Problem Set 1 - Problem 2

Linghui Feng

January 11 2020

1 Make your own model

There are many factors that may affect one's decision on whether to get married on not. For simplicity, I will focus on factor that can be quantified.
In my model, I will include the following variables:

- The number of years an individual receive formal education (EDU)
- Age (AGE)
- Gender (GENDER = 0 if male, or 1 if female)
- Whether the individual is single or has a partner (PAIR = 0 if single, or 1 if has a partner)
- Whether to get married or not (MARRIED = 1 if get married, or 0 if not get married)

$$MARRIED = \beta_0 + \beta_1 * AGE + \beta_2 * GENDER + \beta_3 * EDU + \beta_4 * PAIR + \epsilon \quad (1)$$

(d) I think the key factors in this model are age, gender, and educational level. Newly graduates of college may less likely to get married than those who already have stable jobs, because those young adults may have more pressure on economics and careers and thus may want to focus on their careers first. For the factor of gender, females may get married early because of their economic needs and familial pressure. For instance, women from Asian family may have more pressure on getting married from their family because their cultures consider the age of a woman has value. But at the same time, men have less pressure on getting married because their families may want them to get success in their careers first. Besides, I believe the educational level also impacts people's decisions on marriage. People with higher educational level may be easier to get a well-paid job so that they do not need to rely on their partners for living.

(e) At first, I have thought of a wide range of factors that may be important, including geographical location, ethnicity, and religious belief. However, I decided not to use these factors because they were categorical data instead

of numerical data. For example, religious beliefs contain Christian, Protestant, Buddhism, and many other different religions. So the factor of religious belief contains lots of different values, and these values are difficult to quantify. If I enumerate each religious belief starting from 1, then once I switch the order of any two values, I would get a totally different result. Therefore, I think these categorical variables are difficult to control and it would be better for me not to include them here.
Those factors I use in my model, such as age, gender, and educational level, are numerical variables or can be easy to enumerate. For example, I can label each object's age by his/her actual age, and I can label the educational level by the actual year of schooling. For factors like gender, and whether the individual is single, I can easily label their values by 0 and 1 because they could only have two possible values.

(f) To check whether or not my factors are significant in real life, I would collect a bunch of data, plug them in the formula and calculate those βs. Then if I assume the significance level α to be 5%, I would compare the p-value of those factors with α. If the p-value of a factor is less than 0.05, then this factor is significant. Otherwise, this factor is not significant.