1. (Wooldridge 1.3) A justification for job training programs is that they improve worker productivity. Suppose that you are asked to evaluate whether more job training makes workers more productive. However, rather than having data on individual workers, you have access to data on manufacturing firms in Ohio. In particular, for each firm, you have information on hours of job training per worker (training) and number of nondefective items produced per worker hour (output).

(a) Carefully state the ceteris paribus thought experiment underlying this policy question.

(b) Does it seem likely that a firm’s decision to train its workers will be independent of worker characteristics? What are some of those measurable and unmeasurable worker characteristics?

(c) Name a factor other than worker characteristics that can affect worker productivity.

(d) If you find a positive correlation between output and training, would you have convincingly established that job training makes workers more productive? Explain.

2. (Wooldridge 2.2) The following table contains the ACT scores and the GPA (grade point average) for eight college students. Grade point average is based on a four-point scale and has been rounded to one digit after the decimal.
(a) Estimate the relationship between GPA and ACT using OLS; that is, obtain the intercept and slope estimates in the equation

\[ \hat{\text{GPA}} = \hat{\beta}_0 + \hat{\beta}_1 \text{ACT} \]

Comment on the direction of the relationship. Does the intercept have a useful interpretation here? Explain. How much higher is the GPA predicted to be if the ACT score is increased by five points?

(b) Compute the fitted values and residuals for each observation, and verify that the residuals (approximately) sum to zero.

(c) What is the predicted value of GPA when ACT=20?

(d) How much of the variation in GPA for these eight students is explained by ACT? Explain.

3. (Wooldridge 2.4) Suppose you are interested in estimating the effect of hours spent in an SAT preparation course (hours) on total SAT score (sat). The population is all college-bound high school seniors for a particular year.

(a) Suppose you are given a grant to run a controlled experiment. Explain how you would structure the experiment in order to estimate the causal effect of hours on sat.

(b) Consider the more realistic case where students choose how much time to spend in a preparation course, and you can only randomly sample sat and hours from the
population. Write the population model as

\[ sat = \beta_0 + \beta_1 \text{hours} = u \]

where, as usual in a model with an intercept, we can assume \( E(u) = 0 \). List at least two factors contained in \( u \). Are these likely to have positive or negative correlation with hours?

(c) In the equation from part (b), what should be the sign of \( \beta_1 \) if the preparation course is effective?

(d) In the equation from part (b), what is the interpretation of \( \beta_0 \)?

4. (Wooldridge 2.5) Consider the savings function

\[ sav = \beta_0 + \beta_1 \text{inc} + u, \quad u = \sqrt{\text{inc}} \cdot e, \]

where \( e \) is a random variable with \( E(e) = 0 \) and \( Var(e) = \sigma_e^2 \). Assume that \( e \) is independent of inc.

(a) Show that \( E(u|\text{inc}) = 0 \), so that the key zero conditional mean assumption (Assumption SLR.4) is satisfied. [Hint: If \( e \) is independent of inc, then \( E(e|\text{inc}) = E(e) \).]

(b) Show that \( Var(u|\text{inc}) = \sigma_e^2 \text{inc} \), so that the homoskedasticity Assumption SLR.5 is violated. In particular, the variance of sav increases with inc. [ Hint: \( Var(e|\text{inc}) = Var(e) \), if \( e \) and \( \text{inc} \) are independent.]

(c) Provide a discussion that supports the assumption that the variance of savings increases with family income.

5. (a) Consider the sample regression

\[ y_i = \hat{\beta}_0 + \hat{\beta}_1 x_i + \hat{u}_i \]

Imposing the restrictions (i) \( \sum \hat{u}_i = 0 \) and (ii) \( \sum \hat{u}_i x_i = 0 \), we obtain the estimators \( \hat{\beta}_0 \) and \( \hat{\beta}_1 \). Show that they are identical with the least-square estimators.
(b) Verify the following equations:

\[
\begin{align*}
\sum_{i=1}^{n} x_i(y_i - \bar{y}) &= \sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y}) , \\
\sum_{i=1}^{n} x_i(x_i - \bar{x}) &= \sum_{i=1}^{n} (x_i - \bar{x})^2
\end{align*}
\]