Problem Set #1

MACS 30000, Dr. Evans
Rodrigo Valdes Ortiz

Problem 1

Part (a). I used the American Economic Journal

Part (b). McGoldrick, K., and Schuhmann, P. W. (2016). The Impact of Challenge Quizzes on Student Knowledge. *The American Economic Review*, 106(5), 373-377.

Part (c). Equation

$$score_i = \beta_0 + \beta_1 male + \beta_2 SAT + \beta_3 classyear + \beta_4 attendance + \beta_5 methods + \beta_6 quizz_i + \beta_7 challengequiz + \epsilon \quad (1)$$

The dependent variable is the exam score. The independent variables are the observable student characteristics: sex, standardized SAT score, class year, attendance, number of study methods used during the semester, the grade of the in-class quiz prior to the examination, and a dummy variable that indicates if the student exercised their option to a challenge quiz before the exam.

Part (d). Exogenous: male, SAT, year, attendance, methods, quizz, and challenge quizz. Endogenous: β_0 to β_7.

Part (e). The model is **static** because it does not consider different moments of the same variables. The model is **linear** because all the relationships between the depend and the independent variables are linear. Finally, the model is **stochastic** because it considers an error term which recognizes some unobserved variables.

Part (f). I might include the number of math classes that students took before they enrolled to Microeconomics 1. While the SAT scores control by the characteristics before college, the math classes will consider the college-level classes that affect student's abilities.

The paper only considers the number of methods that students use to prepare exams. However, maybe it is not just the number of methods, but some particular way of studying that makes the difference. Then, I propose to use dummy variables for each method and analyse the effects of each one. The hypothesis is that one o two approaches can explain most of the variation, but they lost that information with the aggregation of those in just one variable (number of methods), instead of regressing each method separately.

Problem 2

Part (a). Write down your own model

$$\begin{aligned} lifespan = \beta_0 + \beta_1 male + \beta_2 alcoholic + \beta_3 tobacco + \beta_4 f(agefirsthit) \\ + \beta_5 harddrugs + \beta_6 dangerousgenre + \beta_7 g(socialmedia) + \xi \end{aligned} \tag{2}$$

Then, lifespan is modeled as a function of the singer's characteristics, sex, alcoholism, tobacco use, a quadratic polynomial of age when they release their first hit represented by *f(agefirsthit)*, hard drug use, dangerous genre, and a function *g(social media)* that indicates the popularity of the singer.

Part (b). The dependent variable is lifespan

Part (c). It is possible to predict the lifespan of a singer with the independent variables. Then, it is a complete data generating process.

Part (d). Why are those the key factors?

I started with a basic model to predict the lifespan of an individual, which include some habits and inherent characteristics, such as gender, alcohol abuse, and tobacco use. Then, I consider variables necessary for musicians, such as a second degree polynomial of the age at their first hit, which will indicate the time where they become famous. I used a second-degree polynomial because people who are famous since childhood may be more affected by the famous issues, while a middle-aged adult can manage that stress in a different way. Furthermore, I consider that the adverse effects of being famous are bigger for someone who becomes famous old, not strong enough to that. Another variable relevant for musicians is hard drugs consumption, such as LCD, methamphetamines, cocaine, or others can be highly addictive and harmful. The dangerous genre indicates if the person sings favorite styles by drug dealers, which sometimes evolves in dangerous situations, like hip-hop, rap or regional styles in South America. Also, there are other hazardous genres, related with the culture they represent because they tend to accidents or suicide, such as punk or metal [1]. Finally, I used the function *g(social media)*, which measure how famous the singer is, wich depends on mentions and followers in twitter, facebook fans and likes, and broadcast success in radio stations, Spotify, Google music, and similars. There is expected that more famous people will develop more psychological and health problems.

The described factors are adequate to predict the lifespan of a singer because those consider the general characteristics that help to forecast general mortality, such as gender and regular drugs usage, with features relevant for musicians, like age at their first hit, hard drugs, music genre, and how famous they are. To sum up, I used the variables related general life expectancy, and those which indicate the additional risks as singers.

Part (e). Why did you decide on those factors and not others?

[1] Kenny, Dianna (2015). Music to die for: how genre affects popular musicians? life expectancy. *The Conversation*, January 8th, 2016, http://theconversation.com/music-to-die-for-how-genre-affects-popular-musicians-life-expectancy-36660

I focused on three characteristics to select variables: measurable, reasonable, and partial evidence of an effect on life expectancy. The variables that I use are measurable with a questionary and data analytics for *g(social media)* (which is a function of digital media). Those are reasonable because they have an effect on singers life expectancy. For instance, gender, alcoholic, tobacco use, and hard drugs have direct consequences on health. Furthermore, age at first hit measures the length of their interaction with fame which indicates the time exposed to famous people issues, such as paparazzi, false friendships, guards, isolation, among others. Dangerous genre considers the contrast between an opera singer and a metal singer, which differ in ideology, habits, fan clubs, drugs' prevalence, among others. The indicator of fame takes into account that the effects of it should be dissimilar for different levels of exposure to celebrity problems and benefits.

I omitted other variables that can be relevant for the general population, such as income. In general, high-income correlates with better health. However, I assumed that all the singers in the sample are rich enough to pay for excellent health services, eat high-quality food, and stay away from dangerous economic activities (for example, mining). Then, I excluded income because it is related with *g(social media)*, the indicator of fame. There are other factors that I did not include due to questionable measurability, such as beliefs about heaven and hell (affects suicide rates), drugs, or friendship with criminal subjects, relationships with significant others (such as mother, girlfriend, boyfriend, others), genetic predisposition to diseases, among others.

Part (f). How could you do a preliminary test whether your factors are significant in real life?

Some studies analyse the relationship between the health and innate variables (male, alcoholic, tobacco, hard drugs) and life expectancy. From those, I will investigate the relevance of that variables. To tests the significance of musics-specific variables (age, *f(agefirsthit)*, dangerous genre, and *g(social media)*, I will read previous studies, and I will look for databases that contain those variables. For each data bank, I will do regressions that help to find evidence on the relevance of the variables. If I can not do regression, I will look for correlations, just to get some initial evidence. Another way will be to build a database with the most recent-death-famous singers with public data. Then, do preliminary tests with that information.