Logistic Regression

Interpreting larger logistic regression models

Jenine Harris
Brown School

Importing and cleaning the data

```
# import the libraries cleaned file
libraries <- read.csv("/Users/harrisj/Box/teaching/Teaching/Fall2020/dat

# change data types
library(package = "tidyverse")
libraries.cleaned <- libraries %>%
  mutate(age = as.numeric(age))
```

Interpreting the results of a larger logistic regression model

- The `summary()` and `odds.n.ends()` output both include values and significance statistics for the predictors.
- The `odds.n.ends()` output includes odds ratios, which are easier to interpret given the form of the logistic function used in computing the results.

Computing odds ratios

```
# run the odds.n.ends code again
lib.model <- glm(formula = uses.lib ~ age + sex + educ + parent + disabl
                 data = libraries.cleaned,
                 na.action = na.exclude,
                 family = binomial("logit"))
odds.n.ends::odds.n.ends(x = lib.model)
```

```
## $`Logistic regression model significance`
## Chi-squared        d.f.           p
##      94.736      12.000       0.000
##
## $`Contingency tables (model fit): percent predicted`
##                   Percent observed
## Percent predicted         1         0       Sum
##               1   0.2648914 0.1744919 0.4393833
##               0   0.2228451 0.3377715 0.5606167
##               Sum 0.4877365 0.5122635 1.0000000
##
## $`Contingency tables (model fit): frequency predicted`
##                   Number observed
## Number predicted    1    0  Sum
##               1    378  249  627
##               0    318  482  800
##               Sum  696  731 1427
##
## $`Predictor odds ratios and 95% CI`
##                                      OR     2.5 %    97.5 %
```

Odds ratio statistical significance

- The statistical significance of an odds ratio is determined by the range of its confidence interval.
- The confidence interval shows where the true or population value of the odds ratio likely lies.
- A confidence interval that includes 1 indicates that the true or population value of the relationship could be 1.
- The interpretation of an odds ratio of 1 is that the odds are 1 times higher or 1 times as high for a change in the predictor.
- This is essentially the same odds.
- When the confidence interval includes 1, the odds ratio could be 1 and this indicates it is not statistically significantly different from 1.

Interpreting odds ratios

- Odds ratios greater than one indicate an increase in the odds of the outcome with a one-unit increase in a numeric variable or in comparison with the reference group for a factor variable.
- If the odds ratio is greater than one **and** the confidence interval does not include one, the odds ratio suggests a statistically significant increase in the odds of the outcome.
- Factor variables with two categories are interpreted with respect to the *reference group* or the group not shown in the output.
- Since the factor variables can only change by going from one group to the other, instead of the odds ratio being for a one-unit change in the predictor, it is the change in the odds of the outcome when moving from the *reference group* to the other group.
- For the sex variable, male is the group shown in the output, so interpret the odds raio for sex as, "Males have 51.1% lower odds of library use compared to females."
    - The 51.1% came from subtracting the odds ratio of .489 from 1, 1 - .489 = .511, and multiplying by 100.
- For factor variables with more than two categories, each odds ratio is interpreted with respect to the *reference group* for that variable.
    - For `educ`, the group not shown is the `< HS` group indicating that `< HS` is the reference group.
    - The odds ratio for the `Four-year degree or more` group is 1.90, so the interpretation would be that individuals with a four-year degree or more have 1.90 times higher odds of

Significant odds ratio interpretation

- A list of the odds ratios where the confidence interval did not include 1:
    - age (OR = .9899; 95% CI: .9835 - 9963)
    - male sex (OR = .49; 95% CI: .39 - .61)
    - Four year degree or more (OR = 1.90; 95% CI: 1.26 - 2.90)
    - Non-Hispanic Black (OR = 1.55; 95% CI: 1.002 - 2.417)
- There were two significant predictors with odds ratios greater than one.
    - People with a four-year degree or more education had 1.90 times higher odds of library use compared to people with less than a high school education (OR = 1.90; 95% CI: 1.26 - 2.90).
    - People who were non-Hispanic Black had 1.55 times higher odds of library use compared to people who were Hispanic (OR = 1.55; 95% CI: 1.002 - 2.417).
- The age variable and male sex both show significant odds ratios lower than 1.
    - For age, the odds of library use are 1% lower for every one year increase in a person's age (OR = .9899; 95% CI: .9835 - 9963).
    - For male sex, the reference group is female and the odds ratio is .49.

Interpreting non-significant odds ratios

- Some odds ratios greater than one were non-significant.
    - For example, the odds ratios for urban and suburban are greater than one, but both of these odds ratios have confidence intervals that include 1.
    - For suburban, the confidence interval is .90 to 1.57 (see odds ratio table in `odds.n.ends()` output).
    - For urban, the confidence interval is .93 to 1.63.
- When the confidence interval includes 1, it is possible that the true value of the odds ratio is one, so the values would be reported without the interpretation of higher odds:
    - The odds of library use were not statistically significantly different for urban residents compared to rural residents (OR = 1.23; 95% CI: .93 - 1.63)*.
- The low-SES category had a non-significant odds ratio of .93 (95% CI: .57 - 1.52).
- The interpretation would be:
    - The odds of library use are not statistically significantly different for those with low SES compared to those in the reference group of high SES (OR = .93; 95% CI: .57 - 1.52).

Using NHST to organize odds ratio interpretation

Sometimes the NHST process can be used to organize the reporting of odds ratios.

NHST Step 1: Write the null and alternate hypotheses

Using sex, for example:

H0: There is no relationship between sex and library use.

HA: There is a relationship between sex and library use.

NHST Step 2: Compute the test statistic

The odds ratio is the test statistic, OR = .49.

NHST Step 3: Compute the probability for the test statistic

The confidence interval shows the probable range of the true or population value of an odds ratio, 95% CI: .39 - .61.

NHST Steps 4 & 5: Interpret the probability and write a conclusion

The odds of library use are 51% lower for males than for females (OR = .49; 95% CI: .39 - .61).

Compute and interpret model fit

- The model correctly predicted 378 of the 765 who use the library and 482 of the 788 who do not use the library.
- Overall it was correct for $\frac{860}{1427}$ of the observations or 60.3% of the time (Count R^2 = .603).
- It was better at classifying those who do not use the library (specificity = 65.9%) than those who use the library (sensitivity = 54.3%).