

CHƯƠNG 3

PHÂN TÍCH THỐNG KÊ

Phân tích thống kê trở nên quan trọng từ khi các kết quả nghiên cứu được dùng làm cơ sở cho các quyết định trị liệu. Bên cạnh việc xác định sự hiệu quả của trị liệu, phân tích thống kê còn khẳng định hay bác bỏ sự ý nghĩa của yếu tố gây nguy cơ hay yếu tố tiên lượng. Ngoài ra, khả năng phát hiện một bệnh không những tùy thuộc đặc tính của các xét nghiệm chẩn đoán mà còn bị ảnh hưởng bởi dung lượng mẫu.

Sai sót trong phân tích thống kê thường là do chọn phương pháp phân tích không phù hợp với loại số liệu hoặc loại bố trí nghiên cứu. Chương này thảo luận cách áp dụng và giải thích các trắc nghiệm thống kê dùng trong dịch tễ học lâm sàng và thảo luận các nguyên tắc hướng dẫn cách chọn phương pháp phân tích thống kê phù hợp.

1. Giải thích kết quả của phân tích thống kê

Trong phần lớn trường hợp, kết quả của các nghiên cứu lâm sàng thường được diễn đạt với các từ '*có sự khác biệt hay không khác biệt*'. Do bởi chúng ta lấy mẫu để tiên đoán một diễn biến thật trong quần thể, cho nên luôn luôn có khả năng đi đến kết luận sai lầm. Khi một trắc nghiệm thống kê được dùng, có 4 kết luận trong đó 2 kết luận đúng và 2 kết luận sai (Bảng 5.1).

Có hai trường hợp kết luận sai. Alpha hay lỗi loại I bị vấp phải khi chúng ta kết luận có sự khác biệt giữa hai kết quả trong khi chúng không khác biệt gì cả. Lỗi loại I tương tự như kết quả dương tính giả của xét nghiệm chẩn đoán. Beta hay lỗi loại II xảy ra khi chúng ta kết luận rằng các kết quả không khác biệt trong khi chúng thật sự khác biệt. Lỗi loại II tương tự kết quả âm tính giả trong chẩn đoán lâm sàng.

Xác suất để đạt kết quả mong đợi cũng là một phương cách khác để diễn đạt alpha và beta. Alpha tương trưng cho xác suất mà '*kết quả đạt được do bởi ngẫu nhiên*'. Mức ý nghĩa 0,05 hoặc ít hơn thường được chọn để giảm thiểu xác suất do ngẫu nhiên (do biến động của mẫu được lấy). Tuy nhiên có thể chọn mức alpha lớn hơn (0,1 hay lớn hơn) trong trường hợp thực hiện các xét nghiệm sàng lọc (screening) để từ đó có định hướng nghiên cứu tiếp. Trị số beta tương trưng cho xác suất '*không phát hiện được sự khác biệt ý nghĩa*' mặc dù sự khác biệt đang hiện diện. Thông thường beta được chọn ở mức gấp 4 lần alpha, do đó nếu $\alpha = 0,05$ thì $\beta = 0,2$. Như vậy $1 - \beta$ được xem là năng lực (power) của bố trí nghiên cứu. Năng lực là xác suất để phát hiện sự khác biệt giữa các kết quả khi sự khác biệt có thật, hoặc là xác suất để khẳng định sự hiện diện của một bệnh. Chúng ta

khó thể giảm alpha và beta cùng lúc bởi vì khi alpha giảm thì beta tăng, khi ấy năng lực phát hiện sự khác biệt sẽ giảm.

Bảng 5.1 Kết luận từ phân tích thống kê và sự khác biệt có thật giữa các kết quả nghiên cứu

		Giả thiết tương đồng H_0 (null hypothesis)*	
		Đúng (true)	Sai (false)
Kết luận từ phân tích thống kê	Chấp nhận	Kết luận đúng	Lỗi loại II (type II error, β)
	Bác bỏ	Lỗi loại I (type I error, α)	Kết luận đúng

* *Thí dụ về giả thiết tương đồng H_0 : $\mu_1 = \mu_2$*

Các trắc nghiệm thống kê trong thú y thường được dùng để bác bỏ giả thiết H_0 , đó là giả thiết rằng không có sự khác biệt giữa hai nhóm theo dõi. Nếu chứng minh có sự khác biệt (bác bỏ H_0), giá trị của Pa (xác suất tương ứng với alpha, dùng Pa để phân biệt với Pb của lỗi loại II) thường được báo cáo. Giá trị Pa thường được xem là có ý nghĩa về thống kê nếu $< 0,05$ (nghĩa là chúng ta mong rằng kết luận chỉ sai tối đa 5% trường hợp nếu cùng loại nghiên cứu được lập lại). Nếu Pa lớn ($\geq 0,5$) thì có nghĩa là '*không phát hiện được sự khác biệt*' mà không là '*không có sự khác biệt*'. Khi không phát hiện được sự khác biệt, cần xem xét lại năng lực của bố trí nghiên cứu vì có thể do bố trí không đúng hoặc dung lượng mẫu nhỏ. Trong các chương trình thanh toán bệnh dựa vào xét nghiệm chẩn đoán để phát hiện thú/đàn thú nhiễm bệnh, beta là yếu tố quan trọng để xác định dung lượng mẫu. Dung lượng mẫu sẽ được thảo luận ở các mục sau.

2. Khoảng tin cậy của tỷ lệ (proportion, rate)

2.1. Khoảng tin cậy của một tỷ lệ

Có thể ước tính khoảng tin cậy của một tỷ lệ, chẳng hạn tỷ lệ bệnh, bằng cách dùng phân bố nhị thức.

$$\text{Phương sai của tỷ lệ bệnh} = \frac{p(1-p)}{n} \quad \text{với } n = \text{dung lượng mẫu, } p = \text{tỷ lệ bệnh}$$

Công thức này được dùng khi giả định rằng mẫu được lấy từ quần thể lớn. Với quần thể nhỏ và dung lượng mẫu lớn (dung lượng mẫu bằng 10% của quần thể hoặc $f = 0,1$) thì tử số phải nhân với $1-f$.

$$\text{Suy ra sai số chuẩn (SE) của tỷ lệ là } \sqrt{\frac{p(1-p)}{n}}$$

Thí dụ, tỷ lệ mắc bệnh trong đàn bò ($n = 171$ con) là 1,2%, do đó:
 Phương sai = $(0,012 \times 0,988)/171 = 0,0000693$ ($\%^2$)

Sai số chuẩn của tỷ lệ = $0,00832 = 0,832\%$

Với 95% độ tin tưởng, khoảng tin cậy là: $1,2 \pm (1,96 \times 0,832) = -0,4\% ; 2,8\%$
Qua thí dụ này, có tỷ lệ bệnh $< 0\%$ trong khoảng tin cậy, đó là do sự phân bố không cân đối của tỷ lệ (tỷ lệ bệnh khá thấp).

Khoảng tin cậy có thể cho biết dung lượng mẫu đủ lớn hay không để có kết quả dương tính (kết quả mong muốn). Nếu mức dưới của khoảng tin cậy lớn hơn trị số ngưỡng nào đó (trị số để được xem là có ý nghĩa về lâm sàng), kết quả chắc chắn dương tính. Nếu mức dưới nhỏ hơn trị số ngưỡng và mức trên lớn hơn trị số ngưỡng, kết quả cũng được xem là dương tính nhưng không chắc chắn, và cần phải tăng dung lượng mẫu. Nếu mức trên của khoảng tin cậy nhỏ hơn trị số ngưỡng thì sao ???

2.2. Khoảng tin cậy của sai biệt giữa 2 tỷ lệ

Khoảng tin cậy của sai biệt giữa 2 tỷ lệ có thể cho thấy sự sai biệt giữa 2 tỷ lệ có ý nghĩa thống kê hay không. Khi **2 nhóm không liên quan**, các bước tính tương tự như cách tính khoảng tin cậy của sai biệt giữa 2 trung bình hoặc 2 trung vị. Sai số chuẩn của sai biệt (SEsb):

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

p_1 = tỷ lệ bệnh ước tính của nhóm 1

p_2 = tỷ lệ bệnh ước tính của nhóm 2

n_1 = dung lượng mẫu của nhóm 1

n_2 = dung lượng mẫu của nhóm 2

Khi ấy, ta có khoảng tin cậy với 95% tin tưởng:

$$(p_1 - p_2) \pm 1,96 \times SEsb$$

Thí dụ: $n_1 = 251, n_2 = 260$

Tỷ lệ huyết thanh dương tính của heo ở vùng 1: $p_1 = 69/251 = 0,275$

Tỷ lệ huyết thanh dương tính của heo ở vùng 2: $p_2 = 61/260 = 0,235$

$$|p_1 - p_2| = 0,04 \text{ và } SEsb = 0,038$$

Khoảng tin cậy với 95% tin tưởng:

$$0,04 \pm (1,96 \times 0,038) = -0,034 ; 0,115 \text{ hay } -3,4\% ; 11,5\%$$

Kết quả cho thấy sự sai biệt giữa 2 tỷ lệ không ý nghĩa ở $\alpha = 0,05$.

3. Tương quan giữa hai biến số

Tương quan tuyến tính giữa 2 biến số có thể được tính bằng hệ số tương quan hay dưới dạng phương trình hồi quy. Khi phương trình hồi quy được lập, hệ số xác định (coefficient of determination, r^2) cho biết mức độ phù hợp của đường hồi quy với số liệu phân tích. Khi hệ số tương quan có ý nghĩa về thống kê, hệ số xác định có thể rất nhỏ, khi

ấy phương trình hồi quy không thể được dùng để tiên đoán biến số phụ thuộc (dependent variable).

4. Chọn lựa trắc nghiệm thống kê thích hợp

Phần lớn các trắc nghiệm thống kê được dùng để ước tính xác suất của lỗi loại I (α), đó là xác suất để kết luận có sự khác biệt dù rằng sự khác biệt không thật sự hiện diện. Giá trị của mỗi trắc nghiệm tùy thuộc vào giả định (assumption) cho số liệu. Nếu số liệu không thỏa mãn giả định được đặt ra, trị số P tìm được có thể không đúng. Sơ đồ 5.1 trình bày sự chọn lựa trắc nghiệm thống kê thích hợp tùy thuộc vào đặc điểm của bố trí nghiên cứu và số liệu phân tích.

Các trắc nghiệm thống kê có thể được chia thành 2 nhóm: trắc nghiệm tham số (parametric) và trắc nghiệm phi tham số (non-parametric) (Bảng 5.2). Trắc nghiệm tham số hiệu lực hơn trắc nghiệm phi tham số do bởi trắc nghiệm tham số có nhiều khả năng hơn trong việc bác bỏ giả thiết H_0 nếu thật sự có sự khác biệt. Trắc nghiệm tham số được dùng khi số liệu thỏa mãn các yêu cầu sau:

- (1) Các nhóm được chọn ngẫu nhiên trong quần thể.
- (2) Số liệu ở dạng khoảng cách/tỷ lệ.
- (3) Số liệu có phân bố chuẩn.
- (4) Phương sai của các nhóm bằng nhau.

Bảng 5.2 Trắc nghiệm tham số và trắc nghiệm phi tham số

Trắc nghiệm phi tham số

Nhi thức (trắc nghiệm cho tỷ lệ)
Chi bình phương (I) (trắc nghiệm tính phù hợp giữa tần số quan sát và tần số kỳ vọng)
Chi bình phương (II) (phân tích bảng ngẫu nhiên)
McNemar
Cohran Q
Kolmogorov-Smirnov
Mann-Whitney U
Sign
Wilcoxon
Kruskal-Wallis
Friedman
Spearman rho (p)*

Trắc nghiệm tham số

t (I) (so sánh mẫu với trung bình của quần thể)
t (II) (trắc nghiệm t không bắt cặp)
t (III) (trắc nghiệm t bắt cặp)
ANOVA một yếu tố
ANOVA nhiều yếu tố
Pearson r *

* Spearman rho và Pearson r trắc nghiệm mối tương quan giữa hai biến số

Nguồn: Sharp, V.F. 1979. *Statistics for the social sciences*. Little, Brown & Co., Boston, M.A.

Số liệu và bố trí nghiên cứu	Loại số liệu	Số nhóm	Đặc tính của nhóm	Số hạng mục	Kích cỡ hạng mục	Số liệu	Trắc nghiệm
	Hạng Mục	Một		Một			Chi square (I)
				Hai	≤ 4		Nhị thức
					≥ 5		Chi square (I)
							Chi square (II)
		Ba, >3		Ba, >3			Chi square (II)
							Chi square (III)
							McNemar
							Chi square (II)
							CochranQ
Số liệu và bố trí nghiên cứu	Thứ tự	Một					Kolmogorov-smirnov
							Mann-Whitney U
		Hai				Dấu	sign
						Số	Wilcoxon
		Ba, >3n					Kruskal-Wallis
							Friedman
Số liệu và bố trí nghiên cứu	Khoảng cách	Một					t (I)
							t (II)
							t (III)
		Hai					ANOVA
		Ba, >3					Khối ngẫu nhiên

(Nguồn : Sharp, V. F. 1979. Statistics for the social science. Little, Brown & Co., Boston, M.A.)

Sơ đồ 5.1 Sơ đồ để chọn lựa trắc nghiệm thống kê thích hợp tùy thuộc vào bố trí nghiên cứu và loại số liệu

Trắc nghiệm phi tham số đòi hỏi ít giả định hơn. Chúng chỉ thỏa mãn yêu cầu thứ nhất như của trắc nghiệm tham số nhưng không cần thỏa mãn các yêu cầu còn lại. Trắc nghiệm phi tham số được dùng khi dung lượng mẫu nhỏ, số liệu phân bố kiểu tự do

và thường ở dạng hạng mục hoặc thứ tự. Trắc nghiệm phi tham số Wilcoxon và Mann-Whitney có thể dùng cho số liệu dạng khoảng cách khi chúng không phân bố chuẩn. Vài thí dụ về chọn loại trắc nghiệm:

- Trường hợp 1 về trị bệnh đau bụng ở ngựa. Đánh giá hiệu lực của thuốc giảm đau bằng cách đo lường mức độ đau vào lúc trước và sau khi dùng thuốc. Mức độ đau được đánh giá bằng thời gian có triệu chứng đau. Như vậy nghiên cứu này theo dõi có 2 nhóm liên quan nhau: trước khi dùng thuốc và sau khi dùng thuốc. Do đó trắc nghiệm t (III) hay trắc nghiệm t bất cặp được dùng để so sánh mức độ đau trước và sau khi dùng thuốc.

- Trường hợp 2 trong phòng bệnh ký sinh trùng đường ruột. Xét ảnh hưởng của phòng bệnh ký sinh trùng lên tăng trọng của 4 nhóm bò dựa vào trọng lượng và cách chữa trị (đối chứng và chữa trị; mức trọng lượng nặng và nhẹ). Trắc nghiệm ANOVA nhiều yếu tố (khối ngẫu nhiên, phân nhánh) thích hợp cho trường hợp này.

- Trường hợp 3 về vô trùng sau khi giải phẫu. Khảo sát được thực hiện để so sánh số vết mổ bị nhiễm trùng khi dùng kháng sinh hay không dùng kháng sinh. Kết quả (nhiễm trùng hoặc không) là số liệu dạng hạng mục ở 2 nhóm thú bệnh độc lập nhau (nhóm dùng kháng sinh và nhóm không dùng). Trắc nghiệm nên là χ^2 để xử lý số liệu.