

Trashball: A Logistic Regression Classroom Activity

Christopher H. Morrell and Richard E. Auer
Loyola College in Maryland

Journal of Statistics Education Volume 15, Number 1 (2007), www.amstat.org/publications/jse/v15n1/morrell.html

Copyright © 2007 by Christopher H. Morrell and Richard E. Auer all rights reserved. This text may be freely shared among individuals, but it may not be republished in any medium without express written consent from the authors and advance notification of the editor.

Key Words:Odds, Odds ratio; Problem solving.

Abstract

In the early 1990's, the National Science Foundation funded many research projects for improving statistical education. Many of these stressed the need for classroom activities that illustrate important issues of designing experiments, generating quality data, fitting models, and performing statistical tests. Our paper describes such an activity on logistic regression that is useful in second applied statistics courses. The activity involves students attempting to toss a ball into a trash can from various distances. The outcome is whether or not students are successful in tossing the ball into the trash can. This activity and the adjoining homework assignments illustrate the binary nature of a response variable, fitting and interpreting simple and multiple logistic regression models, and the use of odds and odds ratios.

1. Introduction

In 1991, [Robert Hogg \(1991\)](#) prepared a summary of what many of the prominent statisticians of the day were saying about the future of statistical education. The National Science Foundation had partially funded a workshop where this conversation had taken place and also funded many other projects on statistical education. Among the proposals suggested by [Hogg \(1991, p342\)](#) was one for student involvement in the entire statistical process saying that “projects give students experience in asking questions ... formulating hypotheses ... collecting data[and] analyzing data.” [Cobb \(1993\)](#) summarized a dozen classroom projects that were funded by the National Science Foundation and, in an appendix, made an incredible amount of information related to in-class projects in statistics available to instructors. [Garfield \(1993\)](#) described cooperative learning in the context of small groups performing statistical activities. [Gnanadesikan, Scheaffer, Watkins, and Witmer \(1997\)](#) described an activity-based statistics course where “hands-on-activities ... promote the teaching of statistics more as an experimental science and less as a traditional course in mathematics.” [Zacharopoulou \(2006\)](#) noted that “such ... activity improves students’ understanding, [and] makes teaching of the subject a pleasant experience and improves the class climate.”

Following these recommendations, we describe an in-class activity that has proved to be very successful in illustrating the concepts of designing statistical studies, fitting logistic regression models, and making accurate interpretations of the ensuing predicted probabilities, odds, and odds ratio. We also describe homework assignments to be given pre-activity and post-activity that expand student involvement further. The activity involves students attempting to toss a ball into a trash can from various distances. The outcome is whether or not students are successful in tossing the ball into the trash can.

The activity was conducted in ST465, *Experimental Research Methods*, in the fall semesters of 2001, 2003, and 2005. This is a junior/senior level course taken by Mathematical Science majors and minors at Loyola College in Maryland. [Roback \(2003\)](#) describes the use of logistic regression in a course that covers similar topics to our ST465.

In many statistical methods or linear models courses, instructors initially concentrate on continuous numerical response variables. Recently, it has become easier to also consider response variables that are either binary or categorical. This has been made possible as more introductory linear models and statistical methods books now include chapters or sections devoted to logistic regression (see [Kleinbaum, Kupper, Muller and Nizam \(1998\)](#), [Kutner, Nachtsheim, and Neter \(2004\)](#), [Moore and McCabe \(2006\)](#) (on CD-ROM and print supplement), [Ott and Longnecker \(2001\)](#), and [Ryan \(1996\)](#)). A number of articles in the *Journal of Statistics Education* have dealt with analyzing data using logistic regression: [Andrews \(2005\)](#), [Duchesne \(2003\)](#), [Johnson and Dasgupta \(2005\)](#), [Love \(1998\)](#), and [Simonoff \(1997, 1998\)](#). In addition, [Willoughby \(2002\)](#) provides a description of logistic regression used in modeling Canadian football as well as citations to other sports examples. Finally, the CBS *Numb3rs* TV show has a web-site containing activities related to the show. The episode “All’s Fair” applies logistic regression to estimate the probability of where a suspect will strike next ([Souhrada, 2006](#)). These papers have described the use of logistic regression to analyze existing data sets. However, the *Journal of Statistics Education* does not currently contain any examples of projects actively involving the students that could be used to motivate logistic regression and, simultaneously, provide data that could be analyzed using a logistic regression model.

Our proposed activity can be viewed as a “striking demonstration” similar to those described in [Sowey \(2001\)](#) and the follow-up letter to the editor in Vol. 10, No. 1, 2002. [Sowey \(2001\)](#) suggests that “intellectual excitement grows from teaching where ... some striking demonstrations are introduced that will arouse students’ curiosity and/or provoke reflection.” Our activity involves student interaction as well as discovery. In addition, principles learned earlier in the course for linear regression models are applied within the context of logistic regression reinforcing these earlier concepts. The activity is clear, self-contained, and can be easily grasped by the audience. The students can immediately understand that linear regression is not appropriate for the binary data that is being collected and students become curious as to which explanatory variables may be important in predicting the binary outcome variable.

In [Section 2](#), pedagogical benefits of the activity are explored. [Section 3](#) introduces a modest pre-activity homework assignment that precedes the activity period. In [Section 4](#), we offer a detailed description of an engaging classroom activity that can be used to motivate the need for and application of the logistic regression model. This activity may also be used to discuss experimental design issues. [Section 5](#) presents the analysis of the data collected from the activity in the fall of 2003. [Section 6](#) describes a post-activity homework assignment that is based on the data that is collected and the models that are fit in class. A summary based on our experiences are provided in [Section 7](#).

2. Benefits of the activity

It is likely that the activity being proposed would best fit a second course in applied statistical methods. Students would have already covered multiple linear regression models and would be prepared to consider a related model that features a binary dependent variable. While focusing on the logistic regression model, the activity would also consider the following issues:

- designing an experiment,
- considering many explanatory variables that may help predict a binary dependent variable,
- specifying null and alternative hypotheses related to these explanatory variables,
- careful collection and computer entry of data,
- learning exactly how the logistic regression model works,
- using a statistical package (like Minitab) to fit the models, and
- calculating and interpreting odds and odds ratios.

An additional issue often underplayed in statistics courses is the concept of operational definitions. [Melton \(2004\)](#) declares that “operational definitions can be loosely described as descriptions that allow two people to look at the same thing at the same time and record the same measurement.” Not only does our proposed activity force students to carefully consider operational definitions, but students also find it educational and enjoyable to use class time to generate real data based on their own performance.

Having conducted the activity in three different years, improvements were made each time. On the basis of what did and did not work well, we suggest one particular way to implement the activity that seems to optimize the benefits of the activity. Using our suggestions as a model, instructors may choose to deviate as they see fit.

3. The pre-activity homework assignment

A few days before the activity is conducted, students are asked to read the section of their textbook on logistic regression.

At this time, students are also given a description of a data collection activity and are asked to answer a set of questions regarding the optimal way for conducting the activity. This involvement ensures that they initially consider many important statistical issues and also orient themselves to the logistic model and the upcoming activity. This pre-activity stage helps to make the in-class activity run smoothly.

The following activity description and questions may be used by the instructor. Note that Appendix A presents possible answers to this pre-activity assignment.

The in-class activity description. “Consider an activity where students throw a ball at a waste paper basket. Statistically, we are interested in what type of explanatory variables may impact the likelihood of making the shot.”

The pre-activity homework assignment. “Make a list of five to ten potential explanatory variables. Include some that you expect to be significantly related to whether a shot is made or not and also consider some that would not be related. On your list, include answers to the following questions for each variable:

- a. Is this explanatory variable categorical, numerical discrete, or numerical continuous?
- b. What null and alternative hypotheses need to be specified regarding the relationship between the explanatory variable and whether a shot is made?
- c. Do you expect to reject or fail to reject the null hypothesis?
- d. How would you crisply define the explanatory variable so that quality data could be collected?”

4. The in-class activity

Because the activity has students attempting to toss a ball into a trash can, the outcome or response variable is whether or not the shot is made. We choose to simply call this binary variable “ShotMade.” In keeping with the notion of operational definitions, the class should discuss the exact definition of these binary outcomes. In our activities, we have decided a ball is not allowed to bounce into the can from the floor, but we do count shots where the ball bounces out of the can. Furthermore, the basket is set flush against a wall and students are allowed to use the “backboard.”

The instructor will surely receive many obvious and even some clever explanatory variables from the homework assignment and may choose to deviate from the variables considered in our paper. From our experience, it is important to not over-complicate the activity nor to use too many explanatory variables. [Gnanadesikan, et al. \(1997\)](#) warned: “If the instructor does not plan carefully, activities can become boring, confusing and a repetitive waste of time.”

For our classes, we have settled on using just three explanatory variables. With more, the activity may fail to be relaxed, enjoyable, and understandable. Distance is used as one of our variables since success in making a shot likely depends on how close the student stands to the trash can. How distance is exactly measured is another issue of operational definitions. Where the student actually stands, how the arm is extended, or if the body is allowed to lean each impact distance. In conducting the activity, the exact location of the trash can set against a wall must be determined. Using a tape measure and masking tape on the floor, distances should be marked off from 5 feet through 12 feet from the trash can. The remaining two explanatory variables included in the design of the experiment include the orientation of the trash can and the gender of the student. Since a rectangular trash can was used, throwing at the narrow side yields a deep target (see [Figure 1](#)). After rotating through 90°, a wide but shallow target is presented. It is hypothesized that the likelihood of making a shot will decrease with increased distance. While it was our prior belief that the probability of making the shot will be lower when tossing at the wide/shallow target compared to the narrow/deep target, instructors are urged to keep this notion open during class discussion. Students should consider their own expectations before the experiment is underway to generate hypotheses that can be tested once the data is collected. While we expect that there would be no differences in tossing skill based on gender, it is recorded primarily to include one factor that would be expected to be an insignificant predictor of ShotMade. It

is a comforting to see that the statistical process is able to eliminate a variable that was thought unlikely to be important. It is also interesting to note that many medical studies used to be conducted exclusively on men. In recent years, to account for possible gender differences, studies are required to include both male and female subjects to allow for the estimation of possible differences in outcome variables between men and women. In 1993, the [NIH Revitalization Act \(2001\)](#) requires that “all NIH-funded clinical research will be carried out in a manner sufficient to ... determine whether the intervention ... being studied affects women or men ... differently.” While students may not expect a significant gender effect in this activity, it has been interesting for the students to record this data and explicitly test the hypotheses.



[Figure 1](#)

Figure 1. Narrow/Deep (Left) and Wide/Shallow (Right) orientation of trash can.

Ideally we would want to achieve a balanced full factorial design in the factors of interest: distance, orientation, and gender. However, due to the make-up of the class it is unlikely that this can be achieved. To ensure a reasonable balance of all the design variables in our classes, we attempted to construct a design that, given the size of the class, endeavored to ensure a partially balanced factorial design. Distance, orientation, and gender are as uniformly distributed as possible. There should be approximately as many tosses at both orientations from each distance and both genders should be as evenly represented at each distance/orientation combination as possible. However, such pre-planning may prove to be difficult to achieve because of uncertainty of the actual attendance on the class day. Despite this, we recommend that the planned settings of the explanatory variables are entered into a computer data file before class time. Adjustments can easily be made to the design as the activity progresses. Given the availability and wide use of the Minitab software package ([Ryan, Joiner, and Cryer 2004](#)), we describe the in-class activity using this software. Instructors are encouraged to make the simple adaptations if they want to use other systems.

[Appendix B](#) presents a sample Minitab worksheet. It lists the planned settings of the three explanatory variables assuming a sample of four male and four female students. This worksheet can be easily adapted to handle more or less students. Having students toss the ball in the order that appears on the worksheet eliminates many class time and design problems and insures maximum simplicity. See [Appendix C](#) for a detailed description of how this worksheet is utilized in class.

If the class size is small only limited data will be obtained if each student tosses the ball once. This is why sample size is

effectively increased by having each student make a number of attempts from varying combinations of distance and orientation. The repeated observations may induce some non-independence and this should be discussed during the execution of the experiment. When the data from conducting the activity is analyzed in [Section 6](#), the reader can see that each student in 2003 made three attempts at tossing the ball into the trash can. It is not surprising that one of the students nicknamed the activity “Trashball.”

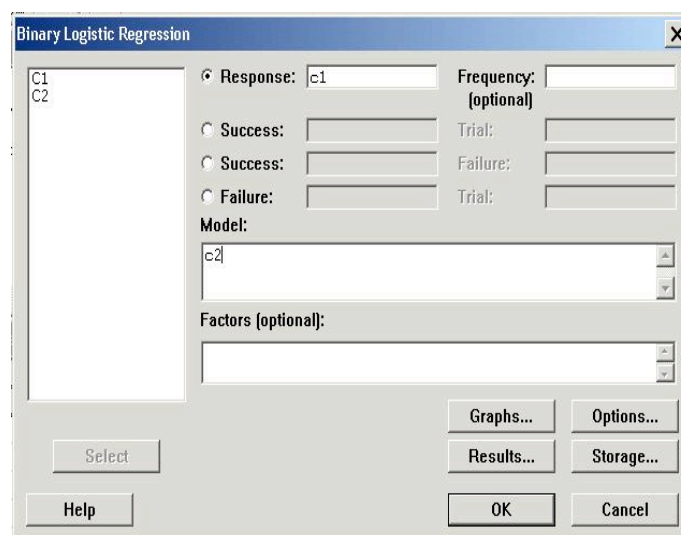
As the activity progresses, the data for the response variable may be entered into the ShotMade column (column 2) of the Minitab worksheet. A 1 is entered when the shot is successfully made and 0 otherwise. Two students should be assigned the job of carefully entering the data. Two students are needed so that one will be on duty when the other is making their tosses. Similarly, two other students should be assigned the job of changing the orientation of the trash can and two more hand out the balls and enforce the distance measurements. To capture the attention of the students, the results are immediately displayed using a classroom projection system.

This activity may be most beneficial to conduct just after completing the topic of multiple linear regression on a continuous numerical dependent variable. When students consider the description of the trash can activity, they may begin to realize that the response variable has only two outcomes. A discussion of the assumptions behind linear regression leads to the realization that linear regression is not appropriate for this data. In addition, it can be pointed out that linear regression may lead to predictions that are negative or greater than one. By now, the students will have discovered that they are actually trying to model the probability of a success and that the results must be values between 0 and 1. Having completed the pre-activity homework, students should be prepared to generate their data and fit logistic regression models.

Once the data is fully entered into the worksheet, students may fit a one-variable model by simply typing:

```
MTB > BLogistic c1 = c2;
SUBC> Logit;
SUBC> brief 2.
```

To include more variables, c3 and c4 may be included just to the right of c2 in the first line. Alternatively, the drop down menu approach may be utilized by clicking on options: Statistics > Regression > Binary Logistic Regression. Entering c1 into the Response box and c2 in the Model box and then clicking on OK will produce the same output (See [Figure 2](#)). Asking students for help during this computer process would be of educational benefit and would also serve to keep the class involved.



[Figure 2](#)

Figure 2. Minitab Dialog Box for Binary Logistic Regression.

5. Results of Conducting the Classroom Activity

In the fall semester of 2003, the activity was conducted and, in this section, we present the analysis of the data obtained. Note that [Tables 1\(a\) - \(c\)](#) display cross tabulations based on a sample of 14 students each of whom made three attempts at the trash can. They illustrate the resulting balance in the explanatory variables. With each count in the tables representing a single shot made toward the trash can, [Tables 1\(a\)](#) and [Table 1\(c\)](#) demonstrate that gender is well balanced across orientation and shot distance. [Table 1\(b\)](#) similarly displays a reasonable balance between orientation and shot distance.

Table 1(a). Cross tabulation of gender by orientation of target.

	Gender		
	Male	Female	Total
Narrow Target	9	12	21
Wide Target	9	12	21
Total	18	24	42

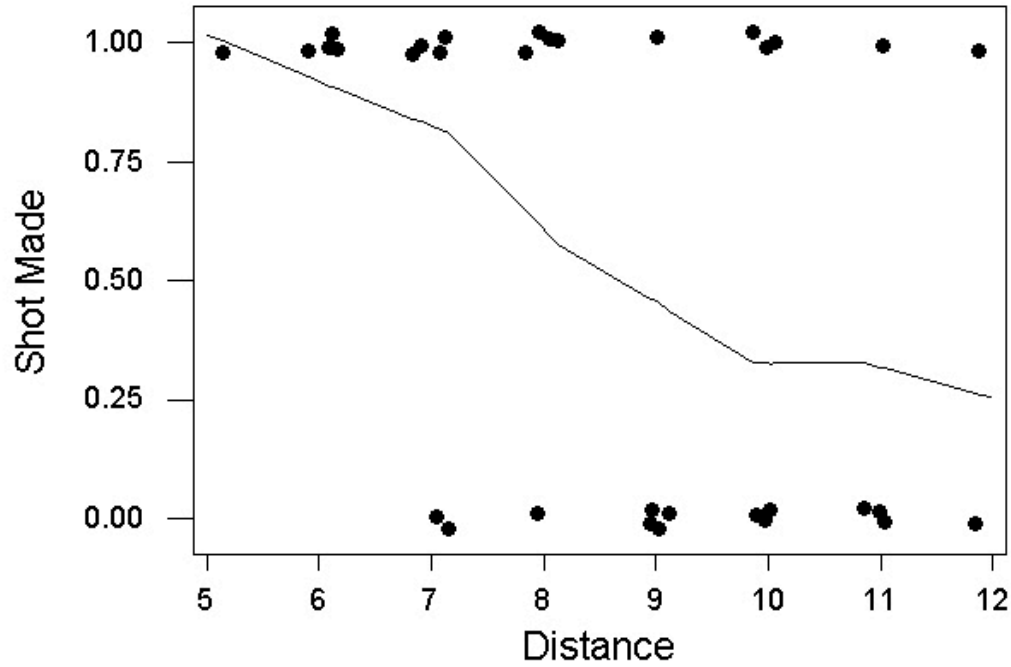
Table 1(b). Cross tabulation of orientation by distance from target.

	Shot Distance (in feet)								Total
	5	6	7	8	9	10	11	12	
Narrow Target	5	0	5	0	5	2	4	0	21
Wide Target	1	4	1	5	0	4	0	6	21
Total	6	4	6	5	5	6	4	6	42

Table 1(c). Cross tabulation of gender by distance from target.

	Shot Distance (in feet)								Total
	5	6	7	8	9	10	11	12	
Male	2	2	3	2	2	3	2	2	18
Female	4	2	3	3	3	3	2	4	24
Total	6	4	6	5	5	6	4	6	42

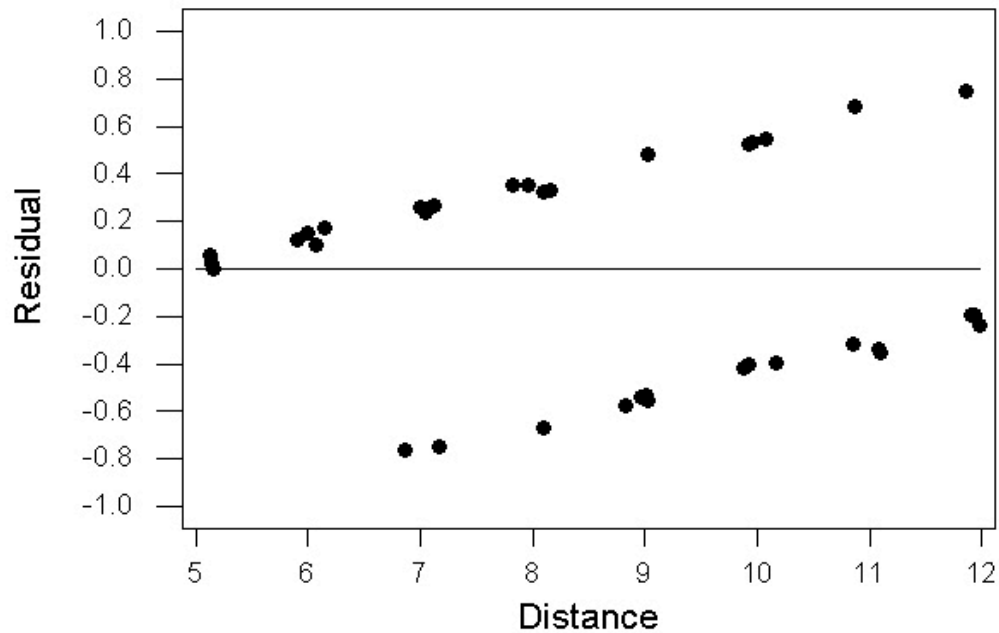
[Figure 3](#) is a plot of ShotMade versus distance (with jitter added to the points) with the Lowess curve overlaid on the plot to illustrate the trend in the data. As one would expect, with increased distance, more shots will be missed. Consequently, one should also expect the predicted probability of making a shot to decline with distance.



[Figure 3](#)

Figure 3. ShotMade versus distance between the thrower and the trash can. Jitter is added to the points to show the repeated observations. The Lowess curve is overlaid.

To demonstrate the inadequacies of the linear regression model, the linear model is fit to the data. [Figure 4](#) shows residuals with a decidedly non-random pattern with two increasing lines of points evident in the plot.



[Figure 4](#)

Figure 4. Residuals vs. Distance from the linear regression model of ShotMade with distance (with jitter).

Note that all of the errors on the top half of the figure result from shots that were made (ShotMade = 1). All predicted probabilities will be less than unity. Given that predicted probabilities will be nearest unity for short distances, these errors will be positive yet small. As the distance increases, the predicted probabilities shrink and the errors grow. For all of the shots that are missed (ShotMade = 0), negative residuals will result, yielding the points on the bottom half of the figure. Since the large distances will yield the smallest predicted probabilities, these will lead to small errors. For the missed shots, the errors become more negative as distance lessens. When such patterns appear on a residual plot, this is usually a sign of an inadequate model. Indeed, the linear model is not appropriate for this binary data, so we turn our attention to fitting logistic regression models. The Minitab output below describes the logistic regression fit for ShotMade as a function of only the distance between the thrower and the trash can.

Minitab Output 1. Logistic regression fit for ShotMade with one explanatory variable: distance between the thrower and the trash can.

```
MTB > BLogistic 'Shot Made' = Distance;
SUBC> Logit;
SUBC> Brief 2.
```

Binary Logistic Regression: ShotMade versus Distance

Link Function: Logit
Response Information

Variable	Value	Count
ShotMade	1	25 (Event)
	0	17
	Total	42

Logistic Regression Table

Predictor	Coef	SE Coef	Z	P	Odds Ratio	95% CI	
						Lower	Upper
Constant	5.204	1.695	3.07	0.002			
Distance	-0.5499	0.1842	-2.98	0.003	0.58	0.40	0.83

Log-Likelihood = -22.294

Test that all slopes are zero: G = 12.102, DF = 1, P-Value = 0.001

Goodness-of-Fit Tests

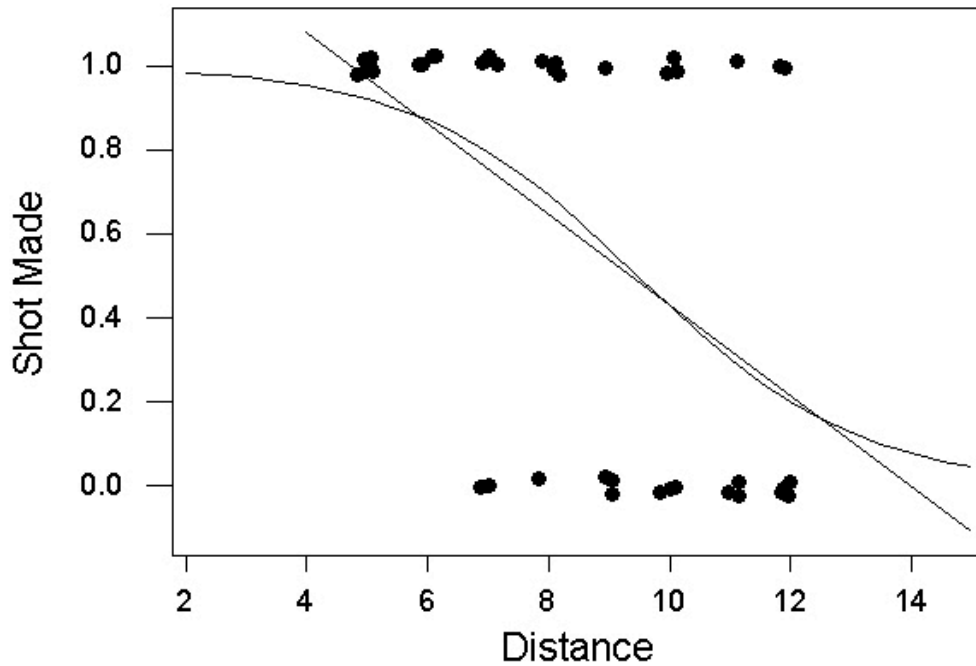
Method	Chi-Square	DF	P
Pearson	5.542	6	0.476
Deviance	6.488	6	0.371
Hosmer-Lemeshow	5.542	6	0.476

At the bottom of this Minitab output, all three goodness of fit tests yield p-values over 0.05 indicating that this model provides an adequate description of the data. Values less than 0.05 would be an indicator of a poorly fitting model. Note, also, that the p-value for the distance variable ($0.003 < 0.05$) suggests a significant predictor variable. The predicted probability of making a shot as a function of distance x becomes:

$$P(\text{the shot is made from distance } x) = \frac{\exp[5.204 - 0.5499x]}{1 + \exp[5.204 - 0.5499x]}$$

[Figure 5](#) illustrates two lines: the fitted linear and logistic regression models. Both models validate our hypothesis that the probability of a shot being made will decrease with distance. But the fitted linear model clearly shows that predictions can fall outside $[0,1]$, the allowable range for probabilities. This concurs with the evidence we found earlier of a poor linear fit. But the two models do agree quite well in the 0.2 to 0.8 range of probabilities. The odds ratio of 0.58 indicates that the odds of making a shot is reduced nearly in half for each additional foot one moves away from the trash can. Had distance not been a significant factor, the odds would have remained constant as the tossing distance increases. This would have yielded an odds ratio of unity. The fact that the 95% Confidence Interval does not contain one also suggests the significant impact of distance.

Note that the odds ratio 0.58 equals $e^{-0.54999}$ where -0.54999 is the coefficient of distance in the logistic model. If one were to find the form of $P(\text{the shot is made from distance } x)$ and $P(\text{the shot is not made from distance } x)$, the ratio of these two answers would yield the odds of making the shot from distance x . Then finding $P(\text{the shot is made from distance } x + 1)$ and $P(\text{the shot is not made from distance } x + 1)$, the ratio of these two answers would yield the odds of making the shot from distance $x + 1$. The odds ratio is defined as the ratio of the latter odds to the former odds. If the mathematical sophistication of the students allows, the instructor may consider asking the students to confirm that the odds ratio can be expressed as $e^{-0.54999}$.



[Figure 5](#)

Figure 5. The fitted linear and logistic regression models. Jitter is included in the observed data points.

Having covered this example of simple logistic regression, the class may now move onto multiple logistic regression by incorporating the additional explanatory variables measured during the experiment. The Minitab output from fitting the logistic model using all of the explanatory variables (distance, gender, and orientation of trash can; but no interactions) is given below. Note that the indicator variable for gender is 1 for females and 0 for males. The indicator variable for trash can orientation is 1 for the narrow/deep target and 0 for the wide/shallow alignment.

Minitab Output 2. Multiple Logistic regression fit for ShotMade with three explanatory variables: distance, gender, and orientation of trash can.

```
MTB > BLogistic 'Shot Made' = Distance Orientation Gender;
SUBC> Logit;
SUBC> Brief 2.
```

Binary Logistic Regression: Shot Made versus Distance, Orientation, ...

Link Function: Logit

Response Information

Variable	Value	Count	
Shot Mad	1	25	(Event)
	0	17	
	Total	42	

Logistic Regression Table

Predictor	Coef	SE Coef	Z	P	Odds		95% CI	
					Ratio	Lower	Upper	
Constant	5.942	1.978	3.00	0.003				
Distance	-0.7422	0.2281	-3.25	0.001	0.48	0.30	0.74	
Orientation	2.3106	0.9831	2.35	0.019	10.08	1.47	69.24	
Gender	-0.1512	0.8266	-0.18	0.855	0.86	0.17	4.34	

Log-Likelihood = -18.667

Test that all slopes are zero: G = 19.357, DF = 3, P-Value = 0.000

Goodness-of-Fit Tests

Method	Chi-Square	DF	P
Pearson	11.862	16	0.753
Deviance	13.061	16	0.668
Hosmer-Lemeshow	5.562	8	0.696

Since gender is the least significant variable (the p-value 0.855 is the largest of the three variable p-values and it is larger than 0.05), it is dropped from the model. Note also that the confidence interval on the gender odds ratio does contain one. This tells us that the odds ratio may equal one. This, in turn, means there would be no difference in the odds of making a shot (while adjusting for other factors) between these men and women students (that is, as the model indicator variable moves from 0 to 1). The next step in the backwards elimination yields all significant variables and, therefore, the final model is summarized in the output below:

Minitab Output 3. Parameter estimates of the final multiple logistic regression fit after backward elimination.

```
MTB > BLogistic 'Shot Made' = Distance Orientation;
SUBC> Logit;
SUBC> Brief 2.
```

Binary Logistic Regression: ShotMade versus Distance, Orientation

Logistic Regression Table

Predictor	Coef	SE Coef	Z	P	Odds		95% CI	
					Ratio	Lower	Upper	
Constant	5.857	1.913	3.06	0.002				
Distance	-0.7425	0.2282	-3.25	0.001	0.48	0.30	0.74	
Orientation	2.3096	0.9827	2.35	0.019	10.07	1.47	69.11	

Log-Likelihood = -18.684

Test that all slopes are zero: G = 19.323, DF = 2, P-Value = 0.000

Goodness-of-Fit Tests

Method	Chi-Square	DF	P
Pearson	3.441	8	0.904
Deviance	3.994	8	0.858
Hosmer-Lemeshow	3.316	7	0.854

The goodness of fit tests again indicates that this model provides an excellent description of this data. The estimated parameters indicate that the probability of making the shot decreases with distance ($p\text{-value} = 0.001 < 0.05$) and that a person has a higher probability of making the shot if the orientation of the trash can ($p\text{-value} = 0.019 < 0.05$) has the narrow/deep target facing the thrower. These findings agree with the expectations described in [Section 4](#). The odds ratio for orientation tells us that the odds of being successful in throwing the ball into the can are 10 times higher if one is throwing at the narrow/deep target versus the wide/shallow target though the confidence interval for this variable is very wide. In addition, the odds are 0.48 times as much for each additional foot one moves away from the trash can. This result is similar to what was found in the one variable model studied earlier. Thus, the predicted probability of making a shot as a function of distance and orientation is given by:

$$P(\text{the shot is made}) = \frac{\exp[5.857 - 0.7425(\text{distance}) + 2.3096(\text{orientation})]}{1 + \exp[5.857 - 0.7425(\text{distance}) + 2.3096(\text{orientation})]}$$

[Table 2](#) contains the observed proportion of shots made by the students in class along with the logistic predicted probabilities for making the shot based on orientation and distance. The modeled probabilities generally conform to the observed proportions in the cells containing data.

Table 2. Observed proportion and modeled probabilities by orientation and distance.

Observed Proportion Logistic Probability	Shot Distance (in feet)							
	5	6	7	8	9	10	11	12
Wide/Shallow Target	1.00 0.895		0.60 0.659		0.20 0.305	0.00 0.173	0.25 0.090	
Narrow/Deep Target	1.00 0.989	1.00 0.976	1.00 0.951	0.80 0.903		0.75 0.677		0.33 0.322

6. The post-activity homework assignment

Having produced the data in class, the next step in this learning experience is another homework set that is meant to link the design and data-collection stages with analysis and conclusions. While there are many questions the instructor could assign at this point, the following problems focus on the application of the logistic model and converting the results into odds and odds ratios. Copies of the logistic regression output should be distributed so that students may perform their analysis. Students will be able to confirm their answers by finding some of them directly on the output. As an aid to instructors wishing to perform this in-class activity, [Appendix D](#) lists the answers to these problems using the output provided in [Section 5](#).

The Post-activity homework assignment. “Perform the following tasks:

- Using the initial model that considers all of the explanatory variables, find the coefficients for distance, orientation, and gender and the standard errors for each coefficient.
- Using these values, confirm the z-scores and p-values assuming a two-sided hypotheses test. Be sure to show all work!
- For distance, orientation, and gender: write the null and alternative hypotheses and describe the conclusion of a test of hypotheses using the p-values from part b). If significant, describe how the SIGN of the coefficient explains HOW the variable is apparently related to ShotMade.
- Using the final model and assuming a female student is throwing at a narrow target, find: The probability of making a shot from 5 feet, the probability of not making a shot from 5

- feet, and the odds of making a shot from 5 feet = $P(\text{making})/P(\text{not making})$.
- Using the final model and assuming a female student is throwing at a narrow target, find: The probability of making a shot from 6 feet, the probability of not making a shot from 6 feet, and the odds of making a shot from 6 feet = $P(\text{making})/P(\text{not making})$.
 - Defining the odds ratio as (the odds at 6 feet)/(the odds at 5 feet), find the odds ratio as you move from 5 feet to 6 feet from the trash can. Confirm your answer by finding it on the Minitab output.
 - Repeat steps d) through f) using the distances 10 feet and 11 feet. Does the odds ratio change?"

7. Summary

When a faculty member attempts to introduce an in-class activity in order to involve the students in the material being taught, there is often fine-tuning that needs to take place before the activity proves to be an organized and successful learning tool. Happily, we offer and recommend a logistic regression activity that has already gone through an evolution of improvement. In the past, we have tried more and different explanatory variables than described here and have learned to keep this part of the project simple. Formerly, the in-class activity was a one day affair. But we now know that the pre-activity homework assignment lays out a nice framework of logistic regression, builds anticipation for the event, and serves to maintain more organization on the day of the activity. And we now realize that the post-activity homework assignment really cements the issues of model interpretation, odds ratio, and significance testing of the variables. We didn't realize how much of the mathematical manipulation of the model was only vaguely understood until we instituted the post-activity assignment.

Textbooks on statistical methods and linear models are more frequently including sections or chapters on logistic regression. Not only do students learn a great deal about logistic regression from this in-class activity, many other aspects of the entire process of conducting statistical research can be experienced first hand. Many statistical topics covered earlier are reinforced including: choice and definition of variables, setting and testing hypotheses, linear regression, probabilities, odds, and odds ratios. As suggested in many publications, bringing statistics alive with this classroom activity has, in our experience, proven very successful in increasing student understanding and motivation to learn.

Appendix A: Answers to the Pre - Activity Homework

Possible explanatory variables include the following:

- Distance between the thrower and the trash can.
 - Numerical continuous.
 - $H_0: \beta_{\text{distance}} = 0$; $H_a: \beta_{\text{distance}} \neq 0$ (Or one could state: $H_a: \beta_{\text{distance}} < 0$).
 - Expect to reject the null hypothesis.
 - Measure, with care, the distance from the facing edge of the trash can to the thrower attack floor to mark the distances. Stand upright with normal arm extension.
- Orientation of a rectangular trash can.
 - Categorical (0 = wide/shallow, 1 = narrow/deep)
 - $H_0: \beta_{\text{orientation}} = 0$; $H_a: \beta_{\text{orientation}} \neq 0$ (or a student may choose the alternative as $<$ or $>$ 0).
 - Expect to reject the null hypothesis.
 - Keep the trash can against the wall; keep the middle of the facing side aligned with the trash can.
- Gender of the thrower.
 - Categorical (0 = Male, 1 = Female).
 - $H_0: \beta_{\text{gender}} = 0$; $H_a: \beta_{\text{gender}} \neq 0$.
 - Expect to fail to reject the null hypothesis
 - Carefully note 0 or 1 for each thrower.

Other variables may include: type of ball (tennis, racquet balls, or table tennis), size of ball, weight of ball, whether a student uses their favored hand (writing hand versus their other hand), size of trash can, level of basketball experience of the thrower, eye strength, self-reported coordination level, and whether a student uses an underhanded or overhanded toss.

Appendix B: A Sample Minitab Worksheet

This sample Minitab worksheet can be altered to apply to any class size and gender distribution. See the text of the article for more information. Instructors should anticipate the number of male and female students expected at the in-class activity. Each row in the worksheet represents a single toss made by a single student. Each student will complete all four tosses before the next student approaches. During class, a "1" should be entered into Column 2 when a shot is successful and a "0" for a missed shot. Note that the indicator variable for gender is set to "1" for females and "0" for males. The indicator variable for trash can orientation is set to "1" for the narrow/deep target and "0" for the wide/shallow alignment.

C1: ID #	C2: ShotMade (1=yes)	C3: Gender (0=F)	C4: Orient (0=Narrow)	C5: Dist (in feet)
1		0	1	5
1		0	0	7
1		0	1	9
1		0	0	11
2		1	1	5
2		1	0	7
2		1	1	9
2		1	0	11
3		0	1	6
3		0	0	8
3		0	1	10
3		0	0	12
4		1	1	6
4		1	0	8
4		1	1	10
4		1	0	12

5		0	0	5
5		0	1	7
5		0	0	9
5		0	1	11
6		1	0	5
6		1	1	7
6		1	0	9
6		1	1	11
7		0	0	6
7		0	1	8
7		0	0	10
7		0	1	12
8		1	0	6
8		1	1	8
8		1	0	10
8		1	1	12

Appendix C: Details on Conducting the Activity

One student will be in charge of changing the orientation of the trash can. The worksheet effectively instructs this student to simply alternate orientation with every throw from wide to narrow. Once half of the entire class of students have made their attempts, this student continues to alternate orientation but switches from narrow to wide. A second student should be assigned the responsibility of inviting two prospective tossers to approach the trash can (one of each gender) and tossing from 5, 7, 9, and 11 feet. Then the next two tossers stand 6, 8, 10, and 12 feet from the trash can. By following the order in the worksheet, the design is kept relatively balanced across the three explanatory variables while simultaneously keeping the activity simple and organized. Organization helps students to focus on watching and understanding the activity and, eventually, on making their own shots.

Even with a pre-set Minitab worksheet, students must still be assigned their own shot settings. Doing this in a relatively random way proved successful when conducting this experiment at a meeting of the Chesapeake Section of the ASA. There are many ways in which such randomization can be handled, but we suggest a plan that simply requires a set of four cards. Each card specifies the distance (in feet) of all four shots to be taken by a single student and also the orientation of the trash can. In groups of four male students, each selects one of these cards. The same is done with female students. This process continues until less than four of each gender are left. The remaining male students then draw one of the four cards; likewise for the female students. The four cards should be prepared as follows:

CARD 1: (5 ft, wide) then (7 ft, narrow), (9 ft, wide), (11 ft, narrow)

CARD 2: (5 ft, narrow) then (7 ft, wide), (9 ft, narrow), (11 ft, wide)

CARD 3: (6 ft, wide) then (8 ft, narrow), and (10 ft, wide), (12 ft, narrow)

CARD 4: (6 ft, narrow) then (8 ft, wide), (10 ft, narrow), (12 ft, wide).

Upon receiving their card, students can check the Minitab worksheet and determine their ID #. In this way, they easily learn the order to follow and the type of shots each of them are to take.

Note that this system of shots has all students throwing from shorter to longer distances. Admittedly, this introduces a learning effect and this concept ought to be discussed. With all students following this pattern, however, no unfair advantage is given to any students. Throwing from 11 or 12 feet without any warmup can be a very difficult proposition. So it may be worth having all students "get a feel" for the activity before they attempt the hardest shots. Also, randomizing on shot distance greatly complicates the activity.

Appendix D: Answers to the Post - Activity Homework

- a) Distance: coefficient = -0.7422, se coefficient = 0.2281
 Orient: coefficient = 2.3106, se coefficient = 0.9831
 Gender: coefficient = -0.1512, se coefficient = 0.8266
- b) Distance: coefficient /se = $-0.7422/0.2281 = -3.2538$ so p-value = $2(0.0006) = 0.0012 \sim 0.001 = t$
 (Note that all p-values in this appendix are found assuming a two-sided alternative hypothesis)
 Orientation: coefficient /se = $2.3106/0.9831 = 2.3503$ so p-value = $2(0.0094) = 0.0188 \sim 0.019 = t$
 Gender: coefficient /se = $-0.1512/0.8266 = -0.1829$ so p-value = $2(0.4274) = 0.8548 \sim 0.855 = t$
- c) Distance:
 $H_0: \beta_1 = 0$; $H_a: \beta_1 \neq 0$; since p-value = $0.001 < 0.05 = \alpha$, we reject the H_0 . Distance appears to be related to ShotMade in this model. The negative coefficient means the further the shot is taken, the less likely the shot is made.
 Orientation:
 $H_0: \beta_2 = 0$; $H_a: \beta_2 \neq 0$; since p-value = $0.019 < 0.05 = \alpha$, we reject the H_0 . Orientation appears to be related to ShotMade in this model. The positive coefficient means, as you move from a wide target to a narrow target, the likelihood of making the shot increases.
 Gender:
 $H_0: \beta_3 = 0$; $H_a: \beta_3 \neq 0$; since p-value = $0.8266 > 0.05 = \alpha$, we fail to reject the H_0 . Gender does not appear to be related to ShotMade.
- d) The gender variable is not in the final model; it is not a significant predictor. We, therefore, ignore the gender of our sample person:
 $P(\text{ShotMade} = 1) = \exp(5.857 - 0.7425(5) + 2.3096) / (1 + \exp(5.857 - 0.7425(5) + 2.3096)) = 0.9885$
 $P(\text{ShotMade} = 0) = 1 - 0.9885 = 0.0115$
 Odds = $0.9885/0.0115 = 85.9565$ (the odds you make this shot are 86 to 1!)
- e) $P(\text{ShotMade} = 1) = \exp(5.857 - 0.7425(6) + 2.3096) / (1 + \exp(5.857 - 0.7425(6) + 2.3096)) = 0.9761$
 $P(\text{ShotMade} = 0) = 1 - 0.9761 = 0.0239$
 Odds = $0.9761/0.0239 = 40.9174$ (the odds you make this shot are 41 to 1!)
- f) The odds ratio is $40.9174/85.9565 = 0.4760 \sim 0.48 =$ the value on the output (the odds drop to about half the size when you move back from 5 feet to 6 feet).
- g) Distance = 10 feet:
 $P(\text{ShotMade} = 1) = \exp(5.857 - 0.7425(10) + 2.3096) / (1 + \exp(5.857 - 0.7425(10) + 2.3096)) = 0.6773$
 $P(\text{ShotMade} = 0) = 1 - 0.6773 = 0.3227$
 Odds = $0.6773/0.3227 = 2.0992$ (the odds you make this shot are 2 to 1!)
- Distance = 11 feet:
 $P(\text{ShotMade} = 1) = \exp(5.857 - 0.7425(11) + 2.3096) / (1 + \exp(5.857 - 0.7425(11) + 2.3096)) = 0.4998$
 $P(\text{ShotMade} = 0) = 1 - 0.4998 = 0.5002$
 Odds = $0.4998/0.5002 = 0.9992$ (the odds you make this shot are 1 to 1! Even chance!)
 The odds ratio is $2.0992/0.9992 = 0.4759 \sim 0.48 =$ the value on the output (the odds drop to about half the size when you move back from 10 feet to 11 feet; NOTE: every time you back up by a foot, the odds are good of making the shot.).
-

Acknowledgements

We thank the editor and referees who provided comments that greatly improved the manuscript.

References

- Andrews, C. (2005) "The Ultimate Flow," *Journal of Statistics Education* [On line], 13(1).
www.amstat.org/publications/jse/v13n1/andrews.html
- Cobb, G.W. (1993) "Reconsidering Statistics Education: A National Science Foundation Conference," *Journal of Statistics Education* [On line], 1(1).
www.amstat.org/publications/jse/v1n1/cobb.html
- Duchesne, P. (2003) "Estimation of a Proportion with Survey Data," *Journal of Statistics Education* [On line], 11(3).
www.amstat.org/publications/jse/v11n3/duchesne.pdf
- Garfield, Joan (1993) "Teaching Statistics Using Small-Group Cooperative Learning," *Journal of Statistics Education* [On line], 1(1).
www.amstat.org/publications/jse/v1n1/garfield.html
- Gnanadesikan, M. and Schaeffer, R. L., Watkins, A. E., and Witmer, J. (1997) "An Activity-Based Statistics Course," *Journal of Statistics Education* [On line], 5(2).
www.amstat.org/publications/jse/v5n2/gnanadesikan.html
- Hogg, R.V. (1991) "Statistical Education: Improvements Are Badly Needed," *The American Statistician*, 45(4), 342-343.
- Johnson, H.D. and Dasgupta, N. (2005) "Traditional versus Non-traditional Teaching: Perspectives of Students in Introductory Statistics Classes," *Journal of Statistics Education* [On line], 13(2).
www.amstat.org/publications/jse/v13n2/johnson.html
- Kleinbaum, D. G., Kupper, L.L., Muller, K.E., and Nizam, A. (1998), *Applied Regression Analysis and Multivariable Methods*, 3rd Edition, Pacific Grove, CA: Duxbury Press.
- Kutner, M.H., Nachtsheim, C.J., and Neter, J.(2004), *Applied Linear Regression Models*, 4th Edition, Boston: McGraw-Hill/Irwin.
- Love, T.E. (1998) "A Project-Driven Second Course," *Journal of Statistics Education* [On line], 6(1).
www.amstat.org/publications/jse/v6n1/love.html
- Melton, K.I. (2004) "Statistical Thinking Activities: Some Simple Exercises With Powerful Lessons," *Journal of Statistics Education* [On line], 12(2).
www.amstat.org/publications/jse/v12n2/melton.html
- Moore, D. S. and McCabe, G. P. (2006), *Introduction to the Practice of Statistics*, 5th Edition, New York: W.H. Freeman and Company.
- NIH Policy and Guidelines on the Inclusion of Women and Minorities as Subjects in Clinical Research - Amended*, October, 2001. (2001)
grants.nih.gov/grants/funding/women_min/guidelines_amended_10_2001.htm
- Ott, R. L. and Longnecker. M. T. (2001) *An Introduction to Statistical Methods and Data Analysis*, 5th Edition, Pacific Grove, CA: Duxbury Press.
- Roback, P. J. (2003) "Teaching an Advanced Methods Course to a Mixed Audience," *Journal of Statistics Education* [On line], 11(2).
www.amstat.org/publications/jse/v11n3/robback.html
- Ryan, B. F., Joiner, B. L., and Cryer, J. D. (2004), *Minitab Handbook*, 5th Edition, Pacific Grove, CA: Duxbury Press.
- Ryan, T. P. (1996), *Modern Regression Methods*, New York: John Wiley and Sons.

Simonoff, J. S. (1997), "The 'Unusual Episode' and a Second Statistics Course," *Journal of Statistics Education* [On line], 5(1).
www.amstat.org/publications/jse/v5n1/simonoff.html

Simonoff, J. S.(1998), "Move Over, Roger Maris: Breaking Baseball's Most Famous Record," *Journal of Statistics Education* [On line], 6(3).
www.amstat.org/publications/jse/v6n3/datasets.simonoff.html

Souhrada, T, (2006,) "Numb3rs Activity: Logging Witnesses. Episode: "Alls Fair,"
www.cbs.com/primetime/numb3rs/ti/activities/Act2_LoggingWitnesses_AllsFair_final.pdf

Sowey, E.R. (2001), "Striking Demonstrations in Teaching Statistics," *Journal of Statistics Education*, 9(1).
www.amstat.org/publications/jse/v9n1/sowey.html

Willoughby, K. A. (2002), "Winning Games in Canadian Football: A Logistic Regression Analysis," *The College Mathematics Journal*, 33, 215-220.

Zacharopoulou, H. (2006), "Two Learning Activities for a Large Introductory Statistics Class," *Journal of Statistics Education* [On line], 14(1).
www.amstat.org/publications/jse/v14n1/zacharopoulou.html

Christopher H. Morrell
Mathematical Sciences Department
Loyola College in Maryland
Baltimore, MD 21210-2699
chm@loyola.edu

Richard E. Auer
Mathematical Sciences Department
Loyola College in Maryland 4501 North Charles Street
Baltimore, MD 21210-2699
rea@loyola.edu

[Volume 15 \(2007\)](#) | [Archive](#) | [Index](#) | [Data Archive](#) | [Information Service](#) | [Editorial Board](#) | [Guidelines for Authors](#) | [Guidelines for Data Contributors](#) | [Home Page](#) | [Contact JSE](#) | [ASA Publications](#)