

PS7<sub>Tao</sub>

jiongliangtao

March 2019

## 1 Summary Table

Table 1:

| Statistic | N     | Mean   | St. Dev. | Min   | Pctl(25) | Pctl(75) | Max    |
|-----------|-------|--------|----------|-------|----------|----------|--------|
| logwage   | 1,669 | 1.625  | 0.386    | 0.005 | 1.362    | 1.936    | 2.261  |
| hgc       | 2,229 | 13.101 | 2.524    | 0     | 12       | 15       | 18     |
| tenure    | 2,229 | 5.971  | 5.507    | 0.000 | 1.583    | 9.333    | 25.917 |
| age       | 2,229 | 39.152 | 3.062    | 34    | 36       | 42       | 46     |

Log wages missing at rate of 0.2512. I think the logwage variable is most likely to be MNAR.

The  $\beta_1$  in first model and third model is 0.062. The  $\beta_1$  in second model is 0.049. Both of those two values are different from the true value. The  $\beta_1$  in first model and third model is closer to the true value. So, the first and third method is better than the second method.

## 2 Question 8

The data that used for this project is collected from Tweeter using API. I have collected the words people use most about this topic. I would like to see what opinions people have on this topic. I may take a linear regression model for this project.

Table 2:

|                         | <i>Dependent variable:</i> |                          |                          |
|-------------------------|----------------------------|--------------------------|--------------------------|
|                         | logwage                    |                          |                          |
|                         | (1)                        | (2)                      | (3)                      |
| hgc                     | 0.062***<br>(0.005)        | 0.049***<br>(0.004)      | 0.062***<br>(0.005)      |
| collegenot college grad | 0.146***<br>(0.035)        | 0.160***<br>(0.026)      | 0.146***<br>(0.035)      |
| tenure                  | 0.023***<br>(0.002)        | 0.015***<br>(0.001)      | 0.023***<br>(0.002)      |
| age                     | −0.001<br>(0.003)          | −0.001<br>(0.002)        | −0.001<br>(0.003)        |
| marriedsingle           | −0.024<br>(0.018)          | −0.029**<br>(0.014)      | −0.024<br>(0.018)        |
| Constant                | 0.639***<br>(0.146)        | 0.833***<br>(0.115)      | 0.639***<br>(0.146)      |
| Observations            | 1,669                      | 2,229                    | 1,669                    |
| R <sup>2</sup>          | 0.195                      | 0.132                    | 0.195                    |
| Adjusted R <sup>2</sup> | 0.192                      | 0.130                    | 0.192                    |
| Residual Std. Error     | 0.346 (df = 1663)          | 0.311 (df = 2223)        | 0.346 (df = 1663)        |
| F Statistic             | 80.508*** (df = 5; 1663)   | 67.496*** (df = 5; 2223) | 80.508*** (df = 5; 1663) |

*Note:*

\* p&lt;0.1; \*\* p&lt;0.05; \*\*\* p&lt;0.01