Odds and Odds Ratio

- The odds in favor of an event or a proposition are the ratio of the probability that an event will happen to the probability that it will not happen.
- 'Odds' are an expression of relative probabilities. Often 'odds' are quoted as odds against, rather than as odds in favor of, because of the possibility of confusion of the latter with the fractional probability of an event occurring.

$$\text{Odds} = \frac{p}{1-p}$$

- Odds value can range from 0 to infinity and tell you how much more likely it is that an observation is a member of the target group rather than a member of the other group.
  - If the probability is 0.80, the odds are 4 to 1 or 0.80/0.20.
  - If the probability is 0.25, the odds are 0.33 (0.25/0.75).
- If the probability of membership in the target group is 0.50, the odds are 1 to 1 (0.50/0.50), as in coin tossing when both outcomes are equally likely.
- Another important concept is the **odds ratio (OR)**, which estimates the change in the odds of membership in the target group for a one unit increase in the predictor. It is calculated by using the regression coefficient of the predictor as the exponent.
  - Suppose we were predicting exam success by a maths competency predictor with an estimate `b = 2.69`.
  - Thus the odds ratio is `exp(2.69)` or `14.73`.
  - Therefore the odds of passing are 14.73 times greater for a student, for example, who had a pre-test score of 5, than for a student whose pre-test score was 4.
- Suppose that the probability of success is 0.8, thus p = 0.8
  - Then the probability of failure is

$$q = 1 - p = 0.2$$

  - The odds of success are defined as

$$\text{Odds}(success) = p/q = 0.8/0.2 = 4$$

    that is, the odds of success are 4 to 1.
  - The odds of failure would be

$$\text{Odds}(failure) = q/p = 0.2/0.8 = .25$$

  - The odds of success and the odds of failure are just reciprocals of one another, i.e., 1/4 = 0.25 and 1/0.25 = 4.

- Logistic regression calculates changes in the log odds of the dependent, not changes in the dependent value as OLS regression does. For a dichotomous variable the odds of membership of the target group are equal to the probability of membership in the target group divided by the probability of membership in the other group.
  - The odds in favor of an event or a proposition are the ratio of the probability that an event will happen to the probability that it will not happen.
  - 'Odds' are an expression of relative probabilities. Often 'odds' are quoted as odds against, rather than as odds in favor of, because of the possibility of confusion of the latter with the fractional probability of an event occurring.

$$\text{Odds} = \frac{p}{1-p}$$

$$p = \frac{\text{Odds}}{1+\text{Odds}}$$

Worked Example: Odds Ratio

Suppose that seven out of 10 males are admitted to an engineering school while three of 10 females are admitted.

- The probabilities for admitting a male are, p = 7/10 = 0.7 (q = 1 - 0.7 = 0.3)
- Here are the same probabilities for females, p = 3/10 = 0.3 (q = 1 - 0.3 = 0.7)

Now we can use the probabilities to compute the admission odds for both males and females,

- odds(male) = 0.7/0.3 = 2.33333
- odds(female) = 0.3/0.7 = 0.42857

Next, we compute the odds ratio for admission,

$$\text{OR} = \frac{2.3333}{0.42857} = 5.44$$

Thus, for a male, the odds of being admitted are 5.44 times as large than the odds for a female being admitted.

Odds Ratios and Logistic Regression

- When a logistic regression is calculated, the regression coefficient (b_1) is the estimated increase in the log odds of the outcome per unit increase in the value of an **exposure** variable (i.e. a relevant non-demographic variable).
- In other words, the exponential function of the regression coefficient (e^{b_1})) is the odds ratio associated with a one-unit increase in the exposure.

- Odds ratios are used to compare the relative odds of the occurrence of the outcome of interest (e.g. disease or disorder), given exposure to the variable of interest (e.g. health characteristic, aspect of medical history).
- The odds ratio can also be used to determine whether a particular exposure is a risk factor for a particular outcome, and to compare the magnitude of various risk factors for that outcome.

* OR= 1 Exposure does not affect odds of outcome
* OR> 1 Exposure associated with higher odds of outcome
* OR< 1 Exposure associated with lower odds of outcome

Confidence Intervals for Odds Ratios

- Many statistical implementations of logistic regression include confidence intervals for the odds ratios.
- The 95% confidence interval (CI) is used to estimate the precision of the OR. A large confidence interval indicates a low level of precision of the OR, whereas a small confidence interval indicates a higher precision of the OR.
- Odds ratios whose confidence limits exclude 1 are statistically significant.
- The odds ratio is referred to in SPSS as `Exp(B)`, the exponentiation of the `B` coefficient.
- It is important to note however, that unlike the $p-$value, the 95% CI does not report a measure's statistical significance.
- In practice, the 95% confidence interval is often used as a proxy for the presence of statistical significance if it does not overlap the null value (e.g. `OR=1`).
- Nevertheless, it would be inappropriate to interpret an OR with 95% confidence interval that spans the null value as indicating evidence for lack of association between the exposure and outcome.