

진단검사 정확도 평가지표의 신뢰구간

오태호 · 박선일*¹

경북대학교 수의과대학, *강원대학교 수의과대학 및 동물의학종합연구소

(Accepted: August 24, 2015)

The Use of Confidence Interval of Measures of Diagnostic Accuracy

Tae-Ho Oh and Son-II Pak*¹

College of Veterinary Medicine, Kyungpook National University, Daegu 702-701, Korea

*College of Veterinary Medicine and Institute of Veterinary Science, Kangwon National University, Chuncheon 200-701, Korea

Abstract : The performance of diagnostic test accuracy is usually summarized by a variety of statistics such as sensitivity, specificity, predictive value, likelihood ratio, and kappa. These indices are most commonly presented when evaluations of competing diagnostic tests are reported, and it is of utmost importance to compare the accuracies of diagnostic tests to decide on the best available test for certain medical disorder. However, it is important to emphasize that specific point values of these indices are merely estimates. If parameter estimates are reported without a measure of uncertainty (precision), knowledgeable readers cannot know the range within which the true values of the indices are likely to lie. Therefore, when evaluations of diagnostic accuracy are reported the precision of estimates should be stated in parallel. To reflect the precision of any estimate of a diagnostic performance characteristic or of the difference between performance characteristics, the computation of confidential interval (CI), an indicator of precision, is widely used in medical literatures in that CIs are more informative to interpret test results than the simple point estimates. The majority of peer-reviewed journals usually require CIs to be specified for descriptive estimates, whereas domestic veterinary journals seem less vigilant on this issues. This paper describes how to calculate the indices and associated CIs using practical examples when assessing diagnostic test performance.

Key words : confidence interval, diagnostic accuracy, sampling variation, agreement.

서 론

대부분의 의학연구에서는 표본을 대상으로 수행되지만 연구의 궁극적인 목적은 표본을 대상으로 얻은 연구결과에 근거하여 이들 표본이 유래한 더 큰 집단 즉 모집단으로 확장하여 통계적 추론(statistical inference)을 유도하는 것이다. 모집단을 대표하지 못하는 표본을 대상으로 수행된 연구결과는 결과를 왜곡시킬 수 있으며, 특히 단일 표본을 대상으로 수행된 결과는 표집오차(sampling variation)에 기인하여 불확실성이 높아지는 단점이 있다. 따라서 연구결과에 대하여 단순히 점추정치(point estimate)와 유의확률(significance probability, p value)을 제시하는 것 보다는 표집오차에 의한 표본 추정치가 가지고 있는 불확실성을 반영하기 위하여 추정치의 신뢰구간(confidence interval)을 제시하는 것이 바람직하다. 그 이유는 유의확률은 귀무가설을 반박하는 증거 혹은 실험군 간 연관성의 강도를 측정하는 수단으로 처리효과의 크기에 대한 정보를 제공하지 못하는 반면에 신뢰구간은 가설검정 결과뿐만 아니라 모수가 위치할 구간의 정밀도

(precision)에 대한 정량적인 정보를 동시에 제공하기 때문이다(12). 대부분의 의학저널에서도 가설검정 결과를 제시하는 수단으로 유의확률 대신에 신뢰구간을 사용할 것을 권고하고 있다(8).

진단검사가 적용되는 최소 단위를 연구의 진단단위(diagnostic unit of study, DUOS)라고 한다(3). 예를 들어 내원한 100두의 개에서 수집된 검체에 대하여 특정 질병 'x'에 대한 항원 검사를 적용하였다면 검체 1건이 최소 진단단위가 된다. 이 경우 진단단위 간 상호 독립성 가정이 성립되며 이러한 자료를 독립형 자료(independent data)라고 한다. 한편 진단단위는 상호 독립적인 경우도 있지만 독립성 가정이 유효하지 못한 경우도 있다. 이를테면 50두의 개를 대상으로 악성 간 종양을 확인하기 위하여 1두 당 1개 이상의 생검(biopsy) 조직에 대하여 병리학적 검사를 수행한다고 가정하면 진단단위 간 상호 독립성 가정이 성립되지 못하며 이러한 자료를 집락형 자료(clustered data)라고 한다. 이 경우 환자인 개를 집락(cluster)이라고 하며, 각각의 개에서 얻은 생검 조직은 집락 내 진단단위가 된다. 민감도, 특이도, 예측도 등과 같이 진단검사의 특성을 추정할 때 신뢰구간을 계산하는 방법은 독립성 가정의 충족 여부에 따라 다르다. 본 연구에서는 진단검사의 특성을 요약하는 다양한 지표에 대

¹Corresponding author.
E-mail : paksi@kangwon.ac.kr

하여 독립성 가정에서 신뢰구간을 계산하는 방법에 대하여 설명한다.

결론

민감도와 특이도

질병상태의 참값을 알고 있는 검체에 대하여 진단검사(index test)를 적용한 결과를 빈도로 요약하면 Table 1과 같다.

Exact binomial method

비율 자료에 대하여 F분포에 근거한 정확 이항(exact binomial) 신뢰구간은 다음과 같이 계산한다(4,6).

$$\text{신뢰하한: } \pi_L = \frac{r}{r + (n-r+1)F_{\alpha/2, 2n-2r+2, 2r}}$$

$$\text{신뢰상한: } \pi_U = \frac{r+1}{r+1 + \frac{n-r}{F_{\alpha/2, 2r+2, 2n-2r}}}$$

여기에서 r = 관심을 두고 있는 사건이 발생한 건수(Table 1의 a 혹은 d), n = 검사 수(Table 1의 n_1 혹은 n_2), $F_{p,a,b}$ = 자유도가 a, b 일 때 F 분포의 값이다.

Wald method

진단단위 간 상호 독립성 가정이 성립될 때 민감도와 특이도를 추정하는 가장 일반적인 방법은 이항분포의 정규분포 근사성(normal approximation to the binomial distribution)을 이용한다(13). 이 경우 민감도(Se)와 특이도(Sp)의 분산과 신뢰구간은 다음과 같다.

$$Se = \frac{a}{n_1}, \text{Var}(Se) = \frac{Se(1-Se)}{n_1} = \frac{ac}{n_1^2}$$

$$Sp = \frac{d}{n_2}, \text{Var}(Sp) = \frac{Sp(1-Sp)}{n_2} = \frac{bd}{n_2^2}$$

$$Se \pm z_{1-\alpha/2} \sqrt{\frac{Se(1-Se)}{n_1}} \text{ (여기에서 } n_1 = a + c)$$

$$Sp \pm z_{1-\alpha/2} \sqrt{\frac{Sp(1-Sp)}{n_2}} \text{ (여기에서 } n_2 = b + d)$$

이 방법은 정규분포 근사성이 유효하고, 표본크기가 크며, 민감도(특이도)가 0이나 1에 접근하지 않을 때 사용하는 것이 바람직하며, 구체적으로 a 와 $N-a$ 가 5 보다 클 때($np > 5$ 와 $nq > 5$) 적절하다(2,6). 만일 민감도(특이도)가 0이나 1에

Table 1. 2×2 frequency contingency table showing responses of a binary diagnostic test, according to the status of each individual samples

Test result	Disease status		Total
	Positive (D+)	Negative (D-)	
Positive (T+)	a	b	m_1
Negative (T-)	c	d	m_2
Total	n_1	n_2	N

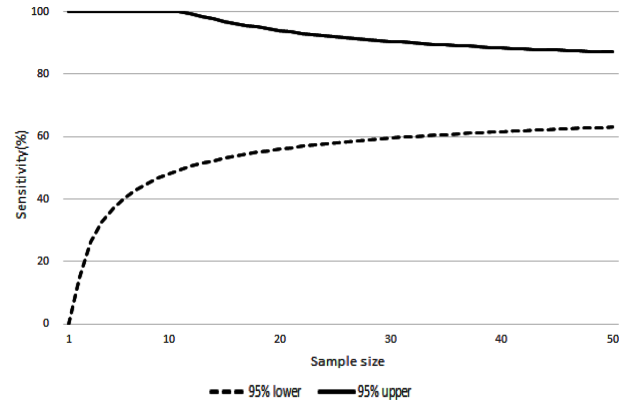


Fig 1. The 95% confidence interval for an observed estimate of 75% by sample size.

접근하는 경우 정규분포 근사성을 이용한 신뢰구간의 상한값이 1을 초과할 수 있어 부정확한 결과를 보이므로 이항분포에 근거하여 정확 신뢰구간을 계산하는 것이 바람직하다. 표본크기가 증가하면 민감도 추정치의 정밀도 역시 증가한다. Fig 1은 표본크기에 따라 민감도가 75%인 검사의 95% 신뢰구간을 나타낸 것으로 표본크기가 10인 경우 95% 신뢰구간은 [48.2%, 100%]이지만, 표본크기가 증가할수록 신뢰구간의 폭이 좁아져 추정치의 정밀도가 향상된다는 것을 알 수 있다.

Agresti-Coull method

전술한 Wald method는 특히 표본크기가 작거나 민감도 혹은 특이도가 0이나 1에 접근할 경우 신뢰구간이 참값을 포함할 가능성이 작아지는 경향이 있으며, 특히 민감도 혹은 특이도가 1에 접근하면 신뢰구간의 상한값이 1을 초과하는 단점이 있다. 이러한 문제를 해결하기 위한 다양한 방법이 제시되어 있는데 첫째는 연속성을 보정하는 방법으로 분할표(Table 2)의 각 셀에 일정한 상수 $k^2/2$ (여기에서 $k = z_{1-\alpha/2}$)를 더해준다(9). Agresti-Coull이 제시한 방법에 의한 민감도(Se_{AC})와 특이도(Sp_{AC})는 다음과 같이 계산한다(1).

$$Se_{AC} = \frac{a+k^2/2}{n'_1}, \text{Var}(Se_{AC}) = \frac{Se_{AC}(1-Se_{AC})}{n'_1+k^2}, a = n_1Se, k = z_{1-\alpha/2}, n'_1 = n_1+k^2$$

$$Sp_{AC} = \frac{d+k^2/2}{n'_2}, \text{Var}(Sp_{AC}) = \frac{Sp_{AC}(1-Sp_{AC})}{n'_2+k^2}, d = n_2Sp, k = z_{1-\alpha/2}, n'_2 = n_2+k^2$$

Zhou-Li method

두 번째 방법은 특히 민감도(특이도)가 0이나 1에 접근할 때 Zhou-Li(ZL) 신뢰구간을 사용할 수 있다(15).

Table 2. Adjusted contingency table using a continuity correction (Agresti-Coull)

Test result	Disease status		Total
	Positive (D+)	Negative (D-)	
Positive (T+)	$a+k^2/2$	$b+k^2/2$	m_1
Negative (T-)	$c+k^2/2$	$d+k^2/2$	m_2
Total	$n'_1 = n_1 + k^2$	$n'_2 = n_2 + k^2$	N

$$LL = \ln \left[\frac{Se}{1-Se} \right] - \frac{1}{\sqrt{n}} \frac{1}{Se(1-Se)} g^{-1}(z_{1-\alpha/2})$$

$$UL = \ln \left[\frac{Se}{1-Se} \right] + \frac{1}{\sqrt{n}} \frac{1}{Se(1-Se)} g^{-1}(z_{\alpha/2})$$

$$g^{-1}(x) = -\sqrt{n} \frac{1}{(\gamma/6)} \left[\left(1 - (\gamma/2) \left(\frac{1}{\sqrt{n}} \frac{1}{x-n} \frac{\gamma}{6} \right) \right)^{1/3} - 1 \right]$$

$$\gamma = \frac{1-2Se}{\sqrt{Se(1-Se)}}$$

$$\text{신뢰구간: } \left[\frac{e^{LL}}{1+e^{LL}}, \frac{e^{UL}}{1+e^{UL}} \right]$$

Wilson's method

이 방법은 score confidence interval이라고도 하며(1,2,11, 14), Table 2의 기호를 사용하여 민감도와 특이도의 신뢰구간은 다음과 같이 계산한다.

$$\begin{aligned} \text{민감도의 신뢰구간: } & [(A-B)/C, (A+B)/C] \\ & A = 2a + z_{1-\alpha/2}^2, \quad B = z_{1-\alpha/2} \sqrt{z_{1-\alpha/2}^2 + 4ac} / (a+c) \\ & C = 2(a+c + z_{1-\alpha/2}^2) \end{aligned}$$

$$\begin{aligned} \text{특이도의 신뢰구간 : } & [(A-B)/C, (A+B)/C] \\ & A = 2d + z_{1-\alpha/2}^2, \quad B = z_{1-\alpha/2} \sqrt{z_{1-\alpha/2}^2 + 4bd} / (b+d) \\ & C = 2(b+d + z_{1-\alpha/2}^2) \end{aligned}$$

특별한 상황으로 $a=0$ 일 때와 $N=a$ 인 경우 신뢰구간은 다음과 같이 계산한다.

$$a=0 \text{ 일 때 신뢰구간 : } \left[0, \frac{z_{1-\alpha/2}^2}{N+z_{1-\alpha/2}^2} \right]$$

$$N=a \text{ 일 때 신뢰구간 : } \left[\frac{N}{N+z_{1-\alpha/2}^2}, 1 \right]$$

예를 들어 질병 A로 진단받은 시료 50개와 대조군 시료 100개에 대하여 진단검사를 적용한 결과가 Table 3과 같다고 할 때 이 검사의 민감도와 특이도의 신뢰구간을 계산하면 다음과 같다.

방법 1: 정규분포 근사성(Wald method)

$$Se = \frac{35}{50} = 0.7, \quad Var(Se) = \frac{35 \times 15}{50^3} = 0.0042,$$

$$\text{민감도의 95\% 신뢰구간: } [0.573, 0.828]$$

$$Sp = \frac{92}{100} = 0.92, \quad Var(Sp) = \frac{8 \times 92}{100^3} = 0.000736$$

$$\text{특이도의 95\% 신뢰구간: } [0.867, 0.974]$$

방법 2: 정확 이항 신뢰구간(exact binomial)

검사 수 $n=50$, 검사음성 수 $r=35$,

Table 3. 2×2 frequency contingency table for a diagnostic test

Test result	Disease status		Total
	Positive (D+)	Negative (D-)	
Positive (T+)	35	8	43
Negative (T-)	15	92	107
Total	50	100	150

$$F_{\alpha/2, 2n-2r+2, 2r} = F_{0.025, 2*50-2*35+2, 2*35} = F_{0.025, 32, 70} = 1.7616$$

$$F_{1-\alpha/2, 2r+2, 2n-2r} = F_{0.025, 2*35+2, 2*50-2*35} = F_{0.025, 72, 30} = 1.916$$

$$\text{민감도 신뢰하한: } \pi_L = \frac{35}{35 + (50-35+1) \times 1.7616} = 0.5539$$

$$\text{민감도 신뢰상한: } \pi_U = \frac{35+1}{35+1 + \frac{50-35}{1.916}} = 0.8213$$

검사 수 $n=100$, 검사음성 수 $r=92$,

$$F_{\alpha/2, 2n-2r+2, 2r} = F_{0.025, 2*100-2*92+2, 2*92} = F_{0.025, 18, 184} = 1.826$$

$$F_{1-\alpha/2, 2r+2, 2n-2r} = F_{0.025, 2*92+2, 2*100-2*92} = F_{0.025, 196, 16} = 2.3597$$

이므로

$$\text{특이도 신뢰하한: } \pi_L = \frac{92}{92 + (100-92+1) \times 1.826} = 0.8484$$

$$\text{특이도 신뢰상한: } \pi_U = \frac{92+1}{92+1 + \frac{100-92}{2.3597}} = 0.9648$$

방법 3: Wilson's method(score confidence interval)

$A = 73.84, B = 13.27, C = 107.68$, 민감도의 95% 신뢰구간: [0.5625, 0.8090]

$A = 187.84, B = 11.31, C = 207.68$, 특이도의 95% 신뢰구간: [0.8500, 0.9589]

진단정확도(진단일치도)

질병 상태의 참값을 알고 있는 검체에 대하여 진단검사 2종으로 평가한 결과가 Table 4와 같다고 할 때 총 검사 건수에서 진양성과 진음성이 차지하는 비율을 진단정확도(diagnostic accuracy)라고 한다. 완벽한 표준검사가 아닌 기존 검사법(index test 1)과 경쟁관계에 있는 신규 검사법(index test 2) 2종의 진단검사를 특성을 비교하는 경우에는 진단일치도(diagnostic agreement)라고 하며 계산공식은 진단정확도와 동일하다.

$$\text{Diagnostic agreement} = \frac{a+d}{N}$$

Score confidence interval을 이용한 진단일치도의 신뢰구간은 다음과 같이 계산한다.

$$A = 2(a+d) + z_{1-\alpha/2}^2$$

$$B = z_{1-\alpha/2} \sqrt{z_{1-\alpha/2}^2 + \frac{4(a+d)(b+c)}{N}} \quad (\text{여기에서 } N = a + b + c + d)$$

$$C = 2(N + z_{1-\alpha/2}^2)$$

$$\text{신뢰구간 : } \left[\frac{A-B}{C}, \frac{A+B}{C} \right]$$

Table 4. 2×2 contingency table showing responses of two binary diagnostic tests

Test result (T_2)	Test result (T_1)		Total
	Positive (T+)	Negative (T-)	
Positive (T+)	50	5	55
Negative (T-)	10	600	610
Total	60	605	665

예를 들어 Table 4의 경우 $A = 1,303.84$, $B = 15.49$, $C = 1,337.68$ 이므로 두 진단검사의 진단일치도에 대한 95% 신뢰구간은 $[0.9631, 0.9863]$ 으로 계산된다.

kappa

Table 1의 기호를 사용하여 2×2 분할표에서 kappa 통계량의 표준오차와 신뢰구간은 다음과 같이 계산한다(7).

$$SE(\kappa) = \frac{1}{\sqrt{n(1-p_e)^2}} \times \left(p_e + p_e^2 - \sum_{i=1}^c [a_i b_i (a_i + b_i)] \right)$$

$$= \frac{1}{(1-p_e)\sqrt{n}} \sqrt{p_e + p_e^2 - \sum_{i=1}^c [a_i b_i (a_i + b_i)]}$$

$$SE(\kappa) = \frac{1}{\sqrt{(1-p_e)^2 n}} \times \left\{ p_e + p_e^2 - \sum_{i=1}^c \frac{r_i c_i}{n^2} (r_i + c_i) \right\}$$

$SE(\kappa)$ 의 근사식: $SE(\kappa) = \sqrt{p_0(1-p_0)/n(1-p_e)^2}$
 신뢰구간: $\kappa \pm z_{1-\alpha/2} SE(\kappa)$

예측도

Table 1의 기호를 사용하여 양성예측도(PPV)와 음성예측도(NPV)의 분산과 신뢰구간을 계산하면 다음과 같다.

$$PPV = \frac{a}{m_1}, \quad Var(PPV) = \frac{ab}{m_1^2} \quad \text{또는} \quad Var(PPV) = \frac{PPV(1-PPV)}{m_1}$$

$$NPV = \frac{d}{m_2}, \quad Var(NPV) = \frac{cd}{m_2^2} \quad \text{또는} \quad Var(NPV) = \frac{NPV(1-NPV)}{m_2}$$

$$PPV \pm z_{1-\alpha/2} \sqrt{\frac{ab}{m_1}} \quad (\text{여기에서 } m_1 = a + b)$$

$$NPV \pm z_{1-\alpha/2} \sqrt{\frac{cd}{m_2}} \quad (\text{여기에서 } m_2 = c + d)$$

한편 분산을 추정하는 방법으로 Delta method (9,10)를 이용한 예측도의 분산은 다음과 같다.

$$Var(PPV) = \frac{[p(1-Sp)(1-p)]^2 \frac{Se(1-Se)}{n_1} + [pSe(1-p)]^2 \frac{Sp(1-Sp)}{n_2}}{[Sep + (1-Sp)(1-p)]^4}$$

$$Var(NPV) = \frac{[Sp(1-p)p]^2 \frac{Se(1-Se)}{n_1} + [(1-Se)(1-p)p]^2 \frac{Sp(1-Sp)}{n_2}}{[(1-Se)p + (Sp(1-p))]^4}$$

예를 들어 질병 A를 검출하기 위한 분변배양검사 결과가 Table 5와 같다고 할 때 분변검사의 민감도, 특이도, 예측도 및 유병률의 참값(true prevalence)에 대한 신뢰구간을 평가하면 다음과 같다.

민감도(Se)와 특이도(Sp)

$$Se = \frac{63}{252} = 0.25, \quad Var(Se) = \frac{Se(1-Se)}{n_1} = \frac{ac}{n_1^2} = 0.000744$$

$$Se \pm z_{1-\alpha/2} \sqrt{\frac{Se(1-Se)}{n_1}} = 0.25 \pm 1.96 \times \sqrt{0.000744} \leftrightarrow [0.20, 0.30]$$

$$Sp = \frac{2299}{2346} = 0.98, \quad Var(Sp) = \frac{Sp(1-Sp)}{n_2} = \frac{bd}{n_2^2} = 8.37E-06$$

$$Sp \pm z_{1-\alpha/2} \sqrt{\frac{Sp(1-Sp)}{n_2}} = 0.98 \pm 1.96 \times \sqrt{8.37E-06} \leftrightarrow [0.97, 0.99]$$

양성예측도(PPV)와 음성예측도(NPV)

$$PPV = \frac{a}{m_1} = \frac{63}{110} = 0.57, \quad Var(PPV) = \frac{ab}{m_1^2} = \frac{63 \times 47}{110^2} = 0.00222$$

Table 5. Fecal culture test result

Test result	Disease status		Total
	Positive (D+)	Negative (D-)	
Positive (T+)	63	47	110
Negative (T-)	189	2,299	2,488
Total	252	2,346	2,598

$$NPV = \frac{d}{m_2} = \frac{2299}{2488} = 0.92, \quad Var(NPV) = \frac{cd}{m_2^2} = \frac{189 \times 2299}{2488^2} = 2.82E-05$$

$$PPV: PPV \pm z_{1-\alpha/2} \sqrt{\frac{ab}{m_1}} = 0.57 \pm 1.96 \times \sqrt{0.00222} \leftrightarrow [0.48, 0.67]$$

$$NPV: NPV \pm z_{1-\alpha/2} \sqrt{\frac{cd}{m_2}} = 0.92 - (1.96 \times \sqrt{2.82E-05}) \leftrightarrow [0.91, 0.93]$$

유병률의 참값(true prevalence, p)

$$p = \frac{252}{2598} = 0.097, \quad Var(p) = \frac{0.097 \times (1-0.097)}{2598} = 3.37E-05$$

$$p: p \pm z_{1-\alpha/2} \sqrt{\frac{p(1-p)}{n}} = 0.097 \pm 1.96 \times \sqrt{3.37E-05} \leftrightarrow [0.086, 0.108]$$

우도비

우도비(likelihood ratio)는 비대칭형의 왜곡된 분포를 보이므로 대수변환 우도비가 정규분포에 더 근사한다. Table 1의 기호를 사용하여 대수변환된 양성우도비[LR(+)]와 음성우도비[LR(-)]의 분산 추정치와 신뢰구간은 다음과 같이 계산한다.

$$\widehat{Var}[\ln(\widehat{LR}+)] = \frac{1-Se}{a} + \frac{Sp}{b}$$

$$LR(+): \exp \left[\ln \frac{Se}{1-Sp} \pm z_{1-\alpha/2} \sqrt{\frac{1-Se}{a} + \frac{Sp}{b}} \right]$$

$$\widehat{Var}[\ln(\widehat{LR}-)] = \frac{Se}{c} + \frac{1-Sp}{d}$$

$$LR(-): \exp \left[\ln \frac{1-Se}{Sp} \pm z_{1-\alpha/2} \sqrt{\frac{Se}{c} + \frac{1-Sp}{d}} \right]$$

예를 들어 Table 5의 자료에 대하여 $LR(+)=12.5$ 의 95% 신뢰구간은 $[8.76, 17.8]$, $LR(-)=0.77$ 의 95% 신뢰구간은 $[0.72, 0.83]$ 로 계산된다.

교차비

교차비(odds ratio, OR)는 우도비와 마찬가지로 왜곡된 분포를 보이므로 대수변환 교차비가 정규분포에 근사한다. 대수변환된 교차비의 분산과 신뢰구간은 다음과 같이 계산된다.

$$Var(\ln(OR)) = \frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}$$

$$\text{신뢰구간: } \frac{ad}{bc} \exp \left[\pm z_{1-\alpha/2} \sqrt{\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}} \right]$$

만일 Table 1의 a, b, c, d 셀 중 어느 하나가 0인 경우 교차비의 추정치가 왜곡될 수 있기 때문에 각 셀에 0.5를 더한 후 교차비와 분산을 계산한다(5). Table 5의 자료에서 교차비(OR = 16.3)의 95% 신뢰구간은 $[10.9, 24.5]$ 로 계산된다.

감사의 글

본 연구는 2014년도 농림축산검역본부 용역연구사업(조류 인플루엔자 유입감시 체계 구축, Project code: Z-1543068-2014-14-01) 및 강원대학교 동물의학종합연구소의 지원으로 수행되었습니다.

참고문헌

1. Agresti A, Coull BA. Approximate is better than exact for interval estimation of binomial proportions. *Am Stat* 1998; 52: 119-126.
2. Altman DA, Machin D, Bryant TN, Gardner MJ. *Statistics with Confidence*. 2nd ed. British Medical Journal, 2000.
3. Beam CA. Analysis of clustered data in receiver operating characteristic studies. *Stat Methods Med Res* 1998; 7: 324-336.
4. Clopper CJ, Pearson ES. The use of confidence or fiducial limits illustrated in the case of the binomial. *Biometrika* 1934; 26: 404-413.
5. Cox DR. *The analysis of binary data*. Methuen, London. 1970.
6. Daly L. Simple SAS macros for the calculation of exact binomial and Poisson confidence limits. *Comput Biol Med* 1992; 22: 351-361.
7. Fleiss JL. *Statistical methods for rates and proportions*. John Wiley & Sons, New York, 1981.
8. ICMJE (International committee of medical journal editors). *Recommendations for the conduct, reporting, editing, and publication of scholarly work in medical journals*. ICMJE, 2014.
9. Mercaldo ND, Lau KF, Zhou XH. Confidence intervals for predictive values with an emphasis to case-control studies. *Stat Med* 2007; 26: 2170-2183.
10. Mossman D, Berger JO. Intervals for posttest probabilities: a comparison of 5 methods. *Med Decis Making* 2001; 21: 498-507.
11. Newcombe RG. Improved confidence intervals for the difference between binomial proportions based on paired data. *Stat Med* 1998; 17: 2635-2650.
12. Pak SI, Lee YW. An introduction to data analysis. *J Vet Clin* 2009; 26: 189-199.
13. Simel DL, Samsa GP, Matchar DB. Likelihood ratios with confidence: sample size estimation for diagnostic test studies. *J Clin Epidemiol* 1991; 44: 763-770.
14. Wilson EB. Probable inference, the law of succession, and statistical inference. *J Am Stat Assoc* 1927; 22: 209-212.
15. Zhou XH1, Li CM, Yang Z. Improving interval estimation of binomial proportions. *Philos Trans A Math Phys Eng Sci* 2008; 366: 2405-2418.