Chương 6

Phân tích ngoại suy cho dữ liệu phân loại

6.1 Phân tích ngoại suy cho một tỷ lệ đơn

6.2 Khác biệt giữa hai tỷ lệ

6.3 Kiểm định độ phù hợp bằng phép kiểm định Chi-bình phương

6.4 Kiểm Định Tính Độc Lập Trong Bảng Hai Chiều

Trong chương này, chúng ta áp dụng các phương pháp và ý tưởng từ Chương 5 vào một số bối cảnh cho dữ liệu phân loại. Chúng ta sẽ bắt đầu bằng cách xem lại những gì đã học đối với một tỷ lệ đơn lẻ, trong đó phân phối chuẩn có thể được sử dụng để mô hình hóa độ bất định trong tỷ lệ mẫu. Tiếp theo, chúng ta áp dụng những ý tưởng tương tự để phân tích sự khác biệt của hai tỷ lệ bằng mô hình chuẩn. Sau đó trong chương, chúng ta sẽ áp dụng các kỹ thuật suy luận vào các bảng dự phòng; mặc dù chúng ta sẽ sử dụng một phân phối khác trong bối cảnh này, nhưng các ý tưởng cốt lõi của kiểm định giả thuyết vẫn giữ nguyên.



Để xem video, slides và các tài nguyên khác, vui lòng truy cập

www.openintro.org/os

6.1 Phân tích ngoại suy cho một tỷ lệ đơn

Chúng ta đã tìm hiểu các phương pháp suy luận đối với một tỷ lệ đơn trong Chương 5, bao gồm ước lượng điểm, khoảng tin cậy, và kiểm định giả thuyết. Trong phần này, chúng ta sẽ ôn tập lại những nội dung đó và đồng thời tìm hiểu cách lựa chọn cỡ mẫu phù hợp khi thu thập dữ liệu trong các tình huống liên quan đến một tỷ lệ đơn.

6.1.1 Xác định khi nào tỷ lệ mẫu có phân phối gần chuẩn

Tỷ lệ mẫu $\hat{p}$ có thể được mô hình hóa bằng phân phối chuẩn khi các quan sát trong mẫu là độc lập và kích thước mẫu đủ lớn.

PHÂN PHỐI LẤY MẪU CỦA $\hat{p}$

Phân phối lấy mẫu của $\hat{p}$ (tỷ lệ mẫu) dựa trên một mẫu có kích thước n, được chọn từ một tổng thể có tỷ lệ thật sự là p, sẽ xấp xỉ phân phối chuẩn khi:

1. Các quan sát trong mẫu là độc lập, ví dụ: được chọn từ một mẫu ngẫu nhiên đơn.
2. Số lần thành công và thất bại kỳ vọng trong mẫu đều ít nhất là 10, tức là: $np \geq 10$ và $n(1-p) \geq 10$. Đây được gọi là điều kiện thành công – thất bại.

Khi các điều kiện này được thỏa mãn, phân phối lấy mẫu của $\hat{p}$ sẽ gần giống phân phối chuẩn với trung bình: p và sai số chuẩn: $SE = \sqrt{\frac{p(1-p)}{n}}$

Thông thường, chúng ta không biết tỷ lệ thật p, nên cần thay thế bằng một giá trị nào đó để kiểm tra điều kiện và ước lượng sai số chuẩn. Với khoảng tin cậy, ta dùng tỷ lệ mẫu $\hat{p}$ để kiểm tra điều kiện thành công – thất bại và tính sai số chuẩn. Còn với kiểm định giả thuyết, ta thường dùng giá trị trong giả thuyết không thay cho p.

6.1.2 Khoảng tin cậy cho một tỷ lệ

Khoảng tin cậy cung cấp một khoảng giá trị hợp lý cho tham số pp, và khi $\hat{p}$ có thể mô hình hóa bằng phân phối chuẩn, khoảng tin cậy cho p có dạng

$$\hat{p} \pm z^{\star} \times SE$$

VÍ DỤ 6.1

Một mẫu ngẫu nhiên đơn gồm 826 người vay tiền nhanh được khảo sát nhằm hiểu rõ hơn về quan điểm của họ đối với các quy định và chi phí liên quan. 70% phản hồi ủng hộ các quy định mới đối với các công ty cho vay tiền nhanh. Liệu có hợp lý không khi mô hình hóa $\hat{p} = 0{,}70$ bằng phân phối chuẩn?

(E)

Dữ liệu được lấy từ một mẫu ngẫu nhiên, do đó các quan sát là độc lập và đại diện cho tổng thể quan tâm. Ta cũng cần kiểm tra điều kiện thành công – thất bại, sử dụng $\hat{p}$ thay cho p khi tính khoảng tin cậy:

Số lượng "ủng hộ": $np \approx 826 \times 0{,}70 = 578$

Số lượng "không ủng hộ": $n(1-p) \approx 826 \times (1-0{,}70) = 248$

Vì cả hai giá trị đều ít nhất là 10, ta có thể sử dụng phân phối chuẩn để mô hình hóa $\hat{p}$.

6.1 PHÂN TÍCH NGOẠI SUY CHO MỘT TỶ LỆ ĐƠN

G

BÀI TẬP THỰC HÀNH 6.2

Ước tính sai số chuẩn của $\hat{p} = 0{,}70$. Vì giá trị thật p không biết, và chúng ta đang tính sai số chuẩn cho khoảng tin cậy, nên dùng $\hat{p}$ thay cho p trong công thức. [1]

E

VÍ DỤ 6.3

Xây dựng khoảng tin cậy 95% cho p, tỉ lệ người vay payday ủng hộ việc siết chặt quy định với các công ty cho vay payday.

Dùng ước lượng điểm là $\hat{p} = 0.70$, hệ số tới hạn $z^\star = 1{,}96$ cho khoảng tin cậy 95%, và sai số chuẩn $SE = 0{,}016$ từ THỰC HÀNH CÓ HƯỚNG DẪN 6.2, ta tính khoảng tin cậy như sau

$$\textit{Ước lượng điểm} \pm z^\star \times SE \rightarrow 0{,}70 \pm 1{,}96 \times 0{,}016 \rightarrow (0{,}669 - \ 0{,}731)$$

Chúng ta tin tưởng 95% rằng tỷ lệ thực sự của người vay payday ủng hộ quy định vào thời điểm khảo sát nằm trong khoảng từ 0,669 đến 0,731.

KHOẢNG TIN CẬY CHO MỘT TỶ LỆ DUY NHẤT

Khi xác định rằng khoảng tin cậy cho một tỷ lệ là phù hợp trong một ứng dụng cụ thể, cần thực hiện **4 bước** để xây dựng khoảng tin cậy:

Chuẩn bị: Xác định giá trị $f(số\ người\ ủng\ hộ)$ và n, và mức độ tin cậy mong muốn.

Kiểm tra: Xác minh điều kiện đảm bảo phân phối xấp xỉ chuẩn. Với khoảng tin cậy một tỷ lệ, dùng $\hat{p}$ thay cho p để kiểm tra điều kiện thành công – thất bại.

Tính toán: Nếu các điều kiện được thỏa mãn, tính SE dùng $\hat{p}$, tìm $z^\star$, và xây dựng khoảng tin cậy.

Kết luận: Diễn giải khoảng tin cậy trong ngữ cảnh của vấn đề.

Để có thêm ví dụ về khoảng tin cậy một tỷ lệ, hãy xem Mục 5.2.

6.1.3 Kiểm định giả thuyết cho một tỷ lệ

Một trong những quy định được đề xuất cho các công ty cho vay payday là yêu cầu kiểm tra tín dụng và đánh giá khả năng trả nợ dựa trên tình hình tài chính của người vay. Chúng ta muốn biết: Liệu người vay có ủng hộ quy định này không?

G

BÀI TẬP THỰC HÀNH 1.11

Thiết lập giả thuyết để đánh giá liệu người vay có đa số ủng hộ hay phản đối loại quy định này. [2]

Để áp dụng khung phân phối chuẩn trong kiểm định giả thuyết về tỷ lệ, cần thỏa mãn hai điều kiện: độc lập và thành công – thất bại. Trong kiểm định giả thuyết, điều kiện thành công – thất bại được kiểm tra dựa trên tỷ lệ giả thuyết p_0: ta cần xác minh $np_0 \geq 10$ và $n(1 - p_0) \geq 10$.

[1] $SE = \sqrt{\frac{p(1-p)}{n}} \approx \sqrt{\frac{0{,}70(1-0{,}70)}{826}} = 0{,}016$

[2] $H_o: p = 0{,}50$. $H_A: p \neq 0{,}50$

6.1 PHÂN TÍCH NGOẠI SUY CHO MỘT TỶ LỆ ĐƠN

BÀI TẬP THỰC HÀNH 6.5

G

Liệu những người vay tiền từ các công ty cho vay ngắn hạn (payday loan) có ủng hộ một quy định yêu cầu bên cho vay phải kiểm tra báo cáo tín dụng và đánh giá khả năng trả nợ không ? Từ một mẫu ngẫu nhiên gồm 826 người vay, có 51% cho biết họ sẽ ủng hộ quy định này. Liệu có hợp lý không khi mô hình hóa $\hat{p} = 0.51$ bằng phân phối chuẩn để thực hiện kiểm định giả thuyết ở đây ? [3]

VÍ DỤ 6.6

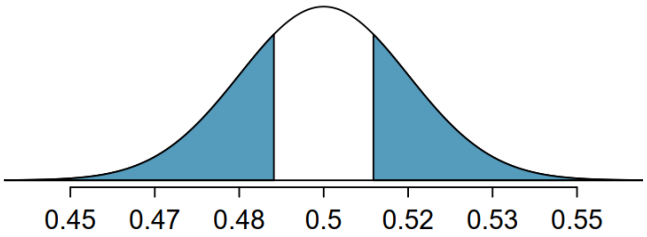
Sử dụng giả thuyết và dữ liệu từ BÀI TẬP THỰC HÀNH 6.4 và 6.5, hãy đánh giá xem liệu cuộc khảo sát có cung cấp bằng chứng thuyết phục rằng đa số người vay tiền payday ủng hộ quy định mới yêu cầu kiểm tra tín dụng và đánh giá khoản nợ không.

Vì giả thuyết đã được thiết lập và các điều kiện đã được kiểm tra, ta có thể chuyển sang bước tính toán. Sai số chuẩn trong bối cảnh kiểm định giả thuyết một tỷ lệ được tính theo giá trị giả thuyết p_0:

$$SE = \sqrt{\frac{p_0(1-p_0)}{n}} = \sqrt{\frac{0{,}5(1-0{,}5)}{826}} = 0{,}017$$

Một biểu đồ của mô hình chuẩn được minh họa bên dưới, với giá trị p-value là vùng được tô xanh.

E



Dựa trên mô hình chuẩn, ta tính thống kê kiểm định như sau:

$$Z = \frac{\textit{Ước lượng điểm} - \textit{Giá trị giả thuyết}}{SE} = \frac{0{,}51 - 0{,}50}{0{,}017} = 0{,}59$$

Diện tích một bên là 0,2776, nên p-value (tổng diện tích hai bên đối xứng) là 0,5552. Vì p-value lớn hơn 0,05, ta không bác bỏ giả thuyết H_0. Cuộc khảo sát không cung cấp bằng chứng đủ thuyết phục rằng đa số người vay payday ủng hộ (hoặc phản đối) quy định liên quan đến kiểm tra tín dụng và đánh giá khoản nợ.

KIỂM ĐỊNH GIẢ THUYẾT CHO MỘT TỶ LỆ

Khi đã xác định rằng kiểm định giả thuyết một tỷ lệ là phương pháp phù hợp, bạn cần thực hiện theo 4 bước sau để hoàn thành bài kiểm định:

Chuẩn bị. Xác định tham số cần quan tâm, liệt kê giả thuyết H_0 và H_A, chọn mức ý nghĩa (α), xác định p_0 và n.

Kiểm tra. Đảm bảo điều kiện để mô hình hoá bằng phân phối chuẩn: dữ liệu độc lập và thỏa mãn điều kiện thành công – thất bại ($np_0 \geq 10$, $n(1-p_0) \geq 10$).

Tính toán. Nếu điều kiện được đáp ứng, tính sai số chuẩn (SE) bằng p_0, tìm giá trị Z, tính p-value.

Kết luận. Diễn giải kết quả trong ngữ cảnh bài toán, dựa vào việc so sánh p-value với mức ý nghĩa α.

Để xem thêm ví dụ về kiểm định giả thuyết một tỷ lệ, hãy tham khảo Mục 5.3.

[3] Điều kiện độc lập được đảm bảo vì khảo sát sử dụng mẫu ngẫu nhiên. Điều kiện thành công – thất bại cũng được thỏa mãn, kiểm tra bằng giá trị giả thuyết $p_0 = 0{,}5$; $np_0 = 826 \times 0{,}5 = 413$; $n(1-p_0) = 826 \times 0{,}5 = 413$.

6.1 PHÂN TÍCH NGOẠI SUY CHO MỘT TỶ LỆ ĐƠN

6.1.4 Khi một hoặc nhiều điều kiện không được đáp ứng

Chúng ta đã dành nhiều thời gian để thảo luận về các điều kiện để tỉ lệ mẫu $\hat{p}$ có thể được mô hình hoá một cách hợp lý bằng phân phối chuẩn. Vậy điều gì xảy ra khi điều kiện thành công – thất bại không được thoả mãn? Còn khi điều kiện độc lập không được đảm bảo thì sao? Trong cả hai trường hợp, các ý tưởng tổng quát về khoảng tin cậy và kiểm định giả thuyết vẫn giữ nguyên, nhưng chiến lược hoặc kỹ thuật được sử dụng để tạo ra khoảng tin cậy hoặc tính p-value sẽ thay đổi.

Khi điều kiện thành công – thất bại không được thoả mãn trong kiểm định giả thuyết, chúng ta có thể mô phỏng phân phối của $\hat{p}$ dựa trên giá trị giả thuyết p_0. Khái niệm mô phỏng này tương tự với các ý tưởng được sử dụng trong nghiên cứu ca bệnh về bệnh sốt rét được trình bày ở Mục 2.3. Một phần trực tuyến trình bày chiến lược này có tại:

www.openintro.org/r?go=stat sim prop ht

Khi muốn xây dựng khoảng tin cậy mà điều kiện thành công – thất bại không được thoả mãn, chúng ta có thể sử dụng **khoảng Clopper-Pearson**. Chi tiết về phương pháp này nằm ngoài phạm vi của cuốn sách, tuy nhiên có rất nhiều tài nguyên trực tuyến nói về chủ đề này.

Điều kiện độc lập là một yêu cầu tinh tế và phức tạp hơn. Khi điều kiện này không được thoả mãn, điều quan trọng là phải hiểu cách thức và nguyên nhân của sự vi phạm đó. Ví dụ, nếu chúng ta thực hiện lấy mẫu theo cụm (cluster sampling, xem Mục 1.3), sẽ có những phương pháp thống kê phù hợp. Tuy nhiên, các phương pháp đó thường nằm ngoài phạm vi của các khoá học thống kê cơ bản hoặc thậm chí cả các khoá học nâng cao. Ngược lại, nếu chúng ta sử dụng lấy mẫu thuận tiện (convenience sampling), thì rất khó để tìm ra bất kỳ phương pháp nào có thể hiệu chỉnh một cách tự tin những thiên lệch cố hữu trong loại dữ liệu này.

Cuốn sách này được thiết kế để bao phủ các vấn đề thống kê có cấu trúc rõ ràng, nhưng bạn hãy nhớ rằng đây chỉ là cuốn sách đầu tiên trong một thư viện rộng lớn của các phương pháp thống kê – phù hợp với nhiều loại dữ liệu và bối cảnh khác nhau.

6.1 PHÂN TÍCH NGOẠI SUY CHO MỘT TỶ LỆ ĐƠN

6.1.5 Chọn cỡ mẫu khi ước lượng một tỉ lệ

Khi thu thập dữ liệu, chúng ta cần lựa chọn một cỡ mẫu phù hợp với mục tiêu của nghiên cứu. Thông thường, điều này đồng nghĩa với việc chọn một cỡ mẫu đủ lớn để **độ lệch chuẩn** (margin of error) – phần được cộng và trừ vào ước lượng điểm trong khoảng tin cậy – đủ nhỏ để mẫu có giá trị sử dụng thực tiễn. Ví dụ, nhiệm vụ của ta có thể là tìm một cỡ mẫu n sao cho tỉ lệ mẫu nằm trong $\pm 0{,}04$ của tỉ lệ thực trong khoảng tin cậy 95%.

ví dụ

VÍ DỤ 6.7

Một tờ báo của trường đại học đang tiến hành khảo sát để xác định tỉ lệ sinh viên ủng hộ việc tăng phí 200 đô la mỗi năm để xây sân vận động bóng bầu dục mới. Cỡ mẫu cần bao nhiêu để đảm bảo sai số nhỏ hơn 0,04, với mức tin cậy 95%?

Biên độ sai số cho một tỷ lệ mẫu là:

$$z^{\star}\sqrt{\frac{p(1-p)}{n}}$$

Mục tiêu của chúng ta là tìm cỡ mẫu n nhỏ nhất sao cho biên độ sai số này nhỏ hơn 0,04. Với mức độ tin cậy 95%, giá trị $z^{\star}$ tương ứng với 1,96:

$$1{,}96 \times \sqrt{\frac{p(1-p)}{n}} < 0{,}04$$

Có hai ẩn số trong phương trình: p và n. Nếu chúng ta có ước lượng p, có thể từ một khảo sát trước đó, chúng ta có thể nhập giá trị đó và giải phương trình tìm n. Nếu không có ước lượng nào như vậy, chúng ta phải sử dụng một giá trị khác cho p. Hóa ra biên độ sai số lớn nhất khi p bằng $0{,}5$, vì vậy chúng ta thường sử dụng giá trị trường hợp xấu nhất này nếu không có ước lượng nào về tỷ lệ:

$$1{,}96 \times \sqrt{\frac{0{,}5(1-0{,}5)}{n}} < 0{,}04$$

$$1{,}96^2 \times \frac{0{,}5(1-0{,}5)}{n} < 0{,}04^2$$

$$1{,}96^2 \times \frac{0{,}5(1-0{,}5)}{0{,}04^2} < n$$

$$600{,}25 < n$$

Chúng ta cần hơn 600,25 người tham gia, nghĩa là chúng ta cần 601 người tham gia trở lên để đảm bảo tỷ lệ mẫu nằm trong khoảng 0,04 so với tỷ lệ thực với độ tin cậy 95%.

Khi có ước lượng về tỉ lệ, ta dùng ước lượng đó thay cho 0.5 để tính toán ra cỡ mẫu cụ thể hơn, thường nhỏ hơn so với khi dùng trường hợp xấu nhất.

6.1 PHÂN TÍCH NGOẠI SUY CHO MỘT TỶ LỆ ĐƠN

BÀI TẬP THỰC HÀNH 6.8

G

Một quản lý sắp giám sát quy trình sản xuất hàng loạt một mẫu lốp xe mới trong nhà máy của mình, và cô ấy muốn ước lượng tỉ lệ lốp bị loại bởi bộ phận kiểm định chất lượng. Đội kiểm định chất lượng đã theo dõi ba mẫu lốp xe trước đó được sản xuất tại nhà máy, với tỉ lệ bị loại như sau: 1,7% đối với mẫu thứ nhất, 6,2% đối với mẫu thứ hai, và 1,3% đối với mẫu thứ ba. Người quản lý muốn kiểm tra đủ số lượng lốp xe để ước lượng được tỉ lệ lốp bị loại của mẫu mới với độ chính xác khoảng 1% và mức tin cậy 90%. Có ba tỉ lệ lỗi khác nhau để lựa chọn (1,7%, 6,2% và 1,3%). Hãy thực hiện tính toán cỡ mẫu tương ứng cho từng tỉ lệ đó, và xác định ba giá trị cỡ mẫu mà người quản lý có thể cân nhắc. [4]

VÍ DỤ 6.9

Các cỡ mẫu trong BÀI TẬP THỰC HÀNH 6.8 có sự khác biệt đáng kể. Vậy trong ba cỡ mẫu đó, bạn sẽ gợi ý sử dụng cỡ mẫu nào? Những yếu tố nào ảnh hưởng đến lựa chọn của bạn?

E

Chúng ta có thể xem mẫu lốp nào giống mẫu mới nhất, rồi chọn cỡ mẫu tương ứng với mẫu đó. Hoặc nếu hai trong ba ước lượng trước đây dựa trên các mẫu nhỏ, còn mẫu còn lại dựa trên mẫu lớn hơn, ta có thể cân nhắc sử dụng giá trị từ mẫu lớn hơn. Ngoài ra cũng có thể có những cách tiếp cận hợp lý khác.

Lưu ý thêm rằng cần kiểm tra điều kiện thành công – thất bại với mẫu cuối cùng. Ví dụ, nếu ta chọn $n = 1584$ lốp xe mà thấy tỉ lệ bị loại chỉ là 0,5%, thì xấp xỉ phân phối chuẩn sẽ không hợp lý, và lúc đó cần các phương pháp thống kê nâng cao hơn để tạo khoảng tin cậy.

BÀI TẬP THỰC HÀNH 6.10

G

Giả sử chúng ta muốn theo dõi thường xuyên mức độ ủng hộ của người vay ngắn hạn đối với việc điều chỉnh quy định đối với bên cho vay, bằng cách thực hiện một cuộc khảo sát mới mỗi tháng. Tuy nhiên, việc thực hiện khảo sát thường xuyên như vậy tốn kém, nên chúng ta quyết định mức sai số rộng hơn – 5% là có thể chấp nhận được cho mỗi cuộc khảo sát đơn lẻ. Dựa trên mẫu ban đầu cho thấy 70% người vay ủng hộ một hình thức điều chỉnh, ta cần chọn cỡ mẫu bao nhiêu cho mỗi khảo sát hàng tháng, để đảm bảo mức sai số là 0,05 với độ tin cậy 95%? [5]

---

[4] Với khoảng tin cậy 90%, $z^* = 1{,}6449$, và vì đã có ước lượng tỷ lệ là 0,017, chúng ta sẽ sử dụng nó trong công thức biên sai số:

$$1{,}6449 \times \sqrt{\frac{0{,}017(1-0{,}017)}{n}} < 0{,}01 \rightarrow \frac{0{,}017(1-0{,}017)}{n} < \left(\frac{0{,}01}{1{,}6449}\right)^2 \rightarrow 452{,}15 < \mathrm{n}$$

Với bài toán tính cỡ mẫu, ta luôn làm tròn lên, do đó mô hình lốp xe đầu tiên cho thấy cần ít nhất 453 chiếc lốp.

Một phép tính tương tự có thể được thực hiện với các tỷ lệ khác như 0,062 và 0,013 cho p, và bạn nên tự kiểm tra rằng các tỷ lệ này sẽ dẫn đến cỡ mẫu tối thiểu lần lượt là 1574 và 348 lốp xe.

[5] Chúng ta thực hiện phép tính tương tự như trước, nhưng lần này sử dụng $0{,}70$ thay vì 0,5:

$$1{,}96 \times \sqrt{\frac{\mathrm{p}(1-\mathrm{p})}{n}} \approx 1{,}96 \times \sqrt{\frac{0{,}70(1-0{,}70)}{n}} \leq 0{,}05 \rightarrow n \geq 322{,}7$$

Vì vậy, một cỡ mẫu từ 323 trở lên là hợp lý. (Lưu ý: luôn làm tròn lên khi tính cỡ mẫu!) Vì chúng ta có kế hoạch theo dõi cuộc khảo sát này theo thời gian, nên cũng có thể cần lặp lại các phép tính này định kỳ, để đảm bảo rằng đề xuất cỡ mẫu luôn phù hợp, trong trường hợp tỷ lệ nền (baseline rate) thay đổi.

6.1 PHÂN TÍCH NGOẠI SUY CHO MỘT TỶ LỆ ĐƠN

Bài tập

6.1 Sinh viên đại học ăn chay. Giả sử rằng 8% sinh viên đại học là người ăn chay. Hãy xác định các phát biểu sau là đúng hay sai, và giải thích lý do.

a) Phân phối của tỷ lệ mẫu sinh viên ăn chay trong các mẫu ngẫu nhiên cỡ 60 xấp xỉ phân phối chuẩn vì $n \geq 30$.
b) Phân phối của tỷ lệ mẫu sinh viên đại học ăn chay trong các mẫu ngẫu nhiên cỡ 50 bị lệch phải.
c) Một mẫu ngẫu nhiên gồm 125 sinh viên đại học, trong đó 12% là người ăn chay, sẽ được coi là bất thường.
d) Một mẫu ngẫu nhiên gồm 250 sinh viên đại học, trong đó 12% là người ăn chay, sẽ được coi là bất thường.
e) Sai số chuẩn sẽ giảm một nửa nếu tăng cỡ mẫu từ 125 lên 250.

6.2 Thanh niên Mỹ, Phần I. Khoảng 77% thanh niên Mỹ nghĩ rằng họ có thể đạt được "giấc mơ Mỹ". Hãy xác định các phát biểu sau là đúng hay sai, và giải thích lý do.[6]

a) Phân phối của tỷ lệ mẫu những thanh niên nghĩ rằng họ có thể đạt được giấc mơ Mỹ trong các mẫu cỡ 20 bị lệch trái.
b) Phân phối của tỷ lệ mẫu trong các mẫu ngẫu nhiên cỡ 40 xấp xỉ chuẩn vì $n \geq 30$.
c) Một mẫu ngẫu nhiên gồm 60 thanh niên, trong đó 85% nghĩ rằng họ có thể đạt được giấc mơ Mỹ, sẽ được coi là bất thường.
d) Một mẫu ngẫu nhiên gồm 120 thanh niên, trong đó 85% nghĩ rằng họ có thể đạt được giấc mơ Mỹ, sẽ được coi là bất thường.

6.3 Mèo mướp cam. Giả sử rằng 90% mèo mướp cam là mèo đực. Hãy xác định các phát biểu sau là đúng hay sai, và giải thích lý do.

a) Phân phối của tỷ lệ mẫu trong các mẫu ngẫu nhiên cỡ 30 bị lệch trái.
b) Sử dụng cỡ mẫu gấp 4 lần sẽ giúp giảm sai số chuẩn của tỷ lệ mẫu xuống một nửa.
c) Phân phối của tỷ lệ mẫu trong các mẫu ngẫu nhiên cỡ 140 xấp xỉ chuẩn.
d) Phân phối của tỷ lệ mẫu trong các mẫu ngẫu nhiên cỡ 280 xấp xỉ chuẩn.

6.4 Thanh niên Mỹ, Phần II. Khoảng 25% thanh niên Mỹ trì hoãn việc lập gia đình do tình trạng kinh tế suy thoái kéo dài. Hãy xác định các phát biểu sau là đúng hay sai, và giải thích lý do.[7]

a) Phân phối của tỷ lệ mẫu thanh niên trì hoãn việc lập gia đình trong các mẫu ngẫu nhiên cỡ 12 bị lệch phải.
b) Để phân phối của tỷ lệ mẫu xấp xỉ chuẩn, cỡ mẫu ngẫu nhiên cần ít nhất là 40.
c) Một mẫu ngẫu nhiên gồm 50 thanh niên, trong đó 20% trì hoãn lập gia đình, sẽ được coi là bất thường.
d) Một mẫu ngẫu nhiên gồm 150 thanh niên, trong đó 20% trì hoãn lập gia đình, sẽ được coi là bất thường.
e) Tăng gấp ba cỡ mẫu sẽ giúp giảm sai số chuẩn của tỷ lệ mẫu xuống một phần ba.

[6] A. Vaughn. "Poll finds young adults optimistic, but not about money". In: *Los Angeles Times* (2011).
[7] Demos.org. "The State of Young America: The Poll". In: (2011).

6.1 PHÂN TÍCH NGOẠI SUY CHO MỘT TỶ LỆ ĐƠN

6.5 Bình đẳng giới. Khảo sát xã hội tổng quát (General Social Survey) đã hỏi một mẫu ngẫu nhiên gồm 1.390 người Mỹ với câu hỏi: "Nhìn chung, bạn có nghĩ rằng chính phủ nên hay không nên có trách nhiệm thúc đẩy bình đẳng giữa nam và nữ?" 82% người tham gia trả lời "nên". Với mức tin cậy 95%, mẫu khảo sát này có biên độ sai số là 2%. Dựa vào thông tin này, hãy xác định các phát biểu sau là đúng hay sai, và giải thích lý do.[8]

a) Chúng ta có thể tin tưởng 95% rằng từ 80% đến 84% người Mỹ trong mẫu khảo sát này cho rằng chính phủ nên thúc đẩy bình đẳng giới.
b) Chúng ta có thể tin tưởng 95% rằng từ 80% đến 84% người Mỹ trong toàn dân số cho rằng chính phủ nên thúc đẩy bình đẳng giới.
c) Nếu chúng ta thực hiện nhiều mẫu ngẫu nhiên gồm 1.390 người Mỹ và tính toán khoảng tin cậy 95% cho mỗi mẫu, thì 95% các khoảng tin cậy này sẽ chứa tỷ lệ thật sự trong dân số cho rằng chính phủ nên thúc đẩy bình đẳng giới.
d) Để giảm biên độ sai số xuống còn 1%, chúng ta cần tăng gấp 4 lần cỡ mẫu.
e) Dựa trên khoảng tin cậy này, có đủ bằng chứng để kết luận rằng đa số người Mỹ cho rằng chính phủ nên thúc đẩy bình đẳng giới.

6.6 Người lái xe cao tuổi. Cuộc khảo sát của Marist Poll báo cáo rằng 66% người trưởng thành trên toàn quốc cho rằng những người có bằng lái xe nên phải thi lại bằng lái khi đến tuổi 65. Cuộc khảo sát thực hiện với 1.018 người Mỹ trưởng thành và biên độ sai số được báo cáo là 3% với mức tin cậy 95%.[9]

a) Kiểm tra xem biên độ sai số được báo cáo bởi Marist Poll có chính xác không.
b) Dựa trên khoảng tin cậy 95%, liệu khảo sát có cung cấp bằng chứng thuyết phục rằng *hơn* 70% dân số nghĩ rằng người có bằng lái nên thi lại khi bước sang tuổi 65?

6.7 Bắn pháo hoa vào ngày 4 tháng 7. Một kênh tin tức địa phương báo cáo rằng 56% trong số 600 cư dân Kansas được chọn ngẫu nhiên có kế hoạch đốt pháo hoa vào ngày 4/7. Hãy xác định biên độ sai số cho ước lượng điểm 56% với mức tin cậy 95%.[10]

6.8 Đánh giá cuộc sống tại Hy Lạp. Hy Lạp đã trải qua cuộc khủng hoảng kinh tế nghiêm trọng từ cuối năm 2009. Một khảo sát của Gallup vào năm 2011 với 1.000 người Hy Lạp được chọn ngẫu nhiên cho thấy 25% trong số họ đánh giá cuộc sống của mình tệ đến mức được xem là "đang chịu khổ".[11]

a) Mô tả tham số dân số cần quan tâm. Giá trị ước lượng điểm của tham số này là bao nhiêu?
b) Kiểm tra xem các điều kiện cần thiết để xây dựng khoảng tin cậy từ dữ liệu này có được thỏa mãn hay không.
c) Xây dựng khoảng tin cậy 95% cho tỷ lệ người Hy Lạp "đang chịu khổ".
d) Không cần tính toán, mô tả điều gì sẽ xảy ra với khoảng tin cậy nếu chúng ta sử dụng mức tin cậy cao hơn.
e) Không cần tính toán, mô tả điều gì sẽ xảy ra với khoảng tin cậy nếu chúng ta tăng cỡ mẫu.

6.9 Du học. Một khảo sát với 1.509 học sinh trung học phổ thông đã thi SAT và hoàn thành một khảo sát web tùy chọn cho thấy 55% cho biết họ tương đối chắc chắn rằng mình sẽ tham gia một chương trình du học khi học đại học.[12]

a) Mẫu khảo sát này có đại diện cho toàn bộ học sinh trung học phổ thông tại Mỹ không? Giải thích lý do.
b) Giả sử các điều kiện để suy luận thống kê được thỏa mãn. Ngay cả khi ở phần (a) bạn cho rằng phương pháp khảo sát không đáng tin cậy, việc phân tích này vẫn có thể có giá trị (dù không nên công bố chính thức). Hãy xây dựng khoảng tin cậy 90% cho tỷ lệ học sinh trung học (trong số những người thi SAT) chắc chắn sẽ tham gia chương trình du học, và giải thích ý nghĩa trong bối cảnh.
c) "Mức tin cậy 90%" nghĩa là gì?
d) Dựa trên khoảng tin cậy ở phần (b), có nên khẳng định rằng đa số học sinh trung học chắc chắn sẽ tham gia chương trình du học ở đại học không?

[8] National Opinion Research Center, General Social Survey, 2018.
[9] Marist Poll, Road Rules: Re-Testing Drivers at Age 65?, March 4, 2011.
[10] Survey USA, News Poll #19333, data collected on June 27, 2012.
[11] Gallup World, More Than One in 10 "Suffering" Worldwide, data collected throughout 2011.
[12] studentPOLL, College-Bound Students' Interests in Study Abroad and Other International Learning Activities, January 2008.

6.1 PHÂN TÍCH NGOẠI SUY CHO MỘT TỶ LỆ ĐƠN

6.10 Hợp pháp hóa cần sa – Phần I. Khảo sát Xã hội Tổng quát đã hỏi 1578 cư dân Hoa Kỳ: "Bạn có nghĩ rằng việc sử dụng cần sa nên được hợp pháp hóa hay không?" 61% người trả lời cho rằng nên được hợp pháp hóa.[13]

a) Giá trị 61% là thống kê mẫu hay tham số tổng thể? Giải thích.
b) Hãy xây dựng khoảng tin cậy 95% cho tỷ lệ cư dân Hoa Kỳ nghĩ rằng cần sa nên được hợp pháp hóa, và diễn giải trong bối cảnh dữ liệu.
c) Một nhà phê bình chỉ ra rằng khoảng tin cậy 95% này chỉ chính xác nếu thống kê tuân theo phân phối chuẩn, hoặc nếu mô hình chuẩn là một xấp xỉ tốt. Điều này có đúng với dữ liệu này không? Giải thích.
d) Một bản tin về kết quả khảo sát này tuyên bố: "Đa số người Mỹ nghĩ rằng cần sa nên được hợp pháp hóa." Dựa trên khoảng tin cậy của bạn, tuyên bố này có được hợp lý không?

6.11 Kế hoạch y tế quốc gia – Phần I. Một cuộc thăm dò của Quỹ Gia đình Kaiser với người trưởng thành tại Hoa Kỳ vào năm 2019 cho thấy: 79% đảng viên Dân chủ, 55% người Độc lập, và 24% đảng viên Cộng hòa ủng hộ một "Kế hoạch Y tế Quốc gia" chung chung. Tổng cộng có 347 người Dân chủ, 298 người Cộng hòa, và 617 người Độc lập được khảo sát.[14]

a) Một chuyên gia phân tích chính trị trên truyền hình tuyên bố rằng đa số người Độc lập ủng hộ Kế hoạch Y tế Quốc gia. Dữ liệu này có cung cấp bằng chứng thuyết phục để ủng hộ tuyên bố này không?
b) Bạn có kỳ vọng rằng một khoảng tin cậy cho tỷ lệ người Độc lập phản đối kế hoạch y tế công sẽ bao gồm 0.5 không? Giải thích.

6.12 Đại học có xứng đáng không? – Phần I. Trong một mẫu ngẫu nhiên đơn giản gồm 331 người Mỹ trưởng thành không có bằng đại học 4 năm và không đang theo học, có 48% cho biết họ quyết định không học đại học vì không đủ khả năng tài chính.[15]

a) Một bài báo tuyên bố rằng chỉ là thiểu số người Mỹ quyết định không học đại học vì lý do tài chính, và sử dụng ước lượng điểm từ khảo sát này làm bằng chứng. Hãy thực hiện kiểm định giả thuyết để xác định liệu dữ liệu này có cung cấp bằng chứng mạnh mẽ ủng hộ tuyên bố đó hay không.
b) Bạn có kỳ vọng rằng một khoảng tin cậy cho tỷ lệ người Mỹ không học đại học vì lý do tài chính sẽ bao gồm 0.5 không? Giải thích.

6.13 Thử nếm. Một số người cho rằng họ có thể phân biệt giữa nước ngọt không đường (diet) và nước ngọt thường (regular) chỉ sau một ngụm. Một nhà nghiên cứu muốn kiểm tra tuyên bố này đã lấy mẫu ngẫu nhiên 80 người. Ông đổ đầy 80 cốc trắng, một nửa chứa nước diet và một nửa chứa nước thường (chia ngẫu nhiên), và yêu cầu mỗi người uống một ngụm và xác định đó là loại nào. Có 53 người trả lời đúng.

a) Dữ liệu này có cung cấp bằng chứng mạnh mẽ rằng những người này phân biệt tốt hơn hoặc kém hơn ngẫu nhiên giữa hai loại nước ngọt không?
b) Hãy diễn giải p-value trong bối cảnh của tình huống này.

6.14 Đại học có xứng đáng không? – Phần II. Bài tập 6.12 trình bày kết quả một cuộc khảo sát cho thấy 48% trong số 331 người Mỹ quyết định không học đại học là do không đủ khả năng tài chính.

a) Hãy tính khoảng tin cậy 90% cho tỷ lệ người Mỹ quyết định không học đại học vì lý do tài chính, và diễn giải trong bối cảnh.
b) Giả sử bạn muốn biên độ sai số với mức tin cậy 90% là khoảng 1.5%. Bạn sẽ đề xuất cỡ mẫu bao nhiêu?

6.15 Kế hoạch y tế quốc gia – Phần II. Bài tập 6.11 trình bày kết quả một khảo sát cho thấy 55% người Độc lập ủng hộ Kế hoạch Y tế Quốc gia. Nếu bạn muốn ước lượng tỷ lệ này trong khoảng sai số 1% với mức tin cậy 90%, bạn sẽ cần bao nhiêu người trong mẫu khảo sát?

6.16 Hợp pháp hóa cần sa – Phần II. Như đã thảo luận trong Bài tập 6.10, Khảo sát Xã hội Tổng quát báo cáo rằng 61% cư dân Hoa Kỳ cho rằng cần sa nên được hợp pháp hóa. Nếu bạn muốn giới hạn biên độ sai số của khoảng tin cậy 95% trong 2%, bạn sẽ cần khảo sát khoảng bao nhiêu người Mỹ?

[13] National Opinion Research Center, General Social Survey, 2018.
[14] Kaiser Family Foundation, The Public On Next Steps For The ACA And Proposals To Expand Coverage, data collected between Jan 9-14, 2019.
[15] Pew Research Center Publications, Is College Worth It?, data collected between March 15-29, 2011.

6.2 Khác biệt giữa hai tỷ lệ

Chúng tôi muốn mở rộng các phương pháp trong Mục 6.1 để áp dụng khoảng tin cậy và kiểm định giả thuyết cho sự khác biệt giữa hai tỷ lệ quần thể: $p_1 - p_2$. Trong quá trình tìm hiểu, chúng ta sẽ xác định một ước lượng hợp lý cho $p_1 - p_2$ dựa trên mẫu, và bạn có thể đã đoán được dạng của nó: $\hat{p}_1 - \hat{p}_2$. Tiếp theo, chúng ta sẽ áp dụng cùng một quy trình như trong trường hợp một tỷ lệ đơn: xác minh rằng ước lượng điểm có thể được mô hình hóa bằng phân phối chuẩn, tính sai số chuẩn của ước lượng, và áp dụng khuôn khổ suy luận thống kê.

6.2.1 Phân phối chọn mẫu của hiệu hai tỷ lệ

Tương tự như với $\hat{p}$, sự khác biệt giữa hai tỷ lệ mẫu $\hat{p}_1 - \hat{p}_2$ có thể được mô hình hóa bằng phân phối chuẩn khi các điều kiện nhất định được thỏa mãn. Trước hết, chúng ta cần một điều kiện độc lập tổng quát hơn, và thứ hai, điều kiện thành công - thất bại phải được thỏa mãn ở cả hai nhóm.

CÁC ĐIỀU KIỆN ĐỂ PHÂN PHỐI LẤY MẪU CỦA $\hat{p}_1 - \hat{p}_2$ CÓ THỂ XEM LÀ PHÂN PHỐI CHUẨN

Hiệu số $\hat{p}_1 - \hat{p}_2$ có thể được mô hình hóa bằng một phân phối chuẩn khi các điều kiện sau được thỏa mãn:

- *Tính độc lập, mở rộng:* Dữ liệu độc lập trong từng nhóm và giữa hai nhóm. Thông thường điều kiện này được đảm bảo nếu dữ liệu được thu thập từ hai mẫu ngẫu nhiên độc lập hoặc từ một thí nghiệm ngẫu nhiên.
- *Điều kiện thành công - thất bại:* Điều kiện thành công - thất bại phải được thỏa mãn cho cả hai nhóm, tức là số lượng "thành công" và "thất bại" trong mỗi nhóm phải đủ lớn (thường là ít nhất 10 mỗi loại). Ta kiểm tra điều kiện này cho từng nhóm một cách riêng biệt.

Khi những điều kiện trên được thoả mãn, sai số chuẩn $\hat{p}_1 - \hat{p}_2$ là:

$$SE = \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$$

với p_1 và p_2 đại diện cho tỷ lệ quần thể, và n_1 và n_2 đại diện cho cỡ mẫu.

6.2.2 Khoảng tin cậy cho $p_1 - p_2$

Ta có thể áp dụng công thức khoảng tin cậy tổng quát cho sự khác biệt giữa hai tỷ lệ, trong đó sử dụng $\hat{p}_1 - \hat{p}_2$ làm ước lượng điểm và thay thế công thức SE:

$$\text{ước lượng điểm } \pm z^\star \times SE \rightarrow \hat{p}_1 - \hat{p}_2 \pm z^\star \times \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$$

Chúng ta cũng có thể tiếp tục sử dụng các bước Chuẩn bị – Kiểm tra – Tính toán – Kết luận để tính khoảng tin cậy hoặc thực hiện kiểm định giả thuyết. Các chi tiết có thay đổi chút ít, nhưng cách tiếp cận tổng thể vẫn giữ nguyên. Hãy ghi nhớ các bước này khi bạn áp dụng các phương pháp thống kê.

6.2 KHÁC BIỆT GIỮA HAI TỶ LỆ

VÍ DỤ 6.11

Chúng ta xem xét một thí nghiệm trên các bệnh nhân bị ngừng tim và được hồi sức tim phổi (CPR), sau đó được nhập viện. Các bệnh nhân này được chia ngẫu nhiên thành hai nhóm: nhóm điều trị được cho dùng thuốc làm loãng máu, và nhóm đối chứng thì không. Biến kết quả là bệnh nhân có sống ít nhất 24 giờ hay không. Kết quả được trình bày trong Hình 6.1. Hãy kiểm tra xem ta có thể mô hình hóa sự khác biệt tỷ lệ mẫu bằng phân phối chuẩn không

Đầu tiên chúng ta kiểm tra điều kiện độc lập: vì đây là thí nghiệm ngẫu nhiên, điều kiện này được thỏa mãn.

Tiếp theo, điều kiện thành công - thất bại: mỗi nhóm đều có ít nhất 10 trường hợp sống sót và tử vong (11, 14, 39, 26), nên điều kiện này cũng được thỏa mãn.

Với cả hai điều kiện được đáp ứng, sự khác biệt trong tỷ lệ mẫu có thể được mô hình hóa bằng phân phối chuẩn đối với dữ liệu này.

	Sống sót	Tử vong	Tổng cộng
Đối chứng	11	39	50
Điều trị	14	26	40
Tổng cộng	25	65	90

Hình 6.1: Kết quả nghiên cứu CPR. Bệnh nhân trong nhóm điều trị được cho dùng thuốc làm loãng máu, còn nhóm đối chứng thì không.

VÍ DỤ 6.12

Tạo và diễn giải khoảng tin cậy 90% cho sự khác biệt tỷ lệ sống sót trong nghiên cứu CPR.

Ta sử dụng p_t cho tỷ lệ sống ở nhóm điều trị và p_c cho nhóm đối chứng:

$$\hat{p}_t - \hat{p}_c = \frac{14}{40} - \frac{11}{50} = 0{,}35 - 0{,}22 = 0{,}13$$

Ta sử dụng công thức SE được đưa ra ở trang 217. Tương tự như trong trường hợp một mẫu, ta sử dụng các ước lượng mẫu của từng tỷ lệ trong công thức khi xây dựng khoảng tin cậy:

$$SE \approx \sqrt{\frac{0{,}35(1-0{,}35)}{40} - \frac{0{,}22(1-0{,}22)}{50}} = 0{,}095$$

Với khoảng tin cậy 90%, ta dùng $z^\star = 1{,}6449$:

$$\text{Ước lượng điểm} \pm z^\star \times SE \rightarrow 0{,}13 \pm 1{,}6449 \times 0{,}095 \rightarrow (-0{,}026 - 0{,}286)$$

Chúng ta tin tưởng 90% rằng thuốc làm loãng máu có ảnh hưởng từ -2,6% đến +28,6% điểm phần trăm lên tỷ lệ sống sót của bệnh nhân tương tự trong nghiên cứu. Vì khoảng này chứa giá trị 0%, chúng tôi không có đủ thông tin để kết luận liệu thuốc làm loãng máu có giúp ích hay gây hại cho bệnh nhân bị nhồi máu cơ tim đã được CPR.

6.2 KHÁC BIỆT GIỮA HAI TỶ LỆ

BÀI TẬP THỰC HÀNH 6.13

Một thí nghiệm kéo dài 5 năm đã được thực hiện để đánh giá hiệu quả của dầu cá trong việc giảm các biến cố tim mạch, trong đó mỗi người tham gia được chia ngẫu nhiên vào một trong hai nhóm điều trị. Chúng ta sẽ xem xét kết quả nhồi máu cơ tim ở các bệnh nhân này:

	Nhồi máu	Không có	Tổng
Dầu cá	145	12788	12933
placebo	200	12738	12938

Hãy tạo khoảng tin cậy 95% cho tác động của dầu cá lên tỷ lệ nhồi máu cơ tim ở những bệnh nhân có đặc điểm tương tự với các đối tượng trong nghiên cứu. Đồng thời, hãy diễn giải khoảng này trong bối cảnh nghiên cứu. [16]

6.2.2 Kiểm định giả thuyết cho sự khác biệt giữa hai tỷ lệ

Mammogram là một quy trình chụp X-quang để phát hiện ung thư vú. Việc có nên sử dụng mammogram hay không vẫn còn gây tranh cãi, và đây là chủ đề trong ví dụ tiếp theo, nơi chúng ta tìm hiểu về kiểm định giả thuyết hai tỷ lệ khi H_0 là $p_1 - p_2 = 0$ (hoặc tương đương là $p_1 = p_2$).

Một nghiên cứu kéo dài 30 năm đã được tiến hành với gần 90000 phụ nữ tham gia. Trong giai đoạn sàng lọc kéo dài 5 năm, mỗi phụ nữ được phân ngẫu nhiên vào một trong hai nhóm: ở nhóm 1, phụ nữ được chụp mammogram định kỳ để sàng lọc ung thư vú; ở nhóm 2, phụ nữ được khám vú định kỳ không dùng mammogram. Không có biện pháp can thiệp nào được thực hiện trong 25 năm tiếp theo. Chúng ta sẽ xem xét tử vong do ung thư vú trong toàn bộ 30 năm. Kết quả được tóm tắt trong Hình 6.2.

Nếu mammogram hiệu quả hơn nhiều so với khám thông thường, ta kỳ vọng sẽ có nhiều ca tử vong hơn trong nhóm đối chứng. Ngược lại, nếu mammogram kém hiệu quả hơn, ta sẽ thấy nhiều ca tử vong hơn trong nhóm được chụp mammogram.

	Tử vong do ung thư vú?	
	Có	Không
Mammogram	500	44425
Đối chứng	505	44405

Hình 6.2: Tóm tắt kết quả cho nghiên cứu ung thư vú.

BÀI TẬP THỰC HÀNH 6.13

Đây là một nghiên cứu thực nghiệm hay nghiên cứu quan sát? [17]

[16] Vì các bệnh nhân được chia ngẫu nhiên, các đối tượng được xem là độc lập, cả trong nhóm lẫn giữa hai nhóm. Điều kiện thành công - thất bại cũng được thỏa mãn đối với cả hai nhóm vì mọi số đếm đều từ 10 trở lên. Điều này đáp ứng các điều kiện cần thiết để mô hình hóa sự khác biệt giữa các tỷ lệ bằng phân phối chuẩn.

Tính các tỷ lệ mẫu $(\hat{p}_{dầu\ cá} = 0{,}0112, ; \hat{p}_{placebo} = 0{,}0155)$, Ước lượng điểm khác biệt $(0{,}0112 - 0{,}0155 = -0{,}0043)$, và sai số chuẩn $\left(SE = \sqrt{\frac{0{,}0112 \times 0{,}9888}{12933} + \frac{0{,}0155 \times 0{,}9845}{12938}} = 0{,}00145\right)$. Áp dụng công thức khoảng tin cậy chung với mức tin cậy 95% và $z^\star = 1{,}96$:

$$-0{,}0043 \pm 1{,}96 \times 0{,}00145 \rightarrow (-0{,}0071;\ -0{,}0015)$$

Chúng ta tin tưởng 95% rằng dầu cá làm giảm nguy cơ nhồi máu cơ tim từ 0,15 đến 0,71 điểm phần trăm (so với tỷ lệ cơ bản khoảng 1,55%) trong vòng 5 năm, đối với các đối tượng tương tự như trong nghiên cứu. Vì toàn bộ khoảng tin cậy nằm dưới 0, dữ liệu cung cấp bằng chứng mạnh rằng bổ sung dầu cá giảm nguy cơ nhồi máu cơ tim ở những bệnh nhân giống như trong nghiên cứu.

[17] Đây là một thí nghiệm. Bệnh nhân đã được phân ngẫu nhiên để nhận mammogram hoặc khám vú tiêu chuẩn không dùng mammogram. Nhờ vậy, chúng ta có thể rút ra các kết luận nhân quả từ nghiên cứu này.

6.2 KHÁC BIỆT GIỮA HAI TỶ LỆ

BÀI TẬP THỰC HÀNH 6.15

Thiết lập các giả thuyết để kiểm định xem có sự khác biệt về tỷ lệ tử vong do ung thư vú giữa nhóm mammogram và nhóm đối chứng hay không.[18]

Trong Ví dụ 6.16, chúng ta sẽ kiểm tra các điều kiện để sử dụng phân phối chuẩn nhằm phân tích kết quả của nghiên cứu. Các chi tiết rất giống với khi xây dựng khoảng tin cậy. Tuy nhiên, khi giả thuyết không là $p_1 - p_2 = 0$, ta sử dụng một tỷ lệ đặc biệt gọi là **tỷ lệ gộp (pooled proportion)** để kiểm tra điều kiện thành công - thất bại:

$$\begin{aligned} p_{pooled} &= \frac{\#\ Số\ bệnh\ nhân\ tử\ vong\ do\ ung\ thư\ vú\ trong\ nghiên\ cứu}{\#\ Số\ bệnh\ nhân\ trong\ nghiên\ cứu} \\ &= \frac{500 + 505}{500 + 44425 + 505 + 44405} \\ &= 0{,}0112 \end{aligned}$$

Đây là ước lượng tốt nhất cho tỷ lệ tử vong do ung thư vú trên toàn bộ nghiên cứu, và là ước lượng tốt nhất của $p_{Mammogram}$ và $p_{\text{Đối chứng}}$ nếu giả thuyết không là đúng là $p_{Mammogram} = p_{\text{Đối chứng}}$. Chúng ta cũng sử dụng tỷ lệ này cũng để tính sai số chuẩn.

VÍ DỤ 6.16

Có hợp lý không khi mô hình hóa sự khác biệt trong tỷ lệ bằng phân phối chuẩn trong nghiên cứu này?

E

Vì bệnh nhân được phân nhóm ngẫu nhiên, họ có thể được xem xét độc lập, cả trong và giữa các nhóm. Chúng ta cũng phải kiểm tra điều kiện thành công-thất bại cho mỗi nhóm. Theo giả thuyết không, tỷ lệ $p_{Mammogram}$ và $p_{\text{Đối chứng}}$ bằng nhau, vì vậy chúng ta kiểm tra điều kiện thành công-thất bại với ước tính tốt nhất của chúng ta về các giá trị này theo H_0, tỷ lệ gộp từ hai mẫu, $\hat{p}_{pooled} = 0{,}0112$:

$\hat{p}_{pooled} \times n_{Mammogram} = 0{,}0112 \times 44925 = 503$
$\hat{p}_{pooled} \times n_{\text{Đối chứng}} = 0{,}0112 \times 44910 = 503$

$(1 - \hat{p}_{pooled}) \times n_{Mammogram} = 0{,}9888 \times 44925 = 44422$
$(1 - \hat{p}_{pooled}) \times n_{\text{Đối chứng}} = 0{,}9888 \times 44910 = 44407$

Vì tất cả các giá trị đều ≥ 10, điều kiện thành công - thất bại được thỏa mãn. Với cả hai điều kiện được đáp ứng, chúng ta có thể mô hình hóa sự khác biệt tỷ lệ bằng phân phối chuẩn.

SỬ DỤNG TỶ LỆ GỘP KHI H_0 LÀ $p_1 - p_2 = 0$

Khi giả thuyết không (H_0) cho rằng hai tỷ lệ bằng nhau, ta sử dụng tỷ lệ gộp $(\hat{p}_{pooled})$ để kiểm tra điều kiện thành công - thất bại và ước lượng sai số chuẩn:

$$\hat{p}_{pooled} = \frac{Số\ trường\ hợp\ "thành\ công"}{Tổng\ số\ trường\ hợp} = \frac{\hat{p}_1 n_1 + \hat{p}_2 n_2}{n_1 + n_2}$$

Trong đó, $\hat{p}_1 n_1$ đại diện cho số trường hợp thành công ở mẫu 1, tính theo công thức:

$$\hat{p}_1 = \frac{Số\ trường\ hợp\ \text{thành công}\ ở\ mẫu\ 1}{n_1}$$

Tương tự, $\hat{p}_2 n_2$ đại diện cho số trường hợp thành công ở mẫu 2.

Trong Ví dụ 6.16, tỷ lệ gộp được sử dụng để kiểm tra điều kiện thành công - thất bại. [19] Ở ví dụ tiếp theo, chúng ta sẽ thấy tỷ lệ gộp được sử dụng ở chỗ thứ hai: tính sai số chuẩn

[18] H_0: Tỷ lệ tử vong do ung thư vú ở bệnh nhân được sàng lọc bằng chụp nhũ ảnh không khác so với tỷ lệ tử vong ở nhóm đối chứng, nghĩa là: $p_{Mammogram} - p_{\text{Đối chứng}} = 0$.
H_A: Tỷ lệ tử vong do ung thư vú ở bệnh nhân được sàng lọc bằng chụp nhũ ảnh khác so với tỷ lệ tử vong ở nhóm đối chứng, nghĩa là:, $p_{Mammogram} - p_{\text{Đối chứng}} \neq 0$.

[19] Để xem một ví dụ về kiểm định giả thuyết hai tỷ lệ mà không cần điều kiện thành công - thất bại, hãy xem Mục 2.3.

6.2 KHÁC BIỆT GIỮA HAI TỶ LỆ

VÍ DỤ 6.17

Tính ước lượng điểm của sự khác biệt tỷ lệ tử vong do ung thư vú giữa hai nhóm, và sử dụng tỷ lệ gộp ước lượng $\hat{p}_{pooled} = 0{,}0112$ để tính sai số chuẩn.

Ước lượng điểm của sự khác biệt tỷ lệ tử vong do ung thư vú là:

$$p_{Mammogram} - p_{\text{Đối chứng}} = \frac{500}{500+44425} - \frac{505}{505+44405} = 0{,}01113 - 0{,}01125 = -0{,}00012$$

Tỷ lệ tử vong do ung thư vú ở nhóm Mammogram là 0,012%, thấp hơn nhóm đối chứng. Tiếp theo, sai số chuẩn được tính bằng cách *sử dụng tỷ lệ gộp*, $\hat{p}_{pooled}$:

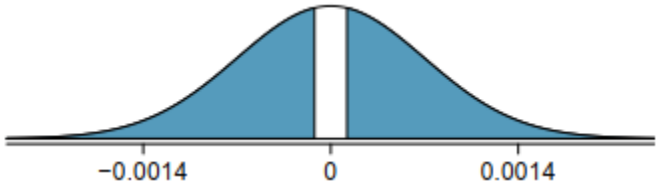
$$SE = \sqrt{\frac{\hat{p}_{pooled}(1-\hat{p}_{pooled})}{n_{Mammogram}} - \frac{\hat{p}_{pooled}(1-\hat{p}_{pooled})}{n_{\text{Đối chứng}}}} = 0{,}00070$$

VÍ DỤ 6.18

Sử dụng ước lượng điểm $p_{Mammogram} - p_{\text{Đối chứng}} = -0{,}00012$ và sai số chuẩn $SE = 0{,}00070$, hãy tính giá trị p cho kiểm định giả thuyết và viết kết luận.

Giống như các kiểm định trước đây, ta đầu tiên tính thống kê kiểm định và vẽ biểu đồ minh họa:

$$Z = \frac{\text{ước lượng điểm} - \text{giá trị không}}{SE} = \frac{-0{,}00012 - 0}{0{,}00070} = -0{,}17$$



Diện tích phía dưới đuôi bên trái là 0,4325, nhân đôi để được p-value = 0,8650. Vì p-value > 0,05, chúng ta không bác bỏ giả thuyết không H_0. Nghĩa là, sự khác biệt về tỷ lệ tử vong do ung thư vú có thể giải thích bằng sự ngẫu nhiên. Chúng ta không quan sát thấy lợi ích hay tác hại rõ ràng từ sử dung phương pháp Mammogramso với khám vú thông thường.

Chúng ta có thể kết luận rằng phương pháp Mammogramso không có lợi hay hại không? Dưới đây là một vài điểm cần cân nhắc khi đánh giá nghiên cứu chụp nhũ ảnh cũng như các nghiên cứu y khoa khác:

- Chúng ta không bác bỏ giả thuyết không, nghĩa là không có đủ bằng chứng để kết luận rằng phương pháp Mammogramsolàm giảm hay tăng tỷ lệ tử vong do ung thư vú.
- Nếu phương pháp Mammogramso thực sự có lợi hoặc gây hại, dữ liệu cho thấy tác động này không lớn.
- Phương pháp Mammogramso có chi phí cao hơn hay thấp hơn khám vú thông thường? Nếu một phương pháp đắt hơn mà không mang lại lợi ích rõ ràng, thì nên ưu tiên phương pháp ít tốn kém hơn.
- Tác giả nghiên cứu cũng phát hiện rằng phương pháp Mammogramso dẫn đến chẩn đoán quá mức bệnh ung thư vú. Tức là, một số trường hợp ung thư được phát hiện (hoặc nghi ngờ phát hiện) thực ra sẽ không gây triệu chứng gì trong suốt đời bệnh nhân. Nghĩa là, bệnh nhân có thể bị điều trị không cần thiết, gây tốn kém và tác hại về thể chất hoặc tinh thần không đáng có.
- Những cân nhắc này cho thấy sự phức tạp của chăm sóc y tế và khuyến nghị điều trị. Các chuyên gia và hội đồng y khoa sử dụng các yếu tố như trên để đưa ra khuyến nghị tốt nhất dựa trên bằng chứng hiện tại.

6.2 KHÁC BIỆT GIỮA HAI TỶ LỆ

6.2.3 Thêm về kiểm định giả thuyết hai tỷ lệ (chủ đề đặc biệt)

Khi tiến hành kiểm định giả thuyết hai tỷ lệ, thông thường H_0 là $p_1 - p_2 = 0$. Tuy nhiên, có những trường hợp hiếm, chúng ta muốn kiểm tra sự khác biệt giữa p_1 và p_2 bằng một giá trị khác 0.

Ví dụ, có thể chúng ta muốn kiểm định giả thuyết $p_1 - p_2 = 0{,}1$. Trong những tình huống như vậy, ta thường sử dụng $\hat{p}_1$ và $\hat{p}_2$ để kiểm tra điều kiện thành công - thất bại và xây dựng sai số chuẩn.

BÀI TẬP THỰC HÀNH 6.19

Một công ty sản xuất máy bay không người lái (quadcopter) đang cân nhắc sử dụng một nhà sản xuất mới cho cánh quạt. Nhà sản xuất mới có chi phí cao hơn, nhưng họ tuyên bố rằng tỷ lệ cánh quạt đạt tiêu chuẩn của họ cao hơn đối thủ 3%.[20]



Hình 6.3: Một chiếu quadcopter Phantom

Ảnh bởi David J (http://flic.kr/p/oiWLNu). CC-BY 2.0 license.
Ảnh đã được cắt và thêm viền.

[20] H_0: Cánh quạt chất lượng cao sẽ vượt qua kiểm định thường xuyên hơn 3% so với cánh quạt chất lượng tiêu chuẩn. $p_{chất\ lượng\ cao} - p_{tiêu\ chuẩn} = 0{,}03$. H_A: Cánh quạt chất lượng cao sẽ vượt qua kiểm định với mức chênh lệch khác 3% so với cánh quạt chất lượng tiêu chuẩn. $p_{chất\ lượng\ cao} - p_{tiêu\ chuẩn} \neq 0{,}03$.

6.2 KHÁC BIỆT GIỮA HAI TỶ LỆ

VÍ DỤ 6.20

Kỹ sư kiểm soát chất lượng trong Bài tập có hướng dẫn 6.19 đã thu thập một mẫu cánh quạt, kiểm tra 1000 cánh từ mỗi công ty. Kết quả cho thấy 899 cánh của nhà cung cấp hiện tại đạt kiểm định, và 958 cánh của nhà cung cấp tiềm năng đạt kiểm định. Hãy đánh giá các giả thuyết từ Bài tập Hướng dẫn 6.19 với mức ý nghĩa 5%.

Trước tiên, chúng ta kiểm tra các điều kiện. Mẫu không nhất thiết được chọn ngẫu nhiên, do đó để tiếp tục, ta phải giả định các cánh quạt là độc lập; đối với mẫu này, ta cho rằng giả định này là hợp lý, nhưng kỹ sư sẽ hiểu rõ hơn liệu điều đó có đúng hay không. Điều kiện thành công-thất bại cũng được thoả mãn với mỗi mẫu. Do đó, chênh lệch giữa hai tỷ lệ mẫu, $0{,}958 - 0{,}899 = 0{,}059$, có thể xem là đến từ một phân phối gần chuẩn.

Độ lệch chuẩn được tính bằng hai tỷ lệ mẫu bởi vì trong bối cảnh này ta không sử dụng tỷ lệ gộp:

(E)

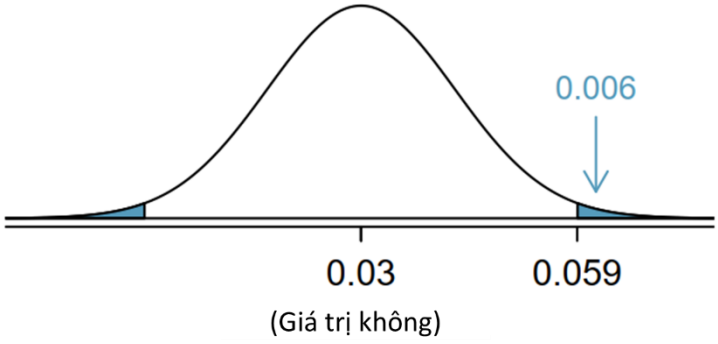
$$SE = \sqrt{\frac{0{,}958(1-0{,}958)}{1000} + \frac{0{,}899(1\ -\ 0{,}899)}{1000}} = 0{,}0114$$

Trong kiểm định giả thuyết này, vì giả thuyết không là $p_1 - p_2 = 0{,}03$, nên tỷ lệ mẫu được dùng để tính độ lệch chuẩn thay vì dùng tỷ lệ gộp.

Tiếp theo, ta tính thống kê kiểm định và dùng nó để tìm p-value, được minh họa ở Hình 6.4.

$$Z = \frac{\text{Ước lượng điểm} - \text{Giá trị không}}{SE} = \frac{0{,}059 - 0{,}03}{0{,}0114} = 2{,}54$$

Sử dụng phân phối chuẩn chuẩn hóa cho thống kê kiểm định này, ta xác định phần diện tích phía bên phải là 0.006, và nhân đôi nó để ra p-value: 0,012. Ta bác bỏ giả thuyết không vì $0{,}012 < 0{,}05$. Vì ta quan sát thấy tỷ lệ vượt kiểm định cao hơn 3%, ta có bằng chứng thống kê có ý nghĩa cho thấy cánh quạt chất lượng cao vượt kiểm *định nhiều hơn* 3% so với cánh quạt đang sử dụng, vượt mức công ty đã tuyên bố.



Hình 6.4: Phân phối của thống kê kiểm định nếu giả thuyết không là đúng. P-value được thể hiện bằng các vùng được tô xanh.

6.2 KHÁC BIỆT GIỮA HAI TỶ LỆ

6.2.4 Xem xét công thức độ lệch chuẩn (chủ đề đặc biệt)

Phần phụ này trình bày các chủ đề lý thuyết chuyên sâu giúp hiểu rõ nguồn gốc của công thức độ lệch chuẩn cho chênh lệch hai tỷ lệ. Cuối cùng, tất cả các công thức độ lệch chuẩn mà ta gặp trong chương này và chương 7 đều có thể được suy ra từ các nguyên lý xác suất trong Mục 3.4.

Công thức độ lệch chuẩn của chênh lệch hai tỷ lệ có thể được phân tích từ các công thức độ lệch chuẩn của từng tỷ lệ mẫu riêng lẻ. Nhớ rằng độ lệch chuẩn của các tỷ lệ mẫu riêng p_1 và p_2 là:

$$SE_{\hat{p}_1} = \sqrt{\frac{p_1(1-p_1)}{n_1}} \qquad SE_{\hat{p}_2} = \sqrt{\frac{p_2(1-p_2)}{n_2}}$$

Độ lệch chuẩn của chênh lệch hai tỷ lệ mẫu có thể được xây dựng từ độ lệch chuẩn của các tỷ lệ mẫu riêng:

$$SE_{\hat{p}_1-\hat{p}_2} = \sqrt{SE_{\hat{p}_1}^2 + SE_{\hat{p}_2}^2} = \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$$

Mối quan hệ đặc biệt này là hệ quả từ lý thuyết xác suất.

BÀI TẬP THỰC HÀNH 6.21

Yêu cầu kiến thức trước: Mục 3.4. Ta có thể viết lại phương trình trên theo cách khác:

$$SE_{\hat{p}_1-\hat{p}_2}^2 = SE_{\hat{p}_1}^2 + SE_{\hat{p}_2}^2$$

Hãy giải thích công thức này đến từ đâu bằng cách sử dụng công thức tính phương sai của tổng hai biến ngẫu nhiên. [21]

[21] Bình phương của sai số chuẩn biểu diễn phương sai của ước lượng. Nếu X và Y là hai biến ngẫu nhiên có phương sai lần lượt là σ_x^2 và σ_y^2 thì phương sai của $X - Y$ là $\sigma_x^2 + \sigma_y^2$. Tương tự, phương sai tương ứng với $\hat{p}_1 - \hat{p}_2$ là $\sigma_{\hat{p}_1}^2 + \sigma_{\hat{p}_2}^2$. Vì $\sigma_{\hat{p}_1}^2$ và $\sigma_{\hat{p}_2}^2$ chỉ là cách viết khác của $SE_{\hat{p}_1}^2$ và $SE_{\hat{p}_2}^2$, nên phương sai tương ứng với $\hat{p}_1 - \hat{p}_2$ có thể được viết là $SE_{\hat{p}_1}^2 + SE_{\hat{p}_2}^2$

6.2 KHÁC BIỆT GIỮA HAI TỶ LỆ

Bài tập

6.17 Thí nghiệm xã hội, Phần I. Một "thí nghiệm xã hội" được thực hiện bởi một chương trình truyền hình đã đặt câu hỏi về việc người ta làm gì khi chứng kiến một người phụ nữ bị bầm tím rõ ràng bị bạn trai của cô ta quấy rối. Tại cùng một nhà hàng, cùng một cặp đôi được thể hiện trong hai tình huống khác nhau: một lần người phụ nữ ăn mặc "khiêu gợi", và một lần ăn mặc "kín đáo". Bảng dưới đây cho biết số lượng người có mặt tại nhà hàng và việc họ có can thiệp hay không.

		Tình huống		
		Khiêu gợi	Kín đáo	Tổng
Can thiệp	Có	5	15	20
	Không	15	10	25
	Tổng	20	25	45

Giải thích vì sao phân phối lấy mẫu của sự khác biệt tỷ lệ can thiệp giữa hai tình huống (ăn mặc khiêu gợi và kín đáo) không tuân theo phân phối gần chuẩn.

6.18 Thành công trong ghép tim. Nghiên cứu ghép tim tại Đại học Stanford được thực hiện để xác định liệu chương trình ghép tim thử nghiệm có giúp kéo dài tuổi thọ không. Các bệnh nhân được phân ngẫu nhiên vào hai nhóm: điều trị (được ghép tim) và kiểm soát (không ghép tim). Dữ liệu như sau. [22]

	kiểm soát	điều trị
sống	4	24
tử vong	30	45

Giả sử bạn muốn ước lượng sự khác biệt trong tỷ lệ sống sót giữa hai nhóm bằng khoảng tin cậy. Giải thích tại sao không thể xây dựng khoảng tin cậy sử dụng xấp xỉ phân phối chuẩn. Điều gì có thể xảy ra sai nếu bạn vẫn tiếp tục xây dựng khoảng tin cậy mặc cho vấn đề này?

6.19 Giới tính và sở thích màu sắc. Một nghiên cứu khảo sát 1.924 nam và 3.666 nữ sinh viên đại học về màu sắc yêu thích. Khoảng tin cậy 95% cho sự khác biệt giữa tỷ lệ nam và nữ thích màu đen (pnam − pnữ là (0.02, 0.06). Dựa trên thông tin này, hãy xác định xem các phát biểu sau về sinh viên đại học có đúng hay sai, và giải thích lý do với mỗi phát biểu bạn cho là sai. [23]

a) Chúng tôi tự tin 95% rằng tỷ lệ thực tế của nam giới mà màu yêu thích là màu đen thấp hơn 2% đến cao hơn 6% so với tỷ lệ thực tế của nữ giới có màu yêu thích là màu đen.
b) Chúng tôi tự tin 95% rằng tỷ lệ thực tế của nam giới mà màu yêu thích là màu đen cao hơn từ 2% đến 6% so với tỷ lệ thực tế của nữ giới có màu yêu thích là màu đen.
c) 95% các mẫu ngẫu nhiên sẽ tạo ra khoảng tin cậy 95% bao gồm sự chênh lệch thực sự giữa tỷ lệ dân số nam và nữ mà màu yêu thích là màu đen.
d) Chúng ta có thể kết luận rằng có sự khác biệt có ý nghĩa giữa tỷ lệ nam và nữ mà màu yêu thích là màu đen và sự khác biệt giữa hai tỷ lệ mẫu là quá lớn để có thể do ngẫu nhiên.
e) Khoảng tin cậy 95% cho $(p_{nữ} - p_{nam})$ không thể được tính chỉ với thông tin được cung cấp trong bài tập này.

[22] B. Turnbull et al. "Survivorship of Heart Transplant Data". In: *Journal of the American Statistical Association* 69 (1974), pp. 74–80.
[23] L Ellis and C Ficek. "Color preferences according to gender and sexual orientation". In: *Personality and Individual Differences* 31.8 (2001), pp. 1375–1379.

6.2 KHÁC BIỆT GIỮA HAI TỶ LỆ

6.20 Government shutdown. Chính phủ liên bang Hoa Kỳ đã đóng cửa từ ngày 22 tháng 12 năm 2018 đến ngày 25 tháng 1 năm 2019, kéo dài 35 ngày. Một cuộc khảo sát của Survey USA với 614 người Mỹ được chọn ngẫu nhiên trong khoảng thời gian này báo cáo rằng 48% những người kiếm dưới $40000/năm và 55% những người kiếm từ $40000 trở lên/năm cho biết việc chính phủ đóng cửa không ảnh hưởng gì đến họ. Một khoảng tin cậy 95% cho $(p < 40K\ -\ p \geq 40K)$, trong đó p là tỷ lệ những người nói rằng việc chính phủ đóng cửa không ảnh hưởng gì đến họ, là $(-0{,}16;\ 0{,}02)$. Dựa trên thông tin này, xác định xem các phát biểu sau là đúng hay sai, và giải thích lý do nếu bạn cho rằng phát biểu đó sai. [24]

a) Ở mức ý nghĩa 5%, dữ liệu cung cấp bằng chứng thuyết phục về sự khác biệt thực sự trong tỷ lệ những người không bị ảnh hưởng cá nhân giữa người Mỹ kiếm dưới $40000/năm và người kiếm $40000/năm.
b) Chúng tôi tự tin 95% rằng có từ 16% nhiều hơn đến 2% ít hơn người Mỹ kiếm dưới $40000/năm không bị ảnh hưởng cá nhân bởi việc chính phủ đóng cửa so với những người kiếm từ $40000 trở lên mỗi năm.
c) Một khoảng tin cậy 90% cho $(p < 40K\ -\ p \geq 40K)$ sẽ rộng hơn khoảng $(-0{,}16;\ 0{,}02)$.
d) Một khoảng tin cậy 95% cho $(p \geq 40K\ -\ p < 40K)$ là $(-0{,}02;\ 0{,}16)$.

6.21 Kế hoạch y tế quốc gia, Phần III. Ví dụ 6.11 trình bày kết quả một cuộc khảo sát đánh giá mức độ ủng hộ đối với một "Kế hoạch Y tế Quốc gia" được gắn nhãn chung tại Hoa Kỳ. 79% trong số 347 người theo Đảng Dân chủ và 55% trong số 617 người độc lập ủng hộ kế hoạch này.

a) Hãy tính khoảng tin cậy 95% cho sự khác biệt giữa tỷ lệ người Dân chủ và người Độc lập ủng hộ Kế hoạch Y tế Quốc gia $(p_D\ -\ p_I)$, và giải thích kết quả trong bối cảnh này. Các điều kiện đã được kiểm tra sẵn.
b) Đúng hay sai: Nếu chúng ta chọn ngẫu nhiên một người Dân chủ và một người Độc lập vào thời điểm khảo sát, thì khả năng người Dân chủ ủng hộ Kế hoạch Y tế Quốc gia cao hơn là đúng.

6.22 Thiếu ngủ, CA vs. OR, Phần I. Theo một báo cáo của Trung tâm Kiểm soát và Phòng ngừa Dịch bệnh (CDC) về tình trạng thiếu ngủ, tỷ lệ người dân California cho biết họ không được nghỉ ngơi hoặc ngủ đầy đủ trong mỗi ngày của 30 ngày gần nhất là 8,0%, trong khi tỷ lệ này ở Oregon là 8,8%. Dữ liệu này dựa trên các mẫu ngẫu nhiên đơn gồm 11545 cư dân California và 4691 cư dân Oregon. Hãy tính khoảng tin cậy 95% cho sự khác biệt giữa tỷ lệ người thiếu ngủ ở California và Oregon, và giải thích kết quả trong bối cảnh dữ liệu. [25]

6.23 Khoan dầu ngoài khơi, Phần I. Một cuộc khảo sát đã hỏi 827 cử tri đăng ký tại California theo mẫu ngẫu nhiên: "Bạn có ủng hộ hay phản đối việc khoan dầu và khí tự nhiên ngoài khơi bờ biển California không? Hay bạn không biết đủ để đưa ra ý kiến?" Dưới đây là phân phối các câu trả lời, được phân chia dựa trên việc người trả lời có tốt nghiệp đại học hay không. [26]

a) Bao nhiêu phần trăm người tốt nghiệp đại học và bao nhiêu phần trăm người không tốt nghiệp đại học trong mẫu này không có đủ thông tin để đưa ra ý kiến về việc khoan dầu và khí tự nhiên ngoài khơi bờ biển California?
b) Tiến hành kiểm định giả thuyết để xác định liệu dữ liệu có cung cấp bằng chứng mạnh mẽ cho thấy tỷ lệ người tốt nghiệp đại học không có ý kiến khác với người không tốt nghiệp đại học hay không.

	Tốt nghiệp ĐH	
	Có	Không
Ủng hộ	154	132
Phản đối	180	126
Không biết	104	131
Tổng	438	389

6.24 Thiếu ngủ, CA vs. OR, Phần II. Ví dụ 6.22 cung cấp dữ liệu về tỷ lệ thiếu ngủ ở California và Oregon. Tỷ lệ người dân California báo cáo thiếu nghỉ ngơi hoặc giấc ngủ trong mỗi ngày của 30 ngày gần nhất là 8,0%, trong khi tỷ lệ này ở Oregon là 8,8%. Dữ liệu được thu thập từ các mẫu ngẫu nhiên đơn gồm 11545 cư dân California và 4691 cư dân Oregon.

a) Tiến hành kiểm định giả thuyết để xác định liệu dữ liệu này có cung cấp bằng chứng mạnh mẽ rằng tỷ lệ thiếu ngủ giữa hai bang là khác nhau hay không. (Lưu ý: hãy kiểm tra các điều kiện cần thiết)
b) Có khả năng kết luận của kiểm định ở phần (a) là sai. Nếu điều đó xảy ra, thì đã mắc loại sai lầm nào?

[24] Survey USA, News Poll #24568, data collected on April 21, 2019.
[25] CDC, Perceived Insufficient Rest or Sleep Among Adults — United States, 2008.
[26] Survey USA, Election Poll #16804, data collected July 8-11, 2010.

6.2 KHÁC BIỆT GIỮA HAI TỶ LỆ

6.25 Khoan dầu ngoài khơi, Phần II. Kết quả một cuộc khảo sát đánh giá mức độ ủng hộ việc khoan dầu và khí tự nhiên ngoài khơi bờ biển California đã được giới thiệu trong Bài tập 6.23.

	Tốt nghiệp đại học	
	Có	Không
Ủng hộ	154	132
Phản đối	180	126
Không biết	104	131
Tổng	438	389

a) Bao nhiêu phần trăm người tốt nghiệp đại học và bao nhiêu phần trăm người không tốt nghiệp đại học trong mẫu này ủng hộ việc khoan dầu và khí tự nhiên ngoài khơi bờ biển California?

b) Tiến hành kiểm định giả thuyết để xác định liệu dữ liệu có cung cấp bằng chứng mạnh mẽ cho thấy tỷ lệ người tốt nghiệp đại học ủng hộ việc khoan ngoài khơi khác biệt với tỷ lệ của người không tốt nghiệp hay không.

6.26 Quét toàn thân, Phần I. Một bài báo đưa tin rằng "Người Mỹ có quan điểm khác nhau về hai biện pháp tiềm ẩn sự bất tiện và xâm lấn mà sân bay có thể áp dụng để phát hiện các cuộc tấn công khủng bố tiềm tàng." Bài viết này dựa trên một khảo sát được thực hiện với một mẫu ngẫu nhiên gồm 1137 người trưởng thành trên toàn quốc, trong đó một câu hỏi được đưa ra là: "Một số sân bay hiện đang sử dụng các máy quét x-quang kỹ thuật số toàn thân để kiểm tra hành khách bằng điện tử "?" Tóm tắt câu trả lời dựa trên đảng phái chính trị. [27]

		Đảng phái		
		Cộng hoà	Dân chủ	Độc lập
Trả lời	Nên	264	299	351
	Không nên	38	55	77
	Không biết/không trả lời	16	15	22
	Tổng	318	369	450

a) Tiến hành kiểm định giả thuyết phù hợp để đánh giá xem có sự khác biệt về tỷ lệ người thuộc Đảng Cộng hòa và Đảng Dân chủ cho rằng việc áp dụng máy quét toàn thân tại sân bay là nên làm hay không. Giả sử tất cả các điều kiện cần thiết đều được đáp ứng.

b) Kết luận của kiểm định ở phần (a) có thể sai, nghĩa là đã mắc lỗi kiểm định. Nếu có lỗi xảy ra, đó là lỗi loại I hay loại II? Giải thích.

6.27 Người làm việc trong ngành giao thông thiếu ngủ. Tổ chức National Sleep Foundation đã thực hiện một khảo sát về thói quen ngủ của một mẫu ngẫu nhiên người lao động trong ngành giao thông và một mẫu đối chứng là người không làm việc trong ngành giao thông. Kết quả khảo sát được trình bày như sau. [28]

		Người lao động trong ngành giao thông			
	Đối chứng	Phi công	Lái xe tải	Lái tài	Lái bus/ taxi/limo
Dưới 6 giờ	35	19	35	29	21
6 đến 8 giờ	193	132	117	119	131
Hơn 8 giờ	64	51	51	32	58
Tổng	292	202	203	180	210

Tiến hành kiểm định giả thuyết để đánh giá liệu dữ liệu này có bằng chứng cho thấy có sự khác biệt về tỷ lệ tài xế xe tải và nhóm đối chứng ngủ ít hơn 6 giờ mỗi ngày, tức là bị thiếu ngủ.

[27] S. Condon. "Poll: 4 in 5 Support Full-Body Airport Scanners". In: *CBS News* (2010).
[28] National Sleep Foundation, 2012 Sleep in America Poll: Transportation Workers' Sleep, 2012.

6.2 KHÁC BIỆT GIỮA HAI TỶ LỆ

6.28 Vitamin trước sinh và chứng tự kỷ. Các nhà nghiên cứu về mối liên hệ giữa việc dùng vitamin trước sinh và chứng tự kỷ đã khảo sát mẹ của trẻ từ 24–60 tháng tuổi bị tự kỷ và một mẫu ngẫu nhiên khác của trẻ phát triển bình thường. Bảng dưới đây cho thấy số lượng người mẹ trong mỗi nhóm có và không sử dụng vitamin trong ba tháng trước khi mang thai (giai đoạn quanh thụ thai). [29]

		Tự kỷ		
		Tự kỷ	Phát triển bình thường	Tổng
Vitamin trước khi sinh dành cho phụ nữ mang thai	Không vitamin	111	70	181
	Vitamin	143	159	302
	Tổng	254	229	483

a) Nêu giả thuyết phù hợp để kiểm định mối liên quan giữa việc dùng vitamin trước khi mang thai và chứng tự kỷ.
b) Thực hiện kiểm định giả thuyết và nêu kết luận phù hợp. (Lưu ý: xác minh các điều kiện cần thiết.)
c) Một bài báo của New York Times đưa tin về nghiên cứu này với tiêu đề: "Prenatal Vitamins May Ward Off Autism" (Vitamin trước khi sinh có thể ngăn ngừa chứng tự kỷ). Bạn có cho rằng tiêu đề bài báo này là phù hợp không? Giải thích và đề xuất một tiêu đề thay thế. [30]

6.29 HIV ở vùng hạ Sahara châu Phi. Vào tháng 7/2008, Viện Y tế Quốc gia Hoa Kỳ đã thông báo dừng sớm một nghiên cứu lâm sàng do kết quả bất ngờ. Nhóm nghiên cứu bao gồm phụ nữ nhiễm HIV ở châu Phi hạ Sahara được cho uống một liều Nevaripine duy nhất khi sinh con để ngăn truyền HIV sang trẻ. Nghiên cứu so sánh ngẫu nhiên giữa việc tiếp tục điều trị bằng Nevaripine sau sinh và điều trị bằng Lopinavir (một thuốc điều trị HIV khác). Có 240 phụ nữ tham gia nghiên cứu, mỗi nhóm gồm 120 người. 24 tuần sau khi bắt đầu điều trị, mỗi người được kiểm tra xem bệnh HIV có trở nên nặng hơn không (gọi là thất bại virus học). 26 trong số 120 người dùng Nevaripine bị thất bại virus học, trong khi chỉ có 10 trong số 120 người dùng thuốc còn lại gặp tình trạng tương tự. [31]

a) Lập bảng hai chiều trình bày kết quả nghiên cứu.
b) Nêu giả thuyết phù hợp để kiểm định sự khác biệt về tỷ lệ thất bại virus học giữa hai nhóm điều trị.
c) Thực hiện kiểm định giả thuyết và nêu kết luận phù hợp. (Lưu ý: xác minh các điều kiện cần thiết.)

6.30 Một quả táo mỗi ngày giúp tránh xa bác sĩ. Một giáo viên thể chất tại một trường trung học muốn nâng cao nhận thức về dinh dưỡng và sức khỏe đã hỏi học sinh vào đầu học kỳ rằng họ có tin câu nói "mỗi ngày ăn một quả táo giúp tránh gặp bác sĩ" không, và 40% học sinh trả lời "có". Suốt học kỳ, cô bắt đầu mỗi buổi học bằng một cuộc thảo luận ngắn về các nghiên cứu cho thấy lợi ích của việc ăn nhiều rau và trái cây. Cuối học kỳ, cô thực hiện lại khảo sát và lần này 60% học sinh trả lời "có". Cô có thể sử dụng phương pháp so sánh hai tỷ lệ để phân tích trường hợp này không? Giải thích lý do.

[29] R.J. Schmidt et al. "Prenatal vitamins, one-carbon metabolism gene variants, and risk for autism". In: *Epidemi- ology* 22.4 (2011), p. 476.
[30] R.C. Rabin. "Patterns: Prenatal Vitamins May Ward Off Autism". In: *New York Times* (2011).
[31] S. Lockman et al. "Response to antiretroviral therapy after a single, peripartum dose of nevirapine". In: *Obstet- rical & gynecological survey* 62.6 (2007), p. 361.

6.3 Kiểm định độ phù hợp bằng phép kiểm định Chi-bình phương

Trong phần này, chúng ta phát triển một phương pháp để đánh giá một mô hình giả thuyết không khi dữ liệu được phân nhóm. Kỹ thuật này thường được sử dụng trong hai trường hợp:

- Khi có một mẫu các đối tượng có thể được phân loại vào nhiều nhóm, ta muốn xác định xem mẫu đó có đại diện cho quần thể chung hay không.
- Đánh giá xem dữ liệu có giống với một phân phối cụ thể nào đó hay không, chẳng hạn như phân phối chuẩn hoặc phân phối hình học.

Cả hai tình huống này đều có thể được xử lý bằng cùng một phương pháp kiểm định thống kê: kiểm định Chi-bình phương.

Trong trường hợp đầu tiên, ta xem xét dữ liệu từ một mẫu ngẫu nhiên gồm 275 thành viên bồi thẩm đoàn trong một quận nhỏ. Các thành viên này tự xác định nhóm chủng tộc của mình như Hình 6.5, và ta muốn xác định xem bồi thẩm đoàn này có đại diện về mặt chủng tộc so với quần thể chung hay không. Nếu bồi thẩm đoàn đại diện cho quần thể, thì tỷ lệ phần trăm trong mẫu nên phản ánh gần đúng tỷ lệ phần trăm trong quần thể đủ điều kiện làm bồi thẩm (tức là cử tri đã đăng ký).

Chủng tộc	Da trắng	Da đen	Hispanic	Khác	Tổng
Số lượng trong bồi thẩm đoàn	205	26	25	19	275
Tỷ lệ trong cử tri đã đăng kí	0,72	0,07	0,12	0,09	1,00

Hình 6.5: Sự phân bố theo chủng tộc trong bồi thẩm đoàn và trong quần thể cử tri của một thành phố.

Mặc dù tỷ lệ trong bồi thẩm đoàn không hoàn toàn giống tỷ lệ trong dân số, nhưng chưa rõ liệu dữ liệu này có cung cấp bằng chứng đủ thuyết phục để nói rằng mẫu không đại diện hay không. Nếu các thành viên bồi thẩm được chọn ngẫu nhiên từ cử tri đã đăng ký, ta vẫn có thể mong đợi có những khác biệt nhỏ do yếu tố ngẫu nhiên. Tuy nhiên, sự khác biệt lớn bất thường có thể là bằng chứng cho thấy bồi thẩm đoàn không đại diện.

Một ứng dụng thứ hai, đánh giá sự phù hợp của phân phối, được trình bày ở cuối phần này. Dữ liệu lợi nhuận hàng ngày từ chỉ số S&P500 trong 25 năm được dùng để đánh giá xem hoạt động chứng khoán mỗi ngày có độc lập với hành vi của những ngày trước hay không.

Trong các bài toán này, ta muốn xem xét tất cả các nhóm cùng một lúc, chứ không chỉ so sánh một hoặc hai nhóm riêng lẻ. Điều này đòi hỏi ta phải phát triển một thống kê kiểm định mới.

6.3.1 Tạo kiểm định thống kê cho bảng một chiều

VÍ DỤ 6.22

Trong số 275 người trong thành phố được chọn làm bồi thẩm đoàn, nếu các cá nhân được chọn ngẫu nhiên thì sẽ có bao nhiêu người da trắng? Bao nhiêu người da đen?

(E)

Khoảng 72% dân số là người da trắng, vì vậy ta kỳ vọng khoảng 72% số bồi thẩm viên là người da trắng: 0,72×275=198.

Tương tự, ta kỳ vọng khoảng 7% số bồi thẩm viên là người da đen, tức khoảng: $0{,}07 \times 275 = 19{,}25$ thành viên da đen.

BÀI TẬP THỰC HÀNH 6.23

(G) 12% dân số là người Hispanic và 9% thuộc các chủng tộc khác. Hãy tính xem trong 275 bồi thẩm viên thì kỳ vọng có bao nhiêu người Hispanic và bao nhiêu người thuộc chủng tộc khác khác. Đáp án có thể tìm thấy ở Hình 6.6.

Tỷ lệ mẫu của từng nhóm chủng tộc trong 275 bồi thẩm viên không hoàn toàn khớp với bất kỳ nhóm dân tộc nào. Mặc dù có sự khác biệt do lấy mẫu, nhưng ta vẫn kỳ vọng các tỷ lệ mẫu sẽ khá tương tự với tỷ lệ trong dân số, nếu không có sự thiên vị trong việc chọn bồi thẩm. Chúng ta cần kiểm định xem liệu sự khác biệt này có đủ lớn để cung cấp bằng chứng

thuyết phục rằng bồi thẩm đoàn không phải là một mẫu ngẫu nhiên hay không. Từ những ý tưởng này, có thể hình thành các giả thuyết:

Chủng tộc	Da trắng	Da đen	Hispanic	Khác	Tổng
Số liệu quan sát	205	26	25	19	275
Số liệu kì vọng	198	19,25	33	24,75	275

Hình 6.6: Thành phần thực tế và kỳ vọng của bồi thẩm đoàn.

H_0: Bồi thẩm đoàn là một mẫu ngẫu nhiên, nghĩa là không có sự thiên vị về chủng tộc trong việc chọn người tham gia bồi thẩm, và các số liệu quan sát phản ánh dao động lấy mẫu tự nhiên.

H_A: Bồi thẩm đoàn không được chọn ngẫu nhiên, nghĩa là có sự thiên vị về chủng tộc trong việc lựa chọn bồi thẩm.

Để đánh giá các giả thuyết này, ta định lượng mức độ khác biệt giữa số liệu quan sát và số liệu kỳ vọng. Bằng chứng mạnh mẽ cho giả thuyết đối sẽ đến từ những sai lệch bất thường lớn trong các nhóm so với những gì kỳ vọng chỉ dựa trên dao động ngẫu nhiên của mẫu.

6.3.2 Thống kê kiểm định Chi-bình phương

Trong các kiểm định giả thuyết trước đây, ta đã xây dựng thống kê kiểm định theo công thức:

$$\frac{\text{Ước lượng điểm } - \text{ giá trị giả thuyết không}}{SE \text{ của ước lượng điểm}}$$

Cấu trúc này dựa trên (1) Xác định sự khác biệt giữa ước lượng điểm và giá trị kỳ vọng nếu giả thuyết gốc đúng và (2) Chuẩn hóa sự khác biệt này bằng sai số chuẩn của ước lượng điểm.Hai ý tưởng này cũng sẽ được áp dụng để xây dựng thống kê kiểm định cho dữ liệu dạng đếm.

Chiến lược của chúng ta: Trước tiên tính toán sự khác biệt giữa số liệu quan sát và số liệu kỳ vọng nếu H_0 đúng, sau đó chuẩn hóa sự khác biệt này:

$$Z_1 = \frac{\text{giá trị quan sát được của nhóm da trắng } - \text{ giá trị giả thuyết không}}{SE \text{ của giá trị quan sát được nhóm da trắng}}$$

Sai số chuẩn của ước lượng điểm là căn bậc hai của số liệu kỳ vọng theo H_0. [32] Do đó:

$$Z_1 = \frac{205 - 198}{\sqrt{198}} = 0{,}05$$

Công thức này tương tự với thống kê kiểm định trước: tính sai lệch trước, rồi chuẩn hóa. Ta thực hiện tương tự cho các nhóm Da đen, Hispanic, và Khác như sau:

Da đen
$$Z_2 = \frac{26 - 19{,}25}{\sqrt{19{,}25}} = 1{,}54$$

Hispanic
$$Z_3 = \frac{25 - 33}{\sqrt{33}} = -1{,}39$$

Khác
$$Z_4 = \frac{19 - 24{,}75}{\sqrt{24{,}75}} = -1{,}16$$

Chúng ta muốn sử dụng một thống kê kiểm định duy nhất để xác định xem bốn sai lệch chuẩn này có cách xa 0 một cách bất thường hay không. Nghĩa là, Z_1, Z_2, Z_3 và Z_4 cần được kết hợp để xem tất cả - xét như một nhóm - có xu hướng cách xa 0 bất thường không. Ý tưởng ban đầu có thể là lấy giá trị tuyệt đối của bốn sai lệch chuẩn hóa rồi cộng lại:

$$|Z_1| + |Z_2| + |Z_3| + |Z_4| = 4{,}58$$

[32] Dựa trên các quy tắc đã học, ta có thể nghĩ sai số chuẩn là $np(1 - p)$, trong đó n là cỡ mẫu và p là tỷ lệ dân số. Điều này đúng nếu chỉ xem một số liệu đếm, nhưng ở đây ta tính nhiều sai lệch chuẩn hóa và cộng chúng lại. Có thể chứng minh – dù không trình bày ở đây – rằng căn bậc hai của số liệu kỳ vọng là cách tốt hơn để chuẩn hóa các sai lệch đếm.

Thật vậy, cách này cho ra một con số tóm tắt mức độ mà các giá trị quan sát thực tế khác so với giá trị kỳ vọng. Tuy nhiên, phổ biến hơn là cộng các giá trị bình phương:

$$Z_1^2 + Z_2^2 + Z_3^2 + Z_4^2 = 5{,}89$$

Việc bình phương từng sai lệch đã chuẩn hóa trước khi cộng lại mang lại hai tác dụng:

- Bất kỳ sai lệch đã chuẩn hóa nào khi bình phương sẽ luôn cho giá trị dương.
- Những sai khác vốn đã trông bất thường - ví dụ, một sai khác đã chuẩn hóa là 2,5 - sẽ trở nên lớn hơn nhiều sau khi bình phương.

Thống kê kiểm định X^2, tức tổng các giá trị Z^2, hường được sử dụng vì những lý do này. Ta cũng có thể viết công thức cho X^2 bằng cách sử dụng số đếm quan sát và số đếm theo giả thuyết không:

$$X^2 = \frac{(\text{Số lượng quan sát được}_1 - \text{Số quan sát giả thuyết không}_1)^2}{\text{Số quan sát giả thuyết không}_1} + \cdots + \frac{(\text{Số lượng quan sát được}_4 - \text{Số quan sát giả thuyết không}_4)^2}{\text{Số quan sát giả thuyết không}_4}$$

Giá trị cuối cùng X^2 tóm tắt mức độ mà các số liệu quan sát được có xu hướng lệch khỏi các giá trị theo giả thuyết không. Trong mục 6.3.4, ta sẽ thấy rằng nếu giả thuyết không đúng, then X^2 sẽ tuân theo một phân phối mới gọi là phân phối *Chi-bình phương*. Sử dụng phân phối này, ta có thể thu được giá trị p-value để đánh giá giả thuyết.

6.3.3 Phân phối Chi-bình phương và cách tìm diện tích

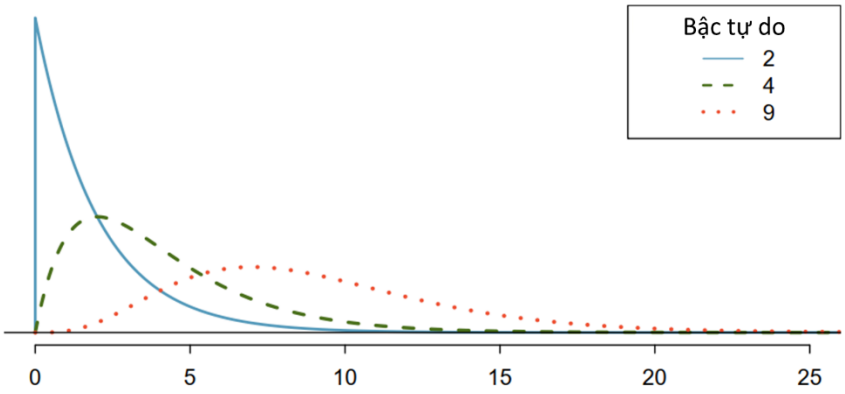
Phân phối **Chi-bình phương** đôi khi được dùng để mô tả các tập dữ liệu và thống kê luôn có giá trị dương và thường lệch phải. Nhớ lại rằng phân phối chuẩn có hai tham số – trung bình và độ lệch chuẩn – dùng để mô tả đặc điểm chính xác của nó. Phân phối Chi-bình phương chỉ có một tham số gọi là **số bậc tự do (degrees of freedom, df)**, ảnh hưởng đến hình dạng, giá trị trung tâm và độ phân tán của phân phối.

thực hành

BÀI TẬP THỰC HÀNH 6.24

Hình 6.7 cho thấy 3 phân phối Chi-bình phương.

a) Trung tâm của phân phối thay đổi như thế nào khi số bậc tự do lớn hơn?
b) Còn độ biến thiên (độ trải rộng) thì sao?
c) Hình dạng của phân phối thay đổi thế nào? [33]



Hình 6.7: Ba kiểm định Chi-bình phương với bậc tự do khác nhau.

[33] (a) Trung tâm trở nên lớn hơn. Nếu quan sát kỹ, chúng ta có thể thấy rằng giá trị trung bình của mỗi phân phối bằng với số bậc tự do của phân phối đó. (b) Độ biến thiên tăng lên khi số bậc tự do tăng. (c) Phân phối lệch rất mạnh khi *df* = 2, và sau đó các phân phối trở nên đối xứng hơn đối với số bậc tự do lớn hơn như *df* = 4 và *df* = 9. Chúng ta sẽ thấy xu hướng này tiếp tục nếu xem xét các phân phối với số bậc tự do còn lớn hơn nữa.

Hình 6.7 và Thực hành có hướng dẫn 6.24 minh họa ba đặc điểm chung của các phân phối Chi bình phương khi số bậc tự do tăng: phân phối trở nên đối xứng hơn, trung tâm dịch sang phải và độ biến thiên tăng lên.

Mối quan tâm chính của chúng ta đối với phân phối Chi bình phương là việc tính toán giá trị p, vốn (như chúng ta đã thấy trước đó) liên quan đến việc tìm diện tích tương ứng ở phần đuôi của phân phối. Những cách phổ biến nhất để thực hiện việc này là sử dụng phần mềm máy tính, máy tính đồ thị hoặc bảng tra cứu. Đối với những ai muốn sử dụng bảng, chúng tôi cung cấp hướng dẫn cách đọc bảng Chi bình phương trong Phụ lục C.3, đây cũng là nơi bạn có thể tìm bảng này. Trong các ví dụ bên dưới, hãy sử dụng phương pháp ưa thích của bạn để xác nhận rằng bạn nhận được cùng một kết quả.

VÍ DỤ 6.25

Hình 6.8(a) minh họa một phân phối Chi-bình phương với 3 bậc tự do và phần đuôi phía trên được tô bóng bắt đầu từ giá trị 6.25. Hãy tìm diện tích phần tô bóng này.

E

Sử dụng phần mềm thống kê hoặc máy tính đồ thị, ta tìm được diện tích phần đuôi phía trên của phân phối Chi-bình phương với 3 bậc tự do (df) và điểm cắt là 6,25 bằng 0,1001. Nói cách khác, phần tô bóng ở đuôi phía trên của Hình 6.8(a) có diện tích là 0.1.

VÍ DỤ 6.26

Hình 6.8(b) minh họa phần đuôi phía trên của phân phối Chi-bình phương với 2 bậc tự do. Điểm cắt cho phần đuôi phía trên này là 4,3. Hãy tìm diện tích phần đuôi.

E

Sử dụng phần mềm, ta tìm được diện tích phần tô bóng ở Hình 6.8(b) bằng 0,1165. Nếu dùng bảng tra, ta chỉ có thể xác định được khoảng giá trị diện tích này: nằm giữa 0,1 và 0,2.

VÍ DỤ 6.27

Hình 6.8(c) minh họa phần đuôi phía trên của phân phối Chi-bình phương với 5 bậc tự do và điểm cắt là 5,1. Hãy tìm diện tích phần đuôi.

E

Sử dụng phần mềm, ta thu được diện tích phần đuôi là 0,4038. Nếu sử dụng bảng ở Phụ lục C.3, ta sẽ xác định được diện tích này lớn hơn 0,3, nhưng không thể đưa ra giá trị chính xác.

BÀI TẬP THỰC HÀNH 6.28

G Hình 6.8(d) có điểm cắt là 11,7 trên phân phối Chi-bình phương với 7 bậc tự do. Hãy tìm diện tích phần đuôi phía trên. [34]

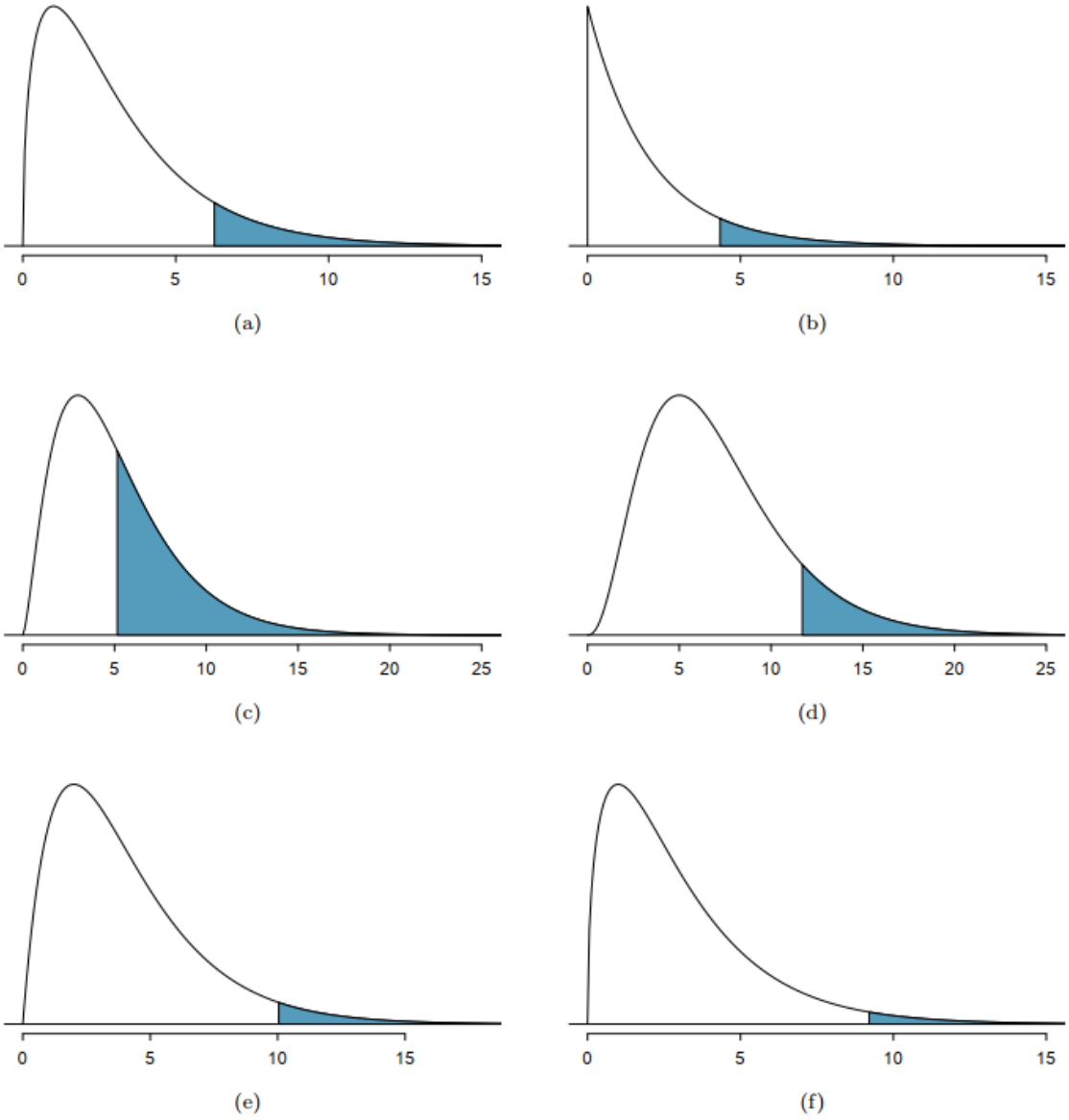
BÀI TẬP THỰC HÀNH 6.29

G Hình 6.8(e) có điểm cắt là 10 trên phân phối Chi-bình phương với 4 bậc tự do. Hãy tìm diện tích phần đuôi phía trên. [35]

BÀI TẬP THỰC HÀNH 6.30

G Hình 6.8(f) có điểm cắt là 9,21 trên phân phối Chi-bình phương với 3 bậc tự do. Hãy tìm diện tích phần đuôi phía trên. [36]

[34] Diện tích là 0,1109. Nếu sử dụng bảng, chúng ta sẽ xác định rằng nó nằm giữa 0,1 và 0,2.
[35] Giá trị chính xác: 0,0404. Nếu sử dụng bảng: nằm giữa 0,02 và 0,05.
[36] Giá trị chính xác: 0,0266. Nếu sử dụng bảng: nằm giữa 0,02 và 0,05



Hình 6.8: (a) Phân phối Chi-bình phương với 3 bậc tự do, phần diện tích bên trên 6,25 được tô bóng. (b) 2 bậc tự do, phần diện tích bên trên 4,3 được tô bóng. (c) 5 bậc tự do, phần diện tích bên trên 5,1 được tô bóng. (d) 7 bậc tự do, phần diện tích bên trên 11,7 được tô bóng. (e) 4 bậc tự do, phần diện tích bên trên 10 được tô bóng. (f) 3 bậc tự do, phần diện tích bên trên 9,21 được tô bóng.

6.3.4 Tìm giá trị p (p-value) cho phân phối Chi-bình phương

Ở mục 6.3.2, chúng ta đã xác định một thống kê kiểm định mới (X^2) trong bối cảnh đánh giá xem có bằng chứng về thiên vị chủng tộc trong cách lựa chọn bồi thẩm đoàn hay không. Giả thuyết không đại diện cho khẳng định rằng bồi thẩm đoàn được chọn ngẫu nhiên và không có thiên vị chủng tộc. Giả thuyết đối là có thiên vị chủng tộc trong cách chọn bồi thẩm đoàn.

Chúng ta xác định rằng một giá trị X^2 lớn sẽ gợi ý bằng chứng mạnh mẽ ủng hộ giả thuyết đối: rằng có sự thiên vị chủng tộc. Tuy nhiên, chúng ta không thể định lượng được khả năng quan sát thấy một thống kê kiểm định lớn như vậy ($X^2 = 5{,}89$) nếu giả thuyết không thực sự đúng. Đây là lúc phân phối Chi-bình phương trở nên hữu ích. Nếu giả thuyết không đúng và không có thiên vị chủng tộc, thì X^2 sẽ tuân theo một phân phối Chi-bình phương, trong trường hợp này với ba bậc tự do. Dưới một số điều kiện nhất định, thống kê X^2 tuân theo phân phối Chi-bình phương với $k - 1$ bậc tự do, trong đó k là số nhóm.

VÍ DỤ 6.31

Có bao nhiêu nhóm trong ví dụ bồi thẩm đoàn? Có bao nhiêu bậc tự do cần được gán cho phân phối Chi-bình phương dùng cho X^2?

(E)

Trong ví dụ bồi thẩm đoàn, có $k = 4$ nhóm: da trắng, da đen, Hispanic và khác. Theo quy tắc trên, thống kê kiểm định X^2 khi đó sẽ tuân theo một phân phối Chi-bình phương với $k - 1 = 3$ bậc tự do nếu H_0 đúng.

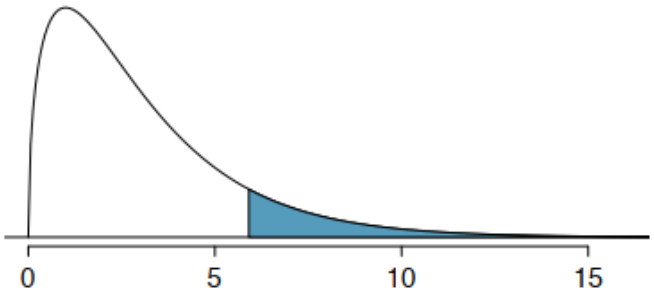
Cũng giống như khi chúng ta kiểm tra điều kiện cỡ mẫu để sử dụng phân phối chuẩn trong các phần trước, chúng ta cũng phải kiểm tra điều kiện cỡ mẫu để áp dụng phân phối Chi-bình phương cho X^2 một cách an toàn. Mỗi tần số kỳ vọng phải ít nhất là 5. Trong ví dụ bồi thẩm đoàn, các tần số kỳ vọng là $198;\ 19{,}25;\ 33$ và $24{,}75$, tất cả đều lớn hơn 5, vì vậy chúng ta có thể áp dụng mô hình Chi-bình phương cho thống kê kiểm định $X^2 = 5{,}89$.

VÍ DỤ 6.32

Nếu giả thuyết không đúng, thống kê kiểm định $X^2 = 5{,}89$ sẽ gắn liền với một phân phối Chi-bình phương có ba bậc tự do. Sử dụng phân phối này và thống kê kiểm định, hãy xác định giá trị p-value.

(E)

Phân phối Chi-bình phương và giá trị p-value được mô tả trong Hình 6.9. Bởi vì giá trị Chi-bình phương lớn hơn đồng nghĩa với bằng chứng mạnh mẽ hơn chống lại giả thuyết không, chúng ta làm đậm phần đuôi trên để biểu thị giá trị p-value. Sử dụng phần mềm thống kê (hoặc bảng trong Phụ lục C.3), chúng ta có thể xác định rằng diện tích này là $0{,}1171$. Thông thường, chúng ta không bác bỏ giả thuyết không với một giá trị p lớn như vậy. Nói cách khác, dữ liệu không cung cấp bằng chứng thuyết phục về sự thiên vị chủng tộc trong việc lựa chọn bồi thẩm đoàn.



Hình 6.9: Giá trị p-value cho kiểm định giả thuyết bồi thẩm đoàn được làm đậm trong phân phối chi-square với $df = 3$.

KIỂM ĐỊNH CHI-BÌNH PHƯƠNG CHO BẢNG MỘT CHIỀU

Giả sử chúng ta cần đánh giá xem liệu có bằng chứng thuyết phục rằng một tập hợp các tần số quan sát được $O_1, O_2, \ldots, O_k$ trong k nhóm khác biệt một cách bất thường so với những gì có thể kỳ vọng theo giả thuyết không hay không. Gọi các *tần số kỳ vọng* dựa trên giả thuyết không là $E_1, E_2, \ldots, E_k$. Nếu mỗi tần số kỳ vọng đều ít nhất là 5 và giả thuyết không đúng, thì thống kê kiểm định dưới đây sẽ tuân theo phân phối Chi-bình phương với $k - 1$ bậc tự do:

$$X^2 = \frac{(O_1 - E_1)^2}{E_1} + \frac{(O_2 - E_2)^2}{E_2} + \cdots + \frac{(O_k - E_k)^2}{E_k}$$

Giá trị p cho thống kê kiểm định này được xác định bằng cách xem xét phần đuôi trên của phân phối Chi-bình phương này. Chúng ta xem xét đuôi trên vì các giá trị X^2 lớn hơn sẽ cung cấp nhiều bằng chứng hơn chống lại giả thuyết không.

ĐIỀU KIỆN ĐỂ THỰC HIỆN KIỂM ĐỊNH CHI-BÌNH PHƯƠNG

Có hai điều kiện cần được kiểm tra trước khi tiến hành kiểm định Chi-bình phương:

Tính độc lập. Mỗi trường hợp đóng góp một giá trị đếm vào bảng phải độc lập với tất cả các trường hợp khác trong bảng.

Kích thước mẫu/phân phối. Mỗi tình huống cụ thể (tức là mỗi ô đếm) phải có ít nhất 5 giá trị kỳ vọng.

Việc không kiểm tra các điều kiện này có thể ảnh hưởng đến tỷ lệ sai số của phép kiểm định.

Khi kiểm tra một bảng chỉ có hai thùng, hãy chọn một thùng và sử dụng các phương pháp một tỷ lệ được giới thiệu trong Phần 6.1.

6.3.5 Đánh giá độ phù hợp cho một phân phối

Mục 4.2 có thể là phần tài liệu nền hữu ích cho ví dụ này, nhưng không phải là điều kiện bắt buộc.

Chúng ta có thể áp dụng khung kiểm định Chi-bình phương cho vấn đề thứ hai trong mục này: đánh giá xem liệu một mô hình thống kê nhất định có phù hợp với một bộ dữ liệu hay không. Dữ liệu lợi nhuận chứng khoán hằng ngày từ S&P500 trong 10 năm có thể được sử dụng để đánh giá xem hoạt động chứng khoán mỗi ngày có độc lập với tình hình chứng khoán trong các ngày trước đó hay không. Đây nghe có vẻ là một câu hỏi phức tạp - và thực tế là như vậy - nhưng kiểm định Chi-bình phương có thể được sử dụng để nghiên cứu vấn đề này. Chúng ta sẽ gán nhãn cho mỗi ngày là Tăng (T) hoặc Giảm (G) tùy thuộc vào việc thị trường hôm đó tăng hay giảm. Ví dụ:

Thay đổi giá	2,52	-1,46	0,51	-4,07	3,36	1,10	-5,46	-1,03	-2,99	1,71
Kết quả	T	G	T	G	T	T	G	G	G	T
Số ngày cho tới Up	1	-	2	-	2	1	-	-	-	4

Nếu các ngày thực sự độc lập, thì số ngày chờ đến một ngày T sẽ tuân theo phân phối hình học. Phân phối này mô tả xác suất phải chờ đến lần thử thứ k^{th} để quan sát thành công đầu tiên. Trong trường hợp này, mỗi ngày T là một thành công, và mỗi ngày G là một thất bại. Trong dữ liệu trên, lần đầu tiên thị trường tăng giá chỉ mất 1 ngày chờ, sau đó mất thêm 2 ngày để có ngày Up tiếp theo, và tiếp tục như vậy. Ta muốn xác định xem những giá trị chờ này $(1, 2, 2, 1, 4, \ldots v.v)$ có tuân theo phân phối hình học hay không. Hình 6.10 cho thấy số ngày chờ để có một ngày tăng giá trong 10 năm đối với S&P500:

Số ngày	1	2	3	4	5	6	7+	Tổng
Số quan sát	717	369	155	69	28	14	10	1362

Hình 6.10: Phân bố quan sát của thời gian chờ cho đến một ngày giao dịch tăng giá đối với chỉ số S&P500.

Chúng ta xem xét số ngày cần chờ để xuất hiện một ngày tăng giá (Up day) trên chỉ số chứng khoán S&P500. Nếu hoạt động của thị trường chứng khoán là độc lập từ ngày này sang ngày khác và xác suất có một ngày giao dịch tăng giá là hằng số, thì thời gian chờ này được kỳ vọng sẽ tuân theo *phân bố hình học*. Có thể sắp xếp vấn đề này thành khung giả thuyết như sau:

H_0: Việc thị trường chứng khoán tăng hoặc giảm vào một ngày nhất định là độc lập với tất cả các ngày khác. Chúng ta sẽ xem xét số ngày trôi qua cho đến khi xuất hiện một ngày tăng giá. Dưới giả thuyết này, số ngày chờ cho đến một ngày tăng giá sẽ tuân theo phân bố hình học.

H_A: Việc thị trường chứng khoán tăng hoặc giảm vào một ngày nhất định không độc lập với tất cả các ngày khác. Vì chúng ta biết rằng số ngày chờ cho đến một ngày tăng giá sẽ tuân theo phân bố hình học dưới giả thuyết không, nên chúng ta sẽ tìm kiếm những sai lệch so với phân bố hình học, điều này sẽ ủng hộ giả thuyết đối.

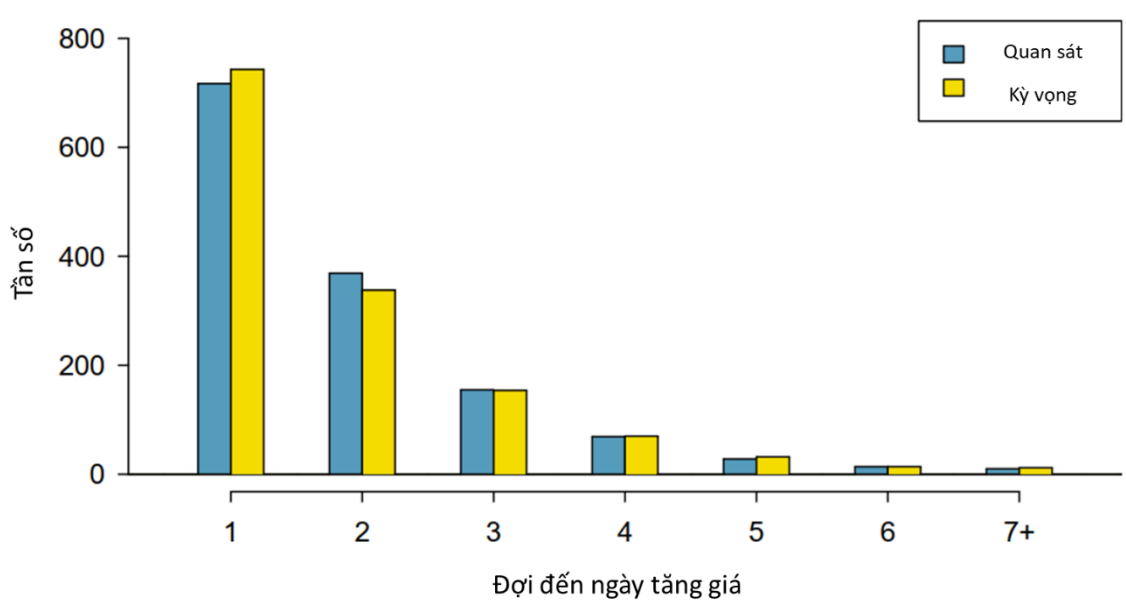
Kết quả này có ý nghĩa quan trọng đối với các nhà giao dịch chứng khoán: nếu thông tin từ các ngày giao dịch trước có thể hữu ích trong việc dự đoán điều gì sẽ xảy ra hôm nay, thì thông tin đó có thể mang lại lợi thế so với những nhà giao dịch khác.

Chúng ta xem xét dữ liệu của S&P500 và tóm tắt thời gian chờ trong Hình 6.11 và Hình 6.12. Chỉ số S&P500 tăng trong 54,5% số ngày.

Vì việc áp dụng khuôn khổ kiểm định Chi-bình phương yêu cầu số đếm kỳ vọng phải ít nhất bằng 5, chúng ta đã gộp tất cả các trường hợp có thời gian chờ từ 7 ngày trở lên để đảm bảo mỗi số đếm kỳ vọng đều lớn hơn mức tối thiểu này. Dữ liệu thực tế, được hiển thị ở hàng *Quan sát* trong Hình 6.11, có thể so sánh với các số đếm kỳ vọng ở hàng *Mô hình hình học*. Phương pháp tính số đếm kỳ vọng được trình bày trong Hình 6.11. Nói chung, số đếm kỳ vọng được xácđịnh bằng cách: (1) Xác định tỉ lệ theo giả thuyết không ứng với mỗi nhóm. (2) Nhân tỉ lệ này với tổng số quan sát để thu được số đếm kỳ vọng. Nói cách khác, phương pháp này xác định tỷ lệ phần trăm của tổng số quan sát mà ta kỳ vọng sẽ nằm trong mỗi nhóm.

Ngày	1	2	3	4	5	6	7+	Tổng
Số quan sát	717	369	155	69	28	14	10	1362
Mô hình hình học	743	338	154	70	32	14	12	1362

Hình 6.11: Phân bố thời gian chờ cho đến một ngày giao dịch tăng giá. Các giá trị kỳ vọng dựa trên mô hình phân bố hình học được thể hiện ở hàng cuối. Để tìm mỗi giá trị kỳ vọng, ta xác định xác suất chờ D ngày dựa trên mô hình phân bố hình học $P(D) = (1-0{,}545)^{D-1}(0{,}545)$ và nhân với tổng số chuỗi là 1362. Ví dụ, việc chờ 3 ngày xảy ra theo mô hình hình học khoảng $0{,}455^2 \times 0{,}545 = 11{,}28\%$ thời gian, tương ứng với $0{,}1128 \times 1362 = 154$ chuỗi.



Hình 6.12: Biểu đồ cột song song của số quan sát thực tế và số kỳ vọng cho mỗi khoảng thời gian chờ.

VÍ DỤ 6.33

Bạn có nhận thấy sự sai lệch bất thường nào trong biểu đồ không? Bạn có thể cho biết liệu những sai lệch này là do ngẫu nhiên chỉ bằng cách quan sát không?

E

Không rõ rằng liệu sự khác biệt giữa các số quan sát được và các số kỳ vọng từ phân phối hình học có khác biệt đáng kể hay không. Nói cách khác, không thể khẳng định liệu những sai lệch này chỉ là do ngẫu nhiên hay chúng đủ lớn để dữ liệu đưa ra bằng chứng thuyết phục chống lại giả thuyết không. Tuy nhiên, chúng ta có thể thực hiện phép kiểm định Chi-bình phương sử dụng các số liệu trong Hình 6.11.

G

BÀI TẬP THỰC HÀNH 6.34

Hình 6.11 cung cấp một tập dữ liệu đếm cho thời gian chờ ($O_1 = 717;\ O_2 = 369;\ \ldots$) và các số kỳ vọng theo phân phối hình học ($E_1 = 743;\ E_2 = 338;\ \ldots$). Hãy tính thống kê kiểm định Chi-bình phương X^2.[37]

G

BÀI TẬP THỰC HÀNH 6.35

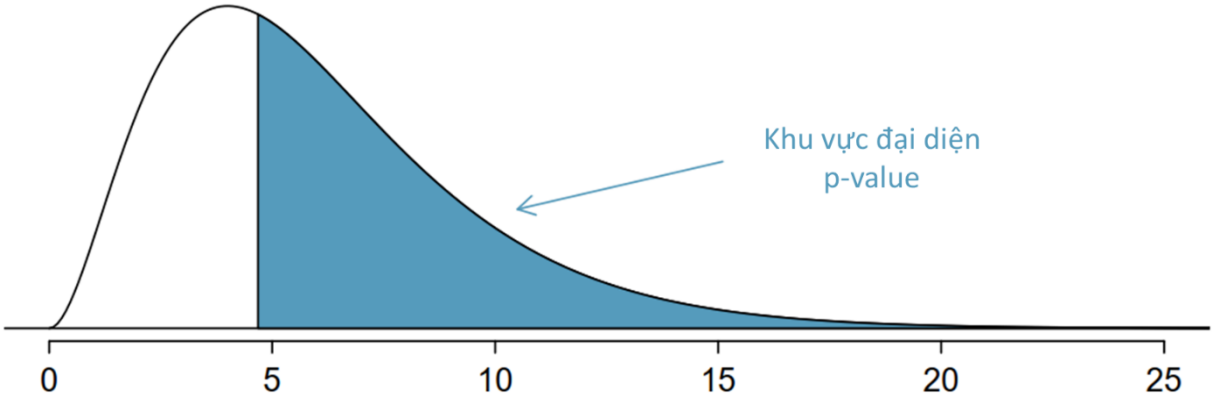
Vì tất cả các số kỳ vọng đều ít nhất là 5, chúng ta có thể áp dụng phân phối Chi-bình phương cho X^2 một cách an toàn. Tuy nhiên, cần sử dụng bao nhiêu bậc tự do? [38]

VÍ DỤ 6.36

Nếu các số đếm quan sát được tuân theo mô hình hình học, thì thống kê kiểm định chi-bình phương $X_2 = 4{,}61$ sẽ xấp xỉ tuân theo phân phối Chi-bình phương với $df = 6$. Dựa trên thông tin này, hãy tính giá trị p-value.

E

Hình 6.13 minh họa phân phối Chi-bình phương, ngưỡng cắt và phần diện tích p-value được tô bóng. Sử dụng phần mềm, ta thu được p-value: 0.5951. Kết luận: chúng ta không có đủ bằng chứng để bác bỏ giả định rằng thời gian chờ tuân theo phân phối hình học trong 10 năm gần đây của S&P500; nói cách khác, không thể bác bỏ giả thuyết rằng các ngày giao dịch là độc lập.



Hình 6.13: Phân phối Chi-bình phương với 6 bậc tự do. Giá trị p-value cho phân tích cổ phiếu được tô bóng.

VÍ DỤ 6.37

Trong Ví dụ 6.36, chúng ta đã không bác bỏ giả thuyết không rằng các ngày giao dịch là độc lập trong 10 năm dữ liệu vừa qua. Điều này quan trọng vì sao?

E

Có thể rất dễ bị cám dỗ để nghĩ rằng thị trường "đến lúc" sẽ có một ngày tăng điểm nếu đã có nhiều ngày giảm liên tiếp. Tuy nhiên, chúng ta chưa tìm thấy bằng chứng mạnh mẽ nào cho thấy tồn tại đặc điểm như vậy, rằng thị trường "đến hạn" điều chỉnh. Ít nhất, phân tích này gợi ý rằng bất kỳ sự phụ thuộc nào giữa các ngày là rất yếu.

[37] $X^2 = \frac{(717-743)^2}{743} + \frac{(369-338)^2}{338} + \cdots + \frac{(10-12)^2}{12} = 4{,}61$

38 Có $k = 7$ nhóm nên ta dùng bậc tực do là $df = 7 - 1 = 6$

Bài tập

6.31 Đúng hay sai, Phần I. Xác định xem các phát biểu dưới đây là đúng hay sai. Với mỗi phát biểu sai, hãy đề xuất một cách diễn đạt khác để biến nó thành phát biểu đúng.

a) Phân phối Chi-bình phương, cũng như phân phối chuẩn, có hai tham số, trung bình và độ lệch chuẩn.
b) Phân phối chi-bình phương luôn lệch phải, bất kể giá trị của tham số bậc tự do.
c) Thống kê chi-bình phương luôn dương.
d) Khi bậc tự do tăng, dạng (độ lệch) của phân phối chi-bình phương trở nên lệch nhiều hơn.

6.32 Đúng hay sai, Phần II. Xác định xem các phát biểu dưới đây là đúng hay sai. Với mỗi phát biểu sai, hãy đề xuất một cách diễn đạt khác để biến nó thành phát biểu đúng.

a) Khi bậc tự do tăng, trung bình của phân phối chi-bình phương tăng.
b) Nếu bạn tìm được $\chi^2 = 10$ với $df = 5$ thì bạn sẽ không thể bác bỏ H_0 ở mức ý nghĩa 5%.
c) Khi tìm p-value của một phép kiểm định Chi-bình phương, chúng ta luôn tô phần diện tích đuôi ở cả hai phía.
d) Khi bậc tự do tăng, độ biến thiên (phương sai) của phân phối chi-bình phương giảm.

6.33 Sách giáo khoa nguồn mở. Một giảng viên sử dụng sách nhập môn thống kê nguồn mở dự đoán rằng 60% sinh viên sẽ mua bản in, 25% sẽ in từ web, và 15% sẽ đọc trực tuyến. Cuối kỳ, thầy yêu cầu sinh viên hoàn thành khảo sát về định dạng sách họ đã sử dụng. Trong 126 sinh viên, 71 nói rằng họ mua bản in, 30 nói rằng họ in từ web, và 25 nói rằng họ đọc trực tuyến.

a) Phát biểu các giả thuyết để kiểm tra xem dự đoán của giảng viên có không chính xác hay không.
b) Giảng viên kỳ vọng bao nhiêu sinh viên sẽ mua sách, in sách, và đọc trực tuyến?
c) Đây là bối cảnh phù hợp cho kiểm định Chi-bình phương. Liệt kê các điều kiện cần cho phép kiểm định và xác minh chúng được thỏa mãn.
d) Tính thống kê Chi-bình phương, số bậc tự do đi kèm, và p-value.
e) Dựa trên p-value ở phần (d), kết luận của phép kiểm định là gì? Diễn giải kết luận trong ngữ cảnh này.

6.34 Hươu Barking. Các yếu tố vi sinh cảnh liên quan đến nơi kiếm ăn và nơi nghỉ ngơi của hươu Barking ở đảo Hải Nam, Trung Quốc đã được khảo sát. Ở vùng này, rừng chiếm 4,8% diện tích đất, bãi cỏ canh tác chiếm 14,7%, và rừng rụng lá chiếm 39,6%. Trong 426 điểm nơi hươu kiếm ăn, 4 được phân loại là rừng, 16 là bãi cỏ canh tác, và 61 là rừng rụng lá. Bảng dưới đây tóm tắt dữ liệu. [39]

Rừng	Bãi cỏ canh tác	Rừng rụng lá	Khác	Tổng
4	16	61	345	426

a) Trình bày các giả thuyết để kiểm tra xem hươu Barking có ưu tiên kiếm ăn ở một số kiểu sinh cảnh nhất định hơn các kiểu khác hay không.
b) Chúng ta có thể dùng loại kiểm định nào để trả lời câu hỏi nghiên cứu này?
c) Kiểm tra xem các giả định và điều kiện cần cho phép kiểm định có được thỏa mãn hay không.
d) Dữ liệu này có cung cấp bằng chứng thuyết phục rằng hươu Barking ưu tiên kiếm ăn ở một số sinh cảnh hơn các sinh cảnh khác không? Thực hiện một phép kiểm định giả thuyết thích hợp để trả lời câu hỏi nghiên cứu này.



Ảnh chụp bởi Shrikant Rao (http://flic.kr/p/4Xjdkk) CC BY 2.0 license

[39] Liwei Teng et al. "Forage and bed sites characteristics of Indian muntjac (Muntiacus muntjak) in Hainan Island, China". In: *Ecological Research* 19.6 (2004), pp. 675–681.

6.4. KIỂM ĐỊNH TÍNH ĐỘC LẬP TRONG BẢNG HAI CHIỀU

6.4. Kiểm Định Tính Độc Lập Trong Bảng Hai Chiều

Chúng ta thường mua các sản phẩm đã qua sử dụng – như ô tô, máy tính, sách giáo khoa, v.v. – và đôi khi mặc nhiên cho rằng người bán sẽ thẳng thắn về những vấn đề tiềm ẩn của sản phẩm. Tuy nhiên, đây không phải là điều có thể coi là hiển nhiên. Trong một nghiên cứu, các nhà nghiên cứu đã tuyển chọn 219 người tham gia với nhiệm vụ bán một chiếc iPod40 [40] vốn được biết là đã từng bị treo hai lần trước đó. Người tham gia được khuyến khích tối đa hóa số tiền bán được, bởi ngoài khoản thù lao cố định 10 USD, họ còn nhận thêm 5% giá trị từ giao dịch. Mục tiêu của nghiên cứu là xác định những loại câu hỏi nào có khả năng thúc đẩy người bán tiết lộ vấn đề treo máy của thiết bị.

Người tham gia trong vai trò người bán không hề biết rằng phía người mua thực chất đang hợp tác với các nhà nghiên cứu nhằm đánh giá tác động của các dạng câu hỏi khác nhau đối với khả năng người bán tiết lộ các vấn đề trước đây của chiếc iPod. Người mua được hướng dẫn theo kịch bản, bắt đầu bằng câu: "*Được rồi, tôi đoán là tôi nên bắt đầu trước. Vậy là bạn đã sử dụng chiếc iPod này được 2 năm ...*" và kết thúc bằng một trong ba dạng câu hỏi.

- Câu hỏi chung: Bạn có thể cho tôi biết điều gì về chiếc máy này?
- Giả định tích cực: Chiếc máy này không gặp vấn đề gì, đúng không?
- Giả định tiêu cực: Chiếc máy này có những vấn đề gì?

Trong nghiên cứu này, loại câu hỏi được xem là biến can thiệp áp dụng cho người bán, còn phản ứng được xác định là việc liệu câu hỏi đó có thúc đẩy họ tiết lộ tình trạng treo máy của iPod hay không. Kết quả được trình bày trong Hình 6.14. Dữ liệu cho thấy câu hỏi *Chiếc máy này có những vấn đề gì*? là cách hiệu quả nhất để khiến người bán thừa nhận vấn đề treo máy trước đó.

	Chung	Giả định tích cực	Giả định tiêu cực	Tổng
Tiết lộ vấn đề	2	23	36	61
Che giấu vấn đề	71	50	37	158
Tổng	73	73	73	219

Hình 6.14. Tóm tắt nghiên cứu về iPod, trong đó một câu hỏi được đặt ra cho người tham gia nghiên cứu trong vai trò người bán.

SỰ KHÁC BIỆT GIỮA BẢNG MỘT CHIỀU VÀ BẢNG HAI CHIỀU

Bảng một chiều mô tả số lượng quan sát cho từng kết quả của một biến duy nhất. Bảng hai chiều mô tả số lượng quan sát cho các tổ hợp kết quả của hai biến. Khi xem xét một bảng hai chiều, chúng ta thường quan tâm đến câu hỏi: liệu hai biến này có liên quan với nhau hay không? Nói cách khác, chúng có phụ thuộc lẫn nhau hay độc lập?

Kiểm định giả thuyết trong thí nghiệm iPod thực chất nhằm đánh giá xem có bằng chứng thống kê đủ ý nghĩa về sự khác biệt trong mức độ thành công của từng loại câu hỏi trong việc khiến người tham gia tiết lộ sự cố của iPod hay không. Nói cách khác, mục tiêu là kiểm tra xem liệu loại câu hỏi của người mua có độc lập với việc người bán có tiết lộ sự cố hay không.

[40] Đối với những độc giả trẻ hơn so với các tác giả, có thể hiểu iPod về cơ bản giống như một chiếc iPhone nhưng không có dịch vụ di động, nếu đó là các thế hệ đời sau. Các thế hệ iPod trước đó thì đơn giản hơn nhiều.

6.4. KIỂM ĐỊNH TÍNH ĐỘC LẬP TRONG BẢNG HAI CHIỀU

6.4.1. Giá trị kỳ vọng trong bảng hai chiều

Tương tự như đối với bảng một chiều, chúng ta cũng cần tính toán các giá trị đếm ước lượng cho từng ô trong bảng hai chiều.

VÍ DỤ 6.38

Từ thí nghiệm, chúng ta có thể tính được tỷ lệ người bán tiết lộ vấn đề treo máy là $61/219 = 0{,}2785$. Nếu thực sự không có sự khác biệt giữa các loại câu hỏi và 27,85% người bán sẽ tiết lộ vấn đề treo máy bất kể câu hỏi nào được đưa ra, thì trong số 73 người thuộc nhóm Chung, chúng ta sẽ kỳ vọng có bao nhiêu người tiết lộ vấn đề treo máy?

E

Chúng ta dự đoán rằng $0{,}2785 \times 73 = 20{,}33$ người bán sẽ tiết lộ vấn đề. Rõ ràng, số quan sát thực tế thấp hơn con số này, tuy nhiên hiện vẫn chưa thể khẳng định đó chỉ là do biến thiên ngẫu nhiên hay do bản chất các loại câu hỏi khác nhau về mức độ hiệu quả trong việc khai thác sự thật.

BÀI TẬP THỰC HÀNH 6.39

G

Nếu giả định rằng các loại câu hỏi thực chất có hiệu quả như nhau, tức là khoảng 27,85% số người tham gia sẽ tiết lộ vấn đề treo máy bất kể họ được hỏi câu gì, thì chúng ta sẽ kỳ vọng có bao nhiêu người bán trong nhóm giả định tích cực che giấu vấn đề treo máy? [41]

Chúng ta có thể tính số lượng kỳ vọng người bán sẽ tiết lộ hoặc che giấu vấn đề treo máy cho tất cả các nhóm, trong trường hợp loại câu hỏi không ảnh hưởng đến việc họ tiết lộ, bằng cách áp dụng cùng một phương pháp như trong ví dụ 6.38 và hướng dẫn thực hành 6.39. Các giá trị kỳ vọng này được sử dụng để xây dựng Hình 6.15, vốn giống với Hình 6.14, chỉ khác là đã bổ sung thêm các giá trị kỳ vọng trong ngoặc đơn.

	Chung	Giả định tích cực	Giả định tiêu cực	Tổng
Tiết lộ vấn đề	2 (20,33)	23 (20,33)	36 (20,33)	61
Che giấu vấn đề	71 (52,67)	50 (52,67)	37 (52,67)	158
Tổng	73	73	73	219

Hình 6.15. Các giá trị quan sát và các giá trị kỳ vọng (đặt trong ngoặc đơn).

Các ví dụ và bài tập nêu trên đã phần nào giúp trong việc tính các giá trị kỳ vọng. Nói chung, các giá trị kỳ vọng trong bảng hai chiều có thể được tính dựa trên tổng số của hàng, tổng số của cột và tổng số chung của bảng. Chẳng hạn, nếu không có sự khác biệt giữa các nhóm, thì khoảng 27,85% của mỗi cột nên xuất hiện ở hàng đầu tiên:

$$0{,}2785 \times (Tổng\ số\ cột\ 1) = 20{,}33$$
$$0{,}2785 \times (Tổng\ số\ cột\ 2 = 20{,}33$$
$$0{,}2785 \times (Tổng\ số\ cột\ 3) = 20{,}33$$

Nhìn lại cách tính $0{,}2785$ – là tỷ lệ người bán tiết lộ vấn đề treo máy $(158/219)$ – ba giá trị kỳ vọng này có thể được tính như sau:

$$\left(\frac{Tổng\ hàng\ 1}{Tổng\ bảng}\right)(\text{Tổng cột } 1) = 20{,}33$$
$$\left(\frac{Tổng\ hàng\ 1}{Tổng\ bảng}\right)(\text{Tổng cột } 2) = 20{,}33$$
$$\left(\frac{Tổng\ hàng\ 1}{Tổng\ bảng}\right)(\text{Tổng cột } 3) = 20{,}33$$

Điều này dẫn chúng ta đến công thức tổng quát để tính các giá trị kỳ vọng trong bảng hai chiều, khi ta muốn kiểm tra xem có bằng chứng mạnh mẽ về mối liên quan giữa biến cột và biến hàng hay không.

[41] Chúng ta sẽ kỳ vọng $(1 - 0{,}2785) \times 73 = 52{,}67$. Không sao cả khi kết quả này, giống như kết quả trong Ví dụ 6.38, là một số thập phân.

6.4. KIỂM ĐỊNH TÍNH ĐỘC LẬP TRONG BẢNG HAI CHIỀU

TÍNH GIÁ TRỊ KỲ VỌNG TRONG BẢNG HAI CHIỀU

Để xác định giá trị kỳ vọng tại hàng thứ i và cột thứ j, ta tính:

$$Giá\ trị\ kỳ\ vọng_{hàng\ i,cột\ j} = \frac{(Tổng\ hàng\ i) \times (Tổng\ cột\ j)}{Tổng\ bảng}$$

6.4.2. Kiểm định chi bình phương cho bảng hai chiều

Giá trị thống kê chi bình phương cho một bảng hai chiều được tính theo cùng phương pháp như đối với bảng một chiều. Đối với mỗi ô trong bảng, ta tính:

$Công\ thức\ chung$	$\frac{(Giá\ trị\ quan\ sát - Giá\ trị\ kỳ\ vọng)^2}{Giá\ trị\ kỳ\ vọng}$
$Hàng\ 1, Cột\ 1$	$\frac{(2-2{,}33)^2}{20{,}33} = 16{,}53$
$Hàng\ 1, Cột\ 2$	$\frac{(23-2{,}33)^2}{20{,}33} = 0{,}35$
⋮	⋮
$Hàng\ 2, Cột\ 3$	$\frac{(37-52{,}67)^2}{52{,}67} = 4{,}66$

Cộng các giá trị đã tính cho từng ô sẽ cho ra giá trị thống kê X^2:

$$X^2 = 16{,}53 + 0{,}35 + \cdots + 4{,}66 = 40{,}13$$

Giống như trước, thống kê kiểm định này tuân theo phân phối chi bình phương. Tuy nhiên, bậc tự do được tính hơi khác đối với bảng hai chiều.[42] Đối với bảng hai chiều, bậc tự do được tính theo công thức:

$$df = (số\ hàng - 1) \times (số\ cột - 1)$$

Trong ví dụ này, tham số bậc tự do là

$$df = (2-1) \times (3-1) = 2$$

Nếu giả thuyết không đúng (tức là các câu hỏi không ảnh hưởng đến người bán trong thí nghiệm), thì thống kê kiểm định $X^2 = 40{,}13$ gần như tuân theo phân phối chi bình phương với 2 bậc tự do. Dựa trên thông tin này, chúng ta có thể tính giá trị p cho kiểm định, như được minh họa trong Hình 6.16.

TÍNH BẬC TỰ DO CHO BẢNG HAI CHIỀU

Khi áp dụng kiểm định chi bình phương cho một bảng hai chiều, chúng ta sử dụng:

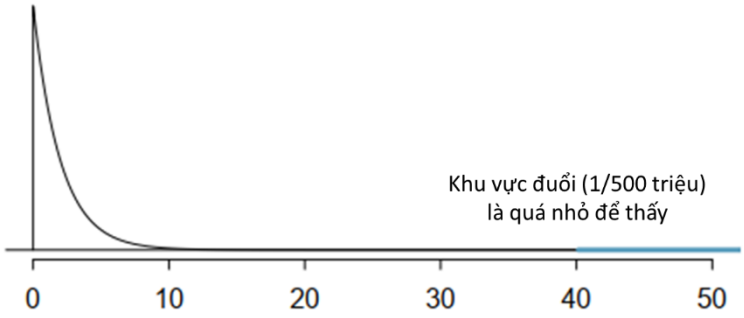
$$df = (R-1) \times (C-1)$$

Trong đó, R là số hàng trong bảng và C là số cột.

Khi phân tích các bảng tần số 2×2, một hướng dẫn là sử dụng các phương pháp hai tỷ lệ đã được giới thiệu trong Mục 6.2.

[42] Nhớ lại: trong bảng một chiều, bậc tự do là số ô trừ đi 1.

6.4. KIỂM ĐỊNH TÍNH ĐỘC LẬP TRONG BẢNG HAI CHIỀU



Hình 6.16: Minh họa giá trị p cho $X^2 = 40{,}13$, khi $df = 2$.

ví dụ

VÍ DỤ 6.40

Tìm giá trị p và rút ra kết luận về việc liệu loại câu hỏi có ảnh hưởng đến khả năng người bán báo cáo vấn đề treo máy hay không.

E

Sử dụng máy tính, chúng ta có thể tính giá trị chính xác cho diện tích phía đuôi lớn hơn $X^2 = 40{,}13$ của phân phối chi bình phương với 2 bậc tự do: 0,000000002. (Nếu sử dụng bảng trong Phụ lục C.3, ta sẽ xác định rằng giá trị p nhỏ hơn 0,001.) Với mức ý nghĩa $\alpha = 0{,}05$, giả thuyết không bị bác bỏ vì giá trị p nhỏ hơn mức ý nghĩa. Nói cách khác, dữ liệu cung cấp bằng chứng thuyết phục rằng loại câu hỏi được đặt ra thực sự ảnh hưởng đến khả năng người bán nói thật về các vấn đề của iPod.

VÍ DỤ 6.41

Hình 6.17 tóm tắt kết quả của một thí nghiệm đánh giá ba phương pháp điều trị bệnh Tiểu đường tuýp 2 ở các bệnh nhân từ 10–17 tuổi đang được điều trị bằng metformin. Ba phương pháp được xem xét là tiếp tục điều trị với metformin (met), điều trị kết hợp metformin và rosiglitazone (rosi) hoặc chương trình can thiệp lối sống. Mỗi bệnh nhân có một kết cục chính là thiếu kiểm soát đường huyết hoặc không thiếu kiểm soát đường huyết. Giả thuyết phù hợp cho kiểm định này là gì?

E

H_0: Không có sự khác biệt về hiệu quả giữa ba phương pháp điều trị.
H_A: Có sự khác biệt về hiệu quả giữa ba phương pháp điều trị, ví dụ như phương pháp rosi có thể hiệu quả hơn so với can thiệp lối sống.

	Thất bại	Thành Công	Tổng
Lối Sống	109	125	234
met	120	112	232
rosi	90	143	233
Tổng	319	380	699

Hình 6.17. Kết quả nghiên cứu về bệnh Tiểu đường tuýp 2.

6.4. KIỂM ĐỊNH TÍNH ĐỘC LẬP TRONG BẢNG HAI CHIỀU

BÀI TẬP THỰC HÀNH 6.42

Một kiểm định chi bình phương cho bảng hai chiều có thể được sử dụng để kiểm tra các giả thuyết trong Ví dụ 6.41. Bước đầu tiên là tính các giá trị kỳ vọng cho từng trong sáu ô của bảng. [43]

BÀI TẬP THỰC HÀNH 6.43

Tính giá trị thống kê Chi bình phương cho dữ liệu trontg Hình 6.17. [44]

BÀI TẬP THỰC HÀNH 6.44

Vì bảng có 3 hàng và 2 cột, bậc tự do cho kiểm định là $df = (3-1) \times (2-1) = 2$. Sử dụng $X^2 = 8{,}16$, $df = 2$, hãy đánh giá xem có nên bác bỏ giả thuyết không hay không với mức ý nghĩa là 0.05. [45]

[43] Giá trị kỳ vọng cho ô hàng 1 / cột 1 được tính bằng cách nhân tổng hàng 1 (234) với tổng cột 1 (319), rồi chia cho tổng bảng (699): $\frac{234 \times 319}{699} = 106{,}8$. Tương tự, hàng 1 và cột 2: $\frac{234 \times 380}{699} = 127{,}2$. Hàng 2: 105,9 và 126,1. Hàng 3: 106,3 và 126,7. [44] Đối với từng ô, tính $\frac{(\text{giá trị quan sát} - \text{giá trị kỳ vọng})^2}{\text{Giá trị kỳ vọng}}$. Ví dụ, hàng 1 và cột 1: $\frac{(109-106{,}8)^2}{106} = 0{,}05$.
Cộng các kết quả của từng ô sẽ cho ra giá trị thống kê $X^2 = 0{,}05 + \cdots + 2{,}11 = 8{,}16$.

[45] Nếu sử dụng máy tính, chúng ta có thể xác định giá trị $p = 0{,}017$. Như vậy, chúng ta bác bỏ giả thuyết không vì giá trị p nhỏ hơn 0,05, và kết luận rằng ít nhất một trong các phương pháp điều trị có hiệu quả hơn hoặc kém hơn các phương pháp còn lại trong việc kiểm soát đường huyết ở bệnh nhân Tiểu đường tuýp 2.

6.4. KIỂM ĐỊNH TÍNH ĐỘC LẬP TRONG BẢNG HAI CHIỀU

BÀI TẬP

6.35 Những người bỏ thuốc lá. Việc tham gia nhóm hỗ trợ có ảnh hưởng đến khả năng bỏ thuốc lá hay không? Một sở y tế quận đã tuyển 300 người hút thuốc tham gia một thử nghiệm ngẫu nhiên. Trong đó, 150 người được phân vào nhóm sử dụng miếng dán nicotine và tham gia nhóm hỗ trợ hàng tuần; 150 người còn lại chỉ nhận miếng dán nicotine mà không tham gia nhóm hỗ trợ.

a) Tạo bảng hai chiều trình bày kết quả nghiên cứu:
b) Trả lời từng câu hỏi dưới giả thuyết không rằng việc tham gia nhóm hỗ trợ không ảnh hưởng đến khả năng bỏ thuốc lá, và xác định xem giá trị kỳ vọng cao hơn hay thấp hơn giá trị quan sát được.
   i. Bạn kỳ vọng bao nhiêu người trong nhóm "dùng miếng dán+ tham giá nhóm hỗ trợ" sẽ bỏ thuốc?
   ii. kỳ vọng bao nhiêu người trong nhóm "chỉ dùng miếng dán" sẽ không bỏ thuốc?

6.36 Chụp toàn thân, Phần II. Bảng dưới đây tóm tắt dữ liệu từ Bài tập 6.26 về quan điểm đối với chụp toàn thân theo xu hướng chính trị. Sự khác biệt giữa các nhóm chính trị có thể xảy ra do ngẫu nhiên. Thực hiện các phép tính dưới giả thuyết không rằng xu hướng chính trị và mức độ ủng hộ chụp toàn thân là độc lập. Trước khi tính toán, có thể thêm một cột tổng theo hàng để thuận tiện cho việc phân tích.

		Đảng phái chính trị		
		Cộng hòa	Dân chủ	Độc lập
Câu trả lời	Cần chụp	264	299	351
	Không cần chụp	38	55	77
	Không biết/Không trả lời	16	15	22
	Tổng	318	369	450

a) Bạn kỳ vọng bao nhiêu người Cộng hòa không ủng hộ việc sử dụng chụp toàn thân?
b) Bạn kỳ vọng bao nhiêu người Dân chủ ủng hộ việc sử dụng chụp toàn thân?
c) Bạn kỳ vọng bao nhiêu người Độc lập không biết hoặc không trả lời?

6.37 Khoan dầu ngoài khơi, Phần III. Bảng dưới đây tóm tắt dữ liệu từ Bài tập 6.23, khảo sát phản hồi của một mẫu ngẫu nhiên gồm những người tốt nghiệp đại học và chưa tốt nghiệp đại học về vấn đề khoan dầu. Hoàn thành phép kiểm định chi-square trên dữ liệu này để xác định xem có sự khác biệt có ý nghĩa thống kê giữa phản hồi của những người tốt nghiệp và chưa tốt nghiệp hay không.

	Tốt nghiệp đại học	
	Có	Không
Ủng hộ	154	132
Phản đối	180	126
Không biết	104	131
Tổng	438	389

6.38 Sâu ký sinh. Bệnh giun bạch huyết là một bệnh do ký sinh trùng gây ra. Các biến chứng của bệnh có thể dẫn đến sưng phù nghiêm trọng và các vấn đề sức khỏe khác. Ở đây, chúng ta xem xét kết quả từ một thử nghiệm ngẫu nhiên so sánh ba phương án điều trị thuốc khác nhau nhằm loại bỏ ký sinh trùng này, mà cộng đồng đang nỗ lực loại trừ hoàn toàn. Kết quả của năm thứ hai trong nghiên cứu được trình bày dưới đây: [46]

	Loại bỏ ký sinh trùng sau năm 2	Chưa loại bỏ ký sinh trùng sau năm 2
Ba loại thuốc	52	2
Hai loại thuốc	31	24
Hai loại thuốc mỗi năm	42	14

a) Thiết lập giả thuyết để đánh giá xem có sự khác biệt nào về hiệu quả giữa các phương pháp điều trị hay không, đồng thời kiểm tra các điều kiện áp dụng kiểm định.
b) Kết quả kiểm định chi bình phương từ phần mềm thống kê, kết quả cho thấy:

$$X^2 = 23{,}7 \qquad df = 2 \qquad p = 7{,}2e-6$$

Sử dụng các kết quả này để đánh giá các giả thuyết ở phần (a), đồng thời đưa ra kết luận phù hợp với bối cảnh nghiên cứu.

[46] Christopher King et al. "A Trial of a Triple-Drug Treatment for Lymphatic Filariasis". In: *New England Journal of Medicine* 379 (2018), pp. 1801–1810.

6.4. KIỂM ĐỊNH TÍNH ĐỘC LẬP TRONG BẢNG HAI CHIỀU

Bài tập cuối chương

6.39 Học tập chủ động. Một giảng viên muốn tăng cường yếu tố học tập chủ động trong khóa học của mình quan tâm đến phản ứng của sinh viên trước các thay đổi mà cô dự định thực hiện. Cô tiến hành một khảo sát trong lớp, hỏi sinh viên liệu họ có tin rằng việc tăng cường học tập chủ động trong lớp (các bài tập thực hành) thay vì chỉ nghe giảng truyền thống có giúp cải thiện việc học của họ hay không. Cô thực hiện khảo sát này vào đầu và cuối học kỳ và muốn đánh giá xem quan điểm của sinh viên có thay đổi trong suốt học kỳ hay không. Liệu cô có thể sử dụng các phương pháp mà chúng ta học trong chương này cho phân tích này không? Giải thích?

6.37 Thí nghiệm trên website. Trang web OpenIntro thỉnh thoảng thử nghiệm các thiết kế và vị trí đặt liên kết. Chúng tôi đã thực hiện một thí nghiệm, kiểm tra ba vị trí khác nhau của liên kết tải về sách giáo khoa này trên trang chính của sách để xem vị trí nào, nếu có, dẫn đến số lượt tải nhiều nhất. Số lượng khách truy cập trang tham gia thí nghiệm là 701 và được thể hiện trong một trong các tổ hợp phản hồi trong bảng dưới đây:

	Tải về	Không tải về
Vị trí 1	13,8%	18,3%
Vị trí 2	14,6%	18,5%
Vị trí 3	12,1%	2,7%

a) Tính số lượng thực tế khách truy cập trang trong mỗi trong sáu loại phản hồi.
b) Mỗi cá nhân tham gia thí nghiệm đều có cơ hội bằng nhau để thuộc vào bất kỳ nhóm thí nghiệm nào trong ba nhóm. Tuy nhiên, chúng ta thấy rằng tổng số khách trong các nhóm hơi khác nhau. Có bằng chứng nào cho thấy các nhóm thực sự không cân bằng không? Hãy nêu rõ giả thuyết, kiểm tra các điều kiện áp dụng, tính toán thống kê kiểm định phù hợp và giá trị p, rồi đưa ra kết luận trong bối cảnh dữ liệu.
c) Thực hiện một kiểm định giả thuyết phù hợp để xác định xem có bằng chứng nào cho thấy tỷ lệ khách truy cập nhấp vào liên kết tải sách giáo khoa cao hơn ở bất kỳ nhóm nào trong ba nhóm hay không.

6.38 Gửi quà trong dịp lễ. Một khảo sát do báo địa phương thực hiện hỏi 500 cư dân Los Angeles được chọn ngẫu nhiên về hãng vận chuyển mà họ ưa chuộng để gửi quà trong dịp lễ. Bảng dưới đây trình bày phân bố phản hồi theo nhóm tuổi, cũng như giá trị kỳ vọng cho từng ô (được ghi trong ngoặc đơn).

		Tuổi						Tổng
		18-34		35-54		55+		
Phương thức gửi quà	USPS	72	(81)	97	(102)	76	(62)	245
	UPS	52	(53)	76	(68)	34	(41)	162
	FedEx	31	(21)	24	(27)	9	(16)	64
	Hãng khác	7	(5)	6	(7)	3	(4)	16
	Không chắc	3	(5)	6	(5)	4	(3)	13
	Tổng	165		209		126		500

a) Nêu giả thuyết không và giả thuyết thay thế để kiểm định sự độc lập giữa độ tuổi và phương thức vận chuyển ưa thích khi gửi quà trong dịp lễ của cư dân Los Angeles.
b) Kiểm tra các điều kiện áp dụng kiểm định chi bình phương.

6.40 Nội chiến Hoa Kỳ. Một khảo sát toàn quốc được tiến hành trên mẫu ngẫu nhiên đơn giản gồm 1.507 người trưởng thành cho thấy 56% người Mỹ cho rằng Nội chiến vẫn có ý nghĩa đối với chính trị và đời sống chính trị Mỹ.[47]

a) Thực hiện kiểm định giả thuyết để xác định xem dữ liệu này có cung cấp bằng chứng mạnh mẽ rằng đa số người Mỹ nghĩ Nội chiến vẫn có ý nghĩa hay không
b) Giải thích ý nghĩa của giá trị p trong bối cảnh này.
c) Tính khoảng tin cậy 90% cho tỷ lệ người Mỹ cho rằng Nội chiến vẫn có ý nghĩa. Giải thích khoảng tin cậy này trong bối cảnh nghiên cứu, và nhận xét xem khoảng tin cậy có phù hợp với kết luận của kiểm định giả thuyết hay không.

[47] Pew Research Center Publications, Civil War at 150: Still Relevant, Still Divisive, data collected between March 30 - April 3, 2011.

6.4. KIỂM ĐỊNH TÍNH ĐỘC LẬP TRONG BẢNG HAI CHIỀU

6.41 Sinh viên đại học hút thuốc. Chúng ta quan tâm đến việc ước lượng tỷ lệ sinh viên tại một trường đại học hút thuốc. Trong một mẫu ngẫu nhiên gồm 200 sinh viên từ trường này, có 40 sinh viên hút thuốc.

a) Tính khoảng tin cậy 95% cho tỷ lệ sinh viên tại trường đại học này hút thuốc, và giải thích ý nghĩa của khoảng tin cậy trong bối cảnh nghiên cứu. (Nhắc nhở: Kiểm tra các điều kiện áp dụng trước khi tính toán.)

b) Nếu muốn sai số biên không vượt quá 2% ở mức tin cậy 95% cho tỷ lệ sinh viên hút thuốc, cần mẫu khảo sát lớn bao nhiêu người?

6.42 Acetaminophen và tổn thương gan. Người ta tin rằng liều cao acetaminophen (thành phần hoạt chất trong các thuốc giảm đau không kê đơn như Tylenol) có thể gây tổn thương gan. Một nhà nghiên cứu muốn tiến hành một nghiên cứu để ước lượng tỷ lệ người sử dụng acetaminophen bị tổn thương gan. Để tham gia nghiên cứu, mỗi đối tượng sẽ được trả 20 USD và nhận khám tư vấn y tế miễn phí nếu bệnh nhân bị tổn thương gan.

a) Nếu ông ấy muốn giới hạn sai số biên của khoảng tin cậy 98% trong vòng 2%, thì số tiền tối thiểu cần dự trữ để trả cho các đối tượng tham gia là bao nhiêu?

b) Số tiền tính được ở phần (a) vượt quá ngân sách, vì vậy ông quyết định sử dụng ít đối tượng hơn. Việc này sẽ ảnh hưởng như thế nào đến độ rộng của khoảng tin cậy?

6.43 Cuộc sống sau đại học. Chúng ta quan tâm đến việc ước lượng tỷ lệ sinh viên tốt nghiệp tại một trường đại học vừa phải tìm được việc làm trong vòng một năm sau khi hoàn thành bằng cử nhân. Giả sử chúng ta tiến hành khảo sát và thấy rằng 348 trong số 400 sinh viên tốt nghiệp được chọn ngẫu nhiên đã tìm được việc làm. Lớp tốt nghiệp đang xét có hơn 4.500 sinh viên.

a) Mô tả tham số tổng thể mà chúng ta quan tâm. Giá trị của ước lượng điểm cho tham số này là bao nhiêu?

b) Kiểm tra xem các điều kiện để xây dựng khoảng tin cậy dựa trên dữ liệu này có được thỏa mãn hay không.

c) Tính khoảng tin cậy 95% cho tỷ lệ sinh viên tốt nghiệp tìm được việc làm trong vòng một năm sau khi hoàn thành bằng cử nhân tại trường này, và giải thích ý nghĩa khoảng tin cậy trong bối cảnh dữ liệu.

d) "Tin cậy 95%" có ý nghĩa gì?

e) Tính khoảng tin cậy 99% cho cùng tham số và giải thích ý nghĩa trong bối cảnh dữ liệu.

f) So sánh độ rộng của khoảng tin cậy 95% và 99%. Khoảng tin cậy nào rộng hơn? Giải thích lý do.

6.44 Bệnh tiểu đường và tình trạng thất nghiệp. Một cuộc khảo sát của Gallup tiến hành trên người Mỹ về tình trạng việc làm và việc có bị tiểu đường hay không. Kết quả khảo sát cho thấy: 1,5% trong số 47 774 người đang đi làm (toàn thời gian hoặc bán thời gian) và 2,5% trong số 5855 người thất nghiệp trong độ tuổi 18–29 bị tiểu đường. [48]

a) Tạo một bảng hai chiều trình bày kết quả của nghiên cứu này.

b) Nêu giả thuyết phù hợp để kiểm định sự khác biệt về tỷ lệ mắc tiểu đường giữa những người Mỹ có việc làm và những người thất nghiệp.

c) Sự khác biệt mẫu khoảng 1%. Nếu chúng ta tiến hành kiểm định giả thuyết, giá trị p rất nhỏ (gần bằng 0), nghĩa là sự khác biệt có ý nghĩa thống kê. Sử dụng kết quả này để giải thích sự khác nhau giữa ý nghĩa thống kê và ý nghĩa thực tiễn.

6.45 Kéo-búa-bao. Kéo-búa-bao là một trò chơi bằng tay được chơi bởi hai hoặc nhiều người, trong đó người chơi chọn ký hiệu tay là kéo, búa hoặc bao. Trong dự án môn thống kê, bạn muốn đánh giá xem người chơi chọn giữa ba tùy chọn này có ngẫu nhiên hay không, hay một số tùy chọn được ưa chuộng hơn các tùy chọn khác. Bạn yêu cầu hai người bạn chơi kéo-búa-bao và đếm số lần mỗi tùy chọn được chọn. Bảng dưới đây tóm tắt dữ liệu:

Búa	Bao	Kéo
43	21	35

Sử dụng dữ liệu này để đánh giá xem người chơi có chọn ngẫu nhiên giữa ba tùy chọn hay không, hay một số tùy chọn được ưa thích hơn. Hãy trình bày rõ ràng từng bước phân tích và giải thích kết quả trong bối cảnh dữ liệu và câu hỏi nghiên cứu

[48] Gallup Wellbeing, Employed Americans in Better Health Than the Unemployed, data collected Jan. 2, 2011 - May 21, 2012.

6.4. KIỂM ĐỊNH TÍNH ĐỘC LẬP TRONG BẢNG HAI CHIỀU

6.46 Luật chăm sóc sức khỏe 2010. Vào ngày 28 tháng 6 năm 2012, Tòa án Tối cao Hoa Kỳ đã giữ nguyên luật chăm sóc sức khỏe 2010 đang gây nhiều tranh cãi, tuyên bố luật này hợp hiến. Một khảo sát Gallup công bố ngày hôm sau cho thấy 46% trong số 1.012 người Mỹ đồng ý với quyết định này. Ở mức tin cậy 95%, mẫu khảo sát này có sai số biên là 3%. Dựa trên thông tin này, xác định các phát biểu sau là đúng hay sai, và giải thích lý do của bạn.[49]

a) Chúng ta tự tin 95% rằng từ 43% đến 49% người Mỹ trong mẫu này ủng hộ quyết định của Tòa án Tối cao Hoa Kỳ về luật chăm sóc sức khỏe 2010.
b) Chúng ta tự tin 95% rằng từ 43% đến 49% tất cả người Mỹ ủng hộ quyết định của Tòa án Tối cao Hoa Kỳ về luật chăm sóc sức khỏe 2010.
c) Nếu chúng ta xét nhiều mẫu ngẫu nhiên gồm 1.012 người Mỹ, và tính tỷ lệ mẫu của những người ủng hộ quyết định của Tòa án Tối cao, thì 95% các tỷ lệ mẫu đó sẽ nằm trong khoảng từ 43% đến 49%.
d) Sai số biên ở mức tin cậy 90% sẽ lớn hơn 3%.

6.47 Duyệt web trên thiết bị di động. Một khảo sát trên 2 254 người trưởng thành tại Mỹ cho thấy 17% chủ sở hữu điện thoại di động duyệt Internet chỉ trên điện thoại của họ, thay vì sử dụng máy tính hoặc các thiết bị khác. [50]

a) Theo một bài viết trực tuyến, một báo cáo từ một công ty nghiên cứu di động cho thấy 38% người dùng Internet di động ở Trung Quốc chỉ truy cập Internet qua điện thoại di động của họ.[51] Tiến hành kiểm định giả thuyết để xác định xem dữ liệu này có cung cấp bằng chứng mạnh mẽ rằng tỷ lệ người Mỹ chỉ dùng điện thoại để truy cập Internet khác với tỷ lệ 38% của Trung Quốc hay không.
b) Giải thích ý nghĩa của giá trị p trong bối cảnh này.
c) Tính khoảng tin cậy 95% cho tỷ lệ người Mỹ chỉ truy cập Internet qua điện thoại di động, và giải thích ý nghĩa khoảng tin cậy trong bối cảnh nghiên cứu.

6.48 Cà phê và trầm cảm. Các nhà nghiên cứu tiến hành một nghiên cứu để khám phá mối quan hệ giữa việc tiêu thụ cà phê có caffein và nguy cơ trầm cảm ở phụ nữ. Họ thu thập dữ liệu trên 50739 phụ nữ không có triệu chứng trầm cảm vào thời điểm bắt đầu nghiên cứu năm 1996, và theo dõi những phụ nữ này đến năm 2006. Các nhà nghiên cứu sử dụng bảng câu hỏi để thu thập dữ liệu về tiêu thụ cà phê có caffein, hỏi từng cá nhân về trầm cảm được chẩn đoán bởi bác sĩ, và cũng hỏi về việc sử dụng thuốc chống trầm cảm. Bảng dưới đây trình bày phân bố tỷ lệ mắc trầm cảm theo mức tiêu thụ cà phê có caffein. [52]

		Mức tiêu thụ cà phê có caffein					
		≤1 cốc/tuần	2–6 cốc/tuần	1 cốc/ngày	2–3 cốc/ngày	≥4 cốc/ngày	Tổng
Triệu chứng trầm cảm	Có	670	373	905	564	95	2 607
	Không	11 545	6 244	16 329	11 726	2 288	48 132
	Tổng	12 215	6 617	17 234	12 290	2 383	50 739

a) Loại kiểm định nào là phù hợp để đánh giá xem có mối liên hệ giữa lượng tiêu thụ cà phê và trầm cảm hay không?
b) Viết giả thuyết cho kiểm định được xác định ở phần (a).
c) Tính tỷ lệ tổng thể của phụ nữ có trầm cảm và không có trầm cảm.
d) Xác định giá trị kỳ vọng cho ô được làm nổi bật, và tính đóng góp của ô này vào thống kê kiểm định, tức là $(giá\ trị\ quan\ sát\ -\ giá\ trị\ kỳ\ vọng)^2/giá\ trị\ kỳ\ vọng$.
e) Thống kê kiểm định là $X^2 = 20{,}93$. Giá trị p là bao nhiêu?
f) Kết luận của kiểm định giả thuyết là gì?
g) Một trong các tác giả của nghiên cứu được trích dẫn trên NYTimes rằng "quá sớm để khuyên phụ nữ uống thêm cà phê" dựa chỉ trên nghiên cứu này. [53] Bạn có đồng ý với phát biểu này không? Giải thích lý do.

[49] Gallup, Americans Issue Split Decision on Healthcare Ruling, data collected June 28, 2012.
[50] Pew Internet, Cell Internet Use 2012, data collected between March 15 - April 13, 2012.
[51] S. Chang. "The Chinese Love to Use Feature Phone to Access the Internet". In: *M.I.C Gadget* (2012).
[52] M. Lucas et al. "Coffee, caffeine, and risk of depression among women". In: *Archives of internal medicine* 171.17 (2011), p. 1571.
[53] A. O'Connor. "Coffee Drinking Linked to Less Depression in Women". In: *New York Times* (2011).