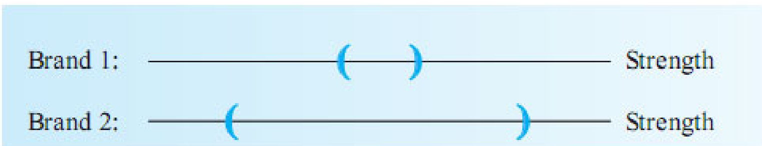
1 Chương 7 KHOẢNG THỐNG KÊ DỰA TRÊN MỘT MẪU ĐƠN

GIỚI THIỆU
Một ước lượng điểm, vì đó là 1 con số, tự bản thân nó không cung cấp bất kỳ thông tin nào về độ chính xác và độ tin cậy của ước lượng đó. Xét ví dụ dùng thống kê $\overline{X}$để tính ước lượng điểm cho giá trị chắc chắn của trung bình thực sự (g) của một thương hiệu khăn giấy xác định, giả sử rằng $\overline{X} = 9322,7$ vì mẫu thay đổi nên hầu như không xảy ra trường hợp$\overline{X} = \mu$. Ước lượng điểm không nói đến việc làm thế nào để tiến gần về μ. Một giải pháp để thay thế đưa ra một giá trị hợp lý duy nhất cho tham số này là tính ước lượng để đưa ra duy nhất một giá trị hợp lý. Cho tham số ước lượng để tính toán và báo cáo toàn bộ khoảng của giá trị hợp lý, một ước lượng khoảng hay khoảng tin cậy (CI) , khoảng tin cậy luôn được tính bằng cách đầu tiên chọn mức độ tin cậy đó là độ đo của khoảng tin cậy, 1 khoảng tin cậy với 95 % độ tin cậy cho trung bình thực sự có giới hạn dưới là 9162,5 và giới hạn trên 9482,9 do đó ở độ tin cậy 95% , bất kỳ giá trị nào của μ giữa 9162,5 và 9482,9 là hợp lý. Độ tin cậy 95% nghĩa là 95% của tất cả mẫu sẽ cho ra một khoảng mà bao gồm cả μ hoặc bất kỳ tham số nào được ước lượng và chỉ 5% của tất cả mẫu nằm trong khoảng sai lầm . khoảng tin cậy thường dùng là 95%, 99% và 90% Cao hơn độ tin cậy, một cách mạnh mẽ hơn ta tin rằng giá trị của tham số được ước lượng trong khoảng (giải thích về độ tin cậy bất kỳ sẽ được đưa ra sau) thông tin về độ chính xác của một khoảng tin cậy được xác định bằng độ rộng của khoảng . nếu độ tin cậy cao và khoảng tin cậy mà ta thu được khá hẹp, ta sẽ dùng kiến thức về giá trị chính xác của tham số hợp lý . Một khoảng tin cậy rất rộng sẽ đưa ra thông tin về sự không chắc chắn trong ước lượng của chúng ta. Hình 7.1 Biểu diễn khoảng tin cậy 95% cho trung bình thực sự của hai nhãn hiệu khăn giấy khác nhau. Một trong những khoảng này cho thấy độ chính xác của μ, ngược lại là gợi ý về giá trị hợp lý của khoảng



Hình 7.1 cho thấy độ chính xác (loại 1) và không chính xác (loại 2) thông tin về μ

2 7.1 Tính chất cơ bản của khoảng tin cậy

Các khái niệm và tính chất cơ bản của khoảng tin cậy (CIs) dễ thấy với trường hợp đơn giản, mặc dù đôi khi không có trong thực tế. xét tham số trung bình μ của một tổng thể.
1. Phân phối của tổng thể là phân phối chuẩn
2. Giá trị độ lệch tiêu chuẩn của tổng thể σ đã biết
Thông thường giả thiết phân phối của quần thể là hợp lý, tuy nhiên nếu chưa biết giá trị của μ mà giá trị σ đã biết thường thì không đủ độ tin cậy (dữ kiện chính về tổng thể thường biết trước các thông tin liên quan) ta sẽ phát triển các phương pháp có giả thiết hạn chế trong phần 7.2 và 7.3
VÍ DỤ 7.1
Kỹ sư công nghiệp chuyên về công thái học quan tâm đến việc thiết kế không gian làm việc thoải mái và các thiết bị công nhân vận hành để đạt được năng suất cao. Bài viết ınghiên cứu về thiết kế bàn phím chữ và sốȷ. (Nhân tố con người ,1985, 175-187) báo cáo về nghiên cứu chiều cao thích hợp cho thí nghiệm bàn phím hỗ trợ cổ tay. Với mẫu n=3 chọn những người

đánh máy đã được đào tạo, và chiều cao bàn phím thích hợp là $\overline{X} = 80.0$ cm. Gỉa sử chiều cao thích hợp có phân phối chuẩn với $\sigma = 2.0cm$ (đây là một trong những giá trị được đề xuất trong bài báo) chiều cao trung bình thích hợp cho tất cả những người có kinh nghiệm đánh máy đạt được một khoảng tin cậy theo μ
Mẫu thực tế quan sát $x_1, x_2, \ldots..x_n$ được giả định là kết quả quan sát của mẫu ngẫu nhiên $X_1, X_2, \ldots..X_n$ từ phân phối chuẩn với trung bình μ và độ lệch tiêu chuẩn σ. Kết quả mô tả trong chương 5 có nghĩa là, không phân biệt cỡ mẫu n bao nhiêu , trung bình mẫu $\overline{X}$ có phân phối chuẩn với kỳ vọng μ và độ lệch tiêu chuẩn $\sigma/\sqrt{n}$, đầu tiên chuẩn hóa $\overline{X}$ bằng cách trừ đi giá trị kỳ vọng sau đó chia cho độ lệch tiêu chuẩn ta được biến chuẩn hóa

$$Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \ (7.1) \tag{7.1}$$

Vì biến chuẩn hóa nằm trong khoảng -1,96 và 1,96 có xác suất .95 nên

$$P(-1,96 < \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} < 1,96) = .95 \tag{7.2}$$

Biến đổi bất đẳng thức bên trong ngoặc của (7.2) sao cho có dạng tương đương $l < \mu < u$, khi đó l và u liên quan đến $\overline{X}$ và $\sigma/\sqrt{n}$ điều này có được thông qua các biến đổi tương đương sau:
1. Nhân cả hai vế với
$\sigma/\sqrt{n}$

$$-1,96.\frac{\sigma}{\sqrt{n}} < \overline{X} - \mu < 1,96.\frac{\sigma}{\sqrt{n}}$$

2. Trừ cả hai vế cho $\overline{X}$

$$-\overline{X} - 1,96.\frac{\sigma}{\sqrt{n}} < -\mu < -\overline{X} + 1,96.\frac{\sigma}{\sqrt{n}}$$

Nhân bất đẳng thức trên cho -1 (lưu ý đảo chiều bất đẳng thức)

$$\overline{X} + 1,96.\frac{\sigma}{\sqrt{n}} > \mu > \overline{X} - 1,96.\frac{\sigma}{\sqrt{n}}$$

Hay

$$\overline{X} - 1,96.\frac{\sigma}{\sqrt{n}} < \mu < \overline{X} + 1,96.\frac{\sigma}{\sqrt{n}}$$

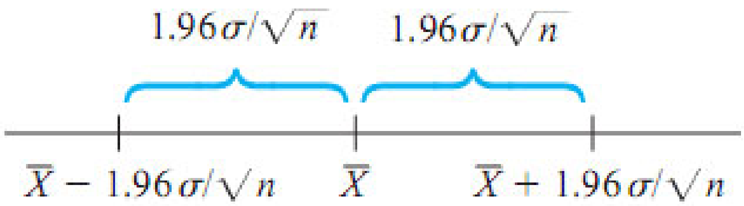
Thay vào (7.2) ta được

$$P(\overline{X} - 1,96.\frac{\sigma}{\sqrt{n}} < \mu < \overline{X} + 1,96.\frac{\sigma}{\sqrt{n}}) = .95 \tag{7.3}$$

Bất đẳng thức bên trong (7.3) chưa được quen thuộc. để đưa về dạng quen thuộc $a \leq Y \leq b$ với a,b là hằng số trong khi bất đẳng thức bên trong (7.3) có số ngẫu nhiên xuất hiện ở 2 điểm kết thúc còn hằng số μ chưa biết lại ở giữa. để giải thích (7.3) ta đưa biến ngẫu nhiên có điểm kết thúc ở bên trái $\overline{X} - 1,96.\frac{\sigma}{\sqrt{n}}$ và điểm kết thúc ở bên phải $\overline{X} + 1,96.\frac{\sigma}{\sqrt{n}}$ về ký hiệu khoảng, tức là

$$(\overline{X} - 1,96.\frac{\sigma}{\sqrt{n}}, \overline{X} + 1,96.\frac{\sigma}{\sqrt{n}}) \tag{7.4}$$

Khoảng trong (7.4)là khoảng ngẫu nhiên vì 2 điểm kết thúc của khoảng liên quan đến biến ngẫu nhiên nó tập trung ở trung bình mẫu $\overline{X}$ và kéo dài đến $(1,96\sigma)/\sqrt{n}$ mỗi bên của $\overline{X}$. Do đó độ rộng của khoảng là $2.(1,96).\frac{\sigma}{\sqrt{n}}$, đó không phải mẫu ngẫu nhiên , chỉ có vị trí của khoảng (nằm giữa $\overline{X}$) là ngẫu nhiên. Có thể giải thích như sau "xác suất là .95 mà bao gồm cả khoảng ngẫu nhiên (7.4) hay bao gồm giá trị đúng của μ " trước khi thực hiện bất kỳ thử nghiệm nào điều đầu tiên là tập hợp dữ liệu sao cho μ sẽ nằm trong khoảng (7.4)



Hình 7.2 khoảng ngẫu nhiên (7.4) tập trung tại $\overline{X}$

ĐỊNH NGHĨA

Nếu sau khi quan sát $X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n$, ta tính được trung bình mẫu được quan sát là $\overline{x}$ sau đó thay $\overline{X}$ trong (7.4) bằng $\overline{x}$ khi đo, khoảng cố định thu được gọi là khoảng tin cậy 95% cho μ. Khi đó khoảng tin cậy được biểu diễn bằng

$$(\overline{x} - 1,96.\frac{\sigma}{\sqrt{n}}, \overline{x} + 1,96.\frac{\sigma}{\sqrt{n}})$$

Là khoảng tin cậy 95% cho μ

Hay

$$\overline{x} - 1,96.\frac{\sigma}{\sqrt{n}} < \mu < \overline{x} + 1,96.\frac{\sigma}{\sqrt{n}}$$

Với độ tin cậy 95% cho μ.

Biểu diễn ngắn gọn cho khoảng như sau $\overline{x} \pm 1.96\frac{\sigma}{\sqrt{n}}$, khi đó dấu ˇ gọi là điểm cuối bên trái (giới hạn dưới) và dấu + là điểm cuối bên phải (giới hạn trên).

VÍ DỤ 7.2

Số lượng cần thiết để tính khoảng tin cậy 95% cho chiều cao trung bình thích hợp là $\sigma = 2.0, n = 31$ và $\overline{x} = 80.0$ khi đó ta có kết quả

$$\overline{x}\frac{\sigma}{\sqrt{n}} = 801.96\frac{2.0}{\sqrt{31}} = 80.7 = (79.3, 80.7)$$

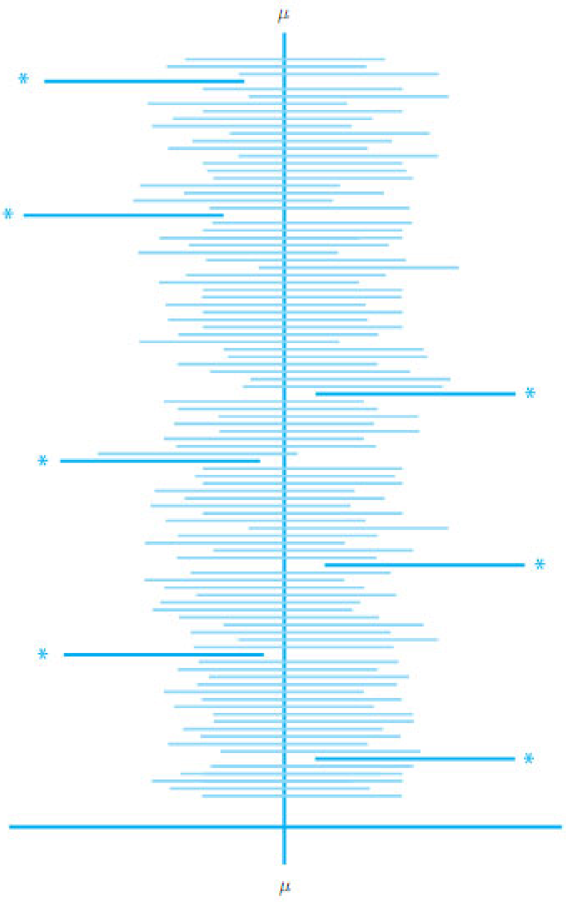
Nghĩa là ta có thể đánh giá được tại mức 95% thì $79.3 < \mu < 80.7$ khoảng cách này tương đối hẹp, tức độ chính xác của ước lượng μ đạt được khá cao.

GIẢI THÍCH VỀ ĐỘ TIN CẬY

Độ tin cậy 95% cho khoảng vừa được định nghĩa được suy ra từ xác suất .95 cho khoảng ngẫu nhiên (7.4) các khoảng có độ tin cậy khác sẽ được giới thiệu ngắn gọn hơn. Hiện tại ta chỉ giải thích cho khoảng tin cậy 95%.

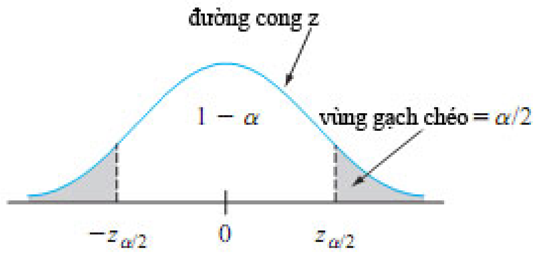
Vì ta bắt đầu với xác suất .95 nên với khoảng ngẫu nhiên trong (7.4) ta tìm được giá trị thực của μ từ đó dựa vào dữ liệu trong ví dụ 7.4 tính được khoảng tin cậy CI $(79.3; 80.7)$. từ đó đưa ra kết luận μ có xác suất .95 trong khoảng cố định này nhưng do ta thay $\overline{X}$bằng $\overline{x}$=80.0 , nên không có sự ngẫu nhiên ở đây. Khoảng $(79.3; 80.7)$ không là khoảng ngẫu nhiên và μ là một

hằng số (chưa biết) do đó khi viết $P(\mu \in (79.3; 80.7)) = .95$ là không chính xác.
Một giải thích chính xác của "độ tin cậy 95% " được dựa trên tần số dài của xác suất, để nói một sự kiện A có xác suất .95 tức nói rằng nếu thử nghiệm trên A được lặp đi lặp lại về lâu dài thì A sẽ xảy ra 95% . Gỉa sử ta lấy 1 mẫu khác của chiều cao thích hợp cho những người đánh máy sau đó tính toán khác với khoảng 95% thì ta xem xét việc lặp lại điều này cho 1 mẫu thứ 3, 1 mẫu thứ 4, 1 mẫu thứ 5,....đặt A là sự kiện mà $\overline{X} - 1,96.\frac{\sigma}{\sqrt{n}} < \mu < \overline{X} + 1,96.\frac{\sigma}{\sqrt{n}}$. Khi đó $P(A) = .95$ theo tính toán thì 95% của khoảng tin cậy sẽ chứa μ , điều này được minh họa trong hình 7.3 khi đường thẳng đứng cắt trục đo tại giá trị μ (chưa biết) chú ý rằng 7 trong 100 khoảng sẽ không chứa μ tức là chỉ 5% của khoảng được biểu diễn như vậy sẽ không chứa μ.
Giải thích thêm cho điều này ,độ tin cậy 95% không đặc biệt như khoảng $(79.3; 80.7)$ mà sẽ liên quan đến số lượng rất lớn của khoảng được xây dựng một cách tương tự.



Hình 7.3 100 khoảng tin cậy với 95% (dấu * xác định khoảng mà nó không chứa μ)
Dùng công thức khoảng tin cậy mặc dù đôi khi không thoả mãn yêu cầu bài toán. Khó khăn ở đây là ta cần giải thích về xác suất khi áp dụng trong một chuỗi dài nhân rộng của thí nghiệm chứ không phải là bản sao duy nhất, một cách tiếp cận khác để xây dựng và giải thích khoảng tin cậy là sử dụng khái niệm về xác suất trong định lý Bayes. Về chi tiết nó vượt quá phạm vi của chủ đề này ta có thể tham khảo sách viết bởi DeGroot, et al (xem danh mục chương 6) khoảng được trình bày ở đây (cũng như những khoảng trình bày sau này) được gọi là khoảng tin cậy "cổ điển" vì giải thích chúng dựa trên khái niệm cổ điển của xác suất.
CÁC MỨC ĐỘ TIN CẬY KHÁC
Độ tin cậy 95% được suy từ xác suất .95 trong bất đẳng thức (7.2). nếu muốn độ tin cậy 99% thì xác suất ban đầu phải thay bằng .99 điều đó đòi hỏi phải thay giá trị đặc biệt Z từ 1.96 đến 2.58, độ tin cậy 99% có được bằng cách thay 2.58 cho 1.96 trong công thức độ tin cậy 95%,

trong thực tế muốn có độ tin cậy bao nhiêu thì ta thay 1.96 hay 2.58 bằng giá trị thích hợp của phân phối chuẩn. hình 7.4 biểu diễn xác suất của 1-α có được bằng cách dùng $Z_{\alpha/2}$ thay cho 1.96.



Hình 7.4 $P(-z_{\alpha/2 \leq Z < z_{\alpha/2}}) = 1-\alpha$

ĐỊNH NGHĨA

Với khoảng tin cậy $100(1-\alpha)\%$ cho trung bình μ của một tổng thể có phân phối chuẩn thì giá trị của σ được cho bởi

$$(\overline{x} - Z_{\alpha/2}.\frac{\sigma}{\sqrt{n}}, \overline{x} + Z_{\alpha/2}.\frac{\sigma}{\sqrt{n}}) \tag{7.5}$$

hay, tương đương với $\overline{x}Z_{\alpha/2}.\frac{\sigma}{\sqrt{n}}$ Công thức (7.5) cho khoảng tin cậy, điều này có thể phát biểu thành lời như sau Ước lượng điểm của μ(giá trị đặc biệt Z) (lỗi chuẩn hóa của kỳ vọng)

VÍ DỤ 7.3

Qui trình sản xuất một loại động cơ kiểm soát nhà ở cụ thể gần đây đã được sửa đổi , lịch sử dữ liệu cho biết phân phối của lỗ ổ cắm trên vỏ có độ lệch tiêu chuẩn 100 mm người ta tin rằng việc sửa đổi này không ảnh hưởng đến phân phối cũng như độ lệch tiêu chuẩn, tuy nhiên giá trị trung bình của đường kính có thể thay đổi, một mẫu gồm 40 nhà ở được chọn,đường kính lỗ được xác định với đường kính trung bình trong 1 mẫu lạ.426mm. hãy tính khoảng tin cậy cho trung bình đường kính lỗ sử dụng độ tin cậy 90% tức $100(1-\alpha) = 90$ suy ra α=10 và $Z_{\alpha/2}$=$Z_{.05}$=1.645 (tương ứng với diện tích của đường cong Z là .9500)

Khi đó xác định được khoảng mong muốn là

$$5.426(1.645).\frac{.100}{\sqrt{40}} = 5.426.026 = (5.400, 5.452)$$

Vì độ tin cậy cao nên ta có thể nói rằng $5.400 < \mu < 5.452$ khoảng này khá hẹp vì vậy một lượng nhỏ giá trị sẽ nằm trong đường kính lỗ($\sigma = .100$).

ĐỘ TIN CẬY, ĐỘ CHÍNH XÁC VÀ CỠ MẪU

Tại sao phải dùng độ tin cậy 95% trong khi độ tin cậy có thể đạt được 99% Vì với độ tin cậy càng cao thì khoảng càng rộng. Khi đó khoảng 95% là khoảng kéo dài của $1.96\frac{\sigma}{\sqrt{n}}$ cho mỗi bên của $\overline{x}$, độ rộng của khoảng là 2.(1.96)$\overline{x}$=3.92$\overline{x}$tương tự ta có độ rộng của khoảng 99% là 2(2.58)$\overline{x}$=5.16$\overline{x}$ nghĩa là ta có nhiều tin cậy hơn vì nó rộng hơn.

Nếu độ rộng của khoảng tin cậy có thể xác định chính xác độ tin cậy thì ngược lại với độ chính xác của nó ước lượng 1 khoảng tin cậy từ độ tin cậy cao mặc dù có thể không chính xác ở điểm kết thúc , khoảng có thể xa nhau trong khi khoảng chính xác có thể dẫn đến độ tin cậy tương đối thấp do đó không thể nói khoảng 99% hợp lý hơn khoảng 95% tức nếu đạt được mong muốn trong độ tin cậy thì đòi hỏi mất mát trong độ chính xác. Để đạt được mong muốn đó thì ta nên xác định độ tin cậy và độ rộng của khoảng sau đó mới xác định cỡ mẫu cần thiết.

VÍ DỤ 7.4

Quan sát thời gian chia sẻ lệnh rộng rãi của một hệ thống máy tính biết rằng thời gian cụ thể cho 1 lệnh chỉnh sửa là phân phối chuẩn với độ lệch tiêu chuẩn là .25 milisec, một hệ điều hành mới được cài đặt, ta ước lượng được thời gian trung bình thực sự là μ, trong điều kiện mới giả sử thời gian luôn có phân phối chuẩn với σ=25, xác định cỡ mẫu cần để có độ tin cậy 95% và độ rộng (tối đa) 10.

Gọi n là cỡ mẫu thỏa yêu cầu

$$10 = 2.(1.96)(25/\sqrt{n})$$

Biến đổi phương trình này ta được

$$\sqrt{n} = 2.(1.96)(25)/10 = 9.80$$

nên

$$n = (9.80)^2 = 96.04$$

Vì n là số nguyên nên cỡ mẫu cần tìm là n=97

Công thức chung tìm cỡ mẫu n với độ rộng khoảng là w là $2.Z_{\alpha/2}.\frac{\sigma}{\sqrt{n}}$ từ đó suy ra n

Cỡ mẫu cần cho khoảng tin cậy trong (7.5) với độ rộng w là

$$n = (2.Z_{\alpha/2}.\frac{\sigma}{w})^2$$

Độ rộng w càng nhỏ thì cỡ mẫu n càng lớn, thêm vào đó,n tăng theo hàm của σ (nhiều biến tổng thể cần 1 cỡ mẫu lớn) có độ tin cậy 100(1-α)(với α giảm $Z_{\alpha/2}$tăng)

Một nửa độ rộng 1.96 $\frac{\sigma}{\sqrt{n}}$ của độ tin cậy 95% đôi khi được gọi là ràng buộc lỗi của ước lượng kết hợp với độ tin cậy 95% nghĩa là với 95% tin cậy,ước lượng điểm $\overline{x}$sẽ gần μ hơn. Trước khi thu thập dữ liệu điều tra viên muốn xác định cỡ mẫu sao cho mỗi giá trị cụ thể phải có một ràng buộc. Lấy ví dụ với μ là trung bình hiệu suất nhiên liệu (mpg) cho tất cả các xe cùng loại, mục tiêu của điều tra viên là có thể ước lượng μ thông qua B(qui định về ràng buộc lỗi của ước lượng) với $100(1-\alpha)$% độ tin cậy,kết quả thu được là cỡ mẫu cần thiết có được bằng cách thay 2/w bằng 1/B trong công thức ở trên.

NGUỒN GỐC CỦA KHOẢNG TIN CẬY

Đặt $X_1, X_2, \ldots..X_n$ là ký hiệu mẫu của khoảng tin cậy dựa trên tham số θ giả sử tìm được biến ngẫu nhiên thỏa 2 tính chất sau

1. Hàm phụ thuộc cả hai biến $X_1, X_2, \ldots..X_n$ và θ

2. Phân phối xác suất của biến không phụ thuộc θ hay bất kỳ tham số nào

Đặt $h(X_1, X_2, \ldots..X_n;\theta)$ là biến ngẫu nhiên. Lấy ví dụ, Nếu phân phối của tổng thể là phân phối chuẩn khi biết σ và $\theta = \mu$, biến $h(X_1, X_2, \ldots..X_n;\mu) = (\overline{X} - \mu)/\sigma/\sqrt{n}$ thỏa cả hai tính chất là hàm phụ thuộc theo μ nhưng phân phối không phụ thuộc μ. Nói chung dạng hàm theo h thường dùng để kiểm tra phân phối của một ước lượng đặc biệt $\hat{\theta}$.

Cho α bất kỳ thuộc 0 và 1 ta tìm được hằng số a và b thỏa

$$P(a < h(X_1, X_2, \ldots..X_n;\theta) < b) = 1 - \alpha \tag{7.6}$$

Theo tính chất 2, a và b không phụ thuộc θ, theo ví dụ chuẩn thì $a = -Z_{\alpha/2}$,$b = Z_{\alpha/2}$ Biến đổi (7.6) theo θ ta được xác suất

$$P(l(X_1, X_2, \ldots..X_n) < \theta < u(X_1, X_2, \ldots..X_n)) = 1 - \alpha$$

Khi đó $l(x_1, x_2, \ldots..x_n)$ và $u(x_1, x2, \ldots..x_n)$ gọi là giới hạn trên và giới hạn dưới của khoảng tin cậy với độ tin cậy$100(1-\alpha)\%$ trong ví dụ chuẩn ta thấy rằng $l(X_1, X_2, \ldots..X_n) = \overline{X} - Z_{\alpha/2}.\sigma/\sqrt{n}$ và $u(X_1, X_2, \ldots..X_n) = \overline{X} + Z_{\alpha/2}.\sigma/\sqrt{n}$

VÍ DỤ 7.5

Mô hình lý thuyết cho rằng thời gian để chất lỏng cách điện ở các điện cực bị phân hủy với một điện áp đặc biệt có phân phối mũ với tham số λ(Xem mục 4.4).Một mẫu ngẫu nhiên có kích thước n=10 với chuỗi thời gian phân hủy có dữ liệu (theo phút) như sau $x_1 = 41.53, x_2 = 18.73, x_3 = 2.99, x_4 = 30.34, x_5 = 12.33, x_6 = 117.52, x_7 = 73.02, x_8 = 223.63, x_9 = 4.00, x_10 = 26.78$. Với độ tin cậy 95% cho λ và cho biết trung bình của thời gian phân hủy được thỏa mãn. Đặt $h(X_1, X_2, \ldots..X_n; \lambda) = 2\lambda \sum X_j$. Nó có thể được biểu diễn bằng một biến ngẫu nhiên có phân phối chi- bình phương với 2n bậc tự do $(df)(\nu = 2n$,khi ν là tham số của phân phối xác suất chi - bình phương như trong chú ý mục 4.4). Bảng phụ lục A.7 là hình minh họa của đường cong mật độ và các giá trị đặc biệt của phân phối chi- bình phương. con số liên quan ở đây là 2(10) = 20. hàng thứ $\nu = 20$ biểu diễn rằng 34.170 giá trị đạt được vùng bên phải .025 và 9.591 giá trị đạt được tại vùng bên trái .025 (vùng bên phải là .975)

Do đó cho n=10

$$P(9.591 < 2\lambda \sum X_j) < 34.170) = .95$$

Chia cả hai vế cho $2\sum X_j$ ta được

$$P(9.591/(2\sum X_j) < \lambda < (34.170/(2\sum X_j)) = .95$$

giới hạn dưới của λ với khoảng tin cậy 95% là $9.591/(2\sum X_j)$,và giới hạn trên là $34.170/(2\sum X_j)$ cho dữ liệu, $\sum X_j = 550.87$, cho khoảng $(.00871, 03101)$.khi đó

Gía trị kỳ vọng của mũ rv là $\mu = 1/\lambda$.khi đó

$$P(2\sum X_j/34.170 < 1/\lambda < 2\sum X_j/9.591) = .95$$

khoảng tin cậy cho trung bình đúng của thời gian phá hủy là $2\sum X_j/34.170, 2\sum X_j/9.591 = (32.24, 114.87)$. Rõ ràng là khoảng này khá rộng , nó phản ánh sự thay đổi đáng kể trong thời gian phá hủy với cỡ mẫu nhỏ.

Nói chung, Kết quả của giới hạn trên và giới hạn dưới của khoảng tin cậy có được từ việc thay dấu < trong (7.6) bằng dấu = và giải tìm θ.Trong ví dụ về sự cách điện của dung dịch chỉ cần xét $2\lambda \sum X_j = 34.170$ suy ra $\lambda = 34.170/(2\sum X_j)$ là giới hạn trên, và giới hạn dưới thu được từ phương trình khác. Chú ý rằng theo ước lượng điểm thì 2 khoảng giới hạn này không bằng nhau. khi đó khoảng không có dạng $\hat{\theta} \pm c$.

BOOTSTRAP KHOẢNG TIN CẬY CỦA CHÍNH NÓ

Kỹ thuật bootstrap đã giới thiệu trong chương 6 như là cách để ước lượng $\sigma_{\hat{\theta}}$.Nó cũng được ứng dụng tìm khoảng tin cậy cho θ.Xét ước lượng có kỳ vọng μ của phân phối chuẩn với σ đã biết. Thay μ bởi θ và dùng $\hat{\theta} = \overline{X}$ như trong ước lượng điểm. Chú ý rằng $1.96\sigma\sqrt{n}$ là bách phân vị thứ 97.5 của phân phối $\hat{\theta} - \theta$ [tức là, $P(\overline{X} - \mu < 1.96\sigma/\sqrt{n}) = P(Z < 1.96) = .9750$].Tương tự , $-1.96\sigma/\sqrt{n}$ là bách phân vị thứ 2.5, vì vậy

.95=P(bách phân vị thứ 2.5 $< \hat{\theta} - \theta <$ bách phân vị thứ 97.5)

$= P(\hat{\theta}$-bách phân vị thứ 2.5 $> \theta > \hat{\theta}$-bách phân vị thứ 97.5)

Nghĩa là, với

$$l = \hat{\theta} - \text{ bách phân vị thứ 97.5 của } \hat{\theta} - \theta l = \hat{\theta} - \text{ bách phân vị thứ 2.5 của } \hat{\theta} - \theta \qquad (7.7)$$

Khoảng tin cậy cho θ là (l, u). Trong nhiều trường hợp, bách phân vị trong (7.7) không thể tính được,nhưng ta có thể ước lượng được từ mẫu của chính nó. Gỉa sử cho $B = 1000$ là mẫu chính nó và tính được $\hat{\theta_1^*}, ..., \hat{\theta_{1000}^*}$,và $\hat{\overline{\theta}^*}$ được suy ra từ 1000 giá trị khác nhau $\hat{\theta_1^*} - \hat{\theta^*}, ..., \hat{\theta_{1000}^*} - \hat{\theta^*}$.giá trị thứ 25 lớn nhất và giá trị thứ 25 nhỏ nhất của những sai khác này là ước lượng của những bách phân vị chưa biết trong (7.7) . Tham khảo thêm trong sách của Devore và Berk hay Efron trong chương 6 để biết thêm thông tin.

BÀI TẬP MỤC 7.1$(1 - 11)$

1. Xét một phân phối chuẩn với giá trị σ đã biết.

a. Xác định độ tin cậy cho khoảng tin cậy $\overline{x} \pm 2.81\sigma/\sqrt{n}$?

b.Xác định độ tin cậy cho khoảng tin cậy $\overline{x} \pm 1.44\sigma/\sqrt{n}$?

c. Tìm giá trị $Z_{\alpha/2}$ trong công thức khoảng tin cậy (7.5) với độ tin cậy 99,7%

d.Trả lời câu hỏi trong phần (c) với độ tin cậy 75%

2.Với mỗi khoảng tin cậy sau cho $\nu =$ trung bình đúng (tức là kỳ vọng của tổng thể) tần số cộng hưởng (Hz) vợt quần vợt của một loại nhất định:

$(114.4, 115.6)$ $\quad$ $(114.1, 115.9)$

a. Tính trung bình mẫu của tần số cộng hưởng

b. Cả hai khoảng đều được tính từ cùng một dữ liệu mẫu. Độ tin cậy cho mỗi khoảng là 90% và những khoảng khác là 99%. Khoảng nào có độ tin cậy 90% và tại sao?

3. Gỉa sử một mẫu ngẫu nhiên gồm 50 chai thuốc xi rô ho được chọn và hàm lượng cồn của mỗi chai được xác định. Đặt μ là trung bình của nồng độ cồn cho tổng thể tất cả các chai của thương hiệu đang được nghiên cứu. Gỉa sử kết quả của độ tin cậy 95% là khoảng tin cậy $(7.8, 9.4)$.

a.Tìm khoảng tin cậy cho độ tin cậy 90% từ mẫu đó và cho biết khoảng này hẹp hơn hay rộng hơn khoảng cho trước? giải thích lý do tại sao.

b.xét phát biểu sau:95% sự thay đổi của μ giữa 7.8 và 9.4 . Đây có là phát biểu đúng không? Tại sao đúng và tại sao không đúng?

c. xét phát biểu sau: Ta có thể rất tự tin cho rằng 95% tất cả các chai si rô ho của cùng một loại có nồng độ cồn nằm trong khoảng 7.8 và 9.4.Đây có là phát biểu đúng không? Tại sao đúng và tại sao không đúng?

d. xét phát biểu sau: Nếu quá trình chọn cỡ mẫu là 50 và tính toán tương ứng với độ tin cậy 95% khoảng tin cậy lặp lại 100 lần, Trong 95 kết quả khoảng đó sẽ bao gồm μ. Đây có là phát biểu đúng không? Tại sao đúng và tại sao không đúng?

4. Một khoảng tin cậy mong muốn cho hao hụt tải trọng có trung bình μ (watts) của một loại động cơ cảm ứng khi tải trọng dòng điện hiện tại nằm trong khoảng 10 amps đến 1500 rpm. Gỉa sử tải trọng mất đi có phân phối chuẩn với $\sigma = 3.0$.

a. Tính khoảng tin cậy 95% cho μ khi $n = 25$ và $\overline{x} = 58.3$

b Tính khoảng tin cậy 95% cho μ khi $n = 100$ và $\overline{x} = 58.3$

c. Tính khoảng tin cậy 99% cho μ khi $n = 100$ và $\overline{x} = 58.3$

d. Tính khoảng tin cậy 82% cho μ khi $n = 100$ và $\overline{x} = 58.3$

e. Tìm n đủ lớn để khoảng tin cậy 99%cho μ có độ rộng 1.0?

5. Gỉa sử độ rỗ của heli (theo phần trăm) của một mẫu than tổ ong của một lớp bất kỳ có phân phối chuẩn với độ lệch tiêu chuẩn .75.

a. Tính khoảng tin cậy cho 95% của lớp xỉ than nếu trung bình độ rỗ cho 20 mẫu từ lớp xỉ than là 4.85.

b. Tính khoảng tin cậy cho 98% cho trung bình đúng của mẫu sỉ than khác dựa trên 16 mẫu với trung bình mẫu 4.56.

c.Tìm n đủ lớn để khoảng tin cậy 95% với độ rộng 4.0?

d. Tìm cỡ mẫu cần để ước lượng trung bình đúng độ xốp bên trong xỉ than là .2 với độ tin cậy 99%

6. Trên cơ sở thử nghiệm rộng rãi, lợi ích của một loại thanh thép nhẹ đặc biệt có phân phối

chuẩn với $\sigma = 100$ đã biết. Thành phần của thanh đã được sửa đổi một chút,nhưng không ảnh hưởng đến chuẩn và giá trị σ.
a. Gỉa sử trong trường hợp này, nếu mẫu thép là 25 dựa trên trung bình mẫu về lợi ích của thép nhẹ 8439 lb. Tính khoảng tin cậy 90% cho lợi ích của những thanh thép nhe.
b. Xác định khoảng tin cậy trong phần (a) với độ tin cậy 92%.
7. Phải tăng cỡ mẫu n bằng bao nhiêu nếu độ rộng của khoảng tin cậy trong (7.5) giảm đi một nửa? Nếu cỡ mẫu tăng thêm 25 lần, độ rộng sẽ bị ảnh hưởng như thế nào? Biện luận cho khẳng định của bạn.
8. Đặt $\alpha_1 > 0, \alpha_2 > 0$, với $\alpha_1 + \alpha_2 = \alpha$. Thì

$$P(-Z_{\alpha_1} < \frac{\overline{X} - \mu}{\sigma/\sqrt{n}}) = 1 - \alpha$$

a. Dùng phương trình này đưa ra một dạng chung với khoảng tin cậy $100(1-\alpha)\%$ cho μ trong đó khoảng trong (7.5) là một trường hợp đặc biệt.
b. Đặt $\alpha = .05$ và $\alpha_1 = \alpha/4, \alpha_2 = 3\alpha/4$,. Kết quả này sẽ rộng hơn hay hẹp hơn khoảng trong (7.5) ?
9. a. Dưới những điều kiện tương tự để có (7.5) ,$P[(\overline{X} - \mu)/(\sigma/\sqrt{n}) < 1.645] = .95$. Dùng kết quả này để rút ra một khoảng cho μ mà có độ rộng được giới hạn và cung cấp một ràng buộc độ tin cậy cho μ.Đó là khoảng nào trong phần bài tập $5(a)$.
b. Tổng hợp kết quả của phần (a) để có chặn dưới với độ tin cậy $100(1-\alpha)\%$
c. Khoảng tin cậy nào trong phần (b) cho biết chặn trên của μ ? Tính cho khoảng tin cậy 99% với dữ liệu trong bài tập $4(a)$
10. Một mẫu ngẫu nhiên $n = 15$ quan sát thời gian hoạt động của máy bơm nhiệt (theo năm) có kết quả:

2.0	1.3	6.0	1.9	5.1	.4	1.0	5.3
15.7	.7	4.8	.9	12.2	5.3	.6	

a. Gỉa sử thời gian hoạt động có phân phối mũ. lập luận tương tự như trong ví dụ (7.5) với khoảng tin cậy 95% cho kỳ vọng thời gian hoạt động (trung bình đúng) b. Khoảng trong phần (a) sẽ thay đổi thế nào để đạt được độ tin cậy 99%?
c. xác định độ lệch chuẩn thời gian hoạt động của phân phối với khoảng tin cậy 95% ? [tức là :độ lệch chuẩn bằng bao nhiêu với kỳ vọng của biến ngẫu nhiên?]
1. Xét thống kê tư vấn cho 1000 khách hàng khác nhau với khoảng tin cậy 95% cho μ. Gỉa sử tập hợp dữ liệu trên khoảng được chọn là độc lập. Có bao nhiêu giá trị kỳ vọng trong 1000 khoảng tương ứng với μ? Xác suất này bằng bao nhiêu khi xét trong khoảng 940 và 960 mà tương ứng với giá trị của μ? [Tức là: đặt Y= Số trong 1000 khoảng bao gồm cả μ biến ngẫu nhiên của Y có dạng thế nào]

3 7.2 CÁC KHOẢNG TIN CẬYCỦA MẪU LỚN CHO TRUNG BÌNH VÀ TỈ LỆ CỦA TỔNG THỂ

Khoảng tin cậy cho μ trong những phần trước đều giả sử tổng thể có phân phối chuẩn với σ đã biết. Bây giờ ta giả sử khoảng tin cậy của mẫu lớn không yêu cầu những giả thiết này. Sau đây ta sẽ lập luận cho điều này để đưa ra khái quát cho những khoảng của mẫu lớn. Ta sẽ thảo luận trên khoảng cho một tỉ lệ tổng thể là p.
MỘT KHOẢNG CỦA MẪU LỚN CHO μ
Đặt $X_1, X_2...., X_n$ là một mẫu ngẫu nhiên từ tổng thể có kỳ vọng μ và độ lệch tiêu chuẩn σ.

Cho n lớn, Định lý giới hạn trung tâm (CLT) chỉ ra rằng $\overline{X}$ dần về phân phối chuẩn tuy nhiên dù tổng thể có phân phối nào. nên có thể nói $Z = \frac{\overline{X}-\mu}{\sigma/\sqrt{n}}$ có phân phối xấp xỉ chuẩn, vì thế

$$P(-Z_{\alpha/2} < \frac{\overline{X}-\mu}{\sigma/\sqrt{n}} < Z_{\alpha}/2) \approx 1-\alpha$$

Lập luận như trong (7.1) cho $\overline{x} \pm z_{\alpha}/2.\sigma/\sqrt{n}$ có khoảng tin cậy của một mẫu lớn cho μ với độ tin cậy xấp xỉ $100(1-\alpha)\%$. Nghĩa là khi n lớn, khoảng tin cậy đã cho trước đó vẫn thỏa với bất kỳ phân phối nào của tổng thể miễn là "xấp xỉ " được đưa vào độ tin cậy.
Khó khăn ở đây là việc tính toán khoảng tin cậy dựa trên giá trị σ mà hiếm khi được biết. Xét biến chuẩn hóa$\frac{\overline{X}-\mu}{S/\sqrt{n}}$, trong biến chuẩn hóa này S thay cho σ. Trước đó chỉ có tử số là ngẫu nhiên theo Z vì $\overline{X}$. Trong biến chuẩn hóa mới này cả$\overline{X}$ và S đêu có giá trị thay đổi từ mẫu này đến mẫu khác. Vì thế ta thấy rằng phân phối của biến mới này trải dài hơn đường cong z phản ánh sự thay đổi trong mẫu số. Điều này đúng khi n nhỏ. Tuy nhiên khi n lớn có một chút thay đổi khi thay biến S cho σ, vì vậy biến này cũng có phân phối xấp xỉ chuẩn. Thao tác trên biến này trong một phát biểu xác suất, như trường hợp của σ đã biết, được một khoảng tin cậy của mẫu lớn cho μ.
MỆNH ĐỀ Nếu n đủ lớn, biến chuẩn hóa

$$\frac{\overline{X}-\mu}{S/\sqrt{n}}$$

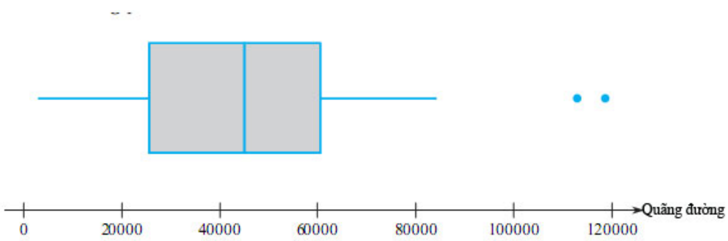
có phân phối xấp xỉ chuẩn, Nghĩa là

$$\overline{x} \pm z_{\alpha}/2.\frac{s}{\sqrt{n}} \tag{7.8}$$

là khoảng tin cậy của mẫu lớn cho μ với độ tin cậy xấp xỉ $100(1-\alpha)\%$. Công thức này đúng với mọi phân phối của tổng thể.
giải thích cho công thức (7.8)
ước lượng điểm của $\mu \pm$ (z là giá trị đặc biệt) (lỗi ước lượng chuẩn của kỳ vọng).
Nói chung, $n > 40$ sẽ thỏa để chứng minh các khoảng này. Điều này có phần hạn chế hơn công thức cho định lý giới hạn trung tâm CLT vì thay biến mới được sử dụng là S thay cho σ.
VÍ DỤ 7.6
Bạn có bao giờ muốn sở hữu một chiếc Porsche? tác giả nghĩ anh ta có khả năng có một chiếc Boxster, dòng xe rẻ nhất. Vì thế anh ta đến www.car.com vào ngày 18 tháng 11 năm 2009, và tìm thấy 1113 chiếc xe được liệt kê. Gía dao động từ $3499 đến $130,000 (giá một trong hai vượt quá $70,000) giá cả làm anh ta chán nản , vì thế anh ta tập trung đọc về đồng hồ đo đường (miles). ở đây là báo cáo của 50 mẫu xe Boxsters:

2948	2996	7197	8338	8500	8759	12710	12925
15767	20000	23247	24863	26000	26210	30552	30600
35700	36466	40316	40596	41021	41234	43000	44607
45000	45027	45442	46963	47978	49518	52000	53334
54208	56062	57000	57365	60020	60265	60803	62851
64404	72140	74594	79308	79500	80000	80000	84000
113000	118634						

Hộp sơ đồ của dữ liệu (hình 7.5) biểu diễn rằng,trừ hai đường thẳng cuối phía trên, phân phối của các giá trị đối xứng hợp lý (thực tế, phân phối chuẩn của các vị trí biểu hiện một mẫu tuyến tính hợp lý, tuy nhiên các điểm phù hợp với hai quan sát nhỏ nhất và lớn nhất đôi khi di chuyển từ một đường thẳng phù hợp đến các điểm còn lại).



Hình 7.5 Hộp dữ liệu của đồng hồ đo đường từ ví dụ 7.6

Tổng hợp số liệu $n = 50, \overline{x} = 45,679.4, x^{\sim} = 45,013.5, s = 26.641.675, f_s = 34,265$ Kỳ vọng và trung vị là hợp lý (nếu hai giá trị lớn nhất giảm đi 30,000,kỳ vọng sẽ giảm đi 44,479.4, trong khi trung vị không chịu ảnh hưởng gì). Hộp sơ đồ và độ lớn của s và f_s có mối quan hệ với kỳ vọng và trung vị cả hai đều cho biết số lượng biến. Độ tin cậy 95% cần cho $z_{.025} = 1.96$, và khoảng là

$$45,679.4 \pm (1.96)(\frac{26,641.675}{\sqrt{50}}) = 45,679.4 \pm 7384.7$$
$$= (38,294.7, 53,064.1)$$

Tức là, $38,294.7 < \mu < 53,064.1$ với độ tin cậy 95%. Khoảng này rộng hơn vì cỡ mẫu là 50, Mặc dù đủ lớn theo qui định, nhưng không đủ lớn để vượt qua thay đổi đáng kể trong mẫu. Ta cũng không ước lượng chính xác kỳ vọng của tổng thể về đồng hồ đo đường.

Khoảng nào mà ta đã tính được 95% trong một thời gian dài bao gồm cả tham số được ước lượng hay nó có nằm trong 5% xấu ?Ngoài việc biết giá trị μ, ta không nói lên điều gì. Nhớ rằng , phần trăm độ tin cậy phù hợp lâu dài khi công thức dùng cho những biến lặp lại; Nó không giải thích cho mẫu đơn và kết quả khoảng.

Không may, Việc chọn cỡ mẫu để có được độ rộng của khoảng không đơn giản khi chỉ biết σ.Vì độ rộng của khoảng trong (7.8) là $2z_{\alpha/2}s/\sqrt{n}$. Khi đó giá trị của s chưa có trước khi thu thập dữ liệu, Độ rộng của khoảng không thể xác định cho việc lựa chọn n , hướng dẫn chọn giá trị s phù hợp, một n lớn hơn yêu cầu sẽ được chọn. Điều tra viên có lý do để sắp xếp giá trị tổng thể chính xác hơn (Khác biệt giữa giá trị lớn và giá trị nhỏ). nếu phân phối của tổng thể không bị lệch quá thì chia giá trị thành 4 phần sẽ cho được kết quả

VÍ DỤ 7.7

Thời gian cho một rẽ cắm sạc (phút) cho việc nung thép carbon trong lò được xác định từ mỗi loại nhiệt có cỡ mẫu n. Nếu điều tra viên tin rằng hầu hết thời gian phân phối trong khoảng từ 320 đến 440. Mẫu nào phù hợp để ước lượng cho trung bình đúng trong thời gian 5 phút. với độ tin cậy 95%

giá trị hợp lý cho s là $(440 - 320)/4 = 30$. Do đó

$$n = [\frac{(1.96)(30)}{5}]^2 = 138.3$$

Do cỡ mẫu là số nguyên nên n=139. chú ý ta đang ước lượng trong 5 phút.Tương tự với độ tin cậy trong thời gian 10 phút.

KHOẢNG TIN CẬY CHO MẪU CHUNG LỚN

Khoảng cho mẫu lớn $\overline{x} \pm z_\alpha/2.\sigma/\sqrt{n}$và $\overline{x} \pm z_\alpha/2.s/\sqrt{n}$ là một trường hợp đặc biệt của khoảng tin cậy của mẫu chung lớn cho tham số θ . Giả sử $\hat{\theta}$ là ước lượng hợp lý thỏa tính chất sau:(1) là phân phối xấp xỉ chuẩn;(2) ;là ước lượng không chệch (ít xấp xỉ nhất) và (3) một biểu diễn của $\sigma_{\hat{\theta}}$, độ lệch tiêu chuẩn $\hat{\theta}$ đã biết. Ví dụ, trong trường hợp $\theta = \mu, \hat{\mu} = \overline{X}$là một ước lượng không chệch của phân phối xấp xỉ chuẩn khi n lớn và $\sigma_{\hat{\mu}} = \sigma_{\overline{X}} = \sigma/\sqrt{n}$. Chuẩn hóa $\hat{\theta}$ thành rv

$Z = (\hat{\theta} - \theta)/\sigma_{\hat{\theta}}$, được xấp xỉ có phân phối chuẩn. Điều này cho ta một phát biểu về xác suất

$$P(-z_{\alpha/2} < \frac{\hat{\theta} - \theta}{\sigma_{\hat{\theta}}} < z_{\alpha/2}) < z_{\alpha/2}) \approx 1 - \alpha \quad (7.9)$$

Đầu tiên giả sử $\sigma_{\hat{\theta}}$ không liên quan đến bất kỳ tham số nào (ví dụ, biết σ trong trường hợp $\theta = \mu$).Thay dấu $<$ trong (7.9) thành dấu $=$ ta có kết quả $\theta = \hat{\theta} \pm z_{\alpha/2}.\sigma_{\hat{\theta}}$, vì vậy giới hạn trên và giới hạn dưới tương ứng với độ tin cậy là $\theta = \hat{\theta} - z_{\alpha/2}.\sigma_{\hat{\theta}}$ và $\theta = \hat{\theta} + z_{\alpha/2}.\sigma_{\hat{\theta}}$.Bây giờ giả sử $\sigma_{\hat{\theta}}$ không liên quan đến θ nhưng có liên quan đến một vài tham số.Đặt $s_{\hat{\theta}}$ là ước lượng của $\sigma_{\hat{\theta}}$ điều này có được bằng cách ước lượng tham số chưa biết.(ví dụ $s/\sqrt{n}$ là ước lượng của $\sigma/\sqrt{n}$) . Dưới những điều kiện chung (cần để $s/\sqrt{n}$ dần về $\sigma_{\hat{\theta}}$ cho hầu hết mẫu). Gía trị khoảng tin cậy là $\hat{\theta} \pm z_{\alpha/2}.s_{\hat{\theta}}$.Khoảng của mẫu lớn$\overline{x} \pm z_{\alpha/2}.s/\sqrt{n}$ là một ví dụ
Cuối cùng, Gỉa sử $\sigma_{\hat{\theta}}$ liên quan đến θ chưa biết. Với trường hợp này,xét ví dụ , khi $\theta = p$,một tỉ lệ tổng thể . Thì $(\hat{\theta} - \theta)/\sigma_{\hat{\theta}} = z_{\alpha/2}$ có thể khó để giải quyết. Một giải pháp xấp xỉ có thể dùng là thay θ bằng $\sigma_{\hat{\theta}}$ rồi ước lượng theo $\hat{\theta}$. Kết quả này là ước lượng cho độ lệch chuẩn $s_{\hat{\theta}}$, khoảng tương ứng là $\hat{\theta} \pm z_{\alpha/2}.s_{\hat{\theta}}$
khi đó khoảng tin cậy là ước lượng điểm của $\theta \pm$ (giá trị điều kiện của z)(ước lượng chuẩn lỗi của người ước lượng)

KHOẢNG TIN CẬY CHO TỈ LỆ TỔNG THỂ

Đặt p là tỉ lệ thành công của tổng thể. với đối tượng có tính chất cụ thể (ví dụ nhóm đối tượng tốt nghiệp từ một trường đại học, máy tính không cần dịch vụ bảo hành,.. v..v).Mẫu ngẫu nhiên của n cá thể độc lập được chọn, và X là số cá thể thành công trong mẫu. Biết n nhỏ so với tổng thể X có thể xem là một biến ngẫu nhiên nhị thức rv với $E(X) = np$ và $\sigma_X = \sqrt{np(1-np)}$. hơn nữa, nếu $np \geq 10, (q = 1-p)$, X có phân phối xấp xỉ chuẩn.
ước lượng tự nhiên của p là $\hat{p} = X/n$,phân số mẫu thành công khi đó $\hat{p}$ có được bằng cách nhân X với hằng số $1/n$ cũng là phân phối xấp xỉ chuẩn. Biểu diễn như trong 6.1 $E(\hat{p}) = p($ không chệch) và $\sigma_{\hat{p}} = \sqrt{p(1-p)/n}$. Độ lệch tiêu chuẩn $\sigma_{\hat{p}}$ liên quan đến tham số p chưa biết. Chuẩn hóa $\hat{p}$ bằng cách trừ cho p và chia cho $\sigma_{\hat{p}}$, suy ra đươc

$$P(-z_{\alpha/2} < \frac{\hat{p} - p}{\sqrt{p(1-p)/n}} < z_{\alpha/2}) \pm 1 - \alpha$$

Tiến hành đề xuất trong mục "nguồn gốc của khoảng tin cậy" (phần 7.1) ,kết quả giới hạn tin cậy có được bằng cách thay $<$ bằng dấu $=$ và giải phương trình theo p .ta được

$$\begin{aligned} p &= \frac{\hat{p} + z_{\alpha/2}^2}{/} 2n1 + z_{\alpha/2}^2/n \pm z_{\alpha/2} \frac{\sqrt{\hat{p}(1-\hat{p})/n + z_{\alpha/2}^2}/4n^2}{1 + z_{\alpha/2}^2/n} \\ &= p^{\sim} \pm \frac{\sqrt{\hat{p}(1-\hat{p})/n + z_{\alpha/2}^2}/4n^2}{1 + z_{\alpha/2}^2/n} \end{aligned}$$

MỆNH ĐỀ

Đặt $p^{\sim} = \frac{\hat{p} + z_{\alpha/2}^2}{/} 2n1 + z_{\alpha/2}^2/n$ thì một khoảng tin cậy chi tỉ lệ của tổng thể p với độ tin cậy xấp xỉ $100(1-\alpha)\%$ là

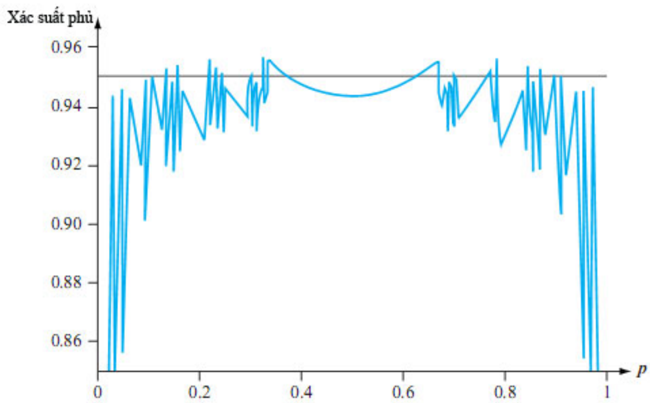
$$p^{\sim} \pm z_{\alpha/2} \frac{\sqrt{\hat{p}\hat{q}/n + z_{\alpha/2}^2}/4n^2}{1 + z_{\alpha/2}^2/n} \quad (7.10)$$

với $\hat{q} = 1 - \hat{p}$ và khi đó dấu - trong 7.10 là giới hạn dưới và dấu + là giới hạn trên.
Đây thường coi như là tỉ lệ ước lượng khoảng cho p

Nếu cỡ mẫu n rất lớn thì $z^2/2n$ không đáng kể (nhỏ) so sánh $\hat{p}$ và z^2/n là so sánh không đáng kể với 1. Từ đó $p^{\sim} \approx \hat{p}$ trong trường hợp này so sánh giữa $z^2/4n^2$ và pq/n cũng không đáng kể (n^2là số chia lớn hơn n(, chiếm ưu thế trong biểu thức $\pm$ là $z_{\alpha/2}\sqrt{\hat{p}\hat{q}/n}$ khi đó tỉ lệ ước lượng của khoảng là xấp xỉ

$$\hat{p} \pm z_{\alpha/2}\sqrt{\hat{p}\hat{q}/n} \tag{7.11}$$

Đây là khoảng cuối cùng có dạng chung là $\hat{\theta} \pm z_{\alpha/2}\hat{\sigma}_{\hat{\theta}}$ của mẫu lớn trong mục cuối này. Khoảng tin cậy xấp xỉ (7.1) đã được giới thiệu nhiều trong những sách về xác suất. Nó rõ ràng là đơn giản và dễ hiểu hơn tỉ số ước lượng. Tại sao phải quan tâm đến trường hợp thứ hai?
Đầu tiên, giả sử ta dùng $z_{.025} = 1.96$ trong công thức truyền thống (7.11), thì trên danh nghĩa độ tin cậy (ta đang nghĩ là sử dụng điều kiện của z) xấp xỉ 95%, vì vậy trước khi chọn mẫu , xác suất là khoảng ngẫu nhiên bao gồm cả giá trị thực của p (tức là xác suất phủ) nên là .95, mà theo biểu diễn của hình 7.6 cho trường hợp n=100, xác suất phủ thực sự cho khoảng có thể khác nhau đáng kể từ xác suất chuẩn .95, đặc biệt khi p không gần với .5 (hình ảnh xác suất phủ so sánh p rất phức tạp vì phân phối xác suất nhị thức cơ bản là rời rạc chứ không liên tục).Nói chung khác biệt của khoảng truyền thống và độ tin cậy thực sự có thể không đáng kể cả khi cho cỡ mẫu lớn. Những nghiên cứu gần đây cũng chỉ ra rằng tỉ lệ khoảng có thể khắc phục cho hầu như tất cả cỡ mẫu và giá trị của p. Độ tin cậy thực sự sẽ ít khác biệt do cách chọn $z_{\alpha/2}$. Điều này thực tế là do tỉ lệ khoảng gần về .5 khi so sánh với khoảng ban đầu. Đặc biệt điểm giữa $p^{\sim}$ của tỉ lệ khoảng gần về .5 hơn điểm giữa $\hat{p}$ của khoảng truyền thống . Đó là điều đặc biệt khi p nằm trong khoảng 0 đến 1.



Hình 7.6 Xác suất phủ thực sự cho khoảng trong (7.1) những giá trị khác nhau của p khi n=100
Thêm vào đó, tỉ lệ khoảng có thể sử dụng cho hầu hết các mẫu và giá trị tham số. chỉ cần kiểm tra điều kiện $n\hat{p} \geq 10$ và $n(1-\hat{p}) \geq 10$ sẽ thỏa yêu cầu của khoảng truyền thống. Câu hỏi đặt ra là khi n đủ lớn cho trong (7.11) có thu được xấp xỉ tốt hơn (7.10), lời khuyên là ta nên dùng tỉ lệ khoảng để kiểm tra. Tính toán sẽ tẻ nhạt hơn khi kiểm tra tính chất của khoảng.
VÍ DỤ 7.8
Bài viết " Sự lặp lại và tái hiện của mật khẩu /dữ liệu lỗi"(J. Kiểm tra và đánh giá $., 1997 : 151 - 153$) báo cáo rằng trong n=48 thử nghiệm ở một phòng thí nghiệm đặc biệt, 16 kết quả trong việc đánh lửa để đốt thuốc từ một loại chất nền đặc biệt. Đặt p là tỉ lệ lâu dài của các thí nghiệm việc đánh lửa. Một ước lượng điểm cho p là $\hat{p} = 16/48 = .333$. Một khoảng tin cậy cho p với độ tin cậy xấp xỉ 95% là

$$\frac{.333 + (1.96)^2/96}{1 + (1.96)^2/48} \pm \frac{\sqrt{(.333)(.667)/48 + (1.96)^2/9216}}{1 + (1.96)^2/48}$$
$$= .345 \pm .129 = (.216, .474)$$

Khoảng này khá rộng vì cỡ mẫu 48 không đủ lớn để ước lượng tỉ lệ.
Khoảng truyền thống là

$$.333 \pm 1.96\sqrt{(.333)(.667)/48} = .333 \pm .133 = (.200, .466)$$

Hai khoảng này thỏa yêu cầu với cỡ mẫu lớn hơn.
Tương tự độ rộng của khoảng tin cậy cho p một độ rộng đặc biệt w cho bởi một phương trình bậc hai với cỡ mẫu n cần thiết để có một khoảng với độ chính xác mong muốn. Rút gọn chỉ số dưới của $z_{\alpha/2}$. Ta được

$$n = \frac{2z^2\hat{p}\hat{q} - z^2w^2 \pm \sqrt{4z^4\hat{p}\hat{q}(\hat{p}\hat{q} - w^2) + w^4z^4}}{w^2} \tag{7.12}$$

Rút gọn tử số theo điều kiện của w^2 ta được

$$n \approx \frac{4z^2\hat{p}\hat{q}}{w^2}$$

Kết quả sau có được từ phương trình về độ rộng khoảng truyền thống theo w.
Điều đáng tiếc là công thức này lại liên quan đến giá trị $\hat{p}$ chưa biết. Cách tiếp cận hay nhất là dùng giá trị lớn nhất của $\hat{p}\hat{q}[= \hat{p}(1\hat{p})]$ khi $\hat{p} = .5$. Do đó nếu $\hat{p} = \hat{q} = .5$ đã dùng trong (7.12) độ rộng này sẽ đạt được tại một điểm w bất kỳ sao cho kết quả giá trị của $\hat{p}$ vẫn nằm trong mẫu. Hay nói cách khác, nếu điều tra viên tin rằng, dựa vào thông tin, mà $p \leq p_0 \leq .5$ thì p_0 có thể thay thế được cho $\hat{p}$. Lập luận tương tự cho $p \geq p_0 \geq .5$.
VÍ DỤ 7.9
Độ rộng khoảng tin cậy 95% như trong ví dụ 7.8 là .258. Gía trị n cần tìm để có độ rộng chắc chắn là .10 với $\hat{p}$ bất kỳ là

$$n = \frac{2(1.96)^2(.25) - (1.96)^2(.01) \pm \sqrt{4(1.96)^4(.25)(.25 - .01) + (.01)(1.96)^4}}{.01} = 380.3$$

Do đó cỡ mẫu cần dùng là 381. Biểu thức của n có được dựa vào khoảng tin cậy truyền thống để có được cỡ mẫu 385.

KHOẢNG TIN CẬY BẤT ĐỐI XỨNG (RÀNG BUỘC CỦA ĐỘ TIN CẬY)

Khoảng tin cậy được nói đến là có cả giới hạn dưới của độ tin cậy và giới hạn trên của độ tin cậy cho ước lượng tham số. Trong một số trường hợp , điều tra viên chỉ mong muốn hai loại ràng buộc.Ví dụ Nhà tâm lý học muốn tính 95% chặn trên của tin cậy cho trung bình đúng cho thời gian phản ứng với một kích thích đặc biệt, hay một kỹ sư mong muốn chỉ có chặn dưới cho trung bình đúng thời gian hoạt động của một loại nhất định. Vì vùng tiêu chuẩn dồn lại ở bên trái đường cong chuẩn với giá trị 1.645 là .95, nên

$$P(\frac{\overline{X} - \mu}{S/\sqrt{n}} < 1.645) \approx .95$$

Dùng bất đẳng thức trong ngoặc đơn biến đổi theo μ. và thay biến chuẩn hóa bằng cách tính toán các giá trị được bất đẳng thức $\mu > \overline{x} - 1.645s/\sqrt{n}$; Vế phải biểu diễn chặn dưới của tin cậy . Với $P(-1.645 < Z) \approx .95$ và biến đổi bất đẳng thức ta được chặn trên của tin cậy. Lập luận tương tự cho mỗi bên của ràng buộc với một độ tin cậy khác.
MỆNH ĐỀ
Một mẫu lớn có chặn trên theo μ là

$$\mu < \overline{x} + z_\alpha.\frac{s}{\sqrt{n}}$$

Một mẫu lớn có chặn dưới theo μ là

$$\mu > \overline{x} + z_{\alpha}.\frac{s}{\sqrt{n}}$$

Mỗi bên ràng buộc tin cậy theo p là kết quả của việc thay $z_{\alpha/2}$ bằng z_{α} và $\pm$ bởi + hay - trong công thức khoảng tin cậy cho p (7.10). Trong tất cả trường hợp độ tin cậy có xấp xỉ $100(1-\alpha)\%$.

VÍ DỤ 7.10

Kiểm tra độ nghiêng là phương pháp được áp dụng rộng rãi nhất khi kiểm tra chất lượng kết cấu của vật liệu bê tông . Bài viết " Kiểm tra sự kết dính giữa vật liệu và nền bê tông"(Vật liệu ACI J., 1996:553-558) Trong một báo cáo của một cuộc điều tra đặc biệt, một mẫu gồm 48 độ bền của vết cắt được trung bình mẫu có độ bền $17.17N/mm^2$ và độ lệch tiêu chuẩn của mẫu $3.28N/mm^2$. Chặn dưới của trung bình đúng độ bền của vết cắt μ với độ tin cậy 95% là

$$17.17 - (1.645)\frac{(3.28)}{\sqrt{48}} = 17.17 - .78 = 16.39$$

Nghĩa là với độ tin cậy 95% ta có $\mu > 16.39$

BÀI TẬP Phần 7.2 $(12 - 27)$

12. Mẫu ngẫu nhiên của 110 tiếng sấm sét trong một khu vực nhất định có thời gian trung bình của máy đo thời lượng tiếng vang .81 giây và độ lệch tiêu chuẩn mẫu là .34 giây (" Sét đánh vào máy bay trong một cơn giông,"của Aaicraft, $1984 : 607 - 611$ Tính khoảng tin cậy 99% (2 bên) của trung bình đúng của thời gian tiếng vang μ và giải thích kết quả cho khoảng đó.

13. Bài viết " Gas nấu ăn, thông gió nhà bếp, và sự tiếp xúc với sản phẩm đốt"(không khí trong nhà; $2006{:}65 - 73$) cho biết một thiết bị theo dõi 50 nhà bếp dùng gas để đun nấu trong một tuần., trung bình mẫu khí CO_2 có mức 654.16 (ppm), và độ lệch tiêu chuẩn là 164.43.
a. Tính và giải thích độ tin cậy 95% (hai bên) cho trung bình đúng mức khí CO_2 trong tổng thể các nhà được chọn trong mẫu .
b. Gỉa sử điều tra viên thực hiện cuộc phỏng đoán sơ bộ 175 giá trị của s trước khi chọn dữ liệu. Tìm cỡ mẫu cần thiết để độ rộng của khoảng là 50 với độ tin cậy 95%?
14. Bài viết " Đánh giá hiệu suất lò nung"(Gốm Amer Soc. Bull.,Aug. $1997 : 59 - 63$) có một số thông tin tóm tắt về độ nứt của n=169 loại gốm được nung trong một lò đặc biệt :$\overline{x} = 89.10, s = 3.73$
a. Tính khoảng tin cậy (hai bên) cho trung bình đúng của độ nứt theo tuổi sử dụng độ tin cậy 95%. Trung bình đúng của độ nứt có thể ước lượng được không?
b. giả sử điều tra viên ưu tiên một tổng thể có độ lệch chuẩn 4MPa. Dựa vào giả thiết đó Tìm độ lớn của mẫu thỏa yêu cầu cho ước lượng μ nằm trong .5MPa với độ tin cậy 95%
15. Xác định độ tin cậy có mẫu lớn ràng buộc một bên như sau:
a. Chặn trên :$\overline{x} + .84s/\sqrt{n}$
b. Chặn dưới :$\overline{x} - 2.05s/\sqrt{n}$
c. Chặn trên :$\overline{x} + .67s/\sqrt{n}$
16. Sự thay đổi điện áp của dòng điện xoay chiều cho biết điện môi của chất lỏng cách điện. Bài viết "Thực hiện thí nghiệm về thay đổi điện áp của dòng điện xoay chiều AC với một số dung môi"(bài báo về vật liệu cách điện $1995 : 21 : 26$) đưa ra kèm theo mẫu quan sát sự cố điện áp của một mạch đặc biệt trong một điều kiện nhất định

62	50	53	57	41	53	55	61	59	64	50	53	64	62	50	68
54	55	57	50	55	50	56	55	46	55	53	54	52	47	47	55

57 48 63 57 57 55 53 59 53 52 50 55 60 50 56 58

a. Xây dựng biểu đồ hộp của dữ liệu và giải thích các đặc điểm của nó

b. Tính và giải thích cho trung bình đúng của sự cố điện áp μ với độ tin cậy 95%. ước lượng theo μ có chính xác không? giải thích

c. Gỉa sử quan sát viên tin rằng hầu như tất cả giá trị của sự cố điện áp thay đổi giữa 40 và 70. Tìm cỡ mẫu để độ rộng của khoảng là 2 kV cho khoảng tin cậy 95% (sao cho μ ước lượng trong 1kV với độ tin cậy 95%)?

17 Bài tập 1.13 đưa ra một mẫu của việc quan sát độ bền của vết nứt cuối. (ksi). Dùng số liệu thống kê kèm theo để từ Minitab để tính chặn dưới cho trung bình đúng của độ bền vết nứt với độ tin cậy 95% và giải thích kết quả

N	Mean	Median	TrMean	StDev	SE Mean
153	135.39	135.40	135.41	4.59	0.37

Minimum	Maximum	Q1	Q3
122.20	147.70	132.95	138.25

18. Bài viết" tải trọng của mỏ neo khi ốc vít giãn nở"(J.Energy Engr.,1993:139-158) cho ra một bảng tóm tắt dữ liệu về sức bền cho mẫu là 3/8. ốc vít mỏ neo : $n = 78, \overline{x} = 4.25, s = 1.30$. Tính chặn dưới dùng độ tin cậy 90% cho trung bình đúng của sức bền.

19. Bài viết " giới hạn của năng lực do thị giác bị khiếm khuyết"(IEEE Trans.on Semiconductor Manuf., 1997:17-23) báo cáo rằng, trong một nghiên cứu của một quá trình kiểm tra đơn giản, 356 trường hợp không vượt qua cuộc thăm dò và 201 trường hợp vượt qua. Gỉa sử quá trình kiểm tra là ổn định , Tính ràng buộc (hai bên) cho độ tin cậy 95% của tỉ lệ không vượt qua thăm dò.

20. Báo đưa tin (tháng 10,2002) cho biết trong cuộc khảo sát 4722 trẻ Mỹ nằm trong độ tuổi 6 đến 19, 15% bị thừa cân (tỉ số khối của cơ thể ít nhất là 30 ; Tỉ số này là biểu diễn mối quan hệ giữa cân nặng và chiều cao) Tính và giải thích một khoảng tin cậy sử dụng độ tin cậy 99% cho tỉ lệ tất cả trẻ bị thừa cân.

21. Trong một mẫu ngẫu nhiên gồm 1000 khách hàng ngẫu nhiên có thể gửi đơn bồi thường sau khi mua một sản phẩm, 250 người nói rằng họ không bao giờ làm.(" bồi thường: những gì bạn xứng đáng được nhận," Báo cáo khách hàng 5/2009:7)lý do được đưa ra như sau, số lượng quá ít,trễ hạn, sợ gửi thư ,mất biên lai, và nghi ngờ việc nhận lại tiền. Tính chặn trên với độ tin cậy 95% của tỉ lệ khách hàng không bao giờ nộp đơn bồi thường.

Dựa trên khoảng này. Bằng chứng này có thuyết phục về tỉ lệ đúng của khách hàng là ít hơn 1/3? giải thích lý do đó.

22. Công nghệ thay thế khớp háng đã có sự thay đổi khi các hoạt động này trở nên phổ biến hơn (trên 250,000 ở Mỹ vào năm 2008). Bắt đầu từ năm 2003 gạc bằng gốm với độ bền cao đã được bán trên thị trường.Không may, quá nhiều bệnh nhân dùng nó với độ bền tăng thì tiếng cót két cũng tăng. Ngày 11/5/2008 Báo cáo nghiên cứu đăng trên New York Times về 143 cá nhân nhận gạc bằng gốm vào những năm 2003 đến 2005, 10 trường hợp bị phát ra tiếng kêu cót két.

a. Tính chặn dưới cho độ tin cậy 95% cho tỉ lệ đúng của các khớp háng phát ra tiếng kêu cót két.

b. giải thích việc sử dụng khoảng tin cậy 95% trong a.

23. Diễn đàn về tôn giáo và đời sống cộng đồng báo cáo vào 9/12/2009 trong một cuộc khảo sát những người trưởng thành ở Mỹ vào năm 2003 25% nói rằng họ tin vào chiêm tinh học.

a. Tính và giải thích một khoảng tin cậy với độ tin cậy 99% cho tỉ lệ những người trưởng thành ở Mỹ tin vào chiêm tinh học.

b. Tìm cỡ mẫu thoả mãn độ rộng tối đa cho khoảng tin cậy 99 % là .05 không phân biệt giá

trị $\hat{p}$.
24. Một mẫu gồm 56 mẫu vải cotton được nghiên cứu có phần trăm của mẫu kéo dài là 8.17 và độ lệch chuẩn là 1.42 (" Mối quan hệ giữa góc xoắn ốcϕ, phần trăm độ giãn E_1, và số chiều của sợi cotton".Nghiên cứu dệt vải J., $1978 : 407 - 410$).Tính trung bình đúng phần trăm co giãn μcho mẫu lớn với khoảng tin cậy 95%.Gỉa thiết nào cho biết phân phối của phần trăm độ co giãn?

25.Nhà lập pháp muốn khảo sát dân cư tại khu vực hành chính của mình để biết về tỉ lệ dân cư dùng ngân sách nhà nước cho việc phá thai.
a. Xác định cỡ mẫu cần thiết nếu trong 95% khoảng tin cậy cho p với độ rộng tối đa.10 không phân biệt p?
b. Nếu nhà lập pháp có lý do mạnh mẽ để tin rằng ít nhất 2/3 dân cư đặt vào địa vị của cô ấy ? Cỡ mẫu cần là bao nhiêu?
26.Quản lý một trường lớn của quận, có một khóa học về xác suất thống kê, tin rằng số lượng giáo viên vắng mặt trong ngày có phân phối poisson với tham số μ. Dùng dữ liệu vắng mặt kèm theo trong 50 ngày để tìm khoảng tin cậy cho mẫu lớn theo μ [ghi chú kỳ vọng và phương sai của một biến poisson cùng bằng μ, vì thế

$$Z = \frac{\overline{X} - \mu}{\sqrt{\mu/n}}$$

có phân phối xấp xỉ chuẩn. tiến hành như trong cách thành lập khoảng p bằng cách đưa ra một phát biểu về xác suất (với xác suất $1-\alpha$) và giải được kết quả theo μ- xem lập luận sau 7.10

Số người vắng mặt	0	1	2	3	4	5	6	7	8	9	10
Thường xuyên	1	4	8	10	8	7	5	3	2	1	1

27. Xét khoảng tin cậy (7.10) cho p, và tập trung vào độ tin cậy 95%. cho thấy giới hạn của độ tin cậy khá tương đồng với khoảng truyền thống (7.11)hai trường hợp thành công và 2 trường hợp thất bại đã được thêm vào mẫu [tức là (7.11) dựa trên $x + 2S'S$ trong n+4 thử nghiệm]. [ghi chú 1.96 $\approx$ 2. Chú ý: Agresti và Coull biểu diễn rằng việc điều chỉnh này khoảng cách truyền thống này cũng tạo ra một độ tin cậy sát thực tế.]

4 7.3 KHOẢNG DỰA TRÊN PHÂN PHỐI CHUẨN CỦA QUẦN THỂ

Khoảng tin cậy cho μ dùng trong phần 7.2 có nghĩa khi n lớn.Kết quả là khoảng có thể sử dụng với bất kỳ phân phối nào của tổng thể. Định lý giới hạn trung tâm CLT không thể dẫn chứng, tuy nhiên,khi n nhỏ trong trường hợp này. Một cách để tiến hành là đưa ra một giả thiết cụ thể về dạng phân phối của tổng thể. sau đó suy ra khoảng tin cậy phù hợp với giả thiết đó. Ví dụt, ta có thể đưa ra một khoảng tin cậy cho μ khi tổng thể được mô tả là có phân phối gamma., một khoảng khác cho trường hợp của phân phối Weibull, ...vv...Các nhà thống kê đã thực hiện điều này cho một số họ phân phối khác nhau. Vì phân phối chuẩn có xấp xỉ quen thuộc như bất kỳ mô hình phân phối khác của tổng thể. Ta sẽ thảo luận sau đây.

Gỉa thiết
Tổng thể có chuẩn tuyệt đối, vì thế nó tạo ra biến ngẫu nhiên $X_1, ...X_n$ từ phân phối chuẩn với μ và σ chưa biết

Kết quả chính của khoảng trong phần 7.2 là cho n lớn, chuẩn hóa rv $Z = (\overline{X} - \mu)/(S/\sqrt{n})$ có phân phối xấp xỉ chuẩn. Khi n nhỏ, S có thể không dần về σ vì thế biến phát sinh Z ngẫu nhiên trong cả tử số và mẫu số . Điều này suy ra rằng phân phối xác suất của$(\overline{X} - \mu)/(S/\sqrt{n})$ dần về phân phối chuẩn nhanh hơn. Kết quả suy luận dựa trên việc giới thiệu một họ mới của phân phối xác suất gọi là phân phối t.
Định lý

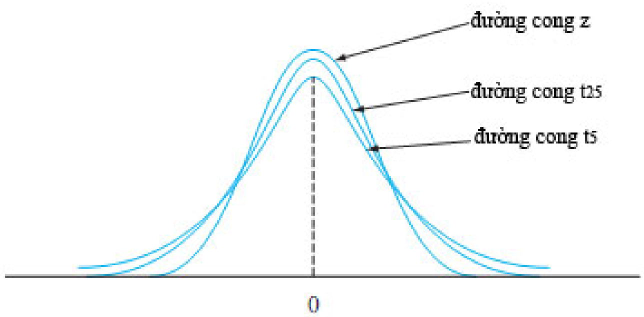
Nếu $\overline{X}$ là trung bình của mẫu ngẫu nhiên có kích thước n từ phân phối chuẩn với kỳ vọng μ, chuẩn hóa

$$T = \frac{\overline{X} - \mu}{S/\sqrt{n}} \tag{7.13}$$

có một phân phối xác suất gọi là phân phối t với n-1 bậc tự do(df)

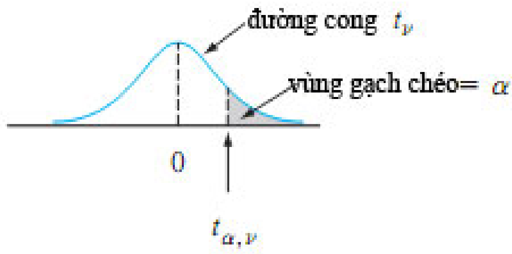
TÍNH CHẤT CỦA PHÂN PHỐI t
Trước khi áp dụng định lý này Các tính chất của phân phối t sẽ được thảo luận theo thứ tự. Mặc dù biến thích hợp vẫn là $(\overline{X} - \mu)/(S/\sqrt{n})$ bây giờ ta sẽ ký hiệu nó là T để nhấn mạnh rằng nó sẽ không có phân phối chuẩn khi n nhỏ. Nhắc lại phân phối chuẩn được xác định bởi hai tham số ; Mỗi sự lựa chọn μ kết hợp với σ cho một phân phối chuẩn đặc biệt.
Bất kỳ phân phối t đặc biệt nào cũng là kết quả từ giá trị xác định của một tham số đơn, được gọi là số của bậc tự do, viết tắt là df. Ta ký hiệu tham số này bằng chữ cái hy lạp ν.Gía trị của ν có thể đạt được là những số dương 1,2,3.... vì vậy đây là phân phối t với 1 bậc tự do, 2 bậc tự do,.....
cho một giá trị ν bất kỳ hàm mật độ nào cũng xác định một đường cong t liên quan thậm chí còn phức tạp hơn hàm mật độ chuẩn. May mắn là ta chỉ cần quan tâm đến một vài tính chất quan trọng của đường cong này.

Tính chất của phân phối t
Đặt t_ν ký hiệu là phân phối t với ν df
1. Đường cong t_ν có dạng chuông và trung điểm tại 0
2. Đường cong t_ν gần với đường cong của phân phối chuẩn (z)
3. Khi ν tăng thì đường cong t_ν giảm tương ứng
4. Khi $\nu \to \infty$, dãy đường cong t_ν xấp xỉ với đường cong của phân phối chuẩn (vì thế đường cong z thường được gọi là đường cong t với bậc tự do $df = \infty$
Hình 7.7 Biểu diễn tính chất chung cho việc chọn giá trị ν



Hình 7.7 đường cong t_ν và z Số bậc tự do df cho T trong (7.13) là $n-1$ vì, mặc dù S dựa trên n độ lệch $X_1 - \overline{X}, ..., X_n - \overline{X}, \sum(X_i - \overline{X}) = 0$ suy ra chỉ tìm được $n-1$ tự do được xác định. số bậc tự do df cho biến t là căn cứ xác định số độ lệch tự do của ước lượng độ lệch chuẩn trong mẫu số của T.
Sử dụng phân phối t sao cho vùng đuôi cong của z_α bắt nguồn từ đường cong z. Bạn có nghĩ có mẹo để tìm t_α. Tuy nhiên, Gía trị mong muốn không chỉ trên một vùng mà còn cả trên df.
Ký hiệu
Đặt $t_{\alpha,\nu}$ là số trên độ đo của trục dưới của đường cong t với nu bậc tự do df về phía bên phải của $t_{\alpha,\nu}$ là α; $t_{\alpha,\nu}$ gọi là giá trị điều kiện t.

ví dụ $t_{.05,6}$ được gọi là giá trị tới hạn được biểu diễn ở vùng bên phải của .05 dưới đường cong t với 6 df.Chú ý được minh họa trong hình 7.8.Vì đường cong t đối xứng qua 0, $-t_{\alpha,\nu}$ biểu diễn ở vùng bên trái của α. Để có được $t_{.05,15} = 1.753$ Ta đi từ cột $\alpha = .05$, chiếu xuống được dòng $\nu = 15$, và đọc được $t_{.05,15} = 1.753$. Tương tự, $t_{.05,22} = 1.717$ (cột .05, dòng $\nu = 22$, và $t_{.01,22} = 2.508$.



Hình 7.8 Minh họa cho giá trị tới hạn t

Gía trị của $t_{\alpha,\nu}$ có được bằng cách di chuyển qua hàng hoặc xuống cột.Cho ν cố định, $t_{\alpha,\nu}$ tăng khi α giảm,khi đó ta phải di chuyển nhanh hơn về phía bên phải của 0 để xác định được khu vực α trong phần đuôi. Cho α cố định, khi ν tăng (nghĩa là ta nhìn xuống bất kỳ cột nào của bảng t) giá trị của $t_{\alpha,\nu}$ giảm. vì thế khu vực đuôi của α không cần xa 0. Thêm vào đó, $t_{\alpha,\nu}$ giảm chậm khi ν tăng. Kết quả là,Bảng giá trị biểu diễn 2 mốc bậc tự do giữa 30 và 40 df và vì thế bước nhảy $\nu = 50, 60, 120$ và cuối cùng là ∞. Vì t_∞ là đường cong của phân phối chuẩn.Họ các giá trị z_α xuất hiện vào dòng cuối của bảng. Công thức được sử dụng trước đó của mẫu lớn CI (nếu $n > 40$ ta được phương trình xấp xỉ chuẩn thông thường và phân phối t cho $\nu \geq 40$.

KHOẢNG TIN CẬY CHO MỘT MẪU t

Gía trị chuẩn hóa T có phân phối t với n-1 bậc tự do df và khu vực dưới tương ứng với đường cong mật độ t giữa $-t_{\alpha/2,n-1}$ và $t_{\alpha/2,n-1}$ là $1-\alpha$(khu vực $\alpha/2$ tại mỗi bên). vì thế.

$$P(-t_{\alpha/2,n-1} < T < t_{\alpha/2,n-1}) = 1 - \alpha \tag{7.14}$$

Biểu thức (7.14) khác với biểu thức trong phần trước trong đó T và$t_{\alpha/2,n-1}$ thay cho Z và $z_{\alpha/2}$. Nhưng nó được thao tác cùng một cách để có được khoảng tin cậy cho μ.
MỆNH ĐỀ
Đặt $\overline{x}$ và s là trung bình mẫu và độ lệch tiêu chuẩn của mẫu được tính từ mẫu ngẫu nhiên có phân phối chuẩn với kỳ vọng μ. Thì một độ $100(1-\alpha)\%$ khoảng tin cậy cho μ là

$$(\overline{x} - t_{\alpha/2,n-1}).\frac{s}{\sqrt{n}}, \overline{x} + t_{\alpha/2,n-1}).\frac{s}{\sqrt{n}}) \tag{7.15}$$

hay, gọn hơn , $\overline{x} \pm t_{\alpha/2,n-1}).s/\sqrt{n}$.
Chặn trên độ tin cậy cho μ là

$$\overline{x} + t_{\alpha/2,n-1}).\frac{s}{\sqrt{n}}$$

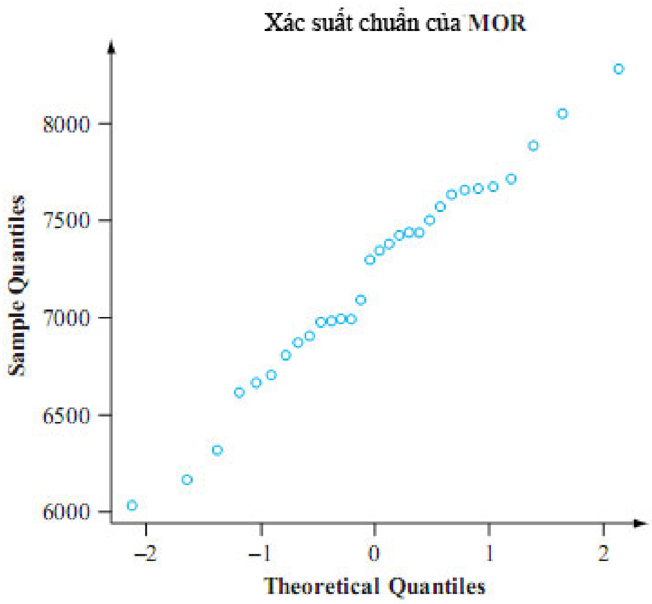
Thay dấu + bằng dấu − trong biểu thức cho ta chặn dưới độ tin cậy cho μ, với độ tin cậy $100(1-\alpha)\%$

VÍ DỤ 7.11

Khi thị trường về các loại gỗ xẻ sweetgum giảm. Phần lớn nhu cầu về các loại gỗ lớn được dùng để xây dựng cầu đường truyền thống càng tăng. Bài viết " Phát triển của nền công nghiệp Laminated Planks từ gỗ xẻ Sweetgum"(J. Kỹ sư cầu đường., 2008 : 64 − 66) mô tả nhà máy và kiểm tra cho thiết kế đầm liên hợp để tăng giá trị cho gỗ xẻ sweetgum có giá trị thấp. Dữ liệu sau đây là trị tuyệt đối của vết cắt.(bài biết tóm tắt giá trị thể hiện bằng MPa):

6807.99	7637.06	6663.28	6165.03	6991.41	6992.23
6981.46	7569.75	7437.88	6872.39	7663.19	6032.28
6906.04	6617.17	6984.12	7093.71	7659.50	7378.61
7295.54	6702.76	7440.17	8053.26	8284.75	7347.95
7422.69	7886.87	6316.67	7713.65	7503.33	7674.99

Hình 7.9 biểu diễn điểm phân phối chuẩn từ phần mềm R.Độ thẳng của các điểm cung cấp một giải thích mạnh mẽ cho phân phối của tổng thể MOR là tại đó ít xấp xỉ chuẩn nhất.



Hình 7.9 Một phân phối chuẩn điểm của dữ liệu trị tuyệt đối vết cắt

Kỳ vọng và độ lệch chuẩn của mẫu tương ứng là 7203.191 và 543.54000, (với bất kỳ tính toán nào, giảm bớt khối lượng tính toán bằng cách trừ đi 6000 từ mỗi giá trị x bao gồm $y_i = x_i - 6000$; thì $\sum y_i = 36,095$ và $\sum yy_i{}^2 = 51,997,668.77$ từ đó $\overline{y} = 1203.191$ và $s_y = s_x$ đã biết).

Bây giờ ta tính khoảng tin cậy cho trung bình đúng MOR dùng độ tin cậy 95%.Khoảng tin cậy này dựa trên $n-1 = 29$ bậc tự do, Gía trị đặc biệt t cần tìm là $t_{.025,29} = 2.045$. Khi đó khoảng ước lượng là

$$\overline{x} \pm t_{.025,29}.\frac{s}{\sqrt{n}} = 7230.191 \pm (2.045).\frac{543.54000}{\sqrt{30}}$$
$$= 7230.191 \pm 202.938 = (7000.253, 7406.129)$$

Khoảng ước lượng là $7000.253 < \mu < 7406.129$ với độ tin cậy 95%. Nếu ta sử dụng cùng công thức cho mẫu sau khi lấy mẫu, về lâu dài có thể tính độ tin cậy 95% cho khoảng tin cậy sẽ bao gồm cả μ. Khi đó μ không có giá trị nào.Ta không biết liệu tính cho khoảng tin cậy là 95% tốt hay 5% xấu . Thậm chí khi cỡ mẫu đủ lớn, khoảng tin cậy khá rộng. Đây là hệ quả của sự thay đổi mẫu đáng kể trong giá trị MOR.
Chặn dưới 95% của khoảng tin cậy cũng là kết quả của giới hạn dưới (với dấu -) và thay 2.045 với $t_{.05,29} = 1.699$.

Không may ở đây là không dễ gì chọn n để có thể có độ rộng thỏa khoảng cho t.Đó là vì độ rộng liên quan đến giá trị chưa biết (trước khi chọn dữ liệu.) s và n không chỉ thông qua $1/\sqrt{n}$ mà còn thông qua $t_{\alpha/2,n-1}$ Như kết quả cho thấy, một xấp xỉ có thể thu được bằng phép thử và sai lầm.
Trong Chương 15 ta đã thảo luận khoảng tin cậy choμ khi mẫu nhỏ mà mà giá trị cung cấp có phân phối tổng thể là đối xứng. một giả thiết yếu hơn chuẩn.Tuy nhiên, khi phân phối tổng thể là chuẩn khoảng t có xu hướng ngắn hơn bất kỳ khoảng nào khác với cùng độ tin cậy.
DỰ ĐOÁN KHOẢNG CHO MỘT GIÁ TRỊ TƯƠNG LAI ĐƠN

Trong nhiều ứng dụng, việc dự đoán một giá trị đơn của một biến quan sát tại một số thời điểm trong tương lai, tốt hơn là ước lượng giá trị trung bình của biến đó
VÍ DỤ 7.12
Xét mẫu sau về hàm lượng chất béo (theo phần trăm) n=10 mẫu xúc xích ngẫu nhiên.("Đánh giá cảm quan và cơ học về chất lượng xúc xích"), J. of texture studies , $1990 : 395 - 409$):
25.2 21.3 22.8 17.0 29.8 21.0 25.5 16.0 20.9 19.5
Gỉa sử những lựa chọn này có được từ phân phối chuẩn của tổng thể. Một khoảng tin cậy CI 95% cho (khoảng ước lượng của) trung bình tổng thể hàm lượng chất béo là:

$$\begin{aligned}\overline{x} \pm t_{.025,9}.\frac{s}{\sqrt{n}} &= 21.90 \pm 2.262.\frac{4.134}{\sqrt{10}} = 21.90 \pm 2.96\\ &= (18.94, 24.86)\end{aligned}$$

gỉa sử, Bạn chỉ ăn duy nhất một loại xúc xích và muốn dự đoán hàm lượng chất béo.Một điểm dự đoán, tương tự ước lượng điểm, $\overline{x}= 21.90$.Dự đoán này lại không mang lại bất kỳ thông tin nào về độ tin cậy hay độ chính xác.
Thiết lập chung sau: Ta có sẵn một mẫu ngẫu nhiên $X_1, X_2, ..., X_n$ từ phân phối chuẩn của tổng thể, và mong muốn dự đoán giá trị quan sát tương lai đơn X_{n+1}(ví dụ như tuổi thọ của một bóng đèn đơn được mua hay hiệu suất nhiên liệu khi thuê một chiếc xe) Một dự báo điểm là $\overline{X}$, và kết quả lỗi dự báo là $\overline{X} - X_{n+1}$.Gía trị kỳ vọng của lỗi dự báo là:

$$E(\overline{X} - \overline{X}_{n+1}) = E(\overline{X}) - E(X_{n+1}) = \mu - \mu = 0$$

Khi đó X_{n+1} là độc lập của $X_1, ...X_n$, nó cũng là độc lập của $\overline{X}$ vì vậy phương sai của lỗi dư báo là

$$v(\overline{X} - \overline{X}_{n+1}) = V(\overline{X}) + V(X_{n+1}) = \frac{\sigma^2}{n} + \sigma^2 = \sigma^2\left(1 + \frac{1}{n}\right)$$

lỗi dự báo là tổ hợp tuyến tính của các biến ngẫu nhiên độc lập, có phân phối chuẩn rv, vì vậy nó cũng là phân phối chuẩn .do đó

$$Z = \frac{(\overline{X} - X_{n+1}) - 0}{\sqrt{\sigma^2\left(1 + \frac{1}{n}\right)}} = \frac{\overline{X} - X_{n+1}}{\sqrt{\sigma^2\left(1 + \frac{1}{n}\right)}}$$

Có phân phối chuẩn, có thể biểu diễn lại bằng cách thay σ bằng độ lệch tiêu chuẩn mẫu S (của $X_1, ...X_n$), kết quả là

$$T = \frac{\overline{X} - X_{n+1}}{S\sqrt{1 + \frac{1}{n}}} \sim t$$

phân phối t với n-1 df
Thao tác biến đổi T thành $T = (\overline{X} - \mu)(S/\sqrt{n})$ với một khoảng tin cậy,có kết quả.
MỆNH ĐỀ
Một khoảng dự đoán (PI) cho một quan sát đơn được chọn từ một tổng thể có phân phối chuẩn là

$$\overline{x} \pm t_{\alpha/2,n-1}.s\sqrt{1 + \frac{1}{n}}$$

Mức dự đoán là $100(1 - \alpha)\%$. Chặn dưới của dự đoán có kết quả từ việc thay $t_{\alpha/2}$ bởi t_α và loại bỏ dấu + trong phần (7.16) làm tương tự cho chặn trên của dự đoán.
giải thích cho việc dự đoán 95% mức dự đoán tương tự độ tin cậy 95%;Nếu khoảng (7.16) được tính cho mẫu sau khi lấy mẫu, về lâu dài 95% khoảng này sẽ bao gồm những giá trị X phù hợp trong tương lai.
VÍ DỤ 7.13 (tiếp ví dụ 7.12)
Với $n = 10, \overline{x} = 21.90, s = 4.134$, và $t_{.025,9} = 2.262$, một khoảng dự đoán cho hàm lượng chất béo cho 1 loại xúc xích là

$$21.90 \pm (2.262)(4.134)\sqrt{1 + \frac{1}{10}} = 21.90 \pm 9.81 = (12.09, 31.71)$$

Khoảng này khá rộng, cho thấy độ chắc chắn không đáng kể về hàm lượng chất béo Chú ý là độ rộng của khoảng dự đoán PI gấp 3 lần so với CI.
Lỗi ước lượng là $\overline{X} - X_{n+1}$, là sai khác giữa hai biến ngẫu nhiên, trong khi ước lượng lỗi là $\overline{X} - \mu$, sai khác giữa hai biến ngẫu nhiên và một giá trị cố định (chưa biết). PI thì rộng hơn CI vì lỗi dự đoán (do X_{n+1}) có nhiều biến hơn ước lượng lỗi.Thật ra, với n lớn tùy ý, khoảng CI thu nhỏ lại giá trị đơn μ, và khoảng PI dần về $\mu \pm z_{\alpha/2}.\sigma$. Khi đó tìm được giá trị đơn X không chắc chắn thậm chí khi đó không cần phải ước lượng
KHOẢNG CHẤP NHẬN ĐƯỢC

Xét một tổng thể ô tô của một loại nhất định,và giả sử dưới những điều kiện xác định, tiết kiệm nhiên liệu (mpg) có phân phối chuẩn với $\mu = 30$ và $\sigma = 2$.Thì khi đó khoảng từ -1.645 đến 1.645 lấy trong 90% vùng bên dươi của đường cong z, 90% các xe ô tô có giá trị tiết kiệm nhiên liệu trong giá trị $\mu - 1.645\sigma$ và $\mu + 1.645\sigma = 33.29$.Nhưng nếu giá trị μ và σ chưa biết thì sao? ta có thể có cỡ mẫu n, xác định mức tiết kiệm nhiên liệu , $\overline{x}$ và s và từ giới hạn dưới là $\overline{x} - 1.645s$ và giới hạn trên là $\overline{x} + 1.645s$.Tuy nhiên vì các biến của mẫu có ước lượng là μ và σ. Thay đổi để kết quảcủa khoảng bao gồm sẽ ít hơn 90% giá trị tổng thể.Bằng trực quan, để có một ưu tiên 95% thay đổi để kết quảcủa khoảng bao gồm sẽ ít hơn 90% giá trị tổng thể, Khi $\overline{x}$ và s được thay cho μ và σ ta cũng nên thay 1.645 bằng một vài số lớn hơn. Ví dụ, khi $n = 20$ giá trị 2.310 và như thế ta có thể đạt 95% tin cậy,với khoảng $\overline{x} \pm 2.310$ sẽ bao gồm ít nhất 90% của giá trị tiết kiệm nhiên liệu trong tổng thể.

Đặt k là số nằm giữa 0 và 100. Khoảng chấp nhận được lấy tại ít nhất k% của giá trị trong phân phối chuẩn của tổng thể với độ tin cậy 95% có dạng là:
$\overline{x} \pm$ (Điều kiện để giá trị chấp nhận được).s
Điều kiện để giá trị chấp nhận được cho $k = 90, 95$ và 99 trong tổ hợp với giá trị cỡ mẫu được

cho trong bảng phụ lục A.6. Bảng này cũng bao gồm giá trị điều kiện cho độ tin cậy 99%(Gía trị đó lớn hơn giá trị phù hợp 95%).Thay $\pm$ bằng + được một chặn trên chấp nhận được và dùng dấu - thay cho $\pm$ được kết quả là chặn dưới chấp nhận được. Gía trị điều kiện để có được mỗi ràng buộc cũng có trong bảng phụ lục A.6

VÍ DỤ 7.14

Một phần trong dự án lớn về nghiên cứu về xử lý tấm kết cấu vỏ chịu lực, Một thành phần kết hợp đang được sử dụng ở Bắc Mỹ, Bài viết "Tính uốn dập phụ thuộc thời gian" (J.of Testing and Eval.,1996:187-193) biết các tính chất cơ học khác nhau của mẫu gỗ thông Scotch. Xét những quan sát về mô dul độ co giãn (Mpa) thu được sau 1 phút chạy trong một cấu hình chắc chắn

10,490	16,620	17,300	15,480	12,970	17,260	13,400	13,900
13,630	13,260	14,370	11,700	15,470	17,840	14,070	14,760

Có một sự tuyến tính của điểm trong xác suất chuẩn của dữ liệu. Tóm tắt có liên quan đến chất lượng là $n = 16, \overline{x} = 14,532.5, s = 2055.67$. cho một độ tin cậy 95%, hai vế chấp nhận được của khoảng cho thấy tại mức ít hơn 95% mô dul giá trị đàn hồi của uốn dập trong tổng thể mẫu sử dụng điều kiện chịu nhiệt có giá trị 2.903. kết quả khoảng là

$$14,532.5 \pm (2.903)(2055.67) = 14,532.5 \pm 5967.6 = (8,564.9, 20,500.1)$$

Ta có thể có độ tin cậy tại mức cao mà ít hơn 95% của tất cả mẫu gỗ vớigiá trị mô dul của độ đàn hồi nằm giữa 8,564.9 và 20,500.1

Khoảng tin cậy CI 95%cho μ là $(13,437.3, 15,627.7)$, và 95% khoảng dự đoán cho mô dul của độ đàn hồi của mẫu gỗ là $(10,017.0, 19,048.0)$ cả hai khoảng dự đoán và khoảng chấp nhận được có độ rộng đáng kể hơn khoảng tin cậy.

KHOẢNG DỰA TRÊN PHÂN PHỐI KHÁC THƯỜNG CỦA TỔNG THỂ

Một mẫu khoảng tin cậy t cho μ từ mạnh mẽ tới nhỏ hay thậm chí lệch ra khỏi bình thường trừ khi n khá nhỏ.Điều này có nghĩa là nếu giá trị điều kiện cho 95% tin cậy, ví dụ , được dùng để tính khoảng, độ tin cậy thực sự gần với độ chuẩn là 95%. Nếu, với n nhỏ và phân phối tổng thể cao hơn chuẩn , thì độ tin cậy thực sự có sự khác biệt đáng kể so với mức giá trị đặc biệt có trong bảng t. nó sẽ gây ra phiền phức để tin rằng độ tin cậy của bạn là 95% trong khi thực ra là chỉ đạt được 88%. Kỹ thuật bootstrap , giới thiệu trong 7.1 tìm ra thành công hơn tại ước lượng tham số với giá trị của độ rộng các biến khác thường.

Ngược lại với khoảng tin cậy giá trị của khoảng dự đoán và chấp nhận được mô tả trong chương này bị ràng buộc chặt để có được giả thiết chuẩn. Những khoảng sau không nên sử dụng với những bằng chứng về chuẩn không rõ ràng. Tài liệu tham khảo về khoảng thống kê, được trích dẫn trong thư mục cuối của chương này,thảo luận về các thủ tục thay thế trong những tình huống khác nhau.

BÀI TẬP 7.3 $(28 - 41)$ 28. Xác định giá trị của những số sau:

a. $t_{.1,15}$

b. $t_{.05,15}$

c.$t_{.05,25}$

d. $t_{.05,40}$

e. $t_{.005,40}$

29 Xác định giá trị điều kiện t (s) mà sẽ cho ra vùng mong muốn của đường cong t trong mỗi trường hợp sau:

a. vùng trung tâm = .95, df=10

b. vùng trung tâm = .95, df=20

c. vùng trung tâm = .99, df=20

d. vùng trung tâm = .99, df=50
e. vùng bên phải =.01,df=25
f. vùng bên trái =.025,df=5
30. Xác định giá trị điều kiện t cho hai bên của khoảng tin cậy trong mỗi tình huống sau:
a. Độ tin cậy=95%,df=10
b. Độ tin cậy=95%,df=15
c. Độ tin cậy=99%,df=15
d. Độ tin cậy=99%,n=5
e. Độ tin cậy=98%,df=24
f. Độ tin cậy=99%,n=38
31. Xác định giá trị điều kiện t cho khoảng chặn trên và chặn dưới của mỗi tình huống mô tả trong bài tập 30.
32. Theo bài viết " Kiểm tra về sức chịu đựng của bao cao su" (Polymer Testing, $2009{:}567-571$)," Kiểm tra việc sử dụng bao cao su hiện tại và những thách thức phải đối mặt khi sử dụng " bao gồm kiểm tra lỗ hổng, kiểm tra độ phồng, kiểm tra niêm phong của gói, và kiểm tra kích thước ,chất lượng chất bôi trơn (miền thích hợp để thực hiện các phương pháp thống kê) điều tra viên xây dựng một cách kiểm tra mới mà thêm chủng tuần hoàn cho khả năng rách. Một mẫu gồm 20 bao cao su của cùng một loại kết quả thu được với số trung bình mẫu của 1584 và một độ lệch tiêu chuẩn mẫu 607. Tính và giải thích khoảng tin cậy tại 99% độ tin cậy cho trung bình đúng số chu kỳ rách [Chú ý: Bài viết hiện tại là kiểm tra giả thiết dựa trên phân phối t; tính hợp lý của chúng phụ thuộc vào giả thiết phân phối chuẩn của tổng thể] 33. Bài viết "Độ đo và những hiểu biết về sự lão hóa của giấy Kraft với máy biến áp điện"(IEEE Electrical Insul.Mag., 1996:28-34) Bao gồm những quan sát saub dựa trên sự trùng hợp của mẫu giấy và độ nhớt theo thời gian được sắp xếp theo thứ tự giảm:

418	421	421	422	425	427	431
434	437	439	446	447	448	453
454	463	465				

a. Xây dựng biểu đồ hộp của dữ liệu và phản hồi về những tính năng có lợi
b. Có hợp lý không khi quan sát mẫu được chọn từ phân phối chuẩn?
c. Tính hai vế cho khoảng tin cậy 95% khoảng tin cậy cho trung bình đúng độ trùng hợp (như ý tác giả bài viết)Khoảng nào được đề nghị sao cho 440 là giá trị hợp lý cho trung bình đúng? độ trùng hợp của 440 là gì?
34. Một mẫu của 14 mẫu chung của một loại xác định cho trung bình mẫu tỉ lệ giới hạn hiệu suất 8.48 Mpa và một độ lệch tiêu chuẩn mẫu .79 Mpa (" Đặc điểm của yếu tố chịu lực trong kết cấu bằng gỗ"J.of Structural Engr., 1997:326-332).
a. Tính và giải thích chặn dưới cho trung bình đúng tỉ lệ giới hạn hiệu suất của mối nối với một độ tin cậy 95%.điều gì xảy ra, nếu bất kỳ giả thiết đã làm về phân phối của tỉ lệ giới hạn hiệu suất.
b. Tính và giải thích 95%chặn dưới cho tỉ lệ giới hạn hiệu suất của mẫu nối đơn cho loại này.
35. Cấy ghép mũi bằng silicon để sửa các dị tật bẩm sinh. Thành công của phương pháp này phụ thuộc vào tính chất của các cơ sinh học của mũi và màng cơ.Bài viết " Phẫu thuật chỉnh hình cơ sinh học " (J. of Med. Engr. and Tech., $2005 : 14-17$) cho biết một mẫu gồm 15 (mới chết) người trưởng thành, Mức độ thất bại trung bình (%) là 25.0 và độ lệch chuẩn là 3.5
a. Gỉa sử một phân phối chuẩn cho mức độ thất bại, ước lượng trung bình đúng cho mức độ theo cách truyền đạt thông tin về độ chính xác và độ tin cậy.
b. Dự đoán mức độ cho một cá thể người trưởng thành theo cách truyền đạt thông tin về độ chính xác và độ tin cậy.So sánh những dự đoán để tính ước lượng trong phần (a)
36. n=26 quan sát thời gian bốc hơi được đưa ra trong bài tập 36 của Chương 1 cho trung bình mẫu và độ lệch tiêu chuẩn mẫu tương ứng là 370.69 và 24.36,
a. Tính một chặn trên của tin cậy cho trung bình tổng thể thời gian giải thoát sử dụng độ tin

cậy 95%
b. Tính chặn trên của dự đoán cho thời gian giải thoát của một công nhân bổ sung đơn lẻ dùng một mức dự đoán 95%. So sánh ràng buộc này với ràng buộc độ tin cậy trong câu (a)
c. Gỉa sử 2 công nhân được thêm vào sẽ được chọn tham gia một bài tập mô phỏng thoát thân. Ký hiệu thời gian giải thoát này là X_{27} và X_{28}, và đặt $\overline{X}_{new}$, Tính 95% hai bên khoảng dựa trên dữ liệu giải thoát đã cho.
37. Nghiên cứu về khả năng đi bộ của một cá nhân trên đường thẳng ("Chúng ta có thể đi thẳng không?" Amer: J. of Physical Anthro.,1992:19-27) Báo cáo kèm theo về dữ liệu nhịp điệu (tốc độ trên mỗi giây) cho một mẫu $n = 20$ lựa chọn ngẫu nhiên sức khỏe người.

.95	.85	.92	.95	.93	.86	1.00	.92	.85	.81
.78	.93	.93	1.05	.93	1.06	1.06	.96	.81	.96

Xác suất chuẩn biểu đồ hỗ trợ đáng kể cho giả thiết mà phân phối tổng thể của nhịp điệu là xấp xỉ chuẩn. Một bảng tóm tắt mô tả dữ liệu từ phần mềm minitab như sau:

```
Variable N  Mean   Median TrMean StDev  SEMean
cadence 20 0.9255 0.9300 0.9261 0.0809 0.0181

Variable    Min    Max     Q1     Q3
cadence   0.7800 1.0600 0.8525 0.9600
```

a. Tính và giải thích 95% khoảng tin cậy cho kỳ vọng nhịp điệu của tổng thể.
b. Tính và giải thích 95% khoảng tin cậy cho kỳ vọng nhịp điệu của một cá thể ngẫu nhiên được chọn từ tổng thể.
c. Tính và giải thích khoảng bao gồm ít nhất 99% của nhịp điệu trong phân phối tổng thể sử dụng độ tin cậy 95%
38. Một mẫu gồm 25 miếng ghép dùng trong bảng mạch của nhà máy được chọn, và số lượng bị biến dạng (trong) điều kiện đặc biệt được xác định cho mỗi miếng, kết quả trong trung bình mẫu biến dạng .0635 và một độ lệch chuẩn .0065.
a. Tính dự đoán cho số lượng biến dạng của một miếng ghép bằng cách cung cấp thông tin về độ chính xác và độ tin cậy.
b. Tính một khoảng mà mức tin cậy ít nhất 95% của tất cả miếng ghép với kết quả trong số biến dạng phải nằm giữa hai giới hạn của khoảng
39. Bài tập 72 của chương 1 có quan sát về một biện pháp ràng buộc cụ thể (Điều chỉnh phân phối âm lượng) cho một mẫu gồm 13 cá nhân khỏe mạnh : 23 39 40 41 43 47 51 58 63 66 67 69 72.
a. Phân phối của tổng thể có hợp lý không từ mẫu được chọn là chuẩn?
b. Tính khoảng mà có thể 95% độ tin cậy mà ít nhất 95% của tất cả cá nhân khỏe mạnh trong tổng thể có phân phối điều chỉnh âm lượng nằm giữa giới hạn của khoảng.
c. Dự đoán phân phối điều chỉnh âm lượng của một cá nhân độc lập bằng cách tính 95% khoảng dự đoán. Khoảng này có độ rộng được so sánh như thế nào với khoảng đã tính trong phần ?(b)
40. Bài tập 13 của Chương 1 trình bày một mẫu n=153 quan sát trên sức căng kéo dài, và bài tập 17 của phần trước cho tóm tắt số lượng và yêu cầu một mẫu lớn cho khoảng tin cậy. Vì cỡ mẫu lớn, không có giả thiết về phân phối của tổng thể được yêu cầu cho tính hợp lệ của khoảng tin cậy CI.
a. Có bất kỳ giả thiết nào về phân phối của sức căng của yêu cầu trước đó để tính chặn dưới của dự đoán cho sức căng của mẫu được chọn mà sử dụng phương pháp mô tả trong phần này không? giải thích.
b. Dùng phần mềm thống kê để xem xét lại tính hợp lý của phân phối chuẩn của tổng thể.
c. Tính một chặn dưới của dự đoán với độ dự đoán 95%cho sức căng kéo dài của mẫu kế tiếp được chọn.
41.Một mở rộng của bảng xếp hạng cho giá trị điều kiện t có trong sách này biểu diễn rằng cho phân phối t với 20df ,vùng bên phải có giá trị là .687 .860 và 1.064 tương ứng với .25, .20 và

.15. Độ tin cậy bằng bao nhiêu cho 3 khoảng tin cậy sau cho kỳ vọng μ của phân phối chuẩn của tổng thể ? Nên sử dụng khoảng nào trong 3 khoảng này, và tại sao?
a.$(\overline{x} - .687s/\sqrt{21}, \overline{x} + 1.725s/\sqrt{21})$
b.$(\overline{x} - .860s/\sqrt{21}, \overline{x} + 1.325s/\sqrt{21})$
c. $(\overline{x} - 1.064s/\sqrt{21}, \overline{x} + 1.064s/\sqrt{21})$

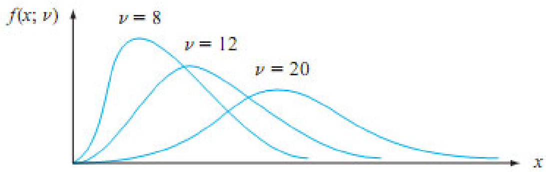
5 7.4 KHOẢNG TIN CẬY CỦA PHƯƠNG SAI VÀ ĐỘ LỆCH CHUẨN CỦA MỘT TỔNG THỂ CHUẨN

Mặc dù những lập luận liên quan đến phương sai của tổng thể σ^2 hay độ lệch chuẩn σ thường ít được quan tâm hơn kỳ vọng hay tỉ lệ của chúng, nhưng đôi khi những thủ thuật như vậy cũng cần đến. Trong trường hợp đó thì phân phối chuẩn của tổng thể sẽ dựa trên kết quả tập trung ở phương sai mẫu S^2.
Định lý
Đặt $X_1, X_2, ..., X_n$ là mẫu ngẫu nhiên từ phân phối chuẩn với tham số μ và σ^2. Thì rv

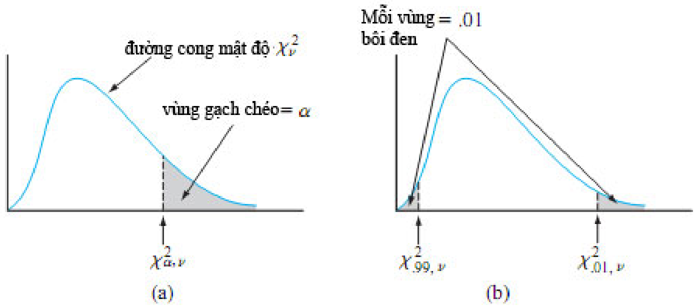
$$\frac{(n-1)S^2}{\sigma^2} = \frac{\sum(X_i - \overline{X})^2}{\sigma^2}$$

có phân phối chi bình phương (χ) với n-1 bậc tự do df

Như đã thảo luận trong phần 4.4 và 7.1 phân phối chi bình phương là phân phối liên tục với tham số đơn ν, gọi là số bậc tự do , với giá trị dương $1, 2, 3....$ Hình ảnh chung hàm mật độ xác suất của χ^2 $(pdf's)$ được minh họa trong hình 7.10 .Mỗi pdf$f(x;\nu)$ có được khi x>0, với mỗi độ lệch dương (kéo dài phần đuôi trên) dù phân phối di chuyển để đối xứng hơn khi ν tăng. Suy luận quá trình này dùng phân phối chi bình phương ta cần biểu diễn tương tự cho một giá trị điều kiện t là $t_{\alpha,\nu}$.



Hình 7.10 Minh họa hàm mật độ chi bình phương
Chú ý
Đặt $\chi^2_{\alpha,\nu}$ gọi là giá trị điều kiện chi bình phương, ký hiệu những số trên trục ngang sao cho α của vùng dưới đường cong chi bình phương với ν df nằm về phía bên phải của $\chi^2_{\alpha,\nu}$
Sự đối xứng của phân phối t cần thiết để xây dựng bảng giá trị điều kiện cho phần đuôi phía trên ($t_{\alpha,\nu}$ cho giá trị α nhỏ). Phân phối chi bình phương không đối xứng , vì thế bảng phụ lục A.7 có chứa giá trị của $\chi^2_{\alpha,\nu}$ cả khi α gần 0 và gần 1. Như minh họa trong 7.11(b). Ví dụ , $\chi^2_{.025,14} = 26.119$, và $\chi^2_{.95,20}$(phần trăm thứ 5) $= 10.851$



Hình 7.11 Minh họa cho chú ý của $\chi^2_{\alpha,\nu}$

rv $(n-1)S^2/\sigma^2$ thỏa hai tính chất mà dựa vào phương pháp tổng quát để có CI: Đó là một hàm của tham số có quan hệ với σ^2. đó chưa phải là phân phối xác suất (chi - bình phương) không phụ thuộc tham số. vùng dưới của đường cong chi - bình phương với ν df qua bên phải của $\chi^2_{\alpha,\nu}$ là $\alpha/2$, cũng như qua trái của $\chi^2_{1-\alpha/2,\nu}$. Do đó vùng có được mà hai giá trị điều kiện là $1-\alpha$. như 1 dãy trong định lý vừa nói đến

$$P\left(\chi^2_{1-\alpha/2,n-1} < \frac{(n-1)S^2}{\sigma^2} < \chi^2_{\alpha/2,n-1}\right) = 1-\alpha \tag{7.17}$$

Bất đẳng thức trong (7.17) tương đương với

$$\frac{(n-1)S^2}{\chi^2_{\alpha/2,n-1}} < \sigma^2 < \frac{(n-1)S^2}{\chi^2_{1-\alpha/2,n-1}}$$

Thay giá trị tính s^2 vào giới hạn của CI cho σ^2, và lấy căn bậc hai cho mỗi khoảng của σ

Một $100(1-\alpha)\%$ khoảng tin cậy cho phương sai σ^2 của một tổng thể chuẩn có giới hạn dưới là

$$(n-1)s^2/\chi^2_{\alpha/2,n-1}$$

Và giới hạn trên

$$(n-1)s^2/\chi^2_{1-\alpha/2,n-1}$$

Một khoảng tin cậy cho σ có giới hạn trên và giới hạn dưới là căn bậc hai của giới hạn tương ứng của khoảng cho σ^2. Một chặn trên hay dưới là kết quả của việc thay $\alpha/2$ với αtrong giới hạn tương ứng của CI.

VÍ DỤ 7.15

Dữ liệu đi kèm về sự phóng điện áp của các mạch điện được đọc từ một xác suất điểm chuẩn mà xuất hiện trong bài viết "thiệt hại của bảng mạch in kết hợp với điện áp của tia sét " (IEEE, Transactions on Components, Hybrids, and Manuf. Tech., 1985 : $214-220$). Từ thảo luận đó ta có đường thẳng của điểm về phóng điện áp có phân phối xấp xỉ chuẩn.

1470 1510 1690 1740 1900 2000 2030 2100 2190

2200 2290 2380 2390 2480 2500 2580 2700

Đặt σ^2 ký hiệu phương sai của phân phối về phóng điện áp. Tính giá trị của phương sai mẫu là $s^2 = 137,324.3$, ước lượng điểm của σ^2. với $df = n-1 = 16$, một 95% CI cần cho $\chi^2_{.975,16} = 6.908$

và $\chi^2_{.025,16} = 28.845$. Khoảng là

$$\left(\frac{16(137,324.3)}{28.845}, \frac{16(137,324.3)}{6.908}\right) = (76,172.3, 318,064.4)$$

Lấy căn bậc hai cho giá trị cuối được $(276.0, 564.0)$ với 95% CI cho σ. Khoảng này khá rộng, phản ánh giá trị đáng kể của phóng điện áp kết hợp với cỡ mẫu nhỏ.
CIs cho σ^2 và σ khi phân phối của tổng thể không là chuẩn có thể khó. trong trường hợp đó, cần tham khảo thêm một số kiến thức về thống kê.

BÀI TẬP Phần 7.4 $(42 - 46)$
42. Xác định giá trị của những số sau:
a. $\chi^2_{.1,15}$
b. $\chi^2_{.1,25}$
c. $\chi^2_{.01,25}$
d. $\chi^2_{.005,25}$
e. $\chi^2_{.99,25}$
f. $\chi^2_{.995,25}$
43 Xác định
a. Bách phân vị thứ 95 của phân phối chi bình phương với $\nu = 10$
b. Bách phân vị thứ 5 của phân phối chi bình phương với $\nu = 10$
c. $P(10.98 \leq \chi^2 \leq 36.78)$ Khi χ^2 là chi bình phương rv với $\nu = 22$
d. $P(\chi^2 < 14.611 \ hay \ \chi^2 > 37.652)$, Khi χ^2 là chi bình phương rv với $\nu = 25$
44. Số lượng của sự nở ngang (mils) được xác định khi cho một mẫu $n = 9$ Xung điện mối hàn hồ quang kim loại. dùng trong bể chứa của tàu LNG Độ lệch chuẩn mẫu là $s = 2.81$ mils. Gỉa sử chuẩn Tìm khoảng tin cậy 95%CI cho σ^2 và cho σ
45. Quan sát sau được dựa trên độ đàn hồi của tấm thép niken maraging[" kiểm tra độ gãy của mối hàn," $ASTM\ Speccial\ Publ.\ No.\ 381,\ 1965:\ 328-356$(trong ksi $\sqrt{in}$, cho bởi giá trị tăng:

69.5	71.9	72.6	73.1	73.3	73.5	75.5	75.7
75.8	76.1	76.2	76.2	77.0	77.9	78.1	79.6
79.7	79.9	80.1	82.2	83.7	93.7		

Tính khoảng tin cậy 99% CI cho độ lệch chuẩn của phân phối độ đàn hồi . Khoảng này có giá trị bao nhiêu bất kể bản chất phân phối là gì? giải thích.
46. Bài viết "áp suất bê tông lên ván khuôn"($Mag.\ Concrete\ Res.,\ 2009:407-417$) cho kết quả quan sát trên áp suất lớn nhất kN/m^2:

33.2	41.8	37.3	40.2	36.7	39.1	36.2	41.8
36.0	35.2	36.7	38.9	35.8	35.2	40.1	

a. Có hợp lý không khi nói mẫu này được chọn từ phân phối chuẩn của tổng thể?
b. Tính chặn trên với độ tin cậy 95%cho độ lệch tiêu chuẩn của tổng thể của áp suất lớn nhất.
BÀI TẬP THÊM $(47 - 62)$

47. Ví dụ 1.11 quan sát về độ bền liên kết có dữ liệu như sau:

11.5	12.1	9.9	9.3	7.8	6.2	6.6	7.0
13.4	17.1	9.3	5.6	5.7	5.4	5.2	5.1
4.9	10.7	15.2	8.5	4.2	4.0	3.9	3.8
3.6	3.4	20.6	25.5	13.8	12.6	13.1	8.9
8.2	10.7	14.2	7.6	5.2	5.5	5.1	5.0
5.2	4.8	4.1	3.8	3.7	3.6	3.6	3.6

a. ước lượng trung bình đúng cho độ bền liên kết bằng cách truyền tải thông tin về độ chính xác và độ tin cậy.[biết $\sum X_i = 387.8$ và $\sum X_i^2 = 4247.08$.]

b. Tính khoảng tin cậy 95% cho tỉ lệ của tất cả những liên kết mà giá trị độ bền vượt quá 10.
48. Một cuộc thi về thể thao ba môn phối hợp bao gồm bơi lội, đạp xe và chạy là một trong những sự kiện thể thao cực kỳ vất vả của vận động viên.Bài viết " Hệ quả phản ứng của tim mạch và nhiệt với thể thao" (medicine and Science in sport and excercise, 1998:385-389) báo cáo một nghiên cứu được thực hiện cho 4 vận động viên nam. tỉ lệ tim đập nhanh nhất (nhịp/giây) được ghi lại trong quá trình thực hiện mỗi 3 sự kiện. Với bơi lội, trung bình mẫu và độ lệch chuẩn tương ứng là 188.0 và 7.2, Gỉa sử tỉ lệ nhịp đập của tim có phân phối là (xấp xỉ) chuẩn, xây dựng khoảng tin cậy 98% cho trung bình đúng nhịp tim đập của vận động viên trong môn bơi lội.
49. Một trong 18 lõi được bảo quản từ bể chứa dầu ướt carbonat, số lượng khí dư bão hòa sau khi bơm dung môi đo được khi nước tràn ra quan sát thấy phần trăm của khối lượng lỗ rỗng là:

23.5	31.5	34.0	46.7	45.6	32.5
41.4	37.2	42.5	46.9	51.5	36.4
44.5	35.7	33.5	39.3	22.0	51.2

(xem " Nghiên cứu về khả năng thẩm thấu dòng nước khí của đá carbonat sau khi bơm dung môi" soc.of petroleum Engineers.J.,1976:23-30)
a. Xây dựng hộp dữ liệu cho dữ liệu này, và giải thích một tính năng thú vị bất kỳ.
b. Có hợp lý không khi mẫu được chọn từ phân phối chuẩn của tổng thể?
c. Tính khoảng tin cậy 95% cho trung bình đúng số lượng khí dư bão hòa.
50. Một bài báo cho rằng một mẫu có kích thước 5 dùng là cơ sở để tính một khoảng tin cậy 95% cho trung bình đúng của tần số riêng (Hz)của dầm phân cách của cùng một loại. Kết quả khoảng là $(229.764, 233.504)$ Độ tin cậy cho quyết định này là 99% phù hợp hơn mức sử dụng là 95%. giới hạn của khoảng 99%là gì? [chú ý: dùng trung tâm của khoảng và độ rộng của nó để xác định $\overline{x}$ và s]
51. Vào tháng 4 năm 2009 khảo sát 2253 người Mỹ trưởng thành được tiến hành bởi Trung tâm nghiên cứu mạng và dự án cuộc sống của người Mỹ tiết lộ rằng 1262 người trả lời đã từng sử dụng mạng không dây cho truy cập trực tuyến.
a. Tính và giải thích một khoảng tin cậy 95% cho tỉ lệ của tất cả người Mỹ trưởng thành những người mà tại thời điểm điều tra có sử dụng mạng không dây cho truy cập trực tuyến
b. Cỡ mẫu nào thỏa mãn nếu độ rộng mong muốn của khoảng tin cậy 95% tối đa là .04, không phân biệt kết quả mẫu như thế nào?
c. giới hạn trên của khoảng trong (a) với 95% chặn trên của độ tin cậy cho tỉ lệ ước lượng là bao nhiêu? giải thích.
52. Nồng độ cao của chất độc trong nguyên tố asen phổ biến ở nước ngầm. Bài viết "Đánh giá hệ thống xử lý loại bỏ asen ra khỏi mạch nước ngầm" (practice Periodical of Hazardous, Toxic, and Radioactive Waste Mgmt., 2005:152-157) báo cáo rằng một mẫu $n = 5$ mẫu nước được chọn cho chữa trị rối loạn đông máu, Trung bình mẫu asen tập trung là $24.3\mu g/L$, và độ lệch chuẩn mẫu là 4.1. Tác giả của bài báo sử dụng phương pháp cơ sở t để phân tích dữ liệu của chúng, Vì thế hy vọng đó là lý do để tin rằng phân phối của sự tập trung asen là chuẩn.
a. Tính khoảng tin cậy 95% cho trung bình đúng tập trung asen trong tất cả mẫu nước ngầm.
b. Tính chặn trên với độ tin cậy 90% cho độ lệch chuẩn của phân phối của tập trung asen
c. Dự đoán tập trung asen cho một mẫu đơn nước ngầm bằng cách cung cấp thông tin về độ chính xác và độ tin cậy.
53. Phá hoại của rệp cây của cây ăn trái có thể diệt bằng cách phun thuốc trừ sâu hay bằng cách ngâm nước với bọ rùa. Trong một vùng đặc biệt, trong 4 vườn khác nhau của cây ăn trái được chọn cho thử nghiệm. Ba vườn đầu tiên được phun thuốc trừ sâu tương ứng 1, 2 và 3, vườn 4 thì dùng bọ rùa, được kết quả cho dưới bảng:

Điều trị	n_i =số cây	$\overline{x}_i$(giạ/cây)	s_i
1	100	10.5	1.5
2	90	10.0	1.3
3	100	10.1	1.8
4	120	10.7	1.6

Đặt μ_i = trung bình đúng vùng (giạ/cây) sau khi diệt, thì:

$$\theta = \frac{1}{3}(\mu_1 + \mu_2 + \mu_3) - \mu_4$$

đo được sự khác biệt của trung bình đúng mỗi vườn giữa việc diệt với thuốc trừ sâu và diệt với bọ rùa. Khi n_1, n_2, n_3 và n_4 lớn , ước lượng $\hat{\theta}$ bằng cách thay mỗi $\overline{X}_i$ bằng μ_i là xấp xỉ chuẩn dùng điều này để lấy khoảng tin cậy cho mẫu lớn $100(1-\alpha)\%$ cho θ, và tính khoảng 95% cho dữ liệu đã có.

54. Mặt nạ dùng cho lính cứu hỏa quan trọng là phải chịu được nhiệt cao vì lính cứu hỏa thường làm việc trong môi trường nhiệt độ từ $200-500^oF$. Trong một kiểm tra của một loại mặt nạ có ống kính bật ra 250^o. Xây dựng khoảng tin cậy 90% cho tỉ lệ đúng của loại mặt nạcó ống kính bật ra 250^o

55. Nhà xuất bản sách đại học có hứng thú trong việc ước lượng độ dài của bìa sách sản xuất bởi một máy đóng gáy đặc biệt. độ dài có thể đo bằng cách ghi lại lực cần thiết để kéo trang sách từ bìa sách lực này được đo bằng pound, Nên kiểm tra bao nhiêu sách để ước lượng trung bình lực cần thiết để xé bìa ra khỏi sách trong .1lb với độ tin cậy 95%? Gỉa sử σ đã biết là .8

56. Tiếp xúc lâu dài với tấm lợp amiăng được biết là có nguy cơ đến sức khỏe. Bài viết "Tác động cấp tính của phơi nhiễm amiăng lên chức năng phổi"(Environ. Research, 1978:360-372) báo cáo kết quả nghiên cứu dựa trên một mẫu các công nhân xây dựng những người tiếp xúc với ami ăng trong một thời gian dài. Trong đó dữ liệu trong bài viết có giá trị như sau của độ giãn nở của phổi (cm^3/cmH_2O) cho 16 đối tượng sau 8 tháng tiếp xúc (độ giãn nở của phổi là thước đo độ đàn hồi của phổi, hay phổi sẽ bị ảnh hưởng thế nào khi hít vào và thở ra)

167.9	180.8	184.8	189.8	194.8	200.2
201.9	206.9	207.2	208.4	226.3	227.7
228.5	223.4	239.8	258.6		

a. Phân phối của tổng thể là chuẩn có hợp lý không?
b. Tính khoảng tin cậy 95% cho trung bình đúng độ giãn nở của phổi sau khi tiếp xúc.
c. Tính một khoảng mà, với độ tin cậy 95%, bao gồm ít nhất 95% của độ giãn nở phổi trong phân phối của tổng thể

57. Trong ví dụ 6.8, ta giới thiệu khái niệm về thí nghiệm kiểm duyệt, trong đó n thành phần được thử và thử nghiệm chấm dứt ngay khi có r thành phần lỗi. Gỉa sử dòng đời của các thành phần là độc lập, mỗi cái có một phân phối mũ với tham số λ.Đặt Y_1 là ký hiệu thời gian mà tại đó xảy ra lỗi.thứ 1. Y_2 là thời gian tại đó lỗi thứ 2 xảy ra...., vì thế $T_r = Y_1 + ... + Y_r + (n-r)Y_r$ là tổng thời gian tích lũy khi chấm dứt kiểm tra.Thì có thể biểu diễn là $2\lambda T_R$ có phân phối chi bình phương với 2r df. Sử dụng thử nghiệm này để xây dựng công thức khoảng tin cậy cho $100(1-\alpha)\%$ có trung bình đúng thời gian tích lũy là $1/\lambda$. Tính khoảng tin cậy 95% từ dữ liệu trong ví dụ 6.8

58. Đặt $X_1, X_2, ..., X_n$ là mẫu ngẫu nhiên từ phân phối xác suất liên tục có trung vị $\mu^{\sim}$(vì thế $P(X_i \leq \mu^{\sim}) = P(X_i \geq \mu^{\sim}) = .5$)
a. Biểu diễn

$$P(min(X_i) < \mu^{\sim} < max(X_i)) = 1 - \left(\frac{1}{2}^{n-1}\right)$$

Vì thế $(min(X_i), max(x_i))$ là khoảng tin cậy $100(1-\alpha)\%$ cho $\mu^{\sim}$ với $\alpha = \left(\frac{1}{2}^{n-1}\right)$ [chú ý bổ sung thêm $min(X_i) < \mu^{\sim} < max(X_i)$ là $\{max(X_i) \leq \mu^{\sim}\} \cup \{min(X_i) \geq \mu^{\sim}\}$. Nhưng $max(X_i) \leq \mu^{\sim}$nếu $X_i \leq \mu^{\sim}$ với mọi i

b. Với mỗi 6 trẻ sơ sinh nam bình thường, lượng axit amin alanine $mg/100mL$ được xác định trong chế độ kiêng một loại isoleucine, kết quả cho bởi dữ liệu sau:

2.84 3.54 2.80 1.44 2.94 2.70

Tính khoảng tin cậy 97% cho trung bình đúng của trung vị số ananine trong chế độ kiêng

c. Đặt $x_{(2)}$ là số thứ hai nhỏ nhất của $x_i^{'}s$ và $x_{(n-1)}$ ký hiệu số thứ hai lớn nhất của $x_i^{'}s$. Tìm độ tin cậy cho khoảng $(x_{(2)}), x_{(n-1)}$ cho $\mu^{\sim}$

59. Đặt $X_1, X_2, ..., X_n$ là biến ngẫu nhiên có cùng phân phối trên khoảng $[0, \theta]$, vì thế

$f(x) = \begin{cases} \dfrac{1}{\theta} \; 0 \leq u \leq 1 \\ \; 0 \text{ khác} \end{cases}$

Thì nếu $Y = max(X_i)$, có thể biểu diễn rv $U = Y/\theta$ có hàm mật độ

$f_U(u) = \begin{cases} nu^{n-1} \; 0 \leq u \leq 1 \\ \quad 0 \text{ khác} \end{cases}$

a. Dùng $f_U(u)$ để kiểm chứng

$$P\left((\alpha/2)^{1/n} < \frac{Y}{\theta} \leq (1-\alpha/2)^{1/n}\right) = 1-\alpha$$

vàxác định một khoảng tin cậy $100(1-\alpha)\%$ cho θ

b. Kiểm chứng $P(\alpha^{1/n} \leq Y/\theta \leq 1) = 1-\alpha$, và xác định một khoảng tin cậy $100(1-\alpha)\%$ cho θdựa trên xác suất này.

c. Khoảng nào trong hai khoảng thu được ngắn hơn? Nếu thời gian chờ xe buýt vào mỗi sáng có phân phối đồng nhất và thời gian quan sát là $x_1 = 4.2 \quad x_2 = 3.5 \quad x_3 = 1.7 \quad x_4 = 1.2$ và $x_5 = 2.4$ Xác định khoảng tin cậy 95% cho θ bằng cách sử dụng khoảng ngắn hơn trong hai khoảng đó.

60. Đặt $0 \leq \gamma \leq \alpha$. Thì một khoảng tin cậy $100(1-\alpha)\%$ cho μ khi n lớn là

$$\left(\overline{x} - z_{\gamma}.\frac{s}{\sqrt{n}}, \overline{x} + z_{\alpha-\gamma}.\frac{s}{\sqrt{n}}\right)$$

chọn $\gamma = \alpha/2$ Dùng khoảng được xác định trong phần 7.2 ; Nếu $\gamma \neq \alpha/2$ khoảng này không đối xứng với $\overline{x}$. Độ rộng của khoảng là $w = s(z_\gamma + z_{\alpha-\gamma})/\sqrt{n}$. Cho thấy w là nhỏ nhất cho lựa chọn $\gamma = \alpha/2$, vì thế khoảng đối xứng là ngắn nhất. [chú ý: (a) do định nghĩa của $z_\alpha, \Phi(z_\alpha) = 1-\alpha$; (b) Mối quan hệ giữa đạo hàm của hàm $y = f(x)$ và nghịch đảo của hàm $x = f^{-1}(y)$ là $(d/dy)f^{-1}(y) = 1/f'(x)$.]

61. Gỉa sử $x_1, x_2, ..., x_n$ là kết quả giá trị quan sát một mẫu ngẫu nhiên đối xứng nhưng có thể phân phối đều đến tận cùng.Đặt x và f_s là ký hiệu của trung vị mẫu tương ứng lan ra gấp 4 lần, Chương 11 của Hiểu biết mạnh mẽ và khám phá việc phân tích dữ liệu (Xem phụ lục chương 6) gợi ý khoảng tin cậy mạnh cho 95% cho trung bình tổng thể (điểm đối xứng):

$$x^{\sim} \pm \left(\frac{\text{giá trị điều kiện t}}{1.075}\right).\frac{f_s}{\sqrt{n}}$$

Lượng giá trị trong ngoặc đơn là 2.10 cho n=10, 194 cho n=20, và 1.91 cho n=30. Tính khoảng tin cậy này cho dữ liệu có trong Bài tập 45, và so sánh khoảng tin cậy cho t với phân phối xấp xỉ chuẩn

62. a. Dùng kết quả trong Ví dụ 7.5 để có được chặn dưới của khoảng tin cậy cho tham số λ của một phân phối mũ, và tính ràng buộc dựa trên dữ liệu cho trong ví dụ.

b.Nếu dòng đời của X có phân phối mũ,xác suất vượt quá dòng đời t là $P(X > t) = e^{-\lambda t}$.

Dùng kết quả của phần (a) để đạt được 95% chặn dưới cho xác suất thời gian vượt quá 100 phút.
Phụ lục

6 CHƯƠNG 8 KIỂM ĐỊNH GIẢ THIẾT DỰA TRÊN MỘT MẪU ĐƠN

GIỚI THIỆU

Một tham số có thể ước lượng từ dữ liệu mẫu hoặc bằng một số duy nhất (ước lượng điểm) hay toàn bộ khoảng của giá trị hợp lý (một khoảng tin cậy). Thường xuyên, tuy nhiên, đối tượng của một cuộc điều tra không phải là ước lượng tham số mà là để quyết định trong hai phát biểu mâu thuẫn về tham số là đúng .Điều này là một phần trong suy luận thống kê gọi là kiểm tra giả thiết. Trong chương này, đầu tiên ta sẽ thảo luận một vài khái niệm cơ bản và thuật ngữ trong kiểm định giả thiết sau đó đưa ra quyết định cho các vấn đề kiểm tra thường gặp nhất dựa trên một mẫu từ một mẫu đơn.

7 8.1 Gỉa thiết và quá trình kiểm tra

Một giả thiết thống kê, hay giả thiết, là một yêu cầu hay sự khẳng định về giá trị của một tham số đơn(đặc điểm của tổng thể hay đặc điểm phân phối xác suất), về giá trị của một số tham số, hay về hình dạng của phân phối xác suất. Một ví dụ của giả thiết là phát biểu $\mu = .75$, Khi μ là trung bình đúng đường kính bên trong của một loại ống PVC xác định. Ví dụ khác là phát biểu $p < .10$, khi p là tỉ lệ mạch bị lỗi trong tất cả các mạch được sản xuất từ một nhà máy. Nếu μ_1 và μ_2 ký hiệu cho trung bình đúng sức mạnh phá vỡ nút thắt của hai loại dây khác nhau, Một giả thiết có khẳng định $\mu_1 - \mu_2 = 0$, và một phát biểu khác là $\mu_1 - \mu_2 > 5$. Một ví dụ khác của giả thiết là khẳng định rằng dưới một điều kiện xác định khoảng cách dừng có phân phối chuẩn. Gỉa thiết của loại thứ 2 sẽ được xét trong Chương 14. Trong chương này và một số chương, ta sẽ tập trung trên giả thiết về tham số.
Trong bất kỳ vấn đề nào về kiểm định giả thiết, cũng đều có hai giả thiết mâu thuẫn được xét. Một giả thiết có thể phát biểu $\mu = .75$ và giả thiết khác là $\mu \neq .75$. Hay hai phát biểu mâu thuẫn có thể là $p \geq 10$ và $p < 10$. Đối tượng để quyết định, dựa trên thông tin về mẫu, giả thiết nào sẽ đúng. giống như trong một phiên tòa xét xử. Một tuyên bố khẳng định cá nhân bị buộc tội là vô tội. trong hệ thống tư pháp U.S, khẳng định ban đầu được cho là đúng. chỉ khi đối mặt với những bằng chứng mạnh mẽ ngược lại bồi thẩm đoàn sẽ bác bỏ tuyên bố này để ủng hộ khẳng định khác mà bị cáo phạm tội. Trong trường hợp này tuyên bố về người vô tội hay bảo vệ giả thiết, và gánh nặng là chứng minh phương án thay thế.
Tương tự, trong kiểm định giả thiết thống kê, vấn đề là xây dựng phát biểu ban đầu thỏa mãn. phát biểu ban đầu này sẽ không bị bác bỏ yêu cầu hay thay thế trừ khi cóbằng chứng mẫu cung cấp hỗ trợ mạnh mẽ cho khẳng định khác.
ĐỊNH NGHĨA Gỉa thiết không, ký hiệu H_0 là phát biểu ban đầu được giả định là đúng (phát biểu trước niềm tin). Gỉa thiết thay thế, ký hiệu là H_a là khẳng định mâu thuẫn với H_0.
Gỉa thiết không sẽ được thay thế bởi giả thiết khác nếu bằng chứng mẫu cho thấy H_0 sai.Nếu không có mâu thuẫn mạnh với H_0, ta sẽ tiếp tục tin vào tính hợp lý của giả thiết không. Hai

kết luận có từ phân tích kiểm định giả thiết sau đó từ chối H_0 hay không từ chối H_0

Một kiểm định của giả thiết là phương pháp sử dụng dữ liệu mẫu để quyết định liệu có nên từ chối giả thiết không . Vì thế ta có thể kiểm định $H_0 : \mu = .75$ ngược lại $H_a : \mu \neq .75$. Chỉ nếu dữ liệu mẫu gợi ý mạnh rằng μ khác .75 thì sẽ bác bỏ giả thiết không. Trong trường hợp không có bằng chứng đó Không bác bỏ H_0. khi đó nó vẫn khá hợp lý.
Đôi khi một điều tra viên nếu không muốn chấp nhận một khẳng định đặc biệt và cho đến khi dữ liệu có thể cung cấp một hỗ trợ mạnh mẽ cho khẳng định đó. Như ví dụ, Gỉa sử một công ty xem xét đặt một loại mới lớp phủ lên ổ đỡ mà có thể tạo ra nó. Trung bình đúng của thời gian có độ mòn với lớp phủ hiện tại là 1000 giờ. Với μ là trung bình đúng của chu kỳ cho lớp phủ mới, công ty sẽ không muốn thay đổi nó nếu không có bằng chứng nào đủ mạnh để cho μ vượt quá 1000. một công thức liên quan sẽ phù hợp với kiểm định $H_0 : \mu = 1000$ ngược lại $H_a : \mu > 1000$ Kết luận mang tính hợp lý được xác định cho H_a, và sẽ có bằng chứng để đưa ra kết luận biện minh cho việc từ chối H_0 và chuyển sang lớp phủ mới.
Nhà khoa học nghiên cứu liên quan việc thử một quyết định liệu lý thuyết cũ có nên thay bởi tính hợp lý hơn và giải thích thỏa đáng hiện tượng điều tra được. Cách tiếp cận bảo thủ xác định từ lý thuyết cũ với H_0 và giải thích theo nghĩa nghiên cứu với H_a việc từ chối lý thuyết cũ sẽ xảy ra chỉ khi có bằng chứng phù hợp hơn với thuyết mới. Trong nhiều tình huống, H_a được gọi là "giả thiết của nhà nghiên cứu" khi đó phát biểu của nghiên cứu viên được xác nhận .Từ không có nghĩa là "không bất kỳ giá trị nào, hiệu quả hay hậu quả," gợi ý rằng H_0 nên xác định giả thiết không đổi (từ ý kiến cũ), không khác , không cải tiến, và ....Gỉa sử, cho ví dụ, 10% bảng mạch được sản xuất bởi cùng một nhà máy trong thời gian gần đây là có khiếm khuyết. Một kỹ sư đề nghị thay đổi dây chuyền sản xuất với niềm tin là nó sẽ có kết quả trong việc giảm tỉ lệ khiếm khuyết. Đặt p là ký hiệu tỉ lệ đúng của khiếm khuyết của mạch kết quả từ việc thay đổi qui trình. sau đó nghiên cứu giả thiết, Đó là gánh nặng để chứng minh, là khẳng định rằng $p < .10$, Do đó giả thiết thay thế là $H_a : p < .10$
Trong cách xử lý của chúng tôi về kiểm định giả thiết , H_0 xem như một phát biểu bình đẳng. Nếu θ ký hiệu là tham số quan tâm, giả thiết không sẽ có dạng $H_0 : \theta = \theta_0$, khi θ_0 là số đặc biệt gọi là giá trị không của tham số (giá trị phát biểu cho θ của giả thiết không). Như ví dụ , Xét tình hình bảng mạch đã được thảo luận . Gỉa thiết được đề nghị thay thế là $H_a : p < .10$, Phát biểu về tỉ lệ khiếm khuyết giảm khi sửa đổi quá trình sản xuất. Một lựa chọn tự nhiên của H_0 trong tình huống này là $p \geq .10$ quá trình mới xảy ra một trong hai hoặc tốt hơn hoặc xấu hơn quá trình cũ. Ta xét giả thiết thay thế $H_0 : p = .10$ ngược lại $H_a : p < 10$ tỉ lệ sử dụng giả thiết không đơn giản là bất kỳ quá trình hợp lý nào để đưa ra quyết định giữa $H_0 : p = .10$ và $H_a : p < .10$ cũng là lý do để đưa ra quyết định cho phát biểu $p \geq .10$và H_a. việc dùng một phát biểu đơn giản H_0 là phù hợp vì đó là kỹ thuật chắc chắn được thêm vào, mà rõ ràng là ngắn gọn hơn.
Việc thay thế giả thiết không $H_0 : \theta = \theta_0$ giống như ba khẳng định sau:
1. $H_a : \theta > \theta_0$(trong những trường hợp tiềm ẩn giả thiết không là $\theta \leq \theta_0$)
2. $H_a : \theta < \theta_0$(trong những trường hợp tiềm ẩn giả thiết không là $\theta \geq \theta_0$) hay
3. $H_a : \theta \neq \theta_0$

Cho ví dụ, đặt σ là ký hiệu độ lệch chuẩn của phân phối đường kính bên trong (inches)tay áo bằng kim loại của một loại xác định. nếu quyết định thực hiện để sử dụng ống tay trừ khi mẫu có bằng chứng kết luận $\sigma > .001$ giả thiết thích hợp sẽ là $H_0 : \sigma = .001$ ngược lại $H_a : \sigma > .001$. Số θ_0 xuất hiện trong cả H_0 và H_a (tách biệt sự lựa chọn từ không) được gọi là giá trị không.
QUÁ TRÌNH KIỂM ĐỊNH
Qúa trình kiểm định là một qui tắc, dựa trên dữ liệu mẫu, cho quyết định có nên bác bỏ H_0. Kiểm định $H_0 : p = .10$ ngược lại $H_a : p < .10$ vấn đề mạch có thể dựa trên kiểm định mẫu ngẫu nhiên của $n = 200$ bảng mạch. Đặt X ký hiệu số bảng mạch bị khiếm khuyết trong mẫu, một biến ngẫu nhiên nhị thức ,x đại diện cho giá trị quan sát của X. Nếu H_0 đúng,

$E(X) = np = 200.(.10) = 20$ trong khi ta mong muốn có ít hơn 20 bảng mạch khiếm khuyết nếu H_a là đúng . Một giá trị x chỉ cần ít hơn 20 không mâu thuẫn mạnh với H_0 đó là lý do bác bỏ H_0 nếu $x \leq 15$ và không bác bỏ H_0 nếu khác đi. Qúa trình này có hai thành phần: (1) một kiểm định thống kê, hay hàm của dữ liệu mẫu dùng để đưa ra quyết định, và (2) vùng bác bỏ bao gồm cả giá trị x cho H_0 sẽ bị bác bỏ H_a. cho qui tắc vừa thảo luận, giá trị bị bác bỏ bao gồm $x = 0, 1, 2, ..., và 15$. H_0 sẽ không bị bác bỏ nếu $x = 16, 17, ...1999, hay 200$.

Một quá trình kiểm định bao gồm những qui định sau:
1. Một kiểm định thống kê , hay hàm của dữ liệu mẫu dùng để đưa ra quyết định (bác bỏ H_0 hay chấp nhận H_0) 2. vùng bác bỏ, một tập hợp tất cả giá trị kiểm định thống kê cho H_0 sẽ bị bác bỏ.

Gỉa thiết không sẽ bị bác bỏ nếu và chỉ nếu quan sát hoặc tính toán được giá trị kiểm định thống kê rơi vào vùng bác bỏ.

Như những ví dụ khác, giả sử một nhà máy sản suất thuốc lá phát biểu rằng lượng nicotin trung bình trong một loại thuốc lá (B) là (ít nhất) 1.5mg. Sẽ không hay nếu bác bỏ phát biểu của nhà máy mà không có bằng chứng mạnh mẽ để chống, vì thế công thức để kiểm tra là $H_0 : \mu = 1.5$ ngược lại $H_a : \mu > 1.5$. Xét một quyết định có qui luật dựa trên việc phân tích một mẫu ngẫu nhiên gồm 32 điếu thuốc, Đặt $\overline{X}$ ký hiệu là trung bình mẫu của lượng nicotine. Nếu H_0 đúng , $E(\overline{X}) = \mu = 1.5$, Nếu H_0 sai ta mong muốn $\overline{X}$ vượt quá 1.5.Với bằng chứng mạnh để H_0 mà được cung cấp bởi giá trị $\overline{x}$ mà vượt quá 1.5. Vì thế ta có thể sử dụng $\overline{X}$ như là một kiểm định thống kê với vùng bác bỏ $\overline{x} \geq 1.60$

Cơ sở để lựa chọn vùng bác bỏ trong việc xét những sai lầm xảy ra mà ta có thể phải đối mặt với việc rút ra kết luận.Xét vùng bác bỏ $x \leq 15$ trong vấn đề bảng mạch, thậm chí khi $H_0 : p = .10$ là đúng, nó có thể là vấn đề bất thường với kết quả mẫu trong $x = 13$, vì thế H_0 sai lầm bị bác bỏ. Một cách khác,thậm chí khi $H_a : p < .10$ là đúng , một mẫu bất thường có thể tìm ra $x = 20$. Trong trường hợp H_0 không bị bác bỏ- sẽ cho ra một quyết định sai lầm. Vì thế có thể nói rằngH_0 có thể bị bác bỏ khi nó đúng hoặc H_0 có thể không bị bác bỏ khi nó sai. Những sai lầm này không phải là kết quả của việc lựa chọn sai vùng bác bỏ, hoặc những sai lầm có kết quả khi vùng $x \leq 14$ được dùng, hay khi bất kỳ vùng hợp lý nào được sử dụng.

ĐỊNH NGHĨA
Sai lầm loại I bao gồm bác bỏ giả thiết không H_0 khi nó đúng. Sai lầm loại II liên quan đến việc chấp nhận H_0 khi H_0 sai.

Trong ví dụ về nicotin, sai lầm loại I bao gồm bác bỏ phát biểu của nhà máy mà $\mu = 1.5$ khi thực tế không đúng như vậy. Nếu vùng bác bỏ $\overline{x} \geq 1.6$ được sử dụng, nó có thể xảy ra $\overline{x} = 1.6$ thậm chí khi $\mu = 1.5$, kết quả trong sai lầm loại I.
Nói cách khác, H_0 có thể sai và chưa quan sát $\overline{x} = 1.52$, dẫn đến không bác bỏ H_0 (sai lầm loại II) Cách tốt nhất có thể, là làm sao để quá trình kiểm định không xảy ra sai lầm. Tuy nhiên, ý tưởng này chỉ đạt được khi quyết định dựa trên kiểm tra đầu vào của tổng thể. Khó khăn với việc sử dụng một quá trình dựa trên dữ liệu mẫu mà còn vì biến của mẫu.Một mẫu không đại diện có thể là kết quả. ví dụ một giá trị của $\overline{X}$ xa μ hay giá trị của $\hat{p}$ khác giá trị ban đầu p. Thay vì đòi hỏi một quá trình không xảy ra sai lầm, ta phải tìm ra một quá trình sao cho cả hai loại lỗi đều không thể xảy ra. Nghĩa là, Một quá trình tốt tức là xác suất của cả hai sai lầm nho.Lựa chọn một vùng bác bỏ có giá trị xác suất phù hợp cho sai lầm loại I và loại II. Những sai lầm trong xác suất được ký hiệu tương ứng bởi α và β. Vì H_0 có giá trị duy nhất là tham số, đó là giá trị đơn α. Tuy nhiên, cũng có một giá trị khác của β cho mỗi giá trị của tham số

phù hợp với H_a

VÍ DỤ 8.1

Một loại ô tô có thiệt hại không nhìn thấy là 25%trong thời gian 10 dặm trên giờ khi kiểm tra va chạm. Một thiết kế đang được nỗ lực đề xuất sửa đổi để tăng phần trăm này. Đặt p ký hiệu tỉ lệ của tất cả va chạm 10 mph với bộ giảm va mới cho kết quả không gây thiệt hại. Gỉa thiết để kiểm tra là $H_0 : p = .25$(không cải tiến) ngược lại $H_a : p > .25$. Kiểm tra này dựa trên kinh nghiệm với $n = 20$những va chạm độc lập với nguyên mẫu của thiết kế mới. Bằng trực giác, H_0 bị bác bỏ nếu số lượng đáng kể va chạm không gây thiệt hại. Xét quá trình kiểm định sau:
Kiểm định thống kê: $X =$ Số va chạm với thiệt hại không nhìn thấy.
Vùng bác bỏ: $R_8 = \{8, 9, 10, ..., 19, 20\}$; nghĩa là, bác bỏ H_0 nếu $x \geq 8$ khi x là giá trị quan sát của kiểm định thống kê.
Vùng bác bỏ gọi là kiểm định bên phải vì nó chỉ bao gồm những giá trị lớn của kiểm định thống kê
Khi H_0 đúng, X có phân phối nhị thức với $n = 20$ và $p = .25$ thì

$$\alpha = P(\text{sai lầm loại I}) = P(\text{bác bỏ } H_0 \text{khi} H_0 \text{đúng})$$
$$= P(X \geq 8 \text{khi} X \sim Bin(20, .25)) = 1 - B(7; 20, .25)$$
$$= 1 - .898 = .102$$

Nghĩa là,khi H_0 thực sự đúng, gần 10 % của tất cả va chạm xảy ra trong đó có 20 va chạm sẽ cho kết quả trong H_0 là bác bỏ không đúng (sai lầm loại I)
Ngược lại với α không có số đơn β khác cho mỗi p khác mà gần .25. Vì thế ta có giá trị của β khi cho $p = .3$ (trong trường hợp $X \sim Bin(20, .3)$), Một giá trị khác của β khi cho $p = .5$ ...v.v...cho Ví dụ,
$\beta(.3) = P$(sai lầm loại II khi p=.3)

$$= P(\text{không bác bỏ} H_0 \text{khi nó sai vì} p = .3)$$

$$= P(X \leq 7 \text{khi} X \sim Bin(20, .3)) = B(7; 20, .3) = .772$$

Khi p có giá trị thực là .3 hơn .25 (nhỏ bé được nhận từ H_0), gần 77% tất cả những loại có kết quả không đúng để bác bỏ H_0
Bảng đính kèm hiển thị giá trị β cho việc lựa chon p (Mỗi phép tính cho ra một vùng bác bỏ R_8). Rõ ràng, β giảm khi giá trị p di chuyển về phía bên phải của giá trị không .25. Bằng trực giác ta thấy, khởi đầu H_0 càng lớn, có ít khả năng phát hiện ra.

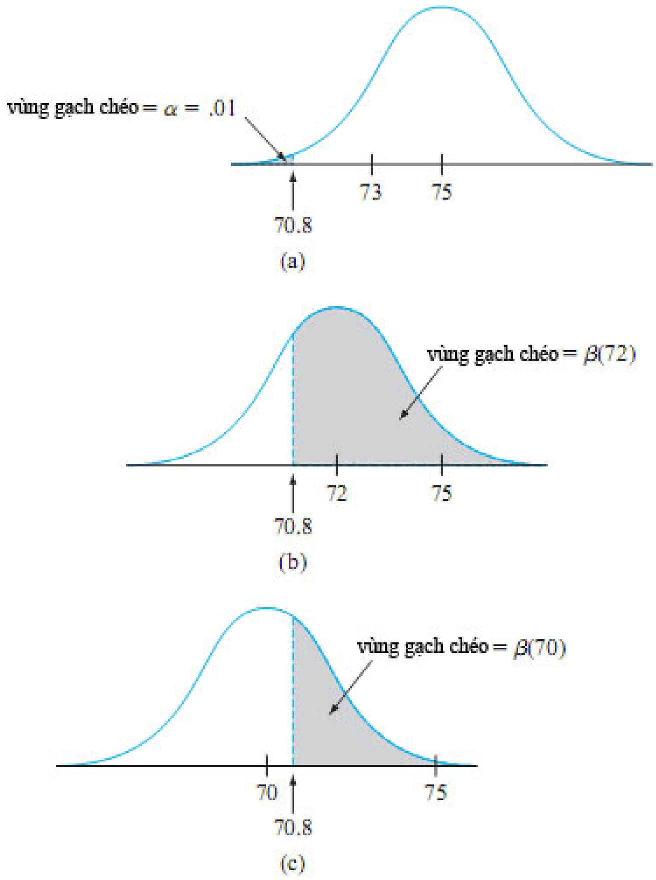
p	.3	.4	.5	.6	.7	.8
$\beta(p)$	.772	.416	.132	.021	.001	.000

Đề xuất quá trình kiểm định là có lý do cho kiểm định giả thiết không trong thực tế mà $p \leq .25$. Trong trường hợp đó, không có giá trị nào dài hơn một giá trị đơn α thay vào đó có một α cho mỗi p là tối đa .25: $\alpha(.25), \alpha(.23), \alpha(.20), \alpha(.15)$...v..v..Dễ dàng chứng minh được rằng $\alpha(p) < \alpha(.25) = .102 \text{nếu} p < .25$. Nghĩa là, giá trị lớn nhất của α xảy ra cho ranh giới giá trị .25 giữa $H_0 \text{và} H_a$. Vì thế nếu α nhỏ đơn giản hóa giả thiết không, nó cũng sẽ nhỏ như vậy hay cho H_0 thực tế nhỏ hơn.

VÍ DỤ 8.2

Thời gian sấy của một loại sơn nhất định được kiểm tra dưới một điều kiện đặc biệt được biết có phân phối chuẩn với giá trị trung bình là 75 phút và độ lệch tiêu chuẩn là 9 phút. Các nhà

hóa học đã đề xuất thêm vào một chất phụ gia mới để giảm thời gian sấy trung bình. Họ tin rằng thời gian sấy khi thêm chất phụ gia cũng sẽ có phân phối chuẩn với $\sigma = 9$. Vì chi phí liên quan đến chất phụ gia thêm vào, bằng chứng là nên đề nghị cải tiến trong thời gian trung bình sấy khô trước khi có kết luận. Đặt μ ký hiệu là trung bình đúng của thời gian sấy khi sử dụng chất phụ gia. Gỉa thiết phù hợp là $H_0 : \mu = 75$ ngược lại $H_a : \mu < 75$. Nếu H_0 bị bác bỏ thì tuyên bố về việc thêm vào chất phụ gia là thành công và được đưa vào sử dụng.
Dữ liệu thử nghiệm bao gồm thời gian sấy của n=25 mẫu kiểm tra. Đặt $X_1, ..., X_{25}$ ký hiệu cho 25 thời gian sấy - một mẫu ngẫu nhiên có kích thước 25 từ phân phối chuẩn với giá trị kỳ vọng μ và độ lệch chuẩn $\sigma = 9$. Trung bình mẫu về thời gian sấy $\overline{X}$ cũng có phân phối chuẩn với giá trị kỳ vọng $\mu_{\overline{X}} = \mu$ và độ lệch chuẩn $\sigma_{\overline{X}} = \sigma/\sqrt{n} = 9/\sqrt{25} = 1.80$. Khi H_0 đúng, $\mu_{\overline{X}} = 75$, vì thế chỉ có một số giá trị $\overline{X}$ ít hơn 75 sẽ mâu thuẫn với H_0. Nguyên nhân vùng bác bỏ có dạng $\overline{x} \leq c$, khi giá trị cắt c được chọn là phù hợp. Xét lựa chọn $c = 70.8$, vì thế quá trình kiểm tra bao gồm kiểm định thống kê $\overline{X}$ và vùng bác bỏ $\overline{x} \leq 70.8$. Vì vùng bác bỏ chỉ bao gồm những giá trị nhỏ của kiểm định thống kê, quá trình kiểm tra này gọi là kiểm tra bên trái .Tính α và β bao gồm cả giá trị chuẩn hóa $\overline{X}$sau bằng cách tham khảo xác suất chuẩn của bảng phụ lục A.3
$\alpha = P(\text{sai lầm loại I}) = P(\text{bác bỏ}H_0\text{khi nó đúng})$
$= P(\overline{X} \leq 70.8\text{khi}\overline{X} \sim \text{chuẩn với}\mu_{\overline{X}} = 75, \sigma_{\overline{X}} = 1.8)$
$= \Phi\left(\frac{70.8-75}{1.8}\right) = \Phi(-2.33) = .01$
$\beta(72) = P(\text{sai lầm loại II khi}\mu = 72)$
$= P(H_0\text{chấp nhận}H_0\text{khi nó sai vì}\mu = 72)$
$= P(\overline{X} > 70.8\text{khi}\overline{X} \sim \text{chuẩn với}\mu_{\overline{X}} = 72\text{và}\sigma_{\overline{X}} = 1.8)$
$= 1 - \Phi\left(\frac{70.8-72}{1.8}\right) = 1 - \Phi(-.67) = 1 - .2514 = .7486$
$\beta(70) = 1 - \Phi\left(\frac{70.8-70}{1.8}\right) = .3300 \qquad \beta(67) = .0174$ Một quá trình kiểm định đặc biệt, chỉ 1%của tất cả thử nghiệm thực hiện như mô tả với kết quả H_0 bị bác bỏ khi nó thực sự đúng. Tuy nhiên, thay đổi của sai lầm loại II là rất lớn khi $\mu = 72$ (một phần nhỏ của H_0) đôi khi ít hơn khi $\mu = 70$, và khá nhỏ khi $\mu = 67$ (mất mát đáng kể từ H_0). Sai lầm xác suất được minh họa trong hình 8.1. Chú ý rằng α được tính bằng cách sử dụng phân phối xác suất của kiểm định thống kê khi H_0 đúng, trong khi xác định β cần phải biết xác suất của kiểm định thống kê khi H_0 sai.



Hình 8.1 α và β minh họa cho ví dụ 8.2 (a)phân phối của $\overline{X}$khi$\mu = 75(H_0)$đúng); (b)phân phối của $\overline{X}$khi$\mu = 72(H_0)$sai)
Như trong ví dụ 8.1, nếu xét càng nhiều giả thiết không trong thực tế $\mu \geq 75$ thì có một α có một giá trị tham số của H_0 là đúng: $\alpha(75), \alpha(75.8), \alpha(76.5), ...$ dễ dàng kiểm chứng rằng $\alpha(75)$ có xác suất lớn nhất trong tất cả sai lầm loại I.tập trung vào giá trị biên để xác định rõ "trường hợp sai"
Đặc điểm của giá trị cắt cho vùng bác bỏ trong ví dụ vừa xét nhận một số giá trị tùy ý. Sử dụng $R_8 = \{8, 9, ...20\}$ trong Ví dụ 8.1 cho $\alpha = .102, \beta(.3) = .772$, và $\beta(.5) = .132$. Nhiều người sẽ nghĩ rằng các lỗi xác suất là không thể chấp nhận được. chúng có thể giảm đi bằng cách thay đổi giá trị điểm cắt.
VÍ DỤ 8.3

Ta hãy sử dụng cùng một thử nghiệm và kiểm định thống kê Xđược mô tả trước đó trong vấn đề về va chạm ô tô nhưng bây giờ xét $R_9 = \{9, 10, ..., 20\}$. Khi X vẫn có phân phối chuẩn với tham số $n = 20$ và p.
$\alpha = P($bác bỏH_0khi$p = .25)$
$= P(X \geq 9$khi$X \sim Bin(20, .25)) = 1 - B(8; 20, .25) = .041$
Sai lầm loại I trong xác suất giảm do sử dụng vùng bác bỏ mới, Tuy nhiên, cái giá phải trả cho sai lầm đó là:
$\beta(.3) = P($chấp nhậnH_0khi$p = .3)$
$= P(X \leq 8$khi$X \sim Bin(20, .3)) = B(8; 20, .3) = .887$
$\beta(.5) = B(8; 20, .5) = .252$
Cả hai β đều lớn hơn lỗi xác suất tương ứng là .772 và .132 cho vùng của R_8. nhìn lại thì cũng không có gì gây ngạc nhiên; α được tính bởi tổng hợp những xác suất trên của giá trị kiểm định thống kê trong vùng bác bỏ. trong khi β là xác suất mà X sai trong phần bổ sung của vùng bác bỏ. Làm cho vùng bác bỏ nhỏ hơn do đó α phải giảm trong khi β tăng cho bất kỳ

$p > .25$

VÍ DỤ 8.4 (Tiếp ví dụ 8.2)
Sử dụng giá trị cắt $c = 70.8$ trong ví dụ thời gian làm sơn khô trong một giá trị rất nhỏ của $\alpha(.01)$ nhưng lại lớn hơn β. Xét cùng một thử nghiệm và kiểm định thống kê $\overline{X}$ với vùng bác bỏ mới $\overline{x} \geq 72$. Vì $\overline{X}$ vẫn có phân phối chuẩn với giá trị kỳ vọng $\mu_{\overline{x}} = \mu$ và $\sigma_{\overline{x}} = 1.8$,

$$\begin{aligned}\alpha &= P(\text{bác bỏ} H_0 \text{khi nó đúng})\\ &= P(\overline{X} \leq 72 \text{khi} \overline{X} \sim N(75, 1.8^2))\\ &= \Phi\left(\frac{72-75}{1.8}\right) = \Phi(-1.67) = .0475 \approx .05\end{aligned}$$

$\beta(72) = P(\text{chấp nhận} H_0 \text{khi} \mu = 72)$
$= P(\overline{X} > 72$ Khi $\overline{X}$chuẩn rv với kỳ vọng 72 và độ lệch chuẩn 1.8)
$= 1 - \Phi\left(\frac{72-72}{1.8}\right) = 1 - \Phi(0) = .5$
$\beta(70) = 1 - \Phi\left(\frac{72-70}{1.8}\right) = .1335 \qquad \beta(67) = .0027$
Những thay đổi trong giá trị cắt làm cho vùng bác bỏ lớn hơn (bao gồm nhiều giá trị $\overline{x}$ hơn), kết quả việc giảm bớt β cho mỗi μ cố định ít hơn 75. Tuy nhiên, α cho vùng mới này tăng từ những giá trị trước đó .01 đến xấp xỉ .05. Nếu sai lầm xác suất loại I nảy lớn có thể chấp nhận, qua đó, vùng thứ hai ($c = 72$) phù hợp hơn vùng đầu tiên ($c = 70.8$) vì đó là nhỏ nhất của β
Kết quả của những ví dụ này có thể tổng quát lại như sau
MỆNH ĐỀ

Xét một thử nghiệm và một cỡ mẫu cố định và kiểm định thống kê được chọn. thì việc giảm kích thước vùng bác bỏ chứa những giá trị nhỏ hơn của α trong giá trị lớn hơn β cho bất kỳ giá trị tham số cụ thể nào phù hợp với H_a

Mệnh đề này nói rằng một kiểm định thống kê và n là cố định, không có vùng kiểm định nào mà đồng thời làm cho cả α và β của nó nhỏ Một vùng được chọn sao cho có sự thỏa mãn giữa α và β
Vì các hướng dẫn đề xuất cho H_0 và H_a, sai lầm loại I thường nghiêm trọng hơn sai lầm loại II (luôn luôn có thể đạt được do việc chọn giả thiết phù hợp) Phương pháp này được tuân thủ bởi hầu hết các bài tập thống kê sau này đều chỉ định giá trị α lớn nhất mà có thể chấp nhận và tìm một vùng bác bỏ để có được giá trị α thay vì bất cứ cái gì nhỏ hơn. Điều này làm cho β càng nhỏ càng tốt tùy thuộc ràng buộc trên α. Kết quả là giá trị của α thường gọi là mức ý nghĩa của kiểm định. Mức truyền thống của ý nghĩa là .10, .05, và .01, thông qua mức này bất kỳ vấn đề cụ thể nào cũng sẽ phụ thuộc tính nghiêm trọng của sai lầm loại I- là sai lầm nghiêm trọng nhất, cái nhỏ hơn được gọi là mức ý nghĩa .
Một quá trình kiểm định tương ứng gọi là mức kiểm định α (ví dụ, một mức kiểm định .05 hay một mức kiểm định .01). Một kiểm định với mức quan trọng α là một trong những sai lầm xác suất loại I bị kiểm soát tại mức quan trọng.
VÍ DỤ 8.5
Tính lại đặt μ ký hiệu cho trung bình đúng lượng nicotin của loại thuốc lá B . Đối tượng kiểm định là $H_0 : \mu = 1.5$ ngược lại $H_a : \mu > 1.5$ dựa vào mẫu ngẫu nhiên $X_1, X_2, ..., X_{32}$ của lượng nicotin. Gỉa sử phân phối của lượng nicotin đã biết là phân phối chuẩn với $\sigma = .20$. Thì $\overline{X}$ có phân phối chuẩn với giá trị kỳ vọng $\mu_{\overline{X}} = \mu$ và độ lệch chuẩn $\sigma_{\overline{X}} = .20\sqrt{32} = .0354$
Thay vì dùng chính $\overline{X}$ để kiểm định, ta đặt biến chuẩn hóa $\overline{X}$,Gỉa sử H_0 đúng
Kiểm định thống kê: $Z = \frac{\overline{X}-1.5}{\sigma\sqrt{n}} = \frac{\overline{X}-1.5}{.0354}$
Z biểu diễn khoảng cách giữa $\overline{X}$ và giá trị kỳ vọng của chính nó khi H_0 đúng như một số của

độ lệch chuẩn. Cho ví dụ, $z = 3$ có kết quả từ một $\overline{x}$ đó là 3 độ lệch chuẩn lớn hơn ta sẽ cóđược kỳ vọng và được H_0 đúng.
Bác bỏ H_0 khi $\overline{x}$ "đáng kể" vượt quá 1.5 thì tương đương với bác bỏ H_0 khi z $\O ngk$ vượt quá 0. Nghĩa là, vùng bác bỏ có dạng là $z \geq c$
Bây giờ xác định c sao cho $\alpha = .05$. Khi H_0 đúng, Z có phân phối chuẩn. Vì vậy

$$\alpha = P(\text{sai lầm loại I }) = P(\text{bác bỏ} H_0 \text{ khi } H_0 \text{đúng})$$
$$= P(Z \geq c \text{ khi } Z \sim N(0,1))$$

Gía trị c phải hiển thị vùng phía bên phải .05 dưới đường cong Z. Xem trong phần 4.3 hay trực tiếp từ phụ lục A.3 $c = z_{.05} = 1.645$
Chú ý rằng $z \geq 1.645$ tương đương với $\overline{x} - 1.5 \geq (.0354)(1.645)$, tức là $\overline{x} \geq 1.56$. Thì βsẽ liên quan đến xác suất mà $\overline{X} < 1.56$ và có thể tính cho μ bất kỳ lớn hơn 1.5

BÀI TẬP Phần 8.1 $(1 - 14)$

1. Với mỗi khẳng định sau, cho biết giả thiết thống kê nào là hợp lý và tại sao:
a $H : \sigma > 100$
b. $H : x^{\sim} = 400$
c. $H : s < .20$
d. $H : \sigma_1/\sigma_2 < 1$
e. $H : \overline{X} - \overline{Y} = 5$
f. $H : \lambda \leq .01$, khi λ là tham số của phân phối mũ dùng trong mô hình thành phần sống.

2. Cho những khẳng định kép sau, chỉ ra rằng sự không tuân thủ qui tắc để thiết lập giả thiết và tại sao (chỉ số dưới 1 và 2 khác nhau giữa số lượng cho hai tổng thể khác hay mẫu khác).
a. $H_0 : \mu = 100, H_a : \mu > 100$
b. $H_0 : \sigma = 20, H_a : \sigma \leq 20$
c. $H_0 : p \neq .25, H_a : p = .25$
d. $H_0 : \mu_1 - \mu_2 = 25, H_a : \mu_1 - \mu_2 > 100$
e. $H_0 : S_1^2 = S_2^2, H_a : S_1^2 \neq S_2^2$
f. $H_0 : \mu = 120, H_a : \mu = 150$
g. $H_0 : \sigma_1/\sigma_2 = 1, H_a : \sigma_1/\sigma_2 \neq 1$
h. $H_0 : p_1 - p_2 = -.1, H_a : p_1 - p_2 < -.1$

3. Để xác định rằng liệu các mối hàn của đường ống trong nhà máy điện hạt nhân có đáp ứng yêu cầu về kỹ thuật, một mẫu ngẫu nhiên các mối hàn được chọn và tiến hành kiểm tra mỗi mối hàn trong cùng một mẫu. Gỉa sử vùng đặc biệt mà trung bình chiều dài của mối hàn vượt quá $100 lb/in^2$; Đoàn thanh tra quyết định kiểm định $H_0 : \mu = 100$ ngược lại $H_a : mu > 100$. giải thích tại sao khi dùng H_a thích hợp hơn $\mu < 100$

4. đặt μ ký hiệu là trung bình đúng mức phóng xạ (picocuries trên lit) . Gía trị 5 pCi/L được xét dòng phân chia giữa nước an toàn và nước không an toàn. Bạn muốn gợi ý kiểm định nào sau đây $H_0 : \mu = 5$ ngược lại $H_a : \mu > 5$ hay $H_0 : \mu = 5$ ngược lại $H_a : \mu < 5$? giải thích cho lý do đó. [tức : Nghĩ về kết quả của sai lầm loại I và sai lầm loại II cho mỗi khả năng]

5. Trước khi đồng ý mua một lượng lớn vỏ polyethylen của một loại dầu cao áp đặc biệt trong tàu ngầm. Một công ty muốn xem bằng chứng để kết luận rằng độ lệch tiêu chuẩn đúng về độ mỏng của vỏ ít nhất là .05 mm. Nên kiểm định giả thiết nào và tại sao ? trong bối cảnh này sai lầm loại I hay loại II sẽ xảy ra?

6. Nhiều căn nhà cũ có hệ thống điện sử dụng cầu chì hơn là bộ ngắt mạch. Một nhà máy sản xuất cầu chì 40 amp muốn chắc rằng trung bình cường độ dòng điện tại đó cầu chì bị cháy trong thực tế là 40. Nếu trung bình cường độ dòng điện thấp hơn 40, khách hàng sẽ phàn nàn vì phải thay cầu chì thường xuyên. Nếu trung bình cường độ dòng điện cao hơn 40, nhà sản xuất có thể phải chịu thiệt hại về hệ thống dòng điện vì cầu chì hư hỏng. để xác monh cường độ dòng điện của cầu chì một mẫu các cầu chì được chọn và kiểm tra Nếu kiểm định giả thiết là phù hợp với kết quả dữ liệu giả thiết không và giả thiết thay thế nào sẽ được đưa vào nhà máy? mô tả sai lầm loại I và sai lầm loại II trong bối cảnh tình hình này.

7. Mẫu nước được lấy từ nước dùng làm mát khi nó đang được thải ra từ một nhà máy điện vào sông: Nó được xác định rằng miễn là nhiệt độ trung bình của nước xả tối đa là 150^oF, sẽ không có tiêu cực nào ảnh hưởng đến hệ sinh thái của sông, việc điều tra liệu nhà máy có tuân thủ các qui định cấm hay không khi nhiệt độ trung bình trên nước xả trên 150^oF, 50 mẫu nước được lấy ngẫu nhiên và ghi lại nhiệt độ của mỗi mẫu nước.Kết quả dữ liệu sẽ được dùng để kiểm định giả thiết $H_0 : \mu = 150^o$ ngược lại $H_a : \mu > 150^o$. Trong bối cảnh của tình huống này mô tả sai lầm loại I và sai lầm loại II. Sai lầm loại nào nghiêm trọng hơn? Giải thích?

8. Một loại laminate hiện đang được sử dụng thường xuyên bởi một nhà máy để sản xuất bảng mạch. Một loại laminate đặc biệt được phát triển để giảm sự cong vênh cho mạch và loại laminate đặc biệt đó có trên một mẫu khác, số lượng bảng mạch bị cong vênh sẽ được xác định cho mỗi mẫu. Nhà sản xuất sẽ chuyển sang sản xuất laminate đặc biệt đó nếu nó có thể chứng minh rằng trung bình đúng số lượng bảng mạch bị cong vênh cho laminate đặc biệt ít hơn laminate thường. Tìm vùng kiểm định thích hợp và mô tả sai lầm loại I và sai lầm loại II trong bối cảnh của tình huống này.

9. Có hai công ty khác nhau cung cấp dịch vụ cab tivi trong một khu vực nhất định. đặt p ký hiệu tỉ lệ khách tiềm năng đang sử dụng dịch vụ của công ty I chuyển sang công ty thứ hai. Xét kiểm định $H_0 : p = .5$ ngược lại $H_a : p \neq .5$, dựa trên mẫu ngẫu nhiên gồm 25 cá nhân. đặt X ký hiệu số người trong mẫu yêu thích công ty thứ I và x đại diện cho giá trị quan sát của X.
a. Vùng bác bỏ nào sau đây phù hợp nhất? Tại sao?
$R_1 = \{x : x \leq 7 \text{ hay } x \geq 18\}$, $R_2 = \{x : x \leq 8\}$, $R_3 = \{x : x \geq 17\}$
b. Trong bối cảnh tình huống này mô tả sai lầm loại I và sai lầm loại II
c. Phân phối xác suất của kiểm định thống kê X là gì khi H_0 đúng ? Dùng nó để tính xác suất sai lầm loại I.
d. Tính xác suất của sai lầm loại II cho vùng đã chọn khi p=.3 ngược lại khi p=.4 và cho cả p=.6 và p=.7
e. Dùng vùng đã chọn, thu được kết luận gì nếu 6 cá nhân trong 25 người chọn công ty thứ I.

10. Một hỗn hợp nhiên liệu bột nghiền và ximăng Porland để tráng vữa có cường độ nén lớn hơn 1300 KN/m^2 Hỗn hợp sẽ không được sử dụng trừ khi có bằng chứng thực nghiệm cho thấy rằng cường độ nén được đáp ứng. Gỉa sử cường độ nén cho mẫu của hỗn hợp là phân phối chuẩn với $\sigma = 60$. đặt μ ký hiệu cho trung bình đúng của cường độ nén
a. Gỉa thiết không hay giả thiết thay thế là thích hợp?
b. đặt $\overline{X}$ ký hiệu là trung bình mẫu của cường độ nén. Xét quá trình kiểm định với kiểm định thống kê $\overline{X}$ và vùng bác bỏ $\overline{x} \geq 1331.26$, Phân phối xác suất của kiểm định thống kê bằng bao nhiêu khi H_0 đúng? xác suất cho quá trình kiểm định đó bằng bao nhiêu ?
c. Phân phối xác suất của kiểm định thống kê bằng bao nhiêu khi $\mu = 1350$? Sử dụng quá trình kiểm định trong câu (b) để kiểm định cho mức ý nghĩa .05? Xác suất của hỗn hợp đó bằng bao

nhiêu thì hỗn hợp đó bị đánh giá không đạt yêu cầu khi thực tế $\mu = 10.1$? khi $\mu = 9.8$?
d.Thay đổi quá trình kiểm định trong câu (b) như thế nào để kiểm định mức ý nghĩa .05? Tác động nào sẽ làm thay đổi sai lầm xác suất trong câu (c)? e. Xét kiểm định thống kê với biến chuẩn hoá $Z = (\overline{X} - 1300)/(\sigma/\sqrt{n}) = (\overline{X} - 1300)/13.42$. Gía trị của Z bằng bao nhiêu để thoả vùng bác bỏ trong câu (b)

11. Hiệu chuẩn của một tỉ lệ bằng cách cân 10kg mẫu kiểm tra trong 25 lần. giả sử rằng kết quả cân những lần khác nhau là độc lập với mẫu khác và cân trên mỗi thử nghiệm có phân phối chuẩn với $\sigma = .200kg$. đặt μ ký hiệu trung bình đúng cân đọc được trên tỉ lệ.
a. Nên kiểm định giả thiết như thế nào?
b. Gỉa sử tỉ lệ đã được hiệu chỉnh lại nếu như $\overline{x} \geq 10.1032$ hoặc $\overline{x} \leq 9.8968$. Xác suất khi thực hiện hiệu chỉnh bằng bao nhiêu ? ,khi nào nó thực sự cần thiết?
c. Xác suất nào của việc hiệu chỉnh được đánh giá không cần thiết khi thực tế $\mu = 10.1$? Khi $\mu = 9.8$
d. đặt $z = (\overline{x} - 10)/(\sigma/\sqrt{n})$. giá trị nào của c là vùng bác bỏ trong câu (b) tương đương với vùng của kiểm định hai bên của $z \geq c$hay $x \leq -c$?
e. Nếu cỡ mẫu chỉ 10 đến 25 thì quá trình kiểm định trong câu (d)sẽ thay đổi thế nào để $\alpha = .05$
f. Dùng kiểm định trong câu (e). kết luận thế nào từ dữ liệu mẫu sau.

9.981	10.006	9.857	10.107	9.888
9.728	10.439	10.214	10.190	9.793

g. Diễn tả lại quá trình kiểm định của câu (b) trong qui định về kiểm định thống kê cho biến chuẩn hoá $Z = (\overline{X} - 10)/(\sigma/\sqrt{n})$

12. Một thiết kế mới cho hệ thống phanh của một loại ô tô nhất định, so với hệ thống cũ trung bình đúng của khoảng cách phanh tại 40 mph, Dưới điều kiện đặc biệt đã biết là 120 ft. Một đề xuất cho thiết kế mới được thực hiện phải chỉ ra rằng từ mẫu dữ liệu cho thấy việc giảm trung bình khoảng cách phanh cho thiết kế mới
a. Xác định tham số quan tâm và vùng kiểm định liên quan
b. Gỉa sử khoảng cách phanh cho hệ thống mới có phân phối chuẩn với $\sigma = 10$. Đặt $\overline{X}$ ký hiệu cho trung bình mẫu của khoảng cách phanh cho mẫu ngẫu nhiên của 36 quan sát. Vùng bác bỏ nào trong ba vùng sau phù hợp
$R_1 = \{\overline{x} : \overline{x} \geq 124.80\{, R_2 = \{\overline{x} : \overline{x} \leq 115.20\}, R_3 = \{\overline{x} : \text{hoặc } \overline{x} \geq 125.13\text{hoặc } \overline{x} \leq 114.87\}$?
c. Mức ý nghĩa nào thích hợp cho vùng bác bỏ của câu (b) vùng đó sẽ thay đổi thế nào khi kiểm định với $\alpha = .001$
d. Xác suất nào cho thiết kế mới không thể thực hiện khi trung bình đúng khoảng cách phanh thực tế là 115 ft và vùng phù hợp dùng trong câu (b)
e. đặt $Z = (\overline{X} - 120)/(\sigma/\sqrt{n})$ mức ý nghĩa nào cho vùng bác bỏ $\{z : z \leq -2.33\}$? cho vùng $\{z : z \leq -2.88\}$?

13. đặt $X_1, ..., X_n$ ký hiệu mẫu ngẫu nhiên từ phân phối chuẩn với σ đã biết
a. Cho kiểm định giả thiết $H_0 : \mu = \mu_0$ ngược lại $H_a : \mu > \mu_0$ (khi μ_0 là một số thoả yêu cầu). Biểu diễn kiểm định mà với kiểm định thống kê $\overline{X}$ và vùng bác bỏ $\overline{x} \geq \mu_0 + 2.33\sigma/\sqrt{n}$ có mức ý nghĩa .01
b. giả sử quá trình của câu (a) dùng kiểm định $H_0 : \mu \leq \mu_0$ ngược lại $H_a : \mu > \mu_0$. Nếu $\mu_0 = 100, n = 25$ và $\sigma = 5$. Xác suất nào cam kết xảy ra sai lầm loại I khi $\mu = 99$? Khi $\mu = 98$? nhìn chung, có thể nói gì về sai lầm xác suất loại Ikhi giá trị thực của μ ít hơn μ_0? Kiểm chứng cho khẳng định của bạn.

14. Xét lại tình huống trong Bài tập 11 và giả sử vùng bác bỏ là $\{\overline{x} : \overline{x} \geq 10.1004 \text{ hay } \overline{x} \leq 9.8940\} = \{z : z \geq 2.51\text{hay } z \leq -2.65\}$.

a.α cho quá trình này bằng bao nhiêu?
b. β bằng bao nhiêu khi $\mu = 10.1$? Khi $\mu = 9.9$? điều này nói lên mong muốn gì?

8 8.2 Kiểm định về một trung bình của tổng thể

Những thảo luận chung trong chương 7 về Khoảng tin cậy cho một tổng thể có kỳ vọng μ tập trung ở ba trường hợp khác nhau. Ta sẽ xây dựng quá trình kiểm định cho những trường hợp này.
Trường hợp I: Một tổng thể chuẩn với Σ đã biết

Mặc dù giả thiết về giá trị σ đã biết rất hiếm gặp trong thực tế, trường hợp này cung cấp một khởi đầu tốt vì dễ thấy với quá trình chung và tính chất của nó có thể xây dựng được. Gỉa thiết không trong cả ba trường hợp nói rằng μ có giá trị đặc biệt, giá trị không, sẽ được ký hiệu là μ_0. Đặt $X_1, ..., X_n$ là một mẫu ngẫu nhiên kích thước n có phân phối chuẩn. Thì trung bình mẫu $\overline{X}$ cũng có phân phối chuẩn với giá trị kỳ vọng $\mu_{\overline{X}} = \mu$ và độ lệch chuẩn $\sigma_{\overline{X}} = \sigma/\sqrt{n}$. Khi H_0 đúng, $\mu_{\overline{X}} = \mu_0$. Bây giờ ta xét thống kê Z chuẩn hóa bởi $\overline{X}$ với giả thiết H_0 đúng.:

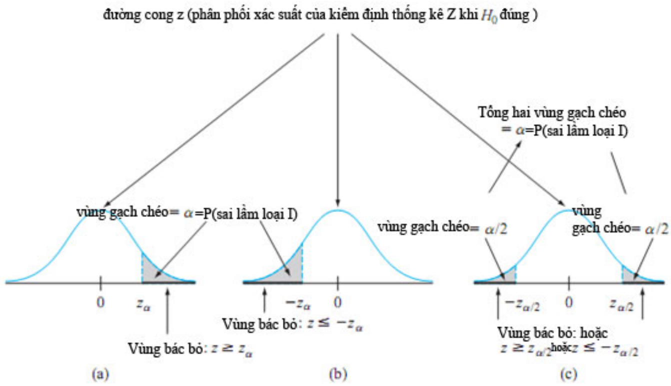
$$Z = \frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}}$$

Tính trung bình mẫu thay thế $\overline{x}$ cho z, khoảng cách giữa $\overline{x}$ và μ_0 được trình bày trong "đơn vị độ lệch chuẩn". Ví dụ, nếu giả thiết không là $H_0 : \mu = 100, \sigma_{\overline{x}} = \sigma/\sqrt{n} = 10/\sqrt{25} = 2.0$, và $\overline{x} = 103$, thì giá trị kiểm định thống kê là một độ đo tự nhiên của khoảng cách giữa $\overline{X}$, ước lượng theo μ, và tính giá trị kỳ vọng của nó khi H_0đúng. Nếu khoảng cách đó là tuyệt vời theo một hướng nhất định với H_a , nên bác bỏ giả thiết không.
Gỉa sử ban đầu giả thiết thay thế có dạng $H_a : \mu > \mu_0$. Thì một giá trị $\overline{x}$ ít hơn μ_0 chắc chắn không thể cung hỗ trợ cho H_a. Vì thế một $\overline{x}$ phù hợp là một giá trị âm của z (khi $\overline{x} - \mu_0$là số âm và số chia $\sigma/\sqrt{n}$ là số âm). Tương tự, một giá trị $\overline{x}$ mà vượt quá μ_0 chỉ một số lượng nhỏ (tương ứng với z, đó là số âm nhưng nhỏ) cũng không gợi ý rằng nên bác bỏ H_0 và ủng hộ H_a. Bác bỏ H_0 phù hợp chỉ khi $\overline{x}$ vượt quá đáng kể μ_0 - nghĩa là, khi giá trị z âm và lớn. Trong tóm tắt, vùng bác bỏ phù hợp, dựa trên kiểm định thống kê Z hơn là $\overline{X}$, có dạng $z \geq c$. Trong lập luận của phần 8.1, giá trị cắt c nên chọn để xác định xác suất của sai lầm loại I tại mức α mong muốn. Điều này dễ dàng làm được vì phân phối của kiểm định thống kê Z khi H_0 đúng là phân phối chuẩn (đó là lý do tại sao μ_0 đã được trừ ra trong chuẩn hóa). Nhắc lại giá trị cắt c là giá trị điều kiện cho z mà được lấy α ở vùng phía bên phảidưới đường cong z. Như một ví dụ, đặt c=1.645, giá trị lấy ở phía vùng .05 ($z_{.05} = 1.645$). Thì

$\alpha = P(\text{sai lầm loại I}) = P(\text{bác bỏ} H_0 \text{khi} H_0 \text{đúng})$
$= P(Z \geq 1.645 \text{ khi} Z \sim N(0,1)) = 1 - \Phi(1.645) = .05$
Nói chung, vùng bác bỏ $z \geq z_\alpha$ có xác suất sai lầm loại I mức α. Qúa trình kiểm định là kiểm định bên phải vì vùng bác bỏ chỉ gồm những giá trị lớn của kiểm định thống kê.
Lập luận tương tự cho giả thiết thay thế $H_a : \mu < \mu_0$ đề nghị vùng bác bỏ có dạng $z \leq c$, khi đó c là số âm được chọn thỏa mãn ($\overline{x}$ở phía dưới μ_0 nếu và chỉ nếu z là số nguyên âm). Vì Z có phân phối chuẩn khi H_0 đúng , lấy $c = -z_\alpha$ $P($ sai lầm loại I $) = \alpha$. Đây là kiểm định bên trái. Lấy ví dụ , $z_{.10} = 1.28$ suy ra vùng bác bỏ $z \leq -128$ kiểm định cụ thể với mức ý nghĩa .10
Cuối cùng, khi giả thiết thay thế là $H_a : \mu \neq \mu_0$, nên bác bỏ H_0 nếu $\overline{x}$ quá xa về phía μ_0. điều này tương đương với bác bỏ H_0 hay nếu $z \geq c$ hoặc nếu $z \leq -c$. Gỉa sử ta muốn $\alpha = .5$. Thì:

$.05 = P(Z \geq c \text{hay} Z \leq -c$ khi Z có phân phối chuẩn)
$= \Phi(-c) + 1 - \Phi(c) = 2[1 - \Phi(c)]$
Do đó tồn tại c sao cho $1 - \Phi(c)$, vùng dưới đường cong z về phía bên phải của c, là .025 (và không .05!). Từ phần 4.3 hay bảng phụ lục A.3, c=1.96 , và với vùng bác bỏ là $z \geq 1.96$ hay $z \geq -1.96$. Cho α bất kỳ , kiểm định hai phía cho vùng bác bỏ $z \geq z_{\alpha/2}$ hai bên của đường cong z). Lặp lại điều này ta thấy, lý do chính của việc sử dụng chuẩn hóa trong kiểm định thống kê Z là vì Z có phân phối đã biết khi H_0 đúng (chuẩn thông thường) khi đó vùng bác bỏ với mong muốn xác suất sai lầm loại I dễ dàng có được bằng cách sử dụng một giá trị điều kiện phù hợp.
Qúa trình kiểm định cho trường hợp I được tóm tắt trong hộp đính kèm, và vùng bác bỏ được minh họa trong hình 8.2
Gỉa thiết không: $H_0 : \mu = \mu_0$
Gía trị kiểm định thống kê: $z = \frac{\overline{x}-\mu_0}{\sigma/\sqrt{n}}$

Gỉa thiết thay thế	Vùng bác bỏ cho kiểm định mức α
$H_a : \mu > \mu_0$	$z \geq z_\alpha$(kiểm định phía bên phải)
$H_a : \mu < \mu_0$	$z \geq z_\alpha$(kiểm định phía bên trái)
$H_a : \mu \neq \mu_0$	hoặc $z \geq z_{\alpha/2}$hoặc $z \leq -z_{\alpha/2}$(kiểm định hai bên)



Hình 8.2 Vùng bác bỏ cho kiểm địnhz:(a) kiểm định bên trái; (b) kiểm định bên phải; (c) kiểm định hai bên
Dùng những bước theo thứ tự được đề nghị sau khi kiểm định giả thiết về tham số.
1. Xác định tham số cần quan tâm và mô tả nó trong bối cảnh xảy ra vấn đề.
2. Xác định giá trị không và vùng của giả thiết không.
3. Vùng phù hợp với giả thiết thay thế
4. Cho một công thức để tính giá trị của kiểm định thống kê (thay thế giá trị không và những giá trị đã biết của bất kỳ những tham số khác, nhưng không dựa trên số lượng mẫu.
5. Vùng bác bỏ cho lựa chọn mức ý nghĩa α
6. Tính số lượng mẫu bất kỳ, thay thế vào công thức cho kiểm định giá trị thống kê, và tính giá trị đó.
7. Quyết định nên bác bỏ H_0, và đưa ra kết luận trong bối cảnh này.
Công thức của giả thiết (Bước 2 và 3) nên thực hiện trước khi kiểm tra dữ liệu.
VÍ DỤ 8.6
nhà sản xuất hệ thống vòi phun nước dùng bảo vệ lửa trong cao ốc văn phòng phát biểu rằng trung bình đúng của hệ thống sẽ được kích hoạt khi nhiệt độ là 130^o. Một mẫu $n = 9$ hệ thống, khi kiểm tra có năng suất của trung bình mẫu kích hoạt với nhiệt độ 131.08^oF. Nếu phân phối của thời gian hoạt động là chuẩn với độ lệch chuẩn 1.5^oF dữ liệu mâu thuẫn nào với phát biểu của nhà sản xuất có mức ý nghĩa $\alpha = .01$?
1. tham số quan tâm :$\mu =$ trung bình đúng nhiệt độ kích hoạt .

2. Gía trị không :$H_0 : \mu = 130$(giá trị không $= \mu_0 = 130$).
3. Gỉa thiết thay thế : $H_a : \mu \neq 130$(Bắt đầu từ giá trị của phát biểu trong cả những liên quan trực tiếp)
4. Gía trị kiểm định thống kê:

$$z = \frac{\overline{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{\overline{x} - 130}{1.5\sqrt{n}}$$

5. Vùng bác bỏ: Dạng của H_a có nghĩa là sử dụng kiểm định hai phía với vùng bác bỏ hoặc $z \geq z_{.005}$ hoặc $z \geq -z_{.005}$. Từ phần 4.3 hay bảng phụ lục A.3, $z_{.005} = 2.58$, vì thế ta bác bỏ H_0 nếu hoặc $z \geq 2.58$ hoặc $z \geq -2.58$.
6. Thay n=9 và $\overline{x} = 131.08$,

$$z = \frac{131.08 - 130}{1.5\sqrt{9}} = \frac{1.08}{.5} = 2.16$$

Nghĩa là,trung bình mẫu quan sát được nhiều hơn 2 độ lệch chuẩn ở trên kỳ vọng bằng bao nhiêu khi H_0 đúng// 7. Tính giá trị $z = 2.16$ không rơi vào vùng bác bỏ $(-2.58 < 2.16 < 2.58)$.Vì thế không thể bác bỏ H_0 tại mức ý nghĩa .01. Dữ liệu không hỗ trợ cho phát biểu mà trung bình đúng khác với giá trị dự kiến 130.
β và xác định cỡ mẫu Kiểm định z cho trường hợp I là một trong số ít trong thống kê cho ra công thức đơn giản sẵn có cho β, xác suất của sai lầm loại II. Xét kiểm định bên phải đầu tiên với vùng bác bỏ $z \geq z_\alpha$. Điều này tương đương với $\overline{x} \geq \mu_0 + z_\alpha.\sigma/\sqrt{n}$, vì thế chấp nhận H_0 nếu $\overline{x} < \mu_0 + z_\alpha.\sigma/\sqrt{n}$. Bây giờ ta đặt μ' ký hiệu một giá trị đặc biệt μ mà gần với giá trị không μ_0. Thì.
$\beta(\mu') = P(\text{chấp nhận} H_0 \text{khi } \mu = \mu')$
$= P(\overline{X} < \mu_0 + z_\alpha.\sigma/\sqrt{n} \text{khi } \mu = \mu')$
$= P\left(\frac{\overline{X}-\mu'}{\sigma/\sqrt{n}} < z_\alpha + \frac{\mu_0-\mu'}{\sigma/\sqrt{n}} \text{khi } \mu = \mu'\right)$
$= \Phi\left(z_\alpha + \frac{\mu_0-\mu'}{\sigma/\sqrt{n}}\right)$
Khi μ' tăng, $mu_0 - \mu'$ trở thành số âm, vì thế $\beta(\mu')$sẽ nhỏ khi μ' vượt quá μ_0(vì giá trị tại Φ được dự đoán sẽ vượt quá). Lỗi xác suất cho kiểm định bên trái và kiểm định hai bên có nguồn gốc giống như vậy.
Nếu σ lớn, xác suất của sai lầm loại II có thể lớn tại một giá trị thay thế μ' đó là điều mà điều tra viên quan tâm đặc biệt. Gỉa sử ta thay α và β bằng một giá trị khác. Trong ví dụ về giọt nước, văn phòng công ty có thể thấy $\mu' = 132$ như một sự hao hụt đáng kể từ $H_0 : \mu = 130$ vì thế muốn $\beta(132) = .10$ phải thêm $\alpha = .01$. Nói chung, khi xét hai hạn chế $P($ sai lầm loại I$) = \alpha$ và $\beta(\mu') = \beta$ để xác định α, μ' và β. thì với một kiểm định bên phải, cỡ mẫu n nên chọn để thỏa mãn là

$$\Phi\left(z_\alpha + \frac{\mu_0 - \mu'}{\sigma/\sqrt{n}}\right) = \beta$$

Suy ra
$-z_\beta = z$giá trị điều kiện mà lấy trong kiểm định bên trái $\beta = z_\alpha + \frac{\mu_0-\mu'}{\sigma/\sqrt{n}}$
giải phương trình này ta sẽ tìm được giá trị n mong muốn. Song song đó ta trình bày sự cần thiết của cỡ mẫu cho kiểm định bên trái và kiểm định hai bên được tóm tắt sau đây

Gỉa thiết thay thế	Sai lầm xác suất loại II cho mức kiểm định α
$H_a : \mu > \mu_0$	$\Phi\left(z_\alpha + \frac{\mu_0-\mu'}{\sigma/\sqrt{n}}\right)$
$H_a : \mu < \mu_0$	$1 - \Phi\left(-z_\alpha + \frac{\mu_0-\mu'}{\sigma/\sqrt{n}}\right)$
$H_a : \mu \neq \mu_0$	$\Phi\left(z_\alpha + \frac{\mu_0-\mu'}{\sigma/\sqrt{n}}\right) - \Phi\left(-z_\alpha + \frac{\mu_0-\mu'}{\sigma/\sqrt{n}}\right)$

khi $\Phi(z)$ = chuẩn thông thường cdf.
Cỡ mẫu n cho mức kiểm định α sao cho $\beta(\mu') = \beta$ tại giá trị thay thế μ' là

n= $\begin{cases} \left[\frac{\sigma(z_\alpha+z_\beta)}{\mu_0-\mu'}\right]^2 \text{cho kiểm định một bên (trái hay phải)} \\ \left[\frac{\sigma(z_{\alpha/2}+z_\beta)}{\mu_0-\mu'}\right]^2 \text{cho kiểm định hai bên (giải pháp xấp xỉ)} \end{cases}$

VÍ DỤ 8.7
Đặt μ ký hiệu là trung bình đúng của một loại lốp xe nhất định.Xét kiểm định $H_0 : Mu = 30,000$ ngược lại $H_a : \mu > 30,000$ dựa trên cỡ mẫu n=16 từ một phân phối chuẩn với $\sigma = 1500$. Yêu cầu kiểm định với $\alpha = .01, z_\alpha = z_{.01} = 2.33$ .Xác suất cho sai lầm loại II khi $\mu = 31,000$ là $\beta(31,000) = \Phi\left(2.33 + \frac{30,000-31,000}{1500/\sqrt{16}}\right) = \Phi(-.34) = .3669$
Khi đó $z_{.1} = 1.28$, mức kiểm định cần thiết .01 như yêu cầu là $\beta(31,000) = .1$ và cỡ mẫu n cần tìm là

$$n = \left[\frac{1500(2.33 + 1.28)}{30,000 - 31,000}\right]^2 = (-5.42)^2 = 29.32$$

Vì cỡ mẫu phải là số nguyên nên n=30.
TRƯỜNG HỢP II: KIỂM ĐỊNH CHO MẪU LỚN

Khi cỡ mẫu lớn, Kiểm định z cho trường hợp I có thể dễ dàng xác định giá trị cho quá trình kiểm định mà không cần yêu cầu là phân phối chuẩn hay σ đã biết. Kết quả chính được dùng trong chương 7 để chứng minh khoảng tin cậy cho mẫu lớn.
Với n lớn suy ra giá trị chuẩn hóa là:

$$Z = \frac{\overline{X} - \mu}{S/\sqrt{n}}$$

có phân phối xấp xỉ chuẩn. Thay giá trị μ bằng giá trị không μ_0 ta được kiểm định thống kê

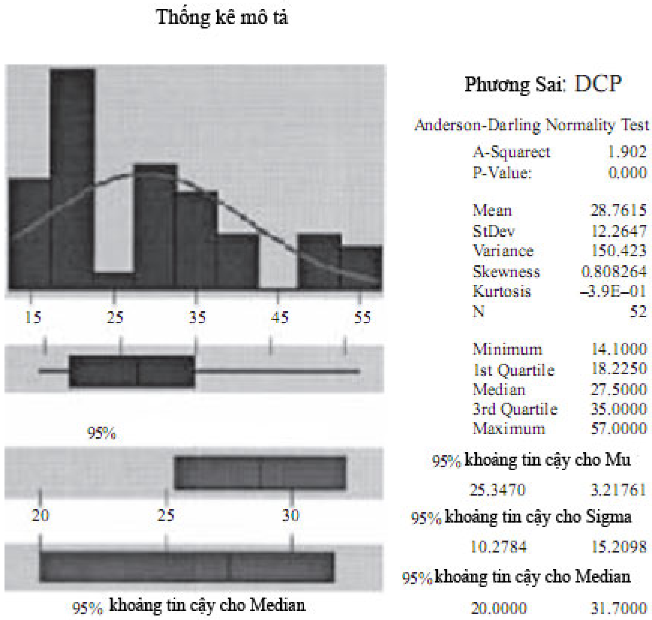
$$Z = \frac{\overline{X} - \mu_0}{S/\sqrt{n}}$$

Có phân phối xấp xỉ chuẩn khi H_0 đúng. Sử dụng vùng bác bỏ trước đây cho trường hợp I (ví dụ như, $x \geq z_\alpha$ khi giả thiết thay thế là $H_a : \mu > \mu_0$ thì kết quả trong quá trình kiểm định có mức ý nghĩa xấp xỉ (chính xác hơn) là α. khi đó qui tắc chung cho n>40 sẽ được sử dụng lại với đặc điểm cỡ mẫu lớn
VÍ DỤ 8.8
Một máy đo trọng lực hình nón (DCP) dùng để đo độ thấm xuyên ($mm/blow$) có hình nón được đưa vào vỉa hè hay lề đường. Gỉa sử cho một ứng dụng đặc biệt mà có giá trị trung bình đúng DCP cho một loại hình nón đặc biệt ít hơn 30. Hình nón này sẽ không được ứng dụng nếu không có bằng chứng kết luận kết luận về đặc điểm kỹ thuật đã được đáp ứng. Đặt bài toán thích hợp về kiểm định giả thiết sử dụng dữ liệu sau("Mô hình xác suất phân tích các giá trị kiểm tra sức cản thấm xuyên của nón trọng lực trong đánh giá cấu trúc vỉa hè")J.of Testing and Evaluation, 1999:7-14)

14.1	14.5	15.5	16.0	16.0	16.7	16.9	17.1	17.5	17.8
17.8	18.1	18.2	18.3	18.3	19.0	19.2	19.4	20.0	20.0
20.8	20.8	21.0	21.5	23.5	27.5	27.5	28.0	28.3	30.0
30.0	31.6	31.7	31.7	32.5	33.5	33.9	35.0	35.0	35.0
36.7	40.0	40.0	41.3	41.7	47.5	50.0	51.0	51.8	54.4
55.0	57.0								

Hình 8.3 Biểu diễn một bảng tóm tắt từ phần mềm minitab. Trung bình mẫu DCP thì ít hơn

30. Tuy nhiên có một lượng lớn giá trị trong dữ liệu (hệ số mẫu của phương sai= $s/\overline{x} = .4265$). Vì thế kỳ vọng ít hơn nói lên rằng việc cắt giảm đặc điểm thiết kế có thể gây ra sự biến đổi mẫu ,chú ý rằng biểu đồ của nó không giống với đường cong chuẩn (và một vùng xác suất chuẩn không thể hiện tính tuyến tính), nhưng kiểm định mẫu lớn z lại không yêu cầu phân phối xác suất chuẩn.



1.μ = giá trị trung bình đúng DCP

2. $H_0 : \mu = 30$

3. $H_a : \mu < 30$(vỉa hè sẽ không được sử dụng nếu không bác bỏ H_0)

4. $z = \frac{\overline{x}-30}{s/\sqrt{n}}$

5. Kiểm định với mức ý nghĩa .05 bác bỏ H_0 khi $z \leq -1.645$(kiểm định phía bên trai1)

6. Với $n = 52, \overline{x} = 28.76$, và $s = 12.2647$,

$$z = \frac{28.76 - 30}{12.2647/\sqrt{52}} = \frac{-1.24}{1.701} = -.73$$

7. Khi đó do $-.73 > -1.645$, không thể bác bỏ H_0. Ta không có bằng chứng thuyết phục ngay cả khi $\mu < 30$; dùng vỉa hè là không hợp lý.

Xác định β và cỡ mẫu cần thiết cho kiểm định mẫu lớn có thể dựa trên việc xác định một giá trị hợp lý σ và dùng công thức trong trường hợp I (có thể dùng s trong kiểm định) hay sử dụng phương pháp ngắn gọn trong trường hợp III sau.

TRƯỜNG HỢP III: PHÂN PHỐI CHUẨN CỦA MỘT TỔNG THỂ

Khi n nhỏ, Định lý giới hạn trung tâm không còn có thể viện dẫn để kiểm định cho mẫu lớn. Ta phải đối mặt với những khó khăn bao gồm một khoảng tin cậy nhỏ CI cho μ xem trong Chương 7. Cách tiếp cận ở đây cũng được dùng ở đây. Ta sẽ giả định rằng phân phối của tổng thể ít nhất xấp xỉ chuẩn và mô tả quá trình kiểm định có hiệu lực dựa trên giá trị giả định này, Nếu quan sát viên có lý do tốt để tin rằng phân phối của tổng thể khá là bất thường một kiểm định cho phân phối tự do từ Chương 15 có thể sử dụng.

Nói cách khác, một thống kê có thể đưa ra giá trị cho quá trình của họ các phân phối khác với họ chuẩn. Hay có thể phát triển bằng quá trình bootstrap.

Kết quả chính trong kiểm định cho một trung bình tổng thể chuẩn dựa vào Chương 7 để lấy được một mẫu t CI: Nếu $X_1, X_2, ..., X_n$ là biến ngẫu nhiên từ phân phối chuẩn, giá trị chuẩn hóa:

$$T = \frac{\overline{X} - \mu}{S/\sqrt{n}}$$

có phân phối t với n bậc tự do (df) . Xét kiểm định $H_0 : \mu = \mu_0$ ngược lại $H_a : \mu > \mu_0$ bằng cách kiểm định thống kê $T = (\overline{X} - \mu_0)/(S/\sqrt{n})$. Nghĩa là, kết quả của kiểm định thống kê từ biến chuẩn hóa $\overline{X}$ với giả thiết này H_0 đúng (dùng $S/\sqrt{n}$, để ước lượng cho độ lệch chuẩn của $\overline{X}$ hơn là dùng $\sigma/\sqrt{n}$. Khi H_0 đúng, kiểm định thống kê có phân phối t với n-1 df. Theo hiểu biết về kiểm định phân phối của thống kê khi H_0 đúng (đó là "phân phối không") chấp nhận cho ta xây dựng vùng bác bỏ của sai lầm xác suất loại I là xác định mức mong muốn. Đặc biệt sử dụng kiểm định bên phải t cho giá trị điều kiện $t_{\alpha,n-1}$ để xác định vùng bác bỏ $t \geq t_{\alpha,n-1}$ suy ra

$P($ sai lầm loại I $) = P($Bác bỏH_0khi nó đúng)

$= P(T \geq t_{\alpha,n-1}$khi T có phân phối t với n-1 df)

$= \alpha$

Kiểm định thống kê này giống như trong trường hợp mẫu lớn nhưng đặt làT để nhấn mạnh rằng phân phối không của nó là một phân phối t với n-1 df hơn là phân phối chuẩn (z). Vùng bác bỏ cho kiểm định t khác hơn từ kiểm định z mà trong đó chỉ có một t với giá trị điều kiện $t_{\alpha,n-1}$ được thay bởi giá trị điều kiện z_α. áp dụng tương tự để thay cho kiểm định bên trái hay kiểm định hai bên là thích hợp

Kiểm định một mẫu t

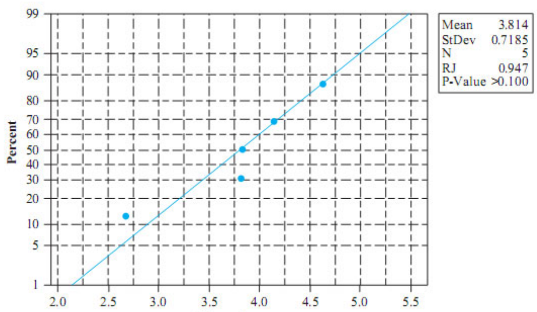
Gỉa thiết không: $H_0 : \mu = \mu_0$

Kiểm định giá trị thống kê: $t = \frac{\overline{x} - \mu_0}{s/\sqrt{n}}$

Gỉa thiết thay thế	Vùng bác bỏ cho kiểm định mức α
$H_a : \mu > \mu_0$	$t \geq t_{\alpha,n-1}$(kiểm định bên phải)
$H_a : \mu < \mu_0$	$t \leq t_{\alpha,n-1}$(kiểm định bên trái)
$H_a : \mu \neq \mu_0$	hoặc $t \geq t_{\alpha,n-1}$ hoặc $t \leq t_{\alpha,n-1}$(kiểm định hai bên)

VÍ DỤ 8.9

Glycerol là sản phẩm chính trong quá trình lên men của rượu vang, góp phần tăng độ ngọt, hương vị, và dậm vị rượu. Bài viết "Phương pháp nhanh và đơn giản để xác định thành phần glycerol, fructose và glucose trong rượu" (American J. of Enology and Viticulture, $2007 : 279 - 283)$ bao gồm những quan sát sau về sự tập trung của glycorol (mg/ml) Lấy ví dụ về chất lượng chuẩn (không chắc) của rượu trắng:2.67, 4.62, 4.14, 3.81, 3.83,.Gỉa sử giá trị tập trung của glycerol mong muốn là 4. Dữ liệu mẫu nào nên lấy mà trung bình đúng của tập trung glycerol đôi khi hơn giá trị mong muốn? Kèm theo sau đây là vùng xác suất chuẩn có từ phần mềm Minitab hỗ trợ mạnh mẽ cho giả thiết phân phối tổng thể của sự tập trung của glycerol là chuẩn.Dùng dữ liệu đó để kiểm định giả thiết xấp xỉ bằng cách dùng kiểm định một mẫu t với mức ý nghĩa .05.



Hình 8.4 Vị trí của phân phối chuẩn có trong dữ liệu của ví dụ 8.9

1. $\mu =$ trung bình đúng của sự tập trung glycerol

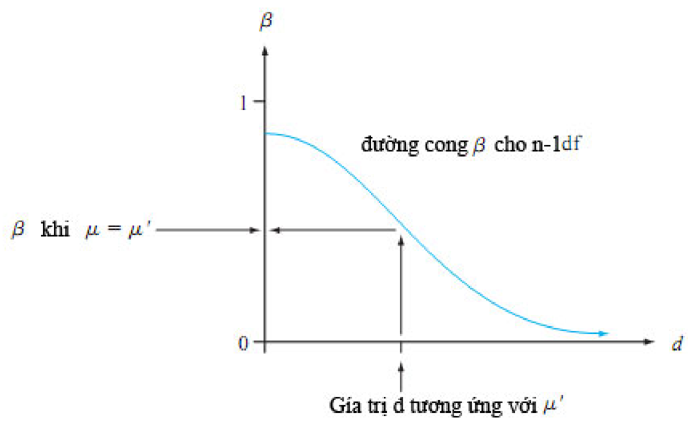
2. $H_0 : \mu = 4$
3. $H_a :\neq 4$
4. $t = \frac{\overline{x}-4}{s/\sqrt{n}}$
5. Bất đẳng thức trong H_a suy ra kiểm định hai bên là xấp xỉ, nhắc lại rằng $t_{\alpha/2,n-1} = t_{2.776}$.Vì thế H_0 sẽ bị bác bỏ nếu $t \geq 2.776$ hoặc $t \leq -2.776$.
6. $\sum x_i = 19.07$, và $\sum x_i^2 = 74.7979$, từ $\overline{x} = 3.814, s = .718$, vàước lượng chuẩn lỗi của trung bình là$s/\sqrt{n} = .321$. giá trị của kiểm định thống kê đó là $t = (3.814 - 4)/.321 = -.58$
7. Rõ ràng $t = -.58$ không nằm trong vùng bác bỏ với mức ý nghĩa .05. Điều này lại hợp lý với $\mu = 4$. Độ lệch của trung bình mẫu là 3.814 có từ giá trị kỳ vọng 4 khi H_0 đúng có thể chỉ dùng để thay đổi mẫu chứ không phải để H_0 sai.
Theo phần mềm Minitab với yêu cầu biểu diễn kiểm định hai bên của một mẫu t biểu diễn và tính những giá trị đồng nhất cho những gì vừa thu được . Thực tế số cuối cùng được xuất ra là giá trị p vượt quá .05 (với bất kỳ lý do nào có được mức ý nghĩa khác) Suy ra không thể bác bỏ giả thiết không. Điều này sẽ được thảo luận chi tiết trong phần 8.4

```
Test of mu = 4 vs not  = 4
Variable    N   Mean  StDev  SE Mean  95% CI              T      P
glyc conc   5  3.814  0.718    0.321  (2.922,  4.706)  -0.58  0.594
```

XÁC ĐỊNH β VÀ CỠ MẪU
Tính β tại giá trị thay thế μ' trong trường hợp I Thực hiện điều này để xác định vùng bác bỏ trong số hạng của $\overline{x}$ (ví dụ , $\overline{x} \geq \mu_0 + z_\alpha.\sigma/\sqrt{n}$ sau đó trừ đi μ' để chuẩn hóa chính xác. Một cách tiếp cận tương đương bao gồm việc ghi nhận rằng khi $\mu = \mu'$, kiểm định thống kê cũng có phân phối chuẩn mà không cần thông qua tiêu chuẩn của phân phối chuẩn.
Vì thế, $\beta(\mu')$ là vùng dưới của đường cong chuẩn tương ứng với giá trị trung bình $(\mu' - \mu_0)/(\sigma/\sqrt{n})$ và phương sai bằng 1. cả α và β đều dựa trên biến của phân phối chuẩn .
Tính $\beta(\mu')$ cho kiểm định t thì phức tạp hơn. Đó là vì phân phối của kiểm định thống kê $T = (\overline{X} - \mu_0)/(S/\sqrt{n})$ khá là rắc rối khi H_0 sai và H_a đúng . Do đó , với một kiểm định bên phải, ta cần xác định

$$\beta(\mu') = P(T < t_{\alpha,n-1} \text{ khi } \mu = \mu' \text{lớn hơn } \mu_0)$$

Liên quan đến việc tích hợp hàm mật độ. Điều này phải được thực hiện bằng số những kết quả được tóm tắt trong đồ thị của β có trong bảng Phụ lục A.17. Có 4 nhóm đồ thị tương ứng với kiểm định một bên tại mức .05 và mức .01 và kiểm định hai bên tại cùng một mức.
Để hiểu cách sử dụng đồ thị như thế nào, đầu tiên chú ý cả β và cỡ mẫu cần thiết trong trường hợp I là không chỉ là hàm với giá trị sai khác $|\mu_0 - \mu'|$ nhưng của $d = |\mu_0 - \mu'|/\sigma$. Gỉa sử , cho ví dụ, Với $|\mu_0 - \mu'| = 10$ xuất phát từ H_0 dễ thấy (với β nhỏ hơn) khi $\sigma = 2$, trong trường hợp mu_0 và μ' có độ lệch chuẩn từng phần là 5, khiđó $\sigma = 10$. Thực tế β cho khẳng định t phụ thuộc vào d thay vì chỉ phụ thuộc $mu_0 - \mu'$ là không phù hợp, khi đó, để sử dụng một đồ thị ta phải tham khảo thêm ý của trung bình đúng của σ
Tính bảo toàn (lớn) dự đoán cho σsẽ mang lại tính bảo toàn (lớn) cho giá trị của $\beta(\mu)$ và một ước lượng cho sự bảo toàn về cỡ mẫu cần thiết theo yêu cầu của α và $\beta(\mu')$.
Một thay thế μ' và chọn được một giá trị σ, sao cho tính được d và giá trị địa phương của nó trên trục ngang của tập đường cong những giá trị β có độ cao n-1 df trên giá trị của d (dùng trực quan khi cần khi n-1 không là giá trị phù hợp có trên đường cong) như minh họa trong hình 8.5



Hình 8.5 Một loại đường cong β cho kiểm định t

Hơn nữa khi n thỏa mãn (ví dụ n-1, vì thế đường cong cụ thể từ β đọc được) , có thể có hai giá trị α (.05 hay .01 ở đây) và một giá trị β cho lựa chọn μ' và σ. Sau khi tính d, điểm (d, β) có vị trí trên tập những đồ thị liên quan . Đường cong phía trên và điểm kết thúc đã cho là n-1 và vì thế tìm được n (ngược lại, dùng phép nội suy khi cần)

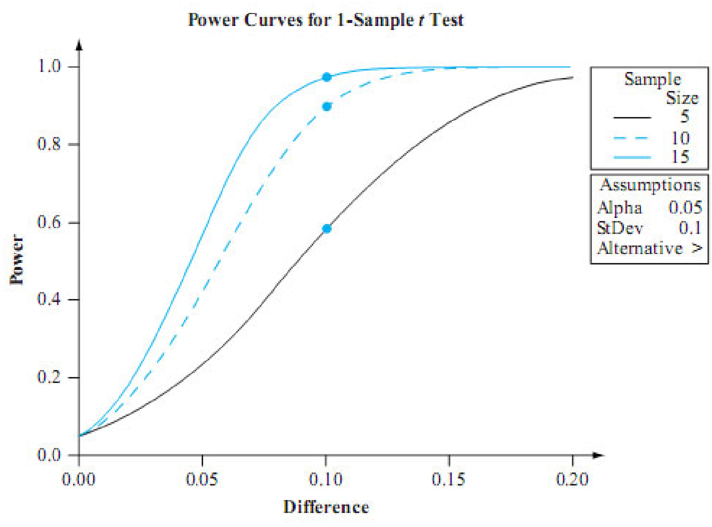
VÍ DỤ 8.10

Trung bình đúng của điện áp giảm từ bộ thu phát của cổng lưỡng cực một loại bóng bán dẫn nhất định.giả sử điện áp đạt ít nhất 2.5 volt . Quan sát viên chọn một mẫu n=10 bóng bán dẫn và dùng kết quả điện áp đó để kiểm định $H_0 : \mu = 2.5$ ngược lại $H_a : \mu > 2.5$ sử dụng kiểm định t với với mức ý nghĩa $\alpha = .05$. Nếu độ lệch chuẩn của phân phối điện áp là $\sigma = .100$, H_0 sẽ thế nào để H_0 không bị bác bỏ với giá trị thực tế $\mu = 2.6$? với $d = |2.5 - 2.6|/.100 = 1.0$, điểm trên đường cong β tại 9 df cho kiểm định một bên với $\alpha = .05$ trên 1.0 có chiều cao xấp xỉ .1,vì thế $\beta \approx .1$. quan sát viên có thể nghĩ rằng đó là giá trị quá lớn của β cho giả thiết H_0 và giá trị mong muốn $\beta = .05$ cho giá trị thay thế của μ. Khi đó $D = 1.0$, điểm $(d, \beta) = (1.0, .05)$ phải được xác định. Điểm này rất gần với 14df của đường cong vì thế khi dùng n=15 sẽ cho $\alpha = .05$ và $\beta = .05$ khi đó giá trị của $\mu = 2.6$ và $\sigma = .10$. Một giá trị σ lớn sẽ cho một giá trị β lớn với việc thay thế này, và một giá trị thay thế của μ gần với 2.5 cũng sẽ cho kết quả β tăng. Hầu hết các phần mềm thống kê được sử dụng rộng rãi đều có khả năng tính xác suất sai lầm loại II chúng thường làm việc dựa trên lũy thừa với trường hợp đơn giản $1 - \beta$. Một giá trị β nhỏ (dần về 0) tương đương với lũy thừa lớn (gần bằng 1) . Kiểm tra lũy thừa là một trong những kiểm định có độ chính xác cao và vì thế đó là khả năng tốt nhất để bảo vệ khi giả thiết không sai.

Như ví dụ, dùng phần mềm Minitab để xác định lũy thừa của kiểm định bên phải trong Ví dụ 8.10 cho ba cỡ mẫu 5,10 và 15 khi $\alpha = .05, \sigma = .10$, và giá trị thực sự của μ là 2.6 hơn nữa giá trị không 2.5-a là sai khác 2.6-2.5=.1. Ta cũng dùng phần mềm này để xác định cỡ mẫu cần thiết cho lũy thừa của .9$(\beta = .1)$ và cũng là .95. Sau đây là kết quả nó xuất ra

```
Power and Sample Size
Testing mean = null (versus > null)
Calculating power for mean = null + difference
Alpha = 0.05 Assumed standard deviation = 0.1
                 Sample
Difference        Size     Power
0.1                5       0.579737
0.1               10       0.897517
0.1               15       0.978916
                  Sample Target
                                              Actual
Difference        Size     Power              Power
0.1               11       0.90               0.924489
0.1               13       0.95               0.959703
```

Lũy thừa cho cỡ mẫu n là một số nhị phân nhỏ hơn .9 . Vì thế nếu ta nhấn mạnh rằng lũy thừa phải ít nhất là .9 ,khi đó cỡ mẫu 11 là bắt buộc và lũy thừa thực tế cho n đó gần bằng .92. Phần mềm nói rằng với mục tiêu của lũy thừa là .95, yêu cầu cho cỡ mẫu sẽ là n=13 trong khi bằng mắt thường ta lại thấy trên đường cong β là15.Với giá trị cho trước, phần mềm này đáng tin cậy hơn là dùng đường cong để xác định cỡ mẫu. Cuối cùng, phần mềm Minitab hiện nay cũng cung cấp đường cong lũy thừa cho cỡ mẫu đặc biệt, như biểu diễn trong hình 8.6. Những đường cong như thế cho thấy lũy thừa tăng như thế nào cho mỗi cỡ mẫu cũng như giá trị thực μ sẽ di chuyển xa và xa hơn giá trị không.



Hình 8.6 đường cong lũy thừa từ phần mềm minitab cho kiểm định t của Ví dụ 8.10

BÀI TẬP PHẦN 8.2(15 – 36)

15. Để kiểm định thống kê Z có phân phối chuẩn khi H_0 đúng. Cho biết mức ý nghĩa của mỗi tình huống sau:

a. $H_a : \mu > \mu_0$, vùng bác bỏ $z \geq 1.88$

b. $H_a : \mu < \mu_0$, vùng bác bỏ $z \leq -2.75$

c. $H_a : \mu \neq \mu_0$, vùng bác bỏ $z \geq .88$ hay $z \leq -2.88$

16. Để kiểm định thống kê T có phân phối t khi H_0 đúng. Cho biết mức ý nghĩa của mỗi tình huống sau:

a. $H_a : \mu > \mu_0, df = 15$, vùng bác bỏ $t \geq 3.733$

b. $H_a : \mu < \mu_0, df = 24$, vùng bác bỏ $t \leq -2.500$
c. $H_a : \mu \neq \mu_0, df = 31$, vùng bác bỏ $t \geq 1.697$ hay $t \leq -1.697$

17. Trả lời những câu hỏi sau cho vấn đề lốp xe trong Ví dụ 8.7
a. Nếu $\overline{x} = 30.960$ và một mức kiểm định $\alpha = .01$ được sử dụng, quyết định nào được đưa ra?
b. Nếu sử dụng một mức kiểm định .01, $\beta(30, 500)$ là gì ?
c. Nếu sử dụng một mức kiểm định .01 và nó cũng yêu cầu $\beta(30, 500) = .05$, cỡ mẫu n cần thiết là gì?
d. Nếu $\overline{x} = 30.960$, α nhỏ nhất bằng bao nhiêu sao cho H_0 bị bác bỏ (dựa vào n=16)?

18. Nhắc lại tình huống sơn khô của Ví dụ 8.2, trong đó thời gian làm khô cho kiểm định một mẫu có phân phối chuẩn với $\sigma = 9$. Gỉa thiết $H_0 : \mu = 75$ ngược lại $H_a : \mu < 75$ là kiểm định một mẫu ngẫu nhiên của n=25 lần quan sát.
a. Có bao nhiêu độ lệch chuẩn (của $\overline{X}$) dưới giá trị không là $\overline{x} = 72.3$
b. Nếu $\overline{x} = 72.3$ Kết luận nào sẽ dùng $\alpha = .01$
c. α bằng bao nhiêu cho quá trình kiểm định để bác bỏ H_0 khi $z \leq -2.88$
d. Cho quá trình kiểm định như câu (c), $\beta(70)$ bằng bao nhiêu?
e. Nếu dùng quá trình kiểm định trong phần (c), n cần thiết bằng bao nhiêu sao cho $\beta(70) = .01$?
f. Nếu dùng một mức kiểm định .01 với $n = 100$, Xác suất sai lầm loại I bằng bao nhiêu khi $\mu = 76$?

19. Độ tan chảy chảy của 16 mẫu của một thương hiệu nhất định dầu thực vật bị hydro hóa được xác định, kết quả trong $\overline{x} = 94.32$. Gỉa sử phân phối của độ tan chảy là chuẩn với $\sigma = 1.20$.
a. Kiểm định $H_0 : \mu = 95$ ngược lại $H_a : \mu \neq 95$ dùng kiểm định hai bên với mức kiểm định .01
b. Nếu dùng mức kiểm định .01, $\beta(94)$ bằng bao nhiêu, xác suất sai lầm loại II bằng bao nhiêu khi $\mu = 94$
c. Gía trị cần thiết của n bằng bao nhiêu để chắc chắn rằng $\beta(94) = .1$ khi $\alpha = .01$?

20. Một loại bóng đèn nhất định được quảng cáo có trung bình thời gian thắp sáng là 750 giờ. Gía của loại bóng này rất ưu đãi, vì thế khách hàng tiềm năng quyết định tiếp tục mua hàng trừ khi nó có thể được chứng minh rằng trung bình đúng của thời gian thắp sáng thì nhỏ hơn những gì đã quảng cáo. Một mẫu ngẫu nhiên gồm 50 bóng đèn được chọn, thời gian thắp sáng của mỗi bóng được xác định, và giả thiết phù hợp được kiểm định bằng phần mềm Minitab, kết quả xuất ra như sau:

Variable	N	Mean	StDev	SEMean	Z	P-Value
lifetime	50	738.44	38.20	5.40	−2.14	0.016

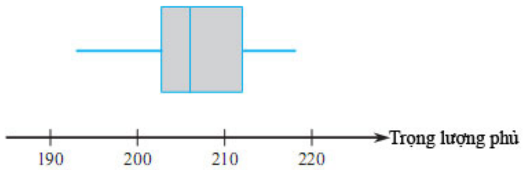
Quyết định phù hợp nào được đưa ra cho một mức ý nghĩa .05? Một mức ý nghĩa .01? Mức ý nghĩa nào và quyết định nào sẽ được gợi ý?

21. Trung bình đúng đường kính vòng bi của một loại xác định được cho là .5. Kiểm định một mẫu t được thực hiện để xem trong trường hợp này. Kết luận nào phù hợp trong mỗi tình huống sau?
a. $n = 13, t = 1.6, \alpha = .05$
b. $n = 13, t = -1.6, \alpha = .05$
c. $n = 25, t = -2.6, \alpha = .01$
d. $n = 25, t = -39$

22. Bài viết "Quan điểm của người quản đốc về quản lý chất lượng" (Quanlity Engr.,

1990 : 257 − 280) mô tả một quan sát trọng lượng lớp phủ cho các đường ống lớn từ quá trình chế biến mạ kẽm. Sản phẩm chuẩn được xét có trung bình đúng trọng lượng là 200 lb trên một ống. Bảng tóm tắt mô tả và sơ đồ hộp từ phần mềm Minitab :

```
Variable N    Mean Median TrMean StDev SEMean
ctg wt   30 206.73 206.00 206.81  6.35   1.16
Variable      Min     Max     Q1      Q3
ctg wt     193.00  218.00 202.75  212.00
```



a. Sơ đồ hộp cho gợi ý gì về phát biểu về đặc điểm kỹ thuật cho trung bình đúng trọng lượng lớp phủ?
b. Vị trí phân phối chuẩn của dữ liệu là khá thẳng hàng. Dùng mô tả mà nó xuất ra để kiểm định giả thiết thích hợp.

23. Bài tập 36 trong Chương I cho n=26 quan sát trên thời gian giải lao (*sec*) của công nhân dầu mỏ trong một bài tập mô phỏng, từ trung bình mẫu và độ lệch chuẩn mẫu tương ứng là 370.69 và 24.36. Gỉa sử điều tra viên tin có một khoảng thời gian mà trung bình đúng thời gian giải lao sẽ có ít nhất là 6 phút. Dữ liệu nào mâu thuẫn với thời gian tin tưởng đó? Gỉa thiết chuẩn, Kiểm định giả thiết thích hợp sử dụng mức ý nghĩa .05?

24. Nhắc lại mẫu quan sát một mẫu về độ nhớt ổn định của nhựa đường được giới thiệu trong Bài tập 46 của Chương I $(2781, 2900, 3013, 2856$và $, 2888)$. Gỉa sử rằng Với một ứng dụng cụ thể với yêu cầu trung bình đúng của độ nhớt là 3000. Yêu cầu này đã được kiểm định chưa? Phát biểu và kiểm định một giả thiết thích hợp.

25. Phần trăm mong muốn của SIO_2 của một loại xi măng nhôm xác định là 5.5. Kiểm định cho trung bình đúng phần trăm là 5.5 của một nhà máy sản xuất cụ thể , Gồm 16 mẫu độc lập được phân tích. Gỉa sử phần trăm của SIO_2 trong một mẫu có phân phối chuẩn với $\sigma = .3$ và $\overline{x} = 5.25$.
a. Điều này có chỉ ra kết luận rằng trung bình đúng phần trăm khác 5.5 không? Thực hiện phân tích này dùng trình tự các bước được đề nghị trong văn bản.
b. Nếu trung bình đúng phần trăm là $\mu = 5.6$ và mức kiểm định $\alpha = .01$ dựa trên $n = 16$ được dùng. Xác suất nào cho kiểm định H_0?
c. Gía trị của n bằng bao nhiêu để thỏa $\alpha = .01$ và $\beta(5.6) = .01$?

26. Để có thông tin về sự chống ăn mòn của một loại ống dẫn bằng thép xác định, 45 mẫu được chôn trong đất trong khoảng 2 năm. Sự thâm nhập tối đa(trong mils) cho mỗi mẫu là đo được, có trung bình mẫu thâm nhập $\overline{x} = 52.7$ và độ lệch chuẩn của mẫu $s = 4.8$. Các ống dẫn được sản xuất với các đặc điểm kỹ thuật mà trung bình đúng thâm nhập ít nhất 50 mils. Ta sẽ dùng nó trừ khi làm rõ kết luận về đặc điểm kỹ thuật có thể gặp. Kết luận đưa ra là gì?

27. Xác định ranh giới tự động của cấu trúc quan trọng trong hình ảnh y tế là lĩnh vực đang được nghiên cứu. Bài báo "Sự phân chia tự động của hình ảnh trong y học bằng ghi nhận hình ảnh: Ứng dụng trong chẩn đoán và mô phỏng" (J. of Medical Engr: and Tech., 2005 : 53 − 63) Thảo luận một kỹ thuật mới cho việc nhận biết. Thước đo sự chính xác của vùng tự động là trung bình khoảng cách tuyến tính (*ALD*). Bài báo cho ALD sau được quan sát từ 49 cuộc giải phẫu (đơn vị của kích thước điểm ảnh)

1.38	0.44	1.09	0.75	0.66	1.28	0.51
0.39	0.70	0.46	0.54	0.83	0.58	0.64
1.30	0.57	0.43	0.62	1.00	1.05	0.82
1.10	0.65	0.99	0.56	0.56	0.64	0.45
0.82	1.06	0.41	0.58	0.66	0.54	0.83
0.59	0.51	1.04	0.85	0.45	0.52	0.58
1.11	0.34	1.25	0.38	1.44	1.28	0.51

a. Tóm tắt/ mô tả dữ liệu
b. Điều đó có hợp lý không khi ALD ít nhất là xấp xỉ phân phối chuẩn? Gỉa định thời gian chuẩn phải thế nào để tính khoảng tin cậy CI cho trung bình đúng ALD hay kiểm định giả thiết về trung bình đúng ALD ? giải thích?
c. Tác giả cho rằng hầu hết những trường hợp ALD thì hơn hoặc phải 1.0. Dữ liệu thực tế nào cần để có bằng chứng mạnh cho kết luận trung bình đúng ALD cho những trường hợp dưới đây ít hơn 1.0? Thực hiện kiểm định xấp xỉ cho giả thiết đó.
d. Tính chặn trên khoảng tin cậy cho trung bình đúng ALD dùng mức tin cậy 95% và giải thích ràng buộc này
28. Tiểu phẫu trên ngựa trong điều kiện đồng ruộng đòi hỏi gây tê ngắn đáng tin cậy tạo sự thư giãn tốt cho cơ bắp, thay đổi tim mạch và hô hấp tối thiểu và hồi phục nhanh, với hiệu quả cao mà không cần giám sát ngựa. Bài viết "Thử nghiệm thực địa gây tê Ketamine trên ngựa" (Equine Vet. J.,1984:176-1794) báo cáo rằng cho một mẫu n=73 con ngựa với Ketamine được quản lý dưới điều kiện nhất định, trung bình mẫu thời gian nằm nghỉ (nằm phía dưới) là 8.6 phút. Dữ liệu cần là bao nhiêu để trung bình đúng thời gian nằm nghỉ dưới điều kiện đó ít hơn 20 phút? Kiểm định giả thiết xấp xỉ với mức ý nghĩa .10?

29. Bài viết "Thay đổi ước lượng với chi phí vòng đời của đường sắt" (J.of Rail and Rapid Transit, 2009) Dữ liệu hiện tại dựa trên thời gian sửa chữa (phút) một đoạn đường sắt bị gãy trong đường sắt cao tốc trên đường cong của một tuyến đường sắt nhất định.
159 120 480 149 270 547 340 43 228 202 240 218
Vị trí phân phối chuẩn của dữ liệu được biểu diễn bằng một mẫu tuyến tính hợp lý, Vì thế Điều này hợp lý với phân phối của tổng thể về thời gian sửa chữa ít nhất là có xấp xỉ chuẩn. Trung bình mẫu và độ lệch chuẩn tương ứng là 249.7 và 145.1
a. Bằng chứng thuyết phục cho kết luận rằng trung bình đúng thời gian sửa chữa vượt quá 200 phút? Hãy kiểm định giả thiết sử dụng mức ý nghĩa .05.
b. Dùng $\sigma = 150$, Xác suất sai lầm loại II trong kiểm định của câu (a) là gì khi trung bình đúng thời gian sửa chữa thực sự là 300 phút? theo đó, $\beta(300)$ bằng bao nhiêu?

30. Bạn có bao giờ thất vọng vì bạn không còn một chút gì trong thùng chứa cuối cùng của nó chưa? Bài viết" sự lắc, vắt và nén lại: Còn lại bao nhiêu trong thùng chứa đó?" (Consumer, Reports, May 2009: 8) Báo cáo trên một quan sát của vấn đề này cho các sản phẩm tiêu dùng khác nhau. Gỉa sử năm 6.0 oz loại kem đánh răng của một nhãn hiệu cụ thể là ngẫu nhiên được chọn và vắt không còn kem đánh răng đi ra. Sau đó mỗi ống kem bị cắt để mở và cân trọng lượng có trong đó, kết quả cho bởi dữ liệu sau (phù hợp với những gì được trích dẫn trong bài báo).53,.65, ,46, .50, .37. Nó có xuất hiện trung bình đúng số lượng bên trái ít hơn 10% của trọng lượng như trong quảng cáo không?
a. kiểm tra giá trị của bất kỳ giả thiết cần thiết cho kiểm định giả thiết xấp xỉ.
b. Thực hiện kiểm định giả thiết xấp xỉ với mức ý nghĩa .05. Quyết định thay đổi thế nào nếu sử dụng mức ý nghĩa .01?
c. Mô tả phần trước sai lầm loại I và loại II, và nói những lỗi có thể làm ảnh hưởng đến kết luận.

31.Một thiết kế hoàn hảo và an toàn của nơi làm việc góp phần đáng kể cho tăng năng suất. Điều quan trọng đặc biệt là công nhân viên không được hỏi để thực hiện bài tập, như là nâng lên, Điều đó vượt quá khả năng của họ. Dữ liệu kèm theo với trọng lượng lớn nhất của sức nâng

(MAWL, trong kg) cho sự lặp lại 4 lần nâng / phút được báo cáo trong bài viết "ảnh hưởng của tốc độ, thói quen, và tải trọng đo được với sàn bàn giao nâng"(Ergonomic, 1992:883-843) chủ đề được chọn ngẫu nhiên từ tổng thể của sức khỏe nam giới tuổi từ 18-30. Gỉa sử MAWL có phân phối chuẩn, dữ liệu nào cần để trung bình tổng thể MAWL vượt quá 25? Thực hiện kiểm định dùng mức ý nghĩa .05

25.8 36.6 26.3 21.8 27.2

32. Chế độ ăn kiêng được đề nghị đối với nam giới hơn 50 tuổi là lượng kẽm phải 15mg/ngày. Bài viết "Chế độ dinh dưỡng và khẩu phần ăn kiêng của người già ở Mỹ: Một nghiên cứu cấp quốc gia" (J. of Gerontology, 1992:M145-150) báo cáo dữ liệu tóm tắt sau trên lượng cho một mẫu nam giới tuổi 65-74 : n=115, $\overline{x} = 11.3$, và $s = 6.43$. Dữ liệu này có cho biết lượng kẽm trung bình hàng ngày trong tổng thể của tất cả nam giới tuổi từ 65-74 giảm xuống mức khuyến cáo chấp nhận được không?

33. Xét dữ liệu mẫu kèm theo về tỉ lệ chi phí (%) cho quĩ hỗ trợ tăng trưởng vốn lớn được giới thiệu đầu tiên trong Bài tập 1.53.

```
0.52 1.06 1.26 2.17 1.55 0.99 1.10 1.07 1.81 2.05
0.91 0.79 1.39 0.62 1.52 1.02 1.10 1.78 1.01 1.15
```

Biểu diễn vị trí phân phối chuẩn là mẫu tuyến tính hợp lý.
a. Bằng chứng thuyết phục cho kết luận trung bình tổng thể của tỉ lệ chi phí vượt quá 1 %. Thực hiện kiểm định của giả thiết liên quan với mức ý nghĩa .01
b. Quay lại câu (a) , mô tả bối cảnh sai lầm loại I và II và nêu ra những lỗi có thể có trong kết luận đó. Nguồn từ dữ liệu có trong báo cáo sao cho $\mu = 1.33$ cho tổng thể của tất cả 762 các quĩ như vậy. Bạn có phạm sai lầm với quyết định đó không?
c. Gỉa sử rằng $\sigma = .5$ xác định và giải thích lũy thừa của kiểm định trong (a) cho giá trị thực của μ vùng trong (b)

34. Một mẫu gồm 12 máy đo hàm lượng radon của một loại xác định được chọn, và mỗi người tiếp xúc với 100 pCi/L của radon.Ta đọc được kết quả sau:
105.6 90.9 91.2 96.6 96.5 91.3
100.1 105.0 99.6 107.7 103.3 92.4
a.Dữ liệu nào được đề nghị để trung bình tổng thể được đọc dưới điều kiện khác từ 100? vùng và kiểm định giả thiết xấp xỉ dùng $\alpha = .05$.
b. Gỉa sử rằng khoảng để thử nghiệm một giá trị của $\sigma = 7.5$ có trong giả thiết. Cần xác định bao nhiêu giá trị phù hợp bao gồm $\beta = .10$ cho giá trị thay thế $\mu = 95$?

35. cho bất kỳ $\Delta > 0$, khi phân phối của tổng thể là chuẩn và σ đã biết, kiểm định hai bên thỏa $\beta(\mu_0 - \Delta) = \beta(\mu_0 + \Delta)$, vì thế $\beta(\mu')$ đối xứng với μ_0

36. Cho giá trị thay thế μ', Biểu diễn $\beta(\mu') \to 0$ khi $n \to \infty$ cho hoặc kiểm định một bên hoặc kiểm định hai bên z trong trường hợp của phân phối chuẩn của tổng thể với σ đã biết.

9 8.3 KIỂM ĐỊNH LIÊN QUAN ĐẾN TỈ LỆ TỔNG THỂ

Đặt p ký hiệu là tỉ lệ của cá thể hoặc đối tượng trong một tổng thể sở hữu một tỉ lệ có tính chất xác định (ví dụ xe hơi với truyền động bằng tay hay người hút thuốc bằng bộ lọc điếu thuốc). Nếu một cá thể hay đối tượng với tính chất được cho là thành công (S) thì p là tỉ lệ thành công của tổng thể . Kiểm định liên quan đến p bằng cách dựa vào mẫu ngẫu nhiên của cỡ mẫu n từ tổng thể. Cho biết n nhỏ liên quan đến kích thước của tổng thể. X (số của s trong mẫu S) có (xấp xỉ) phân phối nhị thức. Thêm vào đó, Nếu bản thân n lớn $[np \geq 10 \text{và} n(1-p) \geq 10]$, cả X và ước lượng $\hat{p} = X/n$ xấp xỉ phân phối chuẩn. Đầu tiên ta xét kiểm định mẫu lớn dựa trên tình huống này sau đó lần lượt với trường hợp mẫu nhỏ mà dùng trực tiếp phân phối nhị thức.

KIỂM ĐỊNH MẪU LỚN

Kiểm định mẫu lớn liên quan đến p là một trường hợp đặc biệt của quá trình mẫu lớn nói chung cho tham số θ. Đặt $\hat{\theta}$ là ước lượng của θ mà (ít xấp xỉ nhất) không chệch và có xấp xỉ phân phối chuẩn giả thiết không có dạng $H_0 : \theta = \theta_0$ khi θ_0 ký hiệu là một số (giá trị không) thỏa mãn trong trường hợp này. Gỉa sử khi H_0 đúng, độ lệch chuẩn của $\hat{\theta}$, $\sigma_{\hat{\theta}}$ liên quan đến tham số chưa biết. Cho ví dụ, Nếu $\theta = \mu$ và $\hat{\theta} = \overline{X}, \sigma_{\hat{\theta}} = \sigma_{\overline{X}} = \sigma/\sqrt{n}$ liên quan đến tham số chưa biết chỉ khi σ đã biết. Một kiểm định cho mẫu lớn có kết quả từ chuẩn hóa $\hat{\theta}$ với giả thiết H_0 đúng (vì thế $E(\hat{\theta} = \theta_0)$)
Kiểm định thống kê $Z = \frac{\hat{\theta} - \theta_0}{\sigma_{\hat{\theta}}}$

Nếu giả thiết thay thế là $H_a : \theta > \theta_0$ một kiểm định bên phải với mức ý nghĩa xấp xỉ α được qui định bởi vùng bác bỏ $z \geq z_\alpha$. Hai giả thiết thay thế khác , $H_a : \theta < \theta_0$ và $H_a : \theta \neq \theta_0$ tương ứng với kiểm định bên trái và kiểm định hai bên cho z.
Trong trường hợp $\theta = p, \sigma_{\hat{\theta}}$ sẽ không liên quan đến bất kỳ tham số chưa biết nào khi H_0 đúng, đây không phải là trường hợp điển hình. Khi $\sigma_{\hat{\theta}}$ liên quan đến tham số chưa biết, thường thì có thể sử dụng ước lượng có độ lệch chuẩn $S_{\hat{\theta}}$ ở vị trí $\sigma_{\hat{\theta}}$ và Z vẫn có phân phối xấp xỉ chuẩn khi H_0 đúng (vì khi n lớn, $s_{\hat{\theta}} \approx \sigma_{\hat{\theta}}$ với hầu hết mẫu) Kiểm định mẫu lớn cho phần trước được trang bị ví dụ này : Vì σ thường không biết, ta dùng $s_{\hat{\theta}} = s_{\overline{X}} = s/\sqrt{n}$ thay cho $\sigma/\sqrt{n}$ trong mẫu của z.
ước lượng $\hat{p}$ là không chệch $(E(\hat{p}) = p)$, có phân phối xấp xỉ chuẩn, và độ lệch chuẩn của nó là $\sigma_{\hat{p}} = \sqrt{p(1-p)/n}$. Thực tế đã được dùng trong Phần 7.2 để có khoảng tin cậy cho p. Khi H_0 đúng, $E(\hat{p}) = p_0$và $\sigma_{\hat{p}} = \sqrt{p_0(1-p_0)/n}$, vì thế $\sigma_{\hat{p}}$ không liên quan đến bất kỳ tham số chưa biết nào. Ta có kiểm định sau khi n lớn và H_0 đúng

$$Z = \frac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)/n}}$$

có phân phối xấp xỉ chuẩn. Nếu giả thiết thay thế là $H_a : p > p_0$ và dùng vùng bác bỏ phía bên phải $z \geq z_\alpha$, thì

$$P(\text{sai lầm loại I}) = P(H_0 \text{bị bác bỏ khi nó đúng })$$
$$= P(Z \geq z_\alpha \text{ khi Z có phân phối xấp xỉ chuẩn }) \approx \alpha$$

Vì thế mức ý nghĩa mong muốn của α đạt được bằng cách sử dụng giá trị tới hạn mà có trong vùng lấy trên đường cong phía bên phải của α. Vùng bác bỏ cho hai giả thiết thay thế khác, bên trái cho $H_a : p < p_0$ và hai bên cho $H_a : p \neq p_0$, là hợp lý trong những loại như vậy.

Gỉa thiết không: $H_0 : p = p_0$
Gía trị kiểm định thống kê: $z = \frac{\hat{p}-p_0}{\sqrt{p_0(1-p_0)/n}}$

Gỉa thiết thay thế	Vùng bác bỏ
$H_a : p > p_0$	$z \geq z_\alpha$(bên phải)
$H_a : p < p_0$	$z \leq z_\alpha$(bên trái)
$H_a : p \neq p_0$	hoặc $z \geq z_\alpha$ hoặc $z \leq -z_{\alpha/2}$(hai bên)

Qúa trình kiểm định này cung cấp giá trị mà $np_0 \geq 10$ và $n(1-p_0) \geq 10$

VÍ DỤ 8.11 Nút chai làm từ các thành phần tự nhiên trong những chai rượu vang có thể bị hư hỏng, và vì thế rượu vang trong chai có thể bị nhiễm bẩn. Bài viết "ảnh hưởng của cách đóng nút chai với nhận thức của người tiêu dùng về chất lượng của rượu vang" (Amer: J. of Enology and Viticulture, $2007 : 182 - 191$) báo cáo rằng trong quảng cáo về mùi vị của rượu làm từ nho Chardonnay, 16 trong 91 chai bị hư hỏng ở mức độ nào đó liên quan đến đặc điểm nút chai. Dữ liệu này có cung cấp bằng chứng mạnh mẽ nào cho kết luận rằng hơn 15% của tất cả các chai bị nhiễm bẩn cũng bằng con đường này không? Hãy thực hiện kiểm định giả thiết dùng mức ý nghĩa .10

1. p= tỉ lệ đúng của tất cả các chai được quảng cáo làm từ nho Chardonnay bị hư hỏng ở một mức nào đó do liên quan đến nút chai.
2. Gỉa thiết không là $H_0 : p = .15$
3. Gỉa thiết thay thế là $H_a : p > .15$, khẳng định rằng phần trăm tổng thể vượt quá 15%
4. Khi đó $np_0 = 91(.15 = 13.65 > 0)$ và $nq_0 = 91(.85) = 77.35 > 10$, kiểm định cho mẫu lớn z có thể sử dụng. Gía trị kiểm định thống kê là $z = (\hat{p} - .15)/\sqrt{(.15)(.85)/n}$
5. Dạng của H_a suy ra rằng kiểm định bên phải là phù hợp: Bác bỏ H_0 nếu $z \geq z_{.10} = 1.28$.
6. $\hat{p} = 16/91 = .1758$, từ đó

$$z = (.1758 - .15)/\sqrt{(.15)(.85)/91} = .0258/.0374 = .69$$

7. Khi đó $.69 < 1.28$, z không nằm trong vùng bác bỏ. Tại mức ý nghĩa .10 không thể bác bỏ giả thiết không. Mặc dù phần trăm của các chai bị nhiễm bẩn có trong mẫu phần nào vượt quá 15%, tỉ lệ như vậy không đủ lớn để kết luận rằng phần trăm của tổng thể vượt quá 15%. Sự khác biệt giữa tỉ lệ mẫu .1758 và giá trị không .15 có thể giải thích đầy đủ bằng cách lấy mẫu biến thiên.

β và xác định cỡ mẫu
Khi H_0 đúng, Kiểm định thống kê Z có xấp xỉ phân phối chuẩn. Bây giờ ta giả sử rằng H_0 không đúng và $p = p'$. Thì Z vẫn có phân phối xấp xỉ chuẩn (vì đó là một hàm tuyến tính của $\hat{p}$), nhưng giá trị kỳ vọng và phương sai không còn là 0 và 1. thay vào được

$$E(Z) = \frac{p' - p_0}{\sqrt{p_0(1-p_0)/n}} \qquad V(Z) = \frac{p'(1-p')/n}{p_0(1-p_0)/n}$$

Xác suất sai lầm loại II cho kiểm định bên phải là $\beta(p') = P(Z > z_\alpha) \text{khi} p = p'$. Điều này có thể tính được bằng cách sử dụng kỳ vọng và phương sai của biến chuẩn hóa sau đó viện dẫn đến tiêu chuẩn cdf. Thêm vào đó, nếu muốn mức kiểm định α cũng có $\beta(p') = \beta$ cho giá trị xác định của β, phương trình này có thể giải với n cần thiết như trong phần 8.2. Nói chung $\beta(p')$ và n được xác định như sau:

Kiểm định giả thiết	$\beta(p')$
$H_a : p > p_0$	$\Phi\left[\frac{p_0 - p' + z_\alpha\sqrt{p_0(1-p_0)/n}}{\sqrt{p'(1-p')/n}}\right]$

$H_a : p < p_0$ $\quad 1 - \Phi\left[\dfrac{p_0 - p' - z_\alpha\sqrt{p_0(1-p_0)/n}}{\sqrt{p'(1-p')/n}}\right]$

$H_a : p \neq p_0$ $\quad \Phi\left[\dfrac{p_0 - p' + z_{\alpha/2}\sqrt{p_0(1-p_0)/n}}{\sqrt{p'(1-p')/n}}\right]$-$\Phi\left[\dfrac{p_0 - p' - z_{\alpha/2}\sqrt{p_0(1-p_0)/n}}{\sqrt{p'(1-p')/n}}\right]$

Cỡ mẫu n cho mức kiểm định α sao cho thỏa $\beta(p') = \beta$ là

$$n=\begin{cases} \left[\dfrac{z_\alpha\sqrt{p_0(1-p_0)+z_\beta\sqrt{p'(1-p')}}}{p'-p_0}\right]^2 \text{ kiểm định một bên} \\ \left[\dfrac{z_{\alpha/2}\sqrt{p_0(1-p_0)+z_\beta\sqrt{p'(1-p')}}}{p'-p_0}\right]^2 \text{ kiểm định hai bên(Một giải pháp xấp xỉ)} \end{cases}$$

VÍ DỤ 8.12

Dịch vụ giao hàng trọn gói quảng cáo rằng ít nhất 90% các gói hàng được mang đến văn phòng từ lúc 9 giờ sáng để giao trong cùng thành phố và giao vào trưa ngày hôm đó. Đặt p ký hiệu tỉ lệ đúng của những gói hàng được giao như quảng cáo và xét giả thiết $H_0 : p = .9$ ngược lại $H_a : p < .9$. Nếu chỉ 80 % gói hàng được giao như quảng cáo, khả năng cho mức kiểm định .01 sẽ như thế nào ? dựa trên n=255 gói hàng được phát hiện bắt đầu từ H_0 Nên chọn cỡ mẫu bằng bao nhiêu để chắc rằng $\beta(.8) = .01$? với $\alpha = .01, p_0 = .9, p' = .8$ và $n = 255$,

$$n = \left[\frac{2.33\sqrt{(.9)(.1)} + 2.33\sqrt{(.8)(.2)}}{.8 - .9}\right]^2 \approx 266$$

KIỂM ĐỊNH MẪU NHỎ

Qúa trình kiểm định khi cỡ mẫu n nhỏ dựa trực tiếp vào phân phối nhị thức hơn là xấp xỉ chuẩn. Xét giả thiết thay thế $H_a : p > p_0$ và lại đặt X là số thành công trong mẫu . Thì là kiểm định thống kê, và vùng bác bỏ phía bên phải có dạng $x \geq c$ khi H_0 đúng. X có phân phối nhị thức với tham số n và p_0, vì thế
P(sai lầm loại I) = $P(H_0$ bị bác bỏ khi nó đúng)
$= P(X \geq c\text{khi}X \sim Bin(n, p_0))$
$= 1 - P(X \leq c - 1\text{khi } X \sim Bin(n, p_0))$
$= 1 - B(c - 1; n, p_0)$
Khi giá trị tới hạn c giảm , nhiều giá trị X sẽ có trong vùng bác bỏ và P(sai lầm loại I) tăng vì X có phân phối xác suất rời rạc , thường không thể tìm được giá trị c chính xác cho P(sai lầm loại I) với mức ý nghĩa mong muốn α (ví dụ, .05 hay .01) , Thêm vào đó vùng bác bỏ lớn nhất có dạng $\{c, c+1, ..., n\}$ thỏa $1 - B(c - 1 : n, p_0) \leq \alpha$ được sử dụng.
Đặt p' ký hiệu cho giá trị thay thế của $p(p' < p)$. Khi p=p' $X \sim Bin(n, p')$, vì thế

$$\beta(p') = P(\text{sai lầm loại II khi p=p'})$$
$$= P(X < c \text{ khi } X \sim Bin(n, p')) = B(c - 1; n, p')$$

Nghĩa là, $\beta(p')$ là kết quả của việc tính xác suất nhị thức cùng cỡ mẫu n cần thiết để chắc rằng một kiểm định mức α cũng thỏa với βgiá trị thay thế cụ thể p' phải được xác định bởi những thử nghiệm và những sai lầm sử dụng trong nhị thức cdf.
Qúa trình kiểm định cho $H_a : p < p_0$ và cho $H_a : p \neq p_0$ tương tự như cách xây dựng trước, vùng bác bỏ xấp xỉ có dạng $x \leq c$ (một kiểm định bên trái). Gía trị tới hạn c là số lớn nhất thỏa $B(c; n, p_0) \leq \alpha$. Vùng bác bỏ khi giả thiết thay thế là $H_a : p \neq p_0$ bao gồm cả giá trị X lớn và nhỏ.

VÍ DỤ 8.13

Một nhà máy sản xuất nhựa dẻo phát triển một loại thùng nhựa mới và đề nghị bán chúng vô điều kiện với bảo hành 6 năm. Để xem điều này có khả thi về mặt kinh tế không, 20 thùng nguyên mẫu được thử nghiệm nhanh hơn chu kỳ sử dụng 6 năm. Đề xuất bảo hành sẽ được sửa đổi chỉ khi mẫu dữ liệu đề xuất mạnh mẽ rằng ít hơn 90% số các thùng còn dùng được sau 6 năm. Đặt p ký hiệu tỉ lệ của tất cả các thùng tăng tốc thử nghiệm. Gỉa thiết cần kiểm định là $H_0 : p = .9$ ngược lại $H_a : p < .9$. Một quyết định sẽ được đưa ra dựa trên thống kê X, số trong 20 cái còn sử dụng được. Nếu mức ý nghĩa mong muốn là $\alpha = .05$ c phải thỏa $B(c; 20, .9) = .133$. Vùng bác bỏ phù hợp là $x \leq 15$. Nếu kết quả đẩy nhanh quá trình kiểm định trong $x = 14$, H_0 sẽ bị bác bỏ và do đó ủng hộ H_a, yêu cầu đề xuất bảo hành phải được sửa đổi, Xác suất của sai lầm loại II cho giá trị thay thế $p' = .8$ là

$$\begin{aligned}\beta(.8) &= P(H_0 \text{ bị bác bỏ khi} X \sim Bin(20, .8)) \\ &= P(X \geq 16 \text{ khi } X \sim Bin(20, .8)) \\ &= 1 - B(15; 20, .8) = 1 - .370 = .630\end{aligned}$$

Nghĩa là khi $p = .8$, 63% của tất cả mẫu n=20 thùng sẽ có kết quả trong bác bỏ không chính xác H_0, sai lầm xác suất cao vì 20 là cỡ mẫu nhỏ và p'=.8 gần với giá trị không $p_0 = .9$

BÀI TẬP PHẦN 8.3 (37 – 46)

37. Đặc điểm chung của những người thừa cân là chỉ số khối cơ thể ít nhất là 30 [BMI= cân nặng/ bình phương chiều cao, trong đó cân nặng tính bằng kg và chiều cao tính bằng m]. Bài viết "tác động của chứng béo phì lên những bệnh chưa biết và năng suất của công nhân trong dân số công nghiệp" (Annals ofEpidemiology, 2008:8-14) báo cáo rằng trong một mẫu công nhân nữ, 262 người có BMI ít hơn 25 và 159 người có BMI đạt ít nhất là 25 nhưng nhỏ hơn 30 và 120 người có BMI vượt quá 30 . Bằng chứng thuyết phục nào để đưa ra kết luận hơn 20% cá nhân có trong mẫu bị béo phì?
a. Phát biểu và kiểm định giả thiết phù hợp sử dụng vùng bác bỏ với mức ý nghĩa .05?
b. giải thích trong bối cảnh này cho sai lầm loại I và loại II
c. Xác suất của kết luận hơn 20% người bị béo phì bằng bao nhiêu khi % thực tế quan sát là 25%

38. Một nhà máy pin niken hydro chọn ngẫu nhiên 100 tấm niken cho việc kiểm tra pin, chúng có một chu kỳ nhất định và xác định được rằng 14 tấm bị phồng rộp.
a. Điều này có cung cấp bằng chứng thuyết phục cho hơn 10% tất cả những tấm dưới vỉ bị như vậy không? Phát biểu và kiểm định giả thiết phù hợp dùng mức ý nghĩa .05. Để đạt được kết luận của bạn, loại sai lầm nào có thể xảy ra cho cam kết này?
b. Nếu thực sự có trường hợp 15% những tấm bị phồng rộp và sử dụng cỡ mẫu 100 thì giả thiết không trong câu (a) sẽ thế nào để không bị bác bỏ với mức ý nghĩa .05? Trả lời câu hỏi này với cỡ mẫu 200.
c. Phải kiểm tra bao nhiêu tấm để có $\beta(.15) = .10$ kiểm định cho câu (a)

39. Một mẫu ngẫu nhiên gồm 150 mẫu máu tình nguyện viên cho ngân hàng máu cho thấy 82 trường hợp có nhóm máu A. Điều này có gợi ý rằng % thực sự của nhóm máu A được quyên góp là 40% không? Phần trăm của tổng thể có nhóm máu A là bao nhiêu? Thực hiện kiểm định giả thiết thích hợp dùng mức ý nghĩa .01 sẽ cho kết luận khác thế nào với việc sử dụng mức ý nghĩa .05?

40. Biết rằng 2/3 con người bị chi phối bởi chân và mắt phải, điều này cũng có nghĩa là

có xu hướng bên phải về hành vi hơn . Bài viết "Hành vi của người trưởng thành đối với việc bất đối xứng đầu" (Nature,2003 : 771) báo cáo rằng một mẫu ngẫu nhiên gồm 24 cặp hôn nhau, cả hai người trong 80 cặp đôi có xu hướng nghiêng về bên phải hơn bên trái.
a. Nếu 2/3 của tất cả các cặp đôi hôn nhau thể hiện hành vi nghiêng về bên phải, Xác suất nào mà trong số 124 người khác từ giá trị kỳ vọng tại đó ít hơn nhiều như giá trị quan sát được?
b. Kết quả thử nghiệm cho thấy rằng 2/3 cho hành vi hôn là không khả thi? Phát biểu và kiểm định giả thiết thích hợp.

41.Bài viết tham khảo trong Ví dụ 8.11 cũng báo cáo rằng trong một mẫu 106 người dùng rượu vang , 22 (20.8%) nghĩ rằng đinh vít là sự thay thế có thể chấp nhận được cho nút chai tự nhiên. Gỉa sử một nhà máy rượu đặc biệt quyết định dùng đinh vít cho một trong những loại rượu vang của họ trừ khi có bằng chứng mạnh mẽ để đề nghị rằng ít hơn 25% người dùng rượu vang chấp nhận.
a. Dùng mức ý nghĩa .10, Bạn muốn giới thiệu cho nhà máy rượu những gì?
b. Cho kiểm định giả thiết trong câu (a) mô tả sai lầm loại I và II trong bối cảnh đó là gì, và sai lầm nào cam kết sẽ xảy ra?

42. Với nguồn cung vật liệu xây dựng trong nước bị chậm lại từ vài năm nay. Gần 60,000 ngôi nhà được xây dựng với tường thạch cao được nhập khẩu từ trung quốc. Theo bài viết "Báo cáo nguồn thạch cao từ Trung Quốc cho vấn đề nhà ở"(New York Times, Nov. $24, 2009$) Cục điều tra liên bang xác định có mối liên hệ giữa hóa chất trong thạch cao với điện năng và cũng có bằng chứng mạnh mẽ về khó khăn trong hô hấp do sự phát tán khí hydro sulfit. Một cuộc điều tra mở rộng gồm 51 ngôi nhà tìm ra 41 ngôi nhà có vấn đề. Gỉa sử 51 mẫu lấy ngẫu nhiên của tổng thể của tất cả nhà có sử dụng thạch cao trung quốc
a. Dữ liệu cung cấp bằng chứng gì cho kết luận hơn 50% tất cả nhà có dùng thạch cao trung quốc có vấn đề về điện hay môi trường? Kiểm định giả thiết kèm theo với $\alpha = .01$?
b. Tính chặn dưới của khoảng tin cậy sử dụng độ tin cậy 99% cho phần trăm của tất cả những nhà có vấn đề về điện hay môi trường
c. Nếu trường hợp thực tế 80% tất cả nhà có vấn đề. Kiểm định trong câu (a) có khả năng thế nào để không có kết luận hơn 50% sẽ xảy ra?

43. Một dự án cho câu lạc bộ du lịch được phát triển bởi 1 hãng hàng không dựa trên giả thiết rằng 5% du khách cũ đủ tiêu chuẩn là thành viên. Một mẫu ngẫu nhiên gồm 500 khách hàng trong đó có 40 người đủ điều kiện
a. Dùng dữ liệu này để kiểm định tại mức .01 giả thiết không của công ty là đúng so với giả thiết thay thế là không đúng
b. Xác suất bằng bao nhiêu khi kiểm định cho câu (a) giả thiết của công ty sẽ được đánh giá đúng khi thực tế 10% của tất cả khách hàng cũ đủ điều kiện

44. Với mỗi nhóm gồm 20 cầu thủ quần vợt được cho 2 vợt trung gian 1 có dây nylon và 1 có sợi tổng hợp khác. sau vài tuần chơi bằng 2 vợt mỗi cầu thủ sẽ được hỏi và nêu rõ sở thích cho mỗi loại trong 2 loại dây đó. Đặt p ký hiệu tỉ lệ tất cả những cầu thủ yêu thích sợi nylon và đặt X là số những cầu thủ trong mẫu yêu thích sợi nylon. Vì sợi đó khá mắc, xét giả thiết không với ít nhất 50% các cầu thủ yêu thích sợi nylon. Ta đơn giản hóa điều này để $H_{)} : p = .5$ dự định từ chối H_0 nếu có bằng chứng mạnh mẽ về tỉ lệ yêu thích sợi nylon.
a. Vùng bác bỏ nào$\{15, 16, 1718, 19, 20\}$ $\{0, 1, 2, 3, 4, 5\}$, hay $\{0, 1, 2, 3, 17, 18, 19, 20\}$ là phù hợp nhất và tại sao 2 vùng còn lại không phù hợp
b. Xác suất sai lầm loại I cho vùng đã chọn trong câu (a). Vùng nào thỏa mức kiểm định .05? Mức kiểm định tốt nhất cho .05 là gì?
c. Nếu có 60% của những người thích sợi đó. Tính xác suất của sai lầm loại II. Sử dụng vùng

xác suất từ câu (a) làm lại với 80%của những người thích sợi đó.
d. Nếu 13 cầu thủ trong số 20 cầu thủ yêu thích loại sợi đó, có nên bác bỏ H_0 dùng mức ý nghĩa .10

45. Một nhà sản xuất đường ống nước phát triển một loại mới của khóa vòi nước. Đặt p=P(Chọn ngẫu nhiên vòi nước của một loại sẽ phát triển với thời gian sử dụng trong thường xuyên trong 2 năm bị rò rỉ) Nhà máy có quyết định tiến hành sản xuất trừ khi nó có thể xác định rằng p có ranh giới khá lớn với việc chấp nhận giá trị p được xác định là 10. Nhà máy quyết định dùng n van nước đã tăng tốc kiểm định (sử dụng thường xuyên khoảng 2 năm) với X= Số các van nước bị rò trước khi có quyết định kiểm định bắt đầu cho sản xuất. trừ khi quan sát X là khá lớn . Điều đó khẳng định rằng nếu p=.10 xác suất không tiến hành sản xuất tối đa là .10. Ngược lại nếu p=.30 xác suất tiến hành sản xuất ít nhất là .10, có thể sử dụng $n = 10$? n=20, n=25? Vùng bác bỏ phù hợp là gì để lựa chọn n và xác suất sai lầm là gì khi sử dụng vùng này.

46. Nhà khoa học nghĩ rằng robot sẽ đóng một vai trò quan trọng trong các nhà máy trong vài thập kỷ tới. Gỉa sử rằng trong một thử nghiệm để xác định liệu việc sử dụng robot để đan cáp máy tính có khả thi không, 1 robot được dùng để đan 500 cáp. cáp được kiểm tra và có 15 cái có khiếm khuyết. Nếu là con người đan thì có khiếm khuyết với tỉ lệ .035(3.5%). Dữ liệu này có hỗ trợ cho giả thiết tỉ lệ khiếm khuyết của robot ít hơn con người dùng mức ý nghĩa .01

10 8.4 Gía trị P

Dùng phương pháp vùng bác bỏ kiểm định cho lựa chọn mức ý nghĩa α. Theo đó tính giá trị của kiểm định thống kê, Gỉa thiết không H_0 bị bác bỏ nếu có giá trị sai trong vùng bác bỏ ngược lại không bác bỏ H_0. Bây giờ ta xét cách khác để có kết luận trong phân tích kiểm định giả thiết. Cách thay thế này dựa trên việc tính toán một xác suất chắc chắn gọi là giá trị P. Một cách tự nhiên ta thấy giá trị P cung cấp một độ đo với chiều dài của bằng chứng trong dữ liệu trước đó của H_0.

ĐỊNH NGHĨA

Gía trị P là một xác suất, được tính bằng cách sử dụng giả thiết không đúng, nhận một giá trị của kiểm định thống kê ít nhất là mâu thuẫn với H_0 như các giá trị được tính từ các mẫu có sẵn.
Định nghĩa này khá ngắn gọn. Ta cần xem những điểm chính sau
. Gía trị P là một xác suất
. Xác suất này tính được bằng cách giả sử H_0 đúng
. Cẩn thận với: Khi H_0 đúng giá trị P không phải là xác suất, cũng không là sai lầm trong xác suất
. Xác định giá trị P,đầu tiên ta phải quyết định giá trị nào của kiểm định thống kê là ít mâu thuẫn với H_0 tính được từ mẫu có sẵn.
VÍ DỤ 8.14
Nước mưa ở đô thị có thể bị ô nhiễm bởi nhiều nguồn, bao gồm cả pin bị bỏ khi nó vỡ ra, những loại pin này giải phóng kim loại vào môi trường . Bài viết "Rác thải đô thị " (J. of Environ. Engr., 2009:46-57) tóm tắt dữ liệu hiện tại cho đặc tính của nhiều loại pin được tìm thấy ở quanh khu đô thị Cleveland. Một mẫu gồm 51 pin Panasonic AAA cho trung bình mẫu của lượng kẽm là 2.06 . Dữ liệu cung cấp này có làm bằng chứng cho kết luận rằng trung bình tổng thể có lượng kẽm vượt quá 2.0 g?

Với μ ký hiệu trung bình đúng khối lượng kẽm cho những loại Pin, giả thiết thích hợp là $H_0 : \mu = 2.0$ ngược lại $H_a : \mu > 2.0$, cỡ mẫu đủ lớn để kiểm định z có thể dùng bất kỳ giả định cụ thể nào về hình dáng phân phối tổng thể. Gía trị kiểm định thống kê là

$$z = \frac{\overline{x} - 2.0}{s/\sqrt{n}} = \frac{2.06 - 2.0}{1.41/\sqrt{51}} = 3.04$$

Bây giờ ta phải tìm giá trị nào của z mà ít mâu thuẫn với H_0. Đầu tiên ta xét bài toán sau: Gía trị nào của $\overline{x}$ mà ít mâu thuẫn với giả thiết không như 2.06. Trung bình mẫu quan sát của ta bằng bao nhiêu? vì dấu $>$ xuất hiện trong H_a nên rõ ràng 2.0 ít mâu thuẫn với H_0 như 2.06, và vì thế thực tế là bất kỳ giá trị $\overline{x}$ nào vượt quá 2.06 sẽ tương ứng với giá trị z mà vượt quá 3.04, vì thế giá trị P là
Gía trị P=$P(Z \geq 3.04 \text{khi } \mu = 2.0)$
Khi đó kiểm định thống kê Z được tạo ra bằng cách trừ đi 2.0 trong phần tử số của giá trị không, khi $\mu = 2.0$ nghĩa là khi H_0 đúng- Z có phân phối xấp xỉ chuẩn như kết quả sau
Gía trị P= $P(Z \geq 3.04$ Khi $\mu = 2.0) \approx$ vùng dưới của đường cong z về phía bên phải của $3.04 = 1 - \Phi(3.04) = .0012$
Ta sẽ minh họa ngắn gọn việc xác định giá trị P cho z bất kỳ hay kiểm định t nghĩa là bất kỳ kiểm định nào khi tham khảo phân phối chuẩn (và đường cong t phù hợp) cho tới thời điểm này, việc lặp lại để đưa ra kết luận tìm được một giá trị P có nghĩa . Vì nó là xác suất, nên giá trị P nằm giữa 0 và 1 .Loại giá trị P nào cung cấp bằng chứng cho việc lặp lại giả thiết không? Xét 2 trường hợp cụ thể.
. Gía trị P=.250 trong trường hợp này toàn bộ 25% của tất cả giá trị có thể của kiểm định giả thiết ít mâu thuẫn với H_0 một trong số đó có trong mẫu của chúng ta vì thế không phải là tất cả dữ liệu đều mâu thuẫn với H_0
. Gía trị P=.0018, ở đây chỉ .18% (ít hơn 1%) của tất cả giá trị kiểm định thống kê có thể ít mâu thuẫn với H_0 như những gì ta có được vì thế xuất hiện mẫu có mâu thuẫn cao hơn với giả thiết không.
Thêm vào đó, giá trị P nhỏ hơn, có nhiều bằng chứng trong dữ liệu mẫu lặp lại trong giả thiết không và cả trong giả thiết thay thế nghĩa là nên bác bỏ H_0 và chấp nhận H_a khi giá trị P đủ nhỏ. vậy "đủ nhỏ " là gì?
Qui tắc quyết định trên giá trị P là gì?
Chọn mức ý nghĩa α (như trước đó, muốn có xác suất sai lầm loại I)
Thì

$$\text{Bác bỏ} H_0 \text{nếu giá trị P} \leq \alpha$$

$$\text{không bác bỏ} H_0 \text{nếu giá trị P} > \alpha$$

Do đó nếu giá trị P vượt quá lựa chọn mức ý nghĩa , giả thiết không không thể bác bỏ tại mức đó, nhưng nếu giá trị P bằng hoặc nhỏ hơn α thì có đủ bằng chứng để chấp nhận H_0. Trong Ví dụ 8.14 ta tính giá trị P=.0012 thì phải sử dụng mức ý nghĩa .01. Ta sẽ bác bỏ giả thiết không và chấp nhận giả thiết thay thế vì $.0012 < .01$. Tuy nhiên, giả sử ta chọn mức ý nghĩa là .001 thì khi đó phải có nhiều bằng chứng hơn từ dữ liệu trước khi H_0 có thể bị bác bỏ. Trong trường hợp này ta không nên bác bỏ H_0 vì $.0012 > .001$
Làm thế nào để so sánh quy tắc quyết định dựa trên giá trị P với quy tắc quyết định tiếp cận thông qua vùng bác bỏ? Hai quá trình của phương pháp vùng bác bỏ và phương pháp giá trị P trên thực tế là như nhau. Bất kể kết luận đạt được bằng cách nào thì vùng bác bỏ với giá trị α cụ thể cũng là kết luận với việc sử dụng giá trị P cũng sử dụng giá trị α đó
VÍ DỤ 8.15

Vấn đề về lượng nicotin đã thảo luận trong ví dụ 8.5 liên quan đến kiểm định $H_0 : \mu = 1.5$

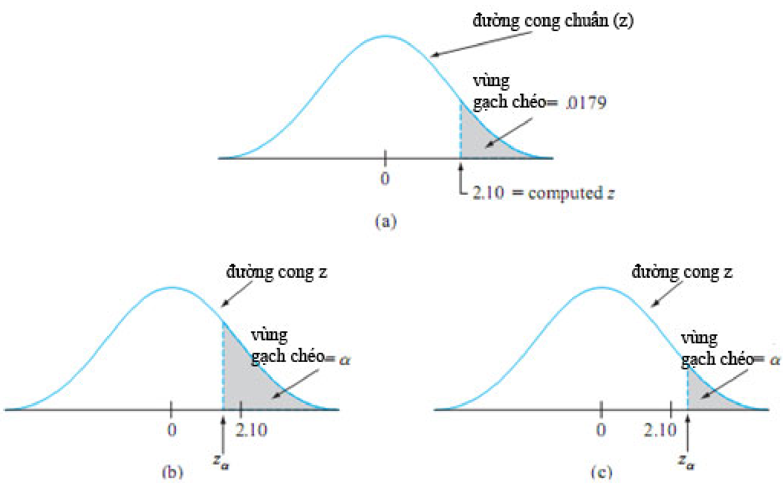
ngược lại $H_a : \mu > 1.5$ dùng kiểm định z (nghĩa là một kiểm định mà sử dụng đường cong z cũng như tham khảo phân phối của nó) Bất đẳng thức trong H_a suy ra rằng vùng bác bỏ phía bên phải $z \geq z_\alpha$ là phù hợp. Gỉa sử z=2.10 thì sử dụng lý do chính xác như trong Ví dụ 8.14 được giá trị P $=1-\Phi(2.10) = .0179$. Bây giờ ta xét những kiểm định với những mức ý nghĩa khác nhau

$$\alpha = .10 \Rightarrow z_\alpha = z_{.10} = 1.28 \Rightarrow 2.10 \geq 1.28 \Rightarrow \text{bác bỏ} H_0$$

$$\alpha = .05 \Rightarrow z_\alpha = z_{.05} = 1.645 \Rightarrow 2.10 \geq 1.645 \Rightarrow \text{bác bỏ} H_0$$

$$\alpha = .01 \Rightarrow z_\alpha = z_{.01} = 2.33 \Rightarrow 2.10 < 2.33 \Rightarrow \text{không bác bỏ} H_0$$

t Vì giá trị P= $.0179 \leq .10$ và $.0179 \leq .05$ Sử dụng kết quả giá trị P trong bác bỏ H_0 cho hai mức ý nghĩa. Tuy nhiên, cho $\alpha = .01, 2.10$ không nằm trong vùng bác bỏ và .0179 thì lớn hơn .01. Hơn nữa, bất cứ khi nào α nhỏ hơn giá trị P .0179, giá trị tới hạn z_α sẽ vượt quá giá trị tính được của z và vì thế không thể bác bỏ H_0 với cả hai phương pháp. Điều này được minh họa trong hình 8.7



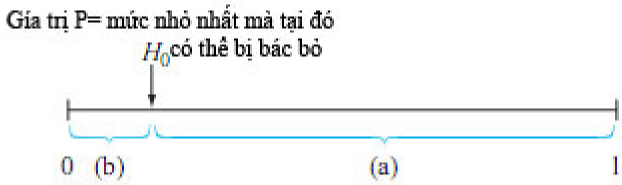
Hình 8.7 Mối quan hệ giữa α và hình ảnh vùng đuôi bằng cách tính z; (a) Hình ảnh vùng đuôi bằng cách tính z; (b) khi $\alpha < .0179, z_\alpha < 2.10$ và H_0 bị bác bỏ; (c) khi $\alpha < .0179, z_\alpha > 2.10$ và H_0 không bị bác bỏ

Xét giá trị P .0012 trong Ví dụ 8.14, H_0 có thể bị bác bỏ nếu $.0012 < \alpha$. Do đó giá trị không có thể bị bác bỏ nếu $\alpha = .05$ hay .01 hay .005 hay .0015 hay .00125. Mức ý nghĩa nào nhỏ nhất ở đây mà H_0 có thể bị bác bỏ.? Đó là giá trị P .0012

MỆNH ĐỀ

Gía trị P là mức ý nghĩa α nhỏ nhất mà tại đó giá trị không có thể bị bác bỏ. Vì vậy giá trị P được gọi là mức ý nghĩa quan sát được (OSL) cho dữ liệu.

Thông thường dữ liệu gọi là ý nghĩa khi đó H_0 bị bác bỏ và ngược lại là không có nghĩa. Gía trị P là mức nhỏ nhất tại dữ liệu là ý nghĩa. Một cách dễ dàng để hình dung là so sánh giá trị P với việc chọn α là vẽ một hình như Hình 8.8. để tính được giá trị P phụ thuộc vào những kiểm định phải, trái, hay hai bên. Tuy nhiên, nên so sánh giữa 1 cái đã tính với α không phụ thuộc vào thử nghiệm đã sử dụng



Hình 8.8 So sánh α và giá trị P: (a) bác bỏ H_0 khi α nằm đó; (b) không bác bỏ H_0 khi α nằm đó

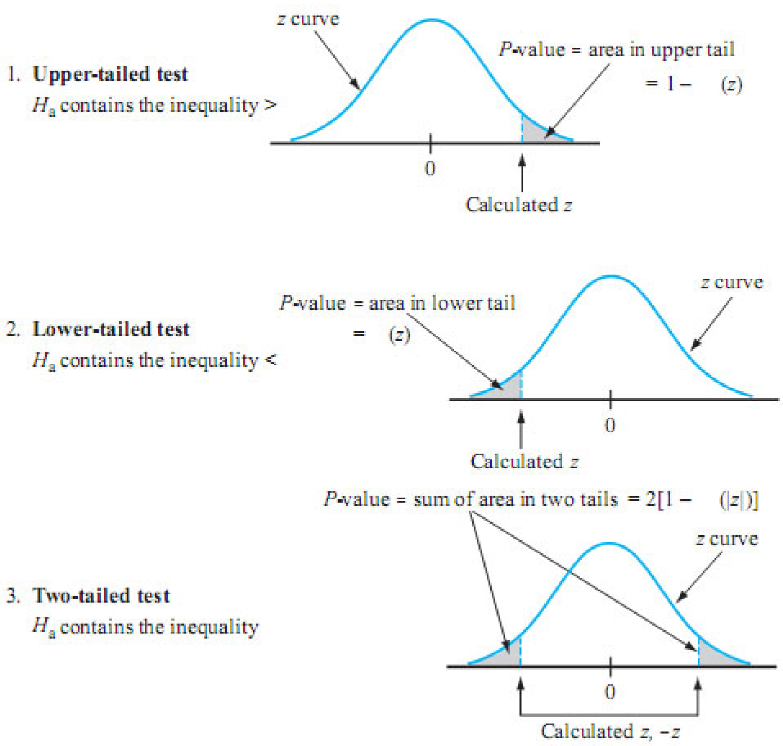
VÍ DỤ 8.16

Trung bình đúng thời gian cho thuốc giảm đau loại được bán chạy nhất được biết là 10 phút. Đặt μ ký hiệu trung bình đúng thời gian cho một loại giảm đau mới của công ty. Gỉa sử rằng khi dữ liệu từ phân tích một thử nghiệm liên quan đến thời gian giảm đau, giá trị P cho kiểm định $H_0 : \mu = 10$ ngược lại $H_a : \mu < 10$ tính được là .0384. Khi đó $\alpha = .05$ lớn hơn giá trị P [.05 nằm trong khoảng (a) của hình 8.8]. H_0 sẽ bị bác bỏ với bất kỳ thử nghiệm nào tại mức .05. Tuy nhiên tại mức .01 H_0 cũng sẽ bị bác bỏ vì .01 nhỏ hơn mức nhỏ nhất (.0384) tại đó H_0 có thể bị bác bỏ.
Phần mềm thống kê được sử dụng rộng rãi nhất trên máy tính đưa vào giá trị P khi thực hiện phân tích một kiểm định giả thiết. Một kết luận có thể vẽ được trực tiếp từ dữ liệu xuất ra, ngoài việc tham khảo bảng của giá trị tới hạn. Có giá trị P trong tay, điều tra viên sẽ nhìn thấy ngay mức ý nghĩa của riêng anh ta hay cô ta. Thêm vào đó, biết được giá trị P cho phép đưa ra quyết định phân biệt giữa cách gọi ngắn gọn (ví dụ $\alpha = .05$,giá trị P=.0498) và có kết luận rõ ràng (ví dụ $\alpha = .05$, giá trị P=.0003) có thể đưa ra phát biểu "H_0 có thể bị bác bỏ tại mức ý nghĩa .05"

GIÁ TRỊ P CHO KIỂM ĐỊNH z

Gía trị P cho kiểm định z (dựa trên kiểm định thống kê có phân phối khi H_0 đúng ít nhất có phân phối xấp xỉ chuẩn) Dễ dàng xác định từ thông tin có trong bảng Phụ lục A.3. Xét một kiểm định bên phải và đặt z là giá trị được tính từ kiểm định thống kê z. giả thiết không bị bác bỏ nếu $z \geq z_\alpha$ và giá trị P là nhỏ nhất của α trong trường hợp này. Khi đó z_α tăng và α giảm. Gía trị P là giá trị của α khi cho $z = z_\alpha$ khi cho $z = z_\alpha$, nghĩa là giá trị P chỉ cần xét trong vùng mà giá trị tính được z nằm ở phía bên phải của đường cong chuẩn. Khi đó vùng tương ứng là $P = 1 - \Phi(z)$
Lập luận tương tự cho kiểm định bên trái biểu diễn cho giá trị P là xét vùng mà giá trị tính được z nằm phía bên trái của đường cong chuẩn.
Cẩn thận với bài toán trong kiểm định hai bên. Gỉa sử đầu tiên xét z là số dương thì giá trị P là giá trị α thỏa $z = z_{\alpha/2}$ (nghĩa là tính z bằng giá trị tới hạn bên phải). Điều này nói lên rằng xét vùng phía bên phải là một nửa của giá trị P, vì thế giá trị P= $2[1 - \Phi(z)]$. Nếu z là số âm, giá trị P là α khi cho $z = -z_{\alpha/2}$, hay, tương đương với $-z = z_{\alpha/2}$,vì thế giá trị P = $2[1 - \Phi(-z)]$. Khi đó $-z = |z|$ khi z là số âm, giá trị P = $2[1 - \Phi(|z|)]$ cho cả z âm và z dương.

Gía trị P: P=$\begin{cases} 1 - \Phi(z) & \text{cho một kiểm định bên phải của z} \\ \Phi(z) & \text{hay một kiểm định bên phải của z} \\ 2[1 - \Phi(|z|)] & \text{một kiểm định hai bên của z} \end{cases}$ Mỗi xác suất sẽ nhận được ít nhất một giá trị cực biên (Gỉa sử H_0 đúng) .Ba trường hợp này được minh họa trong hình 8.9
Ví dụ tiếp theo minh họa cách sử dụng giá trị P với kiểm định giả thiết bằng kỳ vọng với những thứ tự xác định từ thứ tự đã có.



Hình 8.9 Xác định giá trị P cho kiểm định z Ví dụ 8.17

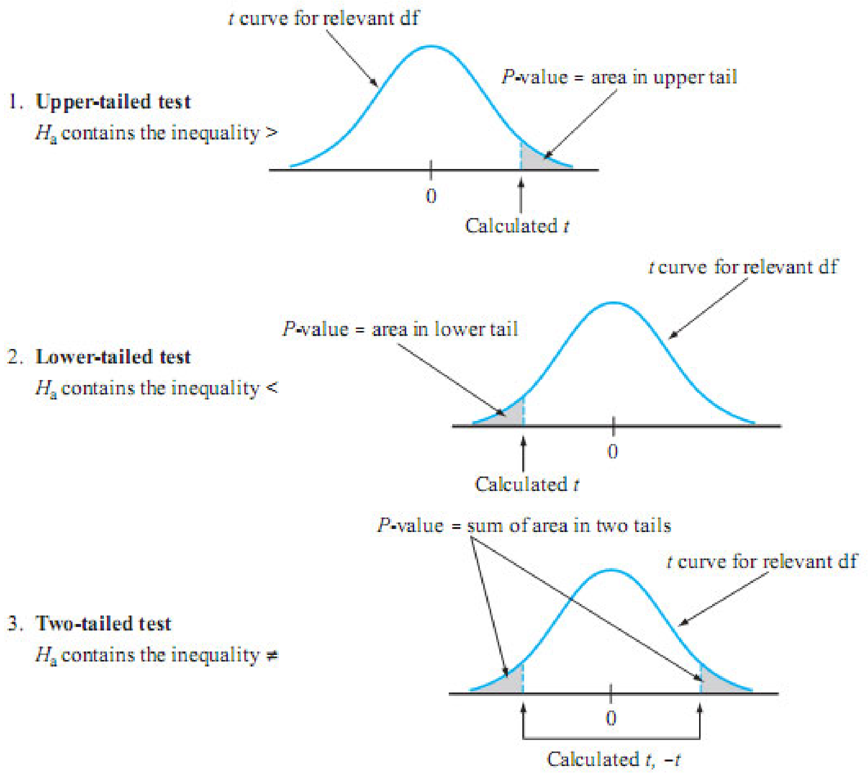
Mục tiêu của độ dày cho các tấm silicon được sử dụng trong một loại mạch tích hợp nhất định là 245 μm. Một mẫu gồm 50 tấm và mỗi tấm có một độ dày xác định , kết quả trung bình mẫu của độ dày là 246.18 μm và độ lệch chuẩn 3.60μm. Dữ liệu này có thấy Trung bình độ dày của tấm khác với giá trị mục tiêu không?
1. Tham số quan tâm: $\mu =$ trung bình đúng độ dày tấm
2. Gỉa thiết không : $H_0 : \mu = 245$
3. Gỉa thiết thay thế : $H_a : \mu \neq 245$
4. Công thức cho giá trị kiểm định thống kê: $z = \dfrac{\overline{x} - 245}{s/\sqrt{n}}$
5. Tính giá trị kiểm định thống kê: $z = \dfrac{246.18 - 245}{3.60/\sqrt{50}} = 2.32$
6. Xác định giá trị P: Vì là kiểm định hai bên nên,

$$\text{Gía trị P} = 2(1 - \Phi(2.31)) = .0204$$

7. Kết luận Sử dụng mức ý nghĩa .01, H_0 sẽ không bị bác bỏ khi .0204 > .01. Tại mức ý nghĩa này, không đủ bằng chứng để kết luận rằng Trung bình đúng độ dày khác với giá trị mục tiêu.

Gía trị P cho kiểm định t

Cũng giống như giá trị P cho kiểm định z là vùng đường cong z, giá trị P cho kiểm định t sẽ là vùng đường cong t. Hình 8.10 sẽ minh họa 3 trường hợp khác nhau đó. số bậc tự do df cho kiểm định một mẫu là n-1



Hình 8.10 Gía trị P cho kiểm định t

Bảng của giá trị tới hạn t dùng ở phần trước cho khoảng tin cậy và dự đoán sẽ không có đủ thông tin về bất kỳ phân phối t cụ thể nào để chấp nhận cho việc xác định chính xác vùng mong muốn vì thế ta có bảng gồm những giá trị t khác nhau trong bảng Phụ lục A.8, trong đó có cả vùng bên phải của đường cong t. Mỗi cột khác nhau của của bảng cho một số df khác nhau và những dòng này dùng để tính giá trị của kiểm định thống kê t được sắp xếp từ 0.0 đến 4.0 tăng dần đến 1. Cho ví dụ, số 074 xuất hiện tại giao của dòng tại 1.6 và cột tại 8 df, vì thế vùng dưới đường cong 8 df về phía bên phải của 1.6(một vùng bên phải) là .074 vì đường cong t đối xứng nên .074 cũng có vùng dưới đường cong về phía bên trái của -1.6 (vùng dưới bên trái)
Gỉa sử, cho Ví dụ, có một kiểm định $H_0 : \mu = 100$ ngược lại $H_a : \mu > 100$ dựa vào phân phối t với 8 df. Nếu tính được giá trị thống kê t là 1.6 thì giá trị P cho kiểm định bên phải là .074 vì .074 vượt quá .05 nên ta không thể bác bỏ H_0 tại mức ý nghĩa .05. Nếu giả thiết thay thế là $H_a : \mu < 100$ và kiểm định dựa trên vùng 20 df với $t = -3.2$, Thì bảng phụ lục A.8 biểu diễn rằng giá trị P có trong vùng bên trái .002. Gỉa thiết không có thể bị bác bỏ tại mức .05 hoặc .01. Xét kiểm định $H_0 : \mu_1 - \mu_2 = 0$ ngược lại $H_a : \mu_1 - \mu_2 \neq 0$ giả thiết không nói rằng trung bình của 2 tổng thể được xác định trong khi giả thiết thay thế nói rằng chúng khác nhau mà không thể xác định được H_0. Nếu một kiểm định t dựa trên 20 df và t=3.2 thì giá trị P cho kiểm định hai bên là 2(.002) = .004. Đây cũng là giá trị P với t=-3.2 vùng đuôi được gấp đôi lên vì giá trị của cả hai lớn hơn 3.2 và nhỏ hơn -3.2 điều này mâu thuẫn với H_0 đã tính (giá trị được sinh ra từ đuôi của đường cong t)

VÍ DỤ 8.18

Trong ví dụ 8.9 ta xét kiểm định $H_0 : \mu = 4$ ngược lại $H_a : \mu \neq 4$ dựa vào mẫu n=5 quan sát từ một tổng thể có phân phối chuẩn. Gía trị kiểm định thống kê là $-.594 \approx -.6$. Nhìn vào cột 4 $(= 5 - 1)$ df của bảng phụ lục A.8 và nhìn xuống hàng 6 ta nhận được giá trị .290 vì kiểm định hai bên vùng bên phải phải gấp đôi bao gồm cả giá trị P, kết quả ta được giá trị

$P \approx .580$. Gía trị P này rõ ràng lớn hơn bất kỳ mức ý nghĩa α nào (.01, .05, hay thậm chí .10), vì thế không có lý do gì để bác bỏ giả thiết không. Phần mềm Minitab trong ví dụ 8.9 cho giá trị P=.594. Gía trị P từ phần mềm này sẽ cho kết quả chính xác hơn trong bảng phụ lục A.8 vì giá trị của t trong bảng chỉ đúng đến chữ số 10.

GIẢI THÍCH THÊM VỀ GIÁ TRỊ P

Kết quả của giá trị P từ việc thực hiện kiểm định trên một mẫu đã chọn không phải là xác suất để H_0 đúng, nó cũng không là xác suất của việc bác bỏ H_0 . Nhắc lại, nó là xác suất để tính toán cho giả thiết H_0 đúng kể cả những giá trị kiểm định thống kê ít mâu thuẫn với giả thiết không hơn là kết quả giá trị thực. Cho ví dụ, xét kiểm định $H_0 : \mu = 50$ ngược lại $H_a : \mu < 50$. sử dụng kiểm định bên trái cho z nếu tính được giá trị kiểm định thống kê là z=-2.00, thì

$$\text{giá trị} P = P(Z < -2.00 \text{ khi } \mu = 50)$$

=Vùng dưới của đường cong z về phía bên phải của -2.00=0.228
Nhưng nếu mẫu thứ hai được chọn, kết quả sẽ cho giá trị của z chắc chắn sẽ khác -2.00, vì thế giá trị P phù hợp có khả năng sẽ khác .0228 vì giá trị kiểm định thống kê bản thân nó thay đổi từ mẫu này đến mẫu khác, giá trị P cũng thay đổi như vậy, nghĩa là, kiểm định thống kê là biến ngẫu nhiên vì thế giá trị P cũng là biến ngẫu nhiên. Mẫu đầu tiên có thể cho giá trị P là .0228, mẫu thứ hai có thể cho giá trị P là .1175, mẫu thứ ba có thể là .0606,....
Nếu H_0 sai, ta hy vọng giá trị P sẽ dần về 0, vì thế giả thiết không có thể bị bác bỏ. Mặt khác, khi H_0 đúng, ta muốn giá trị P vượt quá mức ý nghĩa đã chọn vì vậy quyết định đúng đắn là không bác bỏ H_0. Ví dụ tiếp theo sẽ trình bày mô phỏng giá trị P sẽ có hình dáng thế nào trong cả hai trường hợp khi giả thiết không đúng và khi giả thiết không sai.

VÍ DỤ 8.19

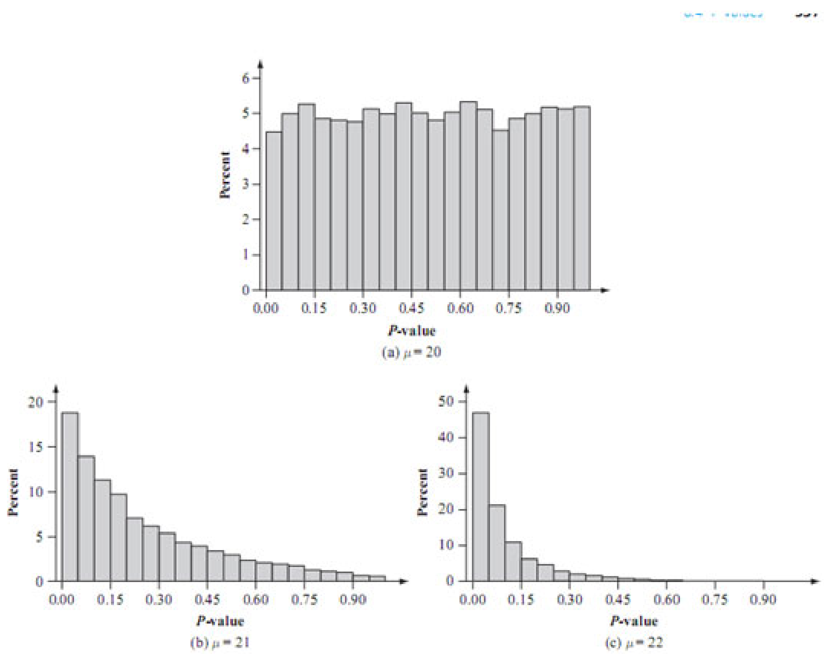
Tiêu hao nhiên liệu của bất kỳ loại xe mới nào trong điều kiện lái xe đúng qui định có thể không giống với EPA xuất hiện trên nhãn xe. Gỉa sử 4 xe khác nhau của 1 loại nhất định được chọn, tài xế phải trải qua một khóa học nhất định, sau đó xác định tiêu hao nhiên liệu của mỗi xe. Đặt μ ký hiệu là trung bình đúng của tiêu hao nhiên liệu dưới những điều kiện đó. Xét kiểm định $H_0 : \mu = 20$ ngược lại $H_a : \mu > 20$ sử dụng kiểm định một mẫu t dựa vào kết quả mẫu. Khi đó, kiểm định dựa trên n-1=3 df, giá trị P cho kiểm định bên phải là vùng dưới đường cong t với 3 df về phía bên phải của giá trị tính được t.
Đầu tiên giả sử rằng giả thiết không là đúng, ta dùng Minitab cho 10,000 mẫu khác nhau, mỗi cái có 4 quan sát từ phân phối chuẩn của tổng thể với giá trị kỳ vọng $\mu = 20$ và độ lệch chuẩn $\sigma = 2$kết quả tóm tắt cho mẫu đầu tiên là

$$x_1 = 20.803, x_2 = 22.232, x_3 = 20.276, x_4 = 17.718$$
$$\overline{x} = 20.264, s = 1.8864, t = \frac{20.264 - 20}{.1.8864/\sqrt{4}} = .2799$$

Gía trị P là vùng dưới của đường cong t về phía bên phải của .2799, dùng phần mềm Minitab được .3989. Dùng mức ý nghĩa .05, giả thiết không sẽ không bị bác bỏ . Gía trị của t cho 4 mẫu khác nhau là -1.5791, .6082, -.7020, và 3.1053, với giá trị P thỏa mãn .912,.293,.773, và .0265. Hình 8.11(a) biểu diễn đồ thị của 10,000 giá trị P từ thực nghiệm, mô phỏng với 4.5% của giá trị P là lớp đầu tiên của khoảng từ 0 đến .05. Do đó, khi sử dụng mức ý nghĩa .05 thì giả thiết không sẽ bị bác bỏ trong khoảng 4.5% của 10,000 kiểm định đó. Nếu ta tiếp tục với mẫu tổng quát và thực hiện kiểm định mỗi mẫu với mức ý nghĩa .05 về lâu dài 5% của giá trị P sẽ nằm

trong lớp đầu tiên của của khoảng. Vì thế, khi H_0 đúng nếu kiểm định với mức ý nghĩa .05, theo định nghĩa ta có xác suất của bác bỏ H_0 là .05
Xem trên biểu đồ, ta thấy, phân phối của giá trị P tương đối phẳng. Thực tế, nó có thể biểu diễn được khi H_0 đúng, phân phối xác suất của giá trị P là phân phối đồng dạng trên khoảng từ 0 đến 1 nghĩa là đường cong mật độ phẳng hoàn toàn trên khoảng này do đó độ cao của nó phải bằng 1 nếu tổng vùng dưới của đường cong là 1. Khi đó vùng dưới của đường cong về phía bên trái của .05 là $(.05)(1) = .05$, Ta có được xác suất của bác bỏ H_0 khi nó đúng là .05, từ đó ta chọn được mức ý nghĩa.
Bây giờ ta xem chuyện gì sẽ xảy ra khi H_0 sai vì $\mu = 21$, ta lại dùng phần mềm Minitab cho mẫu tổng quát gồm10,000 mẫu khác nhau với kích thước 4 (mỗi mẫu đều có phân phối chuẩn với $\mu = 21$ và $\sigma = 2$), Tính $t = (\overline{x} - 20)/(s/\sqrt{4})$ cho mỗi mẫu sau đó xác định giá trị P.Mẫu đầu tiên có kết quả với $\overline{x} = 20.6411, s = .49637, t = 2.5832$, giá trị$P = .0408$. Hình 8.11$(b)$ cho ta biểu đồ giá trị P. Hình ảnh của biểu đồ khác với hình 8.11(a) nó có xu hướng lớn hơn khi giá trị P nhỏ (gần về 0) khi $\mu = 21$ lớn hơn khi $\mu = 20$. H_0lại bị bác bỏ tại mức ý nghĩa .05 bất cứ khi nào giá trị P đạt tối đa .05 (trong khoảng của lớp đầu tiên). Không may vì đây là trường hợp chỉ cho 19% của giá trị P. vì vậy chỉ 19% trong 10,000 được kiểm định chính xác và bác bỏ H_0 với 81% còn lại tức phạm phải sai lầm loại II, sự khác biệt về cỡ mẫu là khá nhỏ và 21 cũng không làm khác với giá trị khẳng định trong giả thiết không.
Hình 8.11(c) minh họa cho giá trị P khi H_0 sai vì $\mu = 22$ (vẫn với $n = 4$ và $\sigma = 2$), khi đó biểu đồ có nhiều giá trị gần 0 hơn. Trường hợp khi $\mu = 21$. Nói chung, khi μ di chuyển về bên phải của giá trị không là 20, phân phối của giá trị P sẽ tập trung nhiều hơn tại những giá trị gần 0. thậm chí ở đây có ít hơn 50% giá trị P nhỏ hơn .05. Vì thế khả năng H_0 không bị bác bỏ là không chính xác. Chỉ khi cho giá trị μ lớn hơn 20 (ví dụ ít nhất là 24 hay 25) giá trị P khi đó sẽ nhỏ hơn .05 và vì thế ta có được kết luận đúng.

Ý tưởng của ví dụ này là vì giá trị của kiểm định thống kê là ngẫu nhiên , giá trị P cũng sẽ là giá trị ngẫu nhiên do đó ta có một phân phối. Hơn nữa khi giá trị thực của tham số lấy từ giá trị xác định của giả thiết không, hơn nữa phân phối của giá trị P sẽ tập trung tại những giá trị gần 0 và có thể kiểm định chính xác hơn việc bác bỏ H_0(phụ thuộc vào β nhỏ hơn).



Hình 8.11 mô phỏng kết quả của giá trị P trong ví dụ 8.19

BÀI TẬP Phần 8.4 (47 – 62)

47. Gía trị P nào sẽ làm cho H_0 bị bác bỏ khi kiểm định tại mức .05?

a. .001 b. .021 c. .078

d. .047 e. .148

48. Với cặp giá trị P và mức ý nghĩa α đã cho. Với mỗi cặp từ giá trị quan sát P có dẫn đến bác bỏ H_0 tại mức ý nghĩa đã cho

a. Gía trịP=.084, $\alpha = .05$

b. Gía trịP=.003, $\alpha = .001$

c. Gía trịP=.0498, $\alpha = .05$

d. Gía trịP=.084, $\alpha = .10$

e. Gía trịP=.039, $\alpha = .01$

f. Gía trịP=.281, $\alpha = .10$

49. Đặt μ là thời gian phản ứng với một kích thích nhất định . Cho kiểm định trên mẫu lớn z của giả thiết $H_0 : \mu = 5$ ngược lại $H_a : \mu > 5$. Tìm giá trị P phù hợp với mỗi giá trị của kiểm định thống kê z

a. 1.42 b. .90 c. 1.96 d. 2.48 e. -.11

50. Lốp xe mới mua của một loại nhất định chịu được áp suất $30lb/in^2$. Đặt μ ký hiệu trung bình đúng của áp suất. Tìm giá trị P thích hợp với mỗi giá trị thống kê z đã cho dưới đây cho kiểm định $H_0 : \mu = 30$ ngược lại $H_a : \mu \neq 30$

a. 2.10 b. -1.75 c. -.55 d. 1.41 e. -5.3

51. Cho biết thông tin về giá trị P của kiểm định t trong mỗi trường hợp sau:

a. Kiểm định bên phải, $df = 8, t = 2.0$

b. Kiểm định bên trái, $df = 11, t = -2.4$

c. Kiểm định hai bên , $df = 15, t = -1.6$

d. Kiểm định bên phải, $df = 19, t = -.4$

e. Kiểm định bên phải, $df = 5, t = 5.0$

f. Kiểm định hai bên, $df = 40, t = -4.8$

52. Sơn dùng vẽ trên đường phố phải có đủ ánh sáng để có thể nhìn thấy rõ vào ban đêm. Đặt μ ký hiệu là trung bình đúng của phản xạ của mắt cho một loại sơn đang được xem xét. Một kiểm định $H_0 : \mu = 20$ ngược lại $H_a : \mu > 20$ được kiểm định dựa trên một mẫu ngẫu nhiên với kích thước n từ phân phối chuẩn. Kết luận nào là phù hợp trong mỗi tình huống sau:

a. $n = 15, t = 3.2, \alpha = .05$

b. $n = 9, t = 1.8, \alpha = .01$

c. $n = 24, t = -2$

53. Đặt μ là ký hiệu trung bình đúng của nồng độ thụ thể huyết thanh của tất cả thai phụ. Trung bình của tất cả phụ nữ được biết là 5.63. Bài viết "Truyền thụ thể huyết thanh để phát hiện thiếu sắt ở phụ nữ mang thai" (Amer: J. of Clinical Nutr., 1991 1077 – 1081) báo cáo rằng giá trị P>.10 cho kiểm định của $H_0 : \mu = 5.63$ ngược lại $H_a : \mu \neq 5.63$ dựa vào n=176 phụ nữ mang thai với mức ý nghĩa .01. Bạn có kết luận gì?

54. Bài viết "Phân tích trữ lượng những chai liên quan: Tại sao phải trả giá cho sự khác biệt đó nếu chỉ dựa vào thông báo về yêu cầu của nhà phê bình " (Chance, Summer 2005,pp.9-15) Báo cáo của một thử nghiệm để điều tra liệu những nhà kinh doanh rượu vang có thể phân biệt được giữa rượu dự trữ mắc hơn rượu thông thường. Rượu thử được chứa trong

4 thùng được đặt tên là A, B, C và D, với 2 thùng rượu riêng biệt và 2 thùng rượu thông thường. Mỗi người nếm thử được chọn ngẫu nhiên trong ba thùng, nếm thử để chọn rượu, và chỉ ra trong 3 thùng anh ta/cô ta tin rằng có 2 thùng khác . Một thử nghiệm n=855 trong đó có 364 kết quả phân biệt đúng (hoặc là rượu dự trữ khác với hai loại rượu thông thường hay một loại rượu thông thường khác với hai loại rượu dự trữ). Điều này có cung cấp bằng chứng hấp dẫn cho kết luận rằng người nếm loại này có khả năng phân biệt giữa rượu thông thường và rượu dự trữ không? phát biểu và kiểm định giả thiết liên quan sử dụng giá trị P. Bạn có ấn tượng gì với khả năng của người thử rượu để phân biệt giữa hai loại rượu này không?

55. Nhà sản xuất aspirin làm đầy lọ thuốc bằng cách cân trọng lượng hơn là đếm.Khi đó mỗi lọ chứa được 100 viên, trung bình trọng lượng mỗi viên là 5 gren (=0,0648 gram) Mỗi 100 viên lấy từ một lô thuốc rất lớn được cân. Kết quả là trung bình mẫu của trọng lượng mỗi viên là 4.87gren và độ lệch chuẩn mẫu là .35 gren. Thông tin nào cung cấp bằng chứng mạnh mẽ cho kết luận rằng công ty không làm đầy lọ thuốc như quảng cáo? Kiểm định giả thiết thích hợp sử dụng .01 bằng cách tính giá trị P ban đầu sau đó so sánh nó để có mức ý nghĩa.

56. Vì thay đổi dây chuyền sản xuất, sản lượng thực tế của một mẫu thép nhẹ chịu lực có năng suất tăng khác thường với năng suất thực tế.Đặt p ký hiệu là trung bình đúng tỉ lệ của mẫu có năng suất trước khi áp dụng năng suất lý thuyết. Nếu dựa vào mẫu đó có thể kết luận rằng hơn 20% của tất cả mẫu có năng suất trước thí điểm lý thuyết, có nên thay đổi dây chuyền sản xuất không?

a. Nếu 15 trong 60 mẫu có năng suất trước năng suất thí điểm lý thuyết. Gía trị P bằng bao nhiêu khi dùng kiểm định xấp xỉ, bạn nên tư vấn gì cho công ty

b. Nếu trung bình đúng của "năng suất đầu tiên" đạt 50% (do đó thí điểm là trung vị của của phân phối năng suất) và mức kiểm định .01 được sử dụng. Xác suất nào để công ty đưa ra kết luận thay đổi dây chuyền là cần thiết.

57. Bài viết "Uống nhiều rượu và dùng thuốc giải rượu trong Sinh viên đại học " (J. of Drug Issues, 2008:445-466) phát biểu rằng 51 trong trong mẫu gồm 462 sinh viên đại học phải kiêng rượu cả đời. Đây có là bằng chứng mạnh mẽ cho kết luận hơn 10% của tổng thể mẫu hoàn toàn không uống rượu ? Kiểm định giả thiết thích hợp dùng phân phối giá trị P [Chú ý: Bài viết sử dụng phương pháp thống kê tự động để nghiên cứu mức độ nghiện rượu trong sinh viên như nhẹ, vừa và nặng]

58. Một mẫu ngẫu nhiên của một mẫu đất xác định, có lượng chất hữu cơ (%) trong đất được xác định cho mỗi mẫu, kết quả cho bởi dữ liệu sau (từ "Tính chất kỹ thuật của đất" soil science, 1998:93-102)

1.10	5.09	0.97	1.59	4.60	0.32	0.55	1.45
0.14	4.47	1.20	3.50	5.02	4.67	5.22	2.69
3.98	3.17	3.03	2.21	0.69	4.47	3.31	1.17
0.76	1.17	1.57	2.62	1.66	2.05		

Gía trị của trung bình mẫu, độ lệch chuẩn mẫu và (ước lượng) lỗi chuẩn hóa của trung bình tương ứng là 2.481 , 1.616 và .295. Dữ liệu nào cho rằng % trung bình đúng của chất hữu cơ trong đất khác hơn 3% thực hiện kiểm định cho giả thiết xấp xỉ tại mức ý nghĩa .10 bằng cách đầu tiên xác định giá trị P. Bạn có kết luận khác không nếu sử dụng $\alpha = .05$? [Chú ý: xác suất chuẩn có những vị trí của dữ liệu được biểu diễn bằng mô hình trong điều kiện đủ ánh sáng và kích thước mẫu lớn]

59. Dữ liệu dưới đây cho biết về cường độ nén (MPa) của một mẫu bê tông khối trong bài viết "Nghiên cứu thử nghiệm về việc tái chế cao su với bê tông chịu cường độ nén cao" (

Magazine of Concrete Res.,2009 : 549 − 556)

112.3	97.0	92.7	86.0	102.0
99.2	95.8	103.5	89.0	86.7

a. Có hợp lý không khi nói cường độ nén của loại bê tông này có phân phối chuẩn?
b. Gỉa sử bê tông được dùng cho một ứng dụng riêng trừ khi có bằng chứng mạnh mẽ rằng trung bình đúng của cường độ nén ít hơn 100 MPa. Có nên dùng loại bê tông đó không? Thực hiện kiểm định của giả thiết xấp xỉ sử dụng phương pháp giá trị P.

60. Một loại viết được thiết kế sao cho trung bình đúng của thời gian sử dụng dưới 1 điều kiện nhất định (liên quan đến việc dùng viết máy) ít hơn 10 h . Một mẫu ngẫu nhiên của 18 cây bút được chọn, thời gian viết của mỗi loại được xác định và phân phối chuẩn có vị trí trong kết quả dữ liệu hỗ trợ cho kiểm định một mẫu t.
a. Nên kiểm định giả thiết nào nếu điều tra viên tin vào dự đoán đặc điểm thiết kế đã thỏa mãn?
b. Kết luận nào là phù hợp nếu giả thiết trong câu (a) là kiểm định $t = -2.3$
c. Kết luận nào là phù hợp nếu giả thiết trong câu (a) là kiểm định $t = -1.8$ và $\alpha = .01$
d. Kết luận nào là phù hợp nếu giả thiết trong câu (a) là kiểm định $t = -3.6$

61. Một quang phổ kế dùng để đo nồng độ CO [ppm (một phần triệu) với thể tích] được kiểm tra tính chính xác bằng cách đọc trên nhà sản xuất gas (gọi là khí gas) trong đó nồng độ CO được kiểm soát chính xác tại 70 ppm. Nếu đọc được cho thấy máy quang phổ hoạt động không bình thường nó phải được hiệu chỉnh lại. Gỉa sử mẫu khí gas có phân phối chuẩn trên cơ sở 6 lần đọc -85,77,82,68, 72 và 69 - việc hiệu chỉnh có cần thiết không? thực hiện kiểm định giả thiết liên quan sử dụng giá trị P với $\alpha = .05$

62. Dẫn xuất tương đối của thiết bị bán dẫn được xác định bởi lượng tạp chất bị pha tạp vào thiết bị trong quá trình sản xuất. cực silicon được sử dụng cho mục đích cụ thể là đòi hỏi cắt giảm điện áp trung bình là 60 Volt,và nếu không đạt được điều đó, lượng tạp chất phải được điều chỉnh. SAS xuất ra kết quả từ yêu cầu kiểm định giả thiết xác suất.

N	Mean	Std Dev	T	Prob. > |T|
15	0.0453333	0.0899100	1.9527887	0.0711

[Chú ý: SAS kiểm định $H_0 : \mu = 0$, vì thế để kiểm định $H_0 : \mu = .60$ giá trị không là .60 phải trừ đi mỗi x_i: Trung bình được báo cáo sau đó là trung bình của giá trị $(x_i - 60)$. Vì thế, giá trị P của SAS luôn là kiểm định hai bên] Kết luận nào sẽ được đưa ra cho mức ý nghĩa .01?.05?.10?

11 8.5 MỘT SỐ NHẬN XÉT VỀ LỰA CHỌN KIỂM ĐỊNH

Một trong những thử nghiệm để trả lời cho câu hỏi cần quan tâm và phương pháp thu thập dữ liệu dữ liệu (thiết kế một thử nghiệm) để xây dựng kiểm định xấp xỉ bao gồm ba điểm chính sau:
1. Chỉ định một thử nghiệm thống kê (hàm của giá trị quan sát sẽ giúp cho việc đưa ra quyết định)
2. Quyết định dạng chung của vùng bác bỏ (hàm của các giá trị quan sát sử dụng cho việc

đưa ra quyết định)
3. Chọn một số những giá trị quan trọng hoặc những giá trị mà sẽ không phụ thuộc vào vùng bác bỏ từ vùng chấp nhận được (bằng cách lấy phân phối của kiểm định thống kê khi H_0 đúng, và sau đó chọn ra một mức ý nghĩa)

Trong những ví dụ trước đó cả bước 1 và bước 2 đều được thực hiện theo một cách thức đặc biệt thông qua trực giác, Ví dụ, giả sử khi tổng thể về cơ bản có phân phối chuẩn với kỳ vọng μ và σ đã biết, từ $\overline{X}$ ta có biến chuẩn hóa cho kiểm định thống kê:

$$Z = \frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}}$$

Cho kiểm định $H_0 : \mu = \mu_0$ ngược lại $H_a : \mu > \mu_0$ bằng trực giác ta đề nghị bác bỏ H_0 khi Z lớn, cuối cùng giá trị quan trọng được xác định bằng mức ý nghĩa của α và dùng giá trị thực Z có phân phối chuẩn, khi H_0 đúng. Độ tin cậy của kiểm định trong việc đưa ra một quyết định đúng có thể đánh giá được bằng những nghiên cứu về sai lầm loại II trong xác suất
Vấn đề cần xem xét ở đây là khi thực hiện các bước từ 1-3 có những câu hỏi cần giải quyết sau.
1.Ý nghĩa thực tế và hệ quả của chọn một mức ý nghĩa cụ thể khi đã xác định được thử nghiệm là gì?
2. Có tồn tại nguyên tắc chung không phụ thuộc vào trực giác mà có thể sử dụng để thu được một quá trình kiểm định tốt nhất hoặc tốt hay không?
3. Khi hai hay nhiều kiểm định là phù hợp trong cùng một bối cảnh phải so sánh những kiểm định đó như thế nào để đưa ra quyết định nên sử dụng cái nào?
4. Nếu một kiểm định xuất phát từ giả thiết có phân phối cụ thể hay từ mẫu tổng thể, kiểm định sẽ có dạng như thế nào khi vi phạm giả thiết?

12 Ý NGHĨA THỐNG KÊ TRONG THỰC TẾ

Mặc dù quá trình đưa ra quyết định bằng cách sử dụng phương pháp luận cổ điển liên quan đến việc chọn mức ý nghĩa và sau đó bác bỏ hoặc không bác bỏ H_0 tại mức α đó, một báo cáo về việc sử dụng α để đưa ra quyết định lại truyền tải ít thông tin về dữ liệu mẫu. Đặc biệt khi kết quả của một thử nghiệm lại truyền đạt đến nhiều người, bác bỏ H_0 tại mức .05 sẽ thuyết phục hơn nếu giá trị quan sát của kiểm định thống kê sẽ vượt quá 5% giá trị điều kiện nếu nó vượt quá giá trị đó sẽ bác bỏ H_0. Điều này đúng với những yêu cầu dẫn đến khái niệm của giá trị P như cách có được mức ý nghĩa mà không cần áp đặt cho nó để có một giá trị α đặc biệt mà người khác muốn đưa ra quyết định cho riêng họ.
Thậm chí nếu giá trị P có trong bảng tóm tắt kết quả, có thể rất khó để giải thích giá trị này để đưa ra quyết định. Đó là vì, một giá trị P nhỏ sẽ chỉ ra một ý nghĩa thống kê, khi đó sẽ đề nghị bác bỏ H_0 và chấp nhận H_a, Khi kết quả là một mẫu lớn bắt đầu từ H_0 thì ít có ý nghĩa trong thực tế hơn. Trong nhiều tình huống, thử nghiệm chỉ bắt đầu từ H_0 lại ít có ý nghĩa trong thực tế trong khi với mức độ nhỏ từ H_0 lại có ý nghĩa thực tế hơn
Xét ví dụ, Kiểm định $H_0 : \mu = 100$ ngược lại $H_a : \mu > 100$ khi μ là kỳ vọng của phân phối chuẩn với $\sigma = 10$, giả sử giá trị đúng của $\mu = 101$, sẽ không thấy được tầm quan trọng của H_0 bằng trực quan ta không bác bỏ H_0 khi $\mu = 101$ sẽ liên quan đến một sai lầm nghiêm trọng. Cho một mẫu lớn hợp lý kích thước n, với giá trị μ này sẽ cho một giá trị $\overline{x}$ gần 101 vì thế ta không muốn có bất kỳ bằng chứng gì về mẫu này để có sự tranh luận mạnh mẽ về việc bác bỏ H_0 khi $\overline{x} = 101$ là biến quan sát được. Khi có nhiều cỡ mẫu bảng 8.1 sẽ ghi lại tất cả giá trị P khi $\overline{x} = 101$ và xác suất của việc không bác bỏ H_0 tại mức .01 khi $\mu = 101$
Cột thứ hai trong bảng 8.1 biểu diễn rằng ngay cả khi cỡ mẫu đủ lớn giá trị P của $\overline{x} = 101$

cũng gây ra tranh cãi mạnh mẽ cho việc bác bỏ H_0, khi $\overline{x}$ là biến quan sát và nó cũng gợi ý rằng giới hạn đúng trong thực tế ít khác với giá trị không $\mu_0 = 100$. Cột thứ ba chỉ ra rằng khi thực tế có sự khác biệt giữa giá trị đúng μ và giá trị không μ_0 cho mức ý nghĩa thích hợp và với một cỡ mẫu lớn sẽ luôn dẫn đến bác bỏ giả thiết không tại mức ý nghĩa đó. Tóm lại, Ta phải đặc biệt cẩn trọng trong việc giải thích bằng chứng khi cỡ mẫu lớn, khi đó bất kỳ những gì liên quan đến H_0 thường sẽ phát hiện bằng một kiểm định nhưng nguyên tắc này thường ít có ý nghĩa thực tế.
Bảng 8.1 Minh họa của việc ảnh hưởng của cỡ mẫu lên giá trị P và β

n	Gía trị P khi $\overline{x} = 101$	β cho mức kiểm định .01
25	.3085	.9664
100	.1587	.9082
400	.0228	.6293
900	.0013	.2514
1600	.0000335	.0475
2500	.000000297	.0038
10,000	$7.69.10^{-24}$	.0000

13 NGUYÊN TẮC VỀ TỈ LỆ HỢP LÝ

Đặt $x_1, x_2, ..., x_n$ là những biến ngẫu nhiên được quan sát có kích thước n từ phân phối xác suất $f(x;\theta)$.Phân phối chung của những giá trị mẫu này là tích $f(x_1;\theta).f(x_2;\theta)......f(x_n;\theta)$. Điều này đã được nói đến về ước lượng của hàm hợp lý, hàm hợp lý là phân phối chung này, được xem như hàm theo θ, xét kiểm định H_0. θ thuộc Ω_0, ngược lại H_a thuộc Ω_a, khi Ω_0 và Ω_a rời nhau(ví dụ, $H_0 : \theta \leq 100$ ngược lại $H_a : \theta \geq 100$). Nguyên tắc tỉ lệ hợp lý cho kiểm định được xây dựng theo quá trình như sau:
1. Tìm một giá trị lớn nhất của hợp lý cho bất kỳ θ thuộc Ω_0 (bằng cách tìm ước lượng lớn nhất thuộc Ω_0 sau đó thay vào hàm hợp lý)
2. Tìm giá trị lớn nhất của hợp lý cho θ bất kỳ thuộc Ω_a
3. Tỉ lệ có dạng:

$$\lambda(x_1, ..., x_n) = \frac{\text{Gía trị hợp lý lớn nhất cho } \theta \in \Omega_0}{\text{Gía trị hợp lý lớn nhất cho } \theta \in \Omega_a}$$

Tỉ lệ $\lambda(x_1, ..., x_n)$ được gọi là giá trị tỉ lệ hợp lý thống kê, quá trình kiểm định bao gồm bác bỏ H_0 khi tỉ lệ này nhỏ, nghĩa là chọn một hằng số k sao cho H_0 bị bác bỏ nếu $\lambda(x_1, ..., x_n) \leq k$. Do đó, H_0 bị bác bỏ khi mẫu số của λ lớn hơn nhiều so với tử số, Điều này chỉ ra rằng dữ liệu phù hợp với H_a hơn là với H_0.
Hằng số k được chọn sao cho có xác suất sai lầm loại I. Thường thì bất đẳng thức $\lambda \leq k$ có thể biến đổi để có điều kiện tương đương mà đơn giản hơn. Ví dụ, Kiểm định $H_0 : \mu \leq \mu_0$, ngược lại $H_a : \mu > \mu_0$. Trong trường hợp thông thường $\lambda \leq k$ tương đương với $t \geq c$. Nên với $c = t_{\alpha,n-1}$ thì kiểm định tỉ lệ hợp lý cũng là kiểm định một mẫu t.
Qui tắc về tỉ lệ hợp lý cũng có thể áp dụng khi $X_i's$ có những phân phối khác và thậm chí ngay cả khi nó độc lập, thông qua hàm hợp lý có thể làm cho phức tạp hơn trong trường hợp này. Một số quá trình kiểm định sẽ được trình bày trong chương sau có thể thu được từ qui tắc kiểm định hợp lý. Kiểm định đó thường xoay quanh việc "giảm thiểu β trong tất cả những kiểm định để có mức α mong muốn", vì thế đó là kiểm định đúng nhất. Để chi tiết hơn, ta có thể tham khảo một số ví dụ trong những tài liệu được liệt kê trong Chương 6.
Để có kiểm định t từ qui tắc tắc tỉ lệ hợp lý ta phải xây dựng một kiểm định thống kê hợp lý, có dạng của một phân phối xác suất, và mẫu phải được chỉ định lấy từ đó. Để lấy kiểm định

t từ qui tắc tỉ lệ hợp lý, Điều tra viên phải giả sử có một chuẩn pdf. Nếu điều tra viên cho rằng phân phối đó đối xứng nhưng không muốn có dạng chính xác (như chuẩn đồng dạng hay cauchy) thì sẽ thất bại vì không có cách nào để liên kết các giá trị cho tất cả các phân phối có tính đối xứng. Trong chương 15 ta sẽ trình bày một kiểm định thống kê phi tham số, vì thế, cái gọi là sai lầm xác suất loại I là sự kiểm soát tất cả những phân phối phụ thuộc khác nhau. Những quá trình này có ích khi điều tra viên có kiến thức hạn chế về phân phối phụ thuộc. Ta cũng sẽ thảo luận nhiều hơn về vấn đề 3 và 4 trong phần đầu.

BÀI TẬP phần 8.5 (63 – 64)
63. Xét vấn đề làm khô sơn đã thảo luận trong Ví dụ 8.2. Gỉa thiết là $H_0 : \mu = 75$ ngược lại $H_a : \mu < 75$, với giả sử σ có giá trị 9.0. Xét giá trị thay thế $\mu = 74$, trong bối cảnh của vấn đề này chắc chắn vi phạm từ H_0 không có trong thực tế
a. Cho mức kiểm định .01, tính β tại giá trị thay thế này cho cỡ mẫu $n = 100, 900$ và 2500.
b. Nếu giá trị quan sát của $\overline{X}$ là $\overline{x} = 74$, Có thể nói gì về kết quả của giá trị P khi $n = 2500$? Dữ liệu này có thỏa mức ý nghĩa tại bất kỳ giá trị chuẩn nào của α hay không?
c. Bạn có thực sự muốn sử dụng cỡ mẫu 2500 với mức kiểm định .01 (không quan tâm đến chi phí thử nghiệm)? giải thích.

64. Xét kiểm định cho mẫu lớn tại mức .01 trong phần 8.3 cho kiểm định $H_0 : p = .2$ với giả thiết $H_a : p > 2$.
a. Cho giá trị thay thế $p = .21$, tính $\beta(.21)$ cho cỡ mẫu $n = 100, 2500, 10,000, 40,000$ và 90,000.
b. Cho $\hat{p} = x/n = .21$, tính giá trị p khi n=100,2500, 10,000 và 40,000.
c. Trong hầu hết các tình huống, có hợp lý không khi sử dụng mức kiểm định .01 kết hợp với cỡ mẫu 40,000? tại sao có và tại sao không?

BÀI TẬP THÊM (65 – 87)
65. Một mẫu gồm 50 tròng kính dùng trong mắt kính có trung bình mẫu của độ dày là 3.05mm và độ lệch chuẩn .34mm. Mức trung bình đúng mong muốn cho độ dày là 3.20mm. Dữ liệu mạnh nào gợi ý rằng trung bình đúng độ dày của những tròng kính khác với giá trị mong muốn? Kiểm định nó sử dụng $\alpha = .05$
66. Trong bài tập 65. Gỉa sử người thử nghiệm tin rằng trước khi tập hợp dữ liệu giá trị σ xấp xỉ là .30. Nếu người thử nghiệm mong muốn xác suất của sai lầm loại II là .05 khi $\mu = 3.00$, là một cỡ mẫu 50 không cần thiết phải lớn ?
67. Một loại sắt xác định được chỉ định rằng nó chứa .85g silicon trong 100g sắt (.85%). Silicon chứa trong mỗi 25 mẫu sắt được chọn ngẫu nhiên được xác định, và theo phần mềm Minitab xuất ra kết quả từ một kiểm định của giả thiết thích hợp

Variable	N	Mean	StDev	SE Mean	T	P
sil cont	25	0.8880	0.1807	0.0361	1.05	0.30

a. Nên kiểm định giả thiết nào?
b. Kết luận nào được đưa ra khi dùng mức ý nghĩa .05 và tại sao? Trả lời cùng câu hỏi cho mức ý nghĩa .10.

68. Một phương pháp làm thẳng dây trước khi cuộn lại để làm một dây cót gọi là " nắn bằng con lăn " . Bài viết "ảnh hưởng của trục lăn và làm thẳng cuộn dây thừng và đặc tính của dây" (Springs, 1987:27-28) báo cáo về đặc tính căng của dây .Gỉa sử một mẫu gồm 16 dây được chọn và mỗi kiểm định được thực hiện để xác định sức căng N/mm^2. Kết quả trung bình mẫu và độ lệch chuẩn tương ứng là 2160 và 30.
a. Trung bình của sức căng cho dây cót sử dụng một sức căng là 2150 N/mm^2. Nên kiểm định giả thiết nào để xác định liệu trung bình sức căng cho phương pháp con lăn có vượt quá 2150?
b. Gỉa sử rằng phân phối của sức căng là xấp xỉ chuẩn. Kiểm định thống kê nào nên sử dụng

để kiểm định giả thiết trong câu (a)?
c. Gía trị của kiểm định thống kê cho dữ liệu này bằng bao nhiêu?
d. Gía trị P cho giá trị của kiểm định được tính trong câu (c) bằng bao nhiêu?
e.Cho mức kiểm định .05, Kết luận nào sẽ đạt được?

69. Ô nhiễm chất rắn trong đất ở Trung Quốc là một vấn đề môi trường nghiêm trọng, Bài viết "Ô nhiễm kim loại nặng trong đất và tích lũy kim loại nặng trong đất hoang của mỏ mangan, miền nam Trung Quốc" (Air, Soil, and Water Res.,$2008 : 31 - 41$) báo cáo rằng, cho một mẫu gồm 3 mẫu đất của một khu vực khai thác mỏ nhất định, trung bình mẫu tập trung cuả toàn bộ lượng đồng là 45.31 mg/kg (ước lượng) với tiêu chuẩn lỗi trung bình tương ứng là 5.26. Nó cũng nói lên rằng giá trị nền tảng của Trung quốc cho sự tập trung này là 20. Kết quả của nhiều thống kê mô tả trong bài viết dực trên giả thiết chuẩn.
a. Dữ liệu nào cung cấp bằng chứng mạnh mẽ cho kết luận rằng trung bình đúng tập trung trong vùng mẫu vượt quá giá trị nền tảng? Thực hiện kiểm định tại mức ý nghĩa .01 sử dụng phương pháp giá trị P. Kết quả này có làm bạn ngạc nhiên không?
b. Trong kiểm định của câu (a), Có khả năng giá trị P ít nhất là .01 khi trung bình đúng tập trung là 50 và độ lệch chuẩn đúng tập trung là 10 không?

70. Bài viết "Quản lý lớp cây trồng sử dụng đất than đá để bảo vệ và chống sương giá" (Argi, and Forest Meteorology, $1998 : 71 - 82$) Báo cáo những giá trị sau cho luồng nhiệt đất của tám ô được phủ bụi than.
34.7 35.4 34.7 37.7 32.5 28.0 18.4 24.9
Trung bình luồng nhiệt trong đất của ô chỉ được phủ bằng cỏ là 29.0. Gỉa sử rằng phân phối của luồng nhiệt là xấp xỉ chuẩn. Nên dùng dữ liệu nào để tăng hiệu quả của hiệu ứng bụi than so với cỏ? Kiểm định giả thiết thích hợp sử dụng $\alpha = .5$

71. Bài viết "Kiến thức về Cafeine, Thái độ, và tiêu thụ ở phụ nữ trưởng thành" (J. of Nutrition Educ., $1992{:}179 - 184$) báo cáo dữ liệu tóm tắt sau về lượng tiêu thụ cà phê hàng ngày của một mẫu gồm những phụ nữ trưởng thành: $n = 47, \overline{x} = 215mg, s = 235mg$, và phạm vi từ 5-1176.
a. Có hợp lý không khi nói phân phối của lượng tiêu thụ cà phê hàng ngày là chuẩn? Có cần giả sử một phân phối chuẩn để kiểm định giả thiết về giá trị của trung bình tổng thể của lượng tiêu thụ? giải thích cho lý do đó.
b. Gỉa sử trước đó tin rằng trung bình lượng tiêu thụ ít nhất là 200 mg. Dữ liệu đã cho có mâu thuẫn với niềm tin trước đó không? Kiểm định giả thiết thích hợp tại mức ý nghĩa .10 bao gồm phân tích giá trị P trong phân tích của bạn.

72. Doanh thu hàng năm của quỹ tương trợ là tỉ lệ phần trăm của tài sản của quỹ được bán ra trong một năm cụ thể. Nói chung, Qũi có giá trị doanh số thấp thì ổn định và ít rủi ro hơn, trong khi quĩ có giá trị doanh số cao cho thấy một lượng đáng kể mua và bán nhằm cố gắng tận dụng những biến động của thị trường tại thời điểm ngắn hạn. Gía trị ở đây của doanh thu cho một mẫu 20 quỹ hỗn hợp vốn lớn (tham khảo Bài tập 1.53 cho một vài thông tin thêm) trích xuất từ Morningstar.com:
1.03 1.23 1.10 1.64 1.30 1.27 1.25 0.87 1.05 0.64
0.94 2.86 1.05 0.75 0.09 0.79 1.61 1.26 0.93 0.84
a. Bạn có sử dụng kiểm định một mẫu t để quyết định liệu có bằng chứng nào hấp dẫn cho kết luận trung bình tổng thể của quĩ ít hơn 100%? giải thích.
b. Xác suất chuẩn của một ô của giá trị 20 ln (doanh số) biểu diễn một mô hình tuyến tính rõ ràng, cho thấy thật hợp lý khi giả sử rằng doanh số có phân phối loga chuẩn. Nhắc lại rằng X có phân phối loga chuẩn nếu $ln(X)$ có phân phối chuẩn với giá trị kỳ vọng μ và phương sai σ^2. Vì μ cũng là trung vị của phân phối $ln(X)$, e^μ là trung vị của phân phối X. Dùng thông

tin này để quyết định liệu có bằng chứng hấp dẫn nào cho quyết định rằng trung vị của phân phối tổng thể quĩ ít hơn 100%.

73. Trung bình đúng của sức phá vỡ của chất cách điện bằng gốm của một loại nhất định giả sử ít nhất là 10 psi. Họ sẽ sử dụng cho một ứng dụng đặc biệt trừ khi dữ liệu mẫu đưa ra kết luận rằng đặc điểm kỹ thuật này chưa được đáp ứng. Một kiểm định giả thiết sử dụng $\alpha = .01$ có được dựa trên một mẫu ngẫu nhiên gồm mười chất cách điện. Gỉa sử rằng sức phá vỡ có phân phối chuẩn với độ lệch chuẩn chưa biết.
a. Nếu độ lệch chuẩn đúng là .80. Có khả năng đó là chất cách điện được đánh giá thỏa mãn khi trung bình đúng của sức phá vỡ thực sự là 9.5 không? là 9.0?
b. Cỡ mẫu nào cần thiết để có 75% thay đổi phát hiện rằng trung bình đúng sức phá vỡ là 9.5 khi trung bình đúng độ lệch chuẩn là .80?

74. Những quan sát kèm theo trên thời gian cháy thừa (*sec*) xử lý dải đồ ngủ của trẻ em được cho trong bài viết "giới thiệu một vài độ chính xác và tính chính xác của vấn đề đo lường" (J.of Testing and Eval., $1982 : 132 - 140$) Gỉa sử trung bình đúng của thời gian cháy tối đa là 9.75 đã được ủy nhiệm. Dữ liệu nào nên lấy mà tình trạng này chưa được đáp ứng? Thực hiện kiểm định xấp xỉ sau suy luận đầu tiên để điều tra tính hợp lý của giả thiết mà dựa vào phương pháp suy luận của bạn.

9.85	9.93	9.75	9.77	9.67	9.87	9.67
9.94	9.85	9.75	9.83	9.92	9.74	9.99
9.88	9.95	9.95	9.93	9.92	9.89	

75.Tỉ lệ mắc một dạng khuyết tật về nhiễm sắc thể ở một tổng thể nam giới trưởng thành ở Mỹ tin rằng nằm ở 1 trên 75.Một mẫu ngẫu nhiên gồm 800 cá nhân ở Mỹ. Các tổ chức tội phạm tiết lộ có 16 người bị khuyết tật đó. Có thể kết luận rằng tỉ lệ mắc khiếm khuyết này trong số các tù nhân khác với tỉ lệ giả định trong tổng thể những nam giới trưởng thành?
a. Phát biểu và kiểm định giả thiết thích hợp sử dụng $\alpha = .05$. Loại sai lầm nào có thể xảy ra trong kết luận đó?
b. Gía trị P nào phù hợp với kiểm định này? Dựa vào giá trị P đó, H_0 có bị bác bỏ tại mức ý nghĩa .20 không?

76. Trong một cuộc điều tra về độc tố sản sinh bởi một loại rắn độc nhất định, Những nhà nghiên cứu chuẩn bị 26 lọ thuốc khác nhau, mỗi lọ chứa 1g chất độc, và sau đó xác định số lượng của thuốc kháng độc tố cần để trung hòa độc tố. Trung bình mẫu lượng thuốc kháng độc tố cần thiết được tìm thấy là 1.89 mg, và độ lệch chuẩn mẫu là .42. Nghiên cứu trước đó đã chỉ ra rằng trung bình đúng của lượng thuốc trung hòa là 1.75 mg/g của độc tố. Dữ liệu mới có mâu thuẫn với giá trị được đề nghị trong nghiên cứu trước đó không? Kiểm định giả thiết phù hợp sử dụng thông qua giá trị P. Gía trị phân tích của bạn có phụ thuộc vào bất kỳ giả thiết về phân phối tổng thể về lượng thuốc trung hòa không? giải thích.

77. Trung bình mẫu sức nén không bị giới hạn cho 45 mẫu của một loại gạch đặc biệt được tính là 3170 psi, và độ lệch chuẩn mẫu là 188. Phân phối của sức nén không giới hạn có thể hơi lệch. Dữ liệu nào biểu thị mạnh mẽ rằng trung bình đúng của sức kéo không giới hạn ít hơn giá trị thiết kế 3200? kiểm định sử dụng $\alpha = .001$.

78. Ngày 30 tháng 12 năm 2009. tờ NewYork Times báo cáo một khảo sát gồm 948 người Mỹ trưởng thành những người nói rằng ít nhất họ cũng quan tâm đến bóng đá trong trường đại học, 597 người nói hệ thống vô địch chung kết bóng đá nên thay thế bằng trận chung kết tương tự hệ thống được sử dụng trong bóng đá của các trường đại học . Điều này có cung cấp

bằng chứng thuyết phục cho kết luận rằng đa số các cá nhân ủng hộ thay B.C.S với trận trung kết? Kiểm định giả thiết thích hợp sử dụng phương pháp giá trị P.

79. Khi $X_1, X_2, ..., X_n$ là những biến Poisson độc lập với tham số μ, và n lớn, trung bình mẫu $\overline{X}$ có phân phối xấp xỉ chuẩn với $\mu = E(\overline{X})$ và $V(\overline{X}) = \mu/n$. Điều này suy ra rằng

$$Z = \frac{\overline{X} - \mu}{\sqrt{\mu/n}}$$

có xấp xỉ phân phối chuẩn. Cho kiểm định $H_0 : \mu = \mu_0$, ta có thể thay μ bằng μ_0 trong phương trình của Z để có được một kiểm định thống kê. này thực sự thích hợp để kiểm định cho mẫu lớn với mẫu số $S/\sqrt{n}$ (khi $X_i's$ là phân phối Poissson) vì nó được chỉnh sửa để thỏa giả thiết Poisson. Nếu số yêu cầu của tư vấn nhận được bằng một thống kê chắc chắn trong 5 ngày làm việc của tuần có phân phối Poisson và tổng số yêu cầu của tư vấn trong vòng 36 tuần là 160, Điều này có gợi ý rằng trung bình đúng số yêu cầu mỗi tuần vượt quá 4.0 không? Kiểm định sử dụng $\alpha = .02$

80.Bài viết vào ngày 11 tháng 11 năm 2005, vấn đề của San Luis Obispo mục diễn đàn báo cáo rằng những nhà nghiên cứu nghiên cứu cách mua hàng ngẫu nhiên ở cửa hàng California Wal-Mart tìm thấy máy quét sai giá tới 8.3% tại thời điểm đó. Gỉa sử điều này dựa trên 200 người mua hàng. Viện nghiên cứu quốc gia và kỹ thuật nói rằng trong khoảng thời gian dài nhất hai trong số 100 mặt hàng máy quét đúng giá.
a. Xây dựng một quá trình kiểm định với mức ý nghĩa (xấp xỉ) .05 và sau đó thực hiện kiểm định để quyết định liệu điểm chuẩn của NIST có không hài lòng ?
b. Cho quá trình kiểm định bạn đã làm trong câu (a). Xác suất của quyết định điểm chuẩn của NIST làm hài lòng khi tỉ lệ lỗi trong thực tế là 5%?

81. Nhà sản xuất bình nước nóng quảng cáo rằng với thiết bị làm nóng của nó, nhiệt độ 100^oF có thể đạt được trong vòng 15 phút. Một mẫu ngẫu nhiên gồm 42 bình được chọn, và thời gian cần thiết để đạt nhiệt độ 100^oF được xác định cho mỗi bình. Trung bình mẫu của thời gian và độ lệch chuẩn mẫu tương ứng là 16.5 phút và 2.2 phút. Dữ liệu nào gây ra nghi ngờ về phát biểu của công ty? Tính giá trị P và sử dụng nó để đưa ra kết luận tại mức .05.

82. Chương 7 trình bày một khoảng tin cậy CI cho phương sai σ^2 của phân phối chuẩn của tổng thể. Kết quả chính là rv $\chi^2 = (n-1)S^2/\sigma^2$ có phân phối Chi bình phương với n-1 df. Xét giả thiết không $H_0 : \sigma^2 = \sigma_0^2$(tương đương với $\sigma = \sigma_0$).Thì khi H_0 đúng, kiểm định thống kê $\chi^2 = (n-1)S^2/\sigma_0^2$ có phân phối chi bình phương với n-1 df. Nếu giả thiết thay thế là $H_a : \sigma^2 > \sigma_0^2$, bác bỏ H_0 nếu $(n-1)s^2/\sigma_0^2$ cho kiểm định với mức ý nghĩa α.
Để đảm bảo tính thống nhất hợp lý cho một ứng dụng cụ thể, mong muốn độ lệch chuẩn đúng của nhiệt độ hóa mềm của một loại dầu khí nhất định đạt tối đa $.50^oC$. Nhiệt độ hóa mềm của 10 mẫu khác nhau được xác định, mang lại một độ lệch chuẩn mẫu $.58^oC$. Điều này có mâu thuẫn với mô tả thống nhất không? Kiểm định giả thiết phù hợp sử dụng $\alpha = .01$?

83. Liên quan đến Bài tập 82, giả sử một điều tra viên muốn kiểm định $H_0 : \sigma^2 = .04$ ngược lại $H_a : \sigma^2 < .04$ dựa vào mẫu quan sát 21. Tính được giá trị của $20s^2/.04$ là 8.58. Tìm giới hạn trên của giá trị P sau đó đưa ra kết luận tại mức .01

84. Khi phân phối của tổng thể là n lớn, độ lệch chuẩn S có phân phối xấp xỉ chuẩn với $E(S) \approx \sigma$ và $V(X) \approx \sigma^2/(2n)$. Ta đã biết trong trường hợp này , cho n bất kỳ, $\overline{X}$ là chuẩn với $E(\overline{X}) = \mu$ và $V(\overline{X}) = \sigma^2/n$.
a. Gỉa sử rằng phân phối cơ bản là chuẩn, Xấp xỉ của ước lượng không chệch tại bách phân vị

thứ 99 $\theta = \mu + 2.33\sigma$ bằng bao nhiêu?
b. Khi $X_i's$ là chuẩn, nó có thể biểu diễn $\overline{X}$ và S là độc lập của rv không? (một độ đo cục bộ trong khi những độ đo khác lan rộng) . Sử dụng để tính $V(\hat{\theta})$ và $\sigma_{\hat{\theta}}$ cho ước lượng $\hat{\theta}$ của câu (a). Lỗi ước lượng chuẩn $\hat{\sigma}_{\hat{\theta}}$ bằng bao nhiêu?
c. Viết kiểm định thống kê cho kiểm định $H_0 : \theta = \theta_0$ có xấp xỉ phân phối chuẩn khi H_0 đúng. Nếu độ Ph của đất có phân phối chuẩn tại một vùng nhất định gồm 64 mẫu đất cho ta $\overline{x} = 6.33, s = .16$, Điều này có cung cấp bằng chứng mạnh mẽ cho kết luận tối đa 99% tất cả các mẫu có độ pH ít hơn 6.75 không? Kiểm định sử dụng $\alpha = .01$

85. Đặt $X_1, X_2, ..., X_n$ là một mẫu ngẫu nhiên của phân phối mũ với tham số λ. Thì có thể biểu diễn $2\lambda \sum X_i$ có phân phối chi bình phương với $\nu = 2n$ (biểu diễn ban đầu là $2\lambda X_i$ có phân phối chi bình phương với $\nu = 2$).
a. Sử dụng điều này để có được một kiểm định thống kê và vùng bác bỏ để xác định một mức kiểm định cho $H_0 : \mu = \mu_0$ ngược lại là một trong ba giả thiết thay thế sau [tức : $E(X_i) = \mu = 1/\lambda$, vì thế $\mu = \mu_0$ tương đương với $\lambda = 1/\mu_0$]
b. Gỉa sử có 10 thành phần giống nhau, mỗi thành phần có phân phối mũ thời gian cho đến khi thất bại, đã được kiểm tra. Kết quả thời gian thất bại là

95	16	11	3	42	71	225	64	87	123

Sử dụng quá trình kiểm định trong câu (b) để đưa ra quyết định liệu dữ liệu này có đề nghị mạnh mẽ rằng trung bình đúng thời gian ít hơn so với giá trị tuyên bố trước đó 75?

86. Gỉa sử phân phối của tổng thể là chuẩn với σ đã biết. Đặt γ sao cho $0 < \gamma < \alpha$.Cho kiểm định $H_0 : \mu = \mu_0$ngược lại $H_a : \mu \neq \mu_0$. Xét một kiểm định mà bác bỏ H_0 nếu hoặc $x \geq z_\gamma$ hoặc $z \leq -z_{\alpha-\gamma}$, khi đó kiểm định thống kê là $Z = (\overline{X} - \mu_0)/(\sigma/\sqrt{n})$.
a. Cho biết $P($ sai lầm loại I$) = \alpha$
b. Rút ra biểu thức cho $\beta(\mu')$. [tức là: Thực hiện kiểm định có dạng " bác bỏ H_0 nếu hoặc $\overline{x} \geq c_1$ hoặc $\leq c_2$"]
c. Đặt $\Delta > 0$. Cho biết giá trị nào của $\gamma($ liên quan đến $\alpha)$ sẽ có $\beta(\mu_0 + \Delta) < \beta(\mu_0 - \Delta)$?

87. Sau một thời gian học việc, một tổ chức cho một bài kiểm tra phải được thông qua để đủ điều kiện là thành viên chính thức. Đặt $p = P($ chọn ngẫu nhiên những người học việc vượt qua). Tổ chức mong muốn bài kiểm tra hầu hết nhưng không phải tất cả đều có thể vượt qua, vì thế quyết định mong muốn là p=.90. Cho một bài kiểm tra đặc biệt, giả thiết thích hợp là $H_0 : p = .90$ giả thiết trái ngược $H_a : p \neq .90$. Gỉa sử có 10 người tham gia kiểm tra và đặt $X =$ số người vượt qua kỳ thi.
a. Vùng của kiểm định bên trái của $\{0, 1, ..., 5\}$ với mức kiểm định.01 là gì?
b. Mặc dù H_a là có hai bên , chỉ ra không có kiểm định hai bên nào tại mức kiểm định .01
c. Phác họa biểu đồ của $\beta(p')$ như một hàm của p' cho kiểm định này. Đó có là điều mong muốn không?