Logistic Regression

Larger logistic model

Jenine Harris
Brown School

Importing and cleaning the data

```r
# 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))
```

A larger logistic regression model with categorical and continuous predictors

- Estimate and interpret the model with all the predictors:
  - $p(uses.lib) = \frac{1}{1+e^{-(b_0+b_1 \cdot age+b_2 \cdot sex+b_3 \cdot educ+b_4 \cdot parent+b_5 \cdot disabled+b_6 \cdot rurality+b_7 \cdot raceth+b_8 ses)}}$

```
# estimate the library use model and print results
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
```

NHST Step 1: Write the null and alternate hypotheses

- Try writing the hypotheses in a more specific way.
- The null and alternate used for the first model would be fine here, but it is also nice to explicitly state what is being tested:
  - H0: A model including age, sex, education, parent status, disability status, rurality, ses, and race-ethnicity is no better than the baseline at explaining library use.
  - HA: A model including age, sex, education, parent status, disability status, rurality, ses, and race-ethnicity is better than the baseline at explaining library use.

NHST Step 2: Compute the test statistic

The chi-squared test statistic of 94.736 was computed by the `odds.n.ends()` function.

NHST Step 3: Compute the probability for the test statistic (p-value)

- The `odds.n.ends()` output also shows the model chi-squared of 94.736 with the corresponding degrees of freedom of 12 and very small p-value.
- Visualizing a chi-squared distribution with 12 degrees of freedom makes it clear why the p-value is so small.
- The probability that the chi-squared would be 94.736 if the full model were no better than the null model is shown as the area under the curve to the right of 94.736.

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

- With a very tiny probability of getting a chi-squared of 94.736 or larger if the null were true, the null hypothesis is rejected.

- Interpretation: A logistic regression model including age, sex, education, parental status, disability status, ses, race-ethnicity, and rurality was statistically significantly better than the baseline probability at predicting library use [χ^2 (12) = 94.736; p < .001].