CHƯƠNG **5**

CÁC PHÂN PHỐI XÁC SUẤT CHUẨN VÀ LIÊN TỤC KHÁC

***Về chương này*:**

Một số biến số ngẫu nhiên rời rạc và các phân phối xác suất của chúng đã được trình bày trong Chương 4. Mục đích của chương này là giới thiệu với các bạn biến số ngẫu nhiên chuẩn, một trong những biến số ngẫu nhiên liên tục quan trọng và thường gặp nhất. Chúng tôi trình bày phân phối xác suất của chúng, và chúng tôi chứng tỏ cách thức mà phân phối xác suất này có thể được sử dụng.

NGHIÊN CỨU ĐIỂN HÌNH

MỘT BÌNH NHIÊN LIỆU ĐÁNG GIÁ BAO NHIÊU?

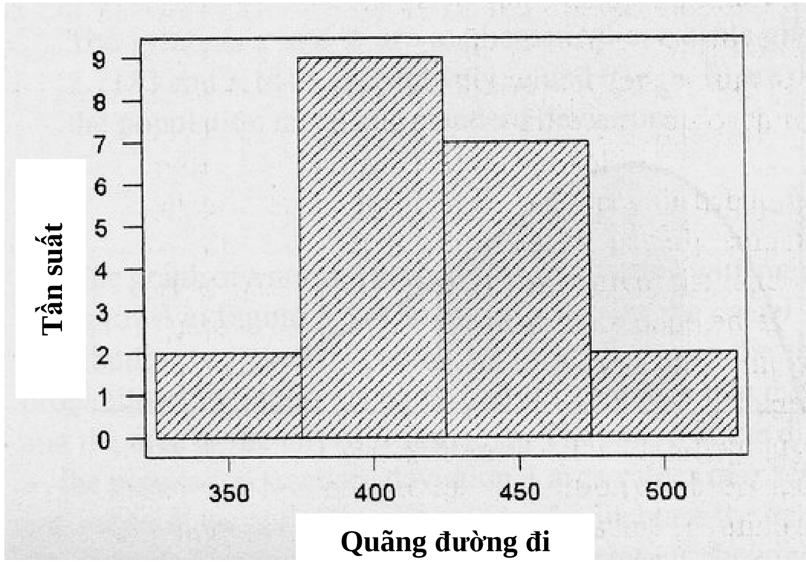
Mua một chiếc xe mới lúc nào cũng là một trải nghiệm hấp dẫn, bởi vì mỗi chúng ta có những kỳ vọng khác nhau về việc chiếc xe mới của chúng ta - bất luận nó có là một chiếc xe con, xe gia đình hay bán tải - sẽ ra sao và nó sẽ vận hành thế nào. Một khi chúng ta đã quyết định về màu sắc, loại xe, và các chọn lựa mà chúng ta muốn có trong chiếc xe mà chúng ta mua, thì chúng ta phải đối mặt với nhiều tiêu chuẩn khác có liên quan đến những quyết định mà chúng ta thực thi. Liệu chiếc xe mà chúng ta chọn lựa có tiết kiệm nhiên liệu không khi di chuyển trong thành phố cũng như trên đường cao tốc? Khác biệt ra sao về khoảng cách phanh khi đường trơn ướt so với khi đường khô ráo? Quãng đường đi (với một bình đầy nhiên liệu) nào của chiếc xe mà chúng ta sẽ lựa chọn?

Khi so sánh số dặm đường trung bình tính trên 1 galông nhiên liệu (mpg) khi lái trong thành phố và ngoài đường cao tốc, thì 20 chiếc xe mà chúng tôi đã chọn từ năm số tạp chí *Consumer Reports (Các Báo cáo về Người tiêu dùng)*(tháng Giêng - tháng Tám 1994) trung bình đạt từ 10 đến 17 mpg khi chạy trong thành phố và từ 21 đến 41 mpg khi lái trên đường cao tốc. Quãng đường đi với một bình nhiên liệu đầy thay đổi từ 350 đến 495 dặm. Trên thực tế, quãng đường đi trung bình là 418,0 dặm, trung vị và trung bình có trọng số lần lượt là 420,0 và 419,1; và độ lệch chuẩn là 45,8 dặm. Bởi vì trung vị và trung bình có trọng số chỉ khác biệt rất ít so với trung bình, chúng ta ắt sẽ kỳ vọng là quãng đường đi được trình bày trong Hình 5.1 sẽ có hình dạng gò và, nếu như có thêm nhiều loại xe hơn nữa được kiểm tra, thì nhiều khả năng được phân phối chuẩn. Các biến số như là những biến số được báo cáo ở đây có xu hướng được phân phối chuẩn, như các biến số khác mà phản ảnh nhiều nhân tố nhỏ nhưng quan trọng sẽ quyết định giá trị của các biến số này.

Ngoài màu sơn và những đồ chọn lựa khác mà bạn có lẽ có khả năng thêm vào chiếc xe mới mua của mình, liệu những chiếc xe có thực sự khác biệt về những đặc trưng này mà rốt cuộc có thể giúp bạn tiết kiệm được một số tiền, và trong trường hợp quãng đường đi, giúp bạn tránh được tình huống khó xử khi bị mắc kẹt với một thùng nhiên liệu trống rỗng?

Đường cong chuẩn, được sử dụng như là một mô hình cho các phân phối tần suất tương đối cho nhiều biến số ngẫu nhiên liên tục, là chủ đề của Chương 5. Chúng ta sẽ khảo cứu các tính chất của chúng, xem cách thức mà chúng có thể được sử dụng để tính toán xác suất, và xem cách mà chúng ta có thể quyết định chọn chiếc xe nào dựa trên những quyết định hợp lý và phân phối xác suất chuẩn.

HÌNH 5.1 **Quãng đường đi trung bình cho *n* =20 chiếc xe đời 1994**



Nguồn: *Báo cáo Người tiêu dùng,* tháng 1-8, 1994

5.1 PHÂN PHỐI XÁC SUẤT CHO CÁC BIẾN SỐ NGẪU NHIÊN LIÊN TỤC

Khi một biến số *x* là rời rạc, chúng ta có thể chỉ định một xác suất dương cho từng giá trị mà *x* có thể có và có được phân phối xác suất cho *x.* Tổng của tất cả các xác suất đi cùng với những giá trị khác nhau của *x* là 1. Tuy nhiên, không phải tất cả các thí nghiệm đều tạo ra những biến số ngẫu nhiên mà rời rạc. **Các biến số ngẫu nhiên liên tục,** ví dụ như chiều cao và cân nặng, vòng đời của một sản phẩm cụ thể, hay khoảng cách thời gian giữa những lần bán hàng, có thể có vô vàn giá trị tương ứng với các điểm trên một đường khoảng cách. Nếu chúng ta cố ấn định một xác suất dương cho mỗi trong số những giá trị không thể đếm được này, thì các xác suất sẽ không còn có tổng là 1 nữa, như với các biến số ngẫu nhiên rời rạc. Do vậy, chúng ta phải sử dụng một cách tiếp cận khác để tạo ra phân phối xác suất cho một biến số ngẫu nhiên liên tục.

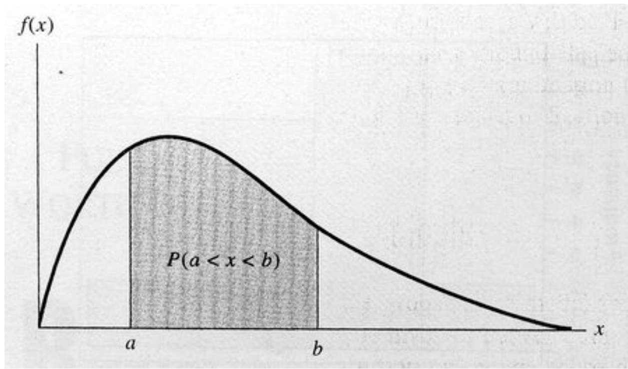
Mô hình xác suất cho phân phối tần suất của một biến số ngẫu nhiên liên tục có liên quan đến sự chọn lựa một đường cong, thường là trơn tru, được gọi là **phân phối xác suất** hay **hàm mật độ xác suất** của biến số ngẫu nhiên đó. Nếu phương trình của phân phối xác suất liên tục này được ký hiệu là $f(x)$, thì xác suất của *x* rơi vào khoảng $a < x < b$ là diện tích bên dưới đường phân phối xác suất $f(x)$ giữa hai điểm *a* và *b* (xem Hình 5.2). Điều này nhất quán với sự biểu diễn của một biểu đồ tần suất tương đối (Chương 2), nơi mà những diện tích nằm phía trên một khoảng trong đồ thị này tương ứng với tỷ lệ của các quan sát rơi vào khoảng đó. Bởi vì số lượng các giá trị mà *x* có thể có là vô cùng lớn và không đếm được, nên xác suất mà *x* bằng với một giá trị cụ thể nào đó, ví dụ *a*, là bằng zêrô. Vì vậy những báo cáo xác suất về các biến số ngẫu nhiên liên tục luôn luôn tương ứng với các diện tích bên dưới phân phối xác suất trong một khoảng, ví dụ từ *a* đến *b*, và được biểu diễn bằng $P(a < x < b)$. Lưu ý rằng xác suất mà $a < x < b$ là bằng với xác suất $a \le x \le b$, bởi vì $P(x = a) = P(x = b) = 0$.

Bằng cách nào mà chúng ta chọn mô hình này - nghĩa là, phân phối xác suất $f(x)$ - phù hợp với thí nghiệm đã biết? Nhiều loại hình đường cong liên tục là sẵn có cho việc mô hình hóa. Một số có hình dạng gò, giống như trong Hình 5.2, nhưng nhiều đường cong khác là không như thế. Nói chung, chúng ta có gắng chọn ra một mô hình mà:

- phù hợp với số lượng tích lũy của dữ liệu

- cho phép chúng ta thực hiện những suy luận có thể có tốt nhất qua việc sử dụng dữ liệu này

HÌNH 5.2 **Phân phối xác suất cho một biến số ngẫu nhiên liên tục**



Mô hình của chúng ta có lẽ không phải lúc nào cũng phù hợp với tình huống thí nghiệm một cách hoàn hảo, nhưng chúng ta cố gắng chọn lựa một mô hình mà *phù hợp tốt nhất* với biểu đồ tần suất tương đối của tổng thể. Mô hình của chúng ta càng ước lượng xấp xỉ với thực tế bao nhiêu thì các suy luận của chúng ta càng tốt hơn bấy nhiêu. May mắn là, chúng ta sẽ tìm ra rằng nhiều biến số ngẫu nhiên liên tục có phân phối tần suất theo hình dạng gò. Một mô hình xác suất mà cung cấp một sự ước lượng xấp xỉ tốt cho một sự phân phối là **phân phối xác suất chuẩn,** chủ đề của Phần 5.2.

5.2 PHÂN PHỐI XÁC XUẤT CHUẨN

Trong Phần 5.1, chúng ta đã thấy rằng mô hình xác suất cho phân phối tần suất của một biến số ngẫu nhiên liên tục có liên quan đến việc lựa chọn đường cong, thường là trơn tru, được gọi là **phân phối xác suất.** Mặc dù những phân phối này có thể có nhiều hình dạng khác nhau, thì một số lớn các biến số ngẫu nhiên quan sát được trong tự nhiên sở hữu một phân phối tần suất mà có hình dạng gần giống quả chuông hay, như một nhà thống kê ắt sẽ nói, là xấp xỉ một phân phối xác suất chuẩn. Công thức mà tạo ra phân phối này được thể hiện dưới đây.

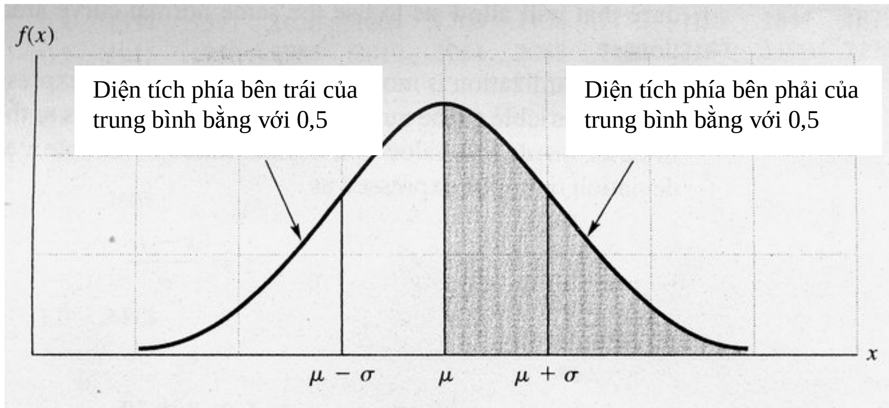
Phân phối Xác suất Chuẩn

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/(2\sigma^2)} \qquad -\infty < x < \infty$$

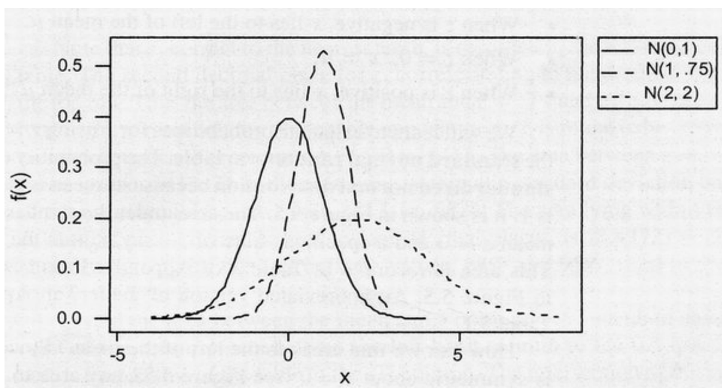
Ký hiệu e và π là các hằng số toán học có giá trị xấp xỉ lần lượt 2,7183 và 3,1416; μ và σ (σ>0) là những tham số đại diện cho trung bình và độ lệch chuẩn của tổng thể.

Đồ thị của một phân phối xác suất chuẩn với trung bình μ và độ lệch chuẩn σ được thể hiện trong Hình 5.3. Trung bình μ nằm ở *trung tâm* của phân phối, và phân phối này có tính *đối xứng* qua trung bình μ của nó. Do tổng diện tích nằm dưới phân phối xác suất chuẩn này là bằng với 1, điều này hàm ý rằng diện tích nằm về phía bên phải của μ là 0,5 và diện tích nằm ở phía bên trái của μ cũng là 0,5. *Hình dạng* của phân phối này được quyết định bởi σ, độ lệch chuẩn của tổng thể. Những giá trị lớn của σ làm giảm đi chiều cao của đường cong và làm tăng bề rộng; các giá trị nhỏ của σ làm tăng chiều cao của đường cong và làm giảm bề rộng. Hình 5.4 là một bản in Executat cho thấy ba phân phối xác suất chuẩn với những trung bình và độ lệch chuẩn khác nhau. Lưu ý những khác biệt về hình dạng và vị trí.

HÌNH 5.3 **Phân phối xác suất chuẩn**



HÌNH 5.4 **Bản in Minitab cho thấy các phân phối xác suất chuẩn với những giá trị khác nhau của μ và σ.**



Trên thực tế, chúng ta ít khi gặp phải những biến số mà trải dài từ các giá trị âm vô cùng lớn đến những giá trị dương lớn vô cùng. Tuy nhiên, nhiều biến số ngẫu nhiên dương (ví dụ như chiều cao, cân nặng và số lần) tạo ra một biểu đồ tần suất mà xấp xỉ rất gần với một phân phối chuẩn. Phép ước lượng xấp xỉ này áp dụng được bởi vì hầu như tất cả các giá trị của một biến số ngẫu nhiên chuẩn nằm trong phạm vi ba lần độ lệch chuẩn của giá trị trung bình, và trong những trường hợp này ($\mu \pm 3\sigma$) hầu như luôn luôn chứa đựng các giá trị dương.

5.3 CÁC DIỆN TÍCH TẠO THÀNH BẢNG CỦA PHÂN PHỐI XÁC XUẤT CHUẨN TẮC

Xác suất mà một biến số ngẫu nhiên liên tục có một giá trị trong khoảng từ *a* đến *b* là bằng với diện tích nằm bên dưới hàm mật độ xác suất giữa các điểm *a* và *b* (xem Hình 5.2). Tuy thế, bởi vì các đường cong chuẩn có những trung bình và độ lệch chuẩn khác nhau (xem Hình 5.4), nên chúng ta có thể tạo ra một số lượng lớn vô cùng các phân phối chuẩn. Từng bảng tính riêng cho những diện tích đối với mỗi trong số các đường cong này rõ ràng là không thực tế. Thay vào đó, chúng ta muốn tạo ra một qui trình chuẩn hóa mà sẽ cho phép chúng ta sử dụng cùng các khu vực đường cong chuẩn cho tất cả những phân phối chuẩn.

Sự chuẩn hóa được thực thi dễ dàng nhất bằng cách thể hiện giá trị của một biến số ngẫu nhiên chuẩn như số lượng các độ lệch chuẩn ở phía bên trái hay bên phải của giá trị trung bình. Nói cách khác, giá trị của một biến số ngẫu nhiên chuẩn x với trung bình μ và độ lệch chuẩn σ có thể được thể hiện như sau:

$$z = \frac{x - \mu}{\sigma}$$

hay, tương đương

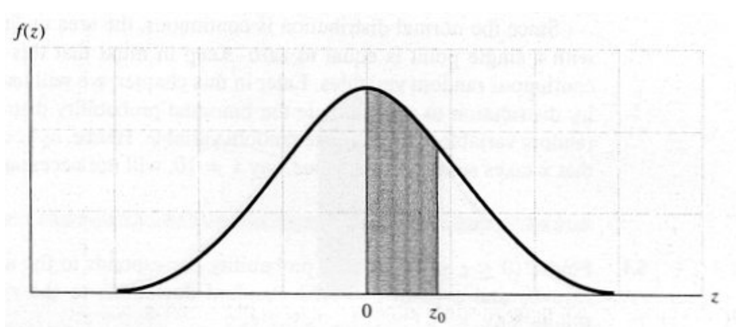
$$x = \mu + z\sigma$$

- Khi z là số âm, thì x nằm ở phía bên trái của trung bình μ.
- Khi $z = 0$, thì $x = \mu$.
- Khi z là số dương, thì x nằm ở phía bên phải của trung bình μ.

Chúng ta sẽ học cách tính toán các xác suất cho x bằng cách sử dụng $z = (x - \mu)/\sigma$, mà được gọi là **biến số ngẫu nhiên chuẩn chuẩn hóa.** Phân phối xác suất cho z được gọi là **phân phối chuẩn chuẩn hóa** bởi vì trung bình của nó bằng zêrô và độ lệch chuẩn của nó bằng 1. Điều này được trình bày trong Hình 5.5. Diện tích bên dưới đường cong chuẩn chuẩn hóa giữa trung bình $z = 0$ và một giá trị được xác định của z, ví dụ, z_0 là xác suất $P(0 \le z \le z_0)$. Diện tích này được trình bày trong Bảng 3 của Phụ lục II và được thể hiện là diện tích được tô đen trong Hình 5.5. Một phiên bản ngắn gọn của Bảng 3 trong Phụ lục II được thể hiện ở đây trong Bảng 5.1.

Bằng cách nào mà chúng ta có thể tìm ra các diện tích ở phía bên trái của trung bình? Bởi vì đường cong chuẩn chuẩn hóa là đối xứng với $z = 0$ (xem Hình 5.5), cho nên bất kỳ diện tích nào ở phía bên trái đều có thể được tìm ra bằng cách sử dụng diện tích tương đương ở phía bên phải của trung bình.

HÌNH 5.5 **Phân phối chuẩn chuẩn hóa**



BẢNG 5.1 **Phiên bản rút gọn của Bảng 3 trong Phụ lục II**

z_0	**0,00**	**0,01**	**0,02**	**0,03**	**0,04**	**0,05**	**0,06**	**0,07**	**0,08**	**0,09**
0,0	0,0000	0,0040	0,0080	0,0120	0,0160	0,1099	0,0239	0,0279	0,0319	0,0359
0,1	0,0398	0,0438	0,0478	0,0517	0,0557	0,0596	0,0636	0,0675	0,0714	0,0753
0,2	0,0793	0,0832	0,0871	0,0910	0,0948	0,0987	0,1026	0,1064	0,1103	0,1141
0,3	0,1179	0,1217	0,1255	0,1293	0,1331	0,1368	0,1406	0,1443	0,1480	0,1517
0,4	0,1554	0,1591	0,1628	0,1664	0,1700	0,1736	0,1772	0,1808	0,1844	0,1879
0,5	0,1915	0,1950	0,1985	0,2019	0,2054	0,2088	0,2123	**0,2157**	0,2190	0,224
0,6	0,2257									
0,7	**0,2580**									

1,0	**0,3413**
2,0	**0,4772**

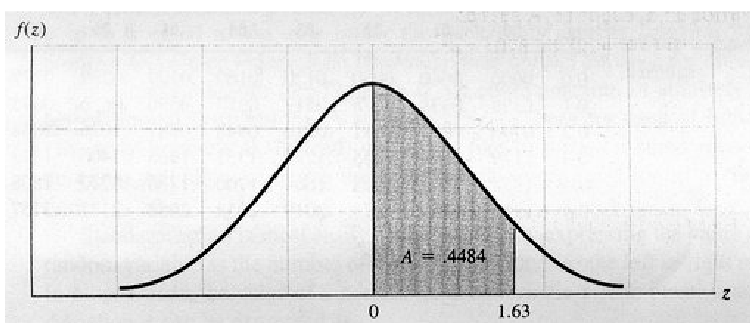
Lưu ý rằng *z*, bằng đúng với một phần mười gần nhất, được ghi vào cột phía bên trái của bảng này. Số thập phân thứ hai cho *z*, tương ứng với một phần trăm, được ghi lại ở hàng trên cùng của bảng. Như vậy, diện tích giữa trung bình và z = 0,7 lần độ lệch chuẩn về phía bên phải, được đọc từ cột thứ hai của bảng so với z = 0,7, được tìm thấy bằng với 0,2580. Tương tự như vậy, diện tích giữa trung bình và z = 1,0 là 0,3413. Diện tích giữa z = -1,0 và trung bình cũng là 0,3413. Vì thế diện tích nằm trong giới hạn một độ lệch chuẩn ở bất kỳ phía nào của trung bình sẽ là hai lần của 0,3413 hay bằng 0,6826. Diện tích nằm trong giới hạn hai độ lệch chuẩn của trung bình, chính xác đến bốn số thập phân, là 2 x 0,4772 = 0,9544. Những con số này nhất quán với các giá trị xấp xỉ gần đúng, 68% và 95%, được sử dụng trong Qui tắc Thực nghiệm trong Chương 2.

Để tìm ra diện tích nằm giữa trung bình và một điểm z = 0,57 độ lệch chuẩn về phía bên phải của trung bình, chúng ta tiến hành dò xuống cột phía bên trái đến hàng 0,5. Sau đó chúng ta chuyển sang hàng trên cùng của bảng đến cột 0,07. Giao điểm của sự kết hợp hàng-cột này cho ta diện tích thích hợp, 0,2157.

Bởi vì phân phối chuẩn là liên tục, cho nên diện tích nằm dưới đường cong này kết hợp với một điểm duy nhất là bằng zêrô. Hãy lưu ý rằng kết quả này chỉ áp dụng được cho các biến số ngẫu nhiên liên tục. Về sau trong chương này, chúng ta sẽ sử dụng phân phối xác suất chuẩn để ước lượng xấp xỉ phân phối xác suất nhị thức. Biến số ngẫu nhiên nhị thức *x* là một biến số ngẫu nhiên rời rạc. Như vậy, như các bạn đã biết, xác suất mà *x* có một giá trị cụ thể nào đó, ví dụ x = 10, sẽ không nhất thiết bằng zêrô.

VÍ DỤ **5.1** Tìm $P(0 \le z \le 1{,}63)$. Xác suất này tương ứng với diện tích giữa trung bình (z = 0) và một điểm z = 1,63 lần độ lệch chuẩn về phía bên phải của trung bình (xem Hình 5.6).

HÌNH 5.6 **Xác suất được yêu cầu trong Ví dụ 5.1**



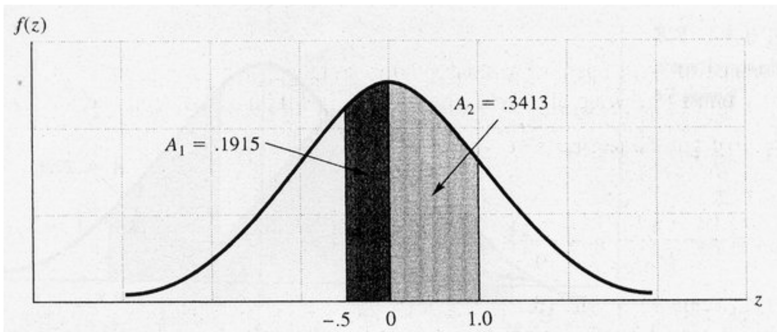
Lời giải Diện tích được bôi đen và chỉ ra bởi ký hiệu *A* trong Hình 5.6. Bởi vì Bảng 3 trong Phụ lục II cho chúng ta các diện tích bên dưới đường cong chuẩn về phía bên phải của trung bình, cho nên chúng ta chỉ cần tìm giá trị ghi trong bảng tương ứng với z = 1,63. Dò xuống cột phía bên trái của bảng đến z = 1,6 và dò ngang hàng trên cùng của bảng đến cột đánh số 0,03. Giao điểm của sự kết hợp hàng-cột này cho chúng ta diện tích A = 0,4484. Vì thế, $P(0 \le z \le 1{,}63) = 0{,}4484$.

VÍ DỤ **5.2** Tìm $P(-0{,}5 \le z \le 1{,}0)$. Xác suất này tương ứng với diện tích giữa z = -0,5 và z = 1,0, như được thể hiện trong Hình 5.7.

Lời giải Diện tích cần thiết bằng với tổng của A_1 và A_2 được thể hiện trong Hình 5.7. Từ Bảng 3 trong Phụ lục II chúng ta đọc thấy $A_2 = 0{,}3413$. Diện tích A_1 bằng với diện tích tương ứng giữa $z = 0$ và $z = 0{,}5$ hay $A_1 = 0{,}1915$. Như vậy, tổng diện tích là:

$$A = A_1 + A_2 = 0{,}1915 + 0{,}3413 = 0{,}5328$$

HÌNH 5.7 **Diện tích bên dưới đường cong chuẩn trong Ví dụ 5.2**



VÍ DỤ **5.3** Tìm giá trị của *z*, ví dụ z_0, để cho chính xác (ở mức bốn số thập phân) 0,95 của diện tích nằm trong giới hạn $\pm z_0$ lần độ lệch chuẩn của trung bình.

Lời giải Một nửa của diện tích 0,95 sẽ nằm về phía bên trái của trung bình và một nửa nằm về phía bên phải bởi vì phân phối chuẩn là đối xứng. Vì thế, chúng ta muốn tìm ra giá trị z_0 tương ứng với một diện tích bằng với 0,475. Tham chiếu Bảng 3 trong Phụ lục II, chúng ta thấy rằng diện tích 0,475 rơi vào hàng tương ứng với $z = 1{,}9$ và cột 0,06. Do đó, $z_0 = 1{,}96$. Lưu ý rằng kết quả này là rất gần với giá trị xấp xỉ gần đúng, $z = 2$ mà được sử dụng trong Qui tắc Thực nghiệm.

VÍ DỤ **5.4** Cho *x* là một biến số ngẫu nhiên phân phối chuẩn với trung bình bằng 10 và độ lệch chuẩn bằng 2. Hãy tìm xác suất để cho *x* nằm trong khoảng 11 đến 13,6.

Lời giải Như là bước đầu tiên, chúng ta phải tính toán các giá trị của *z* tương ứng với $x = 11$ và $x = 13{,}6$. Như vậy:

$$z_1 = \frac{x_1 - \mu}{\sigma} = \frac{11-10}{2} = 0{,}5 \qquad z_2 = \frac{x_2 - \mu}{\sigma} = \frac{13{,}6-10}{2} = 1{,}8$$

Xác suất mong muốn vì vậy là $P(0{,}5 \le z \le 1{,}8)$ và là diện tích nằm giữa z_1 và z_2, như được thể hiện trong Hình 5.8. Diện tích giữa $z = 0$ và z_1 là $A_1 = 0{,}1915$, và diện tích nằm giữa $z = 0$ và z_2 là $A_2 = 0{,}4641$; chúng ta có được những diện tích này từ Bảng 3. Xác suất mong muốn bằng với chênh lệch giữa A_2 và A_1; nghĩa là,

$$P(0{,}5 \le z \le 1{,}8) = A_2 - A_1 = 0{,}4641 - 0{,}1915 = 0{,}2726$$

VÍ DỤ **5.5** Các nghiên cứu chứng tỏ rằng việc sử dụng nhiên liệu cho các xe hơi cỡ nhỏ bán tại Hoa Kỳ là có phân phối chuẩn, với mức sử dụng trung bình là 30,5 dặm mỗi galông nhiên liệu (mpg) và một độ lệch chuẩn là 4,5 mpg. Tỷ lệ phần trăm của xe cỡ nhỏ đạt mức 35 mpg hay cao hơn là bao nhiêu?

Lời giải Tỷ lệ của xe cỡ nhỏ đạt được mức 35 mpg hay cao hơn được cho bởi diện tích bôi đen trong Hình 5.9.

Chúng ta phải tìm ra giá trị *z* tương ứng với $x = 35$. Thay thế vào công thức tính *z*, chúng ta có:

$$z = \frac{x-\mu}{\sigma} = \frac{35-30,5}{4,5} = 1,0$$

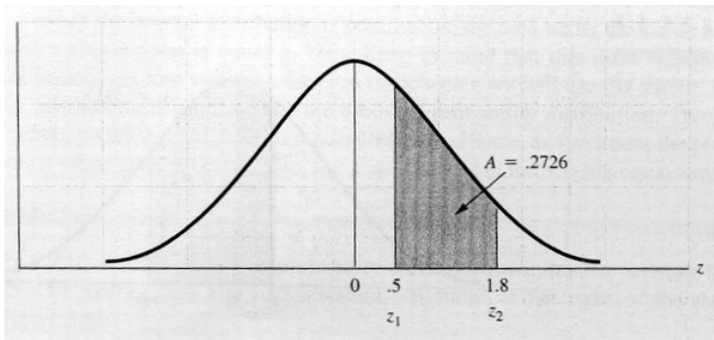
Diện tích *A* nằm về phía bên phải của trung bình, tương ứng với z = 1,0 là 0,3413 (từ Bảng 3). Sau đó tỷ lệ phần trăm xe cỡ nhỏ có hệ số mpg bằng hay lớn hơn 35 là bằng với toàn bộ diện tích nằm về phía bên phải của trung bình, 0,5, trừ đi cho diện tích *A*:

$$P(x \geq 35) = 0,5 - P(0 \leq z \leq 1) = 0,5 - 0,3413 = 0,1587$$

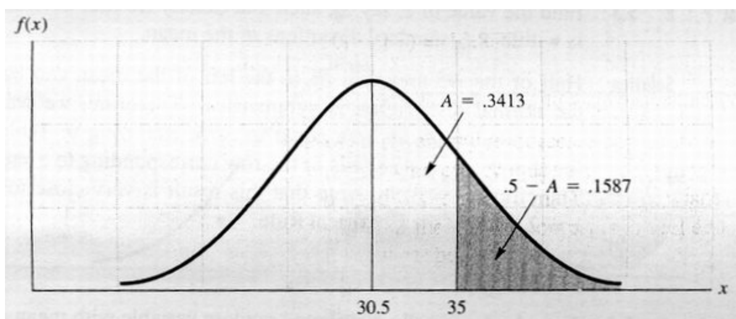
Tỷ lệ phần trăm vượt mức 35 mpg là:

100 (0,1587) = 15,87%

HÌNH 5.8 **Diện tích bên dưới đường cong chuẩn trong Ví dụ 5.4**

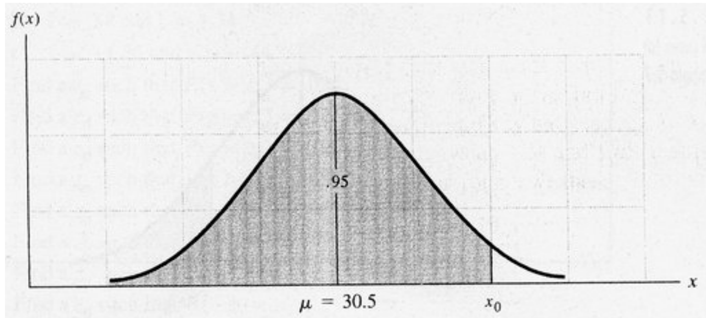


HÌNH 5.9 **Diện tích bên dưới đường cong chuẩn trong Ví dụ 5.5**



VÍ DỤ **5.6** Trở lại Ví dụ 5,5. Trong những thời điểm khan hiếm nguồn năng lượng, thì một nhà sản xuất xe hơi mà có thể chế tạo ra một chiếc xe đạt được mức tiêu hao nhiên liệu tiết kiệm hơn so với xe của các đối thủ cạnh tranh khác sẽ có lợi thế cạnh tranh. Nếu một nhà sản xuất mong muốn phát triển một chiếc xe cỡ nhỏ mà vượt trội hơn 95% các xe cỡ nhỏ hiện hành về phương diện tiết kiệm nhiên liệu, thì mức sử dụng nhiên liệu của chiếc xe mới đó sẽ là bao nhiêu?

HÌNH 5.10 **Vị trí của x_0 để cho** $P(x < x_0 = 0{,}95)$



Lời giải Đặt *x* là một biến số ngẫu nhiên phân phối chuẩn, với trung bình là 30,5 và độ lệch chuẩn là 4,5. Như được biển diễn trong Hình 5.10, chúng ta muốn tìm ra giá trị x_0 để cho:

$$P(x \le x_0) = 0{,}95$$

Bước đầu tiên, chúng ta tìm

$$z_0 = \frac{x_0 - \mu}{\sigma} = \frac{x_0 - 30{,}5}{4{,}5}$$

và lưu ý rằng xác suất mong muốn của chúng ta là giống như diện tích nằm về phía bên trái của z_0 đối với phân phối chuẩn chuẩn hóa. Như vậy,

$$P(z \le z_0) = 0{,}95$$

Diện tích nằm ở phía bên trái của trung bình là 0,5. Diện tích nằm ở phía bên phải của trung bình giữa z_0 và trung bình là 0,95 - 0,5 = 0,45. Như thế, từ Bảng 3, chúng ta tìm thấy rằng z_0 là giữa 1,64 và 1,65. Lưu ý rằng diện tích 0,45 chính xác nằm ở giữa các diện tích đối với z = 1,64 và z = 1,65; nghĩa là z_0 = 1,645.

Thay thế z_0 = 1,645 vào phương trình cho z_0 chúng ta có:

$$1{,}645 = \frac{x_0 - 30{,}5}{4{,}5}$$

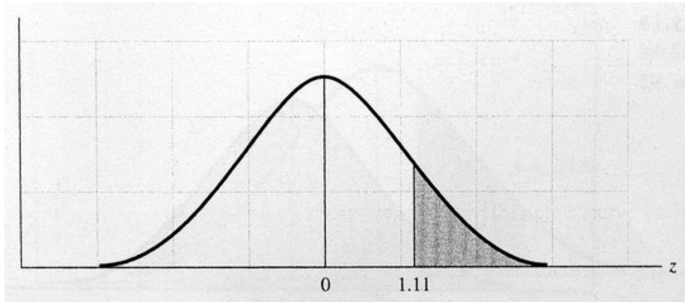
Từ đó, chúng ta có được

$$x_0 = (1{,}645)/(4{,}5) + 30{,}4 = 37{,}9$$

Chiếc xe cỡ nhỏ mới của nhà sản xuất này vì thế phải đạt được mức tiêu thụ nhiên liệu là 37,9 để có thể vượt qua 95% các xe cỡ nhỏ hiện có tại thị trường Hoa Kỳ về phương diện tiết kiệm nhiên liệu.

VÍ DỤ **5.7** Mức lương của những người tốt nghiệp Thạc sĩ Quản trị Kinh doanh (MBA) mà tham gia vào lĩnh vực dịch vụ tiếp thị bình quân xấp xỉ 45.000 USD, với độ lệch chuẩn là 2.250 USD. Nếu các mức lương này được phân phối chuẩn, thì tỷ lệ phần trăm của những người tốt nghiệp MBA mà tham gia vào lĩnh vực dịch vụ tiếp thị có mức lương vượt quá 47.500 USD, mà là mức lương trung bình cho những người tốt nghiệp đại học tham gia vào lĩnh vực quản trị nhãn hiệu/sản phẩm, là bao nhiêu?

HÌNH 5.11 **Diện tích bên dưới đường cong chuẩn cho Ví dụ 5.7**



Lời giải Nhằm tìm ra tỷ lệ phần trăm những người tốt nghiệp MBA mà có mức lương vượt quá 47.500 USD, chúng ta cần có giá trị của *z* tương ứng với mức 47.500 USD. Với μ = 45.000 USD và σ = 2.250 USD,

$$z = \frac{x-\mu}{\sigma} = \frac{47.500 - 45.00}{2.250} = 1{,}11$$

Tiếp đến, chúng ta cần tìm ra diện tích nằm dưới một đường cong chuẩn về bên phải của *z* = 1,1, như được biểu diễn trong Hình 5.11. Như vậy, diện tích cần thiết là bằng với 0,5, tổng diện tích về phía bên phải của zêrô, *trừ đi* diện tích giữa 0 và 1,1.

$$P(x \geq 47.500) = 0{,}5 - 0{,}3665 = 0{,}1335$$

Do đó, 13,35% những người tốt nghiệp MBA có mức lương vượt quá 47.500 USD.

Bài tập

Các Kỹ thuật Cơ bản

5.1 Sử dụng Bảng 3 trong Phụ lục II, tính toán diện tích nằm dưới đường cong chuẩn giữa những giá trị *z* này.

a *z* = 0 và *z* = 1,6

b *z* = 0 và *z* = 1,83

5.2 Lặp lại Bài tập 5.1 cho những giá trị *z* sau.

a *z* = 0 và *z* = 0,90

b *z* = 0 và *z* = -0,90

5.3 Lặp lại Bài tập 5.1 cho những giá trị *z* sau.

a *z* = -1,3 và *z* = 1,8

b *z* = 0,6 và *z* = 1,2

5.4 Lặp lại Bài tập 5.1 cho những giá trị *z* sau.

a *z* = -1,4 và *z* = 1,4

b *z* = -2,0 và *z* = 2,0

c *z* = -3,0 và *z* = 3,0

5.5 Lặp lại Bài tập 5.1 cho những giá trị *z* sau.

a *z* = -1,43 và *z* = 0,68

b z = 0,58 và z = 1,74

c z = -1,55 và z = -0,44

5.6 Tìm một giá trị z_0 để cho $P(z > z_0) = 0{,}025$

5.7 Tìm một giá trị z_0 để cho $P(z < z_0) = 0{,}9251$

5.8 Tìm một giá trị z_0 để cho $P(z < z_0) = 0{,}2981$

5.9 Tìm một giá trị z_0 để cho $P(z > z_0) = 0{,}6985$

5.10 Tìm một giá trị z_0 để cho $P(-z_0 < z < z_0) = 0{,}4714$

5.11 Tìm một giá trị z_0 để cho $P(z < z_0) = 0{,}05$

5.12 Tìm một giá trị z_0 để cho $P(-z_0 < z < z_0) = 0{,}90$

5.13 Tìm một giá trị z_0 để cho $P(-z_0 < z < z_0) = 0{,}99$

5.14 Một biến số z được phân phối chuẩn với trung bình μ = 10 và độ lệch chuẩn σ = 2. Tìm những xác suất này.

a $P(x > 13{,}5)$

b $P(x < 8{,}2)$

c $P(9{,}4 < x < 10{,}6)$

5.15 Một biến số x được phân phối chuẩn với trung bình μ = 1,20 và độ lệch chuẩn σ = 0,15. Tìm xác suất để cho x rơi vào khoảng cho trước này.

a 1,00 < x < 1,10

b x > 1,38

c 1,35 < x < 1,50

5.16 Một biến số x được phân phối chuẩn với trung bình μ chưa biết và độ lệch chuẩn σ = 2. Nếu xác suất để cho x vượt quá 7,5 là 0,8023, hãy tìm μ.

5.17 Một biến số x được phân phối chuẩn với trung bình μ chưa biết và độ lệch chuẩn σ = 1,8. Nếu xác suất để cho x vượt quá 14,4 là 0,3, hãy tìm μ.

5.18 Một biến số x được phân phối chuẩn với trung bình và độ lệch chuẩn chưa biết. Xác suất để cho x vượt quá 4 là 0,9772, và xác suất để cho x vượt quá 5 là 0,9332. Hãy tìm μ và σ.

Các Ứng dụng

5.19 Một phương pháp để đi đến các dự báo kinh tế là sử dụng một cách tiếp cận chuyên gia. Một dự báo có được từ mỗi trong số một số lượng lớn các nhà phân tích; trung bình của những dự báo của mỗi cá nhân này là dự báo đồng thuận. Giả định rằng các dự báo về lãi suất cơ bản tháng Giêng của mỗi cá nhân của tất cả các nhà phân tích kinh tế được phân phối xấp xỉ chuẩn với trung bình bằng với 7,75% và một độ lệch chuẩn bằng với 1,6%. Một nhà phân tích duy nhất được lựa chọn ngẫu nhiên từ trong nhóm này.

a. Xác suất để cho dự báo của nhà phân tích này về lãi suất cơ bản sẽ vượt quá 9% là bao nhiêu?

b. Xác suất để cho dự báo của nhà phân tích này về lãi suất cơ bản sẽ thấp hơn 6% là bao nhiêu?

5.20 Giả định rằng bạn phải thiết lập các qui định liên quan đến số lượng người tối đa mà có thể chứa trong một cái thang máy. Một nghiên cứu về khả năng chứa của thang máy chỉ ra rằng nếu tám người có trong thang máy, thì phân phối xác suất của tổng trọng lượng của tám người này có một giá trị trung bình là 1.200 pao và một phương sai tương đương 9.800 pao. Xác suất để cho tổng trọng lượng của tám người lớn hơn 1.300 pao là bao nhiêu? 1.500 pao là bao nhiêu? (Giả định rằng phân phối xác suất là xấp xỉ chuẩn).

5.21 Việc thải ra các chất rắn lơ lửng từ một mỏ phốt-phát có phân phối chuẩn, với một mức thải trung bình hàng ngày là 27 mg/l và một độ lệch chuẩn là 14 mg/l. Tỷ lệ phần trăm những ngày mà lượng chất thải hàng ngày vượt quá 50 mg/l là bao nhiêu?

5.22 Những người sưu tầm tem thường mua tem ở mức giá bán lẻ hay gần như vậy, nhưng khi họ bán thì giá này thấp hơn nhiều. Ví dụ, có thể là một điều hợp lý khi giả định rằng (tùy thuộc vào bộ sưu tập, tình trạng của nó, nhu cầu, điều kiện kinh tế, v.v) một bộ sưu tập có thể được kỳ vọng có thể bán ở mức x phần trăm của giá bán lẻ, trong đó x được phân phối chuẩn với trung bình bằng với 45% và độ lệch chuẩn bằng với 4,5%. Một nhà sưu tầm tem có một bộ sưu tập để bán mà có một giá trị bán lẻ là 30.000 USD.

a. Xác suất để cho nhà sưu tầm tem này nhận được nhiều hơn 15.000 USD cho bộ sưu tập này là bao nhiêu?

b. Xác suất để cho nhà sưu tầm tem này nhận được ít hơn 15.000 USD cho bộ sưu tập này là bao nhiêu?

c. Xác suất để cho nhà sưu tầm tem này nhận được ít hơn 12.000 USD cho bộ sưu tập này là bao nhiêu?

5.23 Bằng cách nào mà Sở Thuế Nội Bộ (IRS) quyết định về tỷ lệ phần trăm của số thu thuế thu nhập để kiểm toán cho từng tiểu bang? Giả định rằng IRS thực hiện điều này bằng cách chọn ngẫu nhiên 50 giá trị từ một phân phối chuẩn với trung bình bằng với 1,55% và độ lệch chuẩn tương đương 0,45%. (Chương trình máy tính là sẵn có cho loại hình chọn mẫu này).

a. Xác suất để cho một tiểu bang cụ thể sẽ có nhiều hơn 2,5% số thu thuế thu nhập của tiểu bang mình được kiểm toán?

b. Xác suất để cho một tiểu bang cụ thể sẽ có ít hơn 1% số thu thuế thu nhập của tiểu bang mình được kiểm toán?

5.24 Trong một nỗ lực để đẩy mạnh chất lượng sản xuất của các công nhân người Mỹ của mình, công ty Saturn Corporation đang thưởng cho nhân công của mình một khoản bình quân 2.800 USD tiền thưởng vào cuối năm cho việc đáp ứng sản xuất đạt chất lượng và mục tiêu lợi nhuận trong năm 1993 ("Saturn Workers (*Công nhân Saturn*)", 1994). Giả định rằng những khoản thưởng này có phân phối xấp xỉ chuẩn với một độ lệch chuẩn là 500 USD.

a. Xác suất để cho một công nhân nhận được một khoản tiền thưởng cuối năm nhiều hơn 3.500 USD là bao nhiêu?

b. Chín mươi lăm phần trăm tất cả công nhân sẽ nhận được khoản tiền thưởng cuối năm nằm trong những giới hạn nào?

5.25 Người tiêu dùng Hoa Kỳ đang trở nên ngày càng quan tâm hơn đến phí tổn của nhiên liệu cho việc sưởi ấm. Khi những chi phí này gia tăng, người tiêu dùng nói chung cân nhắc các nhiên liệu thay thế, những cải tiến việc cách nhiệt của ngôi nhà, và những hệ thống sưởi ấm mới. Giả định rằng phí tổn của khí tự nhiên mỗi bộ khối (MCF) có phân phối chuẩn với giá trị trung bình là 6 USD và một độ lệch chuẩn tương đương 1.20 USD.

a. Xác suất để cho phí tổn của khí thiên nhiên mỗi MCF cho một người tiêu dùng cụ thể nằm trong khoảng 7.60 đến 8.00 USD là bao nhiêu?

b. Phí tổn trung vị mỗi MCF cho khí thiên nhiên là bao nhiêu?

c. Các phân vị một phần tư trên và dưới cho phí tổn mỗi MCF khí thiên nhiên là bao nhiêu?

5.4 ƯỚC LƯỢNG XẤP XỈ CHUẨN CHO PHÂN PHỐI XÁC SUẤT NHỊ THỨC

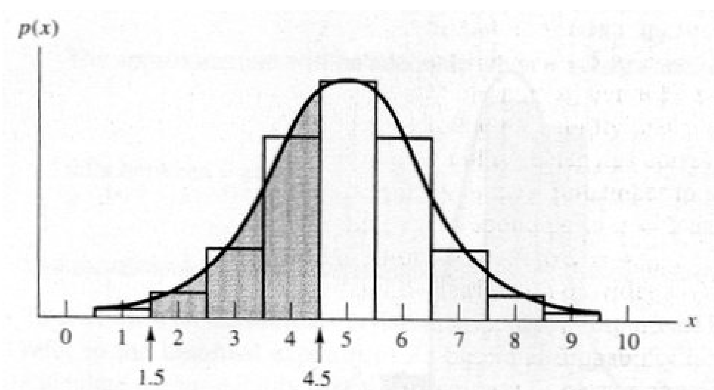
Nhiều phân phối xác suất sở hữu một đặc trưng hữu ích. Khi một số điều kiện được thỏa mãn, thì những phân phối này trở nên xấp xỉ chuẩn về hình dạng. Phân phối xác suất nhị thức là một trong số này. Cụ thể là, khi một số *n* các lần thử trong một thí nghiệm nhị thức là lớn và *p* không quá gần với 0 hay 1, thì phân phối xác suất nhị thức có một hình dạng mà xấp xỉ gần đúng với một đường cong chuẩn với trung bình $\mu = np$ và độ lệch chuẩn $\sigma = \sqrt{npq}$. Tính chất đặc biệt này của phân phối xác suất nhị thức là quan trọng khi chúng ta phải tính toán các xác suất nhị thức *p(x)* cho những giá trị lớn của *n*. Công việc tẻ ngắt và tốn công gặp phải trong những sự tính toán này có thể tránh được bằng cách sử dụng đường cong xấp xỉ chuẩn.

Bởi vì cách thức tốt nhất để chứng tỏ bằng cách nào và tại sao việc ước lượng xấp xỉ chuẩn vận hành được là sử dụng các đồ thị và một giá trị nhỏ của *n*, cho nên chúng ta sẽ minh họa quá trình này cho một phân phối xác suất nhị thức với $n = 10$ và $p = 1/2$. Biểu đồ xác suất cho một phân phối xác suất nhị thức, $n = 10$ và $p = 1/2$, được biểu diễn trong Hình 5.12 cùng với một đường cong xấp xỉ chuẩn với

$$\mu = np = 10\left(\frac{1}{2}\right) = 5$$

$$\sigma = \sqrt{npq} = \sqrt{(10)\left(\frac{1}{2}\right)\left(\frac{1}{2}\right)} = \sqrt{2{,}5} = 1{,}58$$

HÌNH 5.12 **So sánh giữa phân phối xác suất nhị thức và phân phối xấp xỉ chuẩn,**
$n = 10, p = 1/2 (\mu = np = 5); \sigma = \sqrt{npq} = 1{,}58$



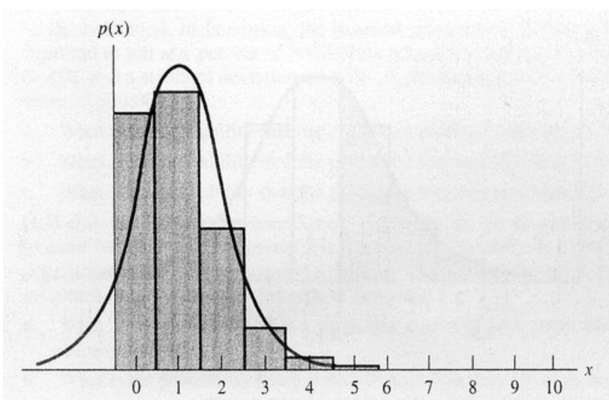
Một khảo sát bằng mắt của hình này gợi ý rằng sự ước lượng xấp xỉ là khá tốt, mặc dù một mẫu nhỏ, $n = 10$, là cần thiết cho sự minh họa bằng đồ thị này.

Giả định rằng chúng ta mong muốn ước lượng xấp xỉ xác suất để cho *x* bằng với 2, 3 hay 4. Bạn có thể thấy trong Hình 5.12 rằng xác suất này bằng đúng với diện tích của ba hình chữ nhật nằm vắt qua $x = 2$, 3 và 4. Chúng ta có thể ước lượng xấp xỉ phân phối này với diện tích nằm dưới đường cong chuẩn từ $x = 1{,}5$ đến $x = 4{,}5$, mà được bôi đen trong Hình 5.12. Lưu ý rằng diện tích nằm dưới đường cong chuẩn từ $x = 2$, 3 và 4 *ắt sẽ không phải* là một sự ước lượng xấp xỉ tốt cho xác suất để cho $x = 2$, 3 và 5 bởi vì công việc này ắt sẽ loại trừ một nửa của các hình chữ nhật xác suất tương ứng với $x = 2$ và $x = 4$. Để có được một sự ước lượng xấp xỉ tốt, bạn phải nhớ ước lượng xấp xỉ toàn bộ các diện tích của những hình chữ nhật xác suất tương ứng với $x = 2$ và $x = 4$ bằng cách thêm vào diện tích nằm dưới đường cong chuẩn từ $x = 1{,}5$ đến $x = 4{,}5$.

Mặc dù phân phối xác suất chuẩn cung cấp cho ta một sự ước lượng xấp xỉ chuẩn tốt cho phân phối xác suất nhị thức trong Hình 5.12, thì điều này không phải luôn luôn đúng. Khi trung bình *np* của một phân phối xác suất nhị thức là gần với zêrô hay *n*, thi phân phối xác suất nhị thức sẽ không có tính đối xứng.[+] Ví dụ, khi *p* là gần với zêrô, thì phần lớn các giá trị của *x* sẽ là nhỏ, qua đó tạo rao một sự phân phối mà được tập trung gần $x = 0$ và có phần đuôi hướng về *n* (xem Hình 5.13). Ắt hẳn là, khi điều này là đúng, thì phân phối chuẩn, đối xứng và có hình dạng quả chuông, sẽ tạo ra một sự ước lượng xấp xỉ tồi cho phân phối xác suất nhị thức. Vậy thì, bằng cách nào mà chúng ta có thể nói rằng liệu *n* và *p* là như vậy để cho phân phối nhị thức là đối xứng?

Nhắc lại Qui tắc Thực nghiệm từ Chương 2, xấp xỉ 95% của những sự đo lường đi cùng với một phân phối chuẩn sẽ nằm trong giới hạn hai độ lệch chuẩn của trung bình và hầu như tất cả đều nằm trong giới hạn ba độ lệch chuẩn. Chúng ta ngờ rằng phân phối xác suất nhị thức ắt sẽ gần như đối xứng nếu như phân phối này có khả năng trải dài trên một khoảng cách bằng với hai độ lệch chuẩn trên bất cứ phía nào của trung bình, và trên thực tế thì điều này là đúng. **Vì vậy, để quyết định khi nào thì sự ước lượng xấp xỉ chuẩn sẽ phù hợp, hãy tính toán $\mu = np$ và $\sigma = \sqrt{npq}$. Nếu khoảng $(\mu \pm 2\sigma)$ nằm bên trong các giới hạn nhị thức, 0 và *n*, thì sự ước lượng xấp xỉ sẽ phù hợp. Sự ước lượng xấp xỉ sẽ tốt nếu khoảng $(\mu \pm 3\sigma)$ nằm trong giới hạn của khoảng từ 0 đến *n*.** Lưu ý rằng chỉ tiêu này được thỏa mãn đối với phân phối xác suất nhị thức của Hình 5.2, nhưng không được thỏa mãn cho phân phối được thể hiện trong Hình 5.13.

HÌNH 5.13 **So sánh một phân phối xác suất nhị thức (được bôi đen) và phân phối xấp xỉ chuẩn,** $n = 10, p = 0{,}1 (\mu = np = 1; \sigma = \sqrt{npq} = 0{,}95)$.



Các công thức cho **ước lượng xấp xỉ chuẩn của sự phân phối xác suất nhị thức** được cung cấp trong phần trình bày dưới đây.

Sự Ước lượng Xấp xỉ Chuẩn cho Phân phối Xác suất Nhị thức

Ước lượng xấp xỉ phân phối xác suất nhị thức bằng cách sử dụng một đường cong chuẩn với

[+] Một phân phối xác suất nhị thức bị lệch có thể được ước lượng xấp xỉ bởi một phân phối xác suất Poisson. Sự ước lượng xấp xỉ này, mà được thảo luận trong Phần 4.3, là thỏa mãn khi *n* là lớn và *p* là nhỏ, ví dụ $np < 7$.

$$\mu = np$$

$$\sigma = \sqrt{npq}$$

trong đó n = số lượng các lần thử; p = xác suất của thành công trong một lần thử duy nhất; $q = 1 - p$.

Ước lượng xấp xỉ sẽ phù hợp khi n lớn và khi khoảng $\mu = np$ rơi vào giữa 0 và n.

VÍ DỤ **5.8** Để xem đường cong chuẩn có thể được sử dụng tốt như thế nào trong việc ước lượng xấp xỉ các xác suất nhị thức, hãy xem lại thí nghiệm nhị thức đã được minh họa trong Hình 5.12, với n = 10, p = 0,5. Tính xác suất để cho x= 2, 3 hay 4, hiệu chỉnh về ba con số thập phân, bằng cách sử dụng Bảng 1 trong Phụ lục II. Sau đó tính toán ước lượng xấp xỉ chuẩn tương ứng cho xác suất này.

Lời giải Xác suất chính xác này có thể được tính toán với n = 10 bằng cách sử dụng Bảng 1 trong Phụ lục II. Như vậy,

$$\sum_{x=2}^{4} p(x) = \sum_{x=0}^{4} p(x) - \sum_{x=0}^{1} p(x) = 0{,}377 - 0{,}011 = 0{,}366$$

Trong Hình 5.12, các hình chữ nhật xác suất nhị thức cho x = 2, 3 và 4 tương ứng với diện tích giữa x_1 = 1,5 và x_2 = 4,5 bên dưới đường cong xấp xỉ chuẩn tương ứng với μ = 1,5 và σ = 1,58. Các giá trị tương ứng của z là:

$$z_1 = \frac{x_1 - \mu}{\sigma} = \frac{1{,}5 - 5}{1{,}58} = -2{,}22$$

$$z_2 = \frac{x_2 - \mu}{\sigma} = \frac{4{,}5 - 5}{1{,}58} = -0{,}32$$

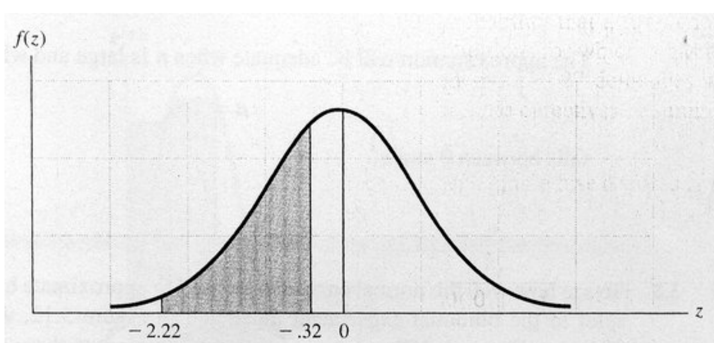
Xác suất này được thể hiện trong Hình 5.14. Diện tích giữa z = 0 và z = 2,22 là A_1 = 0,4868. Tương tự như vậy, diện tích giữa z = 0 và z = 0,32 là A_2 = 0,1255. Từ Hình 5.14,

$$P(-2{,}22 \le z \le 0{,}32) = 0{,}4868 - 0{,}1255 = 0{,}3613$$

Lưu ý rằng sự ước lượng xấp xỉ chuẩn là rất gần với phân phối nhị thức chính xác, 0,366 mà có được từ Bảng 1.

Bạn phải cẩn trọng để không loại trừ một nửa của hai hình chữ nhật xác suất ở hai thái cực khi sử dụng ước lượng xấp xỉ chuẩn đối với phân phối xác suất nhị thức. **Các giá trị *x* được sử dụng để tính toán những giá trị *z* luôn luôn kết thúc ở 0,5.**

HÌNH 5.14 **Diện tích được yêu cầu cho Ví dụ 5.8**



Điều chỉnh cho tính Liên tục

Qui trình của việc cộng thêm hay trừ bớt 0,5 trong việc điều chỉnh các giá trị của *x* cho phân phối nhị thức đối với các giá trị cho việc ước lượng xấp xỉ phân phối chuẩn được gọi là **sự điều chỉnh cho tính liên tục.**

Để chắc chắn rằng bạn đã thực hiện việc điều chỉnh phù hợp cho sự liên tục, hãy luôn luôn vẽ ra một phác thảo nháp tương tự như Hình 5.12.

VÍ DỤ **5.9** Độ tin cậy của một cầu chì điện là xác suất để cho cầu chì đó, được chọn ngẫu nhiên từ số sản phẩm sản xuất ra, sẽ hoạt động được trong những điều kiện mà qua đó nó được thiết kế. Một mẫu ngẫu nhiên gồm 1.000 cầu chì được kiểm tra và $x = 27$ cầu chì có lỗi được quan sát. Tính xác suất của việc quan sát thấy 27 hay nhiều hơn số cầu chì bị lỗi, bằng cách giả định rằng độ tin cậy của cầu chì là 0,98.

Lời giải Xác suất của việc quan sát một sản phẩm bị lỗi khi một cầu chì duy nhất được kiểm tra là $p = 0{,}02$, khi đã biết độ tin cậy của cầu chì là 0,98. Sau đó

$$\mu = np = 1000(0{,}02) = 20$$
$$\sigma = \sqrt{npq} = \sqrt{1000(0{,}02)(0{,}98)} = 4{,}43$$

Xác suất của 27 hay nhiều hơn số cầu chì bị lỗi, khi đã biết $n = 1000$ là

$$\begin{aligned} P &= P(x \geq 27) \\ &= p(27) + p(28) + p(29) + ... + p(999) = p(1000) \end{aligned}$$

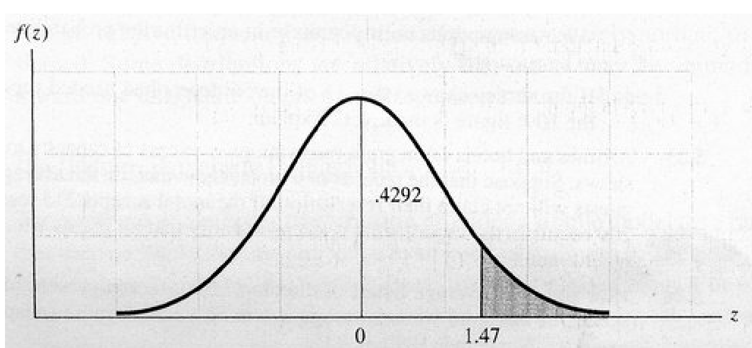
Ước lượng xấp xỉ chuẩn của *P* là diện tích nằm dưới đường cong chuẩn về phía bên phải của $x = 26{,}5$. Khi thực hiện sự điều chỉnh cho tính liên tục, chúng ta phải sử dụng $x = 26{,}5$ hơn là $x = 27$ nhằm để thêm vào toàn bộ hình chữ nhật xác suất đi cùng với $x = 27$. Giá trị *z* tương ứng với $x = 26.5$ là:

$$z = \frac{x-\mu}{\sigma} = \frac{26{,}5-20}{4{,}43} = \frac{6{,}5}{4{,}43} = 1{,}47$$

và diện tích giữa $z = 0$ và $z = 1{,}47$ là bằng với 0,4292, như được thể hiện trong Hình 5.15. Bởi vì toàn bộ diện tích về phía bên phải của trung bình là bằng với 0,5, cho nên

$$P(x \geq 27) \approx 0{,}5 - 0{,}4292 = 0{,}0708$$

HÌNH 5.15 **Ước lượng xấp xỉ chuẩn cho phân phối nhị thức trong Ví dụ 5.9**



Bài tập

Các Kỹ thuật Cơ bản

5.26 Đặt x là một biến số ngẫu nhiên nhị thức với $n = 25$, $p = 0{,}3$.

a. Sử dụng Bảng 1 trong Phụ lục II để tìm ra $P(8 \le x \le 10)$.

b. Tìm μ và σ cho phân phối nhị thức, và sử dụng ước lượng xấp xỉ chuẩn để tìm ra $P(8 \le x \le 10)$. So sánh sự ước lượng xấp xỉ này với giá trị chính xác được tính trong phần (a).

5.27 Tìm ước lượng xấp xỉ chuẩn cho $P(x \ge 6)$ cho phân phối xác suất nhị thức với $n = 10$, $p = 5$.

5.28 Tìm ước lượng xấp xỉ chuẩn cho $P(x > 6)$ cho phân phối xác suất nhị thức với $n = 10$, $p = 5$.

5.29 Tìm ước lượng xấp xỉ chuẩn cho $P(x > 22)$ cho phân phối xác suất nhị thức với $n = 100$, $p = 0{,}2$.

5.30 Tìm ước lượng xấp xỉ chuẩn cho $P(x \ge 22)$ cho phân phối xác suất nhị thức với $n = 100$, $p = 0{,}2$.

5.31 Đặt x là một biến số ngẫu nhiên nhị thức với $n = 25$, $p = 0{,}2$.

a. Sử dụng Bảng 1 trong Phụ lục II để tính $P(4 \le x \le 6)$.

b. Tìm μ và σ cho phân phối nhị thức, và sử dụng ước lượng xấp xỉ chuẩn để tìm ra $P(4 \le x \le 6)$. Lưu ý rằng giá trị này là một sự ước lượng xấp xỉ tốt cho giá trị chính xác của $P(4 \le x \le 6)$.

5.32 Xem xét một thí nghiệm nhị thức với $n = 20$, $p = 0{,}4$. Hãy tính $P(x \ge 10)$ bằng cách sử dụng:

a. Bảng 1 trong Phụ lục II.

b. Ước lượng xấp xỉ chuẩn cho phân phối xác suất nhị thức.

5.33 Tìm ước lượng xấp xỉ chuẩn cho $P(355 \le x \le 360)$ cho một phân phối xác suất nhị thức với $n = 400$ và $p = 0{,}9$.

Các Ứng dụng

5.34 Một số nhãn hiệu của các đồ gia dụng chính yếu là đáng tin cậy hơn những nhãn hiệu khác. Ví dụ, xấp xỉ 10% máy sấy Maytag được mua trong giai đoạn 1986 đến 1992 chưa bao giờ cần đến sự sửa chữa ("Getting Things Fixed (*Sửa chữa Đồ đạc*)", 1994). Giả định rằng một nhóm người tiêu dùng điều tra 56 người sở hữu máy sấy Maytag.

a. Xác suất để cho mười hay nhiều hơn máy sấy chưa bao giờ cần đến sự sửa chữa là bao nhiêu?

b. Xác suất để cho có ít hơn năm máy sấy chưa bao giờ cần đến sự sửa chữa là bao nhiêu?

c. Giả định nào mà bạn cần thực hiện để cho các xác suất tìm thấy trong các phần (a) và (b) là chính xác?

d. Nếu cuộc điều tra này cho thấy rằng 15 trong số 56 máy sấy chưa bao giờ cần đến sự sửa chữa, liệu bạn có nghi ngờ rằng con số 10% là không chính xác không? Hãy giải thích.

5.35 Các hãng hàng không và khách sạn thường chấp thuận việc đặt phòng trước vượt quá năng lực phòng nhằm giảm thiểu những tổn thất do việc đã đặt phòng nhưng không sử dụng. Giả định rằng ghi nhận của một khách sạn dọc đường cho thấy rằng, tính trung bình thì có 10% khách sắp đến của họ sẽ không yêu cầu đặt chỗ trước. Nếu khách sạn này chấp nhận 215 chổ đặt trước và chỉ có 200 phòng trong khách sạn đó, thì xác suất mà tất cả khách đến yêu cầu một phòng sẽ nhận được một phòng là bao nhiêu?

5.36 Độ tuổi trung bình của hội đồng quản trị là bao nhiêu? Sáu mươi bảy phần trăm các định chế tài chính có hội đồng quản trị có tuổi trung bình là 57 hay nhiều hơn (lấy từ *American Demographics (Nhân khẩu học nước Mỹ)*, tháng Mười Một 1990, trang 22).

a. Trong một mẫu ngẫu nhiên gồm n = 400 định chế tài chính, xác suất để cho 300 hay nhiều hơn hội đồng quản trị có tuổi trung bình là 57 hay nhiều hơn là bao nhiêu?

b. Nếu con số 67% này là chính xác, thì số lượng hội đồng quản trị với độ tuổi trung bình là 57 hay nhiều hơn nên nằm trong hai giá trị nào với xác suất là 95%? (Không sử dụng việc điều chỉnh cho sự liên tục).

5.37 Dịch vụ và sự hỗ trợ đã trở thành một vấn đề quan trọng cho những người sử dụng máy tính cá nhân (PC) khi giá cả của máy tính cá nhân đang ngày càng trở nên giống nhau. Các công ty có sự hỗ trợ kỹ thuật chuyên môn nhanh chóng nhận thấy rằng các khách hàng của mình được thỏa mãn, ngay cả khi họ có vấn đề với máy tính của mình. Ví dụ, 82% số khách hàng mà đã có các vấn đề với máy tính để bàn Dell sẽ sẵn sàng mua một máy PC khác từ công ty Dell (Amirrezvani, 1994). Giả định rằng một mẫu ngẫu nhiên gồm 200 khách hàng của Dell mà đã có các vấn đề với các máy tính để bàn của mình được phỏng vấn; hãy sử dụng sự ước lượng xấp xỉ chuẩn cho phân phối nhị thức để trả lời các câu hỏi sau đây.

a. Xác suất của việc quan sát chỉ có 160 khách hàng mà sẵn lòng mua máy tính cá nhân Dell khác là bao nhiêu?

b. Bạn kỳ vọng rằng số lượng khách hàng sẵn sàng mua một máy tính cá nhân Dell khác rơi vào trong những giới hạn nào mà?

5.38 Phân phối theo độ tuổi của những chủ hộ là một công cụ quan trọng cho những nhà tiếp thị quan tâm đến việc quảng cáo phù hợp với độ tuổi cho một sản phẩm cụ thể mà họ mong muốn tung ra thị trường. Một nghiên cứu do Joint Center for Housing Studies (*Trung tâm Chung Nghiên cứu Nhà ở*) thực hiện đã ước tính rằng vào năm 1995 thì 31% tất cả các chủ hộ sẽ năm trong độ tuổi từ 45 đến 64 (Darnay, 1994). Giả định một mẫu gồm 500 chủ hộ được lấy trong năm 1995.

a. Xác suất để cho có ít hơn 135 chủ hộ nằm trong độ tuổi tữ 45 đến 64 là bao nhiêu?

b. Xác suất để cho có từ 135 đến 180 chủ hộ sẽ nằm trong độ tuổi từ 45 đến 64 là bao nhiêu?

5.39 Vào quí đầu tiên của năm 1994, thu nhập trung vị toàn quốc tại Hoa Kỳ là 39.900 USD ("Midwest, South (*Miền Trung Tây, Miền Nam*),", 1994). Giả định rằng 25 người làm công ăn lương được chọn ngẫu nhiên và thu nhập của họ được ghi nhận.

a. Sử dụng Bảng 1 trong Phụ lục II để tìm ra xác suất để cho có ít nhất 20 người làm công ăn lương có thu nhập vượt quá mức trung vị của toàn quốc.

b. Sử dụng ước lượng xấp xỉ chuẩn cho phân phối nhị thức để ước lượng xấp xỉ xác suất được tìm ra trong phần (a). Ước lượng xấp xỉ này của bạn so với xác suất thực sự là như thế nào?

c. Nếu mẫu mà bạn chọn bị hạn chế ở những người làm công ăn lương sống trong một khu vực địa lý nhất định, thì xác suất tính được trong phần (a) có thể hàm ý điều gì về tính đại diện của mẫu bạn chọn?