

HỌC THỐNG KÊ QUA TRUYỆN TRANH

CHƯƠNG 6:

HÃY XEM XÉT MỐI QUAN HỆ GIỮA HAI BIÊN

Hệ số Cramer được tính theo từ bước 1 đến bước 5 như dưới đây.



Bước 1

Lập bảng chéo. Các giá trị được bao quanh bởi khung in đậm được gọi là *tần số thực tế*.

		Phương pháp 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

Bước 2

Tính toán ở bảng bên dưới. Các giá trị được bao quanh bởi khung in đậm được gọi là *tần số dự kiến*.

		Phương pháp mong muốn được hỏi			Tổng
		Điện thoại	Email	Trực tiếp	
Giới tính	Nữ	148 x 72	148 x 101	148 x 127	148
		300	300	300	
	Nam	152 x 72	152 x 101	152 x 127	152
		300	300	300 ↑	
Tổng		72	101	127	300

Số người là nam x Số người muốn hỏi trực tiếp
Tổng số người

Công thức A



Nếu giới tính và cách muốn được hỏi không có mối quan hệ, tỷ lệ giữa điện thoại, Email, trực tiếp

$$72:101:127 = \frac{72}{72+101+127} : \frac{101}{72+101+127} : \frac{127}{72+101+127}$$

$$= \frac{72}{300} : \frac{101}{300} : \frac{127}{300}$$

Cả nam và nữ, theo cột tổng trong Bảng ở bước 2. Do đó, tần suất dự kiến của chúng tôi (công thức A) cho thấy số lượng nam dự đoán muốn được hỏi trực tiếp khi không có mối quan hệ giữa giới tính và cách muốn được hỏi là $152 \times (127:300) = (152 \times 127) : 300$, hoặc:

$$152 \times \frac{127}{300} = \frac{152 \times 127}{300} = 64.3$$



Bước 3

Tính $\frac{(\text{Tần số thực tế} - \text{Tần số dự kiến})^2}{\text{Tần số dự kiến}}$ Cho mỗi ô

		Phương pháp mong muốn được hỏi			Tổng
		Điện thoại	Email	Trực tiếp	
Giới tính	Nữ	$\left(\frac{34 - \frac{148 \times 72}{300}}{\frac{148 \times 72}{300}} \right)^2$	$\left(\frac{61 - \frac{148 \times 101}{300}}{\frac{148 \times 101}{300}} \right)^2$	$\left(\frac{53 - \frac{148 \times 127}{300}}{\frac{148 \times 127}{300}} \right)^2$	148
	Nam	$\left(\frac{38 - \frac{152 \times 72}{300}}{\frac{152 \times 72}{300}} \right)^2$	$\left(\frac{40 - \frac{152 \times 101}{300}}{\frac{152 \times 101}{300}} \right)^2$	$\left(\frac{74 - \frac{152 \times 127}{300}}{\frac{152 \times 127}{300}} \right)^2$	152
Tổng		72	101	127	300



Khoảng cách giữa tần số thực tế và tần số dự kiến càng lớn, các giá trị trong mỗi ô vuông càng lớn.

➤➤➤ HỌC THỐNG KÊ QUA TRUYỆN TRANH

Bước 4

Tính tổng giá trị bên trong khung in đậm trong bảng ở Bước 3. Giá trị này được gọi là kiểm định Chi bình phương của Pearson. Từ bây giờ được viết là χ_0^2

$$\chi_0^2 = \frac{\left(34 - \frac{148 \times 72}{300}\right)^2}{\frac{148 \times 72}{300}} + \frac{\left(61 - \frac{148 \times 101}{300}\right)^2}{\frac{148 \times 101}{300}} + \frac{\left(53 - \frac{148 \times 127}{300}\right)^2}{\frac{148 \times 127}{300}} + \frac{\left(38 - \frac{152 \times 72}{300}\right)^2}{\frac{152 \times 72}{300}} + \frac{\left(40 - \frac{152 \times 101}{300}\right)^2}{\frac{152 \times 101}{300}} + \frac{\left(74 - \frac{152 \times 127}{300}\right)^2}{\frac{152 \times 127}{300}} = 8.0091$$

Như có thể hiểu từ Bước 3, các đại lượng thực tế càng khác với tần số dự kiến của chúng, hoặc mối tương quan giữa giới tính và cách hỏi càng lớn, thì giá trị kiểm định Chi bình phương của Pearson χ_0^2 càng lớn.



Bước 5

Tính hệ số Cramer

$$\sqrt{\frac{\chi_0^2}{\text{Tổng các giá trị } x \text{ (giá trị nhỏ hơn của (số lượng dòng trong bảng chéo, số lượng cột trong bảng chéo) - 1)}}$$

min {a, b} nghĩa là "giá trị nào nhỏ hơn, a hoặc b"

$$\sqrt{\frac{8.0091}{300 \times \min\{2,3\} - 1}} = \sqrt{\frac{8.0091}{300 \times (2 - 1)}} = \sqrt{\frac{8.0091}{300}} = 0.1634$$

Do đó, hệ số Cramer là 0,1634



Như đã giải thích trước đó, hệ số Cramer là giá trị nằm giữa 0 và 1. Mối tương quan giữa hai biến càng mạnh, giá trị càng gần 1 và mối tương quan càng yếu, giá trị càng gần 0. Xem bảng chéo (mong muốn theo chiều ngang) dưới đây để biết thêm chi tiết.



Dưới đây là bảng chéo giữa giới tính là phương pháp muốn được hỏi (mong muốn theo chiều ngang) khi giá trị của Hệ số Cramer bằng 1.

		Phương pháp mong muốn được hỏi			Tổng
		Điện thoại	Email	Trực tiếp	
Giới tính	Nữ	17%	83%	0%	100%
	Nam	0%	0%	100%	100%

Hệ số Cramer = 1 ↔ Mong muốn giữa nam và nữ hoàn toàn khác nhau

Dưới đây là bảng chéo giữa giới tính là phương pháp muốn được hỏi (mong muốn theo chiều ngang) khi giá trị của Hệ số Cramer bằng 0.

		Phương pháp mong muốn được hỏi			Tổng
		Điện thoại	Email	Trực tiếp	
Giới tính	Nữ	17%	48%	35%	100%
	Nam	17%	48%	35%	100%

Hệ số Cramer = 0 ↔ Mong muốn giữa nam và nữ giống nhau



Thật không may, không có tiêu chuẩn thống kê nào như "hai biến có mối tương quan mạnh nếu hệ số Cramer là trên một điểm chuẩn nhất định". Tuy nhiên, tiêu chuẩn không chính thức được đưa ra dưới đây.

Tiêu chuẩn không chính thức của hệ số Cramer

Hệ số Cramer		Mô tả chi tiết	Mô tả đơn giản
1.0-0.8	⇒	Rất chặt chẽ	
0.8-0.5	⇒	Chặt chẽ	Chặt chẽ
0.5-0.25	⇒	Bình thường	
Dưới 0.25	⇒	Không chặt chẽ	Không chặt chẽ

Vì vậy, như kết luận cho ví dụ này, chúng ta có thể nói rằng chúng có mối liên hệ không chặt chẽ.

Bài học hôm nay kết thúc tại đây.

Tiếng vỗ tay!

Em hiểu điều đó.

Cảm ơn anh.

Trong phần cuối bài học hôm nay, anh đã dạy em về hệ số Cramer.

Dựa trên những nội dung anh đã dạy em hôm nay, chúng ta sẽ nghiên cứu các bài kiểm tra về sự độc lập trong bài học tiếp theo.

Kiểm định tính độc lập?

Kiểm tra tính độc lập thường xuyên được sử dụng trong phân tích điều tra.

Một khi em đã thành thạo, em sẽ nắm vững các nguyên tắc cơ bản của thống kê.

Điều đó có nghĩa là bài học tiếp theo của chúng ta sẽ là bài học cuối cùng.

Kết thúc!

Đúng, trong thời điểm này.

Bài tập và trả lời



Bài tập

Công ty X điều hành một nhà hàng ăn uống bình dân. Tình trạng tài chính của nhà hàng gần đây bị sụt giảm. Do đó, Công ty X quyết định nghiên cứu nhu cầu của khách hàng và thực hiện khảo sát những người được chọn ngẫu nhiên, từ 20 tuổi trở lên, sống tại Nhật Bản. Bảng dưới đây thể hiện kết quả của cuộc khảo sát này.

Người trả lời	Bạn thường dùng đồ ăn gì trong nhà hàng	Nếu được phục vụ miễn phí đồ uống sau bữa ăn, bạn thích uống gì? Trà hay cà phê?
1	Trung Quốc	Cà phê
2	Châu Âu	Cà phê
...	...	...
250	Nhật Bản	Trà

Dưới đây là bảng chéo được sử dụng từ bảng trên.

		Ưu tiên trà hay cà phê		Tổng
		Cà phê	Trà	
Loại đồ ăn thường lựa chọn	Nhật Bản	43	33	76
	Châu Âu	51	53	104
	Trung Quốc	29	41	70
Tổng		123	127	250

Tính hệ số Cramer cho loại đồ ăn thường xuyên được lựa chọn trong nhà hàng và đồ uống miễn phí được ưu tiên là trà hay cà phê.

Trả lời

Bước 1:

Lập bảng chéo

		Ưu tiên trà hay cà phê		Tổng
		Cà phê	Trà	
Loại đồ ăn thường lựa chọn	Nhật Bản	43	33	76
	Châu Âu	51	53	104
	Trung Quốc	29	41	70
Tổng		123	127	250

Bước 2:

Tính tần số dự kiến

		Ưu tiên trà hay cà phê		Tổng
		Cà phê	Trà	
Loại đồ ăn thường lựa chọn	Nhật Bản	76×123 250	76×127 250	76
	Châu Âu	104×123 250	104×127 250	104
	Trung Quốc	70×123 250	70×127 250	70
Tổng		123	127	250

➤➤➤ HỌC THỐNG KÊ QUA TRUYỆN TRĂNG

Bước 3:

Tính toán

$$\frac{(\text{Tần số thực tế} - \text{Tần số dự kiến})^2}{\text{Tần số dự kiến}}$$

Cho từng ô vuông

		Ưu tiên trà hay cà phê		Tổng
		Cà phê	Trà	
Loại đồ ăn thường lựa chọn	Nhật Bản	$\left(43 - \frac{76 \times 123}{250}\right)^2$	$\left(33 - \frac{76 \times 127}{250}\right)^2$	76
		$\frac{76 \times 123}{250}$	$\frac{76 \times 127}{250}$	
	Châu Âu	$\left(51 - \frac{104 \times 123}{250}\right)^2$	$\left(53 - \frac{104 \times 127}{250}\right)^2$	104
		$\frac{104 \times 123}{250}$	$\frac{104 \times 127}{250}$	
	Trung Quốc	$\left(29 - \frac{70 \times 123}{250}\right)^2$	$\left(41 - \frac{70 \times 127}{250}\right)^2$	70
		$\frac{70 \times 123}{250}$	$\frac{70 \times 127}{250}$	
Tổng		123	127	250

Bước 4:

Tính tổng các giá trị trong khung in đậm ở bước 3, đây là giá trị kiểm định Chi bình phương của Pearson (χ_0^2).

$$\begin{aligned} \chi_0^2 &= \frac{\left(43 - \frac{76 \times 123}{250}\right)^2}{\frac{76 \times 123}{250}} + \frac{\left(33 - \frac{76 \times 127}{250}\right)^2}{\frac{76 \times 127}{250}} \\ &+ \frac{\left(51 - \frac{104 \times 123}{250}\right)^2}{\frac{104 \times 123}{250}} + \frac{\left(53 - \frac{104 \times 127}{250}\right)^2}{\frac{104 \times 127}{250}} \\ &+ \frac{\left(29 - \frac{70 \times 123}{250}\right)^2}{\frac{70 \times 123}{250}} + \frac{\left(41 - \frac{70 \times 127}{250}\right)^2}{\frac{70 \times 127}{250}} \\ &= 3.3483 \end{aligned}$$

Bước 5:

Tính hệ số Cramer

$$\sqrt{\frac{\chi_0^2}{\text{Tổng các giá trị } x \text{ (giá trị nhỏ hơn của (số lượng dòng trong bảng chéo, số lượng cột trong bảng chéo) - 1)}}$$

$$\sqrt{\frac{3.3483}{250 \times (\min[3, 2] - 1)}} = \sqrt{\frac{3.3483}{250 \times (2 - 1)}} = \sqrt{\frac{3.3483}{250}} = 0.1157$$

Tóm tắt



- Chỉ số được sử dụng để mô tả mức độ tương quan giữa dữ liệu định lượng và dữ liệu định tính là **hệ số tương quan**.
- Chỉ số được sử dụng để mô tả mức độ tương quan giữa dữ liệu định lượng và dữ liệu định tính là **tỷ số tương quan**.
- Chỉ số được sử dụng để mô tả mức độ tương quan giữa dữ liệu định tính và dữ liệu định tính là **hệ số Cramer** (thường được gọi là là V cramer hoặc hệ số độc lập).
- Các đặc điểm của hệ số tương quan, tỷ số tương quan, hệ số Cramer được thể hiện như bảng dưới đây.

	Giá trị nhỏ nhất	Giá trị lớn nhất	Giá trị khi 2 biến không có mối liên hệ	Giá trị khi 2 biến có mối liên hệ chặt chẽ
Hệ số tương quan	-1	1	0	-1 hoặc 1
Tỷ số tương quan	0	1	0	1
Hệ số Cramer	0	1	0	1

- Không có tiêu chuẩn thống kê cho hệ số tương quan, tỷ số tương quan, hệ số Cramer, như là "hai biến có mối tương quan mạnh nếu giá trị là trên một chuẩn mực nhất định".

Biên dịch: Lan Phương