DIRECTIONS:

- This assignment uses the supplementary textbook Chapter 17: Logistic Regression. **Find this chapter as a PDF file on the main course webpage.**
- Some of the exercises listed below include special instructions which modify or clarify textbook instructions.
- Data sets for some exercises are available on the MINITAB Data Sets page.

A. Mail-Order Sales

- Open the PDF file for the supplementary textbook Chapter 17: Logistic Regression on the main course webpage. Read the description of mail-order sales on the first page.

  Answer these questions:

  (a) There are 5 predictor variables and a response variable described for this logistic regression. Let's call them x_1, x_2, x_3, x_4, x_5, and y. Write an English definition for all 6 variables, including the meaning of "0" and "1" for binary variables.

  (b) Which of the variables are *binary*? Which are *quantitative* (i.e., *numerical*)?

B. College Student Drinking

- The textbook introduces logistic regression with the same data on drinking as used in Example 2 of the Topic 10 Notes. Take five minutes to review the answers you wrote in class for Example 2 on pages 8-11 of the Topic 10 Notes.
- Now open page 17-2 of the textbook Chapter 17 PDF file. Read pages 17-2 and 17-3 at a leisurely pace. "Soak up" what the textbook says and occasionally glance at the Topic 10 Notes to confirm that the book's answers are basically the same as ours.
- Next read Model for logistic regression at the bottom of page 17-4 and all of page 17-5.

  Then read Fitting and interpreting the logistic regression model on pages 17-6 and 17-7.

  Pay particular attention to the way the book calculates b_0 and b_1 at the top of 17-7. Are these the same estimates (except for rounding) that MINITAB logistic regression produces in the output shown on page 9 of the Topic 10 Notes?
- Lastly, read Section 17.1 Summary at the bottom of page 17-8 and top of page 17-9.
- Open MINITAB and type the drinking data into the worksheet:

C1	C2	C3	C4
	Gender	**Binge**	**Total**
Men	1	1630	7180
Women	0	1684	9916

- Reproduce the MINITAB output shown on page 9 of the Topic 10 Notes. (Follow the MINITAB steps shown in the notes.)

  Notice that there's some extra MINITAB output at the bottom — just ignore it!

- Now see what happens when you switch the coding for Gender. Define the variable as

$$x = \begin{cases} 1 & \text{if student is female} \\ 0 & \text{if student is male} \end{cases}$$

  Change your MINITAB worksheet accordingly and re-run MINITAB:

C1	C2	C3	C4
	Gender	**Binge**	**Total**
Men	0	1630	7180
Women	1	1684	9916

  Answer these questions:

  (a) What is the new odds ratio?

  (b) Interpret the new odds ratio.

  (c) How is the new odds ratio related to the original odds ratio?

C. Successful Franchises

- Exercise 17.1 (page 17-4)
- Exercise 17.3
- **Followup Exercise:** Perform logistic regression for these data with MINITAB, as described below:

  (a) Define the (binary) response and predictor variables.

  (b) Give the theoretical equation.

  (c) Give the fitted (estimated) equation from MINITAB.

  Hint: The hardest part of using MINITAB is to figure out how to put the data into the worksheet. Compare how you did it for Gender/Binge Drinking and do something similar for this problem.

  (d) Suppose that you were given the estimated equation from part (c), but not the original data. Use the equation to find the odds of success and probability of success for firms with nonexclusive territories.

  (e) Suppose that you were given the estimated equation from part (c), but not the original data. Use the equation to find the odds of success and probability of success for firms with exclusive territories.

  (f) Are your answers for (d) and (e) the same as the ones you calculated in Exercise 17.1?

D. Stock Options

- Exercise 17.19 (p. 17-19)
- Exercise 17.20
- **Followup Exercise:** Perform logistic regression for these data with MINITAB. The MINITAB output should provide the same answers as those you calculated in the above exercises. Give MINITAB's answers to the following (use all available decimal places):
  (a) $b_0 = \widehat{\beta}_0$ and $b_1 = \widehat{\beta}_1$
  (b) the estimated odds ratio
  (c) a 95% CI for the odds ratio
  (d) Based on the 95% CI for the odds ratio, is there overwhelming evidence at 5% significance that the odds of offering stock options differ between high-tech and non-high-tech firms?
  (e) At 5% significance, does the probability of offering stock options differ between high-tech and non-high-tech firms?

E. Blood Pressure and Cardiovascular Disease

- Exercise 17.25
  Directions: Ignore Questions (a)-(c) in the textbook. Instead analyze the data with MINITAB and then answer these questions:
  (a) Define the response and predictor variables.
  (b) What is MINITAB's value for the odds ratio?
  (c) Use Four Steps to determine whether low vs. high blood pressure is related to cardio (heart) disease.
  (d) Find and interpret a 95% CI for the odds ratio.

F. Healthy Versus Failed Companies

- Refer to Example 3 on page 16 of the Topic 10 Notes. Open the `CACL` data file and reproduce the MINITAB output shown on page 17. (Use the MINITAB steps on page 16.)

  Questions:
  (a) Refer to your notes. What are the odds that a firm is healthy if the CACL is 1.5?
  (b) Refer to your notes. What are the odds that a firm is healthy if the CACL is 2.5?
  (c) Recall that the MINITAB output shows an odds ratio of 14.68. Show how to calculate the answer to (b) from the answer to (a) without using the fitted logistic regression equation.
  (d) What is the probability that a firm is healthy if the CACL is 3.5?
  (e) Find the CACL score which indicates that there is just as much chance that a firm will fail as there is that it will succeed.

Reminder:

- *All* predictors used in a model must be entered into the Model box.
- MINITAB also requires you to enter all *binary* variables into the Factors box.

G. Pizza Hut

Undergraduate students at Miami University in Oxford, Ohio were surveyed in order to evaluate the effects of price and gender on the purchase of pizza from Pizza Hut. 220 students were asked to suppose that they were going to have a large two-topping pizza delivered to their residence. They were asked to select from either Pizza Hut or from another pizza shop of their choice.

The price they would have to pay to get a Pizza Hut pizza varied from survey to survey. For example, some surveys used the price currently charged by the Oxford Pizza Hut, \$11.49. Other prices investigated were \$8.49, \$9.49, \$10.49, \$12.49, \$13.49, and \$14.49. The data set contains three variables: Gender (1 = male, 0 = female), Price (in dollars), and Purchase (1 = student selected Pizza Hut, 0 = student selected another pizza shop). Find the data as `Pizzahut` on the website.

(a) Define the response and predictor variables. Which of the variables are *binary*? Which are *quantitative*?

(b) Fit a logistic model which predicts the probability that a student selects Pizza Hut pizza, based on Gender.

1. Give the theoretical equation for the log odds.
2. Find the fitted equation for the log odds.
3. Find and interpret the odds ratio.

(c) Fit a logistic model which predicts the probability that a student selects Pizza Hut pizza, based on Price.

1. Give the theoretical equation for the log odds.
2. Find the fitted equation for the log odds.
3. Find and interpret the odds ratio.

(d) Fit a logistic model which predicts the probability that a student selects Pizza Hut pizza, based on both Gender and Price.

1. Give the theoretical equation for the log odds.
2. Find the fitted equation for the log odds.
3. Find and interpret the odds ratios for both predictor variables.

(e) At 5% significance, which of these three models would you choose as the "best conservative" model, and why? **Use the chosen model to answer all remaining questions.**

(f) Interpret the odds ratios for the two predictor variables in the database.

(g) Estimate the probability that a female student will select Pizza Hut pizza if the price is \$8.99.

(h) Estimate the probability that a male student will select Pizza Hut pizza if the price is \$11.49.

(i) What price should Pizza Hut charge to ensure that 10% of undergrads choose its pizza?

H. Acceptable Cheese

Read the description of the Cheese data on page A-1 in the Data Appendix near the end of the textbook. Download the data set `Cheese`. For purposes of this problem, suppose that the three variables Acetic, H2S, and Lactic are measured as *percents* of the raw cheese mixture. Suppose further that a particular cheese mixture is considered to be *acceptable* if its taste index is at least 37.

To perform logistic regression, we need to first create a new column in MINITAB:

(Type the title Accept above column C6.) > Data > Code > Numeric to Numeric > (Select Taste for *Code data from columns* and Accept for *Into columns*) > (Type the following:)

Original values	New
0:36.99	0
37:100	1

> OK

Do you see why we created this new column and what it means?

Questions:

(a) Identify the response and predictor variables. Which variables are quantitative? Which are binary?

(b) Use the Drop Method to choose a final model, using $\alpha = .10$. (Use this model for all subsequent questions.) What is the fitted regression equation?

(c) Find and interpret the odds ratios for all potential predictor variables.

(d) Find and interpret 95% confidence intervals for the odds ratios of all predictors used in the final model.

(e) Estimate the probability that a cheese mixture which contains 4% acetic acid, 7% hydrogen sulfide, and 1% lactic acid is acceptable.

(continued)

I. Lawn Service

The marketing manager for a nationally-franchised lawn service company would like to study the characteristics that differentiate homeowners who do and do not use a lawn service. A random sample of 30 homeowners located in a suburb of Chicago was selected; 15 homeowners did not use a lawn service (code 0) and 15 did (code 1).

Additional information available concerning these 30 homeowners includes family income (in thousands of dollars), lawn size (in thousands of square feet), attitude toward working outdoors (0 = unfavorable, 1 = favorable), number of teenagers in the household, and age of head of household, in years. See the data set `Lawn`.

(a) Identify the response and predictor variables. Which are quantitative? Which are binary?

(b) Suppose one purpose of the study is to understand the relationship between using a lawn service and the number of teenagers. Interpret the odds ratio generally recommended as most accurate for this purpose.

(c) Suppose one purpose of the study is to understand the relationship between using a lawn service and age. Interpret the odds ratio generally recommended as most accurate for this purpose.

(d) Use the Drop Method to choose a final model, using 10% significance. (Use this model for all remaining questions.) What is the fitted regression equation?

(e) Briefly describe how the variables in your chosen model affect the odds (and probability) of using a lawn service, based on the signs (positive or negative) of their estimated slopes.

(f) Find and interpret the odds ratios for all potential predictor variables.

(g) Estimate the probability of purchasing lawn service for a 48-year-old homeowner with a family income of \$100,000, a lawn size of 2500 square feet, a negative attitude toward working outdoors, and one teenager in the household.

(end of homework)