1. Consider the following model to explain birth weight \((bwght)\)

\[
\log(bwght) = \beta_0 + \beta_1 cigs + \beta_2 \log(faminc) + \beta_3 parity + \beta_4 male + \beta_5 white + u
\]

where \(cigs\) is the average number of cigarettes the mother smoked per day during pregnancy, \(faminc\) is annual family income, \(parity\) is the birth order of this child, \(male\) is a dummy variable indicating whether the child is male, and \(white\) is another dummy indicating whether the child is classified as white. Using data on 1388 births from the datafile BWGHT from Wooldridge book we have obtained the following results:

\[
\begin{align*}
\hat{\log(bwght)} &= 4.66 - 0.0044 \; \text{cigs} + 0.0093 \log(\text{faminc}) + 0.016 \; \text{parity} \\
&\quad + 0.027 \; \text{male} + 0.055 \; \text{white} \\
&\quad (0.022) \quad (0.00085) \quad (0.0059) \quad (0.0056) \quad (0.010) \quad (0.013)
\end{align*}
\]

\(n = 1388, \quad R^2 = 0.047\)

(a) Holding other factors fixed, what is the effect on birth weight from smoking 10 more cigarettes per day?

(b) Holding other factors fixed, how much more is a white child predicted to weigh than a nonwhite child? Is the difference statistically significant?

(c) Holding other factors fixed, how much more is a male child predicted to weigh than a female child? Is the difference statistically significant?

(d) Holding other factors fixed, how much more is a white-male child predicted to weigh than a nonwhite-female child?

2. In order to analyse the wage of a university professor, the following model is used:

\[
wage = \beta_0 + \beta_1 \text{Male} + \beta_2 \text{White} + \beta_3 (\text{Male} \times \text{White}) + \beta_4 \text{exper} + u
\]

where \(wage\) is the annual wage of the professor, \(exper\) are the years of experience, \(Male\) is a dummy variable the takes value 1 if the professor is male and 0 if she is female, \(White\) is another dummy variable the takes value 1 if the professor is white and 0 otherwise.

(a) Determine the average wage for:

- a1) White males.
- a2) White females.
- a3) Nonwhite males.
- a4) Nonwhite females.

(b) What is the difference in the average wage between

- b1) white males and white females with the same years of experience?
- b2) white males and nonwhite males with the same years of experience?
- b3) white females and nonwhite females with the same years of experience?
- b4) nonwhite males and nonwhite females with the same years of experience?
How would you test the hypothesis that the difference in the average wage between whites and nonwhites with the same years of experience is identical for males and females?

3. Suppose you collect data from a survey on wages, education, experience, and gender. In addition, you ask for information about marijuana usage. The original question is: “On how many separate occasions last month did you smoke marijuana?”

(a) Write an equation that would allow you to estimate the effects of marijuana usage on wage, while controlling for other factors. You should be able to make statements such as, “Smoking marijuana one more time per month is estimated to change wage by x%.”

(b) Write a model that would allow you to test whether drug usage has different effects on wages for men and women. How would you test that there are no differences in the effects of drug usage for men and women?

(c) Suppose you think it is better to measure marijuana usage by classifying people into one of the four following categories: nonuser, light user (1 to 5 times per month), moderate user (6 to 10 times per month), and heavy user (more than 10 times per month). Now write a model that allows you to estimate the effects of marijuana usage on wage.

(d) Using the model in part c, explain in detail how to test the null hypothesis that marijuana usage has no effect on wage. Be very specific and indicate the null and alternative hypothesis, the test statistic and its distribution under the null, and the critical region.

4. To test the effectiveness of a job training program on the subsequent wages of workers, we specify the model

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

where train is a binary variable equal to unity if a worker participated in the program. Think of the error term u as containing unobserved worker ability. If less able workers have a greater chance of being selected for the program, what can you say about the likely bias of the OLS estimator of \( \beta_1 \)?

5. Suppose we are interested in analyzing potential differences in beer consumption by gender. To do so we specify the following model:

\[ \text{beer} = \beta_0 + \beta_1 \text{income} + \beta_2 \text{female} + \beta_3 (\text{femaleincome}) + u \]

where beer is annual beer expenditure in euros, income is annual income in thousands of euros, Female is a dummy variable that takes value 1 for females and femaleincome = female * income. Using a sample of 34 individuals we have obtained the following results:

\[ \hat{\text{beer}} = 186.47 + 2.3 \text{income} - 126.00 \text{female} - 1.3 (\text{femaleincome}) \]

\[ n = 34, \quad R^2 = 0.5055 \]

Moreover, using the same sample we have also estimated the following models

\[ \text{beer} = \beta_0 + \beta_1 \text{income} + \beta_3 (\text{femaleincome}) + u \]

finding \( R^2 = 0.3445 \)

\[ \text{beer} = \beta_0 + \beta_1 \text{income} + \beta_2 \text{female} + u \]

finding \( R^2 = 0.2903 \)

\[ \text{beer} = \beta_0 + \beta_1 \text{income} + u \]

finding \( R^2 = 0.1355 \).
(a) What is the estimated difference in beer expenditure between males and females with annual income of 25000 euros?.

(b) Assuming that the errors are normal, test the following hypothesis
   b1) Once we control for income, there are no differences in beer expenditures by gender.
   b2) Once we control for income, the marginal propensity to consume beer is higher for males than for females.

6. Consider the following model

\[ sat = \beta_0 + \beta_1 hsize + \beta_2 hsize^2 + \beta_3 female + \beta_4 black + \beta_5 female \cdot black + u \]

where \( sat \) is the combined SAT score, \( hsize \) is size of the student’s high school graduating class, in hundreds, \( female \) is a gender dummy variable, and \( black \) is a race dummy variable equal to one for blacks, and zero otherwise.

Using the data in GPA2 from Wooldridge book we have obtained the following results:

\[
\begin{align*}
\hat{sat} &= 1028.1 + 19.3 hsize - 2.19 hsize^2 - 45.09 \text{female} \\
&\quad - 169.8 \text{black} + 62.30 \text{female} \cdot \text{black} \\
&\quad (6.29) \quad (3.83) \quad (0.527) \quad (4.29) \quad (12.71) \quad (18.15) \\
n &= 4137, \quad R^2 = 0.0858
\end{align*}
\]

(a) Is there strong evidence that \( hsize^2 \) should be included in the model? From this equation, what is the optimal high school size?

(b) Holding \( hsize \) fixed, what is the estimated difference in SAT score between nonblack females and nonblack males? Is this difference statistically significant?

(c) Holding \( hsize \) fixed, what is the estimated difference in SAT score between nonblack males and black males? Is this difference statistically significant?

(d) Holding \( hsize \) fixed, what is the estimated difference in SAT score between black females and nonblack females? What would you need to do to test whether the difference is statistically significant?

7. Consider the following model

\[ \log(salary) = \beta_0 + \beta_1 \log(sales) + \beta_2 finance + \beta_3 consprod + \beta_4 utility + u \]

where \( salary \) is the annual wage of the executive director of the firm in thousands of dollars, \( sales \) are the annual sales of the firm in millions of dollars and \( finance, consprod \) and \( utility \) are binary variables indicating the sector of the firm (the financial sector, consumption sector and the service sector, respectively). The omitted sector is the industrial sector.

(a) Estimate the model using the data in CEOSAL1 from Wooldridge book and present the results in an equation.

(b) Test whether the wage of the executive directors depends on the sector of the firm.

(c) Holding the sales fixed, compute the average percentage difference in the estimated wage between the service sector and the industrial sector. Is this difference statistically significant at 1%?

(d) Holding sales fixed, what is the average difference in the estimated wage between the consumption sector and the financial sector? Test if this difference is statistically significant.
8. Consider the model

\[
\text{colgpa} = \beta_0 + \beta_1 \text{hsize} + \beta_2 \text{hsize}^2 + \beta_3 \text{hsperc} + \beta_4 \text{sat} + \beta_5 \text{female} + \beta_6 \text{athlete} + u
\]

where \( \text{colgpa} \) is the cumulative grade point average at university, \( \text{hsize} \) is the number of students in the cohort of high school (in hundreds), \( \text{hsperc} \) is the percentile in the high school graduating class (defined so that, for example, \( \text{hsperc} = 5 \) means the top 5% of the class), \( \text{sat} \) is the mark of the SAT test of school aptitude, \( \text{female} \) is a dummy variable taking value 1 if the student is a female, \( \text{athlete} \) is another dummy variable taking value 1 if the student is an athlete.

(a) Estimate the model using the data in file GPA2 of Wooldridge and present the results in an equation. What is the estimated difference in the average mark of university between athletes and those that are not athletes? Is this difference statistically significant?

(b) Omit \( \text{sat} \) from the model and reestimate the equation. What is the estimated difference now due to the fact that the individual is an athlete? Explain why this estimate is different from the one obtained in part (a).

(c) Now consider a model which allows for the effect of being an athlete on the average mark to be different according to the sex of the student and test the null hypothesis that, keeping the rest of the factors fixed, the difference in the average grade between athletes and non-athletes is the same for males and females (include also \( \text{sat} \) in the equation)

(d) Consider now a model which allows for the effect of \( \text{sat} \) on the average mark at university to be different for males and females, is there enough evidence to state that the effect of \( \text{sat} \) on the mark at the university is different for males and females?

9. Consider the following model relating the time devoted to sleep and the time devoted to work, together with other factors affecting sleep:

\[
\text{slepph} = \beta_0 + \beta_1 \text{totwrkh} + \beta_2 \text{educ} + \beta_3 \text{age} + \beta_4 \text{age}^2 + \beta_5 \text{yngkid} + u
\]

where time devoted to sleep (\( \text{slepph} \)) and the total time devoted to work (\( \text{totwrkh} \)) are measured in hours per week (take into account that the variables in the file are \( \text{slepp} \) and \( \text{totwrkh} \), both measure in minutes per week so that the variables \( \text{slepph} \) and \( \text{totwrkh} \) must be generated before estimating the model). The level of education (\( \text{educ} \)) and age (\( \text{age} \)) are measured in years and \( \text{yngkid} \) is a binary variable taking value 1 if the individual has kids below 3 years old.

(a) Estimate separately the equation for males and females using the data in file SLEEP75 of Wooldridge and present the results in an equation. Are there large differences between the two estimated equations?

(b) Compute the Chow test for the equality of parameters in the equation for females and males. What are the degrees of freedom relevant for this test? Should \( H_0 \) be rejected at 5%?

(c) Estimate the model including all the variables, the dummy for male (or for female) and the interactions between all the variables and the dummy for male (or for female) and check that the \( SSR \) of this model coincides with the \( SSR \) of the equations separately estimated for males and females obtained in part (a).

(d) Allowing the constant terms to be different for males and females, test now if there are differences in the slopes.