

ORDINAL REGRESSION.

1. Motivation.

Many variables of interest are ordinal, *that is, you can rank the values, but the real distance between categories is unknown.* For example a Likert “scale”;

- *Strongly disagree*
- *Disagree*
- *Neutral*
- *Agree*
- *Strongly agree*

2. The Ordinal Logit Model.

2.1. Recall - Binary logistic regression.

- A binary logistic regression model- you estimate a set of regression coefficients that predict the probability of the outcome of interest.

$$\ln\left(\frac{\text{prob}(\text{event})}{1 - \text{prob}(\text{event})}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k$$

The quantity to the left of the equal sign is called a logit *the log of the odds that an event occurs.* The odds that an event occurs is the ratio of the number of people who experience the event to the number of people who do not. The coefficients in the logistic regression model tell you how much the logit changes based on the values of the predictor variables.

When you have more than two events, you can extend the binary logistic regression model, however for ordinal categorical variables, the drawback of the multinomial regression model is that the ordering of the categories is ignored.

2.2. Ordinal logistic regression. Ordinal logistic regression or (*ordinal regression*) is used to predict an ordinal dependent variable given one or more independent variables. For example we could use ordinal logistic regression to predict the belief that “people who study Statistics are weird”, this is the ordinal **dependent** variable measure on the 5-point Likert scale given above, based on some **independent** variables such as *sex of respondent*, or *respondents degree studied* etc.

Ordinal regression will be enable us to determine which of our independent variables (if any) have a statistically significant effect on our dependent variable.

For categorical independent variables (e.g., “*degree studied*”), we can interpret the odds that one “group” (e.g., *Art students*) have a higher or lower score on our dependent variable (*a higher value may be stating that they “Strongly agree” that “people who study Statistics are weird” rather than stating that they “Disagree”*). For continuous independent variables we are able to interpret how a single unit increase or decrease in that variable, is associated with the odds of our dependent variable having a higher or lower value We can also determine how well our ordinal regression model predicts the dependent variable.

2.2.1. Assumptions.

- The **dependent** variable is measured on an ordinal level.
- One or more of the **independent** variables are either **continious**, **categorical** or **ordinal**.
- No Multi-collinearity - i.e. when two or more independent variables are highly correlated with each other.
- **Proportional Odds** - i.e. that each independent variable has an identical effect at each cumulative split of the ordinal dependent variable.

These assumptions should be tested in order as if a violation to the assumption is not correctable, you will no longer be able to use ordinal regression. If these assumptions are violated the results you get when running ordinal regression may not be valid.