Homework 2 Answers

Note: Please respect intellectual property and do not make this document available in the public domain

1. 17.8 months more than Person B's

2. 15.18 months of education

3. Holding average fees charged by schools in the person's neighbourhood, the distance of the individual's residence to the nearest tertiary institution, the education of the individual's father, and the individual's gender and race constant, at the 1% significance level, the results suggest that income of parents is not associated with the amount of education a person receives.

Elaboration: The question asks students to identify "which of the following statements is incorrect". Many students answered that "none of the other options is incorrect". However, that is not the answer. The answer is that "Holding average fees charged by schools in the person's neighbourhood, the distance of the individual's residence to the nearest tertiary institution, the education of the individual's father, and the individual's gender and race constant, at the 1% significance level, the results suggest that income of parents is not associated with the amount of education a person receives". This statement is incorrect because although the coefficient on *high_income* itself is seemingly only significant at the 1.2% level and the coefficient on *int_fhighinc* itself is seemingly only significant at the 20% level, the joint F-test testing the coefficients on *high_income* and *int_fhighinc* being both equal to zero is rejected at the 1% level (the F-statistic is 11.71 and the p-value of the F-statistic is =0.0000; rounded to 4 decimal places). Hence, evaluated at the 1% significance level, the F-test result tells us that there *is* evidence that either *high_income* is non-zero or that *int_fhighinc* is non-zero (note that this interaction term also comprises high_income) or that both of them are non-zero. Since *high_income* and *int_fhighinc* are both functions of parental income, this tells us that at the 1% significance level, parental income *is* associated with education. The reason the terms *high_income* and *int_fhighinc* seem to be insignificant individually at the 1% level is because of imperfect multicollinearity (these two variables are highly correlated). However, the F-test is able to see through this imperfect multicollinearity and provides the correct conclusion regarding the association between parental income and education.

4. The coefficient on *fees* and *fees2* would now be 3.712135 and -0.1967065 respectively.

5. 0.1110

6. 8.095

7.
$\hat{\beta}_{lndistance}$ suffers from a positive bias when *minority* is omitted from the regression

Elaboration: The estimated coefficient on *lndistance* changes from -0.5243943 to -0.7379352 (i..e becomes more negative) when *minority* is included as a regressor. This suggests that $\hat{\beta}_{lndistance}$ suffers from a positive bias when *minority* is omitted from the regression (i.e the coefficient on *lndistance* is more positive than it should be when *minority* is omitted from the regression).

8. It suggests that the variables *minority* and *lndistance* are negatively correlated.

Elaboration: From Q7, we know that $\hat{\beta}_{lndistance}$ suffers from a positive bias when *minority* is omitted from the regression (*lndistance* and the error term are positively related when *minority* is omitted from the regression). At the same time, we are now told that the true causal effect of being of a minority race is to decrease educational attainment, and this means that the true coefficient on *minority* is *negative*. If *minority* is omitted from the regression, then it goes into the error term with a negative sign. If the relationship between *lndistance* and the error term is positive, and minority enters the error term negatively, this together means that *lndistance* and *minority* are negatively correlated.

9. Coefficient and standard error on *low_income* would be -3.46022 and 1.380232 respectively. Coefficient and standard error on *int_flowinc* would be -2.358989 and 1.842266 respectively.

10. None of the other options are true.

Elaboration: Here, holding the other variables constant, a 1% decrease in the distance of a person's home to the nearest tertiary institution is on average associated with a (0.738 x 0.01 =) 0.00738 month increase in education received. The negative relationship is statistically significant at the 5% level. The key thing here is to be aware that you have to multiply the coefficient on *lndistance* by 0.01 since we have a linear-log model.

11. (12.18,   15.08)

12. Holding the other variables constant, at the 5% significance level, we cannot reject that the effect of having a university educated father on educational attainment is the same as the effect of having a university educated mother.

13. Science test score is predicted to decrease by 0.52%.

14. None of the other options are correct.

Science test score is predicted to fall by 0.65%

15. None of the other options are correct.

Elaboration: many students are inclined to choose the option where *pared* is removed. We should never do that as this may cause us to fall into a trap of a man-made omitted variables bias situation. Remember, that the true functional form relating parental education and Science test score is unknown. We are guessing that they are related by a quadratic, but this itself could be incorrect. The true functional form could be some other function like a cubic or a quartic, or even some other non-linear functional form. When we run regressions, we always include all the powers, not just pared2, without pared. Suppose pared did exist in the true model, this means that it is determinant of the dependent variable. At the same time, pared2 and pared would necessarily be correlated since we are just taking the same variable to a different power, so we can expect a lot of correlation between them. If pared did exist in the true model and we unwittingly omit it from the regression, this would cause the estimator of the coefficient on pared2 to be biased. Hence, when running polynomial regressions, always include all the powers. If the coefficient on a certain variable is zero, we'll just see it's coefficient to be statistically insignificant, but at least, we'll be safe from the danger of inadvertently creating a man-made omitted variables bias situation for ourselves.

16. Holding socioeconomic status and father's education constant, at the 5% significance level, there is evidence that the effect of out of school study-time on Science test score depends on whether the student is an immigrant.

17. An increase in the real beer tax by $2 per case is associated with an increase in the traffic fatality rate by 0.0709 deaths per 10,000 people. The association between traffic fatality and beer tax is however not significant at the 80% level.

18. The estimated coefficient on the beer tax variable is -0.40559. However, the relationship between *beertax* and *fatalityrate* is not significant at the 20% level.

19. on average, women with more than two children work are estimated to work 0.92 weeks per year more compared to women with only two children. However, the estimated effect is not statistically significant at the 20% level.

20. 97.8667