Week Twelve

Last Week

- Lab: Count Regression for Bike Data
- Lecture: Multicategory Regression Models (Theory)

This Week: Multicategory Regression Models

- Today: Activity
- Thursday: Lab

Next Week: Generalized Linear Mixed Models

- Tuesday: No Class - Veteran's Day
- Thursday: Activity / Lecture

Multicategory Logit Models

Recall Logistic regression for binary data

$$Y \sim Multinomial(n, \pi)$$

$$\log\left(\frac{\pi_i}{\pi_J}\right) = \alpha_i + \beta_j x$$

- We can also directly estimate $\pi_j(x)$ for any set of covariates.

$$\pi_j = \frac{\exp(\alpha_j + \beta_j x)}{\sum_h \exp(\alpha_h + \beta_h x)},$$

where α_h and $\beta_h = 0$ for the reference category.

Data Analysis

The data set contains variables on 200 high school senior students. This dataset was collected by the National Opinion Research Center with funding from the National Center for Education Statistics.

We will treat `prog` as the outcome variable, where academic is a college preparatory program, general is a basic high school program, and vocation/technical focus on vocational paths.

```
library(foreign)
hsb <- read.dta("https://stats.idre.ucla.edu/stat/data/hsbdemo.dta")
```

1. Let's initially consider the relationship between program type and social economic status. Create a contingency table, run a χ^2 test, and summarize the results.

2. Now let's consider a multicategory GLM for program, using SES. For a ML approach you can use `nnet::multinom`. You can also use `brms::brm()` with `family = categorical(link = "logit")` for a Bayesian approach, but it might require compiling a stan program. Fit the model, interpret the coefficients, and summarize the results graphically.

3. Now let's consider a multicategory GLM for program, using a continuous variable: math. Fit the model, interpret the coefficients, and summarize the results graphically.

Ordinal Models

We can also use the `hsb` dataset to fit ordinal regression models.

1. Create a figure to display ses vs. math scores.

2. Fit an ordinal regression model. For a ML approach you can use `MASS::polr`. You can also use `brms::brm()` with `family = cumulative(link = "logit")` for a Bayesian approach, but it might require compiling a stan program.