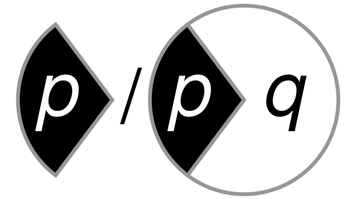
Probability, log-odds, and odds

Yujin Kim

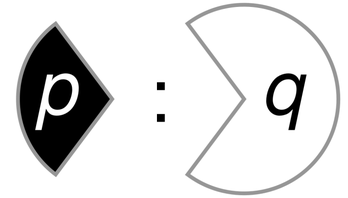
Definition

- Probability is simply how likely something is to happen, the ratio of the number of favorable outcomes to the total number of outcomes of an event.
- In statistics, **odds** are an expression of relative probabilities, generally quoted as the odds *in favor*. They are calculated as the ratio of the number of events that produce that outcome to the number that do not.
- The probability of an event can be calculated from the odds, and vice versa.

$$\text{Probability or risk} = \frac{p}{p+q}$$



$$\text{Odds} = p : q$$



Odds ratio

- An odds ratio (OR) is a statistic that quantifies the strength of the association between two events, A and B.

- The odds ratio is defined as the ratio of the odds of A in the presence of B and the odds of A in the absence of B, or equivalently (due to symmetry), the ratio of the odds of B in the presence of A and the odds of B in the absence of A.

- The odds are a ratio of probabilities; an odds ratio is a ratio of odds, that is, a ratio of ratios of probabilities.

work with a range of values on one scale and convert it to the other

If the probability of an event is 0.2, that

1. The odds of the event occurring are $$odds = \frac{0.2}{0.8} = 0.25$$
2. The log-odds of the event occurring are $$ln\left(\frac{0.2}{0.8}\right) = -1.3863$$
3. The probability can be reconstructed as $$\frac{odds}{1 + odds} = \frac{0.25}{1.25} = 0.2$$
4. The probability can also be reconstructed as

$$\frac{exp(ln(odds))}{1 + exp(ln(odds))} = \frac{exp(-1.3683)}{1 + exp(-1.3683)} = \frac{0.25}{1.25} = 0.2$$

Applications

- In probability theory and statistics, odds and similar ratios may be more natural or more convenient than probabilities. In some cases the **log-odds** are used, which is **the logit of the probability**. Most simply, odds are frequently multiplied or divided, and log converts multiplication to addition and division to subtractions.

Logit function

$$\text{logit}(p) = \log(\frac{p}{1-p})$$

p = probability;

$\frac{p}{1-p}$ = corresponding odds

$$ln(Odds) = ln(\frac{p}{1-p}) = ln(\frac{\frac{1}{1+e^{-(\beta_0+\beta_1 X_1+\cdots+\beta_p X_p)}}}{1-\frac{1}{1+e^{-(\beta_0+\beta_1 X_1+\cdots+\beta_p X_p)}}}) = \beta_0+\beta_1 X_1+\cdots+\beta_p X_p$$

Interpretation of logistic regression

$\text{logit}(p) = \beta_0 + \beta_1 * \textbf{female}$

```
Logistic regression                                     Number of obs   =        200
                                                        LR chi2(1)      =       3.10
                                                        Prob > chi2     =     0.0781
Log likelihood = -109.80312                             Pseudo R2       =     0.0139

------------------------------------------------------------------------------
         hon |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      female |   .5927822   .3414294     1.74   0.083    -.0764072    1.261972
   intercept |  -1.470852   .2689555    -5.47   0.000    -1.997995   -.9437087
------------------------------------------------------------------------------
```

- Get the odds ration by exponentiating the coef. for remale; math.exp(0.5927822) = 1.809015 (Odds Ratio)
- Odds of being female on P (DV) has 80% higher than odds of being male.

- References

- https://www.montana.edu/rotella/documents/502/Prob_odds_log-odds.pdf

- https://en.wikipedia.org/wiki/Odds_ratio

- https://stats.oarc.ucla.edu/other/mult-pkg/faq/general/faq-how-do-i-interpret-odds-ratios-in-logistic-regression/