Errata

Chapter 4: Fitting a Line

Page 118: The entry in the fifth row, fourth column of table 4.2 should be "−10,200" instead of "10,200". The corrected table is

Table 4.2

Predicted values and prediction errors for the regression in table 4.1

Observation	bx_i	$\hat{y}_i = a + bx_i$	$e_i = y_i - \hat{y}_i$	e_i^2
1	66,600	27,800	-7,800	60,840,000
2	77,700	38,900	14,100	198,810,000
3	88,800	50,000	-15,000	225,000,000
4	99,900	61,100	18,900	357,210,000
5	111,000	72,200	−**10,200**	104,040,000
Sums	250,000	250,000	0	945,900,000

Page 122: The heading for the third column in table 4.4, should be

$$\hat{y}_i = a + bx_i$$

rather than

$$y_i = a + bx_i.$$

The corrected table is

Table 4.4

R^2 calculation, rooms and rent

Observation	bx_i	$\hat{y}_i = a + bx_i$	$e_i = y_i - \hat{y}_i$	e_i^2	$(y_i - \bar{y})^2$
1	96.65	411.04	-211.04	44,539.26	209,306.25
2	193.30	507.69	382.31	146,158.67	54,056.25
3	193.30	507.69	32.31	1,043.74	13,806.25
4	193.30	507.69	-87.69	7,690.06	56,406.25
5	193.30	507.69	-277.69	77,113.38	182,756.25
6	193.30	507.69	192.31	36,982.00	1,806.25
7	289.95	604.34	-204.34	41,755.92	66,306.25
8	289.95	604.34	-254.34	64,690.19	94,556.25
9	289.95	604.34	75.66	5,724.03	506.25
10	289.95	604.34	75.66	5,724.03	506.25
11	386.60	700.99	179.01	32,043.73	49,506.25
12	386.60	700.99	-20.99	440.68	506.25
13	386.60	700.99	-250.99	62,997.17	43,056.25
14	386.60	700.99	219.01	47,964.34	68,906.25
15	386.60	700.99	-290.99	84,676.56	61,256.25
16	386.60	700.99	-250.99	62,997.17	43,056.25
17	386.60	700.99	139.01	19,323.12	33,306.25
18	483.25	797.64	502.36	252,363.49	412,806.25
19	579.90	894.29	405.71	164,599.17	412,806.25
20	869.85	1,184.24	-354.24	125,486.60	29,756.25
Sum			0	1,284,313.32	1,834,975.00

Chapter 5: From Sample to Population

Page 145: Equation 5.6 would be clearer if it read as

$$V(\varepsilon_i) = E\left[\left(\varepsilon_i - E(\varepsilon_i)\right)^2\right] = E\left(\varepsilon_i^2\right) = \sigma^2.$$

rather than as

$$V(\varepsilon_i) = E\left(\varepsilon_i - E(\varepsilon_i)\right)^2 = E(\varepsilon_i)^2 = \sigma^2.$$

Page 186: Exercise 5.13b should refer to equation (5.34) rather than equation (5.43). It should read as "Apply equation **(5.34)** to the denominator ..." rather than as "Apply equation (5.43) to the denominator ...".

Page 186: Exercise 5.13c requires some preliminary steps. As given in the solution,

"We need a couple of preliminary steps before we can do this. First, we recognize that

$$\begin{aligned} E\left(\sum_{i=1}^{n}(x_i - \bar{x})\varepsilon_i\right) &= \sum_{i=1}^{n} E\left((x_i - \bar{x})\varepsilon_i\right) \\ &= \sum_{i=1}^{n}(x_i - \bar{x})E(\varepsilon_i) \\ &= \sum_{i=1}^{n}(x_i - \bar{x})(0) \\ &= 0. \end{aligned}$$

We get the first equality from equation (5.13), the second from equation (5.34) and the third equality from equation (5.5). With this result,

$$\begin{aligned} E\left(\left(\sum_{i=1}^{n}(x_i - \bar{x})\varepsilon_i\right)^2\right) &= E\left(\left(\sum_{i=1}^{n}(x_i - \bar{x})\varepsilon_i - 0\right)^2\right) \\ &= E\left(\left(\sum_{i=1}^{n}(x_i - \bar{x})\varepsilon_i - E\left(\sum_{i=1}^{n}(x_i - \bar{x})\varepsilon_i\right)\right)^2\right). \end{aligned}$$

Equation (5.4) reminds us that

$$E\left(\left(\sum_{i=1}^{n}(x_i-\bar{x})\varepsilon_i - E\left(\sum_{i=1}^{n}(x_i-\bar{x})\varepsilon_i\right)\right)^2\right)=V\left(\sum_{i=1}^{n}(x_i-\bar{x})\varepsilon_i\right).$$

Therefore, the answer to 5.13b can be rewritten as

$$\frac{1}{\left(\sum_{i=1}^{n}(x_i-\bar{x})x_i\right)^2}E\left(\left(\sum_{i=1}^{n}(x_i-\bar{x})\varepsilon_i\right)^2\right)=\frac{1}{\left(\sum_{i=1}^{n}(x_i-\bar{x})x_i\right)^2}V\left(\sum_{i=1}^{n}(x_i-\bar{x})\varepsilon_i\right).$$

At last, we can apply equation (5.46)."

Chapter 6: Confidence Intervals and Hypothesis Tests

Page 213: Equation 6.29 should have "SD(d)" rather than "SD(δ)" in the denominator. It should read as

$$1-\alpha=P\left(\frac{d-\delta}{SD(d)}<t_{\alpha}^{(df)}\right).$$

rather than as

$$1-\alpha=P\left(\frac{d-\delta}{SD(\delta)}<t_{\alpha}^{(df)}\right).$$

Page 229: Exercise 6.16 does not specify the relationship between SD(d) and n, which is necessary in order to arrive at the solution. Assume that when n increases, SD(d) declines.

Page 229: Exercise 6.17 should refer to confidence "interval" rather than "test". It should read as "Demonstrate that any specified value for δ_0 in excess of $d+t_{\alpha/2}^{(df)}$SD(d), the upper bound for the confidence **interval** ..." rather than as "Demonstrate that any specified value for δ_0 in excess of $d+t_{\alpha/2}^{(df)}SD(d)$, the upper bound for the confidence test..."

Chapter 7: Inference in Ordinary Least Squares

Pages 263-4: The equations from middle of page 263 to top of page 264 have a negative sign, "−" where they need a positive sign, "+". This passage should read as follows:

"Regrouping,

$$V(\hat{y}_0) = E\left[(a-\alpha) + x_0(b-\beta)\right]^2.$$

Squaring, we obtain

$$V(\hat{y}_0) = E\left[(a-\alpha)^2 + x_0^2(b-\beta) + 2x_0(a-\alpha)(b-\beta)\right]$$

We now have the expected value of a sum of three terms. Again, equation 5.13 allows us to rewrite the expected value of a sum as the sum of the expected values:

$$V(\hat{y}_0) = E(a-\alpha)^2 + x_0^2 E(b-\beta)^2 + E\left[2x_0(a-\alpha)(b-\beta)\right].$$

Equation 5.34 reminds us that the expected value of a constant times a random variable is the constant times the expected value of the random variable, so the third term can be rewritten to produce

$$V(\hat{y}_0) = E(a-\alpha)^2 + x_0^2 E(b-\beta)^2 + 2x_0 E\left[(a-\alpha)(b-\beta)\right].$$

Here, we recognize that E(a)=α from equation 5.42 and E(b)=β from equation 5.37. Therefore, equation 5.4 tells us that the first term is V(a) and the second term is V(b). Equation 5.8 identifies the expectation in the third term as COV(a,b). Therefore

$$V(\hat{y}_0) = V(a) + x_0^2 V(\beta) + 2x_0 COV(a,b)$$

It is now evident ...".

Page 276: Exercise 7.3c should refer to confidence "level" rather than "interval". It should read as "90% confidence **level**" rather than as "90% confidence interval".

Chapter 9: What If the Disturbances are Correlated?

Page 358: Parts b, c and d of exercise 9.5 would be clearer if it were explicit that the "relationships" to which they refer are "population relationships", as follows:

"b. Multiply the **population** relationship for the period $i-4$ by γ, as in equation (9.30).

c. Subtract the population relationship for the period $i-4$, multiplied by γ, from the **population** relationship for the period i, as in equations (9.31) and (9.32).

d. Analyze the random component of the resulting **population** relationship. What are its properties?"

Page 360: Exercise 9.9 refers to "best consistent" estimators, which aren't defined until chapter 15.

Chapter 10: What If the Disturbances and the Explanatory Variables are Related?

Page 392: The headings for the first columns of tables 10.7 and 10.8 should read as "Standard deviation of measurement error, SD(ν_i)" rather than as "Variance of measurement error, V(ν_i)". The corrected tables are

---

Table 10.7

Population standard deviations for b and asymptotic standard deviations for b_{IV}

Simulation	**Standard deviation** of measurement error, **SD**(ν_i)	Population SD(b)	CORR(z_i,x_i)	Asymptotic SD(b_{IV})
1	1	25.649	.89473	28.667
2	1.5	24.112	.79070	30.495
3	2	22.361	.68000	32.883
4	2.5	20.585	.57627	35.720
5	3	18.898	.48571	38.908
6	3.5	17.355	.40964	42.367

---

and

Table 4.8

Estimated standard deviations for b and b_{IV} in simulated data

Simulation	**Standard deviation** of measurement error, **SD**(v_i)	OLS estimate of SD(b)	IV estimate of SD(b_{IV})
1	1	25.909	28.982
2	1.5	24.589	31.277
3	2	23.110	34.473
4	2.5	21.562	38.377
5	3	19.993	43.435
6	3.5	18.432	47.974

Page 402: Exercise 10.2b should read as "Demonstrate that the term to the **right** of the equality ..." rather than as "Demonstrate that the term to the **left** of the equality ...".

Page 407: Exercise 10.16b mistakenly reverses the standard deviations in the numerator and denominator. It should read as

$$b_{IV} \approx \beta + \frac{CORR(z_i, \varepsilon_i)}{CORR(z_i, x_i)} \frac{SD(\varepsilon_i)}{SD(x_i)}$$

rather than as

$$b_{IV} \approx \beta + \frac{CORR(z_i, \varepsilon_i)}{CORR(z_i, x_i)} \frac{SD(x_i)}{SD(\varepsilon_i)}.$$

Chapter 11: What If There is More than One x?

Page 447: Exercise 11.5c should refer to equation (4.15) rather than equation (4.18). It should read as “If we apply the procedure of equations (4.13) through (**4.15**) to the sum of squared errors ...” rather than as “If we apply the procedure of equations (4.13) through (4.18) to the sum of squared errors ...”.

Chapter 12: Understanding and Interpreting Regression With Two x’s?

Page 495: Exercise 12.11 should refer to exercise 4.3 rather than exercise 4.4. It should read as “In exercise **4.3** ...” rather than as “In exercise 4.4 ...”.

Page 496: Unfortunately, the notation in exercise 12.11c is incorrect. The correct formulas are

$$a = \frac{\left(x_{12}x_{23} - x_{13}x_{22}\right)y_1 - \left(x_{23} - x_{22}\right)y_2 + \left(x_{13} - x_{12}\right)y_3}{x_{11}x_{22} - x_{11}x_{23} - x_{12}x_{21} + x_{12}x_{23} + x_{13}x_{21} - x_{13}x_{22}},$$

$$b_1 = -\frac{\left(x_{11}x_{23} - x_{13}x_{21}\right)y_1 - \left(x_{23} - x_{21}\right)y_2 + \left(x_{13} - x_{11}\right)y_3}{x_{11}x_{22} - x_{11}x_{23} - x_{12}x_{21} + x_{12}x_{23} + x_{13}x_{21} - x_{13}x_{22}}$$

and

$$b_2 = \frac{\left(x_{11}x_{22} - x_{12}x_{21}\right)y_1 - \left(x_{22} - x_{21}\right)y_2 + \left(x_{12} - x_{11}\right)y_3}{x_{11}x_{22} - x_{11}x_{23} - x_{12}x_{21} + x_{12}x_{23} + x_{13}x_{21} - x_{13}x_{22}}.$$

Worse, there doesn’t seem to be any way to be certain that these are the correct formulas without actually replacing the sample values in equations (11.28), (11.37) and (11.38), why these formulas are correct. It may be best to move on to the next part of this exercise.

Chapter 13: Making Regression More Flexible

Page 533: Exercise 13.9e should refer to "GNI per capita" rather than to "child mortality". It should read as "What can we conclude about the joint effects of the CPI and its square on **GNI per capita**" rather than as "What can we conclude about the joint effects of the CPI and its square on child mortality".

Page 538: Exercise 13.17d should refer to part a rather than part c. It should read as "Based on the answer to part **a** ..." rather than as "Based on the answer to part c ...".

Chapter 14: More Than Two Explanatory Variables

Page 555: The constant in equation (14.23) should be δ_1 rather than δ_0 in order to be consistent with equation (14.25). Equation (14.23) should read as

$$y_i = \delta_1 + \sum_{l=2}^{k} \delta_l x_{li} + \varepsilon_i$$

rather than as

$$y_i = \delta_0 + \sum_{l=2}^{k} \delta_l x_{li} + \varepsilon_i .$$

Page 555: The constant in equation (14.24) should be μ_1 rather than μ_0 in order to be consistent with equation (14.25). Equation (14.24) should read as

$$y_i = \mu_1 + \sum_{l=2}^{k} \mu_l x_{li} + \varepsilon_i .$$

rather than as

$$y_i = \mu_0 + \sum_{l=2}^{k} \mu_l x_{li} + \varepsilon_i .$$

Page 562: Table 14.3 incorrectly presents the slopes for Hispanic, American Indian or

Alaskan Native, Asian, Native Hawaiian and other Pacific Islander in the female regression as positive. They are actually negative. Table 14.3 should read as follows:

Table 14.3

Earnings regressions stratified by sex, sample from section 7.6

Explanatory variables	Pooled regression	Regression for men	Regression for women
Intercept	−10,899	−24,146	−12,669
	(18.85)	(25.88)	(20.80)
Female	−18,720	-	-
	(94.01)		
Years of school	3,472.2	4,213.6	2,565.7
	(118.04)	(86.17)	(83.04)
Age	266.78	411.81	84.432
	(31.38)	(28.82)	(9.58)
Black	−10,305	−18,601	−2,147.9
	(24.63)	(26.04)	(5.04)
Hispanic	−5,875.2	−8,447.7	**−3,384.1**
	(22.35)	(19.18)	(12.41)
American Indian or Alaskan Native	−11,040	−15,114	**−7,030.6**
	(12.71)	(10.24)	(7.92)
Asian	−5,139.2	−9,313.4	**−1,451.3**
	(15.74)	(16.46)	(4.42)
Native Hawaiian, other Pacific Islander	−7,052.9	−10,752	**−3,766.1**
	(4.59)	(4.21)	(2.34)
R^2	.1565	.1405	.1057
Adjusted R^2	.1564	.1404	.1057
F-statistic	4162.4	2127.9	1494.3
prob-value for F-statistic	<.0001	<.0001	<.0001
Sum of squared errors	3.187676×10^{14}	2.317906×10^{14}	8.272367×10^{13}
Observations	179,549	91,075	88,474

Page 573: Exercise 14.1a isn't entirely correct. The beginning of exercise 12.10a proves only that, with two explanatory variables,

$$\bar{\hat{y}} = a + b_i\bar{x}_1 + b_2\bar{x}_2 = \bar{y}.$$

With k explanatory variables, this result must be generalized to

$$\bar{\hat{y}} = a + \sum_{i=1}^{k} b_i\bar{x}_i = \bar{y}$$

in order to complete exercise 14.1a.

Chapter 15: Categorical Dependent Variables

Page 621: The notation for exercise 15.2 could use a little elaboration. $N_{.33}$ represents the critical value for the standard normal random variable for which the probability of observing outcomes above this value is 33%.

Page 622: Exercise 15.4b needs a "P" towards the end of the first line. It should read as "The probability of the observed sequence is *P(sequence)=**P**(*g_1 *and* g_2 *and* f_3 *and* f_4 *and* f_5 *and* g_6 *and* f_7 *and ...)"* rather than as "The probability of the observed sequence is *P(sequence)=(*g_1 *and* g_2 *and* f_3 *and* f_4 *and* f_5 *and* g_6 *and* f_7 *and ...)*".