Ordered logistic regression

Violetta Korsunova
Dept. of Sociology, HSE in St. Petersburg

Previously

We have discussed how to regress binary and multinomial outcomes

The main idea is to compare the probabilities of the given categories

Ordinal response

Sometimes the response variable is measured by ***ordinal scale (Likert scale, attitudes, ordered levels)***

Here the outcomes are still ***categorical*** but the distribution might be considered ***continuous***

Ordinal response: what to do?

Use linear regression:

- if the N of categories >= 5
- if the distribution or log of distribution is similar to normal

Dichotomize the outcome variable and apply ***binary regression*** with probit link function:

- if the division really make sense
- if the numbers of observations in groups will be meaningful and comparable

Unfortunately, it’s not always the case

What to do if the previous suggestions fail

Fit an ordered logistic regression

Here it is assumed that there is a ***latent continuous variable*** behind our ***ordinal response***

To look at this latent variable we need cumulative probability:

the probability of ***all higher categories*** comparing to ***all lower probabilities***

Now take a deep breath and calm down

Some math

Cumulative probability: $\gamma_{ij} = P(Y_i \leq j) = \pi_{i1} + \ldots + \pi_{ij}$

Cumulative logit: $logit(\gamma_{ij}) = logit(P(Y_i \leq j)) = log(\frac{P(Y_i \leq j)}{1 - P(Y_i \leq j)})$

Ordered logistic model:

$$logit(\gamma_{ij}) = \theta_j - x_i^T \beta$$

where $\boldsymbol{x_i}$ is the vector of independent variables, for the *i*th observation, $\boldsymbol{\beta}$ is the vector of regression coefficients, and $\boldsymbol{\theta_j}$ is (J-1) length vector of intercepts, with **j** = number of categories of ***y***.

Example

Let the proportions of students who answered *likely, somewhat likely*, and *unlikely* are respectively π_1, π_2, and π_3

$$logit(\gamma_{Likely}) = logit(\pi_1) = log(\frac{\pi_1}{\pi_2 + \pi_3}) = \theta_{Likely} - x_i^T \beta$$

$$\begin{aligned} logit(\gamma_{Likely+Smw.Likely}) &= logit(\pi_1 + \pi_2) \\ &= log(\tfrac{\pi_1+\pi_2}{\pi_3}) \\ &= \theta_{Likely+Smw.Likely} - x_i^T \beta \end{aligned}$$