Qian Zhang 14/12/2020

(a) $\frac{\partial \log(wage)}{\partial educ} = \beta_1$

$\partial \log(wage) = \beta_1 \partial educ$

$100\% \Delta wage \approx \beta_1 \Delta educ$

Therefore, if we want to measure wage in pence rather than in pounds, the coefficient of wage on educ will not change, since it measures the percentage of change of wage. However, this change will impact the intercept by $\beta/100$.

(b). If we increase the sample size, the degree of freedom $(n-k-1)$ will increase. Therefore, as the formula for the estimator of σ^2 is $\frac{SSR}{n-k-1}$, it will decrease when the df increase, hence, the variance of OLS estimators will increase. The unbiasedness will also increase, because as the number of sample increases, the OLS regression will get closer to the population regression Function.

(c). Because LPM is a linear case, therefore, 0.073 is simply the coefficient on the marginal effect of gender on the propability of smoking. However, both probit and logit models are nonlinear, and the marginal effect depends on all the estimates and their values. Therefore, we need to consider the average partial effect, instead of the coefficient only, if we want to estimate the marginal effect of gender on the probability of smoking in the two non-linear models.

(d) The formula of R-squared is $\frac{SSE}{SST} = 1 - \frac{SSR}{SST}$.

where $SSR = \sum_{i=1}^{n} \hat{u}_i^2$ and $SST \sum_{i=1}^{n} (y_i - \bar{y})^2$

where $\hat{u}_i = y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \cdots \hat{\beta}_k x_{ik}$, therefore, SSR could not be larger than SST, and R-squared can't be < 0.

The function of R-adjust is $1 - \frac{SSR/(n-k-1)}{SST/(n-1)} = 1 - \frac{SSR}{SST} \cdot \frac{n-1}{n-k-1}$

Depending on the values of n, k, it's possible that R-adjust < 0

Question 2:

(a) $\frac{\partial cor}{\partial inc} = 0.0124 - 2 \times 0.000062 inc$

$\partial cor = 0.0124 - 2 \times 0.000062 \times 80000/1000$

$= 0.00248$ Therefore, the probability decrease by 248%

(b). $T\text{-test} = \frac{0.0035}{0.0025} = 1.4$ The critical value is 1.96

$1.4 < 1.96$

Therefore, we fail to reject the null hypothesis, and the gender difference is statistically insignificant at 5% level

(c). In this regression model the base group is single female, and the cofficients on the dummy of a group is the difference in the intercepts between that group and the base group. Therefore, if we want to compare the coefficient of married female and a single male, we should set up the null hypothesis as

$H_0: \beta_6 - \beta_8 = 0$

$H_1: \beta_6 - \beta_8 \neq 0$

Question 3

(a). Holding other factors fixed. for those who have had two years of formal training and have passed the test for the RMD. their hourly earnings will earn 20% more

(b) Regression model (2)

$$\log(he) = \beta_0 + \beta_1 EXP + \beta_2 RMC + \beta_3 RMD$$

To test whether the earnings of RMD are no higher than those with RMC, we could run the linear hypothesis $H_0: \beta_3 \le 0$ $H_1: \beta_3 > 0$

(c). The model is still $\log(he) = \beta_0 + \beta_1 EXP + \beta_2 RMC + \beta_3 RMD$

The difference between RMC and RMD: $\beta_2 - \beta_3$

The difference between RMC and people with no formal training: β_2

Therefore, the null hypothesis: $H_0: \beta_2 - \beta_3 = \beta_2$

$$H_1: \beta_2 - \beta_3 \neq \beta_2$$

And we use linear hypothesis to test the hypothesis above

(d). Unrestricted model: $\log(he) = \beta_0 + \beta_1 EXP + \beta_2 Training + \beta_3 RMC + \beta_4 RMD$ (3)

restricted model: $\log(he) = \beta_0 + \beta_1 EXP + \beta_2 training$ (1)

$H_0: \beta_2 = 0$

$$F = \frac{(R_{ur}^2 - R_r^2)/q}{(1-R_{ur}^2)/(n-k-1)} = \frac{(0.42-0.35)/2}{(1-0.35)/(1454-1)} = \frac{0.035}{0.00034} = 10.29$$

$c=3$, therefore, $F > c$, Ho is rejected, and RMC and RMD are jointly statistically significant and at least one of the variables has impact on y

0.7, -0.1

Question 4:

The OLS estimator

(1). $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_i + \hat{\beta}_2 z_i$

(2). $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 v_i + \hat{\beta}_2 w_i = \hat{\beta}_0 + \hat{\beta}_1(\hat{x}_i + \hat{z}_i) + \hat{\beta}_2(x_i - z_i) = \hat{\beta}_0 + (\hat{\beta}_1 + \hat{\beta}_2) x_i + (\hat{\beta}_1 + \hat{\beta}_2) z_i$

(3). $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 v_i = \hat{\beta}_0 + \hat{\beta}_1(\hat{x}_i + \hat{z}_i) = \hat{\beta}_0 + \hat{\beta}_1 \hat{x} +$

(4). $\hat{y} = \hat{\beta}_0 + \hat{\beta}_2 z_i + \hat{\beta}_3 v_i = \hat{\beta}_1(z_i) + \hat{\beta}_2(x_i + z_i) = \hat{\beta}_2 x_i + (\hat{\beta}_1 + \hat{\beta}_2) z_i$

from (2). we could see that $\hat{\beta}_1 + \hat{\beta}_2 = 0.6$, $\hat{\beta}_1 - \hat{\beta}_2 = 0.8 \Rightarrow \hat{\beta}_1 = 0.7$, $\hat{\beta}_2 = -0.1$

Therefore: A = 0.7, C = -0.1

$R^2 = 1 - \frac{SSR}{SST} \Rightarrow 0.60 = 1 - \frac{200}{SST}$

$\frac{200}{SST} = 0.4$ $\quad SST = 500$

Since SST is constant

$G = 1 - \frac{220}{500} = 0.56$

so G = 0.56

From (4). $\hat{\beta}_2 = 0.6$, $\hat{\beta}_1 + \hat{\beta}_2 = 0.8$

A = 0.7
B
C = -0.1
D
E
F
G = 0.56
H
I
J
K
L
M