Soc 73994, Homework #5
Ordinal and Multinomial Models

Richard Williams, University of Notre Dame, https://www3.nd.edu/~rwilliam/
Last revised October 3, 2024

All answers should be typed and submitted via Canvas. Be sure your response includes your name, the date, and a clear title, e.g. Homework # 5. If there is a huge amount of output for any analyses you run yourself, you may want to be selective in what you copy and paste into your assignment (but make sure you include enough so it is clear what commands you executed, e.g. you might show all the commands but only parts of the output).

This assignment has multiple parts. The first three questions ask you about basic principles of ologit, mlogit, and interval regression models, based on code and data that are provided to you. Then, in Question 4, using a data set of your choice, you will conduct and interpret your own ordinal logit analyses. Remember, if the data set you want to use does not have a good ordinal dependent variable, you could always recode a continuous dependent variable so it was ordinal. Finally, if you wish, in Questions 5 & 6 you can run mlogit or interval regression models on your own dataset.

NOTE: If you haven't read it before, the Margins05 handout (Adjusted Predictions & Marginal Effects for Multiple Outcome Models & Commands) at https://www3.nd.edu/~rwilliam/xsoc73994/Margins05.pdf may be very helpful.

1. This problem is adapted from *Ordinal Regression Models* by Williams & Quiroz (2019). See https://methods.sagepub.com/Foundations/ordinal-regression-models. They used data from Great Britain. In this problem you will use data from France instead.

The European Social Survey (ESS) is a cross-national study that has been conducted every two years across Europe since 2001. For this example we use the 2012 ESS survey for France (ESS Round 6: European Social Survey Round 6 Data (2012)). The study has 1,968 respondents, of which 1,939 (98.5%) had complete data for the variables used in this analysis. Although cases have unequal probabilities of selection, weighting had little effect on the results so to simplify the presentation they were not used.

Respondents were asked the extent to which they agreed or disagreed with the following statement: "Gay men and lesbians should be free to live their own life as they wish." The possible responses were 1 = Strongly Disagree, 2 = Disagree, 3 = Neither Agree nor Disagree, 4 = Agree, and 5 = Strongly Agree. We use this as our response variable, HMSFREE. Thus, the higher the reported value, the more supportive the person is of gay and lesbian rights.

The explanatory variables are the responses to the following questions. In some cases the original coding has been modified or reversed to make interpretation easier.

- AGEDECADE - Age of respondent (in decades, e.g. a value of 3.4 means 34 years old)
- FEMALE - Gender of respondent (1 = Female, 0 = Male)

- LIFEWORSE - "For most people in this country life is getting worse rather than better" (coded 1 = Strongly Disagree to 5 = Strongly Agree)
- HINCFEL - "Which of the descriptions on this card comes closest to how you feel about your household's income nowadays?" (1 = Living comfortably on present income, 2 = Coping on present income, 3 = Finding it difficult on present income, 4 = Finding it very difficult on present income)
- FEELECON - "On the whole how (dis)satisfied are you with the present state of the economy in this country?" (11 point scale where 0 = extremely satisfied and 10 = extremely dissatisfied).

Age and gender are demographic variables. We might reasonably expect that age is related to attitudes about gays and lesbians because of changing attitudes across time. Gender might also be related if women tend to be more tolerant than men of different lifestyles. HINCFEL and LIFEWORSE measure different aspects of satisfaction with one's life. The theoretical argument for including them as explanatory variables is less clear. Dissatisfaction with one's life could lead to less tolerance towards others, but not necessarily. FEELECON measures satisfaction with the entire economy but not necessarily satisfaction with one's own life.

```
* Adapted from Ordinal regression models by Williams & Quiroz
* https://methods.sagepub.com/Foundations/ordinal-regression-models
* Extended ologit using European Social Survey France 2012

* Open the data, do descriptive stats, drop cases with missing
use https://www3.nd.edu/~rwilliam/statafiles/ESS2012FR.dta, clear
des, short
keep if !missing(hmsfree, lifeworse, hincfel, feelecon, agedecade, female)
des, short
fre

*** Problem 1 Ologit
* Constrained model -- demographic vars only
ologit hmsfree agedecade i.female , cformat(%9.3f) nolog
est store constrained
* Unconstrained model -- Attitudes + demographic vars
ologit hmsfree lifeworse hincfel feelecon  agedecade i.female, cformat(%9.3f)  nolog
est store unconstrained
* Test to see if attitude vars improve the model
lrtest constrained unconstrained, stats
* Now get margins, adjusted predictions. margins & mtable
* tell you pretty much the same thing but the output from mtable is
* much easier to read.
margins female
mtable, at(female=(0 1)) dec(4)
margins, dydx(female)
mtable, dydx(female) dec(4)
* Test model assumptions
brant, detail
```

Based on the results from the above code, answer the following questions (be sure to explain how the results justify your conclusions):

a. Explain why the unconstrained model (i.e. the model that includes all the independent variables) should be preferred. Justify your decision using LR, BIC, and AIC measures. Then, using only the output from the unconstrained model, interpret the results. Indicate whether variables have statistically significant effects and what the sign of their effects is.

b. Use the margins and/or mtable commands to interpret the effect that gender has on feelings toward gays and lesbian.
c. Is the parallel lines/ proportional odds assumption violated? If so, which variables appear to be problematic?
d. (Optional) Do at least one graph of results. (You will have to write the code yourself.) Make sure the graph is legible – since there are multiple outcomes the graphs can become hard to read if you try to include too much information. Discuss the additional insights that are gained thanks to your commands, e.g. how do they make the effects of gender, race, or income clearer?

2. Now run this code. Explain what the results from the two mlogtest commands tell you.

```
***  Problem 2 Mlogit
mlogit hmsfree lifeworse hincfel feelecon  agedecade i.female, cformat(%9.3f)  nolog b(1)
mlogtest, lr
mlogtest, lrcomb
```

3. Now run this code. Explain why dlow and dhigh were computed. Based on the results, do you think interval regression would be a good choice for this analysis?

```
* Shuttle using Interval Regression
use https://www3.nd.edu/~rwilliam/statafiles/shuttle2.dta, clear
recode distress (1 = .)(2=1)(3=3), gen(dlow)
recode distress (1=0)(2=2)(3=.), gen(dhigh)
intreg dlow dhigh date temp, cformat(%9.4f) nolog
oprobit distress date temp, nolog cformat(%9.4f)
```

4. With your chosen data set, use the ologit command to estimate an ordinal model. You should have at least least three or four independent variables in the model, at least one of which is categorical (e.g. gender). Use the svy: prefix if appropriate. Do the following:

a. Looking only at the output from ologit, interpret the results. Indicate whether variables have statistically significant effects and what the sign of their effects is.

b. Use commands like we have discussed before – e.g. `margins`, `mtable`, `mchange` – to make your results more interpretable. For example, you might estimate the marginal effects or adjusted predictions for gender. Or, you could compute predicted values for prototypical cases, e.g. young woman, young man, old woman, old man.

c. Test the parallel lines/ proportional odds assumption. Remember that if you are using svy your options are more limited. If not using svy be sure to use the brant command with the detail option, and indicate what variables, if any, may be problematic.

d. (Optional) Do at least one graph of results. Make sure the graph is legible – since there are multiple outcomes the graphs can become hard to read if you try to include too much information. Discuss the additional insights that are gained thanks to your commands, e.g. how do they make the effects of gender, race, or income clearer?

5. [Optional – Do this either because you want the experience or because you think it is the most appropriate method for your problem.] Now use the mlogit command to estimate a model. You can use the same variables you used for problem 4 or you can use different ones. If some/many commands do not work with the svy prefix, you might have to not use it. (Let me know which ones do not work, as there are many I have not tried.)

a. If using the same dependent and independent variables as before, briefly comment on how results differ from ologit, especially if the proportional odds assumption was violated. If using a different dependent variable, again do analyses to make your results more interpretable. As before, discuss the additional insights that are gained thanks to your commands, e.g. how do they make the effects of gender, race, or income clearer?

b. Use the following commands from spost13. Explain what the purpose of each command is and what the results tell you.

1. mlogtest, with the lr or wald option
2. mlogtest, with the lrcomb or combine option
3. (Optional) Any other spost13 commands you think are useful.

6. [Optional – Do this either because you want the experience or because you think it is the most appropriate method for your problem.] If you have an appropriate dependent variable, use the `intreg` command to estimate a model for it. Also estimate an `oprobit` model to see if your variable seems to work ok with intreg. Briefly interpret the results. If you recoded a continuous variable to be ordinal in question 4, you could use that same variable here.