô tả mối liên hệ giữa hai biến được gọi là **phân tích hồi quy tuyến tính (linear regression analysis)**.

Trong các phần tới, ta sẽ tóm tắt lại các luận điểm chính đã đề cập và tìm hiểu cách sử dụng Microsoft Excel để thực hiện phân tích hồi quy tuyến tính.

## 6.5 PHÂN TÍCH HỒI QUY

### 6.5.1 TÓM LƯỢC

Thuật ngữ **liên hệ (association)** được sử dụng để nói về mối quan hệ giữa các biến số. Để phục vụ mục đích phân tích thống kê, ta xác định hai khía cạnh của mối quan hệ này. Thuật ngữ **hồi quy (regression)** được dùng để mô tả và mô hình hóa *kiểu hình* của mối quan hệ, còn thuật ngữ tương quan (**correlation**) dùng để mô tả *độ mạnh* của mối quan hệ đó.

Quy trình chung khi phân tích quan hệ giữa các biến là sử dụng một mẫu các giá trị tương ứng của biến để thiết lập bản chất của bất kỳ mối quan hệ nào. Từ đó, ta có thể phát triển phương trình toán học, hoặc **mô hình**, để mô tả mối quan hệ này. Trên quan điểm toán học, phương trình tuyến tính là dễ thiết lập và phân tích nhất. Do đó, trừ trường hợp đã thấy rõ quan hệ tuyến tính, thông thường ta sẽ mô tả mối quan hệ giữa các biến thông qua một mô hình tuyến tính. Quy trình này được gọi là **hồi quy tuyến tính (linear regression)**.

Giá trị đo lường mức độ thích hợp của mô hình tuyến tính đối với dữ liệu là chỉ số cho biết độ mạnh của liên hệ tuyến tính giữa các biến, từ đó ta biết được mức độ đáng tin cậy của bất kỳ phép ước tính nào sử dụng mô hình này. Số đo độ mạnh của liên hệ tuyến tính giữa các biến được gọi là **hệ số tương quan (correlation coefficient)**. Ngoài việc mô tả được mối quan hệ giữa hai biến số, mục đích chính của loại hình phân tích này là cho phép ta ước tính giá trị của biến phụ thuộc. Nếu biết giá trị của biến độc lập, ta có thể sử dụng mô hình hồi quy tuyến tính để ước tính giá trị của biến phụ thuộc. Do đó, nếu cho rằng chi phí quảng cáo hàng tháng có liên hệ mật thiết với doanh thu bán hàng, thì ta có thể thiết lập mô hình thích hợp và sử dụng chi phí quảng cáo trong một tháng để ước tính doanh thu bán hàng sẽ đạt được.

### 6.5.2 MÔ HÌNH HỒI QUY TUYẾN TÍNH ĐƠN GIẢN

Hồi quy tuyến tính đơn giản liên quan đến phân phối của hai biến. Ta quan tâm đến việc liệu có tồn tại liên hệ tuyến tính giữa hai biến hay không. Mật độ của các phân phối được tạo thành từ các cặp giá trị biến, chứ không từ giá trị của một biến đơn lẻ. Ví dụ, ta có thể quan tâm đến chiều cao và cân nặng của một nhóm người, giá cả và số lượng đã bán của một sản phẩm, độ tuổi và mức lương của người lao động, độ tuổi và cân nặng của gà, chi phí hành chính và giờ làm việc hàng tuần, hoặc khoảng cách và thời gian di chuyển.

Bước đầu tiên khi phân tích là xét đến tính chất của các biến, nghĩa là chưa xét đến dữ liệu số. Ta đang quan tâm đến các yếu tố nào? Ta nghĩ các yếu tố đó sẽ ảnh hưởng lẫn nhau ra sao? Ví dụ, một người nông dân muốn đoán biết cân nặng của lứa gà đang nuôi.

Cân nặng là biến số ta cần ước tính; do đó cân nặng là **biến phụ thuộc**. Ta sẽ biểu diễn biến phụ thuộc trên **trục y**.

Người nông dân biết rằng cân nặng phụ thuộc vào độ tuổi của gà. Độ tuổi ở đây được cho là **biến độc lập**. Đây là biến số có giá trị mà ta giả định đã biết, mà ta có thể sử dụng để ước tính cân nặng. Biến độc lập sẽ được biểu diễn trên **trục x**. Nếu có thể thiết lập bản chất của mối liên hệ giữa độ tuổi và cân nặng của gà, thì ta có thể đoán được cân nặng của một con gà khi đến tuổi nhất định. Có thể khảo sát bất cứ con gà nào có cân nặng khác hẳn so với đã dự đoán.

Ta sẽ dự kiến cân nặng của gà thay đổi theo độ tuổi ra sao? Đầu tiên, dự kiến cân nặng sẽ tăng theo độ tuổi, nhưng khi con gà đã trưởng thành thì ta dự kiến cân nặng sẽ duy trì với mức dao động nhỏ, tùy theo dinh dưỡng hoặc thời điểm trong năm. Mức tăng cân và cân nặng trưởng thành cũng sẽ phụ thuộc vào giống gà, cách cho ăn và chăn nuôi. Rõ ràng bên cạnh độ tuổi, có rất nhiều yếu tố sẽ ảnh hưởng đến cân nặng. Như vậy, quy trình cân nhắc các mối quan hệ khả thi, quyết định biến phụ thuộc (y) và biến độc lập (x), suy tính các yếu tố khác có thể ảnh hưởng đến mối quan hệ... là một phần rất quan trọng của mọi quy trình thiết lập mô hình. Mục đích của chúng ta không phải “để thực hiện được hồi quy tuyến tính”, mà để cố hiểu và giải thích được biến động cân nặng thông qua việc thiết lập mô hình. Chỉ dựa vào độ tuổi, liệu chúng ta có biết được cân nặng của một con gà hay không?

Từ phần thảo luận ở trên, ta rút ra kết luận: Có thể kết hợp nhiều yếu tố để xác định cân nặng của một con gà cụ thể, nhưng tổng quan chung là:



**Hình 6.5** Quan hệ khả thi giữa cân nặng và độ tuổi của một giống gà cụ thể.

Giờ chúng ta sẽ đến giai đoạn thu thập mẫu dữ liệu để kiểm chứng các giả thuyết trên.

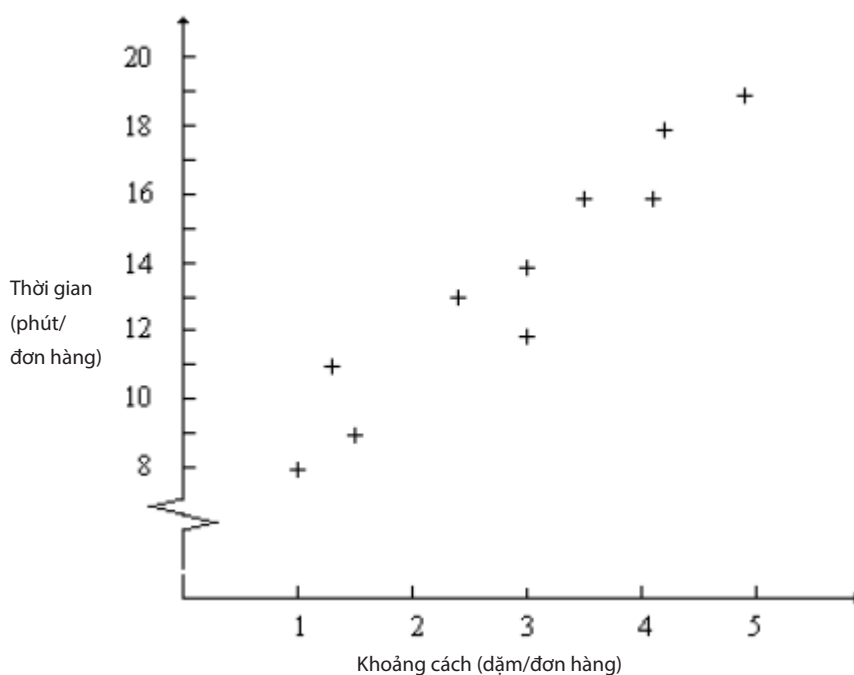
### 6.5.3 BIỂU DIỄN THIẾT LẬP MÔ HÌNH HỒI QUY TUYẾN TÍNH ĐƠN GIẢN

Ta xét lại **Ví dụ 6.1** đã được mô tả ở phần trước. Dữ liệu được nhắc lại dưới đây để tiện theo dõi. Xem **Bảng 6.2**.

Khoảng cách (dặm)	3,5	2,4	4,9	4,2	3,0	1,3	1,0	3,0	1,5	4,1
Thời gian (phút)	16	13	19	18	12	11	8	14	9	16

**Bảng 6.2** Dữ liệu mẫu về khoảng cách và thời gian vận chuyển.

Ta cần diễn tả sự biến động trong khoảng thời gian vận chuyển (biến phụ thuộc  $y$ ), thông qua biến độc lập  $x$  là khoảng cách. Thông thường, ta dự đoán thời gian vận chuyển sẽ tăng khi khoảng cách tăng. Tuy nhiên, như đã giải thích ở trên, thì đầu tiên ta cần biểu diễn dữ liệu mẫu để biết được bất cứ mối liên hệ nào có thể tồn tại giữa các biến số. Biểu đồ phân tán được tái hiện ở **Hình 6.6**.



**Hình 6.6** Biểu đồ thời gian vận chuyển tương ứng với khoảng cách cho một mẫu đơn hàng ngẫu nhiên.

Biểu đồ phân phối cho thấy xu hướng tăng chung của thời gian tương ứng với khoảng cách vận chuyển. Hơn nữa, dường như các điểm được biểu diễn tập trung quanh một

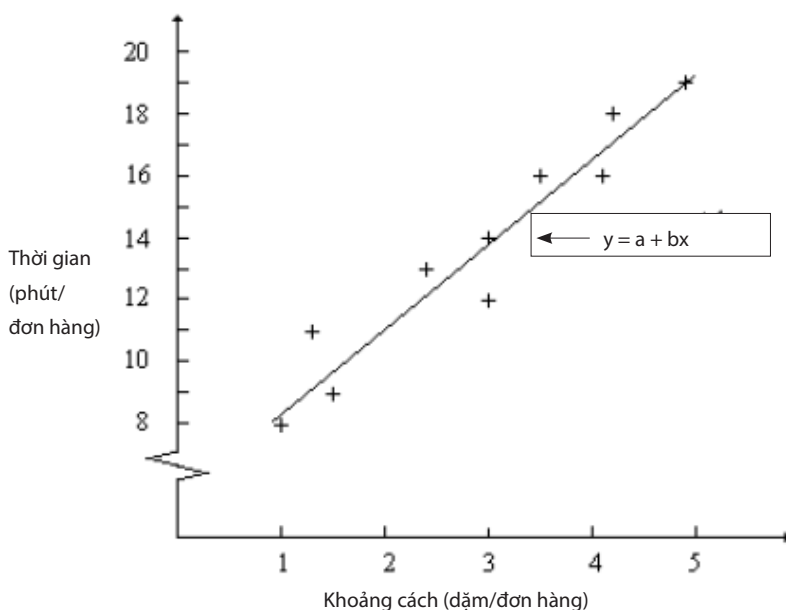
đường thẳng. Điều này có nghĩa là ta có thể sử dụng mô hình tuyến tính để mô tả quan hệ giữa hai biến. Như đã đề cập ở trên, các điểm này không hẳn nằm trọn vẹn trên một đường thẳng. Ta cũng không kỳ vọng đạt được điều này, khi biết rằng tất cả các yếu tố khác có thể ảnh hưởng đến thời gian vận chuyển. So với mối quan hệ thực chất giữa thời gian và khoảng cách vận chuyển thì mô hình tuyến tính chỉ là phương pháp tương đối, nhưng biểu đồ phân tán đã cho thấy đây là lựa chọn sẵn có tốt nhất.

Như vậy, ta đã xác định mô hình tốt nhất để mô tả mối quan hệ giữa thời gian và khoảng cách vận chuyển là tuyến tính. Giờ ta cần một phương pháp để xác định đường thẳng thích hợp nhất với các điểm mẫu. Đường thẳng được cho là **phù hợp nhất (best fit)** sẽ trùng khớp gần nhất có thể với tất cả các điểm dữ liệu. Các điểm này nên tập trung ngẫu nhiên xung quanh đường thẳng được thể hiện ở **Hình 6.7**.

Đường thẳng dựng được qua các điểm dữ liệu là mô hình tuyến tính khả thi nhằm mô tả mối liên hệ giữa các biến. Phương trình đường thẳng này có thể viết như sau:

$$y = a + bx$$

trong đó  $a$  là hằng số (đôi khi gọi là hệ số chặn trên trục  $y$ ) và  $b$  là hệ số góc của đường thẳng. Hệ số góc của đường thẳng thường được gọi là **hệ số hồi quy (regression coefficient)**. (Đừng nhầm với hệ số tương quan nhé).



**Hình 6.7** Biểu đồ thời gian vận chuyển tương ứng với khoảng cách cho một mẫu đơn hàng ngẫu nhiên.



Ta có thể sử dụng lệnh “Trend line” (*đường xu hướng*) trong Microsoft Excel để tìm giá trị của a và b.

Để tiếp tục thảo luận, ta giả sử đã tìm ra được các giá trị này là:

$$a = 5,91$$

$$b = 2,66 \quad (\text{cả hai giá trị đều được làm tròn đến 2 chữ số thập phân})$$

Thay các giá trị này vào công thức mô hình tuyến tính, ta được:

$$y = 5,91 + 2,66x \text{ hoặc}$$

$$\text{Thời gian vận chuyển (phút)} = 5,91 + 2,66 \text{ Khoảng cách vận chuyển (dặm)}$$

Giá trị của hằng số “a” có thể có hoặc không mang ý nghĩa tượng hình rõ ràng.

Trong khi đó, giá trị hệ số hồi quy thì luôn như vậy. Đó là khoảng tăng của giá trị biến ở trục y khi giá trị biến ở trục x tăng thêm 1 đơn vị.

Trong ví dụ này, hệ số góc của đường hồi quy (2,66 phút trên mỗi dặm) có thể coi là số phút ước tính cần thiết cho mỗi dặm đường vận chuyển. Tương tự vậy, hằng số (5,91 phút) có thể coi là thời gian ước tính cần thiết để chuẩn bị đơn hàng trước khi vận chuyển, nghĩa là toàn bộ thời gian cần thiết để vận chuyển đơn hàng không tính thời gian đi đường.

## 6.6 HỆ SỐ XÁC ĐỊNH

Chúng ta đã thảo luận về cách đo độ mạnh của liên hệ tuyến tính thông qua hệ số tương quan, ký hiệu là r.

Song đôi khi thật khó diễn giải rõ ràng giá trị r, do phạm vi đo lường không tuyến tính. Tuy nhiên, nếu bình phương hệ số tương quan ta sẽ được số đo là một tỷ lệ. Tỷ lệ này được gọi là **hệ số xác định (coefficient of determination)** và được ký hiệu là  $r^2$ . Bởi là tỷ lệ nên  $r^2$  có thể nhận giá trị từ 0 đến 1, dù thường được diễn đạt dưới dạng phần trăm. Tỷ lệ phần trăm này cho thấy mức độ biến động của giá trị y trong tương quan tuyến tính với x. Nếu giữa x và y có quan hệ tuyến tính tuyệt đối thì  $r^2 = 1$  hoặc 100%. Không tồn tại bất cứ quan hệ tuyến tính nào có nghĩa  $r^2 = 0$  hoặc 0%.

Sử dụng Microsoft Excel, có thể xác định  $r^2$  thông qua hàm RSQ() hoặc đơn giản là bình phương chính giá trị của r.

Trong ví dụ của chúng ta, hệ số xác định ( $r^2 \times 100\%$ ) cho biết phần trăm biến động toàn bộ trong thời gian vận chuyển đã được giải thích trong mối quan hệ tuyến tính với khoảng cách vận chuyển. Ví dụ này có hệ số xác định khá cao.

$$r^2 = 0,958^2 \times 100 = 91,8\%$$

Mô hình mẫu:

$$\text{Thời gian (phút)} = 5,91 + 2,66 \text{ Khoảng cách (dặm)}$$

đã giải thích được 91,8% biến động trong thời gian đã quan sát. Mô hình mẫu vẫn chưa giải thích 8,2% biến động còn lại trong thời gian vận chuyển. 8,2% biến động là do tác động của tất cả các yếu tố khác đến thời gian vận chuyển mà chưa được thể hiện trong mô hình.

## 6.7 ƯỚC TÍNH SỬ DỤNG MÔ HÌNH HỒI QUY TUYẾN TÍNH

Ta có thể sử dụng mô hình để ước tính thời gian vận chuyển trung bình cho khoảng cách nhất định. Giả sử khoảng cách là 4,0 dặm thì thời gian vận chuyển trung bình ước tính là:

$$y = 5,91 + 2,66 \times 4,0 = 16,6 \text{ nghĩa là khoảng 17 phút}$$

Có lưu ý quan trọng ở đây. Theo thông lệ thì ta không nên sử dụng mô hình để ước tính cho các giá trị biến độc lập nằm ngoài khoảng dữ liệu đã dùng để thiết lập mô hình. Trong mẫu các đơn hàng này, khoảng cách dao động từ 1,0 đến 4,9 dặm. Ta không có bằng chứng đảm bảo là mô hình vẫn đúng ở ngoài khoảng này. Quan hệ giữa thời gian và khoảng cách có thể thay đổi khi khoảng cách vận chuyển tăng. Ví dụ, với quãng đường dài hơn ta phải di chuyển bằng đường cao tốc, trong khi dữ liệu mẫu chỉ nhắc đến hành trình chậm hơn trong phạm vi thành phố. Tương tự như vậy, quãng đường dài có thể phát sinh thời gian ăn uống hoặc nghỉ ngơi, gia tăng đáng kể thời gian thực hiện đơn hàng. Như vậy thông tin chúng ta đã phân tích chỉ là một mẫu các đơn hàng có khoảng cách từ 1 đến 4,9 dặm trong toàn bộ tổng thể. Nếu muốn ngoại suy các khoảng cách ngoài phạm vi này, ta nên thu thập nhiều dữ liệu hơn. Trong trường hợp không thể, ta phải vô cùng cẩn thận khi sử dụng mô hình để ước tính thời gian giao hàng bất kỳ. Giá trị ước tính như vậy thường không đáng tin cậy.

### 6.7.1 ĐÁNH GIÁ MỨC ĐỘ ĐÁNG TIN CẬY CỦA GIÁ TRỊ ƯỚC TÍNH

Các giá trị chúng ta đoán được đáng tin cậy đến mức nào? Tuy không có câu trả lời đơn giản, song vào cách mà mô hình phỏng đoán các giá trị dữ liệu đã biết, chúng ta có thể nói đôi điều về tính phù hợp của mô hình, hay nói cách khác là năng lực phỏng đoán. Hai khía cạnh thể hiện điều này là:

#### a) Hệ số xác định ( $r^2$ )

“phần trăm biến động của biến phụ thuộc  $y$  được giải thích thông qua dữ liệu đã biết về biến độc lập  $x$ ”

Như trong ví dụ về thời gian vận chuyển,  $r^2 = 0,918$ . Do đó khoảng 92% biến động về thời gian vận chuyển có thể được giải thích thông qua dữ liệu đã biết về (biến động trong) khoảng cách vận chuyển. Giá trị ước tính thường đáng tin cậy.

b) Vị trí của giá trị ước tính trong tập dữ liệu

Theo kinh nghiệm và thường thức, giá trị ước tính ở gần khoảng giữa của tập dữ liệu thường đáng tin cậy hơn giá trị ước tính ở các khoảng đầu và cuối phân phối. Khoảng cách 4,0 dặm nằm ở gần cuối khoảng dữ liệu, tuy nhiên vẫn nằm trong phạm vi chính nên ta hoàn toàn có thể kết luận giá trị ước tính này là đáng tin cậy.

## 6.8 TÀI LIỆU THAM KHẢO CHƯƠNG

1. Morris, C, Quantitative Approaches in Business Studies (*tạm dịch*: Phương pháp định lượng trong nghiên cứu kinh doanh), tái bản lần thứ 8, FT/Prentice Hall, 2012 (eBook, tái bản lần thứ 7, 2008), Chương 3.
2. Oakshott, L, Essential Quantitative Methods for Business (*tạm dịch*: Phương pháp nghiên cứu định lượng thiết yếu trong kinh doanh, quản lý & tài chính), tái bản lần thứ 5, Palgrave Macmillan, 2012, Chương 4.
3. Wisniewski, M, Quantitative Methods for Decision Makers (*tạm dịch*: Phương pháp nghiên cứu định lượng phục vụ ra quyết định), tái bản lần thứ 5, FT/Prentice Hall, 2009 (có bản eBook), Chương 3.



## Chương 7

# CHUỖI THỜI GIAN VÀ DỰ BÁO

### 7.1 GIỚI THIỆU

Mọi doanh nghiệp đều phải lập kế hoạch hoạt động cho tương lai. Bất kể kế hoạch ngắn hạn hay dài hạn, ban quản lý sẽ phải dự báo giá trị các biến số quan trọng trong tương lai như doanh thu, lãi suất, chi phí...

Để đưa ra dự báo, đầu tiên ta phải thiết lập mối quan hệ giữa các biến cần xét – tức biến phụ thuộc – với các biến khác ảnh hưởng đến giá trị của biến phụ thuộc – tức biến độc lập.

Trong **phân tích chuỗi thời gian (time series analysis)**, các mô hình được thiết lập sử dụng thời gian là biến độc lập. Bất cứ tập dữ liệu nào có các giá trị biến được thu thập, ghi lại hoặc quan sát trong khoảng thời gian liên tiếp (ngày, tuần, tháng, năm...) được gọi là **chuỗi thời gian (time series)**. Việc phân tích chuỗi thời gian nhằm xác định kiểu hình có trong diễn biến của biến phụ thuộc theo thời gian, từ đó dùng làm tiêu chí dự báo cho các giá trị tương lai. Quy trình sử dụng kiểu hình diễn biến của dữ liệu cũ để đoán biết diễn biến tương lai được gọi là **dự báo (forecasting)**.

Ví dụ, ta muốn lập mô hình sự biến động của doanh thu thông qua cách thức biến động trong vòng hai năm vừa qua, mà không tính đến các yếu tố khác. Nếu xác định được kiểu hình cũ của doanh thu, ta có thể sử dụng chính kiểu hình đó để dự báo các giá trị tương lai. Cần thận trọng vì kiểu hình cũ thường sẽ không còn phù hợp với yêu cầu dự báo cụ thể hiện tại. Hiện ta đang giả thiết rằng kiểu hình cũ vẫn đang tiếp diễn. **Lưu ý: đây là một giả thiết quan trọng, mà ta luôn cần tính đến tính hiệu lực.** Một công ty có thể lên kế hoạch thay đổi khi kiểu hình kinh doanh gây ra thua lỗ. Nhiều yếu tố ngoại cảnh khác cũng có thể thay đổi hoàn toàn kiểu hình; có thể là thay đổi lớn trong giá nguyên liệu thô, lạm phát toàn cầu tăng đột biến hay thiên tai ảnh hưởng đến việc kinh doanh mà không thể đoán trước. Kể cả khi không có thay đổi lớn trong các điều kiện liên quan đến chuỗi thời gian, ta cũng khó có thể xác định kiểu hình cố định, và các mô hình thu được thì vô cùng phức tạp.

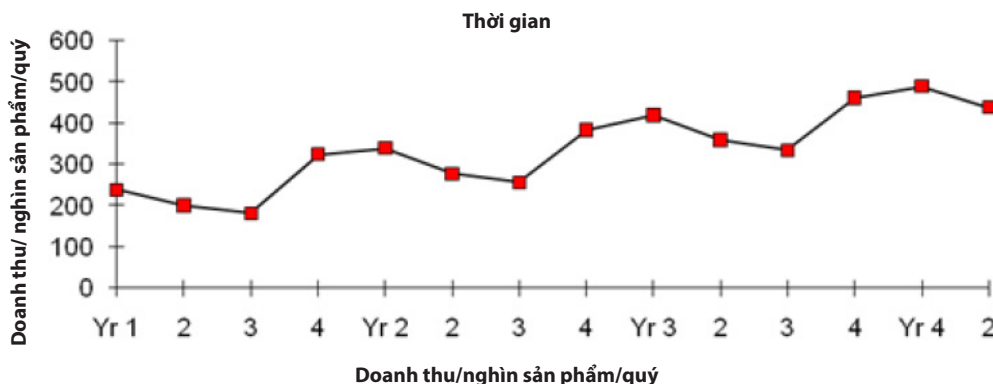
Trong **phần 7.2**, ta sẽ bắt đầu với việc mô tả chuỗi thời gian theo các yếu tố thành phần, như là:

- Xu hướng (trend)
- **Biến động theo mùa vụ** (seasonal variation)
- **Biến động theo chu kỳ** (cyclical variation)
- **Yếu tố bất thường** (irregular) hoặc **ngẫu nhiên** (random)

Ba yếu tố đầu tiên có thể được kết hợp bằng nhiều cách để tạo thành một mô hình chuỗi thời gian. Xét về khái niệm, ta không thể mô hình hóa yếu tố bất thường. Ta sẽ xét một mô hình cụ thể gọi là mô hình kết hợp yếu tố (additive component model), nghĩa là các yếu tố được gộp vào với nhau. Lưu ý rằng các kỹ thuật được mô tả không phải là phương pháp dự báo duy nhất, cũng không hẳn là tốt nhất với mọi tình huống. Có rất nhiều kỹ thuật thống kê tinh vi hơn. Ngoài kỹ thuật định lượng, ta phải sử dụng kỹ thuật định tính khi có rất ít hoặc không có dữ liệu cũ. Kỹ thuật Delphi sử dụng dự báo của giới chuyên gia để phán đoán tình hình tiếp theo và kỹ thuật viết kịch bản là vài ví dụ về kỹ thuật định tính.

## 7.2 YẾU TỐ CHUỖI THỜI GIAN

Kiểu hình chung của chuỗi thời gian sẽ dễ xác định nhất nếu biến phụ thuộc được biểu diễn tương ứng với thời gian trên biểu đồ phân phối. Người ta thường nối các điểm được biểu diễn thành đường thẳng để dễ nhìn ra kiểu hình hơn. **Hình 7.1** biểu diễn một chuỗi thời gian điển hình. Ta có thể thấy số liệu kinh doanh theo quý đang tăng dần, nhưng vẫn tồn tại một số biến động đáng kể.

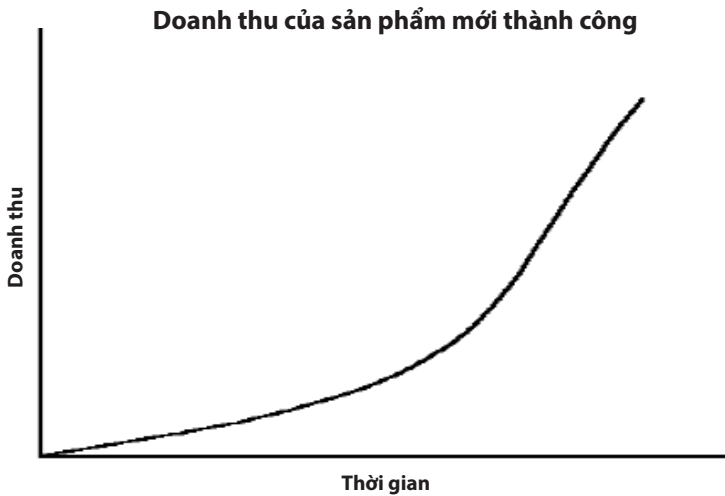


**Hình 7.1** Một chuỗi thời gian điển hình.

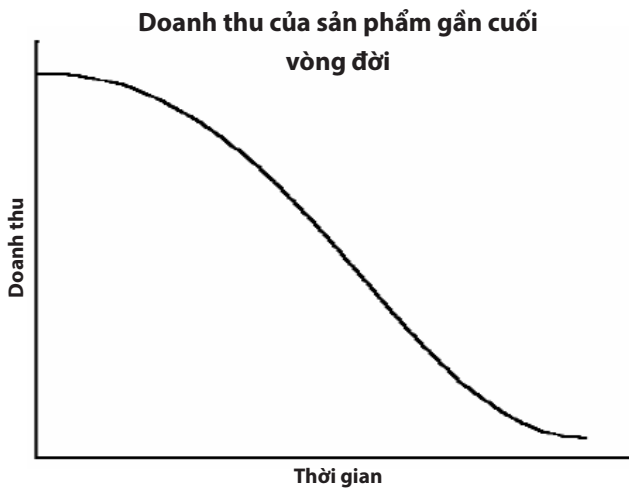
Giá trị của một biến, giả dụ doanh thu, sẽ thay đổi theo thời gian do một vài yếu tố. Ví dụ, công ty có thể mở rộng thị trường cho một sản phẩm mới, do đó doanh thu sản phẩm bán ra sẽ có biến động hướng lên theo thời gian. Thay đổi tổng quan, lâu dài trong giá trị của một biến theo thời gian được gọi là xu hướng (trend), ký hiệu là  $T$ . Các ví dụ mà ta sẽ sử dụng sau đây đều có xu hướng tuyến tính, nghĩa là mức độ thay đổi của

biến phụ thuộc tương đối cố định giữa khoảng thời gian này đến khoảng thời gian kế tiếp. Trên thực tế, có khả năng không xác định được xu hướng, khi các giá trị của biến phụ thuộc dao động trong khoảng một giá trị cố định, hoặc thường thấy hơn là một xu hướng phi tuyến tính. Chúng ta sẽ gặp khó khăn hơn trong thiết lập mô hình và dự báo xu hướng tương lai, nếu sử dụng một xu hướng phi tuyến tính.

Đồ thị dưới đây biểu diễn xu hướng trong doanh thu tại các giai đoạn vòng đời của một sản phẩm. Sản phẩm mới ra mắt thường có xu hướng đi lên, trong khi sản phẩm cũ có đường cong đi xuống cho đến khi kết thúc vòng đời. Khi đó, ta khó mà thiết lập một mô hình phù hợp với các đường cong xu hướng này.



**Hình 7.2** Vòng đời của sản phẩm mới.



**Hình 7.3** Vòng đời của sản phẩm ở cuối vòng đời.

Xem lại **Hình 7.1**, ta có thể thấy một biến động tuần hoàn rõ ràng trong các giá trị cũng như trong xu hướng đi lên. Nếu các biến động thường xuyên như vậy xảy ra trong thời gian ngắn (thường là dưới một năm), thì được gọi là **biến động theo mùa vụ (seasonal variation)**, ký hiệu là S. Biến động dài hạn được gọi là **biến động theo chu kỳ (cyclical variation)**.

**Yếu tố mùa vụ (seasonal component)** trong chuỗi thời gian được định nghĩa là “một kiểu hình dạng sóng **trong vòng một năm và tự lặp lại hàng năm**”, ví dụ biến động “hàng quý”. Tuy nhiên, thuật ngữ tác động “mùa vụ” cũng có thể được hiểu là biến động hàng tháng, hàng tuần hoặc hàng ngày. Ví dụ, trong dữ liệu hàng ngày, tác động “mùa vụ” có thể là một kiểu hình hàng ngày được lặp lại liên tục trong vòng một tuần. (Tương tự như *kiểu hình hàng ngày lặp lại trong một tuần; kiểu hình hàng tuần lặp lại trong một tháng; kiểu hình hàng tháng/hàng quý lặp lại trong một năm...*)

Trong **Hình 7.1**, khoảng thời gian chính là một năm, và ta có thể thấy doanh thu biến động theo chu kỳ trong vòng bốn quý của năm. Hàng năm, doanh thu giảm từ Quý 1 sang Quý 2, sang Quý 3, rồi tăng nhanh trở lại vào Quý 4. Kiểu hình này được lặp lại hàng năm với chiều hướng dần đi lên trong tổng doanh thu. Ví dụ, việc doanh thu sụt giảm từ Quý 2 đến Quý 3 không đáng lo, do ta biết là điều này xảy ra hàng năm. Tuy nhiên, ta nên cân nhắc nếu doanh thu Quý 3 của một năm thấp hơn cùng kỳ năm trước đó.

Nếu sử dụng dữ liệu hàng năm, chúng ta sẽ không thể xác định được kiểu hình mùa vụ. Mọi biến động về xu hướng theo dữ liệu hàng năm sẽ được mô tả bằng yếu tố tuần hoàn. Biến động tuần hoàn là kiểu hình từng bậc theo chu kỳ về xu hướng. Ta thường áp dụng kiểu hình này cho biến động trong môi trường kinh tế vĩ mô. Ta chỉ thấy được tác động của mô hình này trong dữ liệu dài hạn – 10 năm, 15 hay 20 năm. Chỉ trong khoảng thời gian dài như vậy, các yếu tố kinh tế quy mô lớn mới có đủ thời gian để biến đổi qua lại, gây ra các biến động về xu hướng. Độ dài của một chu kỳ được gọi là giai đoạn – tức thời gian giữa hai đỉnh hoặc đáy liên tiếp. Trong một tập dữ liệu hàng năm, kiểu hình tuần hoàn trong kinh tế thường có xu hướng lặp lại sau 4-5 năm. Để xác định được một yếu tố như vậy thường cần đến ít nhất ba chu kỳ dữ liệu (tức là 12 năm). Trong phần này, ta sẽ không xét đến cách xác định tác động tuần hoàn trong dữ liệu chuỗi thời gian.

Thành phần cuối cùng của chuỗi thời gian là các yếu tố **sai số (error)**, **số dư (residual)**, **giá trị bất thường/ngẫu nhiên (irregular/random)** – phần dữ liệu còn “thừa lại” của các giá trị quan sát thực tế sau khi ta đã cân nhắc đến xu hướng, biến động theo mùa vụ và biến động theo chu kỳ. Như vậy ta có:

Sai số hoặc số dư = giá trị thực – giá trị dự báo.



Số dư được sinh ra trong các tình huống hiếm gặp, ngoại lệ không thể dự đoán trước. Do đó, các giá trị này thường không thể hiện kiểu hình nào và rất nhỏ. Số dư lớn có thể cho thấy lỗi trong quá trình thu thập hoặc tính toán, vì thế cần được kiểm tra lại. Ngoài ra, số dư lớn có thể biểu thị một tình huống bất thường nào đó mà ta cần xác định. Khi đó, ta phải quyết định xem tình huống có gây ra khoảng biến động lớn trong chuỗi thời gian đến mức cần loại bỏ phần này trong tập dữ liệu hay không.

### 7.2.1 MÔ HÌNH CHUỖI THỜI GIAN

Khi phân tích chuỗi thời gian, ta cố xác định các thành phần và thiết lập mô hình kết hợp chúng lại một cách hợp lý.

Ví dụ, nếu chuỗi thời gian cho thấy kích cỡ biến động theo mùa vụ không phụ thuộc vào kích cỡ các giá trị thực, thì việc thể hiện các thành phần theo mô hình sau sẽ phù hợp:

$$A = T + S + E$$

trong đó A là giá trị quan sát thực tế, T là xu hướng, S là biến động theo mùa vụ và E là số dư hoặc sai số.

Kiểu mô hình này phù hợp nếu ta xét doanh thu bán một sản phẩm đang tăng dần theo thời gian, nhưng mức tăng doanh thu luôn lớn hơn 100 đơn vị so với xu hướng của quý 4, bất kể giá trị doanh thu đạt được là bao nhiêu. Ví dụ, nếu doanh thu là 1000 đơn vị, thì tác động mùa vụ trong quý 4 sẽ làm tăng doanh thu thành 1100 đơn vị. Nếu doanh thu đã là 10.000 đơn vị thì tác động mùa vụ như vậy sẽ làm tăng doanh thu quý 4 thành 10.100 đơn vị.

Tuy nhiên, nếu kích cỡ của biến động theo mùa vụ thay đổi nhỏ hơn xu hướng và biến động tuần hoàn, thì việc áp dụng phép nhân cho các yếu tố trong mô hình sẽ phù hợp hơn:

$$A = T S E$$

Trong trường hợp này, nếu xét đến ví dụ doanh thu bán một sản phẩm, thì kích cỡ biến động theo mùa vụ là một tỷ trọng cố định của doanh thu. Nếu doanh thu là 100 đơn vị, thì tác động mùa vụ lên quý 4 có thể làm tăng doanh thu thêm 10% thành 110 đơn vị. Nếu doanh thu ban đầu là 1000 đơn vị, thì tác động mùa vụ như vậy sẽ làm tăng doanh thu quý 4 thành 1100 đơn vị.

Hai kiểu mô hình này lần lượt được gọi là **mô hình cộng tính (additive component model)** và **mô hình nhân tính (multiplicative component model)**.

### 7.2.2 PHÂN TÍCH MÔ HÌNH CỘNG TÍNH

Khi phân tích bất kỳ chuỗi thời gian nào, bước đầu tiên là vẽ biểu đồ dữ liệu để xác định các thành phần khả thi và chọn một mô hình phù hợp.

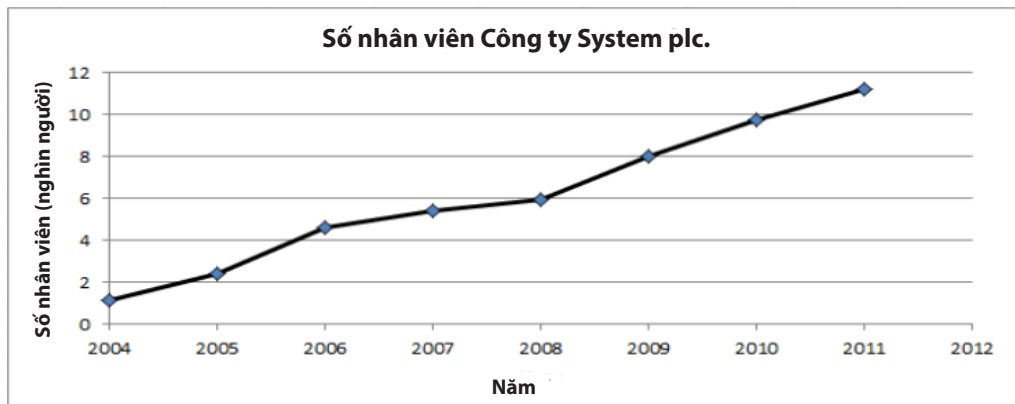
#### Ví dụ 7.1 – Phân tích chuỗi thời gian

Cho biết số nhân viên Công ty Systems plc. từ năm 2004 đến 2011 trong **Bảng 7.1**.

Năm	Số nhân viên (nghìn người)
2004	1,1
2005	2,4
2006	4,6
2007	5,4
2008	5,9
2009	8,0
2010	9,7
2011	11,2

**Bảng 7.1** Số nhân viên Công ty Systems plc. từ năm 2004 đến 2011.

**Lời giải:** Do đây là những số liệu thường niên, nên dữ liệu sẽ không mang yếu tố mùa vụ và chưa đủ để xác định bất kỳ biến động tuần hoàn nào, vì thế các thành phần cần có trong mô hình là xu hướng và số dư. Vẽ biểu đồ dữ liệu. Xem **Hình 7.4**.



**Hình 7.4** Biểu đồ chuỗi thời gian cho Ví dụ 1.

Biểu đồ phân phối trên cho thấy số nhân viên Công ty System plc. hàng năm đã tăng tương đối đều. Vì thế ta có thể sử dụng mô hình tuyến tính để mô tả xu hướng. Số dư biến động tương đối đồng đều tạo nên một mô hình khá tuyến tính, nên ta giả định mô hình chuỗi thời gian thích hợp là:

$$A = T + E$$

và xu hướng có thể được biểu diễn dưới dạng một đường thẳng. Xu hướng tuyến tính có thể được biểu diễn là:

$$A = a + bt$$

trong đó A là số nhân viên ước tính (nghìn người), t là thời gian theo năm ( $t = 1$  tại năm 2004), a và b là hằng số (b là hệ số góc của đường thẳng).

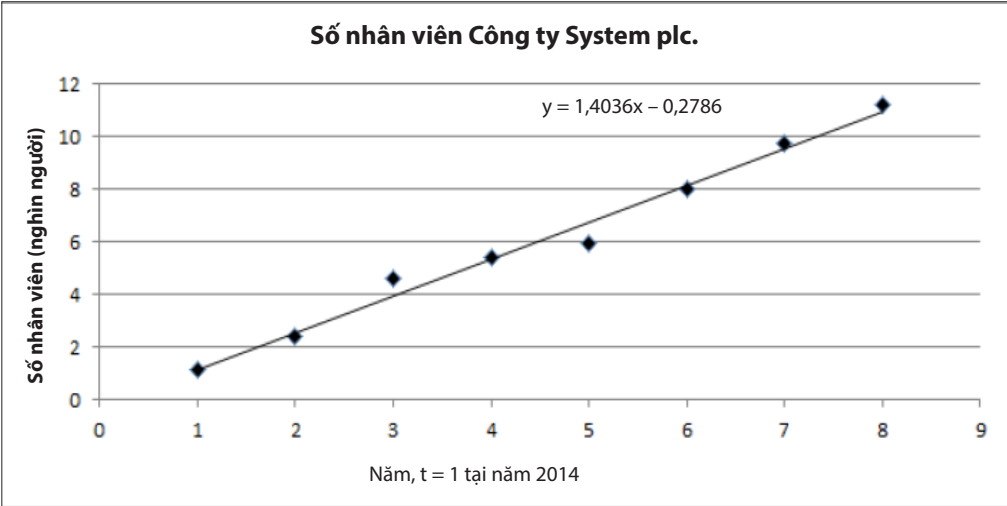
Có thể xác định giá trị của a và b thông qua công cụ Microsoft Excel để tìm đường xu hướng thích hợp cho tập dữ liệu. Phương trình đường thẳng là một tùy chọn khi thực hiện lệnh này. Sử dụng dữ liệu theo định dạng sau với năm 2004 là năm 1:

Thời gian, t (năm)	Số nhân viên (nghìn người), A
1	1,1
2	2,4
3	4,6
4	5,4
5	5,9
6	8,0
7	9,7
8	11,2

**Bảng 7.2** Số nhân viên Công ty Systems plc. theo thời gian.

thì phương trình đường xu hướng là:

$A = -0,279 + 1,404t$  (nghìn nhân viên). Xem hình dưới đây.



**Hình 7.5** Biểu đồ chuỗi thời gian cho Ví dụ 1.

Ta có thể sử dụng mô hình này để ước lượng số nhân viên mỗi năm và từ đó tính biến động số dư. Xem **Bảng 7.3**.

Biến động số dư  $E = \text{giá trị thực} - \text{giá trị dự báo}$ .

Kích thước lớn nhất của số dư là khoảng 17% với cả giá trị dương và giá trị âm. Điều này cho biết mô hình đã xác định là khá hợp lý, nhưng không hoàn toàn phù hợp với dữ liệu. Giá trị ước tính rút ra được từ mô hình này khá đáng tin cậy – chênh lệch  $\pm 17\%$ .

Thời gian, t (năm)	Số nhân viên, A (nghìn người)	Số nhân viên ước tính, $\hat{A}$ (nghìn người)	Số dư, $E = A - \hat{A}$ (nghìn người)
1	1,1	1,13	-0,03
2	2,4	2,53	-0,13
3	4,6	3,93	0,67
4	5,4	5,34	0,06
5	5,9	6,74	-0,84
6	8,0	8,14	-0,14
7	9,7	9,55	0,15
8	11,2	10,95	0,25

**Bảng 7.3** Bảng tính số nhân viên ước tính và biến động số dư.

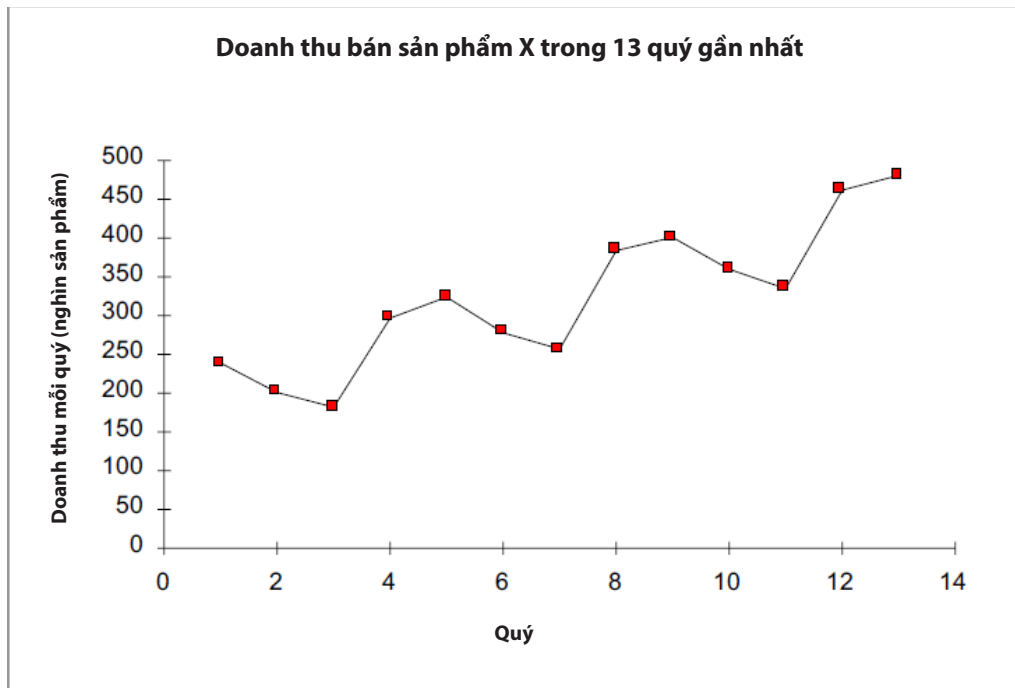
**Ví dụ 7.2 – Phân tích chuỗi thời gian**

Dữ liệu trong **Bảng 7.4** cho biết doanh thu bán sản phẩm X của Công ty Statplan plc. trong vòng 13 quý đến hết tháng 3/2014.

<b>Thời gian</b>	<b>Quý số</b>	<b>Doanh thu hàng quý (nghìn sản phẩm)</b>
Tháng 1-3/2011	1	239
Tháng 4-6	2	201
Tháng 7-9	3	182
Tháng 10-12	4	297
Tháng 1-3/2012	5	324
Tháng 4-6	6	278
Tháng 7-9	7	257
Tháng 10-12	8	384
Tháng 1-3/2013	9	401
Tháng 4-6	10	360
Tháng 7-9	11	335
Tháng 10-12	12	462
Tháng 1-3/2014	13	481

**Bảng 7.4** Doanh thu bán sản phẩm X trong 13 quý đến tháng 3/2014.

**Lời giải:** Bước đầu tiên ta cần làm là vẽ biểu đồ dữ liệu. Xem **Hình 7.6**.



**Hình 7.6** Biểu đồ chuỗi thời gian cho Ví dụ 2.

Từ **Hình 7.6** ta có thể thấy doanh thu bán hàng theo quý của Công ty Statplan plc. có xu hướng đi lên với kiểu hình mùa vụ bốn quý. Kích cỡ biến động theo mùa vụ có vẻ không bị ảnh hưởng bởi kích cỡ doanh thu bán hàng, nên mô hình cộng tính có vẻ thích hợp trong trường hợp này. Do đó, ta sẽ giả sử mô hình chuỗi thời gian thích hợp là:

$$A = T + S + E$$

Giờ ta cần phân tích dữ liệu để rút ra các thành phần mùa vụ và xu hướng.

Đối với cả mô hình cộng tính và nhân tính, quy trình chung đều giống nhau:

*Bước 1:* Tính thành phần mùa vụ.

*Bước 2:* Loại bỏ thành phần mùa vụ khỏi giá trị thực. Quá trình này gọi là san bằng tính mùa vụ (deseasonalize) của dữ liệu. Mô hình hóa xu hướng từ các số liệu đã được san bằng tính mùa vụ.

*Bước 3:* Sử dụng mô hình xu hướng để tính số liệu xu hướng dự báo, sau đó lấy số liệu đã được san bằng trừ đi số liệu xu hướng dự báo này để ra sai số hoặc số dư.

Có thể tách thành phần mùa vụ khỏi xu hướng bằng một kỹ thuật gọi là **trung bình trượt (moving averages)**. Kỹ thuật này giúp phát hiện xu hướng đã từng có bằng cách

sàng lọc loại bỏ các biến động theo mùa vụ. Tuy nhiên, người ta không dùng xu hướng trung bình trượt để dự báo các giá trị xu hướng tương lai, do tồn tại quá nhiều bất trắc liên quan tới việc kéo rộng chuỗi thời gian như vậy.

### 7.2.3 TÍNH THÀNH PHẦN MÙA VỤ CHO MÔ HÌNH CỘNG TÍNH

*Bước 1:* Ta sẽ tiếp tục với **Ví dụ 7.2** trong phần trước nói về doanh thu bán hàng theo quý của Công ty Statplan plc. Ta đã xác định rằng mô hình cộng tính là phù hợp cho tập dữ liệu này; vì vậy doanh thu thực tế có thể được mô hình hóa như sau:

$$A = T + S + E$$

Để loại bỏ thành phần mùa vụ, ta sẽ dùng phương pháp trung bình trượt.

Nếu cộng 4 điểm dữ liệu đầu tiên lại, ta được tổng doanh thu bán hàng năm 2011, chia cho 4 ta sẽ có doanh thu trung bình quý của năm 2011 như sau:

$$(239 + 201 + 182 + 297)/4 = 229,75.$$

Số liệu này không chứa thành phần mùa vụ, do ta đã lấy giá trị trung bình của cả năm. Giờ ta cần ước tính giá trị xu hướng cho giai đoạn giữa năm, nghĩa là điểm trung bình của quý 2 và quý 3. Dịch lên 3 tháng, ta sẽ tính được số liệu trung bình quý cho giai đoạn tháng 4/2011 – tháng 3/2012 (251), giai đoạn tháng 7/2011 – tháng 6/2012 (270,25)... Quy trình này cho ta giá trị trung bình trượt 4 điểm cho doanh thu bán hàng quý. Tập hợp các giá trị trung bình trượt thể hiện rõ nhất ước tính xu hướng trong nhu cầu thị trường.

Giờ ta sử dụng số liệu xu hướng để đưa ra ước tính về thành phần mùa vụ. Ta tính như sau:

$$A - T = S + E$$

Không may là, các giá trị xu hướng được ước tính thông qua kỹ thuật trung bình trượt 4 điểm là của các điểm thuộc thời gian lại khác với giá trị xu hướng của dữ liệu thực. Giá trị đầu tiên – 229,74 – biểu diễn điểm nằm giữa năm 2011, chính giữa quý 2 và quý 3. Giá trị 251 nằm giữa số liệu thực của quý 3 và quý 4. Như vậy, ta cần giá trị trung bình đã san bằng tính mùa vụ, phản ánh số liệu thực cho quý tương ứng. Vị trí của giá trị trung bình đã san bằng tính mùa vụ dịch chuyển theo cặp giá trị dùng tính trung bình. Với cặp giá trị thứ nhất và thứ hai được tính trung bình, đưa giá trị trung bình về giai đoạn quý 3/2011:

$$(229.75 + 251)/2 = 240.4$$

là giá trị trung bình đã san bằng tính mùa vụ cho quý 3/2011 (tháng 7 – tháng 9). Giá trị như vậy được gọi là trung bình trượt có tâm (centred moving average), có thể so sánh

trực tiếp với giá trị thực của quý 3/2011 là 182. Lưu ý rằng khi áp dụng phương pháp này ta sẽ không có số liệu xu hướng ước tính cho hai quý đầu tiên hoặc hai quý cuối cùng trong chuỗi thời gian. Kết quả tính toán được thể hiện ở **Bảng 7.5**.

Quý số	Doanh thu mỗi quý, A (nghìn sản phẩm)	Tổng 4 quý (nghìn sản phẩm)	Trung bình trượt 4 quý (nghìn sản phẩm)	Trung bình trượt có tâm, T (nghìn sản phẩm)	Thành phần mùa vụ ước tính $A-T=S+E$ (nghìn sản phẩm)
1	239		-	-	
2	201		-	-	
		919	229,75		
3	182			240,4	-58,4
		1004	251,00		
4	297			260,6	+36,4
		1081	270,25		
5	324			279,6	+44,4
		1156	289,00		
6	278			299,9	-21,9
		1243	310,75		
7	257			320,4	-63,4
		1320	330,00		
8	384			340,3	+43,8
		1402	350,50		
9	401			360,2	+40,8
		1480	370,00		
10	360			379,8	-19,8
		1558	389,50		
11	335			399,5	-64,5
		1638	409,50		
12	462	-	-	-	
13	481	-	-	-	

**Bảng 7.5** Tính xu hướng trung bình trượt 4 điểm có tâm theo mô hình  $A = T + S + E$ .



Có thể sử dụng Microsoft Excel để tính ví dụ này như sau:

	A	B	C	D	E	F
1	Quý số	Doanh thu mỗi quý, A (nghìn sản phẩm)	Tổng 4 quý (nghìn sản phẩm)	Trung bình trượt 4 quý (nghìn sản phẩm)	Trung bình trượt có tâm, T (nghìn sản phẩm)	Thành phần mùa vụ ước tính $A-T=S+E$ (nghìn sản phẩm)
2	1	239		-	-	
3	2	201		-	-	
4			= sum(B2:B7) (919)	= C4/4 (229,75)		
5	3	182			= (D4+D6)/2 (240,4)	=B5-E5 (-58,4)
6			= sum(B3:B9) (1004)	= C6/4 (251,00)		
7	4	297			= (D6+D8)/2 (260,6)	=B7-E7 (+36,4)
8			= sum(B5:B11) (1081)	= C8/4 (270,25)		
9	5	324			= (D8+D10)/2 (279,6)	=B9-E9 (+44,4)
10			= sum(B7:B13) (1156)	= C10/4 (289,00)		
11	6	278			= (D10+D12)/2 (299,9)	=B11-E11 (-21,9)
12			= sum(B9:B15) (1243)	= C12/4 (310,75)		
13	7	257			= (D12+D14)/2 (320,4)	=B13-E13 (-63,4)
14			= sum(B11:B17) (1320)	= C14/4 (330,00)		

	A	B	C	D	E	F
1	Quý số	Doanh thu mỗi quý, A (nghìn sản phẩm)	Tổng 4 quý (nghìn sản phẩm)	Trung bình trượt 4 quý (nghìn sản phẩm)	Trung bình trượt có tâm, T (nghìn sản phẩm)	Thành phần mùa vụ ước tính $A-T=S+E$ (nghìn sản phẩm)
15	8	384			= (D14+D16)/2 (340,3)	=B15-E15 (+43,8)
16			= sum(B13:B19) (1402)	= C16/4 (350,50)		
17	9	401			= (D16+D18)/2 (360,2)	=B17-E17 (+40,8)
18			= sum(B15:B21) (1480)	= C18/4 (370,00)		
19	10	360			= (D18+D20)/2 (379,8)	=B19-E19 (-19,8)
20			= sum(B17:B23) (1558)	= C20/4 (389,50)		
21	11	335			= (D20+D22)/2 (399,5)	=B21-E21 (-64,5)
22			= sum(B19:B24) (1638)	= C22/4 (409,50)		
23	12	462	-	-	-	
24	13	481	-	-	-	

**Bảng 7.6** Ví dụ sử dụng Microsoft Excel để tính các chỉ số.

Dữ liệu trong Bảng 7.6 là dữ liệu từ Bảng 7.5

Ô C1:F24 chứa hàm cần nhập và giá trị sẽ tính ra tại ô này (trong ngoặc đơn).

Như vậy ta đã có giá trị mùa vụ ước tính của mỗi quý trong năm, nhưng vẫn còn chứa sai số và số dư. Ta còn xử lý hai bước tính toán nữa để có được giá trị thành phần mùa vụ hữu dụng.

Đầu tiên, ta tính trung bình các giá trị mùa vụ ước tính trong **Bảng 7.5** cho mỗi mùa. Quá trình này sẽ giúp loại bớt một vài sai số. Cuối cùng, ta điều chỉnh giá trị trung bình bằng cách thêm hoặc bớt cùng một số cho đến khi tổng của chúng bằng 0. (Chú ý: Các yếu tố mùa vụ luôn phải có tổng bằng 0). Cần thực hiện bước này vì yêu cầu của chúng ta là phải cân bằng các giá trị thành phần mùa vụ trong suốt cả năm, nghĩa là tổng tác động mùa vụ nên bằng 0. Quy trình thực hiện được trình bày trong **Bảng 7.7**.

Năm	Quý trong năm				
	1	2	3	4	
2011	-	-	-58,4	+36,4	
2012	+44,4	-21,9	-63,4	+43,8	Tổng = -0,25
2013	+40,8	-19,8	-64,5	-	Điều chỉnh = +0.0625
Tổng	+85,2	-41,7	-186,3	+80,2	
Trung bình	+85,22 = +42,6	-41,72 = -20,85	-186,33 = -62,1	+80,22 = +40,1	Tổng = 0 (đã tính làm tròn)
Thành phần mùa vụ ước tính	+42,6	-20,85	-62,1	+40,1	(điều chỉnh sao cho tổng thành phần mùa vụ bằng 0)
Thành phần mùa vụ đã điều chỉnh	+42,7	-20,8	-62,0	+40,2	

**Bảng 7.7** Tính thành phần mùa vụ trung bình.

Thành phần mùa vụ ước tính là:

tháng 1-3	+42700 sản phẩm	tháng 4-6	- 20800 sản phẩm
tháng 7-9	- 62000 sản phẩm	tháng 10-12	+40200 sản phẩm

#### 7.2.4 SAN BẰNG TÍNH MÙA VỤ DỮ LIỆU ĐỂ XÁC ĐỊNH XU HƯỚNG

*Bước 2:* San bằng tính mùa vụ dữ liệu cơ bản. Quy trình được biểu diễn như ở **Bảng 7.7** đã giảm trừ yếu tố mùa vụ thích hợp từ doanh thu thực mỗi quý theo công thức sau:

$$A - S = T + E$$

Các giá trị xu hướng ước tính lại chứa biến động số dư (sai số) có thể được sử dụng để thiết lập mô hình xác định xu hướng ẩn. Giá trị này được thêm vào biểu đồ phân phối ban đầu. Xem **Hình 7.8**.

Thời gian	Quý số	Doanh thu mỗi quý, A (nghìn sản phẩm)	Thành phần mùa vụ, S (nghìn sản phẩm)	Doanh thu đã san bằng tính mùa vụ (nghìn sản phẩm), $A-S=T+E$
Tháng 1-3/2011	1	239	+42,7	196,3
Tháng 4-6	2	201	-20,8	221,8
Tháng 7-9	3	182	-62,0	244,0
Tháng 10-12	4	297	+40,2	256,8
Tháng 1-3/2012	5	324	+42,7	281,3
Tháng 4-6	6	278	-20,8	298,8
Tháng 7-9	7	257	-62,0	319,0
Tháng 10-12	8	384	+40,2	343,8
Tháng 1-3/2013	9	401	+42,7	358,3
Tháng 4-6	10	360	-20,8	380,8
Tháng 7-9	11	335	-62,0	397,0
Tháng 10-12	12	462	+40,2	421,8
Tháng 1-3/2014	13	481	+42,7	438,3

**Bảng 7.8** Tính dữ liệu đã san bằng tính mùa vụ.



**Hình 7.7** Doanh thu thực tế và doanh thu đã san bằng tính mùa vụ cho Ví dụ 2.

Doanh thu đã san bằng tính mùa vụ cho ta thấy xu hướng gần như tuyến tính. Ta có thể thiết lập mô hình tuyến tính cho xu hướng này. Phương trình của đường xu hướng là:

$$T = a + b \times \text{số quý}$$

trong đó a và b lần lượt là hệ số chặn và hệ số góc của đường thẳng.

Áp dụng vào phương trình đường xu hướng tuyến tính (như trong Ví dụ 7.1), ta tìm được:

$$b = 19,973 \text{ và } a = 180,038$$

Do đó phương trình của mô hình xu hướng có thể viết như sau:

$$\text{Xu hướng doanh thu (nghìn sản phẩm)} = 180,0 + 20,0 \times \text{số quý}$$

### 7.2.5 TÍNH BIẾN ĐỘNG SỐ DƯ

*Bước 3:* Tính sai số hoặc số dư. Mô hình cần áp dụng là:

$$A = T + S + E$$

Ta đã tính được S và T trong các phần trước. Giờ ta có thể lấy doanh thu thực tế A trừ đi các kết quả đó để tìm số dư trong mô hình. Xem **Bảng 7.9**.

Thời gian	Quý số	Doanh thu mỗi quý, A (nghìn sản phẩm)	Thành phần mùa vụ, S (nghìn sản phẩm)	Thành phần xu hướng, T (nghìn sản phẩm) (từ mô hình)	Số dư, E (nghìn sản phẩm) $A - S - T = E$
<b>Tháng 1-3/2011</b>	1	239	+42,7	200	-3,7
<b>Tháng 4-6</b>	2	201	-20,8	220	+1,8
<b>Tháng 7-9</b>	3	182	-62,0	240	+4,0
<b>Tháng 10-12</b>	4	297	+40,2	260	-3,2
<b>Tháng 1-3/2012</b>	5	324	+42,7	280	+1,3
<b>Tháng 4-6</b>	6	278	-20,8	300	-1,2
<b>Tháng 7-9</b>	7	257	-62,0	320	-1,0
<b>Tháng 10-12</b>	8	384	+40,2	340	+3,8
<b>Tháng 1-3/2013</b>	9	401	+42,7	360	-1,7
<b>Tháng 4-6</b>	10	360	-20,8	380	+0,8
<b>Tháng 7-9</b>	11	335	-62,0	400	-3,0

Thời gian	Quý số	Doanh thu mỗi quý, A (nghìn sản phẩm)	Thành phần mùa vụ, S (nghìn sản phẩm)	Thành phần xu hướng, T (nghìn sản phẩm) (từ mô hình)	Số dư, E (nghìn sản phẩm) $A - S - T = E$
<b>Tháng 10-12</b>	12	462	+40,2	420	+1,8
<b>Tháng 1-3/2014</b>	13	481	+42,7	440	-1,7

**Bảng 7.9** Tính số dư cho mô hình thành phần cộng tính.

Tất cả số dư đều nhỏ, chỉ khoảng 1 đến 2% doanh thu thực tế. Điều này cho thấy kiểu hình cũ có tính ổn định cao và có thể đưa ra được một dự báo ngắn hạn.

Lưu ý là tổng các thành phần mùa vụ đều bằng 0:

$$\text{nghĩa là } 42,7 + -20,8 + -62,0 + 40,2 = 0$$

### 7.2.6 DỰ BÁO SỬ DỤNG MÔ HÌNH CỘNG TÍNH

Dự báo cho mô hình cộng tính được thiết lập ở **Ví dụ 7.2** là:

$$F = T + S \text{ nghìn sản phẩm mỗi quý}$$

Khi thành phần xu hướng  $T = 180 + 20 \times \text{số quý}$  (nghìn sản phẩm), và thành phần mùa vụ S là:

$$\begin{array}{ll} +42,6 \text{ cho tháng 1-3} & -20,7 \text{ cho tháng 4-6} \\ -62,0 \text{ cho tháng 7-9} & +40,1 \text{ cho tháng 10-12} \end{array}$$

Số quý cho giai đoạn 3 tháng tiếp theo (tháng 4-6/2014) là 14, cho nên xu hướng dự báo là:

$$T_{14} = 180 + 20 \times 14 = 460 \text{ (nghìn sản phẩm mỗi quý)}$$

Thành phần mùa vụ tương ứng là -20,7 (nghìn sản phẩm). Do đó giá trị dự báo cho quý này là:

$$F \text{ (tháng 4-6/2014)} = 460 - 20,7 = 439,3 \text{ (nghìn sản phẩm)}$$

Cần nhớ rằng dự báo cho thời điểm càng xa thì tính tin cậy càng thấp. Chúng ta đang giả sử là kiểu hình cũ vẫn được duy trì và không bị ngắt đoạn. Giả định này có thể đúng trong ngắn hạn, nhưng càng về sau thì sẽ càng ít chính xác hơn.

### 7.3 ĐÁNH GIÁ VÀ SO SÁNH CÁC MÔ HÌNH

Nhiệm vụ chính của người sử dụng mô hình dự báo (hay bất cứ mô hình nào khác) là phải đánh giá mô hình. Khi đó, ta cần cân nhắc đến vấn đề: “Mô hình này phù hợp đến đâu, nếu ta dùng để đánh giá những gì đã xảy ra?”

Chúng ta thường dùng nhiều loại mô hình dự báo khác nhau trong cùng một tình huống cụ thể. Lúc này vấn đề nảy sinh là: “Mô hình nào có tác dụng tốt nhất?”, và “Mô hình nào ta nên áp dụng cho các dự báo tương lai?”

Chúng ta sẽ sử dụng nhiều phép **thống kê sai số (error statistics)** để so sánh các mô hình. Đó là các phương pháp tóm lược để xác định xem mô hình dự báo chính xác đến đâu so với thực tế đã xảy ra.

Ta sẽ tính trung bình các sai số có thể xảy ra. Trong phần này, ta xét:

**Sai số trung bình (mean error, ME)**

$$ME = \frac{\text{tổng các sai số}}{n}$$

**Trung bình độ lệch tuyệt đối (mean absolute deviation, MAD)**

$$MAD = \frac{\text{tổng các sai số không xét đến dấu}}{n}$$

**Căn bậc hai của sai số bình phương trung bình (root mean square error, RMSE)**

$$RMSE = \sqrt{\frac{\text{tổng các (sai số)}^2}{n}}$$

trong đó  $n$  là số khoảng thời gian ta vừa quan sát vừa dự báo các giá trị.

Tính căn bậc hai của sai số bình phương trung bình (RMSE) là phương pháp đo lường độ lệch giữa các giá trị dự đoán dựa vào mô hình và giá trị quan sát trên tình huống thực tế. RMSE là phương pháp đánh giá có tính chính xác khá cao; các giá trị chấp nhận được phụ thuộc vào mục đích thiết lập mô hình và tình huống đang được mô hình hóa.

Ví dụ, trong trường hợp mô hình hóa phản ứng của xe cứu thương thuộc Dịch vụ Y tế Anh quốc (*National health service, NHS*), bạn cần một mô hình càng chính xác càng tốt và tỷ lệ sai số chấp nhận được sẽ rất nhỏ, cùng lắm là 2%. Mặt khác, nếu mô hình hóa phong độ thi đấu tương lai của đội bóng yêu thích thì tỷ lệ sai số chấp nhận được có thể cao hơn, đến 5 hay 10%. Nhìn chung, giá trị chấp nhận được phụ thuộc vào tình huống đang được mô hình hóa và mức độ chính xác mà bạn chấp nhận.

## 7.4 TÀI LIỆU THAM KHẢO CHƯƠNG

1. Hanke, JE, Reitsch, AG, Business Forecasting (*tạm dịch*: Dự báo kinh tế), tái bản lần thứ 9, Prentice Hall, 2015, Chương 4.
2. Oakshott, L, Essential Quantitative Methods for Business (*tạm dịch*: Phương pháp nghiên cứu định lượng thiết yếu trong kinh doanh, quản lý & tài chính), tái bản lần thứ 5, Palgrave Macmillan, 2012, Chương 4.
3. Wisniewski, M, Quantitative Methods for Decision Makers (*tạm dịch*: Phương pháp nghiên cứu định lượng phục vụ ra quyết định), tái bản lần thứ 5, FT/Prentice Hall, 2009 (có bản eBook), Chương 3.



## Chương 8

# MÔ HÌNH HÓA TÀI CHÍNH

### 8.1 GIỚI THIỆU

Giờ chúng ta sẽ tập trung vào một số khía cạnh căn bản về Mô hình hóa tài chính (Financial Modelling), giúp ta hiểu rõ hơn môi trường tài chính hiện đại.

Trong chương này, ta sẽ củng cố thêm kiến thức và kỹ năng mô hình hóa thông qua các biến tài chính, cũng như xây dựng giải pháp cho các vấn đề có liên quan đến các biến kinh doanh đơn giản và phức tạp.

### 8.2 CHI PHÍ, DOANH THU, LỢI NHUẬN VÀ SỐ DƯ ĐẢM PHÍ

#### 8.2.1 CHI PHÍ

Chi phí (*cost*) phát sinh trong mọi tổ chức dưới nhiều hình thức. Nhìn chung, chúng ta thường dành thời gian và công sức để giảm thiểu chi phí có thể xảy ra. Ít nhất, chúng ta phải đảm bảo là các chi phí được giải trình rõ ràng. Trong Kế toán kinh doanh, chúng ta sẽ thảo luận về cách thức, thời điểm phát sinh cũng như ghi nhận chi phí. Trong Phân tích kinh doanh, chúng ta sẽ tập trung vào việc mô hình hóa mối liên hệ giữa một vài loại chi phí với các biến kinh doanh khác.

Đa số chi phí có thể được chia làm hai phần: phần cố định (*fixed*) và phần biến đổi (*variable*). Chi phí sản xuất hoặc bán một sản phẩm sẽ bao gồm:

- **chi phí cố định** (*fixed cost*): chi phí phát sinh bất kể số lượng sản phẩm mà bạn sản xuất hay bán ra – bao gồm chi phí mặt bằng, hạ tầng, nhân công...
- **chi phí biến đổi** (*variable cost*): chi phí chỉ phát sinh khi bạn sản xuất hoặc bán ra một sản phẩm – bao gồm chi phí nguyên vật liệu thô, đóng gói...

Do đó, có thể lập công thức đơn giản cho tổng chi phí (*total cost*), viết tắt là TC:

$$TC = \text{tổng chi phí cố định} + \text{tổng chi phí biến đổi}$$

Nếu ta sản xuất được  $q$  sản phẩm/ngày với chi phí  $v\text{£}$  mỗi sản phẩm, thì tổng chi phí biến đổi mỗi ngày sẽ là  $v\text{£}xq$ .

$$TC = \text{tổng chi phí cố định} + vxq$$

Nếu ta sản xuất được  $q$  sản phẩm/ngày với chi phí  $2\text{£}$  mỗi sản phẩm, thì tổng chi phí biến đổi mỗi ngày sẽ là  $2\text{£}xq$ .

Nếu chi phí cố định cho việc sản xuất là  $300\text{£}/\text{ngày}$ , thì tổng chi phí (TC) mỗi ngày có thể được mô tả thông qua mô hình sau:

$$TC = 300 + 2q \quad \text{£/ngày}$$

Ta có thể dùng mô hình trên để tính tổng chi phí mỗi ngày phát sinh trong nhiều cấp độ sản xuất.

### 8.2.2 DOANH THU

Như ta đã biết, doanh thu (*revenue*) là tổng số tiền thu về từ việc bán sản phẩm hay dịch vụ. Ở mức độ đơn giản nhất, doanh thu phụ thuộc vào:

số sản phẩm bán được,  $q$

giá bán mỗi sản phẩm,  $p$

Từ đó ta có thể mô hình hóa doanh thu (R) như sau:

$$R = \text{giá bán mỗi sản phẩm} \times \text{số sản phẩm bán được}$$

$$R = p \times q \quad \text{£/ngày}$$

trong đó  $p$  là giá mỗi sản phẩm tính theo  $\text{£}$  và  $q$  là số lượng sản phẩm bán được mỗi ngày.

Nếu ta bán được  $q$  sản phẩm/ngày với giá  $5\text{£}$  mỗi sản phẩm, thì tổng doanh thu mỗi ngày là  $5\text{£}xq$ .

Mô hình doanh thu như sau:

$$R = 5xq \quad \text{£/ngày.}$$

Ta có thể dùng mô hình này để tính tổng doanh thu mỗi ngày ở nhiều cấp độ bán hàng.

### 8.2.3 LỢI NHUẬN

Mục đích của nhiều tổ chức là tạo ra lợi nhuận (*profit*). Kể cả các tổ chức phi lợi nhuận cũng thường yêu cầu ít nhất phải hòa vốn.

Ở cấp độ đơn giản nhất, lợi nhuận có thể được tính là chênh lệch từ doanh thu trừ đi **tất cả** chi phí (ở cấp độ này, ta không xét đến các vấn đề liên quan đến thuế).

Do đó, mô hình đơn giản cho lợi nhuận (P) là:

$$P = \text{Tổng doanh thu} - \text{Tổng chi phí} = R - TC$$

$$P = p \times q_s - (\text{chi phí cố định} + \text{chi phí biến đổi})$$

$$P = p \times q_s - (\text{chi phí cố định} + v \times q_p)$$

trong đó  $q_s$  là số lượng bán ra và  $q_p$  là số lượng sản xuất trong một giai đoạn cho trước.

Ở trường hợp đơn giản nhất thì  $q_s = q_p$ , nghĩa là ta bán được toàn bộ sản phẩm đã sản xuất.

Nếu ta sản xuất được  $q$  sản phẩm/ngày với giá 2£ mỗi sản phẩm, và bán toàn bộ với giá 5£ mỗi sản phẩm, thì mô hình lợi nhuận tổng thể là:

$$P = R - TC = 5 \times q - (FC + 2 \times q) \quad \text{£/ngày}$$

(FC viết tắt cho tổng chi phí cố định mỗi ngày)

$$P = (5 - 2) \times q - FC = 3 \times q - FC \quad \text{£/ngày}$$

Ta có thể dùng mô hình này để tính tổng lợi nhuận mỗi ngày thu được từ nhiều cấp độ sản xuất và bán hàng.

#### 8.2.4 SỐ DƯ ĐẢM PHÍ

Về cơ bản, số dư đảm phí (*contribution*) là phần chênh lệch giữa doanh thu và tổng chi phí **biến đổi (variable)**. Thông qua các hoạt động thương mại, số dư đảm phí được sử dụng để bù đắp chi phí cố định và tạo ra lợi nhuận (nếu cần thiết).

Như vậy, mô hình đơn giản cho số dư đảm phí là:

Số dư đảm phí = tổng doanh thu – tổng chi phí biến đổi, do đó:

$$\text{Số dư đảm phí} = p \times q_s - v \times q_p$$

trong đó  $q_s$  là số lượng bán ra và  $q_p$  là số lượng sản xuất trong một giai đoạn cho trước.

Ở trường hợp đơn giản nhất thì  $q_s = q_p$ , nghĩa là ta bán được toàn bộ sản phẩm đã sản xuất.

Nếu thu mua  $q$  sản phẩm/ngày với giá 2£ mỗi sản phẩm, rồi bán *tất cả* với giá 5£ một sản phẩm, thì ta có mô hình tổng số dư đảm phí như sau:

Số dư đảm phí =  $R - \text{tổng chi phí biến đổi} = 5 \times q - 2 \times q$       £/ngày

$$\text{Số dư đảm phí} = (5 - 2) \times q = 3 \times q \quad \text{£/ngày}$$

Ta có thể dùng mô hình này để tính tổng số dư đảm phí mỗi ngày tại nhiều cấp độ mua hàng/sản xuất và bán hàng.

### 8.3 MINH HỌA MÔ HÌNH TÀI CHÍNH BẰNG ĐỒ THỊ

Với nhiều cấp độ sản xuất và bán hàng, việc nắm được thông tin về chi phí, doanh thu, lợi nhuận và số dư đảm phí rõ ràng là có ích. Ví dụ, có thể ta muốn biết số sản phẩm phải bán được để đạt đến điểm hòa vốn. Hay ta có thể muốn xác định sản lượng sẽ tối đa hóa được lợi nhuận, hoặc giảm thiểu chi phí trung bình trên mỗi sản phẩm...

Nếu giả sử ta có thể bán hết mọi sản phẩm thu mua về hay sản xuất, thì cách đơn giản để khảo sát chi phí, doanh thu, lợi nhuận và số dư đảm phí là vẽ đồ thị.

Chúng ta cùng xét ví dụ dưới đây.

#### 8.3.1 MÔ HÌNH TUYẾN TÍNH ĐƠN GIẢN

##### Ví dụ 8.1 – Vẽ đồ thị phương trình tuyến tính

Công ty Dickson Mills plc. sản xuất nhiều loại vải. Chi phí sản xuất vải cotton poplin có phương trình sau đây:

$$C = 500 + 2q \quad (\text{£/ngày})$$

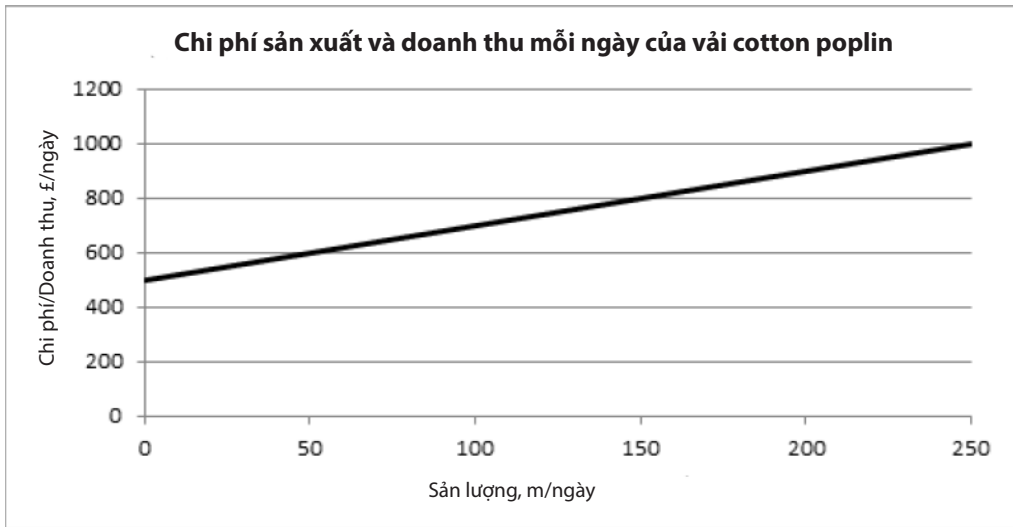
trong đó  $q$  là số mét vải sản xuất mỗi ngày. Phạm vi sản xuất mỗi ngày dao động từ 0 đến 200m.

**Lời giải:**  $q$  là biến độc lập và có phạm vi được xác định – từ 0 đến 200.

Vậy chi phí cố định và chi phí biến đổi đơn vị của sản phẩm này là gì?

(Lưu ý: Môn học không đòi hỏi tìm ra bằng chứng, tuy nhiên nếu muốn bạn có thể hỏi giáo viên hướng dẫn.)

Sử dụng Microsoft Excel, ta vẽ được đồ thị như ở **Hình 8.1**.



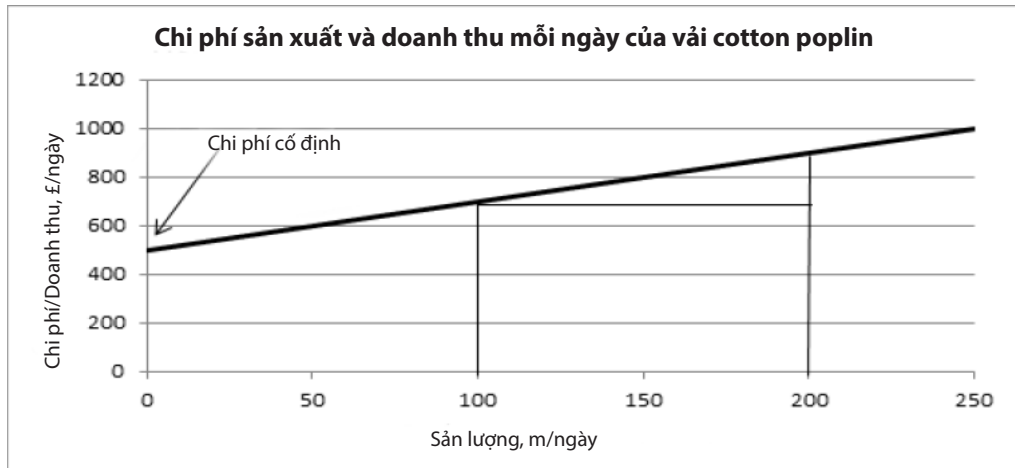
**Hình 8.1** Chi phí sản xuất của Công ty Dickson Mills plc.

- Làm sao để ta xác định giá trị của chi phí cố định từ đồ thị trên?
- Làm sao để ta tìm giá trị của chi phí biến đổi đơn vị từ đồ thị?

$$\text{Nếu } C = 500 + 2q \quad (\text{£/ngày})$$

thì  $C = 500$  (chi phí cố định) khi  $q = 0$ .

Như vậy, nếu không sản xuất sản phẩm nào thì tổng chi phí sẽ bằng chi phí cố định. Do đó chi phí cố định được xác định tại điểm đường thẳng cắt trục tổng chi phí. Chi phí biến đổi đơn vị là chi phí sản xuất 1m vải, nghĩa là mức tăng trong tổng chi phí khi sản xuất thêm mỗi mét – tương đương với hệ số góc của đường thẳng. Như vậy, hệ số góc của đường thẳng chính là chi phí biến đổi đơn vị của quá trình sản xuất.



$$\text{Hệ số góc} = 200/100 = 2$$

**Hình 8.2** Chi phí sản xuất của Công ty Dickson Mills plc.

Có thể sử dụng đồ thị tuyến tính cho nhiều mục đích khác nhau trong việc ra quyết định kinh doanh. Đa phần người sử dụng sẽ thấy đồ thị dễ dùng hơn so với phương trình đường thẳng, dù kết quả là giống nhau. Việc thể hiện số liệu bằng hình ảnh thường có ích trong việc tìm hiểu kết cấu của vấn đề.

Trở lại **Ví dụ 1**, ta có thể sử dụng đồ thị để xác định các giá trị kiểu như chi phí sản xuất 75m vải hay sản lượng thu được với chi phí 600£.

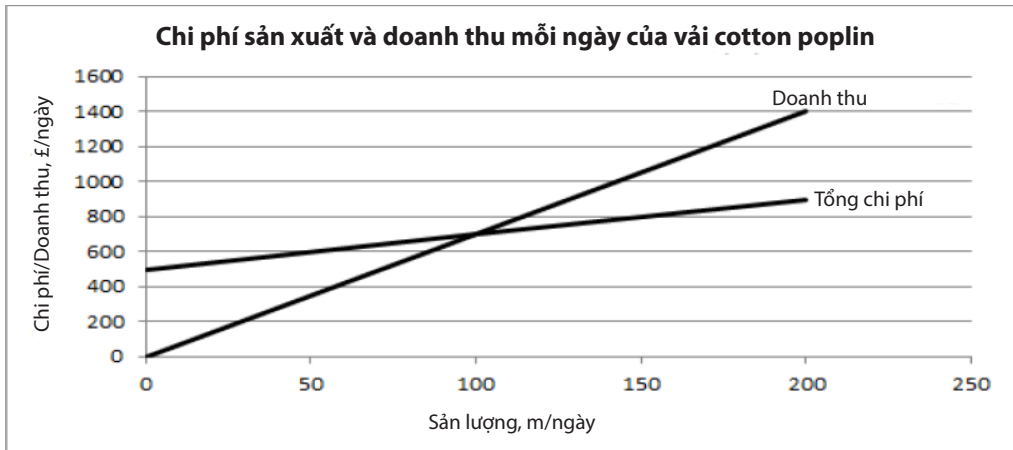
### 8.3.2 ĐỒ THỊ ĐIỂM HÒA VỐN

Xét lại **Ví dụ 8.1**, nếu biết số vải được bán với giá 7£/m, ta có thể sử dụng đồ thị để xác định sản lượng cần thiết để Công ty Dickson **đạt điểm hòa vốn (break-even)**, nghĩa là tại đó toàn bộ chi phí được bù đắp và có thể bắt đầu tạo ra lợi nhuận. Giả sử toàn bộ hàng sản xuất ra đều được bán hết.

Doanh thu R từ số vải được xác định như sau:

$$R = \text{giá/m} \times \text{số m vải bán mỗi ngày} = 7q \text{ (£/ngày)}$$

Đồ thị này là một đường thẳng. Đặt đường thẳng này vào đồ thị trong **Hình 8.2**. Hãy xem **Hình 8.3**.



**Hình 8.3** Chi phí sản xuất và doanh thu của Công ty Dickson Mills plc.

Chúng ta sẽ chỉ ra thời điểm công ty đạt điểm hòa vốn của sản phẩm này bằng cách nào?

Điểm hòa vốn là giao điểm của hai đường thẳng, khi:

$$\text{Doanh thu} = \text{Tổng chi phí}$$

$$\text{Lợi nhuận} = 0$$

Trong ví dụ này, điểm hòa vốn xảy ra khi  $q = 100\text{m/ngày}$ . Do đó, Công ty Dicksons đạt điểm hòa vốn khi họ sản xuất và bán được 100m vải mỗi ngày. Nếu sản xuất và bán được trên 100m thì họ sẽ có lãi, ngược lại dưới 100m thì sẽ bị lỗ.

Đồ thị điểm hòa vốn (*break-even chart*) thường được lập với chi phí cố định tách riêng.

### Ví dụ 8.2 – Đồ thị điểm hòa vốn

Xét lại Ví dụ 8.1. Phương trình tính chi phí của Công ty Dickson Mills plc. là:

$$C = 500 + 2q \quad \text{£/ngày}$$

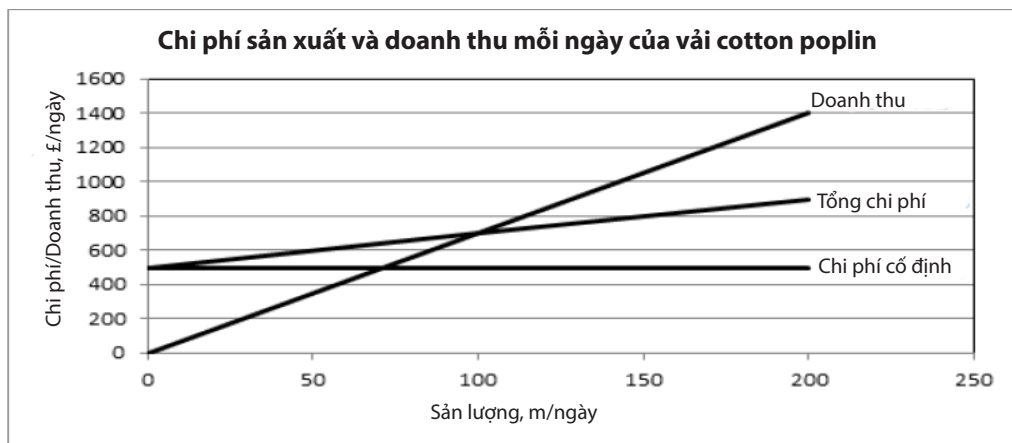
Chi phí này bao gồm một thành phần cố định:

$$FC = 500\text{£/ngày}$$

và một thành phần biến đổi:

$$VC = 2q\text{£/ngày}.$$

Việc vẽ đồ thị cho tổng chi phí, chi phí cố định và doanh thu giúp ta có được đồ thị điểm hòa vốn như trong **Hình 8.4**.



**Hình 8.4** Chi phí sản xuất và doanh thu của Công ty Dickson Mills plc.

Mở rộng đồ thị điểm hòa vốn, ta có thể minh họa được số dư đảm phí.

**Đồ thị số dư đảm phí (contribution chart)** thể hiện đường doanh thu, đường tổng chi phí và đường chi phí biến đổi. Như đã nhắc đến ở trên, điểm hòa vốn là giao điểm giữa đường doanh thu và đường tổng chi phí. Đường tổng chi phí và đường chi phí biến đổi song song với nhau, mà khoảng cách giữa chúng là chi phí cố định. Khoảng cách giữa đường doanh thu và đường chi phí biến đổi là số dư đảm phí.

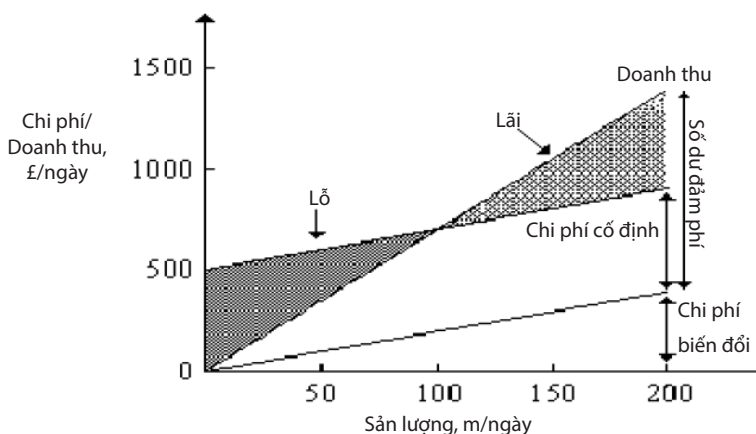
### Ví dụ 8.3 – Đồ thị số dư đảm phí

Xét lại **Ví dụ 8.1**. Vẽ đường doanh thu và đường tổng chi phí như trước, rồi thêm đường chi phí biến đổi, ta được **Hình 8.5**.

Điểm hòa vốn là điểm mà số dư đảm phí bằng chi phí cố định, tức tại điểm  $q = 100$ .

Do đó, như đã phân tích ở trên, Công ty Dicksons hòa vốn khi họ sản xuất và bán được 100m vải mỗi ngày.





**Hình 8.5** Đồ thị số dư đảm phí của Công ty Dickson Mills plc.

Mở rộng đến cùng đồ thị này, ta có **đồ thị lợi nhuận-sản lượng (profit-volume)**. Loại đồ thị này thường được sử dụng trong công tác quản trị, vì thể hiện mức độ biến động lợi nhuận căn bản tương ứng với từng giai đoạn sản xuất.

#### **Ví dụ 8.4 – Đồ thị lợi nhuận-sản lượng**

Lợi nhuận sẽ như thế nào tại điểm hòa vốn?

Nếu điểm hòa vốn xảy ra khi doanh thu = tổng chi phí, nghĩa là tổng doanh thu vừa đủ bù đắp tất cả chi phí phát sinh, khi đó:

$$\text{Doanh thu} = \text{Tổng chi phí}$$

$$\text{Doanh thu} - \text{Tổng chi phí} = 0$$

Lợi nhuận = 0, nghĩa là không phát sinh lợi nhuận.

Xác định điểm không phát sinh lợi nhuận là cách thay thế cho việc xác định điểm hòa vốn. Ta xét lại **Ví dụ 8.1**. Phương trình lợi nhuận là:

$$\text{Lợi nhuận } P = \text{Doanh thu} - \text{Tổng chi phí}$$

$$P = 7q - (500 + 2q) \quad \text{£/ngày}$$

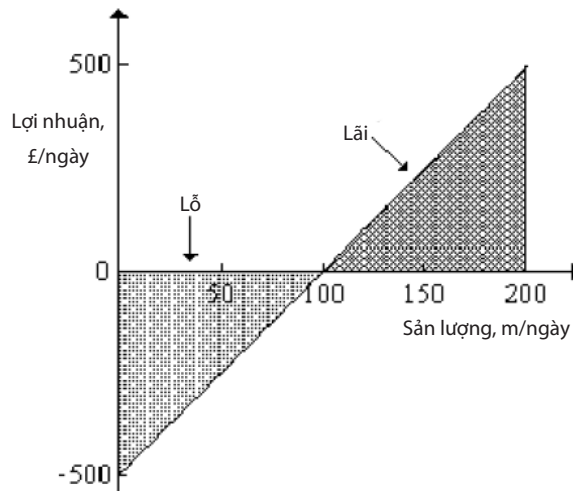
$$P = 7q - 500 - 2q \quad \text{£/ngày}$$

$$P = 5q - 500 \quad \text{£/ngày}$$

Hãy vẽ đường lợi nhuận. Xem **Hình 8.6**.

Khi sản lượng sản xuất bằng 0, công ty lỗ 500£/ngày, nghĩa là chi phí cố định vẫn phải chi trả nhưng không có doanh thu.

Khi sản lượng sản xuất bằng 100, lợi nhuận bằng 0, tại điểm hòa vốn với  $q = 100\text{m}/\text{ngày}$ . Khi  $q > 100$  thì công ty bắt đầu thu về lợi nhuận.



**Hình 8.6** Đồ thị lợi nhuận-sản lượng của Công ty Dickson Mills plc.

## 8.4 LIÊN HỆ PHI TUYẾN TÍNH

Rõ ràng không phải mọi mối liên hệ kinh doanh đều là tuyến tính. Khi đó, việc mô hình hóa một mối liên hệ kinh doanh nhanh chóng trở nên phức tạp. Trong các phần tiếp theo, ta sẽ tìm hiểu một số ý tưởng cơ bản.

### 8.4.1 CHI PHÍ

Trong các ví dụ trước, ta đã giả sử chi phí biến đổi đơn vị luôn cố định với bất kể sản lượng nào. Trên thực tế thường không như vậy. Trong sản xuất thực tế, chi phí biến đổi đơn vị thường giảm khi sản lượng tăng – theo kinh tế quy mô. Tác động này có thể đưa yếu tố thứ ba vào mô hình chi phí, khiến tổng chi phí giảm khi sản lượng tăng. Chúng ta sẽ xét một yếu tố như vậy trong ví dụ dưới đây.

Xét lại **Ví dụ 1** (Công ty Dickson Mills plc.). Chi phí sản xuất vải cotton poplin có mô hình:

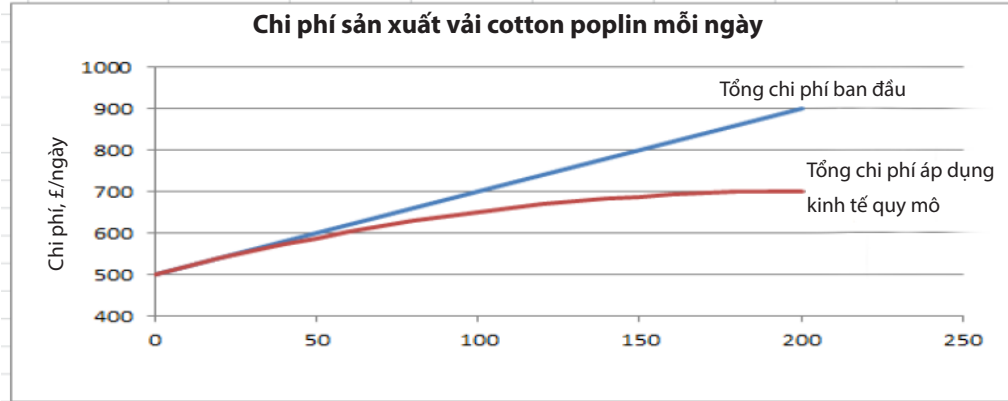
$$C = 500 + 2q \quad (\text{£/ngày})$$

trong đó  $q$  là sản lượng theo ngày (m/ngày). Áp dụng kinh tế quy mô, ta có thể điều chỉnh mô hình này như sau:

$$C = 500 + 2q - 0,005q^2 \quad (\text{£/ngày})$$

Ban đầu, ở mức sản lượng thấp,  $q^2$  có ảnh hưởng rất ít đến tổng chi phí, nhưng khi  $q$  tăng lên thì ngày càng làm giảm chi phí nhiều hơn.

Mức độ ảnh hưởng được thể hiện trong **Hình 8.7**.



**Hình 8.7** Sản xuất tại Công ty Dickson Mills plc.

#### 8.4.2 DOANH THU

Trong các ví dụ trước, ta đã giả sử là giá đơn vị là cố định và không phụ thuộc vào khối lượng bán hàng. Trên thực tế, giá bán luôn phụ thuộc vào nhu cầu sản phẩm. Hoặc ngược lại, khối lượng sản phẩm bán được luôn phụ thuộc vào giá bán.

Giả sử Công ty Dickson Mills phát hiện ra nhu cầu mua vải ( $q$ ) và giá bán ( $p$ ) liên hệ với nhau theo công thức sau:

$$q = 700 - 100p \text{ (m/ngày)}$$

hoặc ngược lại:

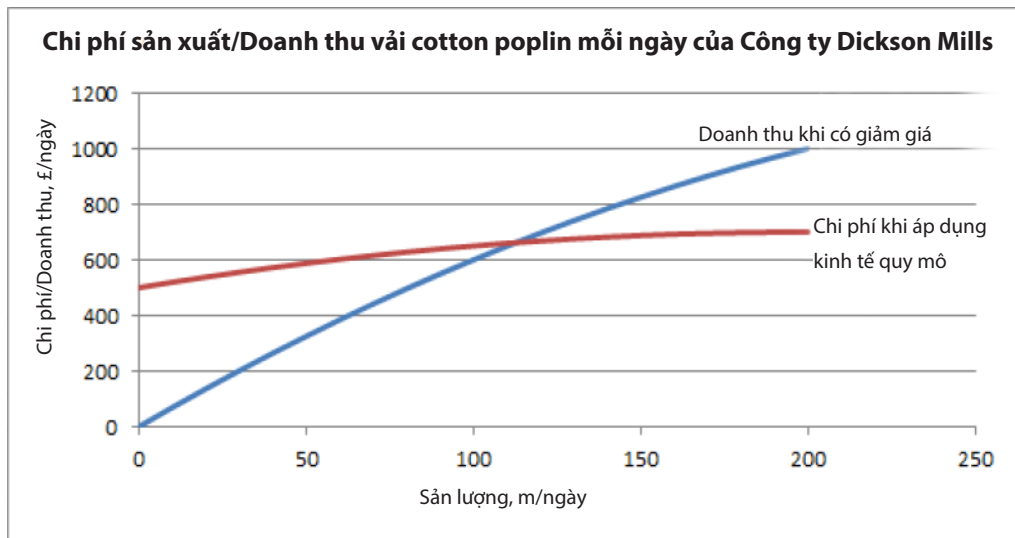
$$p = 7 - 0,01q \text{ (£/ngày)}.$$

Khi đó doanh thu ( $R$ ) là:

$$\begin{aligned} R &= pxq = (7 - 0,01q) \times q \\ &= 7q - 0,01q^2 \text{ (£/ngày)} \end{aligned}$$

$$\text{hoặc } R = pxq = px(700 - 100p) = 700p - 100p^2.$$

Thêm mô hình doanh thu này vào đồ thị chi phí mới, ta có **Hình 8.8**.



**Hình 8.8** Thêm mô hình doanh thu vào đồ thị chi phí mới.

### 8.4.3 LỢI NHUẬN

Giả sử tại Công ty Dickson Mills plc., tổng chi phí sản xuất (TC) là tuyến tính nhưng giá bán phụ thuộc vào nhu cầu mua hàng. Khi đó:

$$TC = 500 + 2q \text{ £/ngày}$$

$$R = 7q - 0,01q^2 \quad \text{£/ngày}$$

Khi đó lợi nhuận (P) là:  $P = R - TC$

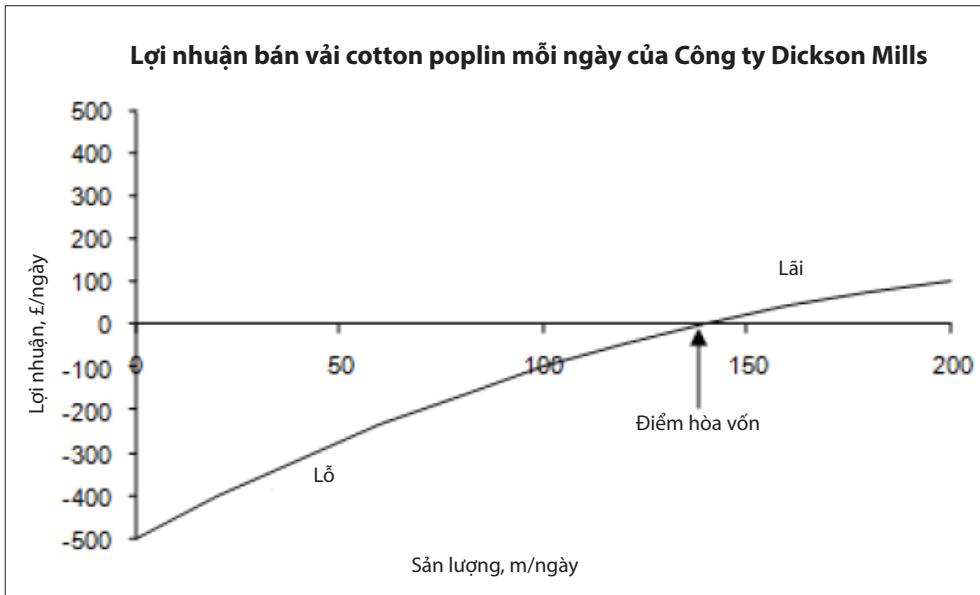
$$P = 7q - 0,01q^2 - (500 + 2q)$$

$$P = 7q - 0,01q^2 - 500 - 2q$$

$$P = 5q - 0,01q^2 - 500 \quad \text{£/ngày}$$

Giờ vẽ đồ thị lợi nhuận tương ứng với lượng hàng sản xuất và bán ra, ta sẽ thấy được biến động của lợi nhuận. Xem **Hình 8.9**.

Như đã thấy, với các giá trị sản lượng và doanh số nhỏ, công ty bị lỗ. Tuy nhiên, mức độ thua lỗ giảm dần khi sản lượng và doanh số tăng lên.



**Hình 8.9** Lợi nhuận tương ứng với lượng hàng sản xuất và bán ra.

Lợi nhuận/lỗ bằng 0 khi công ty sản xuất và bán được khoảng 140m/ngày. Trên mức này công ty sẽ có lãi.

### Ví dụ 8.5 – Tối ưu hóa lợi nhuận

Công ty Dickson Mills còn sản xuất một loại vải có hàm lượng cotton cao khác. Sử dụng thông tin do bộ phận kế toán và marketing cung cấp, ta lập được mô hình tổng chi phí (TC) và doanh thu (R) cho phạm vi sản xuất từ 0 – 200m/ngày như sau:

$$TC = 100 + 3q \quad \text{£/ngày}$$

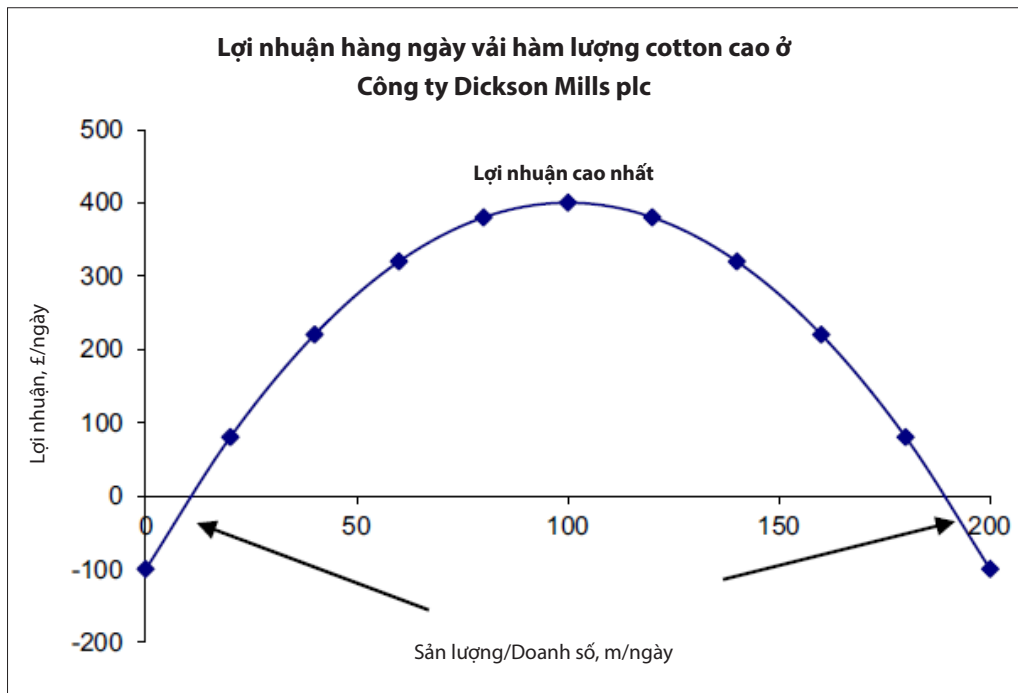
$$R = 13q - 0,05q^2 \quad \text{£/ngày}$$

Khi đó lợi nhuận hàng ngày là:

$$P = R - TC = 13q - 0,05q^2 - (100 + 3q)$$

$$P = 10q - 0,05q^2 - 100 \quad (\text{£/ngày})$$

Xem **Hình 8.10**.



**Hình 8.10** Lợi nhuận hàng ngày.

Ta có thể thấy ở mức sản lượng cao và thấp thì đều lỗ, nhưng trong phần lớn phạm vi đã cho thì công ty thu được lợi nhuận.

Mức lợi nhuận cao nhất thu được khi sản lượng và doanh số là 100m/ngày.

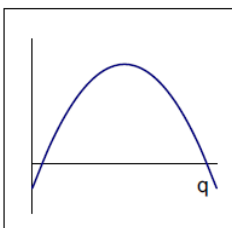
Ta cũng có thể thấy là công ty đạt điểm hòa vốn với sản phẩm này ở 2 mức sản lượng là 10,5m/ngày và 189,5m/ngày. Giữa hai giá trị này, công ty sẽ có lãi; còn ngoài khoảng này thì công ty sẽ bị lỗ.

Lưu ý: Chỉ trong dạng mô hình này thì điểm lợi nhuận cao nhất mới nằm ở chính giữa hai điểm hòa vốn.

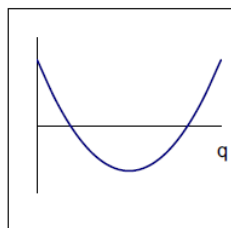
Nói chung, nếu đồ thị được dựng từ công thức có dạng

$$aq^2 + bq + c \quad \text{trong đó } a, b, c \text{ là số}$$

thì



sẽ có dạng này (khi  $a < 0$ )



hoặc dạng này  
(khi  $a > 0$ ).

Điểm quay đầu (tức điểm cực đại hoặc điểm cực tiểu) sẽ nằm ở giữa hai giao điểm của đường cong với trục  $q$ .

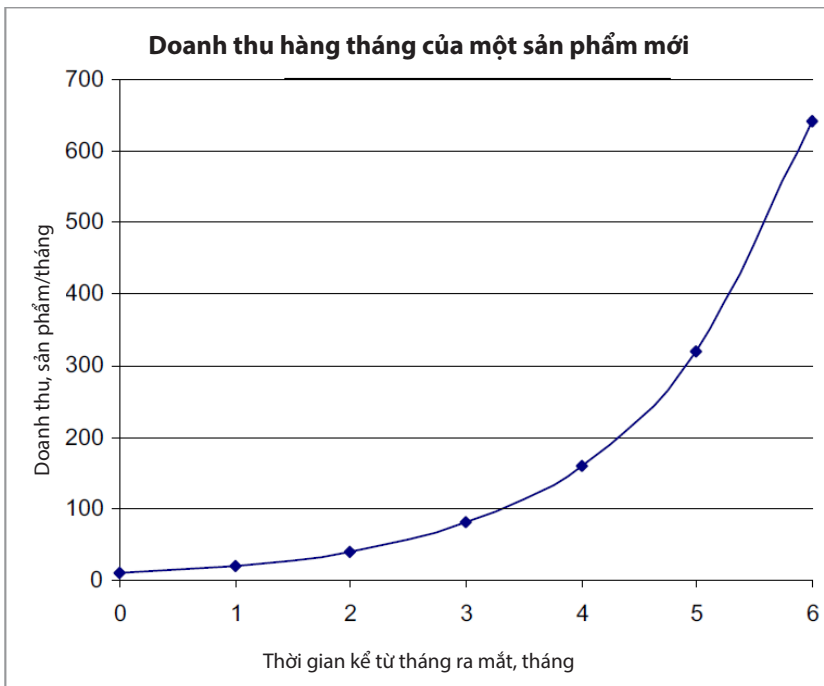
## 8.5 CÁC KIỂU LIÊN HỆ HỮU DỤNG KHÁC

Có rất nhiều kiểu mô hình diễn ra trong các bối cảnh kinh doanh. Trong phần này, ta sẽ chỉ xét đến một kiểu mô hình thôi.

Giả sử một công ty ra mắt thành công một sản phẩm mới và nhận thấy doanh thu của sản phẩm này mỗi tháng đều tăng gấp đôi. Khi đó mỗi liên hệ giữa doanh thu ( $S$ ) và thời gian có thể được mô hình hóa thành:

$S = 10 \times 2^t$  trong đó  $S$  là lượng hàng bán ra trong tháng  $t$

Xem **Hình 8.11**.



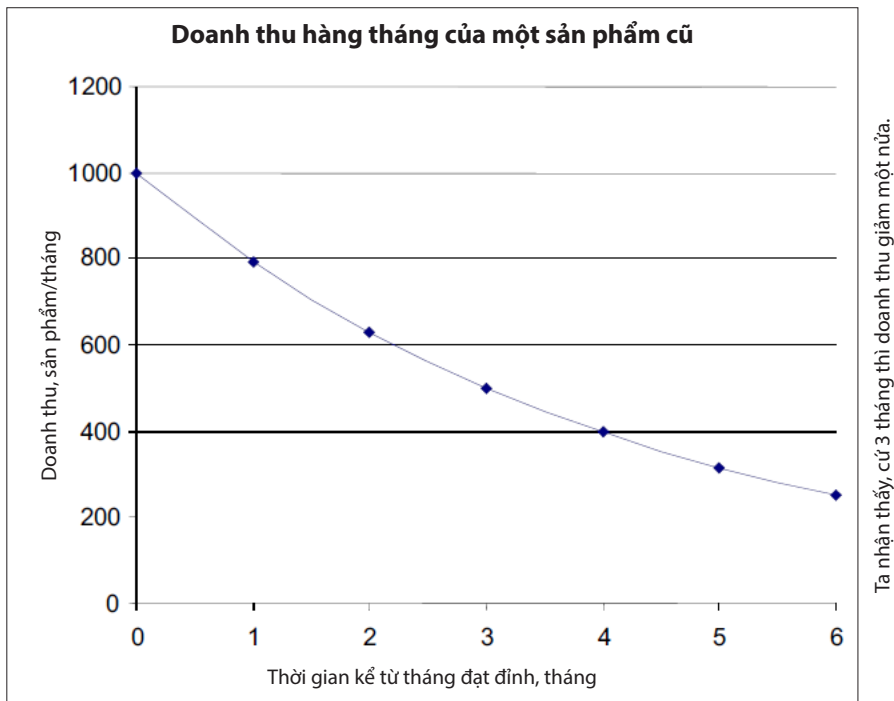
**Hình 8.11** Mối liên hệ giữa doanh thu và thời gian bán hàng.

Bạn có thể thấy số lượng sản phẩm bán được tăng gấp đôi mỗi tháng.

Giả sử một sản phẩm khác đã kết thúc vòng đời bán hàng và doanh thu ( $S$ ) giảm dần theo mô hình:

$S = 1000 \times 2^{-t/3}$  trong đó  $S$  là lượng sản phẩm bán được trong tháng  $t$ .

$t = 0$  khi doanh thu đạt đỉnh. Xem **Hình 8.12**.



**Hình 8.12** Doanh thu của một sản phẩm cũ.

Kiểu liên hệ này được gọi là liên hệ **hàm mũ (exponential)**, mà người ta có thể sử dụng khi các biến phụ thuộc tăng gấp đôi hoặc giảm một nửa tại các khoảng đều nhau.

## 8.6 TÀI LIỆU THAM KHẢO CHƯƠNG

1. Atrill, P. và McLaney, E., Accounting and Finance for Non-Specialists (*tạm dịch*: Kế toán và tài chính cho người không chuyên), tái bản lần thứ 5, FT/Prentice Hall, 2006, Chương 7.
2. Oakshott, L, Essential Quantitative Methods for Business (*tạm dịch*: Phương pháp nghiên cứu định lượng thiết yếu trong kinh doanh, quản lý & tài chính), tái bản lần thứ 5, Palgrave Macmillan, 2012.
3. Wisniewski, M, Quantitative Methods for Decision Makers (*tạm dịch*: Phương pháp nghiên cứu định lượng phục vụ ra quyết định), tái bản lần thứ 5, FT/Prentice Hall, 2009 (có bản eBook).



## Chương 9

# PHÂN TÍCH TÀI CHÍNH TIỀN TỆ - LÃI SUẤT VÀ GIÁ TRỊ THỜI GIAN

### 9.1 GIỚI THIỆU

Bà dì giàu có đã quyết định tặng bạn một món quà là 1000£, với điều kiện bạn không được tiêu số tiền này trong vòng 3 năm. Ngay lúc này, bạn sẽ làm gì với món tiền?

Có rất nhiều khả năng. Với tấm séc này, bạn có thể:

- i) cất trong một chiếc két sắt và giấu dưới nệm giường;
- ii) tiết kiệm qua tài khoản thanh toán thông thường ở ngân hàng hoặc hiệp hội xây dựng;
- iii) tiết kiệm qua tài khoản tiền gửi ở ngân hàng, hoặc tiền gửi lãi suất cao tại hiệp hội xây dựng;
- iv) đầu tư vào chứng khoán và cổ phiếu, tài khoản tiết kiệm cá nhân miễn thuế (ISA), quỹ ủy thác đầu tư... (dĩ của bạn coi đây là hành vi “tiêu tiền”).

Bạn nghĩ hướng hành động nào là hợp lý nhất?

Là nhà lãnh đạo doanh nghiệp trong tương lai, hy vọng bạn đã chọn cách giữ tiền sao cho tiền vẫn sinh **lãi suất (interest)** – và chọn cân bằng giữa tỷ suất lãi có thể đạt được với rủi ro liên quan.

Không chịu thua kém, ông chú của bạn cũng tuyên bố tặng bạn 1000£ nhưng phải đến ngày sinh nhật của bạn vào 3 năm tới thì mới trao. Giả sử là một người nhay bén, liệu ông chú của bạn phải thực sự có sẵn 1000£ lúc này để có thể tặng món tiền đó vào 3 năm sau không? Giờ nếu đầu tư 800£ với lãi suất 10% mỗi năm thôi, thì ông chú của bạn sẽ thu về nhiều hơn số 1000£ cần có cho 3 năm sau.

Tuy nhiên, đứng trên quan điểm cá nhân, liệu lợi ích bạn thu được từ 1000£ nhận vào 3 năm tới có tương đương lợi ích bạn thu được từ 1000£ nhận ngay bây giờ? Với 1000£ nhận vào 3 năm sau, liệu bạn có mua được nhiều thứ như với 1000£ trong tay lúc này? Nói cách khác, liệu giá trị của số tiền có giữ nguyên không? Khả năng cao là “không”. Trong tương lai, số tiền đó sẽ có sức mua nhỏ hơn so với bây giờ.

Hai tình huống này minh họa cho hai khía cạnh về tiền tệ mà mọi nhà quản lý đều phải cân nhắc:

- a) khả năng tiền sinh **lãi** (interest),
- b) **giá trị thời gian** (time value) của tiền.

Sự tăng hoặc giảm giá trị của một khoản tiền là yếu tố quan trọng trong nhiều hoạt động kinh doanh. Điều này ảnh hưởng đến mọi khía cạnh quản lý, chứ không chỉ mỗi chức năng kế toán. Giá trị thực của một tài sản hoặc một khoản tiền, nghĩa là sức mua của tiền, sẽ thay đổi theo thời gian. Xin nhắc lại đây là một yếu tố vô cùng quan trọng trong kinh doanh. Việc không cân nhắc đến “**giá trị thời gian**” (**time value**) của tiền là nguyên nhân gây ra nhiều vấn đề trong một tổ chức.

Trong kế toán, ta thường phải cân nhắc đến giá trị của tài sản hoặc nhiều khoản tiền tại những thời điểm khác nhau. Ví dụ, nếu ký một hợp đồng thuê tài sản, bạn sẽ trả 20£/tháng trong vòng 2 năm tới. Ở thời điểm hiện tại, bạn sẽ thấy là giá trị của khoản trả phí thuê cuối cùng sẽ khác với giá trị của khoản trả phí thuê đầu tiên.

Giờ ta sẽ tìm hiểu chi tiết hơn về hai khía cạnh của tiền – lãi và giá trị thời gian.

**Lưu ý:** Trong các ví dụ dưới đây, ta coi tất cả phép tính đều được Microsoft Excel thực hiện một cách dễ dàng; trừ một số rất ít (nếu có) được thực hiện thủ công.

## 9.2 LÃI SUẤT

Một khoản tiền được đầu tư sẽ sinh lãi bằng nhiều cách. Ta sẽ xét đến 3 trong số những cách đó. Trong mỗi trường hợp, ta giả dụ đầu tư 200£ trong vòng 3 năm với lãi suất 10%/năm.

$A_0$  – khoản tiền đầu tư ban đầu

$A_n$  – khoản tiền thu được sau khoảng thời gian  $n$  (thường là số năm)

$r$  – lãi suất hàng năm, %

### 9.2.1 LÃI KÉP

Lãi kép (*compound interest*) phát sinh khi lãi suất sinh ra được gộp vào tổng số vốn đầu tư; và tổng số tiền mới thu được lại là khoản vốn để tính lãi suất cho giai đoạn kế tiếp.

Đây là cách tính lãi phổ biến nhất trên thực tế. Như đã nói, không có nhiều trường hợp sử dụng lãi đơn (*simple interest*) để tính lãi suất. Trong các ví dụ dưới đây, ta sẽ coi lãi suất là lãi kép, trừ trường hợp được chú thích khác.

**Ví dụ 9.1 – Đầu tư với lãi kép**

200£ được đầu tư với lãi kép 10%/năm. Ta sẽ thu được bao nhiêu tiền khi kết thúc kỳ hạn 3 năm?

**Lời giải:** Công thức chung để tính số tiền thu được sau  $n$  năm là:

Đầu năm 1,

$$A_0 = 200\text{£}$$

Cuối năm 1, lãi suất là 10% của 200£, tức 20£ được thêm vào, ta có  $(200 + 20)$

$$A_1 = 220\text{£}$$

Cuối năm 2, lãi suất là 10% của 220£, tức 22£ được thêm vào, ta có  $(220 + 22)$

$$A_2 = 242\text{£}$$

Cuối năm 3, lãi suất là 10% của 242£, tức 24,2£ được thêm vào, ta có  $(242 + 24,2)$

$$A_3 = 266,20\text{£}$$

Do đó tổng số tiền thu được sau 3 năm là 266,20£.

**9.2.2 TÍNH LÃI KÉP**

Dựa vào chuỗi tính đã thực hiện ở Ví dụ 1, nhìn chung ta thấy số tiền thu được sau  $n$  năm với lãi kép  $r\%/năm$  là:

$$A_n = A_0 (1 + r\%)^n$$

Mô hình này được gọi là “Công thức lãi kép” (Compound Interest Formula). Ta có thể thiết lập mô hình này một cách khá đơn giản trong Microsoft Excel.

	<b>A</b>	<b>B</b>	<b>C</b>	<b>D</b>
<b>1</b>	<b>Cuối năm</b>	<b>Số tiền thu được (£)</b>	<b>Lãi suất hàng năm =</b>	
<b>2</b>	<b>0</b>			
<b>3</b>	<b>1</b>	=B2*(1+\$D\$1)		
<b>4</b>	<b>2</b>	=B3*(1+\$D\$1)		
<b>5</b>	<b>3</b>	=B4*(1+\$D\$1)		
<b>6</b>	<b>...</b>	<b>...</b>		

**Bảng 9.1** Tính lãi kép.

Nhập khoản tiền đầu tư ban đầu vào ô B2 và lãi suất hàng năm dưới dạng phần trăm vào ô D1.

Từ đó, công thức được nhập ở ô B3 sẽ tính ra số tiền thu được vào cuối năm đầu tiên. Giá trị thu được sẽ là số tiền đầu tư ban đầu cho năm thứ 2...

Bảng trên được thiết lập dựa vào chuỗi tính đã sử dụng ở **Ví dụ 9.1**.

Ngoài ra, ta có thể lập bảng như sau:

	<b>A</b>	<b>B</b>	<b>C</b>	<b>D</b>
<b>1</b>	<b>Cuối năm</b>	<b>Số tiền thu được (£)</b>	<b>Lãi suất hàng năm =</b>	
<b>2</b>	<b>0</b>			
<b>3</b>	<b>1</b>	$=B\$2*(1+\$D\$1)^{A3}$		
<b>4</b>	<b>2</b>	$=B\$2*(1+\$D\$1)^{A4}$		
<b>5</b>	<b>3</b>	$=B\$2*(1+\$D\$1)^{A5}$		
<b>6</b>	<b>...</b>	<b>...</b>		

**Bảng 9.2** Bảng tính lãi kép khác.

Nhập số tiền đầu tư ban đầu vào ô B2 và lãi suất hàng năm dưới dạng phần trăm vào ô D1 như trên.

Bảng này được thiết lập dựa vào “Công thức lãi kép” chung.

Hãy xem kỹ cả 2 mô hình và đảm bảo rằng bạn hiểu rõ cách tính ở cả 2 mô hình này.

### 9.2.3 LÃI SUẤT DANH NGHĨA VÀ LÃI SUẤT THỰC

Khi tính lãi tại thời điểm ngắn hạn hơn so với thời hạn trong lãi suất niêm yết, ta cần biết lãi niêm yết là lãi suất **danh nghĩa (nominal)** hay **thực tế (actual)**.

Ví dụ, với một loại tài khoản nhất định tại hiệp hội xây dựng, lãi suất được tính cho tài khoản theo tháng. Lãi niêm yết **danh nghĩa** là 6%/năm.

Như vậy có nghĩa là tiền gửi sinh lãi ở mức  $6/12 = 0,5\%$ /tháng.

Thực tế, số tiền thu được trong cả năm sẽ lớn hơn so với 6% của tổng số tiền gửi đầu năm.

Điều này rất quan trọng trong tình huống đi vay tiền, do số nợ phải trả tích lũy có thể lớn hơn nhiều so với dự kiến,

Luật quy định mọi tổ chức có hoạt động trả hoặc thu lãi đều phải niêm yết **Lãi suất tương đương theo năm (Annual Equivalent Rate, viết tắt là AER)** được áp dụng, nghĩa là khoản tăng phần trăm **thực tế (actual)** của tổng số tiền theo năm.

Nếu lãi suất **danh nghĩa** là  $r\%/năm$ , tính lãi theo quý, thì tổng số tiền đầu tư sẽ tăng  $(r/4)\%/quý$ .

Tuy nhiên, nếu lãi suất **thực tế** là  $r\%/năm$ , tính lãi theo quý, thì tổng số tiền đầu tư sẽ tăng  $i\%/quý$  với  $i$  được xác định theo công thức:

$$(1 + i\%)^4 = 1 + r\%$$

### Ví dụ 9.2 – Lãi suất danh nghĩa và lãi suất thực tế

Đầu tư 100£ với lãi suất  $10\%/năm$

- i) lãi kép;
- ii) lãi kép, tính lãi theo quý.

Tổng số tiền thu được sau 5 năm là bao nhiêu?

**Lời giải:**

- i) Lãi kép:  $A_5 = A_0(1 + 10\%)^5 = A_0(1 + 0,1)^5$   
Do đó  $A_5 = 100(1,1)^5 = 100 \times 1,61051 = 161,051$

Như vậy tổng số tiền thu được sau 5 năm là 161,05£.

Sử dụng cách khác, ta có thể thiết lập chuỗi tính bằng Microsoft Excel:

	A	B	C	D
1	Cuối năm	Số tiền thu được (£)	Lãi suất hàng năm =	10%
2	0	100		
3	1	=B2*(1+\$D\$1)		
4	2	=B3*(1+\$D\$1)		
5	3	=B4*(1+\$D\$1)		
6	4	=B5*(1+\$D\$1)		
7	5	=B6*(1+\$D\$1)		

**Bảng 9.3** Tính lãi kép.

Chuỗi tính này cho kết quả như sau:

	A	B	C	D
1	Cuối năm	Số tiền thu được (£)	Lãi suất hàng năm =	10%
2	0	100,00		
3	1	110,00		
4	2	121,00		
5	3	133,10		
6	4	146,41		
7	5	161,05		

**Bảng 9.4** Giá trị tính lãi kép.

ii) Lãi kép, tính lãi theo quý:

- a) Giả sử lãi suất **danh nghĩa** là 10%/năm, thì lãi mỗi quý sẽ là  $(10/4) = 2,5\%$ .

Có 4 đợt trả lãi mỗi năm, nghĩa là 20 đợt trả lãi trong 5 năm.

$$\text{Do đó, } A_5 = 100(1 + 2,5\%)^{20} = 100(1,025)^{20} = 163,86$$

Tổng số tiền thu được sau 5 năm là 163,86£; cao hơn khi ta sử dụng cùng mức lãi suất danh nghĩa nhưng thanh toán mỗi năm một lần.

Một lần nữa, ta cũng có thể thiết lập chuỗi tính bằng Microsoft Excel – giá trị nhập vào ở ô D1 ở trên đổi thành 2,5% và công thức được điền tự động sao cho giá trị ở cột A là 20.

- b) Giả sử lãi suất **thực tế** là 10%/năm, thì tổng số tiền thu được sau 5 năm sẽ giống như ở trường hợp i) là 161,05£; tuy nhiên lãi suất được tính vào tài khoản cuối mỗi quý thay vì mỗi năm. Điều này sẽ có lợi hơn cho nhà đầu tư, vì trong năm họ sẽ có nhiều thời điểm rút tiền hơn mà không lo mất lãi tại thời điểm rút.

Lãi suất theo quý  $i\%$  được tính như sau:

$$(1 + i\%)^4 = (1 + 10\%) = 1,10$$

$$(1 + i\%) = \sqrt[4]{1,10} = 1,02411 \text{ (làm tròn đến 5 chữ số thập phân)}$$

$$\text{Do đó } i/100 = 0,02411$$

Như vậy, lãi suất theo quý là 2,41% (làm tròn đến 2 chữ số thập phân).

## 9.3 ỨNG DỤNG TÍNH LÃI SUẤT

### 9.3.1 LÃI SUẤT TƯƠNG ĐƯƠNG THEO NĂM (AER)

Lãi suất tương đương theo năm (*Annual Equivalent Rate, AER*) có tên gọi cũ là Lãi suất phần trăm hàng năm (*Annual Percentage Rate, APR*).

Khái niệm AER đã được đề cập đến trong Phần 9.2.3. Chúng ta sẽ làm rõ hơn về tầm quan trọng của giá trị này.

Trong quảng cáo cho vay tiền mua tài sản giá trị như xe ô tô, trong tờ rơi ngân hàng hay hiệp hội xây dựng có thông tin về tài khoản tiết kiệm hoặc tài sản thế chấp, ta thường thấy hai loại lãi suất – một loại nổi bật thu hút ánh nhìn (lãi suất theo tháng hoặc quý hoặc lãi suất danh nghĩa theo năm) và AER (lãi suất tương đương theo năm). Cần có AER, do khoản thanh toán hoặc chi trả thường được thực hiện ở những giai đoạn dưới một năm, như vậy khoản nợ (hoặc tiết kiệm) có thể tăng lên với tốc độ nhanh hơn lãi suất được quảng cáo. Ví dụ, nếu lãi phát sinh từ khoản nợ quá hạn thể tín dụng là 2%/tháng thì khoản nợ sẽ tăng với tốc độ thực tế là gần 27%/năm.

### 9.3.2 KHẤU HAO

Khấu hao (*depreciation*) là sự mất giá trị của tài sản theo thời gian. Trong quá trình sử dụng, hầu hết các loại tài sản giá trị lớn đều giảm dần giá trị, cũng bởi tài sản đó không còn là kiểu mẫu thời thượng nhất nữa. Ta thường thấy tình trạng này ở xe ô tô và thiết bị vi tính.

Sau khi được mua về thì thiết bị sẽ trở thành một phần tài sản của công ty và có một giá trị nhất định tại bất cứ thời điểm nào. Rõ ràng giá trị này sẽ giảm đi theo thời gian khi thiết bị được sử dụng. Kế toán công ty sẽ tính khoản giá trị mất đi (mà ta gọi là khấu hao) vào một tài khoản. Chi phí ban đầu được “khấu trừ” dần theo vòng đời của thiết bị như một giá trị khấu hao nhất định mỗi năm. Khấu hao có thể làm giảm lợi nhuận, do đó công ty có thể tiết kiệm dần để mua thay thế khi thiết bị đã đến cuối vòng đời.

Có nhiều cách xác định khấu hao. Hai phương pháp được sử dụng rộng rãi nhất là:

- phương pháp đường thẳng
- phương pháp số dư giảm dần

Phương pháp đường thẳng (*straight line method*) coi tài sản khấu hao một khoản cố định mỗi năm. Như vậy, nếu ban đầu ta mua thiết bị này với giá trị  $V_0$  và dự kiến bán lại sau  $n$  năm với giá trị  $V_n$ , thì:

$$\text{Khấu hao hàng năm} = \frac{\text{giá trị mua thiết bị} - \text{giá trị bán lại}}{\text{tuổi thọ ước tính của thiết bị}}$$

$$\text{Khấu hao hàng năm} = \frac{V_0 - V_n}{n}$$

Phần lớn thiết bị điện tử sẽ mất một phần lớn giá trị ban đầu chỉ sau một vài năm, nghĩa là khoảng thời gian rất nhỏ nếu tính đến hết vòng đời thực tế. Trong trường hợp này, phương pháp đường thẳng không phản ánh được giá trị thực tế của tài sản vào thời điểm cuối mỗi năm.

Với phương pháp số dư giảm dần (*reducing balance method*), ta giả định giá trị của tài sản sẽ mất đi một số phần trăm cố định qua mỗi năm.

Nếu  $V_0$  là giá trị ban đầu của tài sản và mức độ khấu hao là  $r\%/năm$ , thì giá trị của tài sản cuối năm đầu tiên bằng:

$$V_1 = V_0 - r\%V_0 = V_0(1 - r\%)$$

Giá trị của tài sản cuối năm thứ 2 bằng:

$$V_2 = V_1 - r\%V_1 = V_1(1 - r\%)$$

Giá trị của tài sản cuối năm thứ 3 bằng:

$$V_3 = V_2 - r\%V_2 = V_2(1 - r\%)$$

và tiếp diễn cho đến  $V_n$ .

Chuỗi tính ở trên có thể thiết lập thành công thức “khấu hao số dư giảm dần” tổng quát:

$$V_n = V_0(1 - r\%)^n$$

Thay vào đó, ta cũng có thể thiết lập mô hình Microsoft Excel để tính giá trị đã khấu hao của tài sản tại thời điểm cuối mỗi năm. Xem **Bảng 9.5**.

	A	B	C	D	E
	Cuối năm	Giá trị đã khấu hao (£)	Khấu hao (£)	Mức khấu hao hàng năm =	
1					
2	0				
3	1	=B2*(1-\$E\$1)	=B2-B3		
4	2	=B3*(1-\$E\$1)	=B3-B4		
5	3	=B4*(1-\$E\$1)	=B4-B5		
6	4	=B5*(1-\$E\$1)	=B5-B6		
7	...	...			

**Bảng 9.5** Giá trị tài sản đã khấu hao.



Nhập giá trị ban đầu của tài sản vào ô B2 và mức khấu hao vào ô E1.

Hãy xem công thức ở ô B3 và C3 trở đi – đảm bảo rằng bạn hiểu cách tính trong các công thức này.

### Ví dụ 9.3 – Tính khấu hao

Người ta dự tính bán lại một chiếc máy có giá 6000£ sau 5 năm sử dụng với giá 2000£. Xác định mức khấu hao mỗi năm, trong trường hợp sử dụng

- phương pháp đường thẳng;
- phương pháp số dư giảm dần với mức khấu hao 20%.

#### Lời giải:

- Phương pháp đường thẳng:

$$\text{Khấu hao hàng năm} = \frac{6000 - 2000}{5} = 800\text{£}$$

Như vậy sau mỗi năm, giá trị chiếc máy giảm đi 800£ trong vòng 5 năm.

- Phương pháp số dư giảm dần:

	A	B	C	D	E
1	Cuối năm	Giá trị đã khấu hao (£)	Khấu hao (£)	Mức khấu hao hàng năm =	20%
2	0	6000			
3	1	=B2*(1-£E\$1)	=B2-B3		
4	2	=B3*(1-£E\$1)	=B3-B4		
5	3	=B4*(1-£E\$1)	=B4-B5		
6	4	=B5*(1-£E\$1)	=B5-B6		
7	5	=B5*(1-£E\$1)	=B6-B7		

**Bảng 9.6** Phương pháp số dư giảm dần.

Hoặc thay vào đó, ta có thể thiết lập mô hình như sau:

	A	B	C	D	E
1	Cuối năm	Giá trị đã khấu hao (£)	Khấu hao (£)	Mức khấu hao hàng năm =	20%
2	0	6000			
3	1	$=B\$2*(1-E\$1)^{A3}$	$=B2-B3$		
4	2	$=B\$2*(1-E\$1)^{A4}$	$=B3-B4$		
5	3	$=B\$2*(1-E\$1)^{A5}$	$=B4-B5$		
6	4	$=B\$2*(1-E\$1)^{A6}$	$=B5-B6$		
7	5	$=B\$2*(1-E\$1)^{A7}$	$=B6-B7$		

**Bảng 9.7** Phương pháp số dư giảm dần thay thế.

Hai mô hình trên cùng cho kết quả là bảng giá trị như sau:

	A	B	C	D	E
1	Cuối năm	Giá trị đã khấu hao (£)	Khấu hao (£)	Mức khấu hao hàng năm =	20%
2	0	6000			
3	1	4800	1200		
4	2	3840	960		
5	3	3072	768		
6	4	2458	614		
7	5	1966	492		

**Bảng 9.8** Giá trị tính theo phương pháp số dư giảm dần.

**Lưu ý:** Giá trị đích không chính xác là 2000£. Đây không phải là bất lợi của phương pháp này, do giá trị cuối thường khó là một con số chắc chắn.

Trong trường hợp thiết lập mô hình số dư giảm dần khi chưa biết mức khấu hao, ta có thể xác định một giá trị thích hợp bằng cách điều chỉnh giá trị nhập vào ô E1 cho đến khi đạt được giá trị đích mong muốn.

### 9.3.3 LẠM PHÁT

Ảnh hưởng của lạm phát (*inflation*) được xác định giống hệt như cách tính lãi kép.

#### Ví dụ 9.4 – Tính lạm phát

Giả sử vào tháng 1/1999 giá cà phê là 3,30£/hũ 200g. Với lạm phát cố định 8%/năm thì giá một hũ cà phê 200g sẽ là bao nhiêu vào tháng 1/2003?

$$V_n = V_0 (1 + (r/100))^n$$

Do đó  $V_4 = 3,30(1 + 0,08)^4 = 3,30 \cdot 1,08^4 = 4,49$

Như vậy, ước tính hũ cà phê sẽ có giá 4,49£ vào năm 2003.

Sử dụng Microsoft Excel để thiết lập mô hình tính, ta có:

	A	B	C	D
1	Cuối năm	Giá (£)	Mức lạm phát hàng năm =	8%
2	0	3,30		
3	1	=B2*(1+\$D\$1)		
4	2	=B3*(1+\$D\$1)		
5	3	=B4*(1+\$D\$1)		
6	4	=B5*(1+\$D\$1)		

**Bảng 9.9** Tính lạm phát.

Giá trị tính được là:

	A	B	C	D
1	Cuối năm	Giá (£)	Mức lạm phát hàng năm =	8%
2	0	3,30		
3	1	3,56		
4	2	3,85		
5	3	4,16		
6	4	4,49		

**Bảng 9.10** Giá trị lạm phát tính được.

## 9.4 GIÁ TRỊ HIỆN TẠI

Trong kinh doanh, chúng ta sẽ thường phải xử lý các khoản tiền phát sinh tại nhiều thời điểm khác nhau. Thời gian phát sinh sẽ ảnh hưởng đến giá trị của khoản tiền tại “**thời điểm thực**”, nghĩa là 100£ ta nhận được vào 2 năm tới sẽ mua được ít đồ hơn (có “giá trị thực” thấp hơn) so với 100£ mà ta đang có bây giờ. Để đánh giá giá trị thực của một khoản tiền, ta thường xác định giá trị của nó tại **thời điểm hiện tại (present time)**.

Giá trị **hiện tại** của một khoản tiền sẽ phát sinh tại một thời điểm nào đó, được gọi là **giá trị hiện tại (present value)** của khoản tiền đó.

Việc tính giá trị hiện tại của một khoản tiền sẽ phát sinh trong tương lai nghĩa là ta sẽ phải giảm giá trị thực tế trong tương lai đến một giá trị tương đương ở thời điểm hiện tại, thông qua một **tỷ lệ chiết khấu (discount rate)** theo năm phù hợp.

Quá trình này được gọi là **chiết khấu (discounting)**.

Xét  $A_n = A_0 (1 + r\%)^n$

với  $A_n$  là khoản tiền sẽ phát sinh trong vòng  $n$  năm tới

$A_0$  là giá trị hiện tại của khoản tiền

$r\%$  là tỷ lệ chiết khấu

Theo đó, ta có giá trị hiện tại là

$$A_0 = \frac{A_n}{(1 + r\%)^n}$$

### Ví dụ 9.5 – Tính giá trị hiện tại

Bà dì đã hứa tặng bạn 1000£ vào sinh nhật 25 tuổi, nghĩa là 5 năm tới kể từ thời điểm hiện tại. Vậy món quà đó sẽ có giá trị hiện tại bằng bao nhiêu, nếu giả sử tỷ lệ chiết khấu hàng năm trung bình là 6%?

**Lời giải:**

$$\text{Giá trị hiện tại } A_0 = \frac{A_n}{(1 + r\%)^n} = \frac{1000}{(1 + 0,06)^5} = 747,26$$

Do đó, giá trị hiện tại của món quà khi tỷ lệ chiết khấu bằng 6%/năm là 747£ nếu làm tròn đến giá trị nguyên gần nhất. Giá trị hiện tại được tính bằng cách nhân  $A_n$  với hệ số sau:

$$\frac{1}{(1 + r\%)^n}$$

Ta gọi hệ số nhân này là **hệ số chiết khấu (discount factor)**.

Việc lựa chọn giá trị cho  $r$  sẽ phụ thuộc vào tình huống chiết khấu. Trong trường hợp như ở **Ví dụ 5**, ta sẽ sử dụng tỷ lệ trung bình cho tài khoản tiết kiệm. Trong hai ứng dụng được thảo luận ở phần tiếp theo, cả hai đều liên quan đến quỹ đầu tư (thảm định niên khoản hoặc dự án vốn), ta sẽ chọn tỷ lệ chiết khấu dựa vào lãi suất của quỹ đầu tư nếu không sử dụng cho mục đích kể trên, cũng như trang trải các khoản thuế, chi phí quản lý...

### Ví dụ 9.6 – Tính tổng giá trị hiện tại của một chuỗi thanh toán

Bố mẹ của Tom đã hứa cứ mỗi 6 tháng sẽ cho anh ta 1000£ trong thời gian học đại học. Giả sử Tom vừa mới bắt đầu khóa học sẽ kéo dài 4 năm.

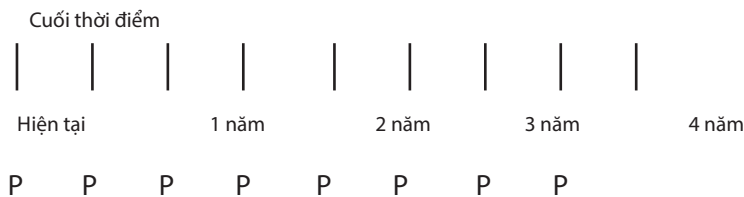
Tổng giá trị hiện tại của khoản tiền cho Tom là bao nhiêu, nếu như bố mẹ anh ta đầu tư số tiền đó vào chỗ khác với lãi suất 3%/6 tháng.

#### Lời giải:

$$\text{Giá trị hiện tại của mỗi khoản trả: } A_0 = \frac{A_n}{(1 + (r / 100))^n} = \frac{1000}{(1 + 0,03)^n}$$

với  $n$  là số giai đoạn nửa năm tương ứng với số thời điểm tính chiết khấu.

Các khoản trả ( $P$ ) được thực hiện như sau:



Tổng giá trị hiện tại (PV) của tất cả các khoản trả bằng:

$$\begin{aligned}
 PV_{3\%/6 \text{ tháng}} &= 1000 + \frac{1000}{1,03} + \frac{1000}{1,03^2} + \frac{1000}{1,03^3} + \frac{1000}{1,03^4} + \frac{1000}{1,03^5} + \frac{1000}{1,03^6} + \frac{1000}{1,03^7} \\
 &= 1000 \times \left( 1 + \frac{1}{1,03} + \frac{1}{1,03^2} + \frac{1}{1,03^3} + \frac{1}{1,03^4} + \frac{1}{1,03^5} + \frac{1}{1,03^6} + \frac{1}{1,03^7} \right) \\
 &= 1000 \times 7,23028 \\
 &= 7230\text{£}
 \end{aligned}$$

Như vậy, tổng giá trị hiện tại của tất cả 8 khoản trả 1000£ mà Tom sẽ nhận là 7230£.

Ta có thể thiết lập mô hình bài toán này với Microsoft Excel như sau:

	A	B	C	D	E	F
1	Cuối nửa năm	Hệ số chiết khấu	Dòng tiền (£)	Giá trị hiện tại (£)	Tỷ lệ chiết khấu =	3%
2	0	$=1/(1+\$F\$1)^{A2}$	1000	$=C2*B2$		
3	1	$=1/(1+\$F\$1)^{A3}$	1000	$=C3*B3$		
4	2	$=1/(1+\$F\$1)^{A4}$	1000	$=C4*B4$		
5	3	$=1/(1+\$F\$1)^{A5}$	1000	$=C5*B5$		
6	4	$=1/(1+\$F\$1)^{A6}$	1000	$=C6*B6$		
7	5	$=1/(1+\$F\$1)^{A7}$	1000	$=C7*B7$		
8	6	$=1/(1+\$F\$1)^{A8}$	1000	$=C8*B8$		
9	7	$=1/(1+\$F\$1)^{A9}$	1000	$=C9*B9$		
10	8					
11			Tổng =	$=SUM(D2:D9)$		

**Bảng 9.11** Tính giá trị hiện tại.

Bảng trên cho kết quả như sau:

	A	B	C	D	E	F
1	Cuối nửa năm	Hệ số chiết khấu	Dòng tiền (£)	Giá trị hiện tại (£)	Tỷ lệ chiết khấu =	3%
2	0	1	1000	1000		
3	1	0.97087	1000	970.874		
4	2	0.94260	1000	942.596		
5	3	0.91514	1000	915.142		
6	4	0.88849	1000	888.487		
7	5	0.86261	1000	862.609		

	A	B	C	D	E	F
8	6	0.83748	1000	837.484		
9	7	0.81309	1000	813.092		
10	8					
11			Tổng =	7230.28		

**Bảng 9.12** Kết quả tính giá trị hiện tại.

**Lưu ý:** Trong ví dụ tính giá trị hiện tại trên, ta giả định dòng tiền phát sinh tại thời điểm cuối mỗi giai đoạn.

### 9.4.1 GIÁ TRỊ HIỆN TẠI THUẦN

Trong quá trình quản trị kinh doanh, ban lãnh đạo sẽ phải ra quyết định đầu tư bằng cách chọn lựa giữa các hoạt động cạnh tranh. Ví dụ, bộ phận R&D đang có dự án phát triển một sản phẩm; bộ phận Marketing có dự án quảng bá sản phẩm trên phạm vi toàn quốc; bộ phận Kỹ thuật đề xuất nâng cấp hệ thống IT. Nếu quỹ hiện tại của công ty chỉ có thể đáp ứng được một trong các hoạt động này, ban quản lý sẽ phải đưa ra quyết định. Một trong các yếu tố ảnh hưởng đến quá trình ra quyết định là **giá trị hiện tại thuần (net present value)**.

Giá trị hiện tại thuần (NPV) của một dự án được tính bằng tổng dòng tiền cuối năm mà dự án tính thu về và trừ đi khoản tiền bỏ ra ban đầu; nghĩa là lấy tổng giá trị hiện tại của dòng tiền sẽ sinh ra cuối năm trừ đi chi phí dự án vốn.

Trong cách tính này, tỷ lệ chiết khấu được sử dụng sẽ tương đương với “chi phí sử dụng vốn” (*cost of capital*) của công ty. Đây là tỷ lệ thu về thấp nhất từ dự án đầu tư mà công ty dự tính sẽ đạt được trước khi quyết định đầu tư. Nói cách khác, đây là chi phí mà công ty mượn vốn đầu tư và do đó khoản thu về từ dự án buộc phải bù đắp được chi phí vốn vay.

Nếu NPV có giá trị dương, nghĩa là số tiền thu về từ dự án tương đương với lãi suất lớn hơn tỷ lệ chiết khấu được sử dụng để tính giá trị hiện tại của dòng tiền, nên đây là một dự án sinh lợi nhuận. Nếu NPV có giá trị âm, nghĩa là số tiền thu về từ dự án tương đương với lãi suất thấp hơn tỷ lệ chiết khấu, nên dự án này không sinh lợi nhuận.

### Ví dụ 9.7 – Thẩm định dự án vốn

Một công ty đang cân nhắc hai dự án khả thi. Chi phí ban đầu và dòng tiền dự tính được đưa ra trong bảng dưới đây.

Công ty thường trông chờ khoản thu về là 15%/năm đối với tiền vốn đầu tư.

Dự án	Chi phí ban đầu (nghìn £)	Dòng tiền thuần cuối năm, nghìn £				
		Năm 1	Năm 2	Năm 3	Năm 4	Năm 5
<b>A</b>	-1000	500	400	300	200	100
<b>B</b>	-1000	200	200	300	400	400

**Bảng 9.13** Dòng tiền cho hai dự án.

Sử dụng giá trị hiện tại thuần để xác định xem công ty nên chọn dự án nào trong hai dự án trên (nếu có).

**Lời giải:** Mô hình Microsoft Excel ta thiết lập được là:

	<b>A</b>	<b>B</b>	<b>C</b>	<b>D</b>	<b>E</b>	<b>F</b>
<b>1</b>	<b>Tỷ lệ chiết khấu</b> =	<b>15%</b>	<b>Dự án A</b>		<b>Dự án B</b>	
<b>2</b>	<b>Cuối năm</b>	<b>Hệ số chiết khấu</b>	<b>Dòng tiền cuối năm (nghìn £)</b>	<b>Giá trị hiện tại (nghìn £)</b>	<b>Dòng tiền cuối năm (nghìn £)</b>	<b>Giá trị hiện tại (nghìn £)</b>
<b>3</b>	0	=1/(1+\$B\$1)^A3	-1000	=C3*B3	-1000	=E3*B3
<b>4</b>	1	=1/(1+\$B\$1)^A4	500	=C4*B4	200	=E4*B4
<b>5</b>	2	=1/(1+\$B\$1)^A5	400	=C5*B5	200	=E5*B5
<b>6</b>	3	=1/(1+\$B\$1)^A6	300	=C6*B6	300	=E6*B6
<b>7</b>	4	=1/(1+\$B\$1)^A7	200	=C7*B7	400	=E7*B7
<b>8</b>	5	=1/(1+\$B\$1)^A8	100	=C8*B8	400	=E8*B8
<b>9</b>						
<b>10</b>			NPV =	=SUM(D3:D8)	NPV =	=SUM(F3:F8)

**Bảng 9.14** So sánh hai dự án.



	A	B	C	D	E	F
1	Tỷ lệ chiết khấu =	15%	Dự án A		Dự án B	
2	Cuối năm	Hệ số chiết khấu	Dòng tiền cuối năm (nghìn £)	Giá trị hiện tại (nghìn £)	Dòng tiền cuối năm (nghìn £)	Giá trị hiện tại (nghìn £)
3	0	1	-1000	-1000.00	-1000	-1000.00
4	1	0.86957	500	434.78	200	173.91
5	2	0.75614	400	302.46	200	151.23
6	3	0.65752	300	197.26	300	197.26
7	4	0.57175	200	114.35	400	228.70
8	5	0.49718	100	49.72	400	198.87
9						
10			NPV =	98.56	NPV =	-50.03

**Bảng 9.15** Giá trị so sánh của hai dự án.

Bảng này cho thấy Dự án A có giá trị NPV dương, do đó khoản thu về so với chi phí ban đầu lớn hơn 15%/năm. Như vậy, dự án A sẽ sinh lãi và thu về từ khoản đầu tư 15%/năm. Dự án B có giá trị NPV âm, nên thu về khoản tiền 15%/năm và không đáp ứng yêu cầu của công ty về khoản thu trên vốn đầu tư cho dự án. Trên thực tế, quyết định đầu tư không chỉ dựa vào giá trị NPV nói riêng, nhưng đây là một yếu tố quan trọng trong quá trình thẩm định các dự án đầu tư vốn.



## Chương 10

# KẾT LUẬN VÀ CÂN NHẮC

Cuốn sách này đã giới thiệu một số phương pháp phân tích dữ liệu thông qua rất nhiều ví dụ mà chúng ta đã thực hiện với các mô hình trên Microsoft Excel.

Các ví dụ này được mô hình hóa dựa trên các tập dữ liệu mà mỗi quản lý thường phải thu thập trong công việc hàng ngày và cần xử lý để hỗ trợ việc ra quyết định. Ở giai đoạn này, việc đưa ra nhận xét về tính hữu dụng và ứng dụng của các mô hình minh họa sẽ rất hữu ích.

Cân nhắc đầu tiên và quan trọng nhất là kết quả của nhiều mô hình được sử dụng cho mục đích ra quyết định chỉ mang tính định hướng – tức là mô hình không thể tự đưa ra quyết định, mà người ra quyết định cuối cùng là nhà quản lý/chuyên gia sử dụng mô hình. Điều này có nghĩa trách nhiệm đưa ra quyết định “không là việc” của mô hình!

Do đó, tại giai đoạn này, chúng tôi sẽ chỉ đưa ra một số cân nhắc liên quan đến việc sử dụng mô hình trong phân tích và ra quyết định kinh doanh.

Thông qua nhiều ví dụ mà chúng ta đã trình bày, có lẽ bạn đã hiểu rõ mô hình với vai trò là một phần trong quá trình ra quyết định về quản lý. Tuy nhiên, chúng tôi sẽ tóm tắt một số điểm cần thiết ở đây, đồng thời đánh giá các ưu, nhược điểm. Mô hình hóa là một trong các cách khảo sát những tùy chọn khả thi khi cần đưa ra quyết định.

Dưới đây là một vài cách diễn hình khi sử dụng mô hình để ra quyết định:

- Mô hình giúp người ra quyết định:
  - rút ra các yếu tố quan trọng trong vấn đề cần quyết định,
  - khảo sát biến động trong nhiều điều kiện/tham số,
  - kiểm tra các giả thuyết.
- Mô hình là công cụ hỗ trợ, nhưng không thay thế, cho tư duy sáng tạo và đánh giá của người quản lý.
- Việc thiết lập mô hình kéo theo việc thảo luận về quyết định và các vấn đề liên quan.

- Mô hình nên đưa ra nhiều tùy chọn để cân nhắc, chứ không phải đáp án cho vấn đề. Phải cân nhắc cả thông tin định tính.
- Việc sử dụng mô hình không có nghĩa là bắt buộc thực hiện một hành động cụ thể nào cả.

Vấn đề có thể phát sinh trong quá trình mô hình hóa thường liên quan đến việc khó thể hiện tình huống phức tạp trên mô hình hoặc khó rút được giá trị đáng tin cậy từ vài tham số (mà ta đã gặp trong ví dụ minh họa của sách này).

Tuy nhiên, nếu vấn đề cần giải quyết mang tính tổng quan và có thể chấp nhận một giải pháp chung, thì mô hình là một giải pháp hữu hiệu và tối ưu chi phí. Mô hình có tính đa dụng – tức là chúng ta có thể đặt nhiều câu hỏi “nếu... thì” và sử dụng mô hình để dự đoán diễn biến của hệ thống trong nhiều điều kiện khác nhau. Nếu không có mô hình, nhà quản lý sẽ phải dựa hoàn toàn vào kinh nghiệm và trực giác cá nhân. Đối với nhà quản lý chưa có kinh nghiệm về vấn đề đang xét đến, thì mô hình là công cụ cần thiết, bởi trực giác có thể là chưa đủ.

Sau đây là phần tổng kết ưu điểm và nhược điểm của mô hình hóa.

**Ưu điểm** của việc sử dụng mô hình bao gồm:

- Khuyến khích thảo luận và sáng tạo
  - đòi hỏi người ra quyết định nghĩ kỹ về bản chất và cấu trúc của tình huống/hệ thống đang được xét,
  - đòi hỏi người ra quyết định khảo sát các tính chất của tình huống/hệ thống,
  - cho phép người ra quyết định chất vấn các quan điểm và giả thuyết “vốn có”.
- Cung cấp khuôn khổ đánh giá dữ liệu.
- Cung cấp phương thức giao tiếp trong nhiều kịch bản kinh doanh khả thi.
- Cho phép đánh giá nhiều tình huống hành động, từ đó có cơ sở hỗ trợ ra quyết định/hành động cụ thể.
- Tính đa dụng – mô hình cho phép chúng ta phân tích giả thuyết “nếu... thì” – nói cách khác, nếu thay đổi các giá trị của một tham số thì kết quả số trong mô hình sẽ ra sao?
- Tính an toàn – trong quá trình thực hiện, việc khảo sát mô hình không ảnh hưởng đến tổ chức do chưa có quyết định được đưa ra.
- Tốc độ – có thể khảo sát tác động của nhiều quyết định hành động một cách nhanh chóng.
- Chi phí thấp – việc phân tích kỹ vấn đề có thể giúp ta tránh mắc lỗi gây hao tổn chi phí.

Nhìn chung, nhờ mô hình hóa ta có thể đưa ra những quyết định thấu đáo, củng cố cơ sở lập luận trong tình huống ra quyết định.

**Nhược điểm** của việc sử dụng mô hình bao gồm:

- Mô hình có thể quá rộng so với phạm vi thực tế.
- Nhiều giá trị tham số có thể chỉ là ước tính, như vậy giới hạn lỗi trong giá trị đầu ra của mô hình tăng lên.
- Chi phí thiết lập và khảo sát mô hình có thể vượt quá lợi nhuận thu được từ những quyết định có rủi ro/chi phí thấp.

Khi cân nhắc việc áp dụng một quy trình mô hình hóa, ta nên đặt các câu hỏi sau:

- Liệu quyết định có thích hợp để mô hình hóa? Liệu có thể sử dụng phương pháp đánh giá số nào khác để xử lý tập dữ liệu không?
- Liệu kết quả của mô hình có giúp ích cho việc đánh giá?
- Sẽ mất bao lâu để thiết lập và khảo sát mô hình?
- Sẽ mất bao chi phí để thiết lập và khảo sát mô hình?

Nếu đa số câu trả lời là “tích cực”, thì mới đáng tiến vào giai đoạn tiếp theo.

