

- 2.10** Consider the standard simple regression model $y = \beta_0 + \beta_1 x + u$ under the Gauss-Markov Assumptions SLR.1 through SLR.5. The usual OLS estimators $\hat{\beta}_0$ and $\hat{\beta}_1$ are unbiased for their respective population parameters. Let $\tilde{\beta}_1$ be the estimator of β_1 obtained by assuming the intercept is zero (see Section 2.6).
- Find $E(\tilde{\beta}_1)$ in terms of the x_i , β_0 , and β_1 . Verify that $\tilde{\beta}_1$ is unbiased for β_1 when the population intercept (β_0) is zero. Are there other cases where $\tilde{\beta}_1$ is unbiased?
 - Find the variance of $\tilde{\beta}_1$. (*Hint*: The variance does not depend on β_0 .)
 - Show that $\text{Var}(\tilde{\beta}_1) \leq \text{Var}(\hat{\beta}_1)$. [*Hint*: For any sample of data, $\sum_{i=1}^n x_i^2 \geq \sum_{i=1}^n (x_i - \bar{x})^2$, with strict inequality unless $\bar{x} = 0$.]
 - Comment on the tradeoff between bias and variance when choosing between $\hat{\beta}_1$ and $\tilde{\beta}_1$.
- 2.11** The data set BWGHT.RAW contains data on births to women in the United States. Two variables of interest are the dependent variable, infant birth weight in ounces (*bwght*), and an explanatory variable, average number of cigarettes the mother smoked per day during pregnancy (*cigs*). The following simple regression was estimated using data on $n = 1,388$ births:

$$\widehat{bwght} = 119.77 - 0.514 \text{ cigs}$$

- What is the predicted birth weight when *cigs* = 0? What about when *cigs* = 20 (one pack per day)? Comment on the difference.
- Does this simple regression necessarily capture a causal relationship between the child's birth weight and the mother's smoking habits? Explain.
- To predict a birth weight of 125 ounces, what would *cigs* have to be? Comment.
- The proportion of women in the sample who do not smoke while pregnant is about .85. Does this help reconcile your finding from part (iii)?

COMPUTER EXERCISES

- C2.1** 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 50¢ contribution by the firm.
- Find the average participation rate and the average match rate in the sample of plans.
 - Now, estimate the simple regression equation

$$\widehat{prate} = \hat{\beta}_0 + \hat{\beta}_1 \text{ mrate},$$

- and report the results along with the sample size and *R*-squared.
- Interpret the intercept in your equation. Interpret the coefficient on *mrate*.
 - Find the predicted *prate* when *mrate* = 3.5. Is this a reasonable prediction? Explain what is happening here.
 - How much of the variation in *prate* is explained by *mrate*? Is this a lot in your opinion?