Criminal Justice Doctoral Comprehensive Exam
Quantitative Methods
August 2018

Instructions: You will have 3 and a half hours to complete this exam. There are two questions, be sure to answer *both* questions. The questions are not related; you do not need to refer to one of them in answering the other. Read the instructions carefully to be sure you are answering the correct part(s) of each question.

Question 1

Using data from serious adolescent offenders, you attempt to predict the subject's level of self-control by examining the social control/bond variables and other key demographic characteristics. You hypothesize that as parental monitoring and attachment increase, adolescent's self-control also increases. You obtain data on respondent's level of self-control (created as a composite scale using multiple item scores and treated as a continuous variable), parental monitoring and warmth (also created as composite scales based on multiple items and treated as continuous variables), race (recoded this categorical variable into multiple binary variables such as White, Black, Hispanic, and Other; and used 'White' as a reference category in the analysis), and gender (recoded this variable into binary variables of Male and Female; and used 'Female' as a reference category in the analysis). With these data you conduct a multiple OLS regression analysis using SPSS, and here are the results:

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.270[a]	.073	.068	.82251

a. Predictors: (Constant), parent_warmth, hispanic, male, other, parental_monitoring, black

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	63.073	6	10.512	15.539	.000[b]
	Residual	799.645	1182	.677		
	Total	862.719	1188			

a. Dependent Variable: self_control

b. Predictors: (Constant), parent_warmth, hispanic, male, other, parental_monitoring, black

Coefficients

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
		B	Std. Error	Beta			Tolerance	VIF
1	(Constant)	1.846	.145		12.764	.000		
	Black	.255	.066	.147	3.875	.000	.543	1.842
	Hispanic	.038	.067	.021	.568	.570	.564	1.772
	Other	-.023	.125	-.006	-.183	.855	.860	1.163
	Male	-.055	.070	-.022	-.788	.431	.967	1.035
	Parent_monitoring	.162	.029	.163	5.677	.000	.946	1.057
	Parent warmth	.168	.036	.134	4.606	.000	.924	1.083

a. Dependent Variable: self_control

Using this output, answer the following questions:

a. What does R square represent? Make sure to discuss it based on the "Sum of Squares" column in the ANOVA table.

b. In the ANOVA table, we can reject the null hypothesis at $p < .01$. Based on the *F*-test results, what would your conclusion about the proposed regression model?
Make sure to state clearly both null and alternative hypotheses in this *F*-test when you interpret the results.

c. Interpret partial slope coefficients and the intercept value (only the ones that are statistically significant at $p < .05$).

d. How are standardized coefficients calculated and what can you infer from the results of Beta estimates?

e. Construct 95% confidence intervals for the slope coefficient of "parent monitoring" variable. Using an alpha level of 0.05, what do you conclude about the statistical significance of the estimated regression coefficient?
Do you reach the same conclusion when you use *p*-value (presented in the table)?

f. What do the results of "Collinearity Statistics" (Tolerance and variance inflation factor: VIF) suggest?

Question 2

You conduct a study on the effect of self-control on subsequent offending behavior. You have a data set collected from a sample of serious adolescent offenders who were adjudicated delinquent or found guilty of a serious offense. The outcome variable (Offending1) was measured by the number of self-reported offending in the past six months. The primary predictor (self_control) was created by taking the average of 15 items (e.g., People who get me angry better watch out). The measure asks participants to rank how much (1= False to 5= True) their behavior in the past six months matches a series of statements. Higher scores indicate higher levels of self control. With these data, you run a negative binomial regression analysis with Offending1 as a dependent variable and the self-control as a primary predictor (and the other IVs are controls). Black, Hispanic, Other are binary variables indicating the race of subjects (reference category = White). "two_parents" indicate if the subjects have both parents in the household or not (no=0, yes=1). Here are the results from Stata.

```
. nbreg Offending1 black hispanic other two_parents self_control , irr nolog

Negative binomial regression                    Number of obs   =        829
                                                LR chi2(5)      =      50.84
Dispersion     = mean                           Prob > chi2     =     0.0000
Log likelihood = -4000.4694                     Pseudo R2       =     0.0063
```

Offending1	IRR	Std. Err.	z	P>|z|	[95% Conf.	Interval]
black	.8994735	.1560813	-0.61	0.541	.6401525	1.263843
hispanic	.7058052	.1263651	-1.95	0.052	.4969222	1.002493
other	.4337002	.1492905	-2.43	0.015	.2208944	.8515195
two_parents	.5033958	.0762793	-4.53	0.000	.3740481	.6774725
self_control	.6775731	.0544808	-4.84	0.000	.5787815	.7932274
_cons	468.7973	133.4311	21.61	0.000	268.3568	818.9505
/lnalpha	1.276062	.0422391			1.193275	1.35885
alpha	3.582506	.1513218			3.297865	3.891714

```
Likelihood-ratio test of alpha=0:  chibar2(01) = 2.7e+05 Prob>=chibar2 = 0.000
```

a. When count outcomes are treated as they are continuous and conventional OLS regression models are applied, what happens to the OLS estimators and Why?

b. To address the issues you discussed above, researchers often rely on either Poisson regression model or Negative binomial regresson model as alternative modeling strategies. Based solely on the Stata output tables, which model is preferred and why?

c. When there is evidence of overdispersion but Poisson regression model is applied, what happens to the estimators?

d. Interpret regression coefficients (except for the intercept value) that are significant at $p < .05$ in terms of IRR (Incidence Rate Ratio).