Week Eleven

Last Week

- Exam Recap
- Probability Distributions for Count Data
- Count Regression

This Week: GLMs for Count Data & Multicategory data

- Bike Data

Next Week: Multicategory Regression

Multicategory Logit Models

Recall Logistic regression for binary data

$$Y \sim Binomial(n, \pi)$$
$$logit(\pi) = \log\left(\frac{\pi}{1-\pi}\right) = X\beta$$

What is an example of multicategory data? Can we come up with both nominal and ordinal categories?

- What is a probability distribution that is appropriate for this setting? Multinomial(n, π)

- What are some predictors for this variable?

- Let's think about a GLM

1. $Y \sim Multinomial(n, \pi)$
2. Linear combination (functional form)
3. Link function

Baseline-Category Logits

Let's arbitrarily set category J to be the baseline. Then we can define a baseline category logit as

$$\log\left(\frac{\pi_i}{\pi_J}\right)$$

this can be formulated in a model as

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

This approach requires J - 1 equations; however, these equations can be used for any pair of categories

$$\log\left(\frac{\pi_a}{\pi_b}\right) = \log\left(\frac{\pi_a/\pi_J}{\pi_b/\pi_J}\right) = \log\left(\frac{\pi_a}{\pi_J}\right) - \log\left(\frac{\pi_b}{\pi_J}\right) = (\alpha_a + \beta_a x) - (\alpha_b + \beta_b x)$$

from which we can find coefficients (α and β for this categories a and b)

- this can give us odds ratios between to categories

- 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.

- Add curve here

Ordinal Data

Now let's consider ordinal responses.

- With this scenario, we are often interested in cumulative probability, the probability that Y falls at or below a particular point

$$P(Y < J) = \pi_1 + ... + \pi_J$$

$$logit[P(Y < J)] = \alpha_j + \beta x$$

note that β does not have a subscript here (fairly strong assumption).

add cumulative logit figure here

add category probability curve too

Latent variable interpretation

A common way to think about (and formulate) ordinal models is with a latent variable approach.

Let Y be some continuous latent variable that has cut points that map to the ordinal values.

A common approach for this model is to assume the latent variable has a standard normal distribution, this results in what is referred to as a probit model.