Problem Set #[1]
MACS 30100, Dr. Evans
Wenxi Xiao

1 Problem 1

1.1 part (a)

The article that I found proposed and tested a model that describes how father household entrances and exits and childrens genetic makeup affect childrens antisocial behavior.

1.2 part (b)

Mitchell, C., McLanahan, S., Notterman, D., Hobcraft, J., Brooks-Gunn, J., & Garfinkel, I. (2015). Family Structure Instability, Genetic Sensitivity, and Child Well-Being1. *American Journal of Sociology, 120(4), 1195-1225.*

1.3 part (c)

The following equation is proposed by Mitchell et al. to model the effects of father household entrances and exits and a childs genetic characteristics on the childs externalizing behavior/antisocial behavior.

$$\begin{aligned} y_{it} = {} & \alpha_0 + \alpha_1 \mathrm{GENES}_i + \alpha_j \mathbf{X}_{ij} + \beta_0 t + \beta_1 \mathrm{GENES}_i t + \beta_j \mathbf{X}_{ij} t \\ & + \gamma_{tt'} w_{it'} + \lambda_{tt'} (\mathrm{GENES} \times w_{it'}) + u_i + v_i t + \varepsilon_{it}, \end{aligned}$$

GENESi are a child s genotype.
X_i is a vector of j control variables, including grandparents characteristics (whether parents were raised in a two-parent household), parents characteristics (race, age, education, employment status, income, health, mental health history, incarceration history, drug and alcohol history), childs geder, parents relationship quality (supportiveness, violence, whether they discussed having an abortion), and childs health (low birth weight, birth order).
y_i represents the childs externalizing behavior.
$\gamma_{tt'}\omega_{it'}$term represents the effect of each previous time t0 father household entry or exit on child externalizing behavior at time t for each ith child, which takes into account of the time varying changes in the father s residential status (i.e., living with or without the mother) on the childs externalizing behavior.
The interaction term $\lambda_{tt'}(GENES \times \omega_{it'})$between genes and family structure changes/-family instability.
ε_i is some measurement error that characterize each child s trajectory of externalizing behaviors.

1.4 part (d)

$GENES_i$, X_i and ω_i are exogenous variables. y_i is an endogenous variable.

1.5 part (e)

This model is dynamic because it represents how childrens externalizing behaviors change over time (i.e., there is a time term in the model). This model is linear because it has a linear format. This model is stochastic because it has an error term.

1.6 part (f)

I think the model would describe child antisocial behavior better if variables of the households geographical and/or neighborhood information were included, with information such as if the child experienced any location changes while growing up and the crime rate and average SES of the neighborhood the child spent most of his/her time growing up. Including these variables that represents the childs external grow-up environment could make the model more precisely describe the child antisocial behaviors and more generlizable. Maybe there is a link between father entrances and exits and family location changes, and family location changes could actually be the direct cause of childs antisocial behavior. However, there is also a tradeoff between including more variables and the model being parsimony. Therefore, researchers could do more analyses and test several models to find the right balance between parsimony and the goodness of fit.

2 Problem 2

2.1 part (a)

$$Y_i = \alpha_{0i} + \beta_1 X_{ij} + \beta_2 G_i + \beta_3 S_i + \beta_4 G_i X S_i + \beta_5 C_{ik} + \beta_6 T_i + \varepsilon_i$$

The above equation is a model that intends to describe how long a musician lives. Y_i is the dependent endogenous variable that represents the musicians predicted lifespan (in years). X_i is a vector of j variables, prospective variables including gender, current age, martial status, race, years of education, if having adverse childhood experience. G_i is a categorical endogenous variable representing the primary music genre the musician produces, inducing rock, pop, hip hop/rap, country, etc.). S_i is a categorical endogenous variable representing the type of the primary substance the musician uses, including alcohol, marijuana, LSD, heroin, angel dust, smoking, and cocaine. T_i is a discrete endogenous variable representing the year the musician has used the substance. C_i is a vector of k variables, prospective variables including average daily working hours, total albums sales, number of songs produced by the musician, and number of major industry awards. The interaction term represents the interactive effect the music genre has on substance choices. ε_i is an error term that is distributed lognormal LN(0,sigma) where sigma is the standard deviation of the log of the error term.

2.2 part (d)

I think all the factors that I entered into my model are important, with the relatively more important ones being music genre (G_i), substance abuse (S_i), the interaction of G_i and S_i, the year of using the abusive substance (T_i), and the career information (C_{ik}) of the musician. As we often see on the news that substance abuse is a key factor that influences musicians lifespan. Also, a musicians chosen music genre often influences his/her lifestyle, healthy or unhealthy, which can have a great impact on the musicians lifespan. The career information can also signal the musician's lifestyle and indirectly measure the level of stress the musician faces.

2.3 part (e)

First, I chose my factors based on news articles and some research articles investigating the lifespan of musicians, in the hope of increasing the validity of my model before I conduct any tests on it. Research and media have told us that substance abuse is a major factor that influences musicians lifespan. And music genre as well as career information such as daily workload and annual income are often correlate with the musicians lifestyle. Secondly, most of the chosen variables include information that is relatively objective and accessible, so that it is feasible to conduct tests on and make predictions with my model. For instance, a given musicians music genre and career information can be found online either from Wikipedia or the musicians social network profile. Even though I think musicians psychological state could be a possible great factor, I did not include that in my model, because there is no very good measurement that has high construct validity on measuring peoples psychological state. Plus, peoples medical records on history of psychological disorder are not always obtainable. Third, I intended to build a model that is highly predictive of lifespan (i.e., can explain much variance of the lifespan). Therefore, I had to give in some parsimony of my model by including a number of variables that I deemed highly relevant. Admittedly, I need to do some further analyses to determine the right balance between parsimony and fit.

2.4 part (f)

I would first randomly select a group of deceased musicians from the database, presumably large enough for me to have good power to conduct statistical analyses. I would then use R to conduct regression analyses using the data I collected online from webpages, presumably by using text extracting/analyses techniques . I would also do some data cleaning beforehand. Then, I would do stepwise regression to better revise my model.