REGRESSION IN R? | R FAQ

The interpretation of coefficients (https://stats.idre.ucla.edu/other/mult-pkg/faq/ologit/) in an ordinal logistic regression varies by the software you use. In this FAQ page, we will focus on the interpretation of the coefficients in R, but the results generalize to Stata (https://stats.idre.ucla.edu/stata/faq/ologit-coefficients/), SPSS and Mplus. For a detailed description of how to analyze your data using R, refer to R Data Analysis Examples Ordinal Logistic Regression. (https://stats.idre.ucla.edu/r/dae/ordinal-logistic-regression/)

Definitions

First let's establish some notation and review the concepts involved in ordinal logistic regression. Let Y be an ordinal outcome with J categories. Then $P(Y \leq j)$ is the cumulative probability of Y less than or equal to a specific category $j = 1, \cdots, J - 1$. Note that $P(Y \leq J) = 1$. The odds of being less than or equal a particular category can be defined as

$$\frac{P(Y \leq j)}{P(Y > j)}$$

for $j = 1, \cdots, J - 1$ since $P(Y > J) = 0$ and dividing by zero is undefined. Alternatively, you can write $P(Y > j) = 1 - P(Y \leq j)$. The **log odds** is also known as the **logit**, so that

$$log\frac{P(Y \leq j)}{P(Y > j)} = logit(P(Y \leq j)).$$

Ordinal Logistic Regression Model

The ordinal logistic regression model can be defined as

$$logit(P(Y \leq j)) = \beta_{j0} + \beta_{j1}x_1 + \cdots + \beta_{jp}x_p,$$

where $\beta_{j0}, \beta_{j1}, \cdots + \beta_{jp}$ are model coefficient parameters (i.e., intercepts and slopes) with p predictors for $j = 1, \cdots, J - 1$. Due to the **parallel lines assumption**, the intercepts are different for each category but the slopes are constant across categories, which simplifies the equation above to

$$logit(P(Y \leq j)) = \beta_{j0} + \beta_1 x_1 + \cdots + \beta_p x_p.$$

How R parameterizes the ordinal regression model

In Stata and R (**polr**) the ordinal logistic regression model is parameterized as

$$logit(P(Y \leq j)) = \beta_{j0} - \eta_1 x_1 - \cdots - \eta_p x_p$$

where $\eta_j = -\beta_j$

where $\eta_i = -\beta_i$.

Suppose we want to see whether a binary predictor parental education (**pared**) predicts an ordinal outcome of students who are *unlikely*, *somewhat likely* and *very likely* to apply to a college (**apply**).

Due to the parallel lines assumption, even though we have three categories, the coefficient of parental education (**pared**) stays the same across the two categories. The the two equations for **pared = 1** and **pared = 0** are

$$logit(P(Y \leq j | x_1 = 1) = \beta_{j0} - \eta_1$$
$$logit(P(Y \leq j | x_1 = 0) = \beta_{j0}$$

Then $logit(P(Y \leq j)|x_1 = 1) - logit(P(Y \leq j)|x_1 = 0) = -\eta_1$.

To run an ordinal logistic regression in R, first load the following libraries:

```
library(foreign)
library(MASS)
```

Now read in the data and run the analysis using **polr**:

```
dat <- read.dta("https://stats.idre.ucla.edu/stat/data/ologit.dta")
m <- polr(apply ~ pared, data = dat)
summary(m)
```

The shortened output looks like the following:

```
Coefficients:
      Value Std. Error t value
pared 1.127     0.2634    4.28

Intercepts:
                          Value    Std. Error t value
unlikely|somewhat likely    0.3768   0.1103     3.4152
somewhat likely|very likely 2.4519   0.1826    13.4302
```

The output shows that for students whose parents attended college, the log odds of being unlikely to apply to college (versus somewhat or very likely) is actually $-\hat{\eta}_1 = -1.13$ or 1.13 points *lower* than students whose parents did not attend college. Recall that $-\eta_i = \beta_i$ for $j = 1, 2$ only since $logit(P(Y \leq 3))$ is undefined. So the formulations for the first and second category becomes:

$$logit(P(Y \leq 1)) = 0.377{-}1.13x_1$$
$$logit(P(Y \leq 2)) = 2.45{-}1.13x_1$$

To see the connection between the parallel lines assumption and the **proportional odds** assumption, exponentiate both sides of the equations above and use the property that $log(b) - log(a) = log(b/a)$ to calculate the odds of **pared** for each level of **apply**.

$$\frac{P(Y \leq 1|x_1 = 1)}{P(Y > 1|x_1 = 1)} = exp(0.377)/exp(1.13)$$

$$\frac{P(Y \leq 1|x_1 = 0)}{P(Y > 1|x_1 = 0)} = exp(0.377)$$

$$\frac{P(Y \leq 2|x_1 = 1)}{P(Y > 2|x_1 = 1)} = exp(2.45)/exp(1.13)$$

$$\frac{P(Y \leq 2|x_1 = 0)}{P(Y > 2|x_1 = 0)} = exp(2.45)$$

From the odds of each level of pared, we can calculate the *odds ratio* of **pared** for each level of **apply**.

$$\frac{P(Y \leq 1|x_1 = 1)}{P(Y > 1|x_1 = 1)} \bigg/ \frac{P(Y \leq 1|x_1 = 0)}{P(Y > 1|x_1 = 0)} = 1/exp(1.13) = exp(-1.13)$$

$$\frac{P(Y \leq 2|x_1 = 1)}{P(Y > 2|x_1 = 1)} \bigg/ \frac{P(Y \leq 2|x_1 = 0)}{P(Y > 2|x_1 = 0)} = 1/exp(1.13) = exp(-1.13)$$

The proportional odds assumption ensures that the *odds ratios* across all $J - 1$ categories are the same. In our example, the proportional odds assumption means that the odds of being unlikely versus somewhat or very likely to apply $(j = 1)$ is the same as the odds of being unlikely and somewhat likely versus very likely to apply $(j = 2)$.

Interpreting the odds ratio

The proportional odds assumption is not simply that the odds are the same but that the *odds ratios* are the same across categories. These odds ratios can be derived by exponentiating the coefficients (in the log-odds metric), but the interpretation is a bit unexpected. Recall that the coefficient $-\eta_1$ represents a one unit change in the log odds of applying for students whose parents went to college versus parents who did not:

$$logit(P(Y \leq j|x_1 = 1)) - logit(P(Y \leq j|x_1 = 0)) = -\eta_1.$$

Since the exponent is the inverse function of the log, we can simply exponentiate both sides of this equation, and by using the property that $log(b) - log(a) = log(b/a)$,

$$\frac{P(Y \leq j|x_1 = 1)}{P(Y > j|x_1 = 1)} \bigg/ \frac{P(Y \leq j|x_1 = 0)}{P(Y > j|x_1 = 0)} = exp(-\eta_1).$$

For simplicity of notation and by the proportional odds assumption, let $\frac{P(Y \leq j|x_1=1)}{P(Y>j|x_1=1)} = p_1/(1 - p_1)$ and $\frac{P(Y \leq j|x_1=0)}{P(Y>j|x_1=0)} = p_0/(1 - p_0)$. Then the odds ratio is defined as

$$\frac{p_1/(1 - p_1)}{p_0/(1 - p_0)} = exp(-\eta_1).$$

However, as we will see in the output, this is *not* what we actually obtain from R!

R

To obtain the odds ratio in R, simply exponentiate the coefficient or log-odds of **pared**. The following code uses **cbind** to combine the odds ratio with its confidence interval. First store the confidence interval in object **ci**,

```
(ci <- confint(m))

    2.5 %     97.5 %
0.6131222 1.6478130
```

Then bind the transpose of the **ci** object with **coef(m)** and exponentiate the values,

```
exp(cbind(coef(m),t(ci)))

                   2.5 %    97.5 %
pared 3.087899 1.846187 5.195605
```

In our example, $exp(-1.127) = 0.324$, which means that students whose parents attended college have a 67.6% lower odds of being less likely to apply to college. However, this does not correspond to the odds ratio from the output! Let's see why.

Since $exp(-\eta_1) = \frac{1}{exp(\eta_1)}$,

$$exp(\eta_1) = \frac{p_0/(1 - p_0)}{p_1/(1 - p_1)}.$$

From the output, $\hat{\eta}_1 = 1.127$, which means the odds ratio $exp(\hat{\eta}_1) = 3.086$ is actually $\frac{p_0/(1-p_0)}{p_1/(1-p_1)}$.

This suggests that students whose parents did *not* go to college have higher odds of being *less* likely to apply.

Another way to look at the odds ratio

Double negation can be logically confusing. Suppose we wanted to interpret the odds of being *more* likely to apply to college. We can perform a slight manipulation of our original odds ratio:

$$
\begin{aligned}
exp(-\eta_1) &= \frac{p_1/(1-p_1)}{p_0/(1-p_0)} \\
&= \frac{p_1(1-p_0)}{p_0(1-p_1)} \\
&= \frac{(1-p_0)/p_0}{(1-p_1)/p_1} \\
&= \frac{P(Y > j | x = 0)/P(Y \le j | x = 0)}{P(Y > j | x = 1)/P(Y \le j | x = 1)}.
\end{aligned}
$$

Since $exp(-\eta_1) = \frac{1}{exp(\eta_1)}$,

$$
\frac{P(Y > j | x = 1)/P(Y \le j | x = 1)}{P(Y > j | x = 0)/P(Y \le j | x = 0)} = exp(\eta).
$$

Instead of interpreting the odds of being in the jth category or less, we can interpret the odds of being greater than the jth category by exponentiating η itself. In our example, $exp(\hat{\eta}) = exp(1.127) = 3.086$ means that students whose parents went to college have 3.086 times the odds of being very likely to apply (vs. somewhat or unlikely) compared to students whose parents did not go to college. The results here are consistent with our intuition because it removes double negatives. As a general rule, it is easier to interpret the odds ratios of $x_1 = 1$ vs. $x_1 = 0$ by simply exponentiating η itself rather than interpreting the odds ratios of $x_1 = 0$ vs. $x_1 = 1$ by exponentiating $-\eta$. However by doing so, we flip the interpretation of the outcome by placing $P(Y > j)$ in the numerator.

Verifying both interpretations of the odds ratio using predicted probabilities

To verify that indeed the odds ratio of 3.08 can be interpreted in two ways, let's derive them from the predicted probabilities in R.

After storing the **polr** object in object **m**, pass this object as well as a dataset with the levels of **pared** into the predict function. Specify **type="p"** for predicted probabilities.

```
newdat <- data.frame(pared=c(0,1))
```

```
                                         (pureu-c(o,±))
(phat <- predict(object = m, newdat, type="p"))

   unlikely somewhat likely very likely
1 0.5931114       0.3275856  0.07930294
2 0.3206801       0.4692269  0.21009300
```

Each row represents the first level ($x_1 = 0$) and second level ($x_1 = 1$) of **pared**, and each column represents $j = 1, 2, 3$ outcome **apply**.

Interpretation 1

The first interpretation is for students whose parents did *not* attend college, the odds of being *unlikely* versus somewhat or very likely (i.e., *less* likely) to apply is 3.08 times that of students whose parents did go to college.

To verify this interpretation, we arbitrarily calculate the odds ratio for the first level of **apply** which we know by the proportional odds assumption is equivalent to the odds ratio for the second level of **apply**. Since we are looking at **pared = 0** vs. **pared = 1** for $P(Y \le 1|x_1 = x)/P(Y > 1|x_1 = x)$ the respective probabilities are $p_0 = .593$ and $p_1 = .321$. Then

$$\frac{p_0/(1 - p_0)}{p_1/(1 - p_1)} = \frac{0.593/(1 - 0.593)}{0.321/(1 - 0.321)} = \frac{1.457}{0.473} = 3.08.$$

Interpretation 2

The second interpretation is for students whose parents *did* attend college, the odds of being *very* or *somewhat* likely versus unlikely (i.e., *more* likely) to apply is 3.08 times that of students whose parents did not go to college.

Here we are looking at **pared = 1** vs. **pared = 0** for $P(Y > 1|x_1 = x)/P(Y \le 1|x_1 = x)$. Then for the first level of **apply** $P(Y > 1|x_1 = 1) = 0.469 + 0.210 = 0.679$ and $P(Y \le 1|x_1 = 1) = 0.321$. Similarly, $P(Y > 1|x_1 = 0) = 0.328 + 0.079 = 0.407$ and $P(Y \le 1|x_1 = 0) = 0.593$. Taking the ratio of the two odds gives us the odds ratio,

$$\frac{P(Y > 1|x_1 = 1)/P(Y \le 1|x_1 = 1)}{P(Y > 1|x_1 = 0)/P(Y \le 1|x_1 = 0)} = \frac{0.679/0.321}{0.407/0.593} = \frac{2.115}{0.686} = 3.08.$$

The odds ratio for both interpretations matches the output of R.

Summary

In general, to obtain the odds ratio it is easier to exponentiate the coefficient itself rather than its negative because this is what is output directly from R (**polr**). The researcher must then decide which of the two interpretations to use:

1. For students whose parents did *not* attend college, the odds of being *less* likely to apply is 3.08 times that of students whose parents did go to college.
2. For students whose parents *did* attend college, the odds of being *more* likely to apply is 3.08 times that of students whose parents did not go to college.

The second interpretation is easier because it avoids double negation.

References

Bilder, C. R., & Loughin, T. M. (2014). *Analysis of categorical data with R*. Chapman and Hall/CRC.

Click here to report an error on this page or leave a comment

How to cite this page (https://stats.idre.ucla.edu/other/mult-pkg/faq/general/faq-how-do-i-cite-web-pages-and-programs-from-the-ucla-statistical-consulting-group/)