1. **Interpreting regression results**
   For each of the equations listed below, interpret $\hat{\beta}_i$. Be sure to pay close attention to the units used in estimation or the appropriate use of percentage change/percentage points.

   - Unem = unemployment rate
   - Uninsure = Uninsurance rate for health insurance
   - Education = Years of education
   - Wages100 = Average monthly income in 100s
   - Wages = Average monthly income in dollars
   - Ln(Wages) = Ln(Average monthly income in dollars)
   - Ln(Price) = Ln(Housing price)
   - Ln(Dist) = Ln(Distance from house to recently built garbage incinerator)

   $\hat{UNINSUR} = 14.37 + 0.30*UNEM$

   $WAGES = 146.95 + 60.21*EDUC$

   $WAGES100 = 1.47 + 0.60*EDUC$

   $LN(WAGES) = 5.97 + 0.06*EDUC$

   $LN(PRICE) = 9.40 + 0.312*LN(DIST)$

2. (Wooldridge C.2.1)
   The data in 401K.RAW are a subset of data analyzed by Papke (1995) to study the relationship between participation in a 401(k) pension plan and the generosity of the plan. The variable *prate* is the percentage of eligible workers with an active account; this is the variable we would like to explain. The measure of generosity is the plan match rate, *mrate*. This variable gives the average amount the firm contributes to each worker’s plan for each $1 contribution by the worker. For example, if *mrate* = 0.50, then a $1 contribution by the worker is matched by a $0.50 contribution by the firm.

   (i) Find the average participation rate and the average match rate in the sample of plans.

   (ii) Now estimate the simple regression equation and report the results along with the sample size and R-squared.

   $prat\hat{e} = \hat{\beta}_0 + \hat{\beta}_1 mrate$

   (iii) Interpret the intercept in your equation. Interpret the coefficient on mrate.

   (iv) Find the predicted *prate* when *mrate* = 3.5. Is this a reasonable prediction? Explain what is happening here.
(v) How much of the variation in prate is explained by mrate? Is this a lot in your opinion?

3. (adapted from Wooldridge C2.2)

The data set in CEOSAL2.RAW 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.

(i) Produce a scatterplot with salary on the vertical axis and ceoten on the horizontal axis. Does this relationship look linear to you?

(ii) Now plot ln(salary) on the vertical axis and ceoten on the horizontal axis. Compare this scatterplot with the one you created in part (i). Which relationship looks more linear to you?

(iii) Why we might prefer to model ln(salary) over salary? (We discussed two reasons in class.)

(iv) Find the average salary and the average tenure in the sample.

(v) How many CEOs are in their first year as CEO (that is, ceoten = 0)? What is the longest tenure as a CEO?

(vi) Estimate the simple regression model

\[ \log(salary) = \beta_0 + \beta_1 \text{ceoten} + \mu \]

and report your results in the form of an equation. In a sentence, interpret the coefficient on ceoten.

4. Data scaling

Use the dataset WAGE2.RAW. We will use the following variables:

WAGE=Monthly earnings
EXPER=Years of work experience

(i) Create the following variables:
WAGE_100=WAGE/100
EXPER_10=EXPER/10

(ii) Run the following regressions and interpret \( \hat{\beta}_1 \) in each. (In this step, you may interpret \( \hat{\beta}_1 \) using whichever units you find to be most appropriate.)

\[ WAGE = \beta_0 + \beta_1 * \text{EXPER} + \mu \]
\[ WAGE_{100} = \beta_0 + \beta_1 * \text{EXPER}_{10} + \mu \]
\[ WAGE = \beta_0 + \beta_1 * \text{EXPER}_{10} + \mu \]
\[ WAGE_{100} = \beta_0 + \beta_1 * \text{EXPER}_{10} + \mu \]

(iii) Now we will use a standard set of units to interpret these results. For each of the regressions, describe how an additional ten years of education impacts wages, in dollars. Is your finding the same for each regression? Why or why not?