**Section 1.2.5 – Odds ratios**

Odds are the probability of success divided by the probability of a failure. With respect to a 2×2 contingency table, we have

|  |  |  |  |
| --- | --- | --- | --- |
|  |  | Response |  |
|  |  | Success | Failure |  |
| Group | 1 | π1 | 1 – π1 | 1 |
| 2 | π2 | 1 – π2 | 1 |

* For row 1, the “odds of a success” are
odds1 = π1/(1-π1).
* For row 2, the “odds of a success” are
odds2 = π2/(1-π2).

Notice that the odds are just a rescaling of the probability of success! For example, if P(success) = 0.75, then the odds are 3 or “3 to 1 odds”. The probability of success is three times as large as the probability of a failure.

The estimated odds are:

 and 

Notice what cells these correspond to in the contingency table:

|  |  |  |
| --- | --- | --- |
|  |  | Response |
|  |  | Success | Failure |
| Group | 1 | w1 | n1 – w1 |
| 2 | w2 | n2 – w2 |

Questions:

* What is the numerical range of an odds?
* What does it mean for an odds to be 1?

To incorporate information from both rows 1 and 2 into a single number, the ratio of the two odds is found. This is called an odds ratio. Formally, it is defined as:



**Odds ratios are VERY useful in categorical data analysis and will be used throughout this course!**

Questions:

* What is the numerical range of OR?
* What does it mean for OR to be 1?
* What does it mean for OR > 1?
* What does it mean for OR < 1?

The estimated odds ratio is



This is the MLE of OR. Notice that the estimate is a product of the counts on the “diagonal” (top-left to bottom-right) of the contingency table divided by a product of the counts on the off diagonal.

Interpretation

The interpretation of OR can be difficult for some people. I recommend that you always go back to the basic property that an odds ratio is the ratio of two odds. Below are some example interpretations:

1. The estimated odds of a success are  times as large for group 1 than for group 2.
2. The estimated odds of a success are  times as large for group 2 than for group 1.

Notice that case 2) inverts the odds ratio so that row 2 is divided by row 1.

Consider the case now of interpreting the odds of a failure (1-π1)/π1. Now, the ratio of group 1 to group 2 becomes

,

which was . This leads to the following additional interpretations:

1. The estimated odds of a failure are  times as large for group 1 than for group 2.
2. The estimated odds of a failure are  times as large for group 2 than for group 1.

What if a cell count in the contingency table is equal to 0? The estimate is 0 or undefined! An ad-hoc solution is to add a small constant to the cell or to all cells of the contingency table even if some are not 0.

Confidence interval

As with other maximum likelihood estimators, we could use a normal approximation with the statistic here and form a Wald confidence interval. However, the large sample normal approximation can be improved upon by working with the natural log transformation of the odds ratio first. Thus, a Wald confidence interval for  is



where

 = 

is derived through using a delta method approximation (see Appendix B).

Of course, we are still interested in OR itself, so we can apply the exponential function to find the Wald confidence interval for OR:



Comments:

* This interval is conservative. What does “conservative” mean here?
* What if a cell count is equal to 0? Same solution as for the estimate.

Example: Larry Bird (Bird.R)

|  |  |  |  |
| --- | --- | --- | --- |
|  |  | Second |   |
|   |   | Made | Missed | Total |
| First | Made | 251 | 34 | 285 |
| Missed | 48 | 5 | 53 |
|   | Total | 299 | 39 | 338 |

> c.table <- array(data = c(251, 48, 34, 5), dim = c(2,2), dimnames = list(First = c("made", "missed"), Second =

 c("made", "missed")))

> OR.hat <- c.table[1,1] \* c.table[2,2] / (c.table[2,1] \*

 c.table[1,2])

> round(OR.hat, 2)

[1] 0.77

> round(1/OR.hat, 2)

[1] 1.3

.

Interpretation:

* The estimated odds of a made second free throw attempt are 0.7690 times as large for when the first throw is made than for when the first free throw is missed.
* The estimated odds of a made second free throw attempt are 1/0.7690 = 1.3 times as large for when the first free throw is missed than for when the first free throw is made.

In actual application, you would only present one of these interpretations. One could also phrase an interpretation as

* The estimated odds of a made second free throw attempt are 30% times larger for when the first free throw is missed than for when the first free throw is made.
* The estimated odds of a made second free throw are reduced by 23% (1 – 0.7690) for when the first free is made than for when the first three is missed.

Below are **INCORRECT** interpretations:

* “The estimated odds of a made second free throw attempt are 1.3 times as **likely** … ” is incorrect because “likely” means probabilities are being compared.
* Replacing “odds” with “probability” in any correct interpretation.
* “The estimated odds are 1.3 times higher …” is incorrect because 1.3 means 30% higher not 130%. Please see the relative risk interpretation in a past set of notes.

To find the confidence interval:

> alpha <- 0.05

> var.log.or <- 1/c.table[1,1] + 1/c.table[1,2] +

 1/c.table[2,1] + 1/c.table[2,2]

> OR.CI <- exp(log(OR.hat) + qnorm(p = c(alpha/2, 1 -alpha/2)) \* sqrt(var.log.or))

> round(OR.CI, 2)

[1] 0.29 2.07

> rev(round(1/OR.CI, 2))

[1] 0.48 3.49

The 95% confidence interval for OR is 0.29 < OR < 2.07. If the interval is inverted, the 95% confidence interval for 1/OR is 0.48 < 1/OR < 3.49.

The interpretation can be extended to:

With 95% confidence, the odds of a made second free throw attempt are between 0.48 and 3.49 times as large for when the first free throw is missed than for when the first free throw is made.

Notice that I do not include the word “estimated” here.

Because 1 is in the interval, there is not sufficient evidence to indicate that the first free throw result has an effect on the second free throw result.

Example: COVID-19 vaccine clinical trial (COVID19vaccine.R)

|  |  |  |  |
| --- | --- | --- | --- |
|  | Positive | Negative |  |
| Vaccine | 8 | 17,403 | 17,411 |
| Placebo | 162 | 17,349 | 17,511 |
|  | 170 | 34,752 | 34,922 |

> c.table <- array(data = c(8, 162, 17403, 17349), dim = c(2,2), dimnames = list(Treatment = c("Vaccine", "Placebo"), Outcome = c("Positive", "Negative")))

> c.table

 Outcome

Treatment Positive Negative

 Vaccine 8 17403

 Placebo 162 17349

> OR.hat <- c.table[1,1] \* c.table[2,2] / (c.table[2,1] \* c.table[1,2])

> round(OR.hat, 4)

[1] 0.0492

> round(1/OR.hat, 4)

[1] 20.313

> alpha <- 0.05

> var.log.OR <- 1/c.table[1,1] + 1/c.table[1,2] + 1/c.table[2,1] + 1/c.table[2,2]

> OR.CI <- exp(log(OR.hat) + qnorm(p = c(alpha/2, 1-alpha/2)) \* sqrt(var.log.OR))

> round(OR.CI, 4)

[1] 0.0242 0.1001

> rev(round(1/OR.CI, 4))

[1] 9.9851 41.3234

Interpretation:

* The estimated odds of infection are 0.0492 times as large for the vaccine group than for the placebo group.
* The estimated odds of being infection free are 20.31 times as large as for the vaccine group than for the placebo group. Notice that I used the “The estimated odds of a failure …” interpretation here.
* One could also phrase the interpretation as

The estimated odds of becoming infected are reduced by 95% when the vaccine is given instead of a placebo.

The 95% confidence interval is 0.02 < OR < 0.10. If the interval is inverted, the 95% confidence interval is 9.99 < 1/OR < 41.32. With 95% confidence, the odds of being infection free are between 9.99 and 41.32 times as large for the vaccine group than for the placebo group.

Questions/comments:

* Would you want to receive the vaccine?
* Why is  similar in value to ? One can show that



Examine what happens when π1 and π2 are small and/or close in value

Other confidence intervals have been proposed.

* A score interval can be calculated using the orscoreci() function of the PropCIs package.
* A LRT can be inverted as well to calculate the interval. This will be discussed further when logistic regression is presented.