1). Do liberal arts colleges pay economists more than they pay other professors? To find out, we looked at a sample of 2,929 small college faculty members and built a model of their salaries that included a number of variables, four of which were:

\[ S_i = 36721 + 817M_i + 426A_i + 406R_i + 3539T_i + \cdots \]

\[ R^2 = 0.77 \quad n = 2929 \]

where
- \( S_i \) = the salary of the \( i \)th college professor,
- \( M_i \) = a dummy variable equal to 1 if the \( i \)th professor is a male and 0 a female,
- \( A_i \) = a dummy variable equal to 1 if the \( i \)th professor is African American and 0 otherwise,
- \( R_i \) = the years in rank of the \( i \)th professor,
- \( T_i \) = a dummy variable equal to 1 if the \( i \)th professor teaches economics and 0 otherwise.

a. Carefully explain the meaning of the estimated coefficient of \( M \).

b. The equation indicates that African Americans earn $426 more than members of other ethnic groups, holding constant the other variables in the equation. Does this coefficient have the sign you expected? Why or why not?

c. Is \( R \) a dummy variable? If not, what is it? Carefully explain the meaning of the coefficient of \( R \).

d. What’s your conclusion? Do economists earn more than other professors at liberal arts colleges? Explain.

e. Assume that your professor is a white, male economist who has been an assistant professor at your college for three years. How much money does the equation predict that he is earning?

2). The Graduate Record Examination (GRE) subject test in economics is a multiple-choice measure of knowledge and analytical ability in economics that is used mainly as an entrance criterion for students applying to Ph.D. Programs in economics. For years, critics have claimed that the GRE, like the Scholastic Aptitude Test (SAT), is biased against women and some ethnic group. To test the possibility that GRE subject test in economics is biased against women, some economists estimated the following equation:

\[ \text{GRE}_i = 172.4 + 39.7G_i + 78.9GP A_i + 0.203 SATM_i + 0.110 SATV_i \]

\[ n = 149 \quad R^2 = 0.46, \]

where
- \( \text{GRE}_i \) = the score of the \( i \)th student in the Graduate Record Examination subject test in economics,
- \( G_i \) = a dummy variable equal to 1 if the \( i \)th student was a male, 0 otherwise,
\[ \text{GPA}_i = \text{the GPA is economics classes of the } i\text{th student (4=A, 3=B, etc.)} \]

\[ \text{SATM}_i = \text{the score of the } i\text{th student on the mathematics portion of the Scholastic Aptitude Test}, \]

\[ \text{SATV}_i = \text{the score of the } i\text{th student on the verbal portion of the Scholastic Aptitude Test.} \]

a. Carefully explain the meaning of the coefficient of \( G \) in this equation.

b. Does this result prove that the GRE is biased against women? Why or why not?

c. Suppose that the authors had defined their gender variables as \( G_i = \) a dummy variable equal to 1 if the \( i\)th student was a female, 0 otherwise. What would the estimated equation have been in that case?

3). Lagged variables make sense in a regression equation whenever the impact of an independent variable on the dependent variable is likely to be delayed for one reason or another. For example, consider the following equation for the supply of cotton

\[
\hat{C}_t = 623 + 3300W_t + 193P_{t-1} - 154S_{t-1} - 120C_{t-1}
\]

\[ R^2 = 0.50, \quad n = 26(\text{annualUS}) \]

where

\[
C_t = \text{the quantity of cotton produced in year } t,
\]

\[
W_t = \text{cotton-growing weather index (hi=good) in year } t,
\]

\[
P_t = \text{the price of cotton in year } t,
\]

\[
S_t = \text{the price of soybeans in year } t,
\]

\[
C_t = \text{the price of farm inputs in year } t.
\]

a. Think through the theory behind cotton supply. What signs do you expect for the coefficients? Explain. Do the estimated signs agree with your expectations? (Hint: Soybeans and cotton can be grown on the same kind of land.)

b. The regression was run on 26 observations, but the data set for this project actually covers 27 different years. Why?

c. Note that \( P \) and \( S \) are lagged but \( W \) and \( C \) are not. Do you agree with these choices? Explain your answer for each variable.

d. Carefully state the meaning of each regression coefficient. In particular, be sure to specify exactly which variables are being held constant. (Hint: Focus on the lag structure, not the units of measurement.)

4). Michael Lovell estimated the following model of the gasoline mileage of various models of cars

\[
\hat{G}_i = 22.008 - 0.002W_i - 2.76A_i + 3.28D_i + 0.415E_i
\]

\[ R^2 = 0.82 \]

where

\[ G_i = \text{miles per gallon of the } i\text{th model as reported by Consumers' Union based on actual road tests}, \]

\[ W_i = \text{the gross weight (in pounds) of the } i\text{th model}, \]
\( A_i \) = a dummy variable equal to 1 if the \( i \)th model has an automatic transmission and 0 otherwise,
\( D_i \) = a dummy variable equal to 1 if the \( i \)th model has a diesel engine and 0 otherwise,
\( E_i \) = the US Environmental Protection Agency’s estimate of the miles per gallon of the \( i \)th model.

a. Hypothesize signs for the slope coefficients of \( W \) and \( E \). Which if any, of the signs of the estimated coefficients are different from your expectations?

b. Carefully interpret the meanings of the estimated coefficients of \( A_i \) and \( D_i \).

c. Lovell included one of the variables in the model to test a specific hypothesis, but that variable wouldn’t necessarily be in another researcher’s gas mileage model. What variable do you think Lovell added? What hypothesis do you think Lovell wanted to test?

5). Your boss is about to start production of her newest box office smash-to-be, when she calls you in and tells you to build a model of the gross receipts of all the movies produced in the last five years. Your regression is

\[
\hat{G}_i = 781 + 15.4T_i - 992F_i + 1770J_i + 3027S_i - 3160B_i + \cdots
\]

\( R^2 = 0.485, \quad n = 254 \)

where

\( G_i \) = the final gross receipts of the \( i \)th motion picture (in thousands of dollars)
\( T_i \) = the number of screens (theaters) on which the \( i \)th film was shown in its first week
\( F_i \) = a dummy variable equal to 1 if the star of the \( i \)th film is a female and 0 otherwise,
\( J_i \) = a dummy variable equal to 1 if the \( i \)th movie was released in June and July and 0 otherwise
\( S_i \) = a dummy variable equal to 1 if the star of the \( i \)th film is a superstar (like Tom Cruise or Milton) and 0 otherwise
\( B_i \) = a dummy variable equal to 1 if at least one member of the supporting cast of the \( i \)th film is a superstar and 0 otherwise.

a. Hypothesize signs for each of the slope coefficients in the equation. Which, if any, of the signs of the estimated coefficients are different from your expectations?

b. Milton, the star, is demanding $4 million from your boss to appear in the new movie. If your estimates are trustworthy, should she say yes or hire Arnold (a nobody) for $500,000?

c. Your boss wants to keep costs low, and it would cost $1.2 million to release the movie on an additional 200 screens. Assuming our estimates are trustworthy, should she spring for the extra screens?

d. The movie is scheduled for release in September, and it would cost $1 million to speed up production enough to allow a July release without hurting quality. Assuming your estimates are trustworthy, is it worth the rush?