

COURSE

Kiểm định thống kê cho biến kết cuộc định tính

Lớp phân tích thống kê cơ bản

Khương Quỳnh Long
Hà Nội, 06-08/06/2020

Nội dung

- Kiểm định so sánh 1 tỉ lệ
- Kiểm định so sánh 2 tỉ lệ
- Kiểm định Chi bình phương
- Kiểm định Fisher

Kiểm định so sánh 1 tỉ lệ

- Một địa phương khảo sát 1000 phụ nữ tuổi 50-54 thì 40 phụ nữ có ung thư vú. Tỉ lệ quốc gia là khoảng 2%.
- So sánh tỉ lệ ung thư khảo sát được so với tỉ lệ quốc gia?
- Giả thuyết Ho?

Kiểm định so sánh 1 tỉ lệ

- Kiểm định Z 1 mẫu trong Stata

`prtesti n psample ppop`

Ví dụ: `prtesti 1000 0.04 0.02`

One-sample test of proportion		x: Number of obs =		1000
	Mean	Std. Err.	[95% Conf. Interval]	
x	.04	.0061968	.0278545	.0521455
p = proportion(x)		z =		4.5175
Ho: p = 0.02				
Ha: p < 0.02		Ha: p != 0.02		Ha: p > 0.02
Pr(Z < z) = 1.0000		Pr(Z > z) = 0.0000		Pr(Z > z) = 0.0000

Kiểm định so sánh 2 tỉ lệ

- Điều trị bệnh X có 2 phác đồ A và B. Trong 250 người điều trị theo phác đồ A có 70% đáp ứng điều trị. Trong 300 người điều trị theo phác đồ B có 60% đáp ứng điều trị.

→ Phác đồ nào có tỉ lệ đáp ứng điều trị cao hơn?

- Giả thuyết H_0 ?

Kiểm định so sánh 2 tỉ lệ

`prtesti n1 p1 n2 p2`

- Ví dụ: `prtesti 250 0.7 300 0.6`

Two-sample test of proportions						x: Number of obs =	250
						y: Number of obs =	300
	Mean	Std. Err.	z	P> z	[95% Conf. Interval]		
x	.7	.0289828			.6431948	.7568052	
y	.6	.0282843			.5445638	.6554362	
diff	.1	.0404969			.0206275	.1793725	
	under Ho:	.0409656	2.44	0.015			
diff = prop(x) - prop(y)						z =	2.4411
Ho: diff = 0							
Ha: diff < 0		Ha: diff != 0		Ha: diff > 0			
Pr(Z < z) = 0.9927		Pr(Z > z) = 0.0146		Pr(Z > z) = 0.0073			

Ví dụ

- Hai nghiên cứu nhằm xác định tỉ lệ ung thư phổi tại HN và TP.HCM.
- ✓ Nghiên cứu 1 tiến hành trên 5000 người cho tỉ lệ là 10%
- ✓ Nghiên cứu 2 tiến hành trên 6000 người cho tỉ lệ là 11.5%
- Kết quả của 2 nghiên cứu trên là tương đồng hay có sự khác biệt?

`prtesti 5000 0.1 6000 0.115`

Kiểm định Chi bình phương & Fisher

- **Bước 1:** Lập bảng $n \times m$ và kiểm tra vọng trị của các ô
- **Bước 2:** chọn lựa kiểm định
 - Nếu có trên 20% số ô có giá trị vọng trị < 5 thì dùng kiểm định Fisher (một số tác giả gợi ý ≥ 25)
 - Nếu tần số trong ô quá nhỏ (< 5) thì nên dùng kiểm định Fisher
 - Nếu không có cả 2 điều kiện trên thì dùng chi bình phương
- **Bước 3:** tiến hành kiểm định với câu lệnh tương ứng

Vọng trị

- Vọng trị # giá trị kỳ vọng # giá trị mong đợi
- ✓ Là trung bình của biến số ngẫu nhiên nếu phép thử được lập lại nhiều lần
- ✓ Tần số mong đợi nếu 2 biến định tính không có mối liên quan

Vọng trị

	Bệnh	Không bệnh	Tổng
Phơi nhiễm	a	b	$a + b \# n_1$
Không PN	c	d	$c + d \# n_2$
Tổng	$a + c \# m_1$	$b + d \# m_2$	N

$$E = \frac{\text{Tổng của cột} \times \text{Tổng của hàng}}{\text{Tổng số chung}}$$

$$E_a = (m_1 \times n_1) / N$$

$$E_b = (m_2 \times n_1) / N$$

$$E_c = (m_1 \times n_2) / N$$

$$E_d = (m_2 \times n_2) / N$$

Vọng trị

	Bệnh	Không bệnh	Tổng
Phơi nhiễm	a	b	$a + b \# n_1$
Không PN	c	d	$c + d \# n_2$
Tổng	$a + c \# m_1$	$b + d \# m_2$	N

$$E = \frac{\text{Tổng của cột} \times \text{Tổng của hàng}}{\text{Tổng số chung}}$$

- Nếu không có liên quan thì tỉ lệ bệnh trong nhóm có phơi nhiễm và không phơi nhiễm là như nhau

$$\frac{a}{a + c} = \frac{b}{b + d} = \frac{a + b}{N}$$

Vọng trị

- Nếu trên 20% giá trị vọng trị < 5 thì dùng kiểm định Fisher (một số tác giả gợi ý ≥ 25)

20.1	15.5
4.1	13.5

- Số vọng trị < 5 : 1 Tổng số ô trong bảng: 4
- Tỷ lệ vọng trị $< 5 = 1 / 4 = 25\% \rightarrow$ Nên dùng Fisher

Vọng trị

- Nếu trên 20% giá trị vọng trị < 5 thì dùng kiểm định Fisher (một số tác giả gợi ý ≥ 25)

20.1	15.5
4.1	13.5
10.2	12.6

- Số vọng trị < 5 : 1 Tổng số ô trong bảng: 6
- Tỉ lệ vọng trị $< 5 = 1 / 6 = 17\% \rightarrow$ có thể dùng Chi²

KĐ Chi bình phương

- Vọng trị:
- TH1: từ số liệu của các ô

`tabi a b \ c d, exp`

a	b
c	d

- TH2: từ biến số

`tab {biếnhàng} {biểncột}, exp`

tabi 49 44 \ 18 73, exp

- Vọng trị?

row	col		Total
	1	2	
1	49	44	93
	33.9	59.1	93.0
2	18	73	91
	33.1	57.9	91.0
Total	67	117	184
	67.0	117.0	184.0

$$E_a = (67 \times 93) / 184 = 33,86$$

$$E_b = (117 \times 93) / 184 = 59.14$$

$$E_c = (67 \times 91) / 184 = 33,14$$

$$E_d = (117 \times 91) / 184 = 57.86$$

KĐ Chi bình phương

- TH1: từ số liệu các ô

`tabi a b \ c d , chi`

a	b
c	d

- TH2: từ biến số

`tab {biên hàng} {biên cột} , chi`

`tabi 49 44 \ 18 73, chi`

- Nhằm so sánh tỉ lệ giữa các nhóm

row	col		Total
	1	2	
1	49	44	93
2	18	73	91
Total	67	117	184

Pearson chi2(1) = 21.5121 Pr = 0.000

- Có thể kết hợp các tùy chọn %

`tabi a b \ c d, co ro cell exp chi`

Kiểm định Chi bình phương

- Một nghiên cứu trên 397 người về tình trạng cao huyết áp.

Có các kết quả như sau:

- Trong số 197 nam, có 48 người cao huyết áp.
- Trong số 200 nữ, có 30 người cao huyết áp.

→ Giới tính có liên quan đến tình trạng cao huyết áp?

tabi 48 149 \ 30 170, exp chi

Kiểm định Fisher

- # Fisher exact
- Chỉ cần thiết nếu như giá trị vọng trị nhỏ vì khi đó phép kiểm Chi bình phương không áp dụng được

Kiểm định Fisher

- TH1: từ số liệu tóm tắt ở các ô

`tabi a b \ c d, exact`

- Ví dụ: `tabi 2 23 \ 5 30, exact`

a	b	a + b
c	d	c + d
a + c	b + d	

row	col		Total
	1	2	
1	2	23	25
2	5	30	35
Total	7	53	60
Fisher's exact =			0.688
1-sided Fisher's exact =			0.375

- TH2: từ số liệu thô

`tab {biểnhàng} {biểncột}, exact`

Kiểm định χ^2 / Fisher

tab {biểnhàng} {biểncột}, co ro exp chi exact

- **co**: phần trăm theo cột
- **ro**: phần trăm theo hàng
- **exp**: vọng trị
- **chi**: Chi bình phương
- **exact**: Fisher

tab gioitinh caoha, co

gioi tinh	cao huyet ap		Total
	khong cao	cao ha	
nam	149 46.71	48 61.54	197 49.62
nu	170 53.29	30 38.46	200 50.38
Total	319 100.00	78 100.00	397 100.00

Kiểm định χ^2 / Fisher

tab {biểnhàng} {biểncột}, co ro exp chi exact

- **co**: phần trăm theo cột
- **ro**: phần trăm theo hàng
- **exp**: vọng trị
- **chi**: Chi bình phương
- **exact**: Fisher

tab gioitinh caoha, ro

gioi tinh	cao huyet ap		Total
	khong cao	cao ha	
nam	149	48	197
	75.63	24.37	100.00
nu	170	30	200
	85.00	15.00	100.00
Total	319	78	397
	80.35	19.65	100.00

Kiểm định χ^2 / Fisher

tab {biếnhàng} {biểncột}, co ro exp chi exact

– **co**: phần trăm theo cột

tab gioitinh caoha, exp ro chi

– **ro**: phần trăm theo hàng

– **exp**: vọng trị

– **chi**: Chi bình phương

– **exact**: Fisher

gioi tinh	cao huyet ap		Total
	khong cao	cao ha	
nam	149	48	197
	158.3	38.7	197.0
	75.63	24.37	100.00
nu	170	30	200
	160.7	39.3	200.0
	85.00	15.00	100.00
Total	319	78	397
	319.0	78.0	397.0
	80.35	19.65	100.00
Pearson chi2(1) = 5.5139			Pr = 0.019

Tóm tắt

- Kiểm định so sánh 1 tỉ lệ

`prtesti n psample Ppopulation`

- Kiểm định so sánh 2 tỉ lệ

`prtesti n1 p1 n2 p2`

- Kiểm định Chi bình phương & Fisher

- Nếu chỉ có dữ liệu tóm tắt

`tabi a b \ c d, co ro exp chi exact`

- Nếu có dữ liệu thô

`tab {biên hàng} {biên cột}, co ro cell exp chi exact`

Nội dung đã học

- Kiểm định so sánh 1 tỉ lệ
- Kiểm định so sánh 2 tỉ lệ
- Kiểm định Chi bình phương
- Kiểm định chính xác Fisher

Thank you!