

Chapter 4

Linear Regression with One Regressor

- 4.1. (a) The predicted average test score is

$$TestScore = 520.4 - 5.82 \times 22 = 392.36$$

- (b) The predicted change in the classroom average test score is

$$\Delta TestScore = (-5.82 \times 19) - (-5.82 \times 23) = 23.28$$

- (c) Using the formula for $\hat{\beta}_0$ in Equation (4.8), we know the sample average of the test scores across the 100 classrooms is

$$\overline{TestScore} = \hat{\beta}_0 + \hat{\beta}_1 \times \overline{CS} = 520.4 - 5.82 \times 21.4 = 395.85.$$

- (d) Use the formula for the standard error of the regression (SER) in Equation (4.19) to get the sum of squared residuals:

$$SSR = (n - 2)SER^2 = (100 - 2) \times 11.5^2 = 12961.$$

Use the formula for R^2 in Equation (4.16) to get the total sum of squares:

$$TSS = \frac{SSR}{1 - R^2} = \frac{12961}{1 - 0.08^2} = 13044.$$

The sample variance is $s_Y^2 = \frac{TSS}{n-1} = \frac{13044}{99} = 131.8$. Thus, standard deviation is $s_Y = \sqrt{s_Y^2} = 11.5$.

- 4.2. The sample size $n = 200$. The estimated regression equation is

$$Weight = (2.15) - 99.41 + (0.31) 3.94 Height, \quad R^2 = 0.81, \quad SER = 10.2.$$

- (a) Substituting $Height = 70, 65,$ and 74 inches into the equation, the predicted weights are 176.39, 156.69, and 192.15 pounds.
- (b) $\Delta Weight = 3.94 \times \Delta Height = 3.94 \times 1.5 = 5.91$.
- (c) We have the following relations: $1 \text{ in} = 2.54 \text{ cm}$ and $1 \text{ lb} = 0.4536 \text{ kg}$. Suppose the regression equation in centimeter-kilogram units is

$$Weight = \hat{\gamma}_0 + \hat{\gamma}_1 Height.$$

The coefficients are $\hat{\gamma}_0 = -99.41 \times 0.4536 = -45.092 \text{ kg}$; $\hat{\gamma}_1 = 3.94 \times \frac{0.4536}{2.54} = 0.7036 \text{ kg per cm}$.

The R^2 is unit free, so it remains at $R^2 = 0.81$. The standard error of the regression is $SER = 10.2 \times 0.4536 = 4.6267 \text{ kg}$.

- 4.3. (a) The coefficient 9.6 shows the marginal effect of *Age* on *AWE*; that is, *AWE* is expected to increase by \$9.6 for each additional year of age. 696.7 is the intercept of the regression line. It determines the overall level of the line.
- (b) *SER* is in the same units as the dependent variable (*Y*, or *AWE* in this example). Thus *SER* is measured in dollars per week.
- (c) R^2 is unit free.
- (d) (i) $696.7 + 9.6 \times 25 = \936.7 ;
(ii) $696.7 + 9.6 \times 45 = \$1,128.7$
- (e) No. The oldest worker in the sample is 65 years old. 99 years is far outside the range of the sample data.
- (f) No. The distribution of earning is positively skewed and has kurtosis larger than the normal.
- (g) $\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$, so that $\bar{Y} = \hat{\beta}_0 + \hat{\beta}_1 \bar{X}$. Thus the sample mean of *AWE* is $696.7 + 9.6 \times 41.6 = \$1,096.06$.
- 4.5. (a) u_i represents factors other than time that influence the student's performance on the exam including amount of time studying, aptitude for the material, and so forth. Some students will have studied more than average, other less; some students will have higher than average aptitude for the subject, others lower, and so forth.
- (b) Because of random assignment u_i is independent of X_i . Since u_i represents deviations from average $E(u_i) = 0$. Because u and X are independent $E(u_i|X_i) = E(u_i) = 0$.
- (c) (2) is satisfied if this year's class is typical of other classes, that is, students in this year's class can be viewed as random draws from the population of students that enroll in the class. (3) is satisfied because $0 \leq Y_i \leq 100$ and X_i can take on only two values (90 and 120).
- (d) (i) $49 + 0.24 \times 90 = 70.6$; $49 + 0.24 \times 120 = 77.8$; $49 + 0.24 \times 150 = 85.0$
(ii) $0.24 \times 10 = 2.4$.