

Empirical Exercise: Chapter 5

The data file **Birthweight_Smoking**, which contains data for a random sample of babies born in Pennsylvania in 1989. The data include the baby's birth weight together with various characteristics of the mother, including whether she smoked during the pregnancy. A detailed description is given in **Birthweight_Smoking_Description**, also available in the folder. In this exercise, you will investigate the relationship between birth weight and smoking during pregnancy.

- a. In the sample:
 - i. What is the average value of *Birthweight* for all mothers?
 - ii. For mothers who smoke?
 - iii. For mothers who do not smoke?
- b.
 - i. Use the data in the sample to estimate the difference in average birth weight for smoking and nonsmoking mothers.
 - ii. What is the standard error for the estimated difference in (i)?
 - iii. Construct a 95% confidence interval for the difference in the average birth weight for smoking and nonsmoking mothers.
- c. Run a regression of *Birthweight* on the binary variable *Smoker*.
 - i. Explain how the estimated slope and intercept are related to your answers in parts (a) and (b).
 - ii. Explain how the $SE(\hat{\beta}_1)$ is related to your answer in b(ii).
 - iii. Construct a 95% confidence interval for the effect of smoking on birth weight.
- d. Do you think smoking is uncorrelated with other factors that cause low birth weight? That is, do you think that the regression error term—say, u_i —has a conditional mean of 0 given *Smoking* (X_i)? (You will investigate this further in *Birthweight* and *Smoking* exercises in later chapters.)