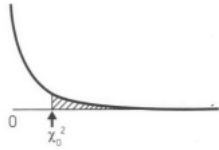


# □□□□ HỌC THỐNG KÊ QUA TRUYỆN TRANH □□□□

Khi  $\chi_0^2$  lớn hơn, xác suất hiển thị ở vùng gạch chéo trong Hình 7-3 tự nhiên sẽ thu hẹp lại hơn.

Hình 7-3: Xác suất tương ứng với  $\chi_0^2$

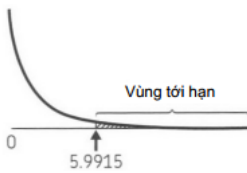


Trong kiểm định Chi bình phương về tính độc lập, nếu xác suất hiển thị là vùng gạch chéo trong Hình 7-3 nhỏ hơn giá trị được gọi là *mức ý nghĩa*, bác bỏ giả thuyết không và kết luận rằng "hệ số Cramer cho tổng thể là lớn hơn 0". Nói chung, mức ý nghĩa (thường gọi là giá trị alpha và ký hiệu bằng biểu tượng  $\alpha$ ) là 0,05 hoặc 0,1.

Tùy thuộc vào nhà phân tích mức độ quan trọng sẽ sử dụng. Giả sử chúng ta quyết định sử dụng 0,05 làm mức ý nghĩa trong trường hợp này. Trên thực tế, mức ý nghĩa là xác suất được biểu thị vùng gạch chéo trong Hình 7-3.

Vùng gạch chéo trong Hình 7-4 gọi là vùng tới hạn.

Hình 7-4: Vùng tới hạn (khi mức ý nghĩa bằng 0,05)



## ! Trả lời

### Bước 1

Xác định tổng thể

Tổng thể là:

Tất cả học sinh trung học phổ thông cư trú tại Nhật Bản.



Trong thử nghiệm này, tổng thể được xác định là "Tất cả học sinh trung học phổ thông cư trú tại Nhật Bản". Do đó, cụ thể trong thử nghiệm này, bước 1 là không cần thiết.

Tuy nhiên, "để kiểm định sự khác nhau giữa tỷ lệ tổng thể trong bảng ở phần đầu chương, tổng thể trong câu hỏi là "cư trú cư trú ở khu vực thành thị" và "cư trú cư trú ở khu vực nông thôn". Các khu vực thành thị chính xác là ở đâu? Đó là Tokyo hay Osaka? Hoặc nó là tỉnh lý/huyện lý? Điều này nhà phân tích phải chỉ ra.

Anh nhắc lại: Khi em thực hiện một kiểm định giả thuyết, em phải xác định tổng thể. Không quan trọng em đang cố gắng thử nghiệm giả thuyết nào, em không được xác định nhầm tổng thể.

Nếu không, em có thể rơi vào một tình huống khiến em bị bối rối, tự hỏi rằng "Mình đang cố gắng để ước tính cái gì?" Rất nhiều số liệu thống kê rơi vào tình huống như thế này, cần chú ý về điểm này.



### Bước 2

Xây dựng một giả thuyết không và một giả thuyết thay thế

Giả thuyết không là: "Hệ số Cramer cho tổng thể = 0. Nói cách khác, giới tính và cách mong muốn được hỏi không có liên hệ tương quan".

Giả thuyết thay thế là: "Hệ số Cramer cho tổng thể > 0. Nói cách khác, giới tính và cách mong muốn được hỏi có liên hệ tương quan".



Giả thuyết không và giả thuyết thay thế sẽ được giải thích ở mục 3 chương này.

### Bước 3

Lựa chọn kiểm định giả thuyết nào để thực hiện.

Mình sẽ thực hiện kiểm định Chi bình phương về tính độc lập.



Thử nghiệm này yêu cầu em thực hiện kiểm định Chi bình phương về tính độc lập. Vì vậy, trong thử nghiệm này, bước 3 là không cần thiết.

(Khi em thực sự làm một kiểm định giả thuyết chứ không phải là thử nghiệm này, em phải tự mình lựa chọn kiểm định giả thuyết phù hợp với mục tiêu phân tích).

### Bước 4

Xác định mức ý nghĩa.

Tôi sẽ sử dụng mức ý nghĩa là 0,05.



Thử nghiệm này sử dụng mức ý nghĩa 0,05, nên bước 4 có thể bỏ qua. Nhưng khi em thực hiện một kiểm định giả thuyết, em cần xác định mức ý nghĩa. Như đã đề cập trước đó, thông thường dùng 0,05 hoặc 0,01. Giá trị p-value được tính toán từ dữ liệu mẫu càng nhỏ thì bằng chứng chống lại giả thuyết không càng mạnh. Nói chung, ký hiệu  $\alpha$  được dùng để hiện mức ý nghĩa (giá trị  $\alpha$ ).

### Bước 5

Tính toán kiểm định thống kê từ dữ liệu mẫu.

Ta thực hiện kiểm định Chi bình phương về tính độc lập. Do đó, kiểm định thống kê là kiểm định thống kê Chi bình phương của Pearson ( $\chi_0^2$ ). Giá trị của  $\chi_0^2$  cho thử nghiệm này đã được tính ở Chương 6,  $\chi_0^2 = 8,0091$ .





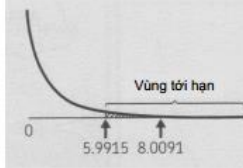
Kiểm định thống kê được lấy từ một hàm tính giá trị từ dữ liệu mẫu. Các loại kiểm định giả thuyết khác nhau có kiểm định thống kê khác nhau. Như đã đề cập trước đó, giá trị cho kiểm định tính độc lập là  $\chi_0^2$ , và trong trường hợp kiểm định về sự tương quan (ở đầu chương), kiểm định thống kê như dưới đây:

$$\frac{\text{Hệ số tương quan} \times \sqrt{(\text{số lượng giá trị} - 2)}}{1 - \sqrt{\text{Hệ số tương quan}^2}}$$

## Bước 6

Xác định kiểm định thống kê ở bước 5 có ở trong vùng tới hạn hay không

Kiểm định thống kê Chi bình phương của Pearson ( $\chi_0^2$ ) là 8.0091. Với mức ý nghĩa ( $\alpha$ ) là 0,05, vùng tới hạn là 5,9915 trở lên, theo bảng phân phối Chi bình phương đã trình bày ở Chương 5. Được hiển thị như hình dưới đây, kiểm định thống kê là trong vùng tới hạn.



Vùng tới hạn thay đổi tùy theo mức ý nghĩa ( $\alpha$ ). Nếu  $\alpha$  trong thử nghiệm này là 0,01 thay cho 0,05, thì vùng tới hạn sẽ là 9,2104 trở lên, theo bảng phân phối Chi bình phương đã trình bày ở Chương 5.

## Bước 7

Nếu kiểm định thống kê là trong vùng tới hạn ở bước 6, em bác bỏ giả thuyết không. Ngược lại, em chưa có cơ sở bác bỏ giả thuyết không. Trong trường hợp này, kiểm định thống kê là trong vùng tới hạn.

Do đó, giả thuyết thay thế, "Hệ số Cramer cho tổng thể là lớn hơn 0", là chính xác!



Em không thể kết luận rằng giả thuyết thay thế là hoàn toàn chính xác trong một kiểm định giả thuyết, ngay cả khi kiểm định thống kê nằm trong vùng tới hạn. Kết luận duy nhất em có thể đưa ra là: "Tôi muốn nói rằng giả thuyết thay thế là hoàn toàn đúng... nhưng, nhiều nhất, có khả năng ( $\alpha \times 100$ )% giả thuyết không sẽ đúng".

Đó là tất cả... Em biết...

Nhưng vẫn có điều gì đó ở bước 7 có vẻ chưa được thuyết phục.

Em cảm thấy có thể nói rằng "giả thuyết không là đúng" nếu kiểm định thống kê không nằm trong vùng tới hạn.

Vùng tới hạn

0 5.9915

Rất tiếc, em không thể. Tất cả những gì em có thể nói là "Tôi chưa có đủ cơ sở để bác bỏ giả thuyết không".

Ví dụ, Giả sử  $\chi_0^2$  là 2,5013 trong thử nghiệm mà chúng ta vừa thực hiện.

Vùng tới hạn

0 2.5013 5.9915

Nó không nằm trong vùng tới hạn ư?

Là sao...?

Là điều hiển nhiên, em không chấp nhận giả thuyết không thì "Hệ số Cramer cho tổng thể lớn hơn 0".

Tuy nhiên, em không có cách nào khẳng định giả thuyết không là "Hệ số Cramer cho tổng thể bằng 0".

Đây là một câu chuyện nhỏ có thể giúp em hiểu rõ hơn về điều này.

Giả sử có ai đó đã lấy trộm mắt chiếc bánh pudding mà em để dành.

Ai đã lấy trộm bánh pudding của tôi!?

Yumi trở thành nghi phạm.

Hãy bỏ qua các chi tiết như các loại kiểm định giả thuyết hay mức ý nghĩa...

Giả thuyết không	Yumi không lấy chiếc bánh
Giả thuyết thay thế	Yumi lấy chiếc bánh

Yumi! Tôi tưởng ấy là bạn của tớ chứ!

Đây chỉ là một câu chuyện giả tạo!

Và chỉ làm một kiểm định giả thuyết chống lại những giả thuyết này.

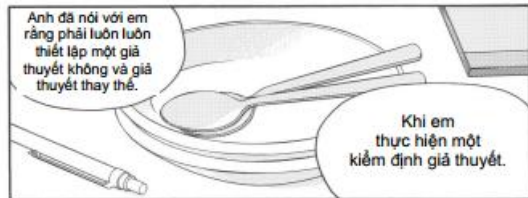
# ➤➤➤ HỌC THỐNG KÊ QUA TRUYỆN TRANH



Ví dụ về kiểm định giả thuyết

Tên	Trường hợp sử dụng
Kiểm định tính độc lập	Ước tính liệu giá trị của hệ số Cramer cho giới tính và cách mong muốn được hỏi có bằng 0 cho tổng thể hay không
Kiểm định tỷ số tương quan	Ước tính liệu giá trị của tỷ số tương quan cho nhân hiệu thời trang yêu thích và tuổi có bằng 0 cho tổng thể hay không
Kiểm định mối tương quan	Ước tính liệu hệ số tương quan cho số tiền chi cho mỹ phẩm và quần áo có bằng 0 cho tổng thể hay không
Kiểm định sự khác nhau giữa các trung bình tổng thể	Ước tính liệu tiền trợ cấp có khác nhau giữa các nữ sinh trung học phổ thông ở Tokyo và Osaka không*
Kiểm định sự khác nhau giữa các tỷ lệ tổng thể	Ước tính liệu sự ủng hộ nội các X có khác nhau giữa các cư trú ở thành thị và nông thôn

\* Chú ý, đây là hai tổng thể đang được xem xét.



## Kiểm định tính độc lập

Giả thuyết không	Hệ số Cramer cho giới tính và cách mong muốn được hỏi cho tổng thể = 0
Giả thuyết thay thế	Hệ số Cramer cho giới tính và cách mong muốn được hỏi cho tổng thể > 0

## Kiểm định tỷ số tương quan

Giả thuyết không	Tỷ số tương quan cho nhân hiệu thời trang yêu thích và tuổi cho tổng thể = 0
Giả thuyết thay thế	Tỷ số tương quan cho nhân hiệu thời trang yêu thích và tuổi cho tổng thể > 0

## Kiểm định mối tương quan

Giả thuyết không	Hệ số tương quan cho số tiền chi cho mỹ phẩm và quần áo cho tổng thể = 0
Giả thuyết thay thế	Hệ số tương quan cho số tiền chi cho mỹ phẩm và quần áo cho tổng thể # 0
	Hoặc
	Hệ số tương quan cho số tiền chi cho mỹ phẩm và quần áo cho tổng thể > 0
Giả thuyết thay thế	Hoặc
	Hệ số tương quan cho số tiền chi cho mỹ phẩm và quần áo cho tổng thể < 0

## Kiểm định sự khác nhau giữa các trung bình tổng thể

Giá thuyết không	Tiền trợ cấp cho nữ sinh trung học phổ thông ở Tokyo và Osaka là như nhau
Giá thuyết thay thế	Tiền trợ cấp cho nữ sinh trung học phổ thông ở Tokyo và Osaka là khác nhau Hoặc Tiền trợ cấp cho nữ sinh trung học phổ thông ở Tokyo lớn hơn so với nữ sinh ở Osaka Hoặc Tiền trợ cấp cho nữ sinh trung học phổ thông ở Tokyo nhỏ hơn so với nữ sinh ở Osaka

## Kiểm định sự khác nhau giữa các tỷ lệ tổng thể

Giá thuyết không	Sự ủng hộ nội các X của cử tri cư trú ở thành thị và nông thôn là như nhau
Giá thuyết thay thế	Sự ủng hộ nội các X của cử tri cư trú ở thành thị và nông thôn là khác nhau Hoặc Sự ủng hộ nội các X của cử tri cư trú ở thành thị cao hơn cử tri cư trú ở nông thôn Hoặc Sự ủng hộ nội các X của cử tri cư trú ở thành thị thấp hơn cử tri cư trú ở nông thôn

Đúng là giải thích hay!



## 4. P-value và quá trình kiểm định giá thuyết



Biên dịch: Lan Phương