

Soc 63993, Advanced Social Statistics II
Homework No. 6
Interaction effects and group comparisons

Problem 1. Download *gender.dta* and/or *gender.sav* 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. (See the hint at the end of this homework if you need some help.)

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, computing any additional variables that may be needed for the analysis:

- 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.

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 *jqqs2.sav* and *jqqs2.sps* from the course web page. Selected variables from The Quality of Employment Survey are contained in *jqqs2.sav*. Run *jqqs2.sps* 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.

NOTE: The information on the printouts about R square change, F change, and Sig F change will make it easy to confirm the accuracy of many of your answers.

Following is a copy of jgqes2.sps:

```
* Problem 2. Quality of Employment survey. Modify file name
* and location if necessary.
Get FILE="D:\soc63993\homework\jgqes2.sav".

* Compute "Good job" variable (professional or managerial).
COMPUTE GDJOB=PROF+MANG.
* Compute dummy variable for white/ nonwhite.
RECODE race (1=1) (else=0) into white.
* hrswk (hours work per week) seems to be off by factor of 10,
* so correct.
Compute hrswk = hrswk/10.
value labels gdjob 0 'Other' 1 'Prof, Manager'/
             white 0 'Nonwhite' 1 'White'.

* Compute race interaction terms.
COMPUTE tenurewh=TENURE*white.
COMPUTE firmszwh=FIRMSIZE*white.
COMPUTE gdjobwh=GDJOB*white.
COMPUTE hrswkwh=HRSWK*white.

* Regressions, set 1. Mean job satisfaction difference between groups.
REGRESSION /VARIABLES=JSAT white GDJOB gdjobwh TENURE tenurewh
           FIRMSIZE firmszwh HRSWK hrswkwh
           /statistics cha def
           /DEPENDENT JSAT
           /METHOD ENTER white.

* Regressions, set 2. Test for any differences between groups.
REGRESSION /VARIABLES=JSAT white GDJOB gdjobwh TENURE tenurewh
           FIRMSIZE firmszwh HRSWK hrswkwh
           /statistics cha def
           /DEPENDENT JSAT
           /METHOD ENTER GDJOB tenure firmsize hrswk
           /method enter white gdjobwh tenurewh firmszwh hrswkwh.

* Regressions, set 3. Test for interaction effects.
REGRESSION /VARIABLES=JSAT white GDJOB gdjobwh TENURE tenurewh
           FIRMSIZE firmszwh HRSWK hrswkwh
           /statistics cha def
           /DEPENDENT JSAT
           /method enter gdjob tenure firmsize hrswk white
           /method enter gdjobwh tenurewh firmszwh hrswkwh.

T-TEST
GROUPS=white(0 1)
/MISSING=ANALYSIS
/VARIABLES=gdjob tenure firmsize hrswk
/CRITERIA=CIN(.95) .
```

Hint for Homework #6, Problem 1

Three regressions are run by the following commands. (These are all models we have estimated before, but this time I am using the `nestreg` command to do it.) For each regression that is estimated, answer the following questions: According to the regression, do the models for blacks and whites differ? If they do differ, how do they differ? If you can answer these questions, then problem #1 should be pretty straightforward.

```
. use "http://www.nd.edu/~rwilliam/xsoc63993/statafiles/blwh.dta"
. gen blacked = black * ed
. gen blackjob = black * job
. nestreg: reg income (educ jobexp) (black) (blacked blackjob)
```

Block 1: educ jobexp

Source	SS	df	MS	Number of obs = 500		
Model	32798.4018	2	16399.2009	F(2, 497)	=	1103.96
Residual	7382.84742	497	14.8548238	Prob > F	=	0.0000
				R-squared	=	0.8163
				Adj R-squared	=	0.8155
				Root MSE	=	3.8542

income	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
educ	1.94512	.0436998	44.51	0.000	1.859261	2.03098
jobexp	.7082212	.0343672	20.61	0.000	.6406983	.775744
_cons	-7.382935	.8027781	-9.20	0.000	-8.960192	-5.805678

Block 2: black

Source	SS	df	MS	Number of obs = 500		
Model	33206.4588	3	11068.8196	F(3, 496)	=	787.14
Residual	6974.79047	496	14.0620776	Prob > F	=	0.0000
				R-squared	=	0.8264
				Adj R-squared	=	0.8254
				Root MSE	=	3.7499

income	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
educ	1.840407	.0467507	39.37	0.000	1.748553	1.932261
jobexp	.6514259	.0350604	18.58	0.000	.5825406	.7203111
black	-2.55136	.4736266	-5.39	0.000	-3.481921	-1.620798
_cons	-4.72676	.9236842	-5.12	0.000	-6.541576	-2.911943

Block 3: blacked blackjob

Source	SS	df	MS	Number of obs =	500
Model	33411.2623	5	6682.25246	F(5, 494) =	487.60
Residual	6769.98696	494	13.7044271	Prob > F =	0.0000
				R-squared =	0.8315
				Adj R-squared =	0.8298
Total	40181.2493	499	80.5235456	Root MSE =	3.7019

income	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
educ	1.893338	.054125	34.98	0.000	1.786994	1.999681
jobexp	.722255	.0396598	18.21	0.000	.6443322	.8001777
black	3.409988	1.756477	1.94	0.053	-.0410984	6.861075
blacked	-.2153886	.1038015	-2.08	0.039	-.4193354	-.0114418
blackjob	-.3002799	.0812705	-3.69	0.000	-.4599584	-.1406015
_cons	-6.461189	1.0479	-6.17	0.000	-8.520079	-4.402298

Block	F	Block df	Residual df	Pr > F	R2	Change in R2
1	1103.96	2	497	0.0000	0.8163	
2	29.02	1	496	0.0000	0.8264	0.0102
3	7.47	2	494	0.0006	0.8315	0.0051