Ulaş Eraslan 25058

ECON 301 - Problem Set 2

Question 1

1- If x1 and x2 are positively correlated with each other and we omit x2 while estimating the regression of y on x1, there will be a bias on β1(Tilda). Because since x1 is positively correlated with x2 and x2 has effect on y, x1 includes the information of x2 as well. Therefore β1(Tilda) will be bigger than β1 (hat).

2- Since x1 and x2 are highly correlated, the R-squared value will be higher, this cause an increase of variance. On the other hand, since sum of squared residuals is smaller due to not omitting x2, this will cause a decrease in variance of error term. Hence, if we consider the formula of standard error (which is the square root of variance), the final result is unclear.

3- If new added independent variables are not linearly dependent with x1 (small correlation) then the R-squared of x1 will be smaller. Therefore standard error of β1(Hat) will be decrease.

4- If the sample size increased by four times, the variance of the error term will be decreased. Because we will divide sum of squared residuals by bigger number. In this case, standard error of β1(Hat) will be decreased by around 50%.

5- t-test shows how many standard deviations of the calculated coefficient is away from zero.

Case 2: Since the final result of standard error of β1(Hat), it is not clear to give a concrete answer to this case. If standard error increases, then t-statistic decreases so it becomes less statistically significant (Wider confidence interval). Else if standard error decreases then t-statistic increases so this means statistically significance (Narrower confidence interval).

Case 3 : With more independent variable which are not highly correlated with x1, standard error of β1(Hat) will be decreased. Hence, t-statistic will be increased which means it becomes more important. In terms of confidence interval, confidence interval will be narrower.

Case 4 : When sample size is increased, standard error of β1(Hat) decreases. So, t-statistic will be higher which leads to a higher statistical significance. In terms of confidence interval, it will be narrower.

Ulaş Eraslan 25058

Question 2

1- psoda = 0.9563196 + 0.1149882 *prpblck + 1.60e-06 * income + u

R^2 = 0.0642
Sample size (n) = 401

Between price of soda and proportion of black people, there is a positive correlation. If the proportion of black people increases one percentage point, price of soda increases by 0.11 dollars.

2- Since t-statistic (4.42) is greater than 1.96 (5% significance level), we can say that it is statistically significance.

If we are looking for the minimum significance level that we can reject the null hypothesis, we need to check p-value. So, the p-value for prpblck is almost 0.

If H0 = 0.1 => t-stat = (β1(Hat) - 0.1)/ se(β1(Hat)) = (0.1149882 - 0.1) / 0.0260006 = 0.57

Since t-statistic is not greater than 1.96 (5% significance level), coefficient is not significantly different from 0.1 at 0.05 significance level.

3- When we omit the income effect, the discrimination effect is smaller. The omitted independent variable correlated with dependent variable, if we omit it and put it to the unobserved variables, its effect will be seen from discrimination effect. (Omitted variable bias)

4- log(psoda) = -0.793768 + 0 .1215803 *prpblck + 0.0765114*log(income) + u

If prpblck increases by 20 percentage points, price of soda increases by 100* 0 .1215803*0.2 = 2.431606 %.

5- psoda = 0.9278289 + 0.0888819 *prpblck + 2.03e-06 *income + 0.157464 *prppov+u,

I would not expect a positive correlation between proportion in poverty and price of soda. Because in poor regions price of soda should be cheaper to be consumed.

”Because income and prppov are so highly correlated, they have no business being in the same regression”

Even they are highly correlated, if they are linearly independent this situation will not cause a collinearity. Instead, it is better to use two related variables separately.

Ulaş Eraslan 25058

Question 3

1- voteA = 45.07893 + 6.083316 * log(expendA) + -6.615417 * log(expendB) + 0.1519574 prtystrA + u

N = 173 R^2 = 0.7926

If campaign expenditures of A is increased by 1%, vote of A increases by 0.06 percentage point. So the direction is positive.
I do not think so that this is a causal effect because to say that we have to be sure about zero conditional mean condition which leads to exogeneity. But in this context there are many unobserved factors that most probably carry information about the expenditures of A. These factors might be the amount of donation to party A or total budget of that candidate's party and etc.

2- R-squared is around 79% which is a relatively high yet this does not imply that there is a causal relationship between independent variable and dependent variable. This is not an evidence for causal relationship because of the endogeneity problem.

3- T-Value

The t-value of expenditure of A is 15.92 which is greater than 1.96 (5% significance level) therefore this estimations is statistically significant.

P-Value

The p-value is almost 0 which is smaller than 0.05 level. Therefore, estimation is statistically significant.

Confidence Interval

Reported confidence interval is (5.328914 - 6.837719), since it does not contain 0 in it self, we can state that estimated coefficient is statistically significant.