to characterize the problem as inferring whether study "causes" work or work "causes" study? Explain.

**COMPUTER EXERCISES**

C1.1 Use the data in WAGE1.RAW for this exercise.
   (i) Find the average education level in the sample. What are the lowest and highest years of education?
   (ii) Find the average hourly wage in the sample. Does it seem high or low?
   (iii) The wage data are reported in 1976 dollars. Using the *Economic Report of the President* (2004 or later), obtain and report the Consumer Price Index (CPI) for the years 1976 and 2003.
   (iv) Use the CPI values from part (iii) to find the average hourly wage in 2003 dollars. Now does the average hourly wage seem reasonable?
   (v) How many women are in the sample? How many men?

C1.2 Use the data in BWGHT.RAW to answer this question.
   (i) How many women are in the sample, and how many report smoking during pregnancy?
   (ii) What is the average number of cigarettes smoked per day? Is the average a good measure of the "typical" woman in this case? Explain.
   (iii) Among women who smoked during pregnancy, what is the average number of cigarettes smoked per day? How does this compare with your answer from part (ii), and why?
   (iv) Find the average of fatheeduc in the sample. Why are only 1,192 observations used to compute this average?
   (v) Report the average family income and its standard deviation in dollars.

C1.3 The data in MEAP01.RAW are for the state of Michigan in the year 2001. Use these data to answer the following questions.
   (i) Find the largest and smallest values of math4. Does the range make sense? Explain.
   (ii) How many schools have a perfect pass rate on the math test? What percentage is this of the total sample?
   (iii) How many schools have math pass rates of exactly 50 percent?
   (iv) Compare the average pass rates for the math and reading scores. Which test is harder to pass?
   (v) Find the correlation between math4 and read4. What do you conclude?
   (vi) The variable expmp is expenditure per pupil. Find the average of expmp along with its standard deviation. Would you say there is wide variation in per pupil spending?
   (vii) Suppose School A spends $6,000 per student and School B spends $5,500 per student. By what percentage does School A's spending exceed School B's? Compare this to $100 \cdot \left(\log(6,000) - \log(5,500)\right)$, which is the approximation percentage difference based on the difference in the natural logs. (See Section A.4 in Appendix A.)
C1.4 The data in JTRAIN2.RAW come from a job training experiment conducted for low-income men during 1976–1977; see Lalonde (1986).

(i) Use the indicator variable \( \text{train} \) to determine the fraction of men receiving job training.

(ii) The variable \( \text{re78} \) is earnings from 1978, measured in thousands of 1982 dollars. Find the averages of \( \text{re78} \) for the sample of men receiving job training and the sample not receiving job training. Is the difference economically large?

(iii) The variable \( \text{unem78} \) is an indicator of whether a man is unemployed or not in 1978. What fraction of the men who received job training are unemployed? What about for men who did not receive job training? Comment on the difference.

(iv) From parts (ii) and (iii), does it appear that the job training program was effective? What would make our conclusions more convincing?