JP Journal of Biostatistics
Volume 2, Number 3, 2008, Pages 185-191
Published Online: January 10, 2009
This paper is available online at http://www.pphmj.com

A NEW MEASURE BASED ON SENSITIVITY AND SPECIFICITY

NOBUKO MIYAMOTO, TOMOAKI NARUSHIMA and SADAO TOMIZAWA

Department of Information Sciences
Faculty of Science and Technology
Tokyo University of Science
Noda City, Chiba, 278-8510
Japan
e-mail: tomizawa@is.noda.tus.ac.jp

Abstract

For contingency tables, there are two measures called the sensitivity and specificity. These are calculated from one contingency table as the separate measures. This paper proposes a new measure which is obtained by combining sensitivity with specificity. By using this measure, we analyze several data on the relation between tuberculosis and diagnosis and we evaluate the performance of diagnostic tests.

1. Introduction

With diagnostic tests for a disease, the two correct diagnoses are a positive test outcome when the subject has the disease and a negative test outcome when a subject does not have it. Given that the subject has the disease, the conditional probability that the diagnostic test is positive is called the *sensitivity*; given that the subject does not have the disease, the conditional probability that the diagnostic test is negative is called the *specificity*. Ideally, these are both high. See Agresti [1, p. 38] and Bishop et al. [2, p. 380].

2000 Mathematics Subject Classification: 62H17.

Keywords and phrases: sensitivity, specificity.

Received July 25, 2008

Consider the data in Table 1, taken from Yerushalmy et al. [3], which describes the relation between tuberculosis and diagnosis by six qualified radiologists and chest specialists designated (A), (B), (C), (D), (E) and (F). These alphabets correspond to the titles of the tables. The row variable shows the diagnosis (positive, negative), where the positive outcome predicts that a subject has inflammatory disease. The column variable shows the true disease status (i.e., whether a subject truly has tuberculosis). We are interested in two measures; sensitivity and specificity.

In Table 1, the sensitivity of (A) is estimated to be 0.733. Of subjects with tuberculosis, 73.3% were diagnosed correctly. The specificity is estimated to be 0.984. Of subjects not having tuberculosis, 98.4% were diagnosed correctly. Similarly, in (B), (C), (D), (E) and (F), the sensitivity are estimated to be 0.567, 0.700, 0.600, 0.733 and 0.733, respectively. The specificity are estimated to be 0.991, 0.991, 0.971, 0.988 and 0.974, respectively.

If we decide who is the best person to diagnose, (A), (E) and (F) are preferable to the others with regard to the sensitivity. Similarly, in the specificity, (B) and (C) are superior compared with the others.

The purpose of this paper is to propose a measure which combines the sensitivity with the specificity.

2. New Measure

For a 2×2 contingency table, let X and Y denote the row and column variables, respectively. Let p_{ij} denote a probability of the ith row and jth column in the table $(i = 1, 2;\ j = 1, 2)$. Then the sensitivity and the specificity are given by

$$p_{1(1)} = \frac{p_{11}}{p_{+1}}$$

and

$$p_{2(2)} = \frac{p_{22}}{p_{+2}},$$

respectively, where $p_{+j} = \sum_{s} p_{sj}$.

Sample distributions use similar notation with $\hat{p}$. The cell frequencies are denoted by $\{n_{ij}\}$, and $n = \sum_i \sum_j n_{ij}$ is the total sample size. Thus,

$$\hat{p}_{ij} = \frac{n_{ij}}{n}.$$

The sample proportion of times that subjects in column j made response i is

$$\hat{p}_{i(j)} = \frac{\hat{p}_{ij}}{\hat{p}_{+j}} = \frac{n_{ij}}{n_{+j}},$$

where $n_{+j} = n\hat{p}_{+j} = \sum_i n_{ij}$.

Consider the following measure:

$$\phi = p_{1(1)} + p_{2(2)} - 1.$$

We see that the measure ϕ must lie between $-1 \leq \phi \leq 1$, and (i) $\phi = -1$ (when the diagnosis is useless absolutely) if and only if $p_{11} = p_{22} = 0$, (ii) $\phi = 1$ (when the diagnosis is certainly successful) if and only if $p_{12} = p_{21} = 0$, and (iii) $\phi = 0$ if and only if X and Y are independent.

The sample version of ϕ, i.e., $\hat{\phi}$, is given by ϕ with $\{p_{ij}\}$ replaced by $\{\hat{p}_{ij}\}$, where $\hat{p}_{ij} = n_{ij}/n$. Using the delta method, $\sqrt{n}(\hat{\phi} - \phi)$ has asymptotically (as $n \to \infty$) a normal distribution with mean zero and variance $\sigma^2[\phi]$, where

$$\sigma^2[\phi] = \frac{p_{11}p_{21}}{p_{+1}^3} + \frac{p_{12}p_{22}}{p_{+2}^3}.$$

Let $\hat{\sigma}^2[\phi]$ denote $\sigma^2[\phi]$ with $\{p_{ij}\}$ replaced by $\{\hat{p}_{ij}\}$. Then $\hat{\sigma}[\phi]/\sqrt{n}$ is an estimated standard error for $\hat{\phi}$, and $\hat{\phi} \pm z_{p/2}\hat{\sigma}[\phi]/\sqrt{n}$ is an approximate $100(1 - p)$ percent confidence interval for ϕ, where $z_{p/2}$ is the percentage point from the standard normal distribution corresponding to a two-tail probability equal to p.

3. Examples

Consider the data in Table 1. Table 2 gives the estimates of the measure ϕ, the estimated approximate standard error for $\hat{\phi}$, the approximate 95% confidence interval for ϕ and the estimated correlation. When we pay attention only to the value of estimated measure, (E) is the best person to diagnose and (B) is the worst. Although (A), (E) and (F) have same value of the estimated sensitivity 0.733, the estimated measure $\hat{\phi}$ indicates that (E) is preferable to others. In addition, all the values of estimated measures are closer to 1 than 0, so the diagnoses are regarded as good basically.

Consider the confidence intervals of the measure ϕ. Since each interval is overlapped, we cannot decide who is superior or inferior from the confidence intervals. But the confidence intervals of (B) and (D) include values which are less than 0.5, although those of the other tables do not include them. Thus (B) and (D) might be not good at this diagnosis.

4. Concluding Remarks

The measure $\hat{\phi}$ could be useful when we want to know who is the best person to diagnose or which is the best diagnostics. For a 2×2 contingency table, the correlation coefficient is described as follows (Bishop et al. [2, p. 381]):

$$\rho = \frac{p_{11}p_{22} - p_{21}p_{12}}{\sqrt{p_{1+}p_{2+}p_{+1}p_{+2}}},$$

where $p_{i+} = \sum_t p_{it}$ $(i = 1, 2)$. Assuming that the marginal probabilities are positive, the row and column variables are independent if and only if $\rho = 0$; $p_{12} = p_{21} = 0$ if and only if $\rho = 1$; and $p_{11} = p_{22} = 0$ if and only if $\rho = -1$.

These are the identical conditions with our new measure and the values of ϕ and ρ are completely same in the case of $p_{12} = p_{21}$. But there is a difference of feature between ϕ and ρ. The proposed measure ϕ is useful when we want to focus on the *sensitivity* and *specificity* and to check an effectiveness of a diagnosing by using a summary measure being a function of sensitivity and specificity, although the correlation ρ focuses on the distance from the independence and cares about the linear relation.

For example, the values of $\hat{\phi}$ for (E) and (F) in Table 1 are close, even though the values of $\hat{\rho}$ are distant each other (see Table 2). Furthermore, although the values of $\hat{\rho}$ for (C) is the highest, the ranking of the values of $\hat{\phi}$ is fourth. The ranking of $\hat{\phi}$ is (E), (A), (F), (C), (D) and (B) from the top. The ranking of $\hat{\rho}$ is (C), (E), (A), (B), (F) and (D). Thus there is a great difference of ranking between $\hat{\phi}$ and $\hat{\rho}$. It is useful to use the measure ϕ in case that we want to know who is the best person to diagnose and we want to focus on the sensitivity and specificity.

References

[1] A. Agresti, Categorical Data Analysis, 2nd ed., Wiley, New York, 2002.

[2] Y. M. M. Bishop, S. E. Fienberg and P. W. Holland, Discrete Multivariate Analysis: Theory and Practice, MIT Press, MA, Cambridge, 1975.

[3] J. Yerushalmy, J. T. Harkness, J. H. Cope and B. R. Kennedy, The role of dual reading in mass radiography, Amer. Rev. Tuberculosis 61 (1950), 443-464.

Table 1. The relation between tuberculosis and diagnosis by six qualified radiologists and chest specialists; from Yerushalmy et al. [3]

(A)

	Disease		
Diagnosis	Yes	No	Total
Positive	22	29	51
Negative	8	1731	1739
Total	30	1760	1790

(B)

	Disease		
Diagnosis	Yes	No	Total
Positive	17	15	32
Negative	13	1745	1758
Total	30	1760	1790

(C)

	Disease		
Diagnosis	Yes	No	Total
Positive	21	16	37
Negative	9	1744	1753
Total	30	1760	1790

(D)

	Disease		
Diagnosis	Yes	No	Total
Positive	18	51	69
Negative	12	1709	1721
Total	30	1760	1790

(E)

	Disease		
Diagnosis	Yes	No	Total
Positive	22	21	43
Negative	8	1739	1747
Total	30	1760	1790

(F)

	Disease		
Diagnosis	Yes	No	Total
Positive	22	46	68
Negative	8	1714	1722
Total	30	1760	1790

Table 2. Estimate of ϕ, estimated approximate standard error for $\hat{\phi}$, approximate 95% confidence interval for ϕ and estimated correlation $\hat{\rho}$, applied to Table 1

Applied data	Estimated measure $\hat{\phi}$	Standard error $\hat{\sigma}[\phi]/\sqrt{n}$	Confidence interval $\hat{\phi} \pm z_{0.025}\hat{\sigma}[\phi]/\sqrt{n}$	Estimated correlation $\hat{\rho}$
Table (A)	0.717	0.081	(0.558, 0.875)	0.553
Table (B)	0.558	0.090	(0.381, 0.736)	0.541
Table (C)	0.691	0.084	(0.527, 0.855)	0.623
Table (D)	0.571	0.090	(0.396, 0.747)	0.381
Table (E)	0.721	0.081	(0.563, 0.880)	0.605
Table (F)	0.707	0.081	(0.549, 0.866)	0.475

■