

# Dạy học thống kê trong đánh giá độ chính xác của một Test chẩn đoán với sự hỗ trợ của phần mềm STATA

Đào Hồng Nam

TS. Trường ĐH Y- Dược TP Hồ Chí Minh

Received: 17/6/2024; Accepted: 20/6/2024; Published: 24/6/2024

**Abstract:** Statistics has many applications in many areas of life in general and medicine in particular. In recent years, much applied statistical knowledge in medicine has been included in the curricula and textbooks of health sciences universities. However, teaching so that students can understand and apply this knowledge in studying, scientific research and in the professional practice of future doctors still has certain limitations. One of those limitations is that the textbooks only present pure mathematical statistical knowledge and illustrate this knowledge through manual formula calculations, without instructions or integration of calculation tools, mathematics and statistical software to quickly solve and draw meaningful conclusions for clinical situations. In this article we will show some types of tasks involved in evaluating the accuracy of a diagnostic test with the help of Stata software.

**Keywords:** Teaching, accuracy, diagnostic tests, Cohen's Kappa

## 1. Đặt vấn đề

Khi phân tích các thể chế dạy học (chương trình, giáo trình) thống kê y học của các trường đại học khoa học sức khỏe, chúng tôi nhận thấy: Các tổ chức toán học liên quan đến kiểu nhiệm vụ T: “Đánh giá độ chính xác của một test chẩn đoán” đã được trình bày một cách tường minh. Tuy nhiên trong các tổ chức toán học liên quan đến kiểu nhiệm vụ T vẫn còn thiếu hụt một ứng viên quan trọng đó là thống kê Kappa bao gồm Cohen's Kappa (cho 2 người hoặc 2 phương pháp (PP) đánh giá) và Fleiss Kappa (cho từ 3 người hoặc từ 3 PP đánh giá trở lên) nhằm đánh giá sự phù hợp của các test chẩn đoán. Trong khuôn khổ bài báo này, tác giả đề xuất một số biện pháp dạy học ước lượng khoảng tin cậy để đánh giá độ chính xác của các test chẩn đoán và thống kê Cohen's Kappa với sự hỗ trợ của phần mềm Stata.

## 2. Nội dung nghiên cứu

### 2.1. Độ chính xác của một xét nghiệm

Một xét nghiệm (test) thể hiện độ chính xác bởi hai thông số: độ nhạy và độ đặc hiệu. Độ nhạy (sn) là xác suất test dương tính đối với người được chẩn đoán mắc bệnh B, trong khi độ đặc hiệu (sp) là xác suất test âm tính đối với người được chẩn đoán không mắc bệnh B.

Trong một số tình huống nghiên cứu để xem xét thay thế một test phức tạp (xâm lấn, rủi ro, tốn kém) bằng một xét nghiệm đơn giản hơn (T) thì nhà nghiên cứu hoặc BS lâm sàng cần phải đánh giá độ chính xác của T, xét nghiệm này sẽ được so sánh với

xét nghiệm chuẩn (gold standard) như trong bảng sau đây:

Một ví dụ: Kết quả Realtime RT-PCR trên hệ thống xét nghiệm Cobas 6800 so với quy trình của WHO trong chẩn đoán xác định SARS-CoV-2. Tính độ chính xác của xét nghiệm theo quy trình của WHO.

WHO	Cobas 6800		Tổng
	Dương tính	Âm tính	
Dương tính	231	1	232
Âm tính	50	27	77
Tổng	281	28	309

Nguồn: Nguyễn Văn Đoàn và các tác giả khác (2024)

$$sn = \frac{231}{281} = 0.822; sp = \frac{27}{28} = 0.964$$

Để đánh giá KTC của các tỷ lệ về độ chính xác của một xét nghiệm, chúng ta có thể sử dụng KTC chuẩn (còn gọi là KTC Wald) như sau:

$$p = \hat{p} \pm 1.96 \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \quad (1)$$

Trong đó  $\hat{p} = \frac{k}{n}$  là tỷ lệ đặc tính X cần ước lượng.

Theo Irene, Mojica (2011), điều kiện để KTC Wald có hiệu lực là  $n \min\{\hat{p}, 1-\hat{p}\} \geq 5(*)$

Nếu điều kiện (\*) không thỏa mãn thì có thể sử dụng các KTC thay thế là Wilson, Jeffreys hoặc Clopper-Pearson. Đặc biệt nếu thử nghiệm trên một cỡ mẫu nhỏ ( $n < 40$ ) và/hoặc xác suất của biến cố là những giá trị cực đoan (rất lớn hoặc rất nhỏ) thì có thể sử dụng KTC Wilson:

$$p = \frac{n_s + \frac{1}{2}z^2 \pm z\sqrt{\frac{n_s \cdot n_F + z^2}{n} + \frac{z^2}{4}}}{n + z^2} \quad (2)$$

Đối với tỷ lệ dương tính thật (độ nhạy):

với  $n = 28; \hat{p} = sn = \frac{231}{281} = 0.822; 1 - \hat{p} = \frac{50}{281} = 0.178$

Điều kiện (\*) thỏa mãn vì  $nmin\{\hat{p}, 1 - \hat{p}\} = 50$ , khi đó Wald cho tỷ lệ dương tính thật ở độ tin cậy 95% theo công thức (1) là:  $0.822 \pm 0.045 = [0.777; 0.867]$

Sử dụng lệnh ước lượng KTC Wald trong Stata: `cii prop 281 231, wald` ta có kết quả như sau:

Variable	Obs	Proportion	Std. err.	Binomial Wald [95% conf. interval]	
	281	.8220641	.0228156	.7773463	.8667818

Đối với tỷ lệ âm tính thật (độ đặc hiệu): với

$n = 28; \hat{p} = sp = \frac{27}{28} = 0.964$

Điều kiện  $nmin\{\hat{p}, 1 - \hat{p}\} \geq 5(*)$  không thỏa mãn vì  $nmin\{\hat{p}, 1 - \hat{p}\} = 1.008$ , khi đó không thể sử dụng KTC 95% của Wald cho tỷ lệ âm tính thật. Trong trường hợp này, có thể sử dụng KTC Wilson theo công thức (2):

$$\frac{27 + 0.5 \times 1.96^2 \pm 1.96 \sqrt{\frac{27 \times 1}{28} + \frac{1.96^2}{4}}}{28 + 1.96^2} [0.823; 0.994]$$

$n_s$  là số ca dương tính,  $n_F$  là số ca âm tính trong tổng số  $n$  là số ca xét nghiệm

với  $n_s = 27$  (số ca âm tính theo WHO) và  $n_F = 1$  (số ca dương tính theo WHO) trong tổng số  $n = 28$  ca có kết quả xét nghiệm Cobas 6800 âm tính.

Sử dụng lệnh ước lượng KTC Wilson trong Stata: `cii prop 28 27, wilson` ta có kết quả như sau:

Variable	Obs	Proportion	Std. err.	Wilson [95% conf. interval]	
	28	.9642857	.0350707	.822878	.9936675

## 2.2. Thống kê Cohen's Kappa

Thống kê Kappa lần đầu tiên được đề xuất bởi Cohen (1960). Cohen (1968) đã khái quát hóa các trọng số phản ánh mức độ nghiêm trọng tương đối của từng bất đồng có thể xảy ra trong đánh giá mức độ phù hợp của các PP chẩn đoán hoặc sự đồng thuận của các bác sĩ trong đánh giá, chẩn đoán.

Tất cả các công thức dưới đây được trình bày trong Fleiss, Levin và Paik (2003).

Gọi  $m$  là số người đánh giá (hoặc số PP đánh giá) và gọi  $k$  là số kết quả đánh giá. Phần dưới đây trình bày công thức với  $m = 2, k = 2$ .

WHO	Cobas 6800		Tổng	WHO	Cobas 6800		Tổng
	Dương tính	Âm tính			Dương tính	Âm tính	
Dương tính	231	1	232	Dương tính	210.977	21.023	232
Âm tính	50	27	77	Âm tính	70.023	6.977	77
Tổng	281	28	309	Tổng	281	28	309

Gọi  $w_{ij}$  ( $i = 1, \dots, k; j = 1, \dots, k$ ) là trọng số đồng thuận và không đồng thuận tương ứng (hàm trong Stata là `wgt()`) với  $w_{ij} = 1 - |i - j| / (k - 1)$ .

Nếu `wgt(2)` được xác định thì  $w_{ij} = 1 - \{|i - j| / (k - 1)\}^2$

Tỷ lệ đồng thuận quan sát là:  $p_0 = \sum_{i=1}^k \sum_{j=1}^k w_{ij} p_{ij}$

trong đó  $p_{ij}$  là tỷ lệ xếp hạng  $i$  của người xếp hạng đầu tiên và  $j$  của người xếp hạng thứ hai. Tỷ lệ thỏa

thuận dự kiến là:  $p_e = \sum_{i=1}^k \sum_{j=1}^k w_{ij} p_{i.} p_{.j}$

Trong đó  $p_{i.} = \sum_j p_{ij}$  và  $p_{.j} = \sum_i p_{ij}$ .

Thống kê Kappa là:  $\hat{\kappa} = \frac{p_0 - p_e}{1 - p_e} \quad (3)$

Sai số chuẩn của  $\hat{\kappa}$  khi so sánh với 0 là:

$$\hat{\sigma}_0 = \frac{1}{(1 - p_e)\sqrt{n}} \sqrt{\left[ \sum_i \sum_j p_{i.} p_{.j} \{w_{ij} - (\bar{w}_{i.} + \bar{w}_{.j})\}^2 \right] - p_e^2}$$

Với  $n$  là số đối tượng được đánh giá,

$\bar{w}_{i.} = \sum_j p_{.j} w_{ij}$  và  $\bar{w}_{.j} = \sum_i p_{i.} w_{ij}$

Khi đó biến ngẫu nhiên  $Z = \frac{\hat{\kappa}}{\hat{\sigma}_0} \sim N(0, 1) \quad (4)$

Trong đó  $p_0, p_e$  là tỷ lệ phù hợp/ đồng thuận quan sát và mong đợi tương ứng

Nếu  $\kappa = 0$  thì sự đồng thuận được cho là sẽ xảy ra một cách ngẫu nhiên, nếu  $\kappa = 1$  thì sự đồng thuận là hoàn hảo. Với các giá trị khác Landis và Koch (1977a, 165) đề xuất những cách giải thích sau về mức độ phù hợp/đồng thuận:

Hệ số Kappa $\hat{\kappa}$	Mức độ
< 0	Yếu
0.00-0.20	Kém
0.21 - 0.40	Trung bình
0.41 - 0.60	Khá
0.61 - 0.80	Đáng kể
0.81 - 1	Gần như hoàn hảo

Trong ví dụ. Đánh giá sự đồng nhất về kết quả realtime RT-PCR trên hệ thống xét nghiệm Cobas 6800 so với quy trình của WHO trong chẩn đoán xác định SARS-CoV-2 trong chẩn đoán xác định SARS-CoV-2

$$p_0 = \frac{231+27}{309} = 0.835; p_e = \frac{210.977+6.977}{309} = 0.705;$$

$$\kappa = \frac{p_0 - p_e}{1 - p_e} = \frac{0.835 - 0.705}{1 - 0.705} = 0.44$$

Sử dụng Stata

Lệnh *diagni 231 50 1 27* để nhập dữ liệu tính độ chính xác và các giá trị liên quan

Kết quả:

True disease status	Test result		Total
	Neg.	Pos.	
Normal	27	1	28
Abnormal	50	231	281
Total	77	232	309

[95% Confidence Interval]

Prevalence	Pr(A)	90.9%	87.2%	93.9%
Sensitivity	Pr(+ A)	82.2%	77.2%	86.5%
Specificity	Pr(- N)	96.4%	81.7%	99.9%
ROC area	(Sens. + Spec.)/2	0.89	0.85	0.93
Likelihood ratio (+)	Pr(+ A)/Pr(+ N)	23.02	3.36	157.86
Likelihood ratio (-)	Pr(- A)/Pr(- N)	0.18	0.14	0.24
Odds ratio	LR(+)/LR(-)	124.74	20.88	.
Positive predictive value	Pr(A +)	99.6%	97.6%	100.0%
Negative predictive value	Pr(N -)	35.1%	24.5%	46.8%

Để tính thống kê Kappa, nhập DL như sau: *tabi 231 1\50 27, replace* kết quả như bảng:

row	col		Total
	1	2	
1	231	1	232
2	50	27	77
Total	281	28	309

WHO	Cobas 6800		Total
	Positive	Negative	
Positive	231	1	232
Negative	50	27	77
Total	281	28	309

Thực hiện đổi tên biến và gán trọng số sau đây, ta được kết quả như bảng:

```
rename row WHO
rename col Cobas6800
label var Cobas6800 "Cobas 6800"
label define assess 1 Positive 2 Negative
label values WHO assess
label values Cobas6800 assess
kap WHO Cobas6800 [fw=pop], tab
```

Agreement	Expected agreement	Kappa	Std. err.	Z	Prob>Z
83.50%	70.54%	0.4398	0.0479	9.17	0.0000

	Cobas 6800		Total
	Positive	Negative	
Positive	231	1	232
Negative	50	27	77
Total	281	28	309

  

Agreement	Expected agreement	Kappa	Std. err.	Z	Prob>Z
83.50%	70.54%	0.4398	0.0479	9.17	0.0000

### 3. Kết luận

Có nhiều PP đánh giá độ chính xác của một test chẩn đoán, trong bài báo này, tác giả đã trình bày PP ước lượng KTC cho tỷ lệ bao gồm tỷ lệ dương tính thật (độ nhạy) và tỷ lệ âm tính thật (độ đặc hiệu). Tác giả cũng đã trình bày một cách ngắn gọn lý thuyết thống kê Kappa để đánh giá hoặc so sánh sự phù hợp của các Test chẩn đoán hoặc sự đồng thuận của hội đồng chẩn đoán.

Các công thức tính toán là khá dài và mất nhiều thời gian tính toán. Tuy nhiên sau khi trình bày lý thuyết và một số ví dụ được tính toán thủ công để SV có thể hiểu sâu phần lý thuyết, GV có thể hướng dẫn SV sử dụng phần mềm Stata để tính toán các thông số này một cách nhanh chóng và hiệu quả.

*Ghi chú: Đề tài này được nhận kinh phí tài trợ từ Trường Đại học Y Dược TP Hồ Chí Minh.*

### Tài liệu tham khảo

1. Nguyễn, V. Đoàn, Nguyễn, T. H. P., Dương, H. Q., Văn, Đình T., Phạm, Ngọc, & Lê, N. M. H (2024). Đánh giá sự đồng nhất về kết quả realtime RT-PCR trên hệ thống xét nghiệm Cobas 6800 so với quy trình của WHO trong chẩn đoán xác định SARS-CoV-2. Tạp chí Truyền nhiễm Việt Nam, 1(45), 44-60.

2. Landis, J. R., and G. G. Koch (1977). *The measurement of observer agreement for categorical data. Biometrics* 33: 159-174. <http://doi.org/10.2307/2529310>

3. Fleiss, J. L., B. Levin, and M. C. Paik (2003). *Statistical Methods for Rates and Proportions. 3rd ed. New York: Wiley. Mojica, Irene, "Exploring Confidence Intervals in the Case of Binomial and Hypergeometric Distributions". Electronic Theses and Dissertations. 6644.*