Assignment #5

ECON 2023 Introductory Econometric

April 2, 2023

1 Multiple Regression

1. We've estimated the equation

$$\begin{aligned}\widehat{sleep} =& 3,638.25 - 0.148 totwrk - 11.13 educ + 2.20 age \\ & (112.28) \quad (0.017) \qquad\qquad (5.88) \qquad (1.45), \\ & n = 706, R^2 = 0.113,\end{aligned}$$

where we now report standard errors along with the estimate

   (a) Is either *educ* or *age* individually significant at the 5% level against a two-sided alternative? Show your work.

   (b) Dropping *educ* and *age* from the equations gives

$$\begin{aligned}\widehat{sleep} =& 3.586.38 - 0.151 totwrk \\ & (38.91) \quad (0.017) \\ & n = 706, R^2 = 0.103,\end{aligned}$$

   Are *educ* and *age* jointly significant in the original equation at the 5% level? Justify your answer.

   (c) Does including *educ* and *age* in the model greatly affect the estimated trade off between sleeping and working?

   (d) Suppose that the sleep equation contains heteroskedasticity. What does this mean about the tests computed in parts (a) and (b)?

2. Consider the multiple regression model with three independent variables, under the classical linear model assumptions MLR. 1 through MLR.6:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + u$$

You would like to test the null hypothesis $H0 : \beta_1 - 3\beta_2 = 1$

   (a) Let $\hat{\beta}_1$ and $\hat{\beta}_2$ denote the OLS estimators of β_1 and β_2. Find $Var(\hat{\beta}_1 - 3\hat{\beta}_2)$ in terms of the variances of $\hat{\beta}_1$ and $\hat{\beta}_2$ and the covariance between them. What is the standard error of $\hat{\beta}_1 - 3\hat{\beta}_2$?

   (b) Write the t statistic for testing $H0 : \beta_1 - 3\beta_2 = 1$.

   (c) Define $\theta_1 = \beta_1 - 3\beta_2$ and $\hat{\theta}_1 = \hat{\beta}_1 - 3\hat{\beta}_2$ Write a regression equation involving $\beta_0, \theta_1, \beta_2$ and β_3 that allows you to directly obtain $\hat{\theta}_1$ and its standard error.

2 Software Problem Set

1. Use the data in **DISCRIM.dta** to answer this question.

   (a) Use OLS to estimate the model

   $$\log(psoda) = \beta_0 + \beta_1 prpblck + \beta_2 \log(income) + \beta_3 prppov + u,$$

   and report the results in the usual form. Is $\hat{\beta}_1$ statistically different from zero at the 5% level against a two-sided alternative? What about at the 1% level?

   (b) What is the correlation between $\log(income)$ and $prppov$? Is each variable statistically significant in any case? Report the two-sided p-values.

   (c) To the regression in part (a), add the variable $\log(hseval)$. Interpret its coefficient and report the two-sided p-value for $H0 : b\beta_{\log(hseval)} = 0$.

   (d) In the regression in part (c), what happens to the individual statistical significance of $\log(income)$ and $prppov$? Are these variables jointly significant? (Compute a p-value.) What do you make of your answers?

   (e) Given the results of the previous regressions, which one would you report as most reliable in determining whether the racial makeup of a zip code influences local fast-food prices?

2. Use the data in **HTV.dta** to answer this question.

   (a) Estimate the regression model

   $$educ = \beta_0 + \beta_1 motheduc + \beta_2 fatheduc1 + \beta_3 abil + \beta_4 abil^2 + u$$

   by OLS and report the results in the usual form. Test the null hypothesis that $educ$ is linearly related to $abil$ against the alternative that the relationship is quadratic.

   (b) Using the equation in part (a), test $H_0 : \beta_1 = \beta_2$ against a two-sided alternative. What is the p-value of the test?

   (c) Add the two college tuition variables to the regression from part (a) and determine whether they are jointly statistically significant.

   (d) What is the correlation between $tuit17$ and $tuit18$? Explain why using the average of the tuition over the two years might be preferred to adding each separately. What happens when you do use the average?

   (e) Do the findings for the average tuition variable in part (d) make sense when interpreted causally? What might be going on?