In **Part II** below, you’ll have a chance to apply logistic regression (LR) to new data sets. But first get comfortable in **Part I** by “walking through” the Topic 10 examples/data and reproducing their graphs/output!

- EX1 (Ames Homes)
- EX2 (Adult Gamers)
- EX3 (Business Firm CACL)

**PART I**

Two of the three Topic 10 examples featured in **Part I** (EX2, EX3) use *simple* LR (one predictor only) while EX1 uses *multiple* LR (two predictors.) **Good Luck!**

**A. EX1 (Ames Homes)**  
*(Follow directions. Answer selected questions (Q) along the way.)*

1. Open Notebook p. 178 to see a “snapshot” of the Ames Homes data that we analyzed previously as multiple regression in Topic 9.
2. Navigate to the MINITAB Data Sets page of the Stats website and open the Ames Homes data file. (It’s found under Topics 7–9.)

**QUESTION 1**

(a) What’s the sales price for the 15th home listed in the data file?  
(b) What are the 15th home’s characteristics (area and age)?

3. Now turn to p. 180. Our goal is to use LR to predict

   \[ z = \text{Premium Sale} \]

   so we need a new MINITAB data column. Apply these MINITAB steps:

   (Type the title **Premium Sale** above column C5) > Data > Code > Numeric to Numeric  
   > (Select **Sales Price** for Code data from columns and **Premium Sale** for Store coded data in columns)  
   > (Type the following):

<table>
<thead>
<tr>
<th>Original values</th>
<th>New</th>
</tr>
</thead>
<tbody>
<tr>
<td>0:124999</td>
<td>0</td>
</tr>
<tr>
<td>125000:1000000</td>
<td>1</td>
</tr>
</tbody>
</table>
   > OK

*(See screenshot next page)*
Here’s a screenshot:

Afterward, Homes 1–2 should show “1” (they’re premium sales) while Homes 3–5 should show “0” (they sell for less than $125,000 and so aren’t premium.)

4. Great, now that we have the data we can do some regression!

Turn to p. 189 and let’s re-create those fitted-line plots from both STANDARD and LOGISTIC regression:

(1) STANDARD Regression Plot:
Stat > Regression > Fitted Line Plot
> (Set response = Premium Sale, predictor = Area)

(2) LOGISTIC Regression Plot:
Stat > Regression > Binary Fitted Line Plot
> (Set response = Premium Sale, predictor = Area)

Did you reproduce the graphs on p. 189? Now let’s run some numbers!
**QUESTION 2 (Software Predictions)**

(a) What probability for a premium sale does STANDARD regression predict for a home with 2400 square feet? Is this a sensible prediction?

(b) What probability for that same home is predicted by LOGISTIC regression?

Use software for QUESTION 2. Recall that to *predict* in MINITAB we first need to *fit the model*. Then we run Regression again to get the prediction:

(a)  
Stat > Regression > Regression > Fit Regression Model  
   > (Response = Premium Sale, Continuous Predictor = Area)

   Stat > Regression > Regression > Predict  
   > (Enter Area = 2400)

   *(Check your predicted probability in HW9 Solution online.)*

(b) Prediction with LOGISTIC regression is quite similar. First fit the model, then predict:

   Stat > Regression > Binary Logistic Regression > Fit Binary Logistic Model  
   > (Choose Response in binary response/frequency format)  
   > (Response: Premium Sales, Continuous predictors: Area) > Results  
   > (Display of results: Expanded tables) > OK > OK

   Stat > Regression > Binary Logistic Regression > Predict  
   > (Enter Area = 2400)

   *(Check your predicted probability online.)*

**QUESTION 3 (Hand Calculations)**

To feel comfortable with computer predictions, we need to know that we can verify those numbers by hand!

(a) Write down the STANDARD regression equation from QUESTION 2:

\[
\hat{p} = b_0 + b_1 x
\]

(except substitute numbers for \(b_0\) and \(b_1\)) and plug in \(x = \text{Area} = 2400\) to *hand-calculate* the probability of a premium sale. *Use at least four decimal places.*

Except for rounding, does your answer match MINITAB from QUESTION 2?

(b) Write down the LOGISTIC regression equation from QUESTION 2:

\[
\log \text{odds}(x) = b_0 + b_1 x
\]

(except substitute numbers for \(b_0\) and \(b_1\)) and plug in \(x = \text{Area} = 2400\) to *hand-calculate* the “log odds” of a premium sale.

Now turn to your notes for part (c) on p. 202 to recall how to convert the log odds first to *odds*, then to *probability*. Calculate the predicted probability!

Except for rounding, does your answer match MINITAB from QUESTION 2?
5. The other predictor available in our Ames Homes data is **Age**. As you did earlier for **Area**, get regression plots that predict Premium Sales based on **Age**:

(1) **STANDARD** Regression Plot: Reproduce Premium Sale vs. Age on p. 181!

(2) **LOGISTIC** Regression Plot: Reproduce Premium Sale vs. Age on p. 182!

**QUESTION 4 (Predicting Premium Sale Based on Age)**

(a) What probability for a premium sale does **STANDARD** regression predict for a home that’s 95 years old? Is this a sensible prediction?

(Use either **MINITAB** or hand calculation, your choice!)

(b) What probability for that same home is predicted by **LOGISTIC** regression?

(Use either **MINITAB** or hand calculation, your choice!)

6. Now let’s use both **Area** and **Age** as predictors!

Turn to p. 186 to fit **multiple** logistic regression. Using the **MINITAB** steps there, do you get the output shown on p. 186? (Prof. Whitten edited the less-useful output out of the Notebook.)

7. Predict the probability of a premium sale for a 30-year-old home with 1500 square feet using the **MINITAB** steps on p. 187. Do you get the “Fitted Probability” shown on p. 187?

8. **Graphs for Fun!**

Make the **Contour Plot** shown on p. 187. Notice the different **probability zones** based on Age and Area that you can see in the green-colored plot! (Read the key on the side: Darker green indicates greater probability.)

9. **Graphs for Fun!**

(1) Make the **Surface Plot** shown on p. 188. Notice the **3D Graph Tools** that pops up!

(2) Click your mouse on the **fifth button** to spin the graph in three dimensions (on the vertical Z axis.)

**BY VIEWING THE GRAPH FROM DIFFERENT ANGLES** it’s easier to show a client (for instance, a realtor!) that larger Area and smaller Age increase the probability of a premium sale!

Imagine how impressed a client might be by your presentation! (and the consulting work and fees that might follow.) Logistic regression is a powerful tool!

(continued)
B. EX2 (Adult Gamers)

1. Turn to page 195. The most striking difference about the MINITAB steps this time is that there’s no data set to open like we did for Ames Homes in Example 1!

We’re told on p. 193 that

A survey of 2054 adults … found that … 481 out of 880 men and 582 out of 1174 women … are gamers.

Since the survey records answers from 2054 individuals, we can be sure that there’s a spreadsheet somewhere that records raw data, such as:

<table>
<thead>
<tr>
<th>Individual</th>
<th>Gender</th>
<th>Gamer?</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>male</td>
<td>yes</td>
</tr>
<tr>
<td>2</td>
<td>female</td>
<td>yes</td>
</tr>
<tr>
<td>3</td>
<td>female</td>
<td>no</td>
</tr>
<tr>
<td>4</td>
<td>male</td>
<td>yes</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>2054</td>
<td>male</td>
<td>no</td>
</tr>
</tbody>
</table>

But in the meantime someone has been nice enough to total these numbers for us! So of course MINITAB enables data entry in this summarized form also.

Can you do it? Open a blank MINITAB worksheet and enter the Gamer survey data into the worksheet exactly as shown on p. 195. Then use the MINITAB steps on p. 195 to fit the logistic regression.

Check to find items \(1—6\) in the output. If you can, you’ve succeeded!

2. In Topic 10 Notes Prof. Whitten defined the \(x = \text{Gender}\) variable as

\[
x = \begin{cases} 
1 & \text{if man} \\
0 & \text{if woman}
\end{cases}
\]

but of course this classification is arbitrary and could just as well have been switched:

\[
x = \begin{cases} 
1 & \text{if woman} \\
0 & \text{if man}
\end{cases}
\]

Let’s try MINITAB again with this second gender classification. To do this, adjust the MINITAB spreadsheet accordingly:

<table>
<thead>
<tr>
<th>C3-T</th>
<th>C4</th>
<th>C5</th>
<th>C6</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Gender</td>
<td>Gamer</td>
<td>Total</td>
</tr>
<tr>
<td>Women</td>
<td>1</td>
<td>582</td>
<td>1174</td>
</tr>
<tr>
<td>Men</td>
<td>0</td>
<td>481</td>
<td>880</td>
</tr>
</tbody>
</table>

and re-run logistic regression again, this time with women coded as “1”!

(See QUESTIONS TO ANSWER FOR EX2 next page)
3. QUESTIONS TO ANSWER FOR EX2:

(a) What’s the value of the odds ratio when women are coded as “1”?

(Record all decimal places from MINITAB answer. Check answer online.)

(b) INTERPRET the odds ratio from the previous question.

(c) Refer to your notes for the odds ratio and its interpretation on p. 194. There’s a mathematical trick to determine the odds ratio for women compared to men from the odds ratio for men compared to women without re-running LR.

Can you guess the trick? (Check or see answer online.)

(d) Use MINITAB to predict the probability that women game. (Record all decimal places of the MINITAB “Fitted Probability.”)

MINITAB’s answer should be the same as the survey sample probability for women that we calculated on p. 194! Is it?

C. EX3 (Business Firm CAACL)

1. • Open the CAACL Ratio data from the MINITAB Data Sets page. Make a Binary Fitted Line Plot. The plot should show that the probability that a firm is healthy increases as CAACL increase! Does it?

• Now get the full LR output by using the MINITAB steps shown on p. 199.

Check to find items (1)—(6) in the output. If you can, you’ve succeeded!

2. QUESTIONS TO ANSWER FOR EX3:

(a) Refer to your hand calculation in (c) on p. 202 for the probability that a firm with CAACL 1.5 remains healthy.

Now use MINITAB to predict the same probability! What’s MINITAB’s answer? (Use all decimal places. Check answer online.)

(b) Similarly use MINITAB to predict the probability that a firm with CAACL 2.5 remains healthy.

(c) What’s the probability of health for a firm whose CAACL is 0.5?

(d) Challenge Question: Find the “tipping point” CAACL that indicates that a firm has a 50% chance to remain healthy. (TIP: See similar calculations in Topic 10 Notes.)

(continued)
PART II

In Part II we’ll analyze three data sets found on the MINITAB Data Sets page for Topic 10:

- Pizza Hut Pizza
- Cereal Purchases
- Lawn Service

D. Pizza Hut Pizza

Undergrads at Miami University in Oxford, Ohio were surveyed to evaluate the impacts of price and gender on the purchase of pizza from Pizza Hut.

220 students were asked to order pizza from either Pizza Hut or from any other competing pizza shop based on the price of Pizza Hut pizza. The hypothetical prices for a large two-topping pizza were $8.49, $9.49, $10.49, $11.49, $12.49, $13.49, and $14.49.

The data contain three variables: Gender (1 = male, 0 = female), Price (in dollars), and Purchase (1 = student selected Pizza Hut, 0 = student selected a competing pizza shop).

QUESTIONS:

1. (Multiple Choice) Open the data file and take a look at the numbers. Which variable is the response variable $z$ for logistic regression?
   (a) Gender
   (b) Price
   (c) Purchase

2. Fit multiple LR. (Use all available predictors.) Write down the LR equation.

3. One of the two predictors is really not helping! (It’s not significant.) Drop that variable and run simple LR with the remaining significant variable. (Be sure to run “Expanded Tables” so you can see the $P$-values.)

   Write down the new LR equation. Use this new model to answer all remaining questions.

4. Interpret the odds ratio for the new model.

5. Predict the probability that a female student will select Pizza Hut pizza if the price is $8.99. (Use the LR equation and hand calculations.)

6. Predict the probability that a female student will select Pizza Hut pizza if the price is $8.99. (Repeat the previous answer but this time use MINITAB Predict.)

7. Predict the probability that a male student will select Pizza Hut pizza if the price is $11.49. (Use hand calculations or MINITAB, your choice.)

8. What price should Pizza Hut charge to ensure that 10% of undergrads choose its pizza?

(continued)
E. Cereal Purchases

Open the CerealPurchase data file.

The data show the choices of 71 families to purchase a brand-name cereal or not, whether the family has children, has seen a TV ad for the cereal, and the family’s take-home annual income (after taxes) in thousands of dollars.

QUESTIONS:

1. (Multiple Choice)
   Open the data file and take a look at the numbers. Which variable is the response variable z?
   (a) Bought
   (b) Income
   (c) Children
   (d) ViewAd

2. Fit the full model. (Use all predictors.) Notice that some predictors aren’t significant.
   
   Apply the Drop Method until all remaining predictors are significant at 5%.
   (Use this model to answer all remaining questions.)

   Write down the LR equation that results from using the Drop Method.

3. Based on the odds ratio for Children, does having children make a family more likely or less likely to purchase the cereal? Also INTERPRET the ratio!

4. Based on the odds ratio for ViewAd, does viewing the ad make a family more likely or less likely to purchase the cereal? Also INTERPRET the ratio!

5. Predict the probability that a family with no children and that hasn’t seen the ad purchases the cereal.

6. Predict the probability that a family with children and that has seen the ad purchases the cereal.

(continued)
The marketing manager for a nationally-franchised lawn service company studied 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.

QUESTIONS:

1. Identify the response variable $z$ in the data file.

2. Use the Drop Method to choose a final model, using 10% significance. (Use this model to answer all remaining questions.) What’s the LR equation?

3. Consider the odds ratios for all predictor variables that survive the Drop Method. INTERPRET each of these odds ratios.

4. Predict 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.

   Use the LR equation to calculate the probability by hand.

5. Make the same prediction but this time use MINITAB! What’s the MINITAB prediction? (Include all available decimal places in your answer.)

(end of HW9)