Solutions to Worktext Exercises

Chapter 19

Visualizing Binary Predictors in Regression

Basic Learning Exercises

1. a) Scps = $\beta_0 + \beta_1$ Tmp + β_2 Ovr40 + ε b) 2 lines c) The two lines will be parallel. If older people eat less ice cream, then the Ovr40 line will be below the other line.

2. a) Scps = 0.320 + 0.0439 (Tmp) - 0.273 Ovr40 + ε b) There are two lines because when Ovr40 is 0 the equation is Scps = 0.320 + 0.0439 (Tmp), but when Ovr40 = 1 then the equation is Scps = (0.320 - 0.273) + 0.0439 (Tmp). c) - 0.273

3. a) Not reject H_0 since the p-value > 0.05. b) People over 40 eat the same amount of ice cream as people 40 and under. c) Reject H_0 since p-value/2 < 0.05. d) People over 40 eat less ice cream than people 40 and under. e) To test if people over 40 eat a *different* amount of ice cream then people 40 and under you would use the hypothesis in a). To test if people over 40 eat *less* ice cream, you would use the hypothesis in c).

4. a) StUn = $\beta_0 + \beta_1$ (CtyUn) + β_2 (Urbn) (CtyUn)+ ε. b) 2 lines c) The 2 lines have the same intercept but different slopes; Urbn = 1 will be steeper then Urbn = 0.

5. StUn = 4.50 + 0.121 (CtyUn) + 0.215 (Urbn) (CtyUn) b) There are two lines because when Urbn is 0 the equation is StUn = 4.50 + 0.121 (CtyUn), but when Urbn = 1 then the equation is StUn = 4.50 + (0.121 + 0.215) (CtyUn). The intercept is 4.5 in both cases but the slopes are 0.121 and 0.336.

6. You would reject H_0 since the p-value < 0.01. Counties that contain an urban area have more of an effect on the state's unemployment rate. Since counties with urban areas have larger populations they will have a larger impact on their state's unemployment rate.

7. a) Inc = $\beta_0 + \beta_1$ (Exp) + β_2 Male +β_3 (Male) (Exp) + ε. b) 2 lines c) The two lines have different intercepts *and* different slopes.

8. a) Inc = 25.6 + 1.33 (Exp) - 1.98 Male + 0.266 (Male) (Exp) b) There are two lines because when Male is 0 the equation is Inc = 25.6 + 1.33 (Exp), but when Male = 1 the equation is Inc = (25.6 - 1.98) + (1.33 + 0.266) (Exp). c) A different intercept means that the starting salary for males and females are different. d) A different slope means that males and females are rewarded differently for each year of experience.

9. a) You would not reject H_0 since the p-value > 0.05. b) That male and female starting salaries are the same. c) No, it would not support the contention of gender discrimination since the starting salaries seem to be the same.

10. a) You would reject H_0 since the p-value < 0.05. b) That males on average earn \$0.266 thousand (\$266) more for every year of experience. c) Yes, this result would support the contention of gender discrimination since on average males seem to earn more for each year of experience. d) Other useful factors would be the education and performance appraisal for each worker.

Intermediate Learning Exercises

11. a) $Clos = \beta_0 + \beta_1 (Pop) + \beta_2 S2 + \beta_3 (S2)(Pop) + \beta_4 S3 + \beta_5 (S3)(Pop) + \beta_6 S4 + \beta_7 (S4)(Pop) + \varepsilon$. b) 4 lines c) The four lines each have a different intercept and slope.

12. a) Clos = 20.3 + 98.7 (Pop) - 3.40 S2 + 17.7 (S2) (Pop) - 29.8 S3 - 0.0878 (S3) (Pop) + 34.4 S4 + 20.2 (S4) (Pop) b) There are four lines: When S2 = S3 = S4 = 0, then Clos = 20.3 + 98.7 (Pop). When S2 = 1, then Clos = 16.9 + 116.4 (Pop). When S3 = 1, then Clos = -9.5 + 98.6122 (Pop). When S4 = 1, then Clos = 54.7 + 118.9 (Pop). c) A different intercept means that a different base amount is sold each quarter regardless of the population. d) A different slope means that population affects clothes sales differently in each quarter.

13. You would not reject H_0 since the p-value < 0.05. That the base sales in the 4^{th} quarter is \$34.4 million greater than in the 1^{st} quarter.

14. a) Reject H_0 since the p-value < 0.05. b) On average each person (Pop and Clos are in millions) spends \$20.20 more on clothes in the 4^{th} quarter than the 1^{st} quarter. You would reject H_0 for β_3 and *not* reject H_0 for β_2, β_4 and β_5. This means that the 3rd period is similar to the 1^{st} period. In contrast, the second period has the same base spending as the 1^{st} period, but on average each person spends \$17.70 more than in the 1^{st} quarter.

15. Clos = 54.7 + 119 (Pop) - 34.4 S1 - 20.2 (S1) (Pop) - 37.8 S2 - 2.50 (S2) (Pop) - 64.2 S3 - 20.3 (S3) (Pop). The estimated model is the same for each quarter. The coefficients are different because they are estimated *relative* to the quarter that has been omitted.

16. a) $Cars = \beta_0 + \beta_1 (Per) + \beta_2 S2 + \beta_3 S3 + \beta_4 S4 + \beta_5 S5 + \varepsilon$. b) 5 lines c) The 5 lines have different intercepts but the same slope (β_3, β_4, and β_5 are negative since sales should be greater on days that the dealerships are open late).

17. a) Cars = 6.89 + 0.21 (Per) – 0.01 S2 – 1.62 S3 – 2.06 S4 – 0.765 S5 b) There are five lines, when S2 = S3 = S4 = S5 = 0, then Cars = 6.89 + 0.21 (Per). When S2 = 1, then Cars = 6.88 + 0.21 (Per). When S3 = 1, then Cars = 5.23 + 0.21 (Per). When S4 = 1, then Cars = 4.83 + 0.21 (Per). When S5 = 1, then Cars = 6.125 + 0.21 (Per). c) The number of cars sold on that day of the week in period 0, i.e. Per = 0. d) They have the same slope because there is no interaction between the binary and Period.

18. $H_0: \beta_1 \leq 0$, $H_a: \beta_1 > 0$. Since the estimate is greater than 0 and the p-value/2 = 0.003 < 0.05, you would reject H_0. The data support the contention that sales increase during the month.

19. a) Since the dealership is open late on Mon. and Tues. nights it would be reasonable to assume that sales would be higher on these two days. b) $H_0: \beta_3 \geq 0$, $H_a: \beta_3 < 0$. $H_0: \beta_4 \geq 0$, $H_a: \beta_4 < 0$. $H_0: \beta_5 \geq 0$, $H_a: \beta_5 < 0$. c) No, in each case you would not reject the hypothesis. Therefore, the data did not support the belief that staying open late increased sales.

20. a) Cars = 6.8632 + 0.21189 Period –1.4874 WThF. b) Reject H_0 since the estimate is less than 0 and the p-value/2 = 0.0475 < 0.05. c) Yes, since sales are significantly higher on Mon. and Tues. d) We have one binary for all three days instead of three binaries.

Advanced Learning Exercises

21. a) 2 b) Equal frequencies c) 15 in both d) There are 15 observations less than or equal to 2594.5 and 15 observations greater than 2594.5. e) This is a median split because the data has been divided into 2 groups, observations below the median and above the median. f) The first one will be omitted.

22. a) Per = $\beta_0 + \beta_1$ (Rsk) + β_2 (Q2) (Rsk) + ε b) 2lines c) Same intercept but different slope. The slope for the wealthy will be greater than the slope for the non-wealthy.

23. a) Per = 7.27 + 0.764 (Rsk) + 0.527 (Q2) (Rsk) b) H_0: $\beta_2 = 0$, H_a: $\beta_2 > 0$. c) Reject H_0 since the p-value/2 < 0.05.

24. a) β_1 is the slope for the non-wealthy and β_2 is the slope for the wealthy. b) The estimate for β_1 should be 0.764 and for β_2 1.291. c) Per = 7.27 + 0.764 (Q1) (Rsk) + 1.29 (Q2) (Rsk) d) 0.764 + 0.527 is rounded to 1.29.

25. This model says that regardless of the risk, wealthy people earn a higher rate of return.

26. Per = 6.29 + 0.951 (Rsk) + 2.48 Q2. This is a two-tailed test. Hence, since the p-value exceeds 0.10 we do not reject H_0. Therefore, wealthy investors, based on these 30 investors, do not earn a different rate of return.

27. a) In order to take increased risk there has to be a *more* than proportional increase in percentage return. b) Since this is a log model, the gap is the same *percentage* of the variable on the vertical axis.

28. a) Since this is a one-tailed test and the p-value/2 < 0.05, we would reject H_0. We would conclude that wealthy investors view risk differently and require a higher return in order to justify taking the added risk.

29. a) 1 is single head of households, 2 are married households with 1 income, and 3 is married head of households with 2 incomes. b) 3 c) One binary was assigned to each categorical variable that was in the data. d) Binary 1 is being omitted.

30. Three lines will be drawn. Based on this sample of 75 families, married households with one income spend out of each additional dollar earned than either of the other two types of households.