Logistic Regression

Predicting probabilities

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))
```

Larger model

```
# 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"))
```

Using the model to predict probabilities

- Logistic regression models are not only useful for examining relationships between predictors and binary outcomes, they can also be used to predict probabilities for hypothetical or new cases that are not in the data frame.
- Try predicting the probability of library use for:
  - a 35 years old, male, with a four-year degree, not a parent, has low socioeconomic status, is non-Hispanic white, lives in a rural area, not disabled.
  - 65 and 68 year olds, female and male, non-Hispanic White, medium SES, with four-year degrees, are parents, live in a rural area, not disabled.

```
# make a new data frame containing the observations of interest
newdata <- data.frame(age = c(35, 65, 68),
                      sex = c("male", "female", "male"),
                      educ = c("Four-year degree or more", "Four-year de
                      disabled = c("no", "no", "no"),
                      parent = c("not parent", "parent", "parent"),
                      rurality = c("rural", "rural", "rural"),
                      raceth = c("Non-Hispanic White", "Non-Hispanic Whi
                      ses = c("low", "medium", "medium"))
```

Interpreting predicted values

```
# use the new data frame to predict
predictions <- predict(lib.model, newdata, type = "response")
predictions
```

```
##         1         2         3
## 0.5135547 0.6466977 0.4648331
```

- The model predicts the 35 year old male has a 51.4% probability of library use, the older female has a 64.7% probability of library use, and the older male has a 46.5% probability of library use.
- The males have very similar probabilities, while the female is much more likely to be a library user.
- The differences in the data between the males are age, parent status, and socioeconomic status.
- The only difference between the older adults is between female and male.
- Looking back at the odds ratios, the odds of library use are 51% lower for males compared to females, so that must be what made the difference from 64.7% for the older female and 46.5% for the older male!