ucation will be $7.64 per hour if we
d predictor. In this case the sample
tor. Among the 1000 workers there
$8.30, so the corrected predictor is

l that $\hat{\sigma}^2$ must be greater than zero
ways to increase the value of the
natural predictor tends to system-
del, and the correction offsets the

een y and its fitted value $\hat{y}$, where $\hat{y}$
of goodness-of-fit that we can use
consider the "best" predictor can
That is, a general goodness-of-fit

246, as compared to the reported
. (In this case since the corrected
correlation is the same for both.)
ssage: R^2-values tend to be small
riations in individual behavior are

MODEL

del. It is the "point" predictor, or
er that is our best prediction of y.
must rely on the natural predictor
tion 4.1, and then take antilogs.
f), where the critical value t_c is
se(f) is given in (4.5). Then a

$t_c \mathrm{se}(f))\Big]$

ge of a worker with 12 years of

$.4905)] = [2.9184, 20.0046]$

or variance, making the t-distribution no

The interval prediction is \$2.92–\$20.00, which is so wide that it is basically useless. What does this tell us? Nothing we did not already know. Our model is not an accurate predictor of individual behavior in this case. In later chapters we will see if we can improve this model by adding additional explanatory variables, such as experience, that should be relevant.

4.5 Exercises

Answer to exercises marked * appear in Appendix D at the end of the book.

4.5.1 PROBLEMS

4.1* (a) Suppose that a simple regression has quantities $\Sigma(y_i - \bar{y})^2 = 631.63$ and $\Sigma\hat{e}_i^2 = 182.85$, find R^2.

(b) Suppose that a simple regression has quantities $N = 20$, $\Sigma y_i^2 = 5930.94$, $\bar{y} = 16.035$, and $SSR = 666.72$, find R^2.

(c) Suppose that a simple regression has quantities $R^2 = 0.7911$, $SST = 552.36$, and $N = 20$, find $\hat{\sigma}^2$.

4.2* Consider the following estimated regression equation (standard errors in parentheses):

$$\hat{y} = 5.83 + 0.869x \quad R^2 = 0.756$$
$$(\mathrm{se})\ (1.23)\ \ (0.117)$$

Rewrite the estimated equation that would result if

(a) All values of x were divided by 10 before estimation.

(b) All values of y were divided by 10 before estimation.

(c) All values of y and x were divided by 10 before estimation.

4.3 Using the data in Exercise 2.1 and only a calculator (show your work) compute

(a) the predicted value of y for $x_0 = 5$.

(b) the se(f) corresponding to part (a).

(c) a 95% prediction interval for y given $x_0 = 5$.

(d) a 99% prediction interval for y given $x_0 = 5$.

(e) a 95% prediction interval for y given $x = \bar{x}$. Compare the width of this interval to the one computed in part (c).

4.4 Given the simple linear model $y = \beta_1 + \beta_2 x + e$, and the least squares estimators, we can estimate $E(y)$ for any value of $x = x_0$ as $\widehat{E(y_0)} = b_1 + b_2 x_0$.

(a) Describe the difference between predicting y_0 and estimating $E(y_0)$.

(b) Find the expected value and variance of $\widehat{E(y_0)} = b_1 + b_2 x_0$.

(c) When discussing the unbiasedness of the least squares predictor we showed that $E(f) = E(y_0 - \hat{y}_0) = 0$, where f is the forecast error. Why did we define unbiasedness in this strange way? What is wrong with saying, as we have in other unbiasedness demonstrations, that $E(\hat{y}_0) = y_0$?

4.5 Suppose you are estimating a simple linear regression model.

(a) If you multiply all the x values by 10, but not the y values, what happens to the parameter values β_1 and β_2? What happens to the least squares estimates b_1 and b_2? What happens to the variance of the error term?

(b) Suppose you are estimating a simple linear regression model. If you multiply all the y values by 10, but not the x values, what happens to the parameter values β_1