Problem Set #[1]

MACS 30100, Dr. Evans
Xuancheng Qian

1. Classify a model from a journal.

Part (a-b). Find a theoretical or statistical model from a recently published article and give a detailed citation of the article.

Answer: McGhee, Eric, Seth Masket, Boris Shor, Steven Rogers, and Nolan McCarty. "A primary cause of partisanship? Nomination systems and legislator ideology." *American Journal of Political Science* 58, no. 2 (2014): 337-351.

Part (c). Write down the mathematical or statistical model.

Answer: $y_{it} = \mathbf{x_{it}^{'}} * \beta + \alpha_i + \delta_t + \epsilon_{it}$ where y_{it} represents legislator ideal points, $x_{it}^{'}$ represents the variables of electoral systems including semiclosed, semiopen, pureopen, and nonpartisan, α_i, δ_t indicate that states and time are fixed effects and ϵ_{it} represents the error term.

Part (d). List which variables are exogenous (determined outside the model, assumed) and which variables are endogenous (determined inside the model, the output of the model).

Answer: Endogenous variable is the legislator ideal points. And the exogenous variables are simply the explanatory variables of electoral systems including semiclosed, semiopen, pureopen, and nonpartisan.

Part (e). Classify the model as static vs. dynamic, linear vs. nonlinear, deterministic vs. stochastic.

Answer: The model is linear, static and stochastic. It is ordinary least square regression, each time period is distinct and the error term introduces the randomness.

Part (f-g). List a variable of feature that you think the model is missing that might be valuable.

Answer: In this paper, the authors study the effect of polarization based on democrats and republicans. The missing feature that might be useful could be the third party's influence as voters are not satisfied with democrats or republicans, they would definitely pay attention to the third party's legislator and this should have an influence on democrats and republicans.

2. Make your own model.

Part (a-c). Write down a model of how long popular musicians live (e.g., rock, pop, hip hop/rap, country).

Answer:

$$predicted\ lifespan_i = \beta_0 + \beta_1 * demographic\ information_i + \beta_2 * personal\ information_i + \beta_3 * drug_i + \beta_4 * musicgenre_i + \beta_5 * family\ information_i + \epsilon_i \quad (1)$$

where the dependent variable is the musician's predicted lifespan in years. The explanatory variables including the demographic information of the musicians, such as

the country development (1= developed, 0 = developing), the living environment factors consisting of air quality, water quality (rating based), the personal information including birth year, gender, marital status, education and income level, the family information including the early phase family financial support, family medical history, the drug information indicates the drug usage and musicians' music genres are also included

Part (d). What do you think are the key factors that influence this outcome?

Answer: The factors included in my model are all the important ones. And in my model, I will assume that most important factors should consist of the living environment factors including air quality, water quality, the musician's personal information of income level, drug usage and music genre. There is no doubt that environment plays an important role on human being's lifespan. Meanwhile, the musician's music genre and income level could indicate the musician's lifestyle and the capacity of health care expenditure. Lastly, the drug usage could also be a key factor in predicting musician's lifespan.

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

Answer: I choose these factors based on previous research and particular assumptions of musicians. The previous reports have shown that external factors including environment such as air quality, water quality can have an impact on human being's predicted lifespan. And music genres, drug usage, income level are particular indicative factors of musician's lifestyle, capacity of health care expenditure which are all important in predicting musician's lifespan

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

Answer: A preliminary test should involve data collection and data testing. I should acquire relevant data related to dead popular musicians from website including musician genre, demographic information, personal information and family information. We have to notice that drug usage may be private so that this could be difficult to collect real data which we could assume from previous medical research. And then we could perform statistical analysis such as ordinary least square regression for a preliminary test.

Reference

1. WHO, Geneva. "WHO methods for life expectancy and healthy life expectancy." (2014).

2. Salomon, Joshua A., Haidong Wang, Michael K. Freeman, Theo Vos, Abraham D. Flaxman, Alan D. Lopez, and Christopher JL Murray. "Healthy life expectancy for 187 countries, 19902010: a systematic analysis for the Global Burden Disease Study 2010." The Lancet 380, no. 9859 (2013): 2144-2162.