15.9 Practical 15

Dataset required: `nhanesglm.dta`

Introduction

In this practical we will use multinomial logistic regression techniques to analyse the NHANES alcohol data introduced in the lecture. The outcome variable is a categorical measure of daily alcohol consumption, in three levels.

The dataset is called `nhanesglm.dta` and contains six variables as below.

Variable	Description
`gender`	1 = male, 2 = female
`ageyrs`	Age in (whole number of) years
`bmi`	Body Mass Index (kg/m^2)
`sbp`	Systolic blood pressure (mmHg)
`ALQ130`	Average number of alcoholic drinks consumed on days when alcohol is consumed
`alccat`	Categorised `ALQ130`: 1 = 1 drink per day 2 = 2-5 drinks per day 3 = 6+ drinks per day

Aims

- Compare the estimates from multinomial logistic regression to those from standard logistic regression.
- Understand how to calculate estimated probabilities from multinomial regression estimates.
- Learn how to use the margins command in Stata to obtain the estimated probabilities.

Analysis

1 Tabulate the categorised alcohol consumption variable (`alccat`) by gender (as in the lecture slides).

2 Fit a (standard) logistic regression model with categorised alcohol consumption as the dependent variable and gender as the only covariate, **omitting the highest drinking category**.

Do this by first using the following code to create a binary outcome that is 1 when `alccat` is equal to 1, 0 when `alccat` is equal to 2 and missing when `alccat` is equal to 3 and then fitting the model.

```
gen alc12 = alccat if alccat < 2.5

replace alc12 = 0 if alc12 == 2

logit alc12 i.gender, nolog
```

Also fit a second analogous logistic regression model, this time omitting the lowest drinking category.

3 Fit a multinomial logistic model to the categorised alcohol consumption variable with gender as the only covariate.

Discuss: Compare the parameter estimates from the logistic regression models with those from the multinomial logistic regression model and check that you understand the interpretation of each of the estimated coefficients for the multinomial logistic regression model.

4 Fit a multinomial logistic model to the categorised alcohol consumption variable with gender and age as covariates. Give an exact interpretation of each estimated coefficient.

Discuss: Compare your interpretations with one or more of your colleagues (in your Breakout Room if online).

5 Use the estimated coefficients from the model to predict the probabilities that a 50 year old female is in each of the three alcohol consumption categories.

Use the `predict` command to generate the fitted probabilities of being in each alcohol consumption category for each participant. You will need to name three variables to be the predicted probabilities.

```
predict pr1 pr2 pr3
```

Check that your predictions for a 50 year old female match those predicted here.

Plot line graphs of these fitted probabilities against age, separately in males and females. Hint: in order to see the lines sort the data by age before plotting the probabilities.

6 The `margins` command can also be used to display fitted probabilities. For example, use the following commands to produce and display the fitted probabilities for the 1 drink per day category by gender, at ages 20, 30, 40, ... and 80 after refitting the multinomial logistic regression model with age and gender as covariates.

```
margins, at(ageyrs=(20(10)80)) over(gender) predict(outcome(1))

marginsplot
```

Discuss: With one or more colleagues discuss the patterns exhibited by the various plots. Write a few sentences that describe how the distribution of the categorised alcohol consumption variable varies by age and gender. Post these in the zoom chat if online.

7 Add an interaction between age and gender to the model in part 4 and use an appropriate test to see whether there is evidence that this improves the fit of the model.

8 (optional) Calculate fitted probabilities for the model in part 7 and use plots analogous to those used earlier to explore whether the predictions are materially altered by the inclusion of the interaction.