Chapter 13: Generalised Linear Models

Exercises

Brian Fogarty

Exercise 1 - Logit Model

a. Using the `VF England.csv` dataset, create new versions of `brexit_vote`, `vfproblem`, and `pid`. For `brexit_vote`, flip the values so that "2. Leave" is the higher category. For `vfproblem`, correctly order the labels/values and then convert it to a numeric variable. For `pid`, re-order the labels/values so "Other" is the first category.

b. Run a logit regression with the new version of `brexit_vote` as the outcome variable and `age` and the new versions of `vfproblem` and `pid` as predictors. Briefly discuss (one sentence) what predictors are statistically significant.

c. Plot the logit regression coefficients using the `ggcoef_model()` function from the `GGally` package.

d. Provide odds ratio as a percentage interpretations for the statistically significant predictors.

e. Use the `ggpredict()` and `ggplot()` functions to create a predicted probability plot for the new version of `vfproblem` (i.e., the new version of `vfproblem` will be the x-axis variable). Briefly discuss the plot (2-3 sentences).

f. Use the `ggpredict()` and `ggplot()` functions to create a predicted probability plot for the new version of `vfproblem` (i.e., the new version of `vfproblem` will be the x-axis variable) and the new version of `pid` (i.e., the lines will be for the three categories of the new version `pid`). Briefly discuss the plot (2-3 sentences).

Exercise 2 - Ordered Logit Model

a. Create a new version of `vfproblem` where the 3 disagree categories are collapsed to 1 category and the 3 agree categories are collapsed to 1 category. Then order the categories as the following: "Disagree", "Neither agree nor disagree", "Agree".

b. Run an ordered logit regression with the new version of `vfproblem` (from '2.a.') as the outcome variable and `age` and the new versions of `brexit_vote` and `pid` (from '1.a.') as predictors. Briefly discuss (one sentence) what predictors are statistically significant.

c. Plot the ordered logit regression coefficients using the `ggcoef_model()` function from the `GGally` package.

d. Test the parallel regression assumption of the ordered logit model. Discuss the test results.

e. Provide odds ratio as a percentage interpretations for the statistically significant predictors.

f. Use the `ggpredict()` and `ggplot()` functions to create a predicted probability plot with `age` on the x-axis and lines for each value of the new version of `vfproblem` (from '2.a.'). Briefly discuss the plot (2-3 sentences).

Exercise 3 - Multinomial Logit Model

a. Run a multinomial logit regression with the new version of `pid` as the outcome variable and `age` and the new versions of `brexit_vote` and `vfproblem` (from '1.a.') as predictors. Briefly discuss (one sentence) what predictors are statistically significant.

b. Plot the multinomial logit regression coefficients using the `ggcoef_multinom()` function from the `GGally` package.

c. Provide odds ratio as a percentage interpretations for the statistically significant comparisons for each predictor (i.e., each predictor will have a coefficient for Conservative vs. Other and a coefficient for UKIP/Brexit vs. Other).

d. Use the `ggpredict()` and `ggplot()` functions to create a predicted probability plot for `age` (i.e., `age` will be the x-axis variable). Briefly discuss the plot (2-3 sentences).

e. Use the `ggpredict()` and `ggplot()` functions to create a faceted predicted probability plot for `age` and the new version of `brexit_vote`. This means each category of `pid` will be its own subplot with `age` as the x-axis variable and the lines will be for the two categories of the new version of `brexit_vote`. Briefly discuss each subplot (2-3 sentences per subplot).

For the answers see **Chapter 13 - Answers to Exercises**.