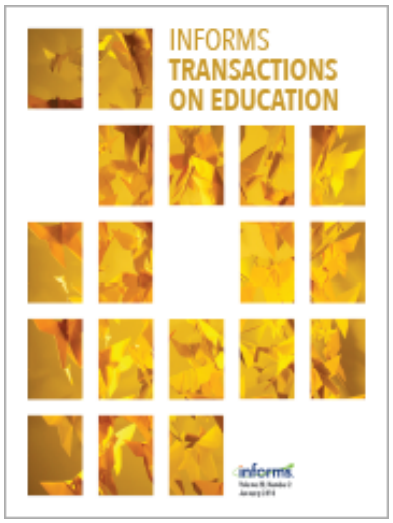
This article was downloaded by: [45.82.243.160] On: 14 July 2023, At: 23:50
Publisher: Institute for Operations Research and the Management Sciences (INFORMS)
INFORMS is located in Maryland, USA



INFORMS Transactions on Education

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Logistic Regression via Excel Spreadsheets: Mechanics, Model Selection, and Relative Predictor Importance

Michael Brusco

To cite this article:
Michael Brusco (2022) Logistic Regression via Excel Spreadsheets: Mechanics, Model Selection, and Relative Predictor Importance. INFORMS Transactions on Education 23(1):1-11. https://doi.org/10.1287/ited.2021.0263

Full terms and conditions of use: https://pubsonline.informs.org/Publications/Librarians-Portal/PubsOnLine-Terms-and-Conditions

This article may be used only for the purposes of research, teaching, and/or private study. Commercial use or systematic downloading (by robots or other automatic processes) is prohibited without explicit Publisher approval, unless otherwise noted. For more information, contact permissions@informs.org.

The Publisher does not warrant or guarantee the article's accuracy, completeness, merchantability, fitness for a particular purpose, or non-infringement. Descriptions of, or references to, products or publications, or inclusion of an advertisement in this article, neither constitutes nor implies a guarantee, endorsement, or support of claims made of that product, publication, or service.

Copyright © 2021 The Author(s)

Please scroll down for article—it is on subsequent pages

With 12,500 members from nearly 90 countries, INFORMS is the largest international association of operations research (O.R.) and analytics professionals and students. INFORMS provides unique networking and learning opportunities for individual professionals, and organizations of all types and sizes, to better understand and use O.R. and analytics tools and methods to transform strategic visions and achieve better outcomes.
For more information on INFORMS, its publications, membership, or meetings visit http://www.informs.org

https://pubsonline.informs.org/journal/ited

INFORMS TRANSACTIONS ON EDUCATION
Vol. 23, No. 1, September 2022, pp. 1–11
ISSN 1532-0545 (online)

Logistic Regression via Excel Spreadsheets: Mechanics, Model Selection, and Relative Predictor Importance

Michael Brusco[a]

[a] Department of Business Analytics, Information Systems, and Supply Chain, Florida State University, Tallahassee, Florida 33206
Contact: mbrusco@fsu.edu, https://orcid.org/0000-0002-1465-6233 (MB)

Received: February 6, 2021
Revised: August 7, 2021; October 7, 2021
Accepted: October 8, 2021
Published Online in Articles in Advance: December 9, 2021

https://doi.org/10.1287/ited.2021.0263

Copyright: © 2021 The Author(s)

Abstract. Logistic regression is one of the most fundamental tools in predictive analytics. Graduate business analytics students are often familiarized with implementation of logistic regression using Python, R, SPSS, or other software packages. However, an understanding of the underlying maximum likelihood model and the mechanics of estimation are often lacking. This paper describes two Excel workbooks that can be used to enhance conceptual understanding of logistic regression in several respects: (i) by providing a clear formulation and solution of the maximum likelihood estimation problem; (ii) by showing the process for testing the significance of logistic regression coefficients; (iii) by demonstrating different methods for model selection to avoid overfitting, specifically, all possible subsets ordinary least squares regression and l_1-regularized logistic regression (lasso); and (iv) by illustrating the measurement of relative predictor importance using all possible subsets.

Open Access Statement: This work is licensed under a Creative Commons Attribution 4.0 International License. You are free to copy, distribute, transmit and adapt this work, but you must attribute this work as "*INFORMS Transactions on Education*. Copyright © 2021 The Author(s). https://doi.org/10.1287/ited.2021.0263, used under a Creative Commons Attribution License: https://creativecommons.org/licenses/by/4.0/."

Supplemental Material: The e-companion is available at https://doi.org/10.1287/ited.2021.0263.

Keywords: logistic regression • OLS regression • all possible subsets • lasso • spreadsheets

Downloaded from informs.org by [45.82.243.160] on 14 July 2023, at 23:50 . For personal use only, all rights reserved.

1. Introduction

Logistic regression is one of the most popular methodological tools in predictive analytics. Although logistic regression, broadly defined, encompasses a variety of models that allow for multiple (possibly ordinal) categories of a dependent variable, attention here is restricted to the case of a binary dependent variable. Thus, the term *logistic regression* when used in this paper refers to what is sometimes called "binary logistic regression." Applications of the method abound in business-related contexts and include prediction of bank failure (Zaghdoudi 2013), loan default (Agbemava et al. 2016), item purchase (Bejaei et al. 2015), complaint behavior (Salamah and Ramayanti 2018), word-of-mouth communication regarding products (Alboqami et al. 2015), and conversion of point differentials or point spreads into win–loss probabilities (Kvam and Sokol 2004, Huggins et al. 2020). Like ordinary least squares (OLS) regression in the case of a continuous dependent variable, logistic regression can be used for purely *predictive* purposes, for purely *explanatory* purposes to identify the key independent variables that drive the dependent variable, or for both predictive and explanatory purposes.

In most business analytics programs, implementation of logistic regression is accomplished using statistical software programs on platforms such as Python, R, SAS, or SPSS. Although these programs are efficient and effective for large-scale applications, the underlying maximum likelihood model for logistic regression and its corresponding estimation are often hidden from students. This observation, which is well articulated by Pinder (2013) to motivate the use of Excel for nonlinear regression applications, is unfortunate for several reasons. First, it means students might be blindly applying a method they do not sufficiently comprehend. A better understanding of maximum likelihood estimation in the context of logistic regression is crucial and could also be beneficial for students who encounter other maximum likelihood problems in their statistical journey, such as those found in confirmatory factor analysis, structural equation modeling, and mixture-model clustering. Second, the principles of model selection to avoid overfitting are also rather murky. Third, systematic approaches to measuring the relative importance of the independent variables with respect to their explanation of the dependent variable are easy to overlook.

Excel spreadsheets are an excellent platform for addressing the three problem areas and can improve student comprehension of logistic regression. To address

Downloaded from informs.org by [45.82.243.160] on 14 July 2023, at 23:50 . For personal use only, all rights reserved.

the first problem, it is not difficult to develop a spreadsheet that incorporates the maximum likelihood formulas necessary for estimating the coefficients of a logistic regression model (see, for example, Carlberg 2013, Pinder 2013, Schield 2017, Ragsdale 2018). In most instances, the generalized reduced gradient (GRG) nonlinear engine of the Excel solver can rapidly find the coefficients that maximize likelihood. YouTube videos (e.g., Ritter 2013) demonstrating this process are available. Testing the significance of the logistic regression coefficients is a little more complicated but can also be accomplished in Excel. The second problem, model selection, can be addressed different ways in a spreadsheet analysis. Examples include l_1-regularized logistic regression (lasso, Tibshirani 1996) and all possible subsets regression (Miller 2002) via a visual basic application (VBA) macro. The all possible subsets approach is also viable for tackling the third problem, measuring relative predictor importance.

The development of Excel spreadsheets to address the aforementioned problems assumes that students have already been exposed to some of the basic principles of logistic regression, such as (i) the difference between relative risk and the odds ratio, specifically noting that the former is a ratio of probabilities rather than odds; (ii) a plot of the logistic function across different values of probability for the dependent variable; (iii) the difference between categorical and continuous independent variables in logistic regression; and (iv) some "toy" examples to introduce basic concepts concerning likelihood and coefficient interpretation. Some good teaching practices for introducing students to logistic regression are available in the literature (Campbell 1998, Morrell and Auer 2007).

The particular Excel workbooks that I have developed are specially designed for teaching logistic regression in a graduate-level business analytics course. They have been prepared to enhance student understanding of the mechanics of logistic regression and, more importantly, to demonstrate the critical nature of good model-building for both predictive and explanatory purposes. The particular context is a multivariable application corresponding to the prediction of the spreading of positive word-of-mouth in terms of recommendation of a restaurant. The importance of fostering "multivariable thinking and modeling" is a key aspect of the Guidelines for Assessment and Instruction in Statistics Education (GAISE): College Report 2016 (GAISE 2016, p 15). Moreover, it leads naturally into issues regarding the critical aspects of model selection (including the potential for *overfitting*) and the measurement of relative predictor importance. As such, the emphasis of the Excel workbooks on multivariable modeling are in accordance with the higher levels of cognitive skills (e.g., *application*, *analysis*, *synthesis*, and *evaluation*) in Bloom's taxonomy (Bloom et al. 1956, Anderson and Krathwohl 2001). The key learning objectives associated with the Excel workbooks are

1. Students should be able to understand and implement maximum likelihood estimation of logistic regression using the Excel solver.
2. Students should be able to understand and implement significance testing of the logistic regression coefficients using Excel.
3. Students should understand the l_1-regularized regression (lasso, Tibshirani 1996) approach to model selection and be able to implement the method using the Excel solver.
4. Students should be able to understand and implement (via an Excel VBA macro provided by the instructor) all possible subsets OLS regression (Miller 2002) as a model-selection heuristic for logistic regression and compare and contrast this approach to l_1-regularized logistic regression.
5. Students should be able to understand and implement (via an Excel VBA macro provided by the instructor) all possible subsets OLS regression to provide a heuristic assessment of the relative importance of the predictors.

Section 2 of this paper provides a brief overview of maximum likelihood estimation of logistic regression coefficients and their significance tests as well as issues associated with model selection and the measurement of predictor importance. Section 3 describes the Excel workbooks for logistic regression with model selection and predictor importance (`LR_MSPI.xlsm`) and logistic regression significance testing (`LR_SigTest.xlsm`), using the word-of-mouth data as an example. The paper concludes in Section 4 with a brief summary.

2. Models and Methods

2.1. Logistic Regression

Thorough treatments of logistic regression are provided by several authors, including Hosmer et al. (2013) and Menard (2010). Here, the goal is to provide a brief description and a linkage to OLS regression. I define n as the number of observations in the training (or estimation) sample, v as the number of predictors, y_i as the binary dependent variable measurement for observation i, and x_{ij} as the independent variable measurement for observation i on predictor j. Moreover, it is assumed that predictor $j = 0$ is the intercept term and column 0 of the $n \times (v+1)$ matrix, $\mathbf{X} = [x_{ij}]$, is just a column of ones (i.e., $x_{i0} = 1$ for all $1 \leq i \leq n$). Denoting β_j as the regression coefficient for variable j (for all $0 \leq j \leq v$), a standard OLS regression model is

$$y_i = \sum_{j=0}^{v} \beta_j x_{ij}. \tag{1}$$

Two problems with the model are typically highlighted: (i) the potential for predicted values outside of the

Downloaded from informs.org by [45.82.243.160] on 14 July 2023, at 23:50 . For personal use only, all rights reserved.

(0, 1) range for the dependent variable and (ii) violations of assumptions that facilitate testing of the regression coefficients, most notably the assumption of constant error variance. It is possible to use OLS regression for two-group discriminant analysis (see Ragsdale and Stam 1992); however, many researchers prefer logistic regression, which models the logarithm of the odds (i.e., logit) of the observations. Dropping the i subscripts momentarily to avoid notational clutter, an observation is modeled by replacing y on the left with the logarithm of its odds:

$$\log \frac{P(\mathbf{x})}{1 - P(\mathbf{x})} = \sum_{j=0}^{v} \beta_j x_j, \tag{2}$$

where $P(\mathbf{x})$ should be interpreted as the probability of $y = 1$ given the vector $\mathbf{x}$ of predictor variable measurements. Accordingly, the left side of (2) is the logarithm of the odds ratio (i.e., the probability of $y = 1$, $P(\mathbf{x})$, divided by the probability of $y = 0$, which is $1 - P(\mathbf{x})$). Taking the exponent of both the left and right sides of (2) yields

$$\frac{P(\mathbf{x})}{1 - P(\mathbf{x})} = \exp\left(\sum_{j=0}^{v} \beta_j x_j\right) = \prod_{j=0}^{v} \exp(\beta_j x_j). \tag{3}$$

The far-right expression in (3) is included only to show the multiplicative (rather than additive) nature of the independent variables. In other words, for a one-unit increase in x_j, we would expect a change in the odds of the binary variable assuming a value of one as being $\exp(\beta_j)$. Multiplying through by $1 - P(\mathbf{x})$ yields

$$P(\mathbf{x}) = \exp\left(\sum_{j=0}^{v} \beta_j x_j\right) - P(x)\exp\left(\sum_{j=0}^{v} \beta_j x_j\right). \tag{4}$$

Collecting the $P(\mathbf{x})$ terms on the left,

$$P(\mathbf{x}) + P(\mathbf{x})\exp\left(\sum_{j=0}^{v} \beta_j x_j\right) = \exp\left(\sum_{j=0}^{v} \beta_j x_j\right). \tag{5}$$

Factor out $P(\mathbf{x})$ and divide to obtain

$$P(\mathbf{x}) = \frac{\exp\left(\sum_{j=0}^{v} \beta_j x_j\right)}{1 + \exp\left(\sum_{j=0}^{v} \beta_j x_j\right)}. \tag{6}$$

Equation (6) is expressed in a slightly more compact way in the Excel spreadsheets:

$$\begin{aligned} P(\mathbf{x}) &= \frac{\exp\left(\sum_{j=0}^{v} \beta_j x_j\right)}{1 + \exp\left(\sum_{j=0}^{v} \beta_j x_j\right)}\left(\frac{\exp -\left(\sum_{j=0}^{v} \beta_j x_j\right)}{\exp -\left(\sum_{j=0}^{v} \beta_j x_j\right)}\right) \\ &= \frac{1}{1 + \exp -\left(\sum_{j=0}^{v} \beta_j x_j\right)}. \end{aligned} \tag{7}$$

The goal in logistic regression is to estimate the β's so as to maximize the likelihood of the data across all observations. Therefore, because likelihood is across all observations, it is necessary to bring back the i subscripts to write the likelihood function as follows:

$$L(\mathbf{X} \mid \boldsymbol{\beta}) = \prod_{i:y_i=1} P(\mathbf{x}_i) \prod_{i:y_i=0} (1 - P(\mathbf{x}_i)). \tag{8}$$

It is much easier to work with the logarithm of the likelihood function, and the estimates that maximize one also maximize the other:

$$\begin{aligned} LL(\mathbf{X} \mid \boldsymbol{\beta}) &= \log(L(\mathbf{X} \mid \boldsymbol{\beta})) \\ &= \sum_{i:y_i=1} \log P(\mathbf{x}_i) + \sum_{i:y_i=0} \log(1 - P(\mathbf{x}_i)). \end{aligned} \tag{9}$$

Estimating the parameters that maximize (9) requires partial differentiation. Unfortunately, unlike standard OLS regression, there is no closed-form solution. Fortunately, the GRG nonlinear engine of the Excel solver is typically effective for completing the estimation process as is demonstrated in the Excel workbook `LR_MSPI.xlsm`. More specifically, the estimates of the coefficients obtained by the GRG nonlinear engine are not overly sensitive to the initial starting values provided for those estimates.

Upon successful estimation, cases for the training sample as well as a holdout (or validation) sample, are often made based on Equation (7) by assigning cases with $P(\mathbf{x}) > 0.5$ to the group associated with $y = 1$ and to the group associated with $y = 0$ otherwise. For predictive applications of logistic regression in which the number of cases for the two groups is markedly imbalanced and/or the costs of misclassification are asymmetric (i.e., it costs more to misclassify a group 1 case as group 2 than it does to misclassify a group 2 case as group 1 or vice versa), a probability other than 0.5 might be used for assigning cases (see, for example, Brownlee 2020).

Testing the significance of the estimated logistic regression coefficients, $\hat{\beta}_j$, requires estimates of their standard errors, $se(\hat{\beta}_j)$. The $se(\hat{\beta}_j)$ values are computed as the square roots of the main diagonal of the covariance matrix, $(\mathbf{X}'\mathbf{V}\mathbf{X})^{-1}$, where $\mathbf{V}$ is an $n \times n$ diagonal matrix with elements $P(\mathbf{x}_i)(1\text{-}P\mathbf{x}_i)$ along the main diagonal. The Wald coefficients, $w_j = (\hat{\beta}_j / se(\hat{\beta}_j))^2$, and z-values, $z_j = (\hat{\beta}_j / se(\hat{\beta}_j))$, can then be computed and their significance tested using the standard normal distribution.

2.2. Model Selection

As in the case of OLS regression, there are different possible approaches to model selection in logistic regression. Here, two approaches are discussed: (1) the lasso, and (2) all possible subsets.

Downloaded from informs.org by [45.82.243.160] on 14 July 2023, at 23:50 . For personal use only, all rights reserved.

2.2.1. The Lasso Method. One approach is to constrain or penalize the coefficients of the model in some fashion. This is the principle used in ridge regression (Hoerl and Kennard 1970) and l_1-regularized (lasso, Tibshirani 1996) methods. Here, attention is restricted to the latter method because of its immense popularity. Using the constrained approach, the goal is to maximize Equation (9) or, equivalently, minimize the *deviance* (-2*LL*) measure subject to the restriction that the sum of the absolute values of the model coefficients does not exceed some threshold, λ. More formally, this constraint is specified as

$$\sum_{j=1}^{v} |\hat{\beta}_j| \leq \lambda. \tag{10}$$

2.2.2. The All Possible Subsets Method. A second approach is to use all possible subsets regression (Miller 2002). With this approach, all 2^v regression models are estimated, and the model obtaining the best fit is stored for each subset size. For OLS regression, the measure of fit is typically the residual sum of squares, whereas for logistic regression the deviance is the common measure. For both models, the measure of fit generally improves (decreases) as the subset size increases. Therefore, it is necessary to employ criteria that can facilitate the selection of the appropriate subset size. As in the case of OLS regression, two of the most popular criteria for logistic regression are Akaike's information criterion (AIC $= -2LL + 2(v+1)$): Akaike 1973) and the Bayesian information criterion (BIC $= -2LL + (v+1)\log(n)$), Schwartz 1978). The best subset size and corresponding model are selected based on the minimum value for the AIC or BIC criterion.

2.3. Relative Predictor Importance

The measurement of the relative importance of predictors is distinct from model selection (Azen and Budescu 2003). Predictor A might be selected for inclusion in a regression model and predictor B excluded, but that does not mean that predictor A is better (or relatively more important) than predictor B. Thus, in regression applications in which explanation is at least as important as prediction, it is beneficial to employ methods that can produce a ranking of predictors with respect to their relative importance. As with model selection, establishing the relative importance of predictors can be accomplished using all possible subsets regression. In particular, the measure of general dominance, which is based on R^2 shares, has proved useful for OLS regression (Lindeman et al. 1980, Budescu 1993). The R^2 share for a predictor corresponds to its average contribution to explained variation across all possible orderings for sequential inclusion of the predictors in the model (see Grömping 2015 for details). An especially desirable property of this measure is that the sum of the R^2 shares across all predictors is equal to R^2 for the full regression model using all predictors. Extensions of this approach to logistic regression are developed by Azen and Traxel (2009).

Given that all possible subsets can be used for both model selection and relative predictor importance, it provides a useful framework for both OLS and logistic regression. However, it is important to recognize that the computational demand for logistic regression is appreciably greater than it is for OLS regression, and accordingly, the feasibility of an Excel spreadsheet implementation of an all possible subsets analysis is less likely to be viable for logistic regression. Therefore, if we can use an Excel spreadsheet application for all possible subsets OLS regression as a surrogate for all possible subsets logistic regression, then that would have practical value. This is explored in the next section.

3. Excel Workbooks

As noted in the introduction, several authors propose the use of Excel spreadsheets for teaching the basic principles of logistic regression (Campbell 1998, Carlberg 2013, Ritter 2013, Schield 2017, Ragsdale 2018). However, the main focus of the `LR_MSPI` workbook presented herein is on a multivariate application that involves both model selection and predictor importance. Because it requires some large-scale matrix operations that can slow down the Excel solver, significance testing is performed using a separate workbook, `LR_SigTest`. This section begins with a description of the data for the example and continues with data analysis pertaining to maximum likelihood estimation, significance testing of the coefficients, model selection, and the measurement of predictor importance.

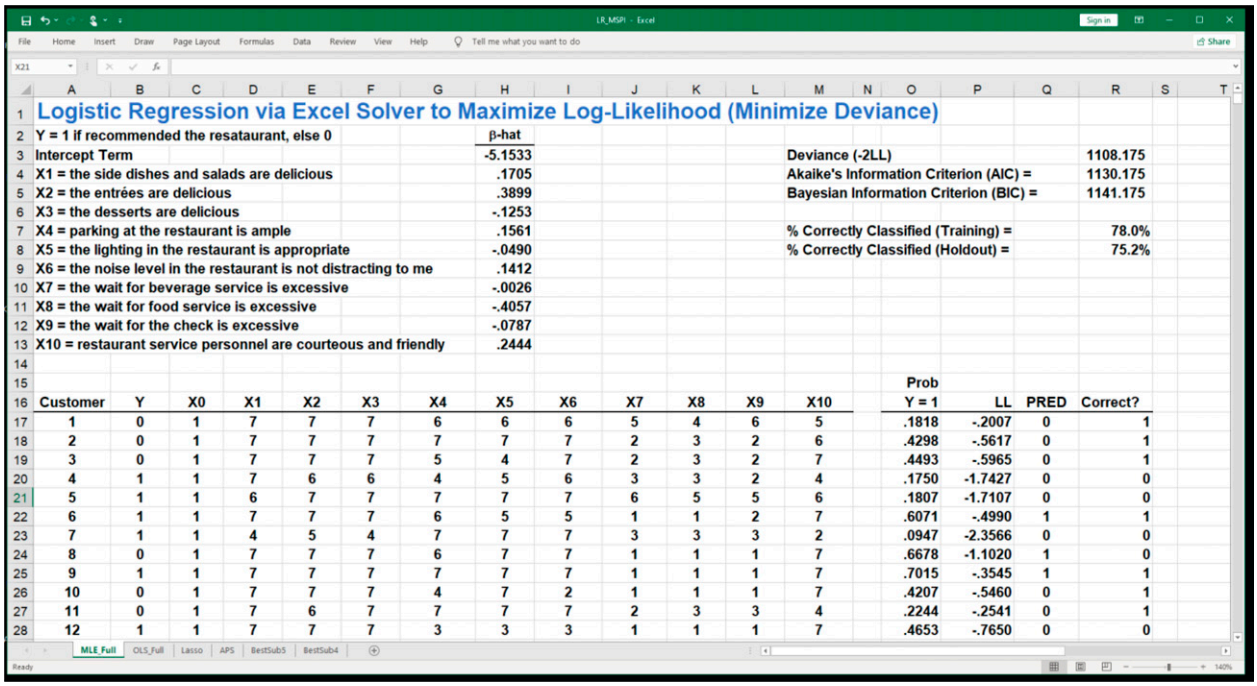
3.1. Example Context

The context of the example is recommendation behavior in a services marketing context. The dependent variable assumes a value of $y_i = 1$ if customer i recommends the restaurant (i.e., engages in positive word-of-mouth behavior) and $y_i = 0$ otherwise. There are 10 predictor variables, each measured on a seven-point Likert scale (1 = strongly disagree to 7 = strongly agree). The predictor variable statements are displayed in cells A4:A13 of the `MLE_Full` worksheet in the `LR_MSPI.xlsm` workbook. A screenshot of a portion of this worksheet is displayed in Figure 1. The training sample consists of $n = 1{,}000$ observations for the dependent and predictor variables, which are contained in cells B17:M1016 of the worksheet. A correlation matrix for the predictors is provided in Table 1.

In this type of application context, the goals of the logistic regression analysis might be both predictive

Downloaded from informs.org by [45.82.243.160] on 14 July 2023, at 23:50 . For personal use only, all rights reserved.

Figure 1. Excel Worksheet "MLE_Full" Used to Implement Maximum Likelihood Estimation of the Parameters of the Logistic Regression Model Using All 10 Predictors



Logistic Regression via Excel Solver to Maximize Log-Likelihood (Minimize Deviance)

		β-hat
2	Y = 1 if recommended the resataurant, else 0	
3	Intercept Term	-5.1533
4	X1 = the side dishes and salads are delicious	.1705
5	X2 = the entrées are delicious	.3899
6	X3 = the desserts are delicious	-.1253
7	X4 = parking at the restaurant is ample	.1561
8	X5 = the lighting in the restaurant is appropriate	-.0490
9	X6 = the noise level in the restaurant is not distracting to me	.1412
10	X7 = the wait for beverage service is excessive	-.0026
11	X8 = the wait for food service is excessive	-.4057
12	X9 = the wait for the check is excessive	-.0787
13	X10 = restaurant service personnel are courteous and friendly	.2444

Deviance (-2LL)	1108.175
Akaike's Information Criterion (AIC) =	1130.175
Bayesian Information Criterion (BIC) =	1141.175
% Correctly Classified (Training) =	78.0%
% Correctly Classified (Holdout) =	75.2%

Customer	Y	X0	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	Prob Y = 1	LL	PRED	Correct?
1	0	1	7	7	7	6	6	6	5	4	6	5	.1818	-.2007	0	1
2	0	1	7	7	7	7	7	7	2	3	2	6	.4298	-.5617	0	1
3	0	1	7	7	7	5	4	7	2	3	2	7	.4493	-.5965	0	1
4	1	1	7	6	6	4	5	6	3	3	2	4	.1750	-1.7427	0	0
5	1	1	6	7	7	7	7	7	6	5	5	6	.1807	-1.7107	0	0
6	1	1	7	7	7	6	5	5	1	1	2	7	.6071	-.4990	1	1
7	1	1	4	5	4	7	7	7	3	3	3	2	.0947	-2.3566	0	0
8	0	1	7	7	7	6	7	7	1	1	1	7	.6678	-1.1020	1	0
9	1	1	7	7	7	7	7	7	1	1	1	7	.7015	-.3545	1	1
10	0	1	7	7	7	4	7	2	1	1	1	7	.4207	-.5460	0	1
11	0	1	7	6	7	7	7	7	2	3	3	4	.2244	-.2541	0	1
12	1	1	7	7	7	3	3	3	1	1	1	7	.4653	-.7650	0	0

and explanatory. Building a good predictive engine requires model selection, and both the lasso and all possible subsets are evaluated for this purpose. There are 500 additional observations in cells B1017:M1516, which are used as a holdout sample to evaluate predictive performance. As noted by Jaggia et al. (2020), the importance of using a holdout sample should be stressed during the instruction process. To evaluate the predictors with respect to their explanatory importance, all possible subsets OLS regression is applied.

3.2. Logistic Regression (Full Model)

The maximum likelihood estimation of the coefficients for the full (i.e., using all 10 predictors) logistic regression model is accomplished in the `MLE_Full` worksheet of the `LR_MSPI.xlsm` workbook. The description of the worksheet is divided into four components: (i) structure of the worksheet for maximum likelihood estimation, (ii) making predictions with the model, (iii) significance testing, and (iv) substantive interpretation of the results.

3.2.1. Structure of the `MLE_Full` Worksheet. The coefficients are in cells H3:H13. These cells are used in conjunction with the predictor variable measurements to populate cells O17:O1516 with the $P(\mathbf{x})$ values associated with Equation (7). The log-likelihoods for each observation are computed in cells P17:P1516 using Equation (9), which requires the $P(\mathbf{x})$ values in O17:O1516 and the dependent variable measures in

Table 1. Correlation Matrix for the Word-of-Mouth Data (Training Sample)

	X1	*X2*	*X3*	*X4*	*X5*	*X6*	*X7*	*X8*	*X9*	*X10*
X1	1.0000									
X2	0.8460	1.0000								
X3	0.8498	0.8412	1.0000							
X4	0.1571	0.1243	0.1477	1.0000						
X5	0.1508	0.0887	0.1207	0.6358	1.0000					
X6	0.1162	0.0789	0.1002	0.6168	0.6123	1.0000				
X7	−0.3368	−0.3183	−0.3485	−0.1034	−0.1254	−0.1441	1.0000			
X8	−0.3516	−0.3157	−0.3415	−0.1360	−0.1343	−0.1485	0.8253	1.0000		
X9	−0.3406	−0.3154	−0.3260	−0.1403	−0.1506	−0.1747	0.8360	0.8214	1.0000	
X10	0.5068	0.5223	0.4997	0.1136	0.1125	0.1011	−0.5215	−0.5126	−0.5165	1.0000

Downloaded from informs.org by [45.82.243.160] on 14 July 2023, at 23:50 . For personal use only, all rights reserved.

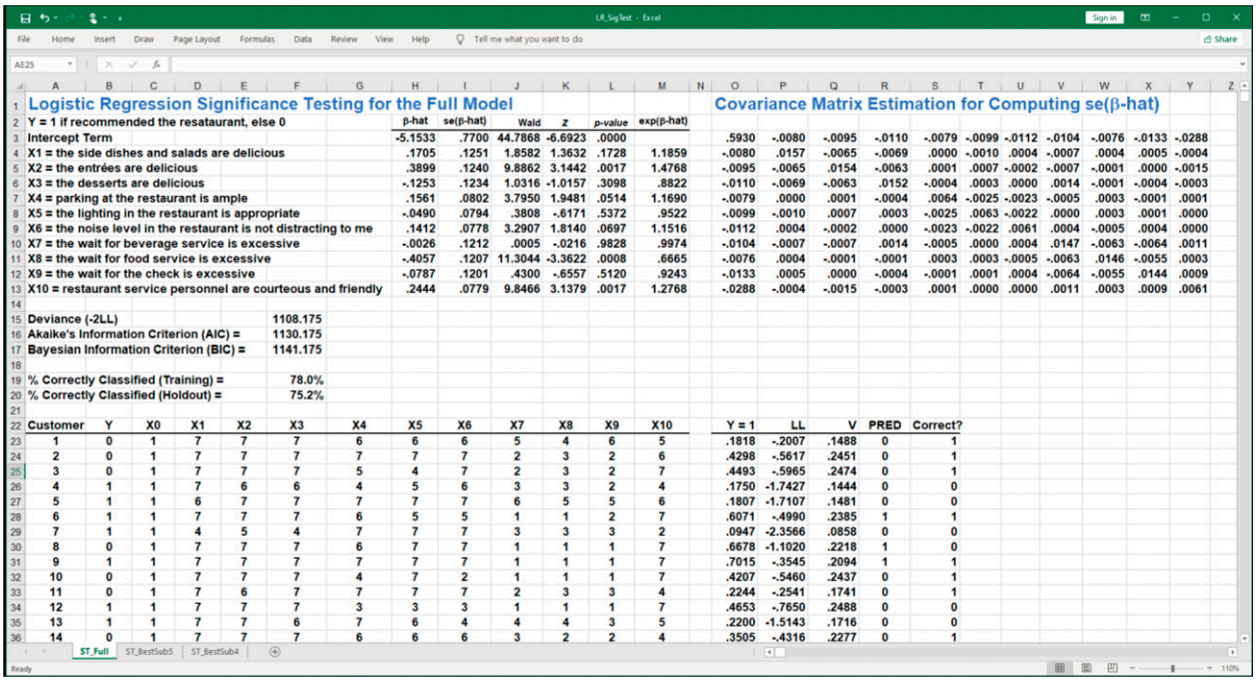
cells B17:B1516. The deviance of the model is computed in cell R3, using the training sample only as -2*SUM(P17:P1016). Cell R3 is the objective function that the Excel solver seeks to minimize by changing the values in H3:H13. The GRG nonlinear engine of the Excel solver is recommended. Initial values for H3:H13 can be established in various ways although there is no guarantee of a globally optimal solution. To obtain the results shown in Figure 1, I initialized all of the values to zero and ran the algorithm. As a check, I obtained the same results using SPSS (IBM Corp. 2019).

3.2.2. Making Predictions. The deviance measure in cell R3 is used to compute the AIC and BIC in cells R4 and R5, respectively. The predictions for the dependent variable are obtained in cells Q17:Q1516 by applying *if* statements to the $P(\mathbf{x})$ values in O17:O1516, and cells R17:R1516 assume values of one if the predictions match the observed y values in column B or zero if the predictions are wrong. Cells R7 and R8 contain the percentage of correct predictions for the training (78.0%) and holdout (75.2%) samples, respectively.

3.2.3. Significance Testing. Although significance testing of the coefficients can be accomplished seamlessly using an Excel add-in, such as XLSTAT (Addinsoft 2021), I want students to understand the mechanics of this process, which is somewhat more involved than obtaining the coefficients themselves. For this reason, it is accomplished in a separate workbook, `LR_SigTest.xlsm`. As shown in Figure 2, the `ST_Full` worksheet of this workbook is similar in appearance to the `MLE_Full` worksheet, but it has additional information in cells I3:M13 regarding the significance tests. The values of the coefficients from cells H3:H13 of the `MLE_Full` worksheet should be copied and pasted to cells H3:H13 of the `ST_Full` worksheet. The $se(\hat{\beta}_j)$ values are computed in cells I3:I13. Their computation requires the diagonal elements of $\mathbf{V}$ in cells Q17:Q1016, which are used to obtain the covariance matrix, $(\mathbf{X}'\mathbf{V}\mathbf{X})^{-1}$, in cells O3:Y13. The matrix computation is accomplished using the MINVERSE, MMULT, and TRANSPOSE functions as well as an *if* statement that produces the diagonal matrix. To assure that the $\mathbf{V}$ matrix is diagonal, the elements in Q17:Q1016 are perturbed via a small random error to guarantee that they are unique. Once the $se(\hat{\beta}_j)$ values are obtained, the w_j and z_j values can be computed in cell ranges J3:J13 and K3:K13, respectively. The p-values for the coefficients are computed using the NORMDIST function in cells L3:L13. Only three predictors (x_2, x_8, and x_{10}) are significant at $\alpha = 0.05$ (x_4 and x_6 are borderline). Finally, the exponents of each regression coefficient are provided in cells M3:M13 for easy interpretation.

3.2.4. Substantive Interpretation. An examination of the solution reveals some interesting findings. For example, for a one-unit increase in the response to the

Figure 2. Excel Worksheet "SigTest_Full" from the `LR_SigTest.xlsm` Workbook, Which Is Used to Perform Significance Testing of the Logistic Regression Coefficients for the 10-Predictor Model



Logistic Regression Significance Testing for the Full Model

Y = 1 if recommended the resataurant, else 0	β-hat	se(β-hat)	Wald	z	p-value	exp(β-hat)
Intercept Term	-5.1533	.7700	44.7868	-6.6923	.0000	
X1 = the side dishes and salads are delicious	.1705	.1251	1.8582	1.3632	.1728	1.1859
X2 = the entrées are delicious	.3899	.1240	9.8862	3.1442	.0017	1.4768
X3 = the desserts are delicious	-.1253	.1234	1.0316	-1.0157	.3098	.8822
X4 = parking at the restaurant is ample	.1561	.0802	3.7950	1.9481	.0514	1.1690
X5 = the lighting in the restaurant is appropriate	-.0490	.0794	.3808	-.6171	.5372	.9522
X6 = the noise level in the restaurant is not distracting to me	.1412	.0778	3.2907	1.8140	.0697	1.1516
X7 = the wait for beverage service is excessive	-.0026	.1212	.0005	-.0216	.9828	.9974
X8 = the wait for food service is excessive	-.4057	.1207	11.3044	-3.3622	.0008	.6665
X9 = the wait for the check is excessive	-.0787	.1201	.4300	-.6557	.5120	.9243
X10 = restaurant service personnel are courteous and friendly	.2444	.0779	9.8466	3.1379	.0017	1.2768

Deviance (-2LL)	1108.175
Akaike's Information Criterion (AIC) =	1130.175
Bayesian Information Criterion (BIC) =	1141.175
% Correctly Classified (Training) =	78.0%
% Correctly Classified (Holdout) =	75.2%

Covariance Matrix Estimation for Computing se(β-hat)

.5930	-.0080	-.0095	-.0110	-.0079	-.0099	-.0112	-.0104	-.0076	-.0133	-.0288
-.0080	.0157	-.0065	-.0069	.0000	-.0010	.0004	-.0007	.0004	.0005	-.0004
-.0095	-.0065	.0154	-.0063	.0001	.0007	-.0002	-.0007	-.0001	.0000	-.0015
-.0110	-.0069	-.0063	.0152	-.0004	.0003	.0000	.0014	-.0001	-.0004	-.0003
-.0079	.0000	.0001	-.0004	.0064	-.0025	-.0023	-.0005	.0003	-.0001	.0001
-.0099	-.0010	.0007	.0003	-.0025	.0063	-.0022	.0000	.0003	.0001	.0000
-.0112	.0004	-.0002	.0000	-.0023	-.0022	.0061	.0004	-.0005	.0004	.0000
-.0104	-.0007	-.0007	.0014	-.0005	.0000	.0004	.0147	-.0063	-.0064	.0011
-.0076	.0004	-.0001	-.0001	.0003	.0003	-.0005	-.0063	.0146	-.0055	.0003
-.0133	.0005	.0000	-.0004	-.0001	.0001	.0004	-.0064	-.0055	.0144	.0009
-.0288	-.0004	-.0015	-.0003	.0001	.0000	.0000	.0011	.0003	.0009	.0061

Customer	Y	X0	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	Y = 1	LL	V	PRED	Correct?
1	0	1	7	7	7	6	6	6	5	4	6	5	.1818	-.2007	.1488	0	1
2	0	1	7	7	7	7	7	7	2	3	2	6	.4298	-.5617	.2451	0	1
3	0	1	7	7	7	5	4	7	2	3	2	7	.4493	-.5965	.2474	0	1
4	1	1	7	6	6	4	5	6	3	3	2	4	.1750	-1.7427	.1444	0	0
5	1	1	6	7	7	7	7	7	6	5	5	6	.1807	-1.7107	.1481	0	0
6	1	1	7	7	7	6	5	5	1	1	2	7	.6071	-.4990	.2385	1	1
7	1	1	4	5	4	7	7	7	3	3	3	2	.0947	-2.3566	.0858	0	0
8	0	1	7	7	7	6	7	7	1	1	1	7	.6678	-1.1020	.2218	1	0
9	1	1	7	7	7	7	7	7	1	1	1	7	.7015	-.3545	.2094	1	1
10	0	1	7	7	7	4	7	2	1	1	1	7	.4207	-.5460	.2437	0	1
11	0	1	7	6	7	7	7	7	2	3	3	4	.2244	-.2541	.1741	0	1
12	1	1	7	7	7	3	3	3	1	1	1	7	.4653	-.7650	.2488	0	0
13	1	1	7	7	6	7	6	4	4	4	3	5	.2200	-1.5143	.1716	0	0
14	0	1	7	7	7	6	6	6	3	2	2	4	.3505	-.4316	.2277	0	1

Downloaded from informs.org by [45.82.243.160] on 14 July 2023, at 23:50 . For personal use only, all rights reserved.

statement "the side dishes and salads are delicious," we expect an 18.6% increase in the odds that the individual engages in positive word of mouth by recommending the restaurant. A stronger increase in the odds of 47.7% would be expected for a one-unit increase in the response to the statement "the entrées are delicious." Contrastingly, in light of its negative coefficient, we expect a decrease of 11.8% for a one-unit increase in the response to the statement "the desserts are delicious." This is somewhat surprising. We expect that greater agreement with the statements regarding side dishes and salads, entrées, and desserts (x_1, x_2, and x_3) would have a positive effect on recommendation behavior. Most likely, the antithetical sign for desserts is attributable to the problem of multicollinearity. Specifically, Table 1 reveals that the pairwise correlations between x_1, x_2, and x_3 all exceed 0.8. An antithetical sign is also observed for x_5: "the lighting in the restaurant is appropriate." In light of these results, interpretation and testing of regression coefficients is likely impaired by the multicollinearity that exists. Moreover, although the predictive results for the training and holdout samples are reasonably good, there is also the potential for overfitting in light of the relatively weak coefficients (and lack of statistical significance) for many of the variables. These issues are explored in greater detail in the model selection section.

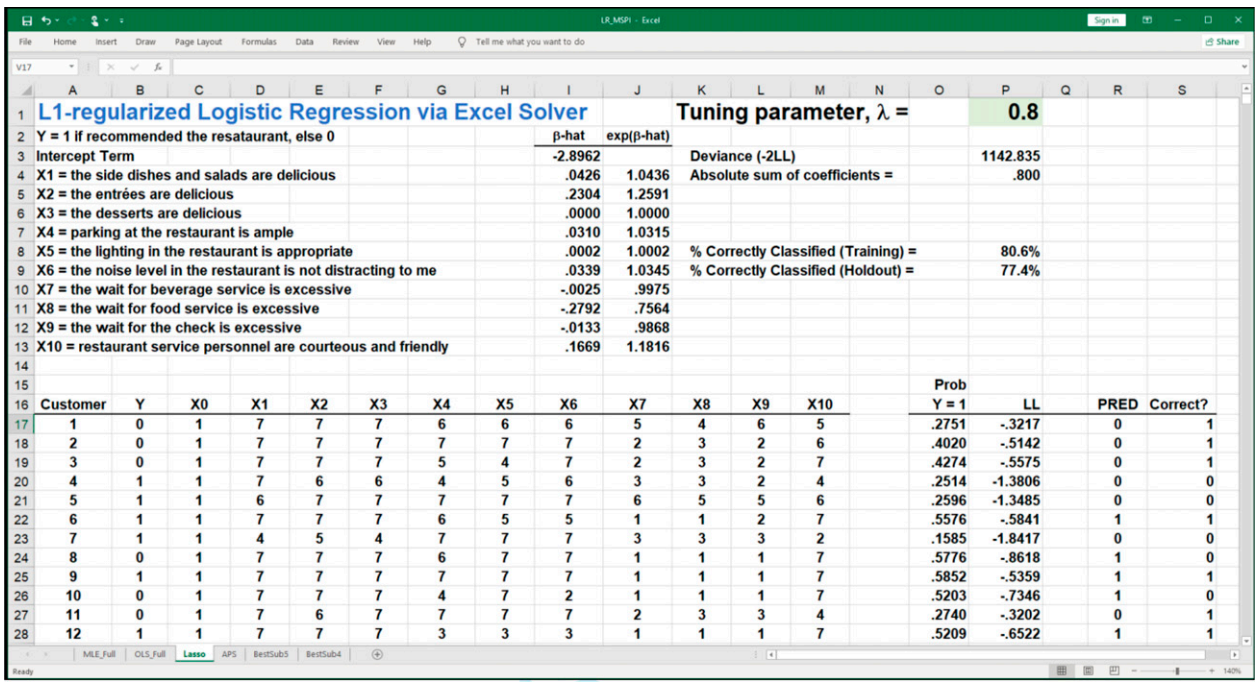
3.3. Model Selection via Lasso

The presentation of the lasso method is divided into three subsections: (i) the basic spreadsheet structure and some of the differences relative to the `MLE_Full` worksheet, (ii) Excel solver recommendations, and (iii) substantive interpretation.

3.3.1. Structure of the `Lasso` Worksheet. I used the `Lasso` worksheet displayed in Figure 3 to investigate the effect on prediction for the holdout sample at different values of λ. The lasso worksheet of the `LR_MSPI.xlsm` workbook is very similar to the `MLE_Full` worksheet. The primary differences are (i) that the significance testing of the coefficients is omitted in the lasso worksheet because the testing process used in `MLE_Full` is no longer viable and (ii) the inclusion of cell P1 to contain the user-specified tuning parameter, λ. In addition, the sum of the absolute values of the predictor-variable coefficients is computed and stored in cell P4. A constraint is added in the Excel solver to assure that cell P4 is equal to or less than cell P1.

3.3.2. Excel Solver Recommendations. Relative to the `MLE_Full` worksheet, the results obtained by the GRG nonlinear engine of the Excel solver for the lasso worksheet are more sensitive to the initial starting values for the coefficients. One approach to this problem is, for each value of λ evaluated, to run the GRG nonlinear engine of the solver multiple times using different (e.g., random) initial values for the coefficients. A second approach, which I use here, is to set the initial values for the logistic regression coefficients equal to those obtained for the unconstrained solution using the `MLE_Full` worksheet.

Figure 3. The Excel Worksheet "Lasso" Used to Implement l_1-Regularized Logistic Regression



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1	L1-regularized Logistic Regression via Excel Solver										Tuning parameter, λ =					0.8			
2	Y = 1 if recommended the resataurant, else 0								β-hat	exp(β-hat)									
3	Intercept Term								-2.8962		Deviance (-2LL)					1142.835			
4	X1 = the side dishes and salads are delicious								.0426	1.0436	Absolute sum of coefficients =					.800			
5	X2 = the entrées are delicious								.2304	1.2591									
6	X3 = the desserts are delicious								.0000	1.0000									
7	X4 = parking at the restaurant is ample								.0310	1.0315									
8	X5 = the lighting in the restaurant is appropriate								.0002	1.0002	% Correctly Classified (Training) =					80.6%			
9	X6 = the noise level in the restaurant is not distracting to me								.0339	1.0345	% Correctly Classified (Holdout) =					77.4%			
10	X7 = the wait for beverage service is excessive								-.0025	.9975									
11	X8 = the wait for food service is excessive								-.2792	.7564									
12	X9 = the wait for the check is excessive								-.0133	.9868									
13	X10 = restaurant service personnel are courteous and friendly								.1669	1.1816									
14																			
15															Prob				
16	Customer	Y	X0	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10		Y = 1	LL		PRED	Correct?
17	1	0	1	7	7	7	6	6	6	5	4	6	5		.2751	-.3217		0	1
18	2	0	1	7	7	7	7	7	7	2	3	2	6		.4020	-.5142		0	1
19	3	0	1	7	7	7	5	4	7	2	3	2	7		.4274	-.5575		0	1
20	4	1	1	7	6	6	4	5	6	3	3	2	4		.2514	-1.3806		0	0
21	5	1	1	6	7	7	7	7	7	6	5	5	6		.2596	-1.3485		0	0
22	6	1	1	7	7	7	6	5	5	1	1	2	7		.5576	-.5841		1	1
23	7	1	1	4	5	4	7	7	7	3	3	3	2		.1585	-1.8417		0	0
24	8	0	1	7	7	7	6	7	7	1	1	1	7		.5776	-.8618		1	0
25	9	1	1	7	7	7	7	7	7	1	1	1	7		.5852	-.5359		1	1
26	10	0	1	7	7	7	4	7	2	1	1	1	7		.5203	-.7346		1	0
27	11	0	1	7	6	7	7	7	7	2	3	3	4		.2740	-.3202		0	1
28	12	1	1	7	7	7	3	3	3	1	1	1	7		.5209	-.6522		1	1

Downloaded from informs.org by [45.82.243.160] on 14 July 2023, at 23:50 . For personal use only, all rights reserved.

Table 2. A Comparison of Best Subsets and Lasso Models

		Best subsets		l_1-regularized logistic regression (lasso)			
	Full model	Bestsub-5	Bestsub-4	$\lambda = 1.3$	$\lambda = 1.1$	$\lambda = 0.9$	$\lambda = 0.8$
b0	−5.1533	−5.3287	−5.0074	−4.8078	−4.0224	−3.2807	−2.8962
b1	0.1705			0.0921	0.0682	0.0514	0.0426
b2	0.3899	0.4261	0.4314	0.3214	0.2871	0.2488	0.2304
b3	−0.1253			0.0000	0.0000	0.0001	0.0000
b4	0.1561	0.1376		0.1137	0.0814	0.0482	0.0310
b5	−0.0490			0.0000	0.0000	0.0000	0.0002
b6	0.1412	0.1287	0.2095	0.1067	0.0778	0.0493	0.0339
b7	−0.0026			−0.0026	−0.0026	−0.0025	−0.0025
b8	−0.4057	−0.4624	−0.4634	−0.3730	−0.3274	−0.2986	−0.2792
b9	−0.0787			−0.0654	−0.0546	−0.0224	−0.0133
b10	0.2444	0.2566	0.2549	0.2251	0.2010	0.1787	0.1669
$-2LL$	1,108.175	1,111.065	1,114.615	1,110.861	1,118.108	1,132.551	1,142.835
AIC	1,130.175	1,123.065	1,124.615	1,132.861	1,140.108	1,154.551	1,164.835
BIC	1,141.175	1,129.065	1,129.615	1,143.861	1,151.108	1,165.551	1,175.835
Training (percentage correct)	78.0	78.2	79.1	78.4	79.5	80.1	80.6
Holdout (percentage correct)	75.2	75.4	76.8	75.2	76.4	76.6	77.4

3.3.3. Lasso Results. In this example, the sum of the absolute values of the predictor variable coefficients for the maximum likelihood logistic regression solution in Figure 1 is 1.764. Therefore, I began with an initial value of $\lambda = 1.7$ in step 2 and systematically decreased the value by 0.1 in step 3. The most interesting range of solutions was over the interval $0.8 \leq \lambda \leq 1.3$. Values of $\lambda > 1.3$ led to results comparable to those for the unconstrained model, whereas values of $\lambda < 0.8$ led to inferior results. Table 2 displays the results for $\lambda = 1.3$, $\lambda = 1.1$, $\lambda = 0.9$, and $\lambda = 0.8$. Although global optimality is not guaranteed for any of these solutions, the monotonic improvement in prediction accuracy for both the training and holdout samples over this range is useful for elucidating the potential benefits of the lasso method to students.

The results in Table 2 pertain to a single run of the lasso method for a given training and holdout sample. It is important to explain to students that a better approach to the selection of the appropriate value of λ is accomplished using some type of sampling approach, such as bootstrapping or k-fold cross-validation. This is not easy to implement in Excel because it is computationally demanding. For each value of λ evaluated, it is necessary to use the GRG nonlinear engine to estimate the lasso model for each sample and obtain an average measure of predictive performance for that value of λ.

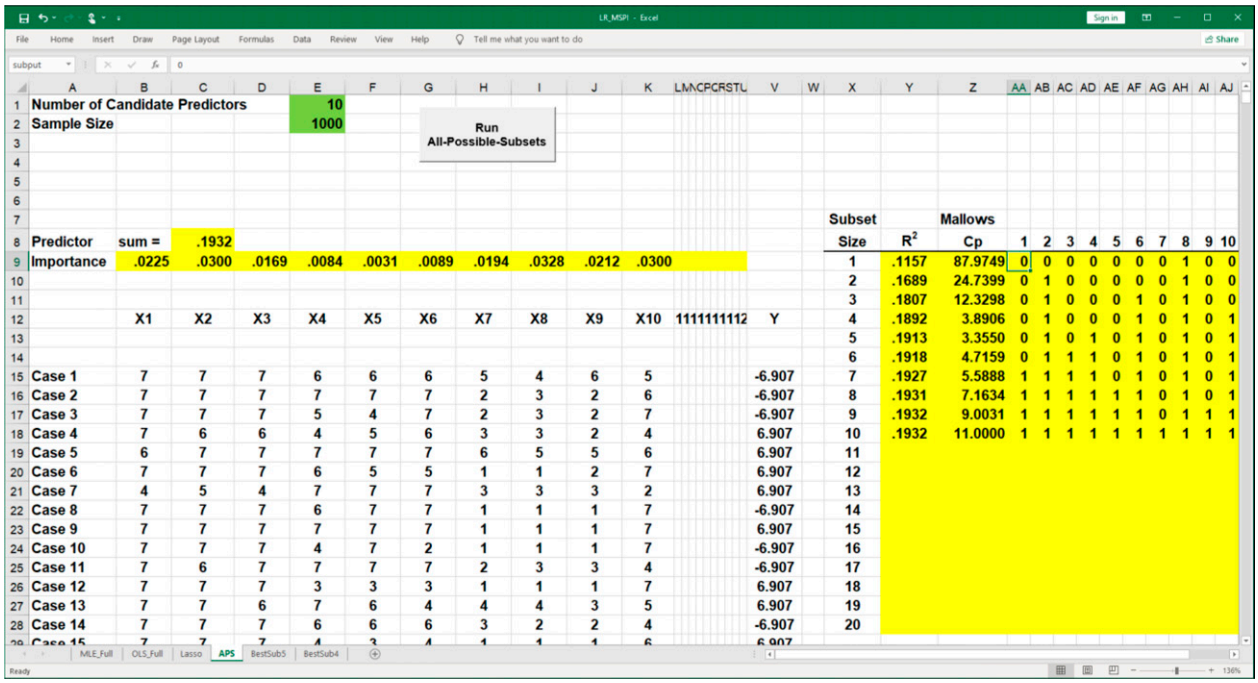
In addition, it should be noted that the lasso does not directly include or exclude variables but merely restricts the sum of absolute values of their coefficients. From the solution in Figure 3 for $\lambda = 0.8$, it does appear that the coefficients for x_3 and x_5 are driven to near zero such that they might be safely excluded from the model. Other variables have very small coefficients yet must be retained in the model.

3.4. Model Selection via All Possible Subsets OLS Regression

3.4.1. Running All Possible Subsets. Unlike the lasso method, the all possible subsets approach can be used to establish the "best subset" for each number of predictors on the interval $1 \leq u \leq v$. Accordingly, it explicitly includes or excludes variables for each subset size. I used the `APS` worksheet of the `LR_MSPI.xlsm` workbook (see Figure 4) to obtain the results for an all possible subsets OLS regression for the 10 predictors. The dependent variable measures for the OLS regression are constructed by "nudging" the binary outcomes as described by Schield (2017). Specifically, $y_i = 1$ is replaced by $y_i = 1 - \varepsilon$ and $y_i = 0$ is replaced by $y_i = \varepsilon$ and the dependent variable measures are $\log(y_i/(1 - y_i))$. I used $\varepsilon = 0.001$ for the analyses. All possible subsets are evaluated by clicking on the button "Run All-Possible-Subsets," which runs a VBA macro that performs sweep operations (Goodnight 1979) on the correlation matrix for the training sample to complete the analysis (see Brusco 2019). The macro reads the number of predictors (cell E1), the number of training sample observations (cell E2), and the training sample measures (B15:K1014 and V15:V1014). Upon evaluation of all possible subsets, the macro writes the best R^2 value for each subset size to cells Y9:Y18. The predictors associated with the best subset for each subset size are marked by a "1" in the cell range AA9:AJ18. The selection of the best subset size can be facilitated by the Mallows' C_p values (Mallows 1973) in cells Z9:Z18. Mallows' C_p is similar to the AIC; in fact, the selection

Downloaded from informs.org by [45.82.243.160] on 14 July 2023, at 23:50 . For personal use only, all rights reserved.

Figure 4. The Excel Worksheet "APS" Used to Implement All Possible Subsets Regression



	A	B	C	D	E	F	G	H	I	J	K	LMNOPQRSTU	V
1	Number of Candidate Predictors				10								
2	Sample Size				1000								
3													
4													
5													
6													
7													
8	Predictor	sum =	.1932										
9	Importance	.0225	.0300	.0169	.0084	.0031	.0089	.0194	.0328	.0212	.0300		
10													
11													
12		X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	1111111112	Y
13													
14													
15	Case 1	7	7	7	6	6	6	5	4	6	5		-6.907
16	Case 2	7	7	7	7	7	7	2	3	2	6		-6.907
17	Case 3	7	7	7	5	4	7	2	3	2	7		-6.907
18	Case 4	7	6	6	4	5	6	3	3	2	4		6.907
19	Case 5	6	7	7	7	7	7	6	5	5	6		6.907
20	Case 6	7	7	7	6	5	5	1	1	2	7		6.907
21	Case 7	4	5	4	7	7	7	3	3	3	2		6.907
22	Case 8	7	7	7	6	7	7	1	1	1	7		-6.907
23	Case 9	7	7	7	7	7	7	1	1	1	7		6.907
24	Case 10	7	7	7	4	7	2	1	1	1	7		-6.907
25	Case 11	7	6	7	7	7	7	2	3	3	4		-6.907
26	Case 12	7	7	7	3	3	3	1	1	1	7		6.907
27	Case 13	7	7	6	7	6	4	4	4	3	5		6.907
28	Case 14	7	7	7	6	6	6	3	2	2	4		-6.907
29	Case 15	7	7	7	4	3	4	1	1	1	6		6.907

Subset Size	R^2	Mallows Cp	1	2	3	4	5	6	7	8	9	10
1	.1157	87.9749	0	0	0	0	0	0	0	1	0	0
2	.1689	24.7399	0	1	0	0	0	0	0	1	0	0
3	.1807	12.3298	0	1	0	0	0	1	0	1	0	0
4	.1892	3.8906	0	1	0	0	0	1	0	1	0	1
5	.1913	3.3550	0	1	0	1	0	1	0	1	0	1
6	.1918	4.7159	0	1	1	1	0	1	0	1	0	1
7	.1927	5.5888	1	1	1	1	0	1	0	1	0	1
8	.1931	7.1634	1	1	1	1	1	1	0	1	0	1
9	.1932	9.0031	1	1	1	1	1	1	0	1	1	1
10	.1932	11.0000	1	1	1	1	1	1	1	1	1	1
11												
12												
13												
14												
15												
16												
17												
18												
19												
20												

of subset size based on the minimum value of Mallows' C_p is equivalent to selection based on the minimum AIC.

3.4.2. APS Predictive Results. The five-predictor subset $\{x_2, x_4, x_6, x_8, x_{10}\}$ yields the minimum value (3.35) of Mallows' C_p in Figure 4; however, the four-predictor subset $\{x_2, x_6, x_8, x_{10}\}$ is very close (3.89). Maximum likelihood estimation for the five- and four-predictor subsets was completed in the worksheets `BestSub5` and `BestSub4` of the `LR_MSPI.xlsm` workbook, respectively. The worksheets `ST_BestSub5` and `ST_BestSub4` of the `LR_SigTest.xlsm` workbook provide the results of the significance tests. Finally, the results for these two models are summarized in the Bestsub5 and Bestsub4 columns in Table 2.

The five-predictor model provides only slightly better prediction of the training and holdout samples than the full logistic regression model. The predictive performance of the four-predictor model is appreciably better and is second only to the lasso result using $\lambda = 0.8$ with respect to the percentage of correct classifications for the holdout sample (76.8% for all possible subsets versus 77.4% for lasso with $\lambda = 0.8$). Also, all predictors are significant ($\alpha = 0.05$) in the four-predictor model, but x_4 and x_6 are not significant in the five-predictor model.

3.5. Relative Predictor Importance via All Possible Subsets OLS Regression

Running the all possible subsets algorithm in the `APS` worksheet also generates the measures of relative predictor importance. The R^2-shares for each variable are shown in cells B9:K9 of Figure 4. The R^2-shares sum to the R^2 value of 0.1932 for the full regression model as shown in cell B8. The three most important explanatory variables based on the rank order of R^2 shares (from least to greatest) are x_8 ("the wait for food service is excessive"), x_2 ("the entrées are delicious"), and x_{10} ("restaurant service personnel are courteous and friendly"). This is not surprising given their importance in all of the predictive models displayed Table 2.

The last three explanatory variables in the rank order of R^2 shares are x_4 ("parking at the restaurant is ample"), x_6 ("the noise level in the restaurant is not distracting to me"), and x_5 ("the lighting in the restaurant is appropriate"). It is not surprising that these three aesthetic items would have less explanatory value for recommendation behavior than the food-quality items (x_1, x_2, x_3) and the service waiting time and quality items (x_7, x_8, x_9, x_{10}). However, there is a critical teaching point here that should not be overlooked. Students should recognize that a variable's exclusion from a predictive model does not necessarily mean that it is less important than some variable that is included in the model. For example, although it is less important as an explanatory variable, the variable x_6 (and also x_4) was often a more important "predictor" of recommendation behavior than x_1, x_3, x_7, and x_9 (notice that none of these four variables was selected for the four- or five-predictor subset). The reason is attributable to multicollinearity. For example, although x_1 and x_3 are good explanatory variables, because of their

Downloaded from informs.org by [45.82.243.160] on 14 July 2023, at 23:50 . For personal use only, all rights reserved.

high correlation to x_2, they add little predictive benefit when x_2 is included in the model. The same holds true for x_7 and x_9 when x_8 is already included in the predictive model. However, x_6 (and/or x_4) are clearly useful for augmenting the predictive model.

3.6. Pedagogical Assessment of the Workbooks

A formal assessment of learning goals was conducted in a master's level business analytics course. A series of videos was prepared to describe the Excel workbooks. Students watched the videos and completed an assignment that required them to modify the content of the workbooks, fit logistic regression models using the workbooks, and compare alternative model-selection approaches. Likert-scale statements to evaluate the learning goals described in the introduction were provided to the class at the end of the semester and the results are displayed in Table 3.

Eighty percent of the students agreed or strongly agreed with the statement pertaining to the assignment affording a greater understanding of maximum likelihood estimation of logistic regression coefficients. More than 75% of the students agreed or strongly agreed with the statements related to greater understanding of significance testing and model selection based on all possible subsets OLS regression. The percentage of students agreeing or strongly agreeing to the statements pertaining to the lasso (64%) and measurement of predictor importance (58%) were somewhat lower, which suggests that more time needs to be spent explaining these concepts and methods in future semesters.

4. Summary

Logistic regression is a critical tool for business analytics. Although the method can be implemented using a variety of software packages, many of these do not facilitate a conceptual understanding of maximum likelihood estimation, significance testing of logistic regression coefficients, or the relationship between OLS and logistic regression. The `LR_MSPI` and `LR_SigTest` Excel workbooks help to strengthen student comprehension of maximum likelihood estimation and testing of logistic regression coefficients.

The `LR_MSPI` Excel workbook also illustrates model selection in an easy-to-understand framework. For example, the workbook includes a worksheet for implementing l_1-regularized logistic regression (i.e., the lasso method), which constrains the sum of the absolute values of the model coefficients to a prescribed threshold. Additionally, an alternative approach based on all possible subsets regression is demonstrated in the workbook. This approach is also useful for better

Table 3. Summary of Results for Pedagogical Assessment: The Values in Each Column of Likert Scale Responses Are the Number of Respondents and, in Parentheses, the Percentage of Respondents

	(1) Strongly disagree	(2) Disagree	(3) Neither agree nor disagree	(4) Agree	(5) Strongly agree	Median
S1: After completing the logistic regression assignment, I am able to understand and implement maximum likelihood estimation of logistic regression coefficients using the Excel solver.	0 (0)	0 (0)	3 (20)	8 (53)	4 (27)	4
S2: After completing the logistic regression assignment, I am able to understand and implement significance testing of logistic regression coefficients using Excel.	0 (0)	0 (0)	3 (23)	9 (69)	1 (8)	4
S3: After completing the logistic regression assignment, I am able to understand and implement the l1-regularized regression (i.e., the lasso method) approach to model selection using the Excel solver.	0 (0)	0 (0)	5 (36)	6 (43)	3 (21)	4
S4: After completing the logistic regression assignment, I am able to understand and implement (via an Excel VBA macro) all possible subsets OLS regression as a model-selection heuristic for logistic regression.	0 (0)	1 (7)	2 (14)	8 (57)	3 (21)	4
S5: After completing the logistic regression assignment, I am able to understand and implement (via an Excel VBA macro) all possible subsets OLS regression to provide a heuristic assessment of the relative importance of the predictors for logistic regression.	0 (0)	1 (7)	5 (36)	5 (36)	3 (21)	4

Downloaded from informs.org by [45.82.243.160] on 14 July 2023, at 23:50 . For personal use only, all rights reserved.

understanding of the relative importance of the independent variables.

Acknowledgments

I sincerely appreciate the constructive comments of two anonymous reviewers and the associate editor, which led to significant improvements in this article.

References

Addinsoft (2021) XLSTAT (version 2021.2). Accessed April 30, 2021, https://www.xlstat.com/en/.

Agbemava E, Nyarko IK, Adade TC, Bediako AK (2016) Logistic regression analysis of predictors of loan defaults by customers of non-traditional banks in Ghana. *Eur. Sci. J.* 12(1):175–189.

Akaike H (1973) Information theory and an extension of the maximum likelihood principle. Petrov BN, Csaki BF, eds. *Second Internat. Sympos. Inform. Theory* (Academiai Kiado, Budapest), 267–281.

Alboqami H, Al-Karaghouli W, Baeshen Y, Erkan I, Evans C, Ghoneim A (2015) Electronic word of mouth in social media: The common characteristics of retweeted and favourited marketer-generated content posted on Twitter. *Internat. J. Internet Marketing Advertising* 9(4):338–358.

Anderson L, Krathwohl DA (2001) *Taxonomy for Learning, Teaching and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives* (Longman, New York).

Azen R, Budescu DV (2003) The dominance analysis approach for comparing predictors in multiple regression. *Psych. Methods* 8(2):129–148.

Azen R, Traxel N (2009) Using dominance analysis to determine predictor importance in logistic regression. *J. Ed. Behav. Statist.* 34(3):319–347.

Bejaei M, Wiseman K, Cheng KMT (2015) Developing logistic regression models using purchase attributes and demographics to predict the probability of purchases of regular and specialty eggs. *British Poultry Sci.* 56(4):1–11.

Bloom BS, Engelhart MD, Furst EJ, Hill WH, Krathwohl DR (1956) Taxonomy of Educational Objectives: The Classification of Educational Goals. *Handbook I: Cognitive Domain* (David McKay Company, New York).

Brownlee, J (2020) *Imbalanced Classification with Python: Better Metrics, Balance Skewed Classes, Cost-Sensitive Learning* (Machine Learning Mastery, San Juan, Puerto Rico).

Brusco M (2019) An Excel spreadsheet and VBA macro for model selection and predictor importance using all-possible-subsets regression. Spreadsheets in Education, 12(1). Accessed April 29, 2019, https://sie.scholasticahq.com/article/8064-an-excel-spreadsheet-andvba-macro-for-model-selection-and-predictor-importance-using-all-possible-subsetsregression.

Budescu DV (1993) Dominance analysis: A new approach to the problem of relative importance of predictors in multiple regression. *Psych. Bull.* 114(3):542–551.

Campbell MJ (1998) Teaching logistic regression. *Proc. Internat. Conf. Teaching Statist.* (International Association for Statistical Education, Singapore), 284–289.

Carlberg C (2013) *Decision Analytics: Microsoft Excel* (Que Publishing Company, Seattle).

GAISE (2016) College report ASA revision committee. Guidelines for assessment and instruction in statistics education college report 2016. Accessed December 30, 2020, http://www.amstat.org/education/gaise.

Goodnight JH (1979) A tutorial on the SWEEP operator. *Amer. Statist.* 33(3):149–158.

Grömping U (2015) Variable importance in regression models. *Wiley Interdisciplinary Rev. Comput. Statist.* 7:137–152.

Hoerl AE, Kennard RW (1970) Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics* 12(1):55–67.

Hosmer DW, Lemeshow S, Sturdivant RX (2013) *Applied Logistic Regression*, 3rd ed. (Wiley, New York).

Huggins E, Bailey M, Guardiola I (2020) Converting point spreads into probabilities: A case study for teaching business analytics. *INFORMS Trans. Ed.* 21(1):61–63.

IBM Corp. (2019) IBM SPSS Statistics for Windows, version 26.0. Accessed December 27, 2020. (IBM Corp., Armonk, NY).

Jaggia S, Kelly A, Lertwachara K, Chen L (2020) Applying the CRISP-DM framework for teaching business analytics. *Decision Sci. J. Innovative Ed.* 18(4):612–633.

Kvam PH, Sokol J (2004) Teaching statistics with sports examples. *INFORMS Trans. Ed.* 5(1):75–87.

Lindeman RH, Merenda PF, Gold RZ (1980) *Introduction to Bivariate and Multivariate Analysis* (Scott, Foresman, Glenview, IL).

Mallows CL (1973) Some comments on *Cp*. *Technometrics* 15(4): 661–675.

Menard S (2010) *Logistic Regression: From Introductory to Advanced Concepts and Applications* (Sage, Thousand Oaks, CA).

Miller AJ (2002) *Subset Selection in Regression*, 2nd ed. (Chapman and Hall, London).

Morrell CH, Auer RE (2007) Trashball: A logistic regression classroom activity. *J. Statist. Ed.* 15(1). Accessed December 30, 2020, https://www.tandfonline.com/doi/full/10.1080/10691898.2007.11889455.

Pinder J (2013) An Excel solver exercise to introduce nonlinear regression. *Decision Sci. J. Innovative Ed.* 11(3):263–278.

Ragsdale CT (2018) *Spreadsheet Modeling and Decision Analysis*, 8th ed. (Cengage, Boston).

Ragsdale CT, Stam A (1992) Introducing discriminant analysis to the business statistics curriculum. *Decision Sci.* 23(3): 724–745.

Ritter B (2013) How to do logistic regression in Excel. Accessed December 29, 2020, https://www.youtube.com/watch?v=rbKtZcrTlr8.

Salamah U, Ramayanti D (2018) Implementation of logistic regression algorithm for complaint text classification in Indonesian ministry of marine and fisheries. *Internat. J. Comput. Techniques* 5(5):74–78.

Schield M (2017) Teaching logistic regression using ordinary least squares in Excel. 2017 JSM Proc. Papers *Presented Joint Statist. Meetings* (American Statistical Association, Baltimore), 2963–2987). Accessed December 29, 2020, http://www.statlit.org/pdf/2017-Schield-ASA.pdf.

Schwartz G (1978) Estimating the dimension of a model. *Ann. Statist.* 6(2):461–464.

Tibshirani R (1996) Regression shrinkage and selection via the Lasso. *J. Roy. Statist. Soc. B.* 58(1):267–288.

Zaghdoudi T (2013) Bank failure prediction with logistic regression. *Internat. J. Econom. Financial Issues* 3(2):537–543.