Soc 63993, Homework #6:
Interaction effects and group comparisons

Richard Williams, University of Notre Dame, https://www3.nd.edu/~rwilliam/
Last revised February 20, 2015

Problem 1. Download *gender.dta* from the course web page. This is the hypothetical data on gender, income, education, and job experience that you used in homework 5. You will once again examine group differences in the parameters of this model, this time using dummy variables and interaction effects.

1. You are interested in the effects of education and job experience on income, and whether and if there are any differences in the models for men and women. Estimate the following three models using dummy variables and interaction effects (use Stata's factor variable notation to do so):

> a. There are no differences by gender – the models are identical for men and women.
>
> b. The intercepts differ by gender, but the effects of education and job experience are the same for both men and women.
>
> c. The intercepts and slopes differ by gender, i.e. all model parameters are free to differ by gender.

2. Indicate which model you think is best, and why. Briefly discuss the substantive interpretation of what you think is the "best" model. Include in your discussion any insights that the model provides concerning gender differences. To help you with the discussion, run the following commands after your preferred model. Note that, in each case, the variable NOT being graphed is set to zero – which means that the (nonexistent in the data) point where income = 0 and jobexp = 0 is included in each graph.

```
quietly margins female, at(educ=(0(1)20) jobexp=0)
marginsplot, noci ytitle("Predicted Income") ylabel(#10) scheme(sj) name(educ)
quietly margins female, at(jobexp=(0(1)20) educ=0)
marginsplot, noci ytitle("Predicted Income") ylabel(#10) scheme(sj) name(jobexp)
```

3. In the models above, the effect of Female changes from negative to positive once interaction terms are added to the model. Explain why this should not concern you. In particular, explain how the interpretation of the coefficient for Female changes once interaction terms are added to the model.

4. Center the continuous variables and rerun the three models. How do your results differ from before? Explain how centering makes it easier to interpret the results.

Problem 2. Get *jgqes2.dta* and *jgqes2.do* from the course web page. Selected variables from The Quality of Employment Survey are contained in jgqes2.dta. Run jgqes2.do and answer the following questions:

1. What is the mean of each group on the dependent variable (jsat = Job Satisfaction)? Is the mean difference between groups statistically significant?

2. Are there any statistically significant differences in the model parameters between groups?

3. If the answer to 2 is yes, are these differences limited to differences in the intercepts? Or are there differences in the effects of the IVs across groups (i.e. are there statistically significant interaction effects? Or is it just the coefficient of the dummy variable for group membership that is statistically significant?)

4. Briefly discuss the substantive interpretation of what you think is the "best" model for the data set. Include in your discussion any insights that the model provides concerning group differences.

5. Examine the compositional differences (i.e. mean differences) between groups on the independent variables. Discuss how these differences help lead to mean differences on the dependent variable.

Following is a copy of jgqes2.do. Note that it computes the interaction terms so it can use the `nestreg` command, but it wouldn't be too hard to rewrite it to use factor variables.

```
version 12.1
* Problem 2. Quality of Employment survey.
use https://www3.nd.edu/~rwilliam/statafiles/jqges2.dta, clear
* Tidy up the data for our purposes
keep jsat prof mang tenure firmsize hrswk race
* Compute "Good job" variable (professional or managerial).
gen goodjob=prof+mang
* Compute dummy variable for white/ nonwhite.
recode race (1=1) (else=0), gen(white)
* hrswk (hours work per week) seems to be off by factor of 10,
* so correct.
replace hrswk = hrswk/10.
label define gdjob 0 "Other" 1 "Prof, Manager"
label values goodjob goodjob
label define white 0 "NonWhite" 1 "White"
label values white white
* Limit to cases with complete data
keep if !missing(jsat, goodjob, tenure, firmsize, hrswk, white)

* Compute race interaction terms.
gen tenurewh=tenure*white
gen firmszwh=firmsize*white
gen goodjobwh=goodjob*white
gen hrswkwh=hrswk*white

* Regressions, set 1. Mean job satisfaction difference between groups.
reg jsat white
* Regressions, set 2. Test for any differences between groups.
nestreg: reg jsat (goodjob tenure firmsize hrswk) (white goodjobwh tenurewh firmszwh hrswkwh)
* Regressions, set 3. More detailed tests for differences in effects.
nestreg: reg jsat (goodjob tenure firmsize hrswk) (white) (goodjobwh tenurewh firmszwh hrswkwh)
* t-tests for compositional differences
ttest goodjob, by(white)
ttest tenure, by(white)
ttest firmsize, by(white)
ttest hrswk, by(white)
ttest jsat, by(white)
```