METHODS/STATISTICS COMPREHENSIVE EXAM

SPRING 2016

You have six (6) hours to complete this examination

Please Answer:

One (1) question from the qualitative methods section

Two (2) questions from the quantitative methods section

Qualitative Methods:

1. In our SOCI606 – Qualitative Methodology class—we discussed four main principles to mixed methods research projects: identifying and choosing the best design, matching your chosen design to your research problem, explaining why mixed methods is necessary and explaining the synergy between the quantitative and qualitative strands.

You have been asked to design a mixed methods research project to investigate the relationship between drugs and crime – at the community level- in Wilmington so that city officials can help stem the direct and collateral damage emanating from drug-involved criminal networks. Please design a mixed methods study, using both quantitative and qualitative data, and addressing each of the four principles described above.

2. In a 1999 paper in *Law & Social Inquiry*, ethnographer Sudhir Venkatesh argued that social scientists could never fully deliver on their human subject promises to research subjects. To what extent do you agree and disagree with this claim? Your answer should discuss the major ethical issues raised in qualitative research and how the investigator tries to handle them, e.g., disclosure, consent forms and various other human subjects. Make sure to address how each ethical precaution permits you to balance the scales of beneficence and justice. Use examples to illustrate your points.

Quantitative Methods

1. Multi-level modeling is a frequently-used strategy, though researchers often estimate and report results in different ways. Variations involve choice of fixed- vs. random-effects models, the use of centering, and reporting of model fit statistics (among others). Discuss these choices by explaining different modeling and reporting strategies, and why a researcher should choose one over another. Specifically, answer the following:

   a. What is the difference between fixed-effects and random-effects analysis, and why should a researcher use one instead of the other?
   b. When performing a random-intercept model, what model fit statistics are important to report, and how does one interpret them?
   c. Is centering necessary or optional in multi-level models? When or why should one center variables? How does one interpret coefficients for centered variables?

Understanding temporal change in human behavior and/or attitudes is a central issue in our discipline. With technological advances, what are being called intensive longitudinal data (ILD) are increasingly being collected, which allow us to describe temporal behavioral/attitudinal changes in detail along with related environmental and psychosocial antecedents and consequences. A few statistical methods have been developed to simultaneously model the effects of both time-stationary (e.g. gender, race) and time-varying covariates on behaviors/attitudes/physical outcomes over time.

a. Provide an overview of at least two different statistical models that could be used to examine the simultaneous effects both time-stationary and time-varying covariates on change in some measure over time. It may be helpful to provide a substantive example for this exercise. Be explicit about how each model determines the effects of both types of variables on change in the outcome variable.
b. Now provide a discussion of which method you would prefer to use for such an endeavor along with the strengths and limitations of this choice compared to the other method.

2. In this question we outline the results growth curve models predicting adolescents' physical aggression trajectories between the ages of 11 and 18 and ask you to interpret the results. First, however, we will describe the study and the variables. The specific questions for this problem follow this discussion and the results.

Adolescents develop within an environment with multiple variables that synergistically influence their behavior. Imagine you are a researcher interested in examining the factors related to aggression during adolescence. You collect survey data at five waves, every 6 months beginning when students were in 6^{th}-8^{th} grade and then ending when they were in 8^{th}-10^{th} grade. Surveys are also collected from parents. In this study, you are measuring both individual variables along with census track data matched to each student's home address. The **dependent variable** measuring physical aggression is an interval index ranging from 0-16 that indicates the extent of **physical aggression** the youth had engaged in during the past 6 months. There are several independent variables including: 1) **neighborhood disadvantage,** measuring economic disadvantage of student's neighborhood; 2) **neighborhood social disorganization,** an index based on parents' assessments of both collective efficacy and disorganization within their neighborhoods; 3) **family conflict,** an index measuring students' reports of conflict and fighting within their family household; 4) **parent-child bonding,** an index based on reports of youth measuring attachments to their parents; 5) **parental control,** an index indicating students' reports about family rules, and whether they know where their children are when not at home, etc., 6) demographic controls including race, parent education, and family structure (single-parent versus two parent).

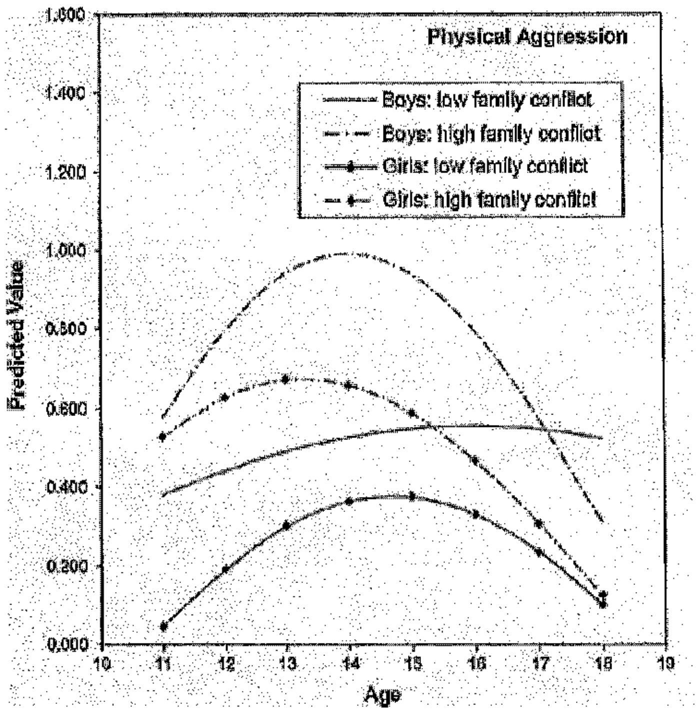
You conduct a multi-level growth curve to model physical aggression. Models estimate changes in aggression outcomes over time (level 1) within individuals (level 2) nested within neighborhoods (level 3). The models also included three random effects (individual intercept, individual linear slope, and neighborhood intercept); level 2 random effects are allowed to correlate. You determine that the five waves of data cannot support additional random effects so all effects of level-2 and level-3 predictor variables are fixed. You then use backwards elimination to remove two-way interactions that are not significant.

The table on the next page presents the associations of neighborhood and family factors with physical aggression trajectories between ages 11 and 18 (all demographics are controlled):

	Physical aggression	
	B	95 % CI
Effects on intercept		
Intercept	0.238**	(0.106, 0.370)
Male	0.153**	(0.048, 0.258)
Socioeconomic disadvantage	0.007**	(0.004, 0.010)
Social disorganization	0.003	(−0.039, 0.045)
Family conflict	0.169**	(0.109, 0.228)
Parent–child bonding	−0.036*	(−0.067, −0.006)
Parental control	−0.060**	(−0.082, −0.038)
Disadvantage × male	−0.004**	(−0.007, −0.010)
Conflict × male	−0.109*	(−0.199, −0.020)
Effects on linear slope		
Age	0.115**	(0.062, 0.167)
Male	−0.012	(−0.082, 0.059)
Conflict	−0.031	(−0.070, 0.009)
Bonding		
Conflict × male	0.078**	(0.021, 0.136)
Effects on quadratic slope		
Age-squared	−0.019**	(−0.027, −0.012)
Male	0.003	(−0.008, 0.014)
Conflict	0.001	(−0.005, 0.008)
Conflict × male	−0.011*	(−0.020, −0.001)

For significant interactions, trajectories are graphed at high and low values of the predictor of interest, using scores of one standard deviation above and below the mean and setting all other predictors to zero.

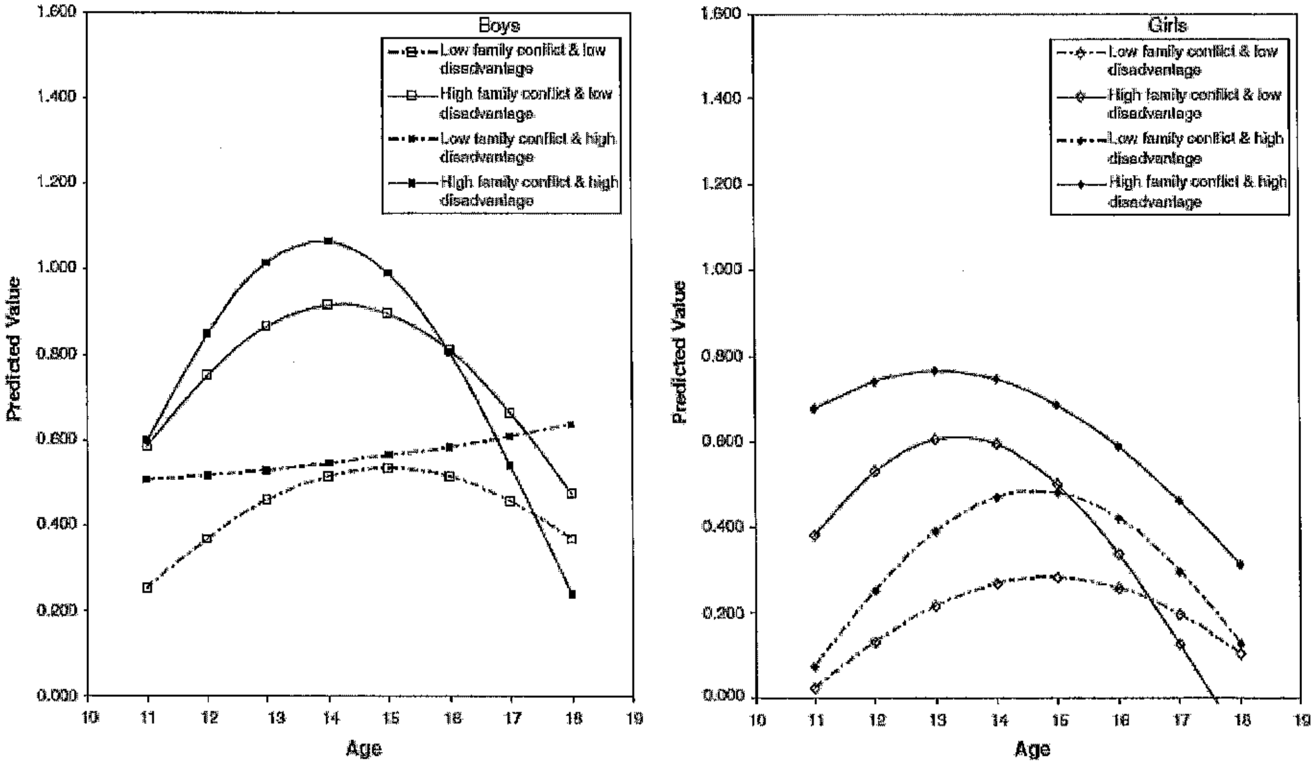
The following graph displays the trajectories of physical aggression as a function of age, gender and family conflict:



While some of the coefficients in the table below replicate the earlier table, the purpose of the table below is to show results of the models examining the three-way interactions of neighborhood disadvantage, family conflict, and gender for physical aggression trajectories between ages 11 and 18:

	Physical aggression	
	B	95 % CI
Effects on intercept		
Intercept	0.232**	(0.100, 0.365)
Male	0.160**	(0.053, 0.267)
Socioeconomic disadvantage	0.007	(−0.002, 0.015)
Social disorganization	0.003	(−0.039, 0.045)
Family conflict	0.163**	(0.103, 0.224)
Parent–child bonding	−0.037*	(−0.067, −0.006)
Parental control	−0.060**	(−0.082, −0.037)
Disadvantage × male	−0.001	(−0.013, 0.011)
Conflict × male	−0.098*	(−0.188, −0.008)
Conflict × disadvantage	0.004	(−0.004, 0.011)
Conflict × disadvantage × male	−0.008	(−0.019, 0.003)
Effects on linear slope		
Age	0.118**	(0.065, 0.172)
Male	−0.017	(−0.088, 0.054)
Disadvantage	0.000	(−0.006, 0.005)
Conflict	−0.025	(−0.065, 0.015)
Disadvantage × male	−0.001	(−0.009, 0.008)
Conflict × male	0.067*	(0.009, 0.125)
Conflict × disadvantage	−0.004	(−0.008, 0.001)
Conflict × disadvantage × male	0.007*	(0.001, 0.014)
Effects on quadratic slope		
Age-squared	−0.020**	(−0.027, −0.012)
Male	0.004	(−0.007, 0.015)
Disadvantage	0.000	(−0.001, 0.001)
Conflict	0.000	(−0.007, 0.007)
Disadvantage × male	0.000	(−0.001, 0.001)
Conflict × male	−0.008†	(−0.018, 0.001)
Conflict × disadvantage	0.001†	(0.000, 0.001)
Conflict × disadvantage × male	−0.001*	(−0.002, 0.000)

To summarize these interrelationships, you predict trajectories for hypothetical groups defined by gender and varying levels of neighborhood disadvantage and family conflict and graph them as follows:



Questions: For this question, your task is discuss all of these findings: 1) interpret all significant coefficients in the first table; 2) interpret the first graph; 3) interpret all significant coefficients in the second table; 4) interpret the second set of graphs, and finally 5) provide a summary discussion of the effects of these neighborhood, family, and individual characteristics on the developmental trajectory of physical aggression for male and female adolescents.

3. Explain the necessity of ordinal logistic regression, that is, when and why should it be used instead of either multinomial logistic regression, negative binomial or Poisson models, or a series of binary logistic regression models. Second, discuss what diagnostic tests one should do before estimating ordinal logistic regression models. Finally, interpret the following ordinal logistic regression results, using a $p<.10$ significance level. This model, estimated in Stata, uses GSS 2014 data to regress political party id on dummy variables for Black and Other race (white is withheld as the contrast), years of education, sex (male), and the interaction of male * years of education. A frequency table for the dependent variable follows, and then the regression output.

```
. tab partyid, missing
```

political party affiliation	Freq.	Percent	Cum.
0. strong democrat	419	16.51	16.51
1. not str democrat	406	16.00	32.51
2. ind,near dem	337	13.28	45.78
3. independent	502	19.78	65.56
4. ind,near rep	249	9.81	75.37
5. not str republican	292	11.51	86.88
6. strong republican	245	9.65	96.53
.	62	2.44	98.98
.d. dk	1	0.04	99.01
.n. na	25	0.99	100.00
Total	2,538	100.00	

```
. ologit partyid black other1 educ male educ_male

Iteration 0:   log likelihood = -4687.0114
Iteration 1:   log likelihood = -4531.9658
Iteration 2:   log likelihood = -4530.4732
Iteration 3:   log likelihood = -4530.4722
Iteration 4:   log likelihood = -4530.4722

Ordered logistic regression                       Number of obs   =       2449
                                                  LR chi2(5)      =     313.08
                                                  Prob > chi2     =     0.0000
Log likelihood = -4530.4722                       Pseudo R2       =     0.0334
```

partyid	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
black	-1.797781	.1082998	-16.60	0.000	-2.010045	-1.585517
other1	-.5270524	.1122192	-4.70	0.000	-.746998	-.3071069
educ	-.0331461	.0155661	-2.13	0.033	-.0636551	-.0026372
male	-.4248837	.3265322	-1.30	0.193	-1.064875	.2151077
educ_male	.0430053	.023545	1.83	0.068	-.003142	.0891525
/cut1	-2.430448	.2263511			-2.874088	-1.986808
/cut2	-1.432572	.2220089			-1.867702	-.9974429
/cut3	-.8025632	.220201			-1.234149	-.3709771
/cut4	.1158571	.219703			-.3147529	.5464671
/cut5	.6652341	.2210121			.2320584	1.09841
/cut6	1.618805	.2259179			1.176014	2.061596