Problem Set #1
MACS 30100, Dr. Evans
Yiqing Zhu

Problem 1 Classify a model from a journal
Part (a)-(b).
The article that I have found is cited as:
Goldrick-Rab S, Kelchen R, Harris D N, et al. Reducing Income Inequality in Educational Attainment: Experimental Evidence on the Impact of Financial Aid on College Completion 1[J]. American Journal of Sociology, 2016, 121(6): 1762-1817.

Part (c).
The authors use the following model to estimate the impact of The Wisconsin Scholars Grant (WSG) offer on on-time (four-year) bachelors degree completion for student i:

$$Y_i = \alpha_{0i} + \alpha_{1i}T_i + \alpha_{2i}X_i + \alpha_{3i}(T_i \times X_i) + \alpha_{4i}C_i + \epsilon_i, \tag{1}$$

where Y_i represents the outcome of interest (graduation), T_i is an indicator for whether the student was assigned to receive the WSG, X_i represents out-of-pocket costs or the demographic measure of interest (race/ethnicity, age, gender, parental education, dependency status, family income, immigration status), $(T_i \times X_i)$ represents the interaction, and C_i is a vector of college fixed effects (undergraduate enrollment, financing, enrollment management).

Part (d).
Y_i is endogenous variable, while T_i, X_i, and C_i are exogenous variables.

Part (e).
This model is static, because it helps in depicting static constituents of the system and in this model, one time period will not affect another.
This model is linear as it is an ordinary least squares regression.
This model is stochastic since it includes the error term ϵ_i.

Part (f).
Instead of evaluating the achievement of students including credits earned, accumulative GPA, and the varying pattern of GPA over time purely as the treatment impact, it can also be regarded as a partially exogenous variable as it has an impact on students expectation on future study life and would contribute to the decision on whether to continue or end school. Moreover, students major, retention rate in his or her social community, economic environment can also be variables of feature that can be considered in this model.

Problem 2 Make your own model
Part (a)-(c).

$$L_i = \beta_0 + \beta_1 \overline{L}_i + \beta_2 G_i + \beta_3 I_i + \beta_4 W_i + \beta_5 M_i + \beta_6 U_i + \beta_7 E_i + \beta_8 D_i + \epsilon_i \qquad (2)$$

I mark a particular musician as i and his/her predicted lifespan (in years) is the dependent endogenous variable L_i. $\overline{L}_i$ is the average expected lifespan predicted from the popularity who match the demographic characteristics with i including born year period, gender, race, nationality, place of residence. G_i is a categorical variable implying the genre of the music of i. I_i indicates the income level of i, being marked between (low income)1 to 10(high income). W_i indicates the workload of i of recent three years, being marked between (low workload)1 to 10(high workload), evaluated from a combination of information about frequency of music/music video release, frequency of concerts/shows, frequency of fan meetings, frequency of oversea work, etc. M_i indicates the number of marriage. U_i indicates drug/alcohol/cigarette use, evaluated by both the effect of drugs and consumption of drugs, marked as between (light user)1 to 10(heavy user). E_i indicates the exercise habit of i, marked between (little exercise)1 to 10(much exercise). D_i indicates the diet habit of i, marked between (healthy diet)1 to 10(unhealthy diet).

Part (d).
The key factors that I think would influence the outcome are the ones that I have included in my model, namely, average lifespan of people with the similar demographic characteristics, the genre of music, income level, workload, the number of marriage, drug/alcohol/cigarette use, exercise habit, and diet habit.

Part (e).
First, the average lifespan of people with the similar demographic characteristics serve as the starting point to predict popular musicians' lifespans and can be obtained by professional census report. Second, though through conjecture, the musicians in the same music genre share some patterns which would affect their lifespans. For example, Rock musicians are known to have more scandals, drug abuse problems, depression problems, which indicates their averagely shorter lifespan than musicians of other genres.Third, to evaluate musicians' lifespans different from general people, I focus on two aspects, mental health and physical health. To evaluate mental health, I take into account three elements: Income level, which is an accessible variable showing the musicians' pressure under popularity since higher income usually indicates more popularity and more accompanying pressure; Workload, which shows anxiety and pressure of work and also serves as an indicator of physical health; The number of marriage, which shows the stability of and support from the family. For physical health, I have considered three elements: Alcohol/drug/smoke use, as musicians are often reported to have alcohol/drug/smoke abuse problems; Exercise and Diet habit, which are important aspects for physical health.

Part (f).
To do a preliminary test, I can use the data from musicians that have passed away. I may choose 50 musicians from different genres and are super stars in their own

domains. I can easily gather their demographic information, genre information, and the number of marriage from Google. To determine the average lifespan of people with similar demographic characteristics, I may need to consult census report. I can figure out estimated work load in his/her last three years by searching their news reports. I can also track the musicians' income level by searching relevant news or estimating from his/her working load. About drug/alcohol/cigarette use, exercise habit and diet habit, I can find information from news or if possible, get information from the musician's families, friends, colleagues, or fans. By testing my model with known lifespan of musicians, I may be able to find whether my factors are significant and find the coefficient of my variables.