



Math for the people, by the people.

odds ratio

Canonical name	OddsRatio
Date of creation	2013-03-22 14:41:07
Last modified on	2013-03-22 14:41:07
Owner	CWoo (3771)
Last modified by	CWoo (3771)
Numerical id	10
Author	CWoo (3771)
Entry type	Definition
Classification	msc 62H20
Classification	msc 62H17
Defines	odds
Defines	log-odds ratio

Suppose the probability of an event A is not 1. The *odds* of the event A is the ratio

$$\text{odds}(A) := \frac{P(A)}{P(A^c)} = \frac{P(A)}{1 - P(A)}.$$

For example, if the odds of landing a head in a coin flip is 2, then the probability of landing a head is twice as likely as that of landing a tail.

In a 2×2 contingency table, with a dichotomous explanatory variable X having levels 1 and 2, and a binary response variable Y with success and failure as two possible levels of outcomes:

	success	failure
1	n_{11}	n_{12}
2	n_{21}	n_{22}

where, given the i th level of X , n_{i1} and n_{i2} are the counts of success and failure, respectively. Two levels of odds for success can be formed:

$$\text{odds}(Y = \text{success} \mid X = i) = \frac{n_{i1}/(n_{i1} + n_{i2})}{n_{i2}/(n_{i1} + n_{i2})} = \frac{n_{i1}}{n_{i2}}, i = 1, 2.$$

We can form the ratio of these two odds, called the *odds ratio*:

$$OR_{XY} = \frac{\text{odds}(Y = \text{success} \mid X = 1)}{\text{odds}(Y = \text{success} \mid X = 2)}.$$

To interpret odds ratio, we look at some hypothetical examples:

1. Suppose that, during an average year, 35 out of 100 young drivers get involved in traffic violations, while 10 out of 100 adult drivers are involved. The odds of a young driver getting involved in traffic violations in a typical year is $35/65$, and for the adults, it is $10/90$. Calculating the odds ratio, $(35/65)/(10/90) \approx 4.85$, and we find that a young driver is almost 5 times as likely to get a traffic ticket as an adult. This indicates that the driver's age (the explanatory variable) and the chance of getting a traffic ticket (the response variable) might be associated.
2. Another example. Suppose a die is tossed 100 times and the face with one dot is observed 20 times. Another die twice the size of the first die is tossed 100 times and one is observed 16 times. The odds ratio of getting

one between the smaller die and the bigger die is $(20/80)/(16/84) \approx 1.31$, which shows that the odds of getting a one is about the same for the smaller die as for the bigger die.

From the two examples above, we see that the odds ratio can be used to test the association of two dichotomous variables. The further away OR is from 1, the higher the association is between the two variables.

On the other hand, the closer to 1 the odds ratio is, the closer to independence between the variables. In fact, odds ratio = 1 iff the two dichotomous variables are independent, as can be readily seen in the following argument:

$$OR_{XY} = 1 \text{ iff } \text{odds}(Y = \text{success} \mid X = 1) = \text{odds}(Y = \text{success} \mid X = 2).$$

From the last equation, we see that

$$\begin{aligned} P(Y = \text{success} \mid X = 1) &= \frac{\text{odds}(Y = \text{success} \mid X = 1)}{1 + \text{odds}(Y = \text{success} \mid X = 1)} \\ &= \frac{\text{odds}(Y = \text{success} \mid X = 2)}{1 + \text{odds}(Y = \text{success} \mid X = 2)} \\ &= P(Y = \text{success} \mid X = 2). \end{aligned}$$

Because the random variable X is dichotomous, we see that

$$P(Y = \text{success}) = P(Y = \text{success} \mid X = i), i=1,2.$$

Remarks

- Since the odds ratio lies in the interval $[0, \infty)$, 1 is highly skewed towards 0. So we see that even though 10 and 0.1 both indicate similar degrees of association, 0.1 is a lot closer to 1 than 10 is to 1.
- By taking the natural log of the odds ratio, $\ln(OR_{XY})$, we see that the left boundary value 0 of the odds ratio is now stretched to $-\infty$, and testing of independence is now boiled down to testing whether the *log-odds ratio*, is 0 or not. It turns out that the log-odds ratio has an asymptotic normal distribution, which can be used to find the confidence interval of OR.
- Another way of transforming the odds ratio so as to get a more symmetrical distribution is:

$$\frac{OR - 1}{OR + 1}.$$

The transformed odds ratio lies in the interval $[-1, 1]$. This transformed value is also known as the *Yule's Q*.

- The use of odds ratios can be generalized to applications of M by N 2-way contingency tables, where $M > 2$ and $N \geq 2$, as well as higher way contingency tables involving more than one explanatory variable.

References

- [1] A. Agresti, *An Introduction to Categorical Data Analysis*, Wiley & Sons, New York (1996).