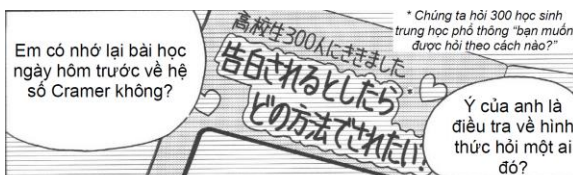
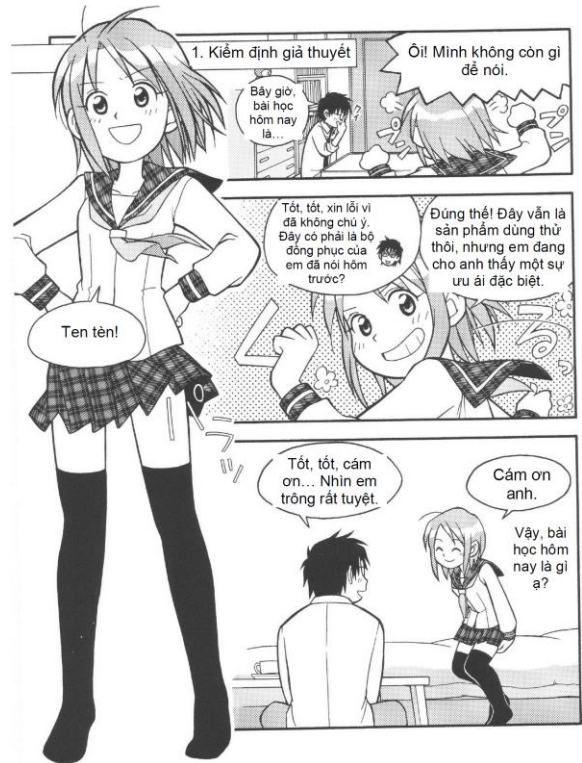


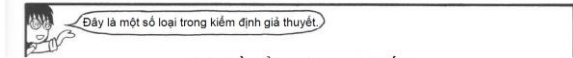
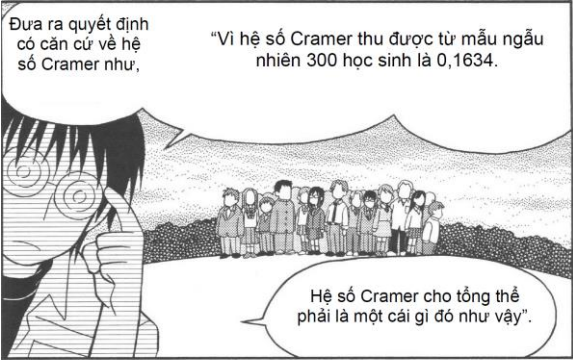
□□□□ HỌC THỐNG KÊ QUA TRUYỆN TRANH □□□□

CHƯƠNG 7:

HÃY TÌM HIỂU VỀ KIỂM ĐỊNH GIẢ THUYẾT



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 hệ 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 cử tri cư trú ở thành thị và nông thôn

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



Quy trình kiểm định giả thuyết	
Bước 1	Xác định tổng thể
Bước 2	Xây dựng một giả thuyết không (H_0) và một giả thuyết thay thế (H_1)
Bước 3	Lựa chọn giả thuyết kiểm định để thực hiện
Bước 4	Xác định mức ý nghĩa (α)
Bước 5	Thu được kiểm định thống kê từ dữ liệu mẫu (P-value)
Bước 6	Xác định xem P-value thu được ở bước 5 nằm trong vùng tới hạn
Bước 7	Nếu P-value nhỏ hơn mức ý nghĩa α , em bác bỏ H_0 , ngược lại, em chưa có cơ sở bác bỏ H_0 .



HỌC THỐNG KÊ QUA TRUYỆN TRANH

2. Kiểm định Chi bình phương về tính độc lập

Bây giờ, anh sẽ giải thích điều quan trọng nhất ngày hôm nay, đó là kiểm định tính độc lập.

Anh đã bảo với em kiểm định tính độc lập là một kỹ thuật phân tích dùng để ước tính liệu hệ số Cramer cho tổng thể có bằng 0 hay không.

Vâng, anh đã bảo với em.

Em hiểu, đó là lý do tại sao nó được áp dụng để phân tích điều tra.

Nói cách khác, đây là một kỹ thuật phân tích dùng để ước tính liệu giữa 2 biến trong bảng chéo là tương quan với nhau.

		Phương pháp muốn được hỏi			
Giới tính		Điện thoại	Email	Mặt đối mặt	Sum
Nam	34	61	53	148	
Nữ	38	40	74	152	
Sum	72	101	127		

Kiểm định tính độc lập này còn được gọi là Kiểm định Chi bình phương.

Không phải Chi... gì đó nữa!

Giải thích
Kiểm định thống kê Chi bình phương của Pearson (χ^2) và phân phối Chi bình phương



Trước khi đưa ra một ví dụ thực tế về kiểm định tính độc lập. Anh muốn giải thích một điều quan trọng - đây là nền tảng cho việc kiểm định tính độc lập. Mặc dù không thể làm điều này trong thực tế, giả sử ví dụ dưới đây được thực hiện.

Bước 1:

Chọn ngẫu nhiên 300 học sinh từ tổng thể "Tất cả học sinh trung học phổ thông cư trú tại Nhật Bản".



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

300 học sinh

Bước 2:

Tiến hành điều tra 300 người được chọn mẫu ở bước 1 để tính được thống kê Chi bình phương (χ^2).

Bước 3:

Đưa 300 người quay trở lại tổng thể.

Bước 4:

Thực hiện lặp lại bước 1 đến bước 3 nhiều lần.

Trong ví dụ này, nếu giá trị của hệ số Cramer cho tổng thể "Tất cả học sinh trung học phổ thông cư trú tại Nhật Bản" bằng 0, đó thì của kiểm định thống kê Chi bình phương là một phân phối Chi bình phương với bậc tự do bằng 2. Nói cách khác, nếu giá trị của hệ số Cramer cho tổng thể "Tất cả học sinh trung học phổ thông cư trú tại Nhật Bản" bằng 0, thì kiểm định thống kê Chi bình phương của Pearson (χ^2) là phân phối Chi bình phương với bậc tự do bằng 2.

Xem chương 6 để biết về kiểm định thống kê Chi bình phương của Pearson (χ^2) và phân phối Chi bình phương với bậc tự do bằng 2.

Chúng ta cùng thực hiện thử nghiệm này. Khi thực hiện, chúng ta đưa ra các hạn chế dưới đây.



- Vì không thể thử nghiệm với tổng thể thực tế của "Tất cả học sinh trung học phổ thông cư trú tại Nhật Bản", thay vào đó, nhóm 10.000 người trong Bảng 7-1 sẽ được coi là "Tất cả học sinh trung học phổ thông cư trú tại Nhật Bản".
- Chúng ta giả sử rằng hệ số Cramer cho "Tất cả học sinh trung học phổ thông cư trú tại Nhật Bản" bằng 0. Điều này có nghĩa là tỷ lệ của những người thích được hỏi qua điện thoại so với những người thích được hỏi qua Email với những người thích được hỏi trực tiếp là bằng nhau giữa nam và nữ. Bảng chéo của Bảng 7-1 là Bảng 7-2.
- Vì nó là vô tận, chúng ta sẽ dừng lặp lại các bước 1 đến bước 3 sau 10.000 lần.

Bảng 7-1: Cách mong muốn được hỏi

Người trả lời	Giới tính	Cách mong muốn được hỏi
1	Nữ	Trực tiếp
2	Nữ	Điện thoại
...	...	...
10.000	Nam	Email

Bảng 7-2: Bảng chéo giữa giới tính và cách mong muốn được hỏi

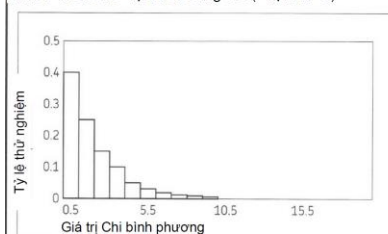
		Cách mong muốn được hỏi			Tổng
		Điện thoại	Email	Trực tiếp	
Giới tính	Nữ	400	1.600	2.000	4.000
	Nam	600	2.400	3.000	6.000
Tổng		1.000	4.000	5.000	10.000

Bảng 7-3 thể hiện kết quả thử nghiệm. Hình 7-1 là biểu đồ theo Bảng 7-3

Bảng 7-3: Kết quả thử nghiệm

Thử nghiệm	Kiểm định thống kê Chi bình phương của Pearson (χ^2)
1	0,8598
2	0,7557
...	...
10.000	2,7953

Hình 7-1: Biểu đồ dựa trên Bảng 7-3 (Phạm vi = 1)



Hình 7-1 thực sự trông tương tự như biểu đồ ở Chương 5, "bậc tự do = 2". Dường như Kiểm định thống kê Chi bình phương của Pearson (χ^2) là phân phối Chi bình phương với bậc tự do bằng 2. Nghĩ rằng điều này không liên quan gì đến thí nghiệm, đây là một điểm cần lưu ý. Bậc tự do bằng 2 được xuất phát như sau:

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

2 loại: Nam và Nữ 3 loại: Điện thoại, Email, trực tiếp



Anh sẽ không giải thích tại sao một phép tính kỳ lạ như vậy lại được áp dụng, vì nó là một chủ đề quá cao cấp đối với cấp độ của cuốn sách này. Nhưng đừng lo lắng - ngay cả khi em không hiểu đầy đủ về tính toán, em cũng sẽ không gặp khó khăn nào.

HỌC THỐNG KÊ QUA TRUYỆN TRANH <<<

Minh thấy rằng giá trị của hệ số Cramer cho "Tất cả học sinh trung học phổ thông cư trú tại Nhật Bản" = 0... nghĩa là không có mối quan hệ giữa giới tính và cách mong muốn được hỏi.

Giả sử...

Tỷ lệ ưa thích là như nhau cho cả nam và nữ!

Do đó, mình điều tra 300 học sinh được chọn từ "Tất cả học sinh trung học phổ thông cư trú tại Nhật Bản".

Lập lại điều đó nhiều lần!

Điều tra

Điều tra

Sau đó, mình tính kiểm định thống kê Chi bình phương của Pearson.

Tần số thực tế - tần số dự kiến ở mỗi ô vuông

Tần số dự kiến

Đồ thị thu được từ kết quả là phân phối Chi bình phương với bậc tự do = 2.

Đã xong...

Có câu trả lời!

Anh ấn tượng với những điều mà em đã học được.

Hoan hô!

Bây giờ, sử dụng nó để trả lời cuộc điều tra như trong ví dụ.

Hãy thử thách kiểm định Chi bình phương về tính độc lập.

Anh! Không, không...

Anh sẽ hướng dẫn em bài học hôm nay theo 3 bước:

Bài tập: Hãy nghĩ về nó

Trả lời

Này-Này

Anh sẽ sử dụng hai con rối này để giải thích phần trả lời

Nhà phân tích

Giáo viên

Bởi vì nó dễ dàng hơn để giải thích câu trả lời theo cách này.

A, sao không phải là em?

Em nhìn tốt hơn nó nhiều!

Anh sẽ tự đóng cả hai vai.

Bài tập

Tạp chí P-Girls quyết định xuất bản một bài báo có tiêu đề "Chúng tôi hỏi 300 học sinh trung học phổ thông, 'bạn muốn được hỏi theo phương pháp nào?'. Để chuẩn bị bài báo, một nhà báo đã chọn ngẫu nhiên 300 người từ tất cả các học sinh cư trú tại Nhật Bản tham gia cuộc khảo sát. Bảng dưới đây là kết quả của cuộc khảo sát này.

Người trả lời	Cách mong muốn được hỏi	Tuổi	Giới tính
1	Trực tiếp	17	Nữ
2	Điện thoại	15	Nữ
...	...	...	...
300	Email	18	Nam

Bảng dưới đây là bảng chéo giữa giới tính và cách mong muốn được hỏi.

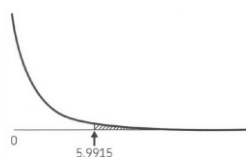
		Cách mong muốn được hỏi			Tổng
		Điện thoại	Email	Trực tiếp	
Giới tính	Nữ	34	61	53	148
	Nam	38	40	74	152
Tổng		72	101	127	300

Sử dụng kiểm định Chi bình phương về sự độc lập, ước tính hệ số Cramer cho giới tính và cách mong muốn được hỏi cho tổng thể "tất cả học sinh trung học phổ thông cư trú tại Nhật Bản" lớn hơn 0. Điều này giống như ước tính với thử nghiệm tính độc lập cho dù giới tính và cách mong muốn được hỏi là tương quan với nhau. Hãy nhớ rằng mức ý nghĩa (giải thích sau) là 0,05.



Hãy nghĩ về nó

Như thử nghiệm ở trên, kiểm định thống kê Chi bình phương của Pearson (χ^2) theo phân phối Chi bình phương với bậc tự do = 2 nếu giả thuyết 0 là giá trị của hệ số Cramer cho tổng thể "tất cả học sinh trung học phổ thông cư trú tại Nhật Bản" = 0. Nếu điều đó là đúng, thì xác suất mà χ^2 có được từ 300 người được chọn ngẫu nhiên là 5,9915 hoặc lớn hơn 0,05.



Điều này thể hiện rõ ràng từ bảng phân phối Chi bình phương.

χ^2 cho thử nghiệm đã được tính bằng 8,0091. Đúng, con số này đã được tính toán dựa trên dữ liệu từ 300 học sinh được chọn ngẫu nhiên, nhưng điều này dường như không quá lớn? Cần nhắc nhận xét ở Chương 6, không phải tự nhiên mà cho rằng hệ số Cramer cho tổng thể "tất cả học sinh trung học phổ thông cư trú tại Nhật Bản" lớn hơn 0?

Hãy nhớ lại quy trình kiểm định Chi bình phương về tính độc lập (không giới hạn trong thử nghiệm) như thế này:

1. Giả sử giả thuyết không rằng "hệ số Cramer cho tổng thể = 0" cho đến lúc này.
2. Tính toán χ^2 từ dữ liệu mẫu.
3. Nếu χ^2 lớn, bác bỏ giả thuyết không và kết luận rằng "hệ số Cramer cho tổng thể là lớn hơn 0".

Biên dịch: Lan Phương