Note: In all the exercises assume that the assumptions of the classical linear model are satisfied.

1. The data set in CEOSAL2 contains information on chief executive officers for U.S. corporations. The variable salary is annual compensation, in thousands of dollars, and ceoten is prior number of years as company CEO.

   (a) Propose a model for which a one-year increase in the post of CEO always implies the same percentage increase in salary.
   
   (b) Estimate the model of this section a and present the parameter estimates, standard errors, the number of observations and the $R^2$ in an equation.
   
   (c) Interpret the slope of the estimated equation.
   
   (d) Test whether a one-year increase in the post of CEO leads to a 1% increase in the CEO’s salary. Calculate the p-value and interpret the result.

2. We are interested in studying cigarette demand by young adults. Let’s consider the following model

   $$ npack = \beta_0 + \beta_1 price + \beta_2 age + \beta_3 income + u $$

where npack is monthly cigarette consumption in packs, price is the price of a pack of cigarettes in dollars, age is age in years, and income is annual income in thousands of dollars.

   (a) What is the expected sign of $\beta_1$?
   
   (b) Based on a sample of 30 individuals, the following estimated model is obtained:

   $$ \hat{npack} = -11.74 - 0.53 price + 0.99 age + 0.08 income $$

   (b1) Interpret the estimated coefficients.
   
   (b2) Test the individual significance of each variable.
   
   (c) Now consider the model

   $$ \log(npack) = \beta_0 + \beta_1 \log(price) + \beta_2 age + \beta_3 income + u $$
c1) Using the same sample above we have obtained the following results:

\[ \log(npack) = 1.17 - 0.69 \log(price) + 0.06 \log(age) + 0.009 \text{income} \]

Interpret the estimated coefficients.

(c2) Test whether there is enough evidence to say that an increase of 1000 dollars in income, holding price and age constant, increases cigarette consumption by less than 1%.

3. The following model allows the return to education to depend upon the total amount of both parents’ education, called \( \text{pareduc} \):

\[ \log(wage) = \beta_0 + \beta_1 \text{educ} + \beta_2 \text{pareduc} \times \text{educ} + \beta_3 \text{exper} + \beta_4 \text{tenure} + u \]

(a) Show that the return of one more year of education in this model is

\[ 100(\beta_1 + \beta_2 \text{pareduc}) \]

What is the expected sign of \( \beta_2 \)? Why?

(b) Using the data in WAGE2 from Wooldridge book, the estimated equation is

\[
\begin{align*}
\log(wage) & = 5.65 + 0.047 \text{educ} + 0.00078 \text{pareduc} \times \text{educ} + 0.019 \text{exper} + 0.010 \text{tenure} \\
\hat{u} & \sim N(0, 0.004) \\
n & = 722, \quad R^2 = 0.169
\end{align*}
\]

(Only 722 observations contain full information on parents’ education). Interpret the coefficient on the interaction term. It might help to choose two specific values for \( \text{pareduc} \)—for example, \( \text{pareduc} = 32 \) if both parents have a college education, or \( \text{pareduc} = 24 \) if both parents have a high school education—and to compare the estimated return to \( \text{educ} \).

(c) When \( \text{pareduc} \) is added as a separate variable to the equation, we get:

\[
\begin{align*}
\log(wage) & = 4.94 + 0.097 \text{educ} + 0.033 \text{pareduc} - 0.0016 \text{pareduc} \times \text{educ} \\
& + 0.020 \text{exper} + 0.010 \text{tenure} \\
\hat{u} & \sim N(0, 0.004) \\
n & = 722, \quad R^2 = 0.174
\end{align*}
\]

Does the return to education now depend positively on parental education? Test the null hypothesis that the return to education does not depend on parental education.

4. Consider the following model to explain the percentage of students receiving a passing score on a math test \( \text{math4} \) at Michigan schools

\[ \text{math4} = \beta_0 + \beta_1 \log(\text{expend}) + \beta_2 \log(\text{enroll}) + \beta_3 \text{lunch} + u \quad (1) \]
where \textit{expend} is total school expenditure, \textit{enroll} is the number of students at the school and \textit{lunch} is the percentage of students who are eligible for the school free lunch program. Using a sample of 1823 schools we have found the following results:

\[
\mathit{math4} = 46.19 + 8.53 \log(\text{expend}) - 13.37 \log(\text{enroll}) - 0.471 \text{lunch}
\]

\[
n = 1823, \quad R^2 = 0.380
\]

(a) Interpret the estimated coefficient on \(\log(\text{expend})\). Why do you think that the estimated coefficient on \(\text{lunch}\) is negative?

(b) Test the overall significance of the regression.

(c) Next year, the number of students is expected to increase by 1\%. An economist states that if the total school expenditure also increases by 1\%, the 1\% increase in the number of students will have no effect on the percentage of students who pass the exam. Write down the null and the alternative hypotheses to test the economist’s statement. What would be the restricted model after imposing the null hypothesis?

(d) Consider now the following estimated model

\[
\mathit{math4} = -6.34 + 11.34 \log(\frac{\text{expend}}{\text{enroll}}) - 0.471 \text{lunch} \quad (2)
\]

If variables \(\log(\frac{\text{expend}}{\text{enroll}})\) and \(\text{lunch}\) are positively correlated and \(\text{lunch}\) is omitted, do you think that the estimated coefficient on \(\log(\frac{\text{expend}}{\text{enroll}})\) will be larger or smaller than the one it appears in equation (2)? Explain.

5. Consider the following model to explain the number of children that a woman has (\textit{children})

\[
\text{children} = \beta_0 + \beta_1 \text{age} + \beta_2 \text{educ} + \beta_3 \text{heduc} + u
\]

where \textit{age} is the age of the woman, \textit{educ} is the years of education of the woman and \textit{heduc} is the years of education of her husband.

(a) Use the data in FERTIL2 from Wooldridge book to estimate the model above and present the results in an equation. Interpret the estimated coefficient on \textit{age}.

(b) Calculate a 95\% confidence interval for \(\beta_1\).

(c) Holding the age of the woman fixed, test whether increasing by one year the education of the woman has the same effect on the number of children than increasing by one year the education of her husband.

(d) Holding the age of the woman fixed, test that neither the years of education of the woman nor the years of education of her husband affect the number of children that the woman has.
(e) Suppose that the variance of the number of children that a woman has depends on her years of education, what are the consequences for the tests above? Explain.

6. Let us consider the following model to explain housing prices

\[ \log(price) = \beta_0 + \beta_1 \text{sqrft} + \beta_2 \text{bdrms} + u \]

where \( price \) is the price of the house in thousands of dollars, \( \text{sqrft} \) is the size of the house in square feet (1 square foot is 0.093 \( m^2 \)), and \( \text{bdrms} \) is the number of bedrooms.

(a) Use the data in the file HPRICE1 of Wooldridge's book to estimate the model and present the results in the usual way.

(b) What is the estimated percentage change in the price of a house with an additional 150 square feet in size, holding the number of rooms constant?

(c) Obtain a 95% confidence interval for the percentage change in the price of a house with an additional 150 square feet in size, holding the number of rooms constant.

(d) What is the estimated percentage change in the price of a house with an additional bedroom that is 150 square feet in size?

(e) Obtain a 95% confidence interval for the percentage change in the price of a house with an additional bedroom that is 150 square feet in size.

7. Consider a model where the return to education depends upon the amount of work experience (and vice versa):

\[ \log(wage) = \beta_0 + \beta_1 \text{educ} + \beta_2 \text{exper} + \beta_3 \text{educ} \times \text{exper} + u \]

where \( wage \) is the monthly salary in hundreds of dollars, \( \text{educ} \) is the years of education of the individual, and \( \text{exper} \) is the years of work experience.

(a) Show that the return to another year of education, holding \( \text{exper} \) constant, is

\[ 100(\beta_1 + \beta_3 \text{exper}) \]

(b) State the null hypothesis that the return to education does not depend on \( \text{exper} \). What do you think is the appropriate alternative?

(c) Use the data from the file WAGE2 of Wooldridge's book to test the null hypothesis of \( b \) against your stated alternative.

(d) Obtain a 95% confidence interval for the return to education when \( \text{exper} = 10 \).
8. Use the data from the file GPA2 of Wooldridge’s book for this exercise.

(a) Estimate the model
\[ sat = \beta_0 + \beta_1 hsize + \beta_2 hsize^2 + u, \]
where \( hsize \) is the size of graduating class (in hundreds) and \( sat \) is the result of a test. Write the results in the usual form.

(b) Test whether the result of the test depends linearly on the size of graduating class versus a two-sided alternative.

(c) Using the estimated equation in a, what would be the optimal class size? Explain your answer.

(d) Find the estimated optimal class size, using \( \log(sat) \) as the dependent variable. Is it much different from what you obtained in part c?

9. The following model can be used to study whether campaign expenditures affect election outcomes:
\[ voteA = \beta_0 + \beta_1 \log(expendA) + \beta_2 \log(expendB) + \beta_3 prtystrA + u \]
where \( voteA \) is the percentage of votes obtained by candidate A, \( expendA \) and \( expendB \) are the campaign expenditures by candidates A and B, and \( prtystrA \) is a measure of party strength for candidate A.

(a) What is the interpretation of \( \beta_1 \)?

(b) Specify, in terms of the model parameters, the hypothesis that the effect of a 1% increase in campaign expenditures by A on the percentage of votes for candidate A is offset by an increase of 1% in campaign expenditures by B.

(c) Estimate the model using data from the file VOTE1 of Wooldridge’s book about the election outcomes in 173 two-party races for the U.S. House of Representatives in 1988 and present the results in the usual way.

(d) How do the campaign expenditures by candidate A affect the candidate’s election outcomes? And the campaign expenditures by B?

(e) Test the hypothesis of b against a two-sided alternative using a \( t \) statistic.

(f) Write the restricted model imposing the restriction in b. Estimate the model and use an \( F \) statistic to test the hypothesis of b again. What is the relationship between the \( t \) statistic of e and the \( F \) statistic you just calculated?
10. Consider the following model for wages

\[
\log(wage) = \beta_0 + \beta_1 \text{educ} + \beta_2 \text{feduc} + \beta_3 \text{meduc} + \beta_4 \text{exper} + u
\]  

(1)

where wage are the monthly income in dollars, educ are years of education, feduc are the years of education of the father, meduc are the years of education of the mother and exper are the years of labour market experience.

(a) Estimate the model using the data in file WAGE2 from Wooldridge’s book and present the results in the usual form.

(b) Interpret the estimated coefficient on feduc and test whether this variable is significant.

(c) Interpret the estimated coefficient on meduc and test whether this variable is significant.

(d) Holding education and experience fixed, test that neither the years of education of the father nor the years of education of the mother affect wages.

(e) Holding education and experience fixed, test, versus a two-sided alternative, that the years of education of both the father and the mother have the same effect on the salary of their children.