SPREADSHEETS in EDUCATION

Bond University
Volume 12 | Issue 1 | 2019

An Excel Spreadsheet and VBA Macro for Model Selection and Predictor Importance Using All-Possible-Subsets Regression

Michael Brusco
Florida State University, Tallahassee, FL USA, mbrusco@business.fsu.edu

---

An Excel Spreadsheet and VBA Macro for Model Selection and Predictor Importance Using All-Possible-Subsets Regression

Abstract

Two of the most challenging aspects of teaching regression analysis pertain to model selection and the relative importance of predictors. All-possible-subsets (APS) regression is a particularly useful tool for addressing both of these topics. When using regression analysis to analyze customer satisfaction data in marketing analytics courses, an Excel spreadsheet for implementing APS regression has led to fruitful discussions regarding: (1) which predictor variables should be retained in the regression model, and (2) which predictor variables are most useful for explaining the dependent variable. The spreadsheet, which uses a VBA macro to run APS regression via sweep operations on the correlation matrix, is scalable for up to $p = 20$ predictors and reports the best subset for all subset sizes on the interval $1 \leq q \leq p$. The selection of an appropriate value for q is facilitated by R^2 and Mallows' C_p information. Relative predictor importance is established via a general dominance measure obtained from the APS regression.

Keywords: Spreadsheets, VBA, all-possible-subsets, model selection, predictor importance, customer satisfaction

1. Introduction

This paper focuses on the multiple linear regression model of the form:

$$y_i = \beta_0 + \beta_1 x_{i1} + \ldots + \beta_p x_{ip} + \varepsilon_i, \qquad \text{for } 1 \le i \le n, \tag{1}$$

where n is the number of cases, $P = \{x_1, \ldots, x_p\}$ is a set of $p \ge 2$ candidate predictor variables, x_{ij} is the measurement for case i on predictor variable j, y_i is the dependent variable measurement for case i, β_0 is an intercept term, β_j ($1 \le j \le p$) are slope coefficients, and ε_i is the error term for case i. Although Excel spreadsheet tools are readily available for model fitting, hypothesis testing and diagnostics for regression analysis [20], the particular interest herein is on the related topics of model selection and relative predictor importance.

Model selection (also known as subset selection or variable selection) in regression analysis is concerned with the choice of q ($1 \le q \le p$) predictors from the full set (P) of candidate predictors. Methods for model selection include stepwise regression [9], all-possible subsets (APS) regression [11, 17, 23, 27], branch-and-bound programming [5, 10], the garotte [3], and the lasso [30]. Miller [21] provides an extended treatment of many of these methods. *Predictor importance* is concerned with determining a ranking of the predictors with respect to their influence on the dependent variable. Numerous measures of predictor importance have been proposed. Bivariate correlations, partial correlations, and standardized regression coefficients are among the simplest measures, but more sophisticated indices are also available. Thorough evaluations of predictor importance measures have been provided by Nimon and Oswald [22] and Grömping [14].

APS regression can provide a unifying framework for model selection and predictor importance analyses [22]. To facilitate model selection, for each possible subset size, q, APS regression obtains the subset that maximizes R^2. The selection of the best value of q can then be based on a variety of alternative indices, such as adjusted-R^2 [31], Akaike's [1, 2] information criterion (AIC), Schwartz' [26] Bayesian information

criterion (BIC), and Mallows' C_p [19]. APS also facilitates the measurement of relative predictor importance by enabling the computation of the general dominance index [6, 18]. The first step in computing the general dominance measure is to find the average improvement in explained variation associated with each predictor for each subset size. The general dominance measure for a given predictor is then computed as the average over all subset sizes. The sum of the general dominance measures over all predictors has the desirable property of equaling R^2.

Although APS regression can enhance student understanding of model selection and predictor importance for a variety of application contexts, the particular interest here pertains to its use in marketing research/analytics courses at both the graduate and undergraduate levels. In particular, we focus on the well-established use of regression analysis in the study of customer satisfaction [7, 13, 16, 25]. In our marketing analytics courses, we have commonly analyzed customer satisfaction data using regression analysis via SPSS. However, an enhanced understanding of model selection and predictor importance has been accomplished using an Excel spreadsheet that includes a VBA macro for implementing APS regression. The distinct benefits of the APS analysis are:

(1) A clear communication that there are 2^p possible regression models (or $2^p - 1$ if the model with no predictors is ignored) for a set of p predictors and that, for a given subset size q, there are $p!/(q!(p-q)!)$ models.

(2) The provision of the best subset for each possible subset size, along with the corresponding R^2 and Mallows' C_p information.

(3) A recognition that stepwise regression does not always obtain the best number of predictors.

(4) A recognition that the q-predictor model obtained via stepwise regression is not necessarily the q-predictor subset yielding the largest R^2.

(5) Building on (2), (3), and (4), there is an understanding that there can often be many competing models yielding roughly the same R^2, which can make it difficult to establish which predictors are most important.

(6) A recognition that standardized slope coefficients are not necessarily the best indicators of the relative importance of predictors.

(7) An understanding that APS regression also yields general dominance measures for establishing the relative importance of the predictors.

2. Background

2.1. Customer Satisfaction in Marketing Research

Students taking an undergraduate course in marketing research or a graduate course in marketing analytics have typically had an undergraduate statistics course where regression analysis was covered. Nevertheless, some reminders of the basics of regression analysis and pertinent issues such as assumptions, diagnostics, outliers, and multicollinearity are advisable. Emphasis swiftly moves to the analysis of customer satisfaction data, where it is clarified that multiple regression is primarily used in an *explanatory* rather than a *predictive* role. That is, the focus is typically on the identification of the key drivers of customer satisfaction rather than actual forecasts of customer satisfaction.

2.2. All-Possible-Subsets Regression

All-possible-subsets regression is explained at a conceptual level using a small number of candidate predictors. For example, for $p = 4$, the $2^4 = 16$ possible subsets are: $\varnothing$, {1}, {2}, {3}, {4}, {1, 2}, {1, 3}, {1, 4}, {2, 3}, {2, 4}, {3, 4}, {1, 2, 3}, {1, 2, 4}, {1, 3, 4}, {2, 3, 4}, {1, 2, 3, 4}. The students immediately recognize that it would be possible to run APS regression for small p; however, it is clarified that the VBA/spreadsheet implementation is possible for roughly $p = 20$, where the number of possible subsets is just over one million.

It is explained to the students that the playing field is level for subsets of the same size and, therefore, the subset yielding the largest value of R^2 can be identified as the "best subset" for each q ($1 \le q \le p$). However, because the addition of a variable to a regression model cannot possibly reduce R^2, we need some basis for comparing the best subsets across different q. One simple approach is to choose q based on inspection of the R^2 values. That is, a good value of q is one where the improvement in R^2 when moving from q-1 to q is large but the improvement when moving from q to q+1 is small. It is then noted that this type of ad hoc rule is often both impractical and suboptimal, but that some formal indices are available. The particular measure used in the APS regression approach is Mallows' C_p:

$$C_p(P_q) = [SSE(P_q)/MSE(P)] - (n - 2(q+1)), \qquad \text{for } 1 \le q \le p, \qquad (2)$$

where P_q is the best subset of q predictors, $SSE(P_q)$ is the error sum of squares for the regression using only the predictors in P_q, $MSE(P)$ is the mean squared error associated with the full regression model using all candidate predictors. We adopt the convention used by Olejnik et al. [23] that the subset yielding the minimum value of $C_p(P_q)$ is selected, while recognizing that it is also desirable for $C_p(P_q)$ to be close to $q + 1$.

2.3. Predictor Importance

Students are advised that a simple and intuitive ranking of the predictors can be based on their bivariate correlation with the dependent variable. However, such an approach completely ignores the multivariate nature of the regression analysis. Perhaps a better approach is to establish a ranking based on the standardized regression coefficients, which are immediately available from SPSS output; however, the use of this measure has also been criticized. An index defined by the product of the bivariate correlation and the standardized regression coefficient, which was originally proposed by Hoffman [15] and is often associated with the work of Pratt [24], has also been proposed and is favorably viewed by some researchers [28, 29]. The APS

implementation allows for the use of far richer measure, the general dominance index [6, 18], which has been well-received in the literature [14, 22].

3. The SWEEP Algorithm

The primary engine for running APS regression is a SWEEP algorithm that operates on the $(p+1) \times (p+1)$ correlation matrix, **R**, associated with the p predictors and the dependent variable (by convention, the dependent variable corresponds to row $p+1$ and column $p+1$ of the matrix). SWEEP algorithms have been widely used for subset selection problems in regression [5, 10, 11], as well as other variable-selection optimization problems in multivariate statistics [4, 8]. An excellent tutorial on the history and applications of the SWEEP algorithm is provided by Goodnight [12].

The SWEEP algorithm transforms a given $(p+1) \times (p+1)$ matrix **R** into a new matrix **S** using the following process.

Step 1. Select one of the predictor variables h $(1 \le h \le p)$ for sweeping.

Step 2. Set the pivot element as $\lambda = 1/r_{hh}$.

Step 3. Set $s_{ij} = r_{ij} - \lambda r_{hj} r_{ih}$ for all $1 \le i \ne h$, $j \ne h \le p+1$.

Step 4. Set $s_{hj} = \lambda r_{hj}$ for all $1 \le j \ne h \le p+1$.

Step 5. Set $s_{ih} = -\lambda r_{ih}$ for all $1 \le i \ne h \le p+1$.

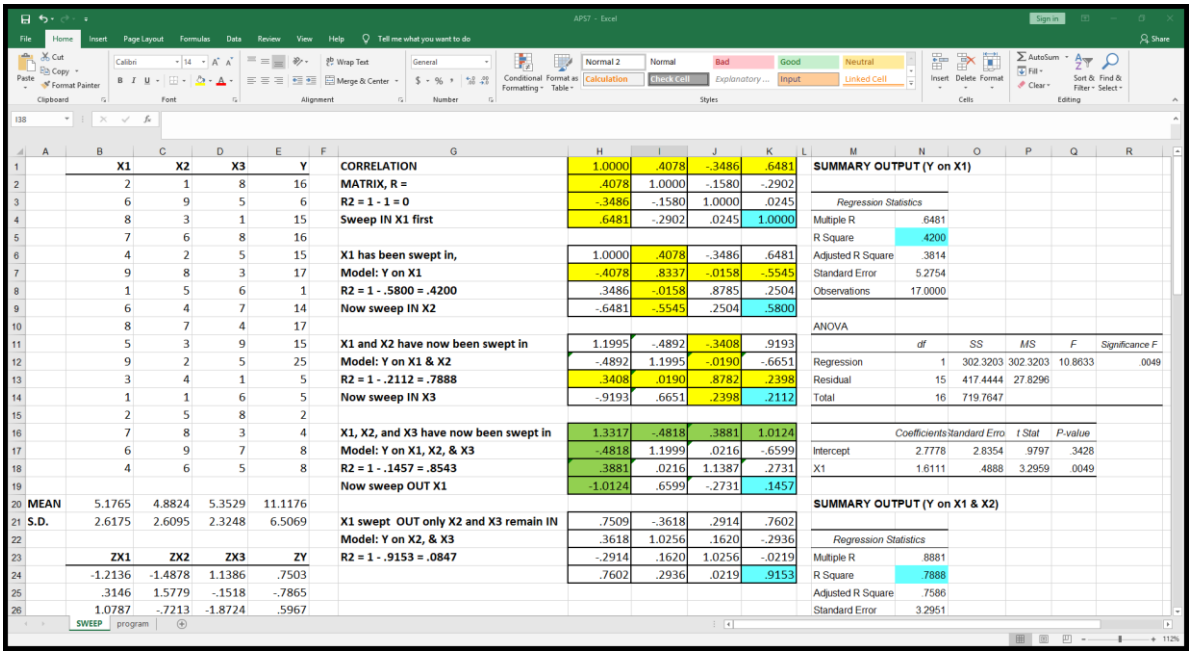
Step 6. $s_{hh} = \lambda$.

Step 7. **R** = **S**.

A predictor variable, h, is selected for a SWEEP operation in Step 1. Step 2 defines the pivot element as the inverse of the main diagonal element of **R** corresponding to predictor h (i.e., $\lambda = 1/r_{hh}$). Step 3 computes all elements of **S** that do not correspond to either row h or column h. Step 4 computes the row h elements of **S**, with the exception of the main diagonal element in that row (i.e., column h in row h). Step 5 computes the column h elements of **S**, with the exception of the main diagonal element in that column (i.e., row h in column h). Step 6 sets the value for column h of row h in **S**. Step 7 resets **R**

to **S**, which enables the exact same set of steps to be repeated for sweeping other variables in or out.

The SWEEP worksheet in the APS workbook (see Figure 1) provides a small numerical example to illustrate the algorithm. The problem consists of $n = 17$ respondents measured on $p = 3$ predictors (x_1, x_2, x_3) and a dependent variable (y). The raw data are provided in cells B2:E18 of the workbook. The means and standard deviations of the dependent variable and predictors are computed in cells B20:E21 to facilitate the computation of z-scores for each variable. The z-scores, which are obtained by differencing the raw variable measures from the mean and dividing by the standard deviation, are displayed in cells B24:E40. Defining the 17×4 matrix in these cells as **Z**, the correlation matrix is computed as $\mathbf{R} = (1/17)\mathbf{Z}^{\mathrm{T}}\mathbf{Z}$. This is accomplished using the MMULT function in Excel and the $(p+1) \times (p+1)$ correlation matrix, **R**, is contained in cells H1:K4. The element in row $p + 1$, column $p + 1$ of the matrix (i.e., cell K4) is of particular interest because one minus this value is equal to R^2 for the regression model. Because none of the predictors have been swept at this point, cells H1:K4 corresponds to an absence of predictors in the model and, therefore, $R^2 = 1 - 1 = 0$.



	B	C	D	E	G	H	I	J	K	M	N	O	P	Q	R
1	X1	X2	X3	Y	CORRELATION	1.0000	.4078	-.3486	.6481	SUMMARY OUTPUT (Y on X1)					
2	2	1	8	16	MATRIX, R =	.4078	1.0000	-.1580	-.2902						
3	6	9	5	6	R2 = 1 - 1 = 0	-.3486	-.1580	1.0000	.0245	*Regression Statistics*					
4	8	3	1	15	Sweep IN X1 first	.6481	-.2902	.0245	1.0000	Multiple R	.6481				
5	7	6	8	16						R Square	.4200				
6	4	2	5	15	X1 has been swept in,	1.0000	.4078	-.3486	.6481	Adjusted R Square	.3814				
7	9	8	3	17	Model: Y on X1	-.4078	.8337	-.0158	-.5545	Standard Error	5.2754				
8	1	5	6	1	R2 = 1 - .5800 = .4200	.3486	-.0158	.8785	.2504	Observations	17.0000				
9	6	4	7	14	Now sweep IN X2	-.6481	-.5545	.2504	.5800						
10	8	7	4	17						ANOVA					
11	5	3	9	15	X1 and X2 have now been swept in	1.1995	-.4892	-.3408	.9193		*df*	*SS*	*MS*	*F*	*Significance F*
12	9	2	5	25	Model: Y on X1 & X2	-.4892	1.1995	-.0190	-.6651	Regression	1	302.3203	302.3203	10.8633	.0049
13	3	4	1	5	R2 = 1 - .2112 = .7888	.3408	.0190	.8782	.2398	Residual	15	417.4444	27.8296		
14	1	1	6	5	Now sweep IN X3	-.9193	.6651	.2398	.2112	Total	16	719.7647			
15	2	5	8	2											
16	7	8	3	4	X1, X2, and X3 have now been swept in	1.3317	-.4818	.3881	1.0124		*Coefficients*	*Standard Erro*	*t Stat*	*P-value*	
17	6	9	7	8	Model: Y on X1, X2, & X3	-.4818	1.1999	.0216	-.6599	Intercept	2.7778	2.8354	.9797	.3428	
18	4	6	5	8	R2 = 1 - .1457 = .8543	.3881	.0216	1.1387	.2731	X1	1.6111	.4888	3.2959	.0049	
19					Now sweep OUT X1	-1.0124	.6599	-.2731	.1457						
20	MEAN	5.1765	4.8824	5.3529	11.1176					SUMMARY OUTPUT (Y on X1 & X2)					
21	S.D.	2.6175	2.6095	2.3248	6.5069	X1 swept OUT only X2 and X3 remain IN	.7509	-.3618	.2914	.7602						
22						Model: Y on X2, & X3	.3618	1.0256	.1620	-.2936	*Regression Statistics*					
23	ZX1	ZX2	ZX3	ZY		R2 = 1 - .9153 = .0847	-.2914	.1620	1.0256	-.0219	Multiple R	.8881				
24	-1.2136	-1.4878	1.1386	.7503			.7602	.2936	.0219	.9153	R Square	.7888				
25	.3146	1.5779	-.1518	-.7865							Adjusted R Square	.7586				
26	1.0787	-.7213	-1.8724	.5967							Standard Error	3.2951				

Figure 1: A worksheet containing a small example to illustrate the sweep operator.

The matrix in cells H6:K9 was obtained by setting $h = 1$ at Step 1 and applying the remaining steps of the SWEEP algorithm. Sweeping on the variable x_1 provides the R^2 value for the regression model for y using only x_1 as predictor, which is equal to one minus the value in cell K9 (i.e., 1 - .5800 = .4200). To verify this result, the data analysis toolpack in Excel was used to run regression analysis on the raw data and the results are displayed in cells M1:R18. The value of $R^2 = .4200$ in cell N5 confirms the results obtained by the SWEEP algorithm.

Next, the predictor x_2 was swept into solution by setting $h = 2$ and applying the SWEEP algorithm to the matrix in cells H6:K9. The result is the matrix in cells H11:K14, which corresponds to the two-predictor model associated with x_1 and x_2. The value of R^2 for this model is found by taking one minus the value in cell K14 (i.e., 1 - .2112 = .7888). The verification of this result was performed using the data analysis toolpack and the results are displayed in cells M20:R38. The value of $R^2 = .7888$ in cell N24 confirms the SWEEP algorithm results.

The predictor x_3 was swept into solution by setting $h = 3$ and applying the SWEEP algorithm to the matrix in cells H11:K14. The result is the matrix in cells H16:K19, which corresponds to the full regression model with all three predictors (x_1, x_2, x_3). The value of R^2 for the full regression model is obtained by taking one minus the value in cell K19 (i.e., 1 - .1457 = .8543). The verification of this result was performed using the data analysis toolpack and the results are displayed in cells M40:R59. The value of $R^2 = .8543$ in cell N44 confirms the SWEEP algorithm results.

At this point, all three predictor variables have been swept into the model. Choosing one of these variables and applying the SWEEP algorithm again will sweep that variable out of the model. To illustrate, we selected variable x_1. The predictor x_1 was swept out of solution by setting $h = 1$ and applying the SWEEP algorithm to the matrix in cells H16:K19. The result is the matrix in cells H21:K24, which corresponds to the two-predictor model associated with x_2 and x_3. The value of R^2 for this model is found

by taking one minus the value in cell K24 (i.e., 1 - .9153 = .0847). The verification of this result was performed using the data analysis toolpack and the results are displayed in cells M61:R79. The value of R^2 = .0847 in cell N65 confirms the SWEEP algorithm results.

The numerical example in the SWEEP worksheet shows that the SWEEP algorithm is an efficient approach for obtaining the R^2 values associated with different regression models. Variables can be swept in (or out) of the model very rapidly. Such an approach is far less computationally intensive than running a separate regression analysis for each possible subset. The VBA macro in the APS workbook uses the SWEEP algorithm to conduct a systematic evaluation of all possible subsets in an efficient manner.

4. The Illustration

4.1. The Data Set and SPSS Results

The example used for illustration consists of n = 60 respondents measured on p = 10 Likert-scale questions regarding a dining experience at a restaurant. Although the VBA macro is dimensioned to accommodate up to n = 2500 cases and p = 20 predictor variables, a 10-variable set of candidate predictors seemed reasonable to illustrate the principles of model selection and predictor importance. Using fewer candidate predictors might result in a problem that lacks a sufficient degree of multivariate complexity, whereas using more might create an overload for students for whom this is their first exposure to the relevant topics.

The students are told that a marketing research firm is interested in determining the key drivers of overall customer satisfaction at a popular restaurant, and that linear regression will be used to determine which drivers are most important. Data were collected from n = 60 patrons of the restaurant. The metric dependent variable, y, is overall satisfaction, which is measured on a scale of 0 to 100 with larger values indicating greater satisfaction. The firm is considering 10 candidate independent variables to predict overall satisfaction. Each candidate predictor variable is a Likert

scale response to a statement ranging from 1 = strongly disagree to 7 = strongly agree. The 10 predictor statements are shown in Table 1

Table 1: Predictor variables for the example.

x_1 = the side dishes and salads are delicious
x_2 = the entrées are delicious
x_3 = the desserts are delicious
x_4 = parking at the restaurant is ample
x_5 = the lighting in the restaurant is appropriate
x_6 = the noise level in the restaurant is not distracting to me
x_7 = the wait for beverage service is excessive
x_8 = the wait for food service is excessive
x_9 = the wait for the check is excessive
x_{10} = restaurant service personnel are courteous and friendly

The full regression model for the customer satisfaction data was fit using SPSS. The value of R^2 is .5837 and the overall F-test for the regression model is significant (p-value < .01). Table 2 provides the estimates of the slope coefficients, the standardized regression coefficients, t-statistics and significance levels for the predictor variables. Interestingly, only one of the 10 predictors (x_{10}) is statistically significant. Moreover, the signs of some of the slope coefficients are discordant with theoretical expectations. For example, variable x_6 has a negative slope coefficient despite the fact that a positive coefficient should be expected (i.e., the more someone agrees with the statement that the noise level is not distracting, the higher satisfaction should be).

Table 2: Regression coefficients for the full regression model (SPSS results)

		Unstandardized Coefficients		Standardized Coefficients		
Model		B	Std. Error	Beta	t	Sig.
	(Constant)	26.8449	11.6630		2.3017	.0256
	x_1	2.7464	1.5306	.2635	1.7943	.0789
	x_2	2.8106	1.6715	.2819	1.6814	.0990
	x_3	2.2305	1.6671	.2155	1.3380	.1871
	x_4	.0045	2.0555	.0004	.0022	.9983
	x_5	3.1823	1.6766	.3200	1.8980	.0636
	x_6	-2.2068	1.9650	-.2109	-1.1231	.2669
	x_7	.7639	1.6636	.0790	.4592	.6481

x_8	-2.0005	1.6010	-.1989	-1.2496	.2174
x_9	-1.6029	1.1427	-.1630	-1.4028	.1670
x_{10}	3.8911	1.2138	.3664	3.2058	.0024

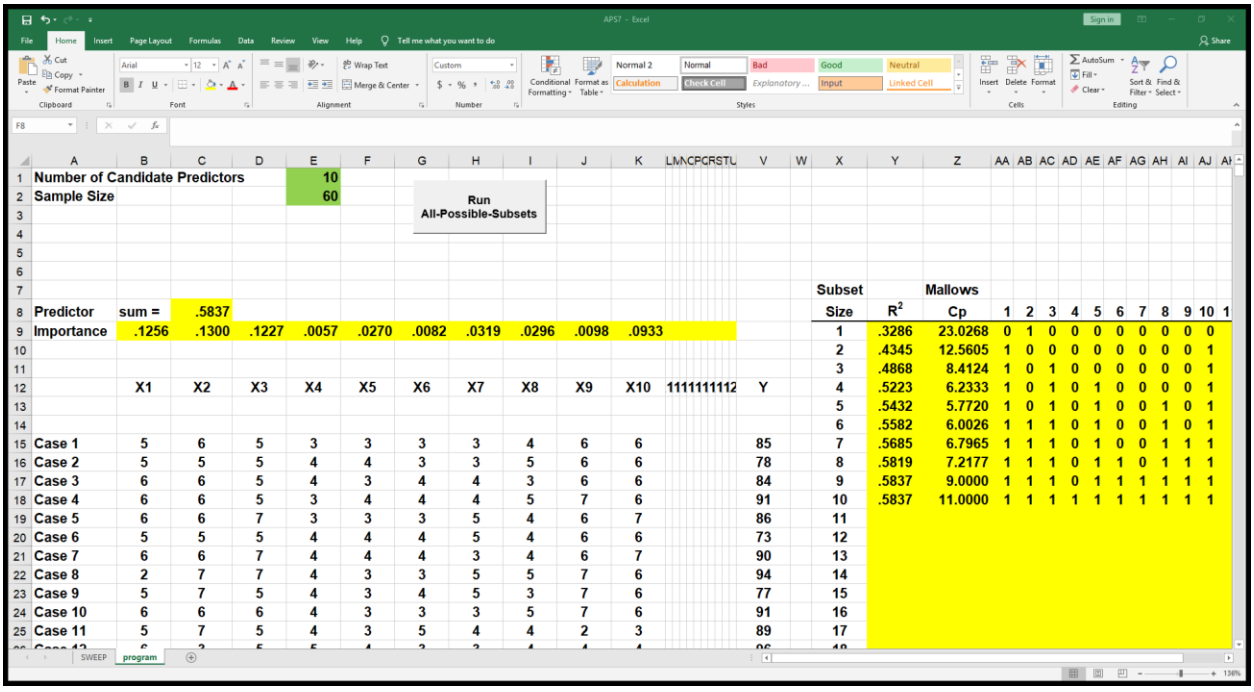
Stepwise regression was applied using SPSS and the results are provided in Table 3. The stepwise approach resulted in a four-predictor subset (x_1, x_2, x_8, x_{10}) with R^2 = .5176. All four predictors were significant at the α = .05 level and have the correct theoretical sign.

Table 3: Regression coefficients for the stepwise regression model (SPSS results)

		Unstandardized Coefficients		Standardized Coefficients		
Model		B	Std. Error	Beta	t	Sig.
	(Constant)	33.4128	8.1683		4.0906	.0001
	x_2	2.9541	1.2906	.2963	2.2889	.0259
	x_{10}	3.8765	1.0337	.3650	3.7502	.0004
	x_1	3.4081	1.3915	.3270	2.4492	.0175
	x_8	-2.0796	1.0151	-.2068	-2.0487	.0453

4.2. The APS Worksheet and VBA Macro

Figure 2 provides a screenshot of the Excel spreadsheet for APS regression. The number of candidate predictors is entered in cell E1 and the sample size in cell E2. The sample measurements for the 10 predictors and the dependent variable must begin in cells B15 and V15, respectively. The button "Run All-Possible-Subsets" runs the VBA macro for APS when clicked. The VBA macro reads n, p, the predictor variable data matrix (**X**) and the dependent variable (**y**) from the worksheet and constructs a p+1 correlation matrix for the predictors and dependent variables. Sweep operations on the correlation matrix are used to compute the value of R^2 for all possible subsets.



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC	AD	AE	AF	AG	AH	AI	AJ	AK
1	Number of Candidate Predictors				10																															
2	Sample Size				60																															
3																																				
4																																				
5																																				
6																																				
7																								Subset		Mallows										
8	Predictor	sum =	.5837																					Size	R^2	Cp	1	2	3	4	5	6	7	8	9	10	1
9	Importance	.1256	.1300	.1227	.0057	.0270	.0082	.0319	.0296	.0098	.0933													1	.3286	23.0268	0	1	0	0	0	0	0	0	0	0
10																								2	.4345	12.5605	1	0	0	0	0	0	0	0	0	1
11																								3	.4868	8.4124	1	0	1	0	0	0	0	0	0	1
12		X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	1111111112										Y		4	.5223	6.2333	1	0	1	0	1	0	0	0	0	1
13																								5	.5432	5.7720	1	0	1	0	1	0	0	1	0	1
14																								6	.5582	6.0026	1	1	1	0	1	0	0	1	0	1
15	Case 1	5	6	5	3	3	3	3	4	6	6											85		7	.5685	6.7965	1	1	1	0	1	0	0	1	1	1
16	Case 2	5	5	5	4	4	3	3	5	6	6											78		8	.5819	7.2177	1	1	1	0	1	1	0	1	1	1
17	Case 3	6	6	5	4	3	4	4	3	6	6											84		9	.5837	9.0000	1	1	1	0	1	1	1	1	1	1
18	Case 4	6	6	5	3	4	4	4	5	7	6											91		10	.5837	11.0000	1	1	1	1	1	1	1	1	1	1
19	Case 5	6	6	7	3	3	3	5	4	6	7											86		11												
20	Case 6	5	5	5	4	4	4	5	4	6	6											73		12												
21	Case 7	6	6	7	4	4	4	3	4	6	7											90		13												
22	Case 8	2	7	7	4	3	3	5	5	7	6											94		14												
23	Case 9	5	7	5	4	3	4	5	3	7	6											77		15												
24	Case 10	6	6	6	4	3	3	3	5	7	6											91		16												
25	Case 11	5	7	5	4	3	5	4	4	2	3											89		17												

Figure 2: The APS regression worksheet.

The cells highlighted in yellow are output regions for the VBA program. Beginning in cells Y9 and Z9, for each subset size, the R^2 and Mallows C_p values are reported, respectively. Beginning in cell AA9, the selected variables are identified (1 if selected, 0 otherwise). For example, the best four-predictor subset consists of {x_1, x_3, x_5, x_{10}}. This four-predictor subset yields R^2 = .5223 and C_p = 6.2333. A plot of the number of predictors versus Mallows' C_p is shown in Figure 3. The general dominance measures of predictor importance begin in cell B9. The sum of these values is displayed in cell C8 and it is noted that this sum is equal to the R^2 value for the full (10-predictor) regression model (see cell Y18).

4.3. Model Selection

The SPSS stepwise regression solution consists of four predictors; however, the Mallows' C_p results from the APS regression analysis suggest that four is probably not the best number of predictors. The C_p index is minimized at q = 5 predictors. Moreover, the value of C_p = 5.7720 for q = 5 predictors is quite close to the number of parameters in the regression model (5 + 1 = 6), which also lends support for q = 5 predictors. The APS

results also reveal that, even if a four-predictor subset is desired, the stepwise regression subset $\{x_1, x_2, x_8, x_{10}\}$ is inferior to the APS subset $\{x_1, x_3, x_5, x_{10}\}$ with respect to explained variation. The stepwise regression subset yields $R^2 = .5176$, whereas the APS subset yields $R^2 = .5223$.

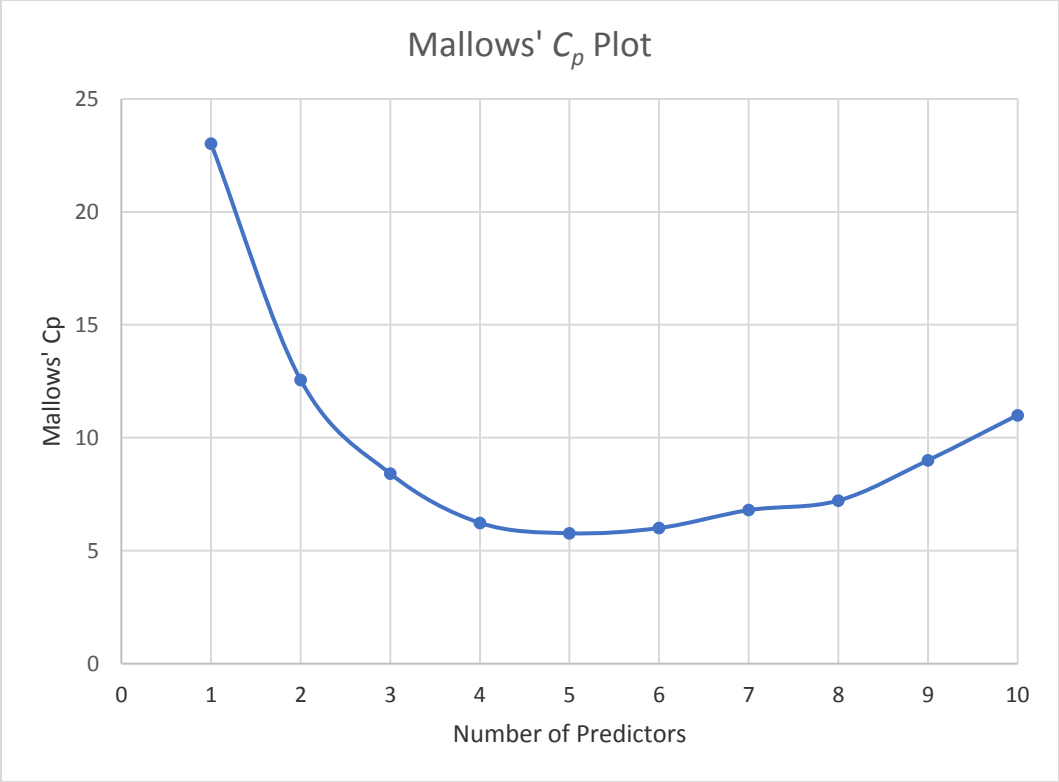


Figure 3: Plot of Mallows' C_p values vs. number of predictors.

To summarize, there are three salient findings from the model selection analysis. First, stepwise regression does not necessarily produce the maximum R^2 value for the number of predictors that it selects. Second, stepwise regression does not necessarily lead to the best choice for the number of predictors. Third, there is the potential for rather different subsets to yield very similar values of R^2. For example, the stepwise and APS results reveal that, augmenting $\{x_1, x_{10}\}$ with $\{x_3, x_5\}$ results in an R^2 value that is only better by .0047 than the R^2 achieved when augmenting with $\{x_2, x_8\}$. Accordingly, model selection results alone are not sufficient to establish which variables are more important than others.

4.4. Predictor Importance

Based on the absolute magnitude of the standardized regression coefficients from the full regression model in Table 2, the rank order of the predictors from greatest to least with respect to importance is x_{10}, x_5, x_2, x_1, x_3, x_6, x_8, x_9, x_7, x_4. This is a somewhat perplexing ranking, as it would seem highly unlikely that the lighting in the restaurant (x_5) would be the second most important driver of satisfaction. The food-quality items (x_2, x_1, x_3) are placed after lighting in the third through fifth positions. Another puzzling aspect of the predictor ranking is that the waiting time measures (x_8, x_9, x_7) occupy three of the last four positions in the ranking.

Contrastingly, based on the general dominance measures from the APS solution, the rank order of the predictors would be: x_2, x_1, x_3, x_{10}, x_7, x_8, x_5, x_9, x_6, x_4. This ranking is far more logical. The food-quality items occupy the first three positions in the ranking. The fourth position is occupied by the courteousness of the server measure (x_{10}). Three of the next four positions are occupied by the waiting time measures (x_7, x_8, x_9). Three of the last four positions are occupied by the aesthetic measures (x_5, x_6, x_4).

The key takeaways from the predictor importance analysis are twofold. First, it would be a mistake to rely solely on *p*-values or standardized regression coefficients from the full regression model to establish the relative importance of predictors. Second, APS regression provides a well-established measure for ranking predictors that, in this example, led to a far more interpretable ranking. The ranking obtained from the general dominance measures reflects the primacy of the food-quality items (x_1, x_2, x_3) and the secondary importance of the service-quality measures (courtesy measure x_{10} and promptness measures x_7, x_8, x_9). The aesthetic measures pertaining to parking (x_4), lighting (x_5), and noise level (x_6) are of lesser importance.

5. Experience with the Excel Workbook

I have used components of this Excel workbook in undergraduate marketing research courses, masters level courses on marketing analytics, and Ph.D. seminars for

quantitative methods in marketing. The computational aspects of the SWEEP algorithm are not discussed in the undergraduate and masters level courses. However, the efficient nature of this approach for evaluating all possible subsets can be discussed at the doctoral level. At all levels, however, there is considerable interest in the restaurant satisfaction example. Students are particularly intrigued by the fact that a popular software system (e.g., SPSS) can produce results that are both suboptimal and potentially misleading. The suboptimality is easily demonstrated by the fact that the four-predictor subset obtained via stepwise regression is not the four-predictor subset that maximizes R^2. The misleading nature of the regression coefficients and significance tests for the full regression model also leads to fruitful discussions. The concept of running all possible subsets of predictor variables can be explained with relative ease to the students, as can the large number of possible subsets when p is large.

Upon completion of the APS analysis, several limitations pertaining to the software program can be provided to the class. The extent to which these issues are addressed may differ between undergraduate and graduate courses. With respect to model selection, three limitations are particularly relevant. First, Mallows' C_p is not the only possible measure for choosing the number of variables. Other options include the adjusted R^2, AIC, and BIC. Second, it might be advantageous to generate the top three or four models for each subset size for comparison purposes. Third, the all-possible subsets algorithm becomes computationally infeasible for large p. Computation time and storage requirements roughly double as p is increased by one. The VBA macro runs almost instantaneously for the dining experience example where there are $p = 10$ predictors. However, it required 17.5 minutes for a customer satisfaction dataset with $p = 20$ predictors. Currently, the VBA macro is dimensioned for up to $n = 2500$ cases and $p = 20$ predictors. Branch-and-bound programming can be used for subset selection for $p \leq 50$ when using compiled code in Fortran or C. For interpreter code as in VBA, scalability is apt to be more problematic. Moreover, because branch-and-bound programming

circumvents the explicit enumeration of all possible subsets, it would not allow the predictor importance indices to be computed.

This exercise is also useful for demonstrating that examination of bivariate correlations or standardized regression coefficients is not an advisable method for measuring the relative importance of predictors. The general dominance measure is arguably one of the best alternatives. Nevertheless, students could be made aware that there are a variety of predictor importance measures that have been described in the literature and that no single measure can be judged the best for all circumstances [22].

References

[1] Aikake, H. (1973). Maximum likelihood identification of Gaussian autoregressive moving average models. *Biometrika*, 60, 255-265.

[2] Aikake, H. (1974). A new look at the statistical identification model. *IEEE Transactions on Automatic Control*, 19, 716-723.

[3] Breiman, L. (1995). Better subset regression using the nonnegative garrote. *Technometrics*, 37, 373-384.

[4] Brusco, M. J., & Steinley, D. (2011). Exact and approximate algorithms for variable selection in linear discriminant analysis. *Computational Statistics and Data Analysis*, 55, 123-131.

[5] Brusco, M. J., Steinley, D., & Cradit, J. D. (2009). An exact algorithm for finding hierarchically well-formulated subsets in second-order polynomial regression. *Technometrics*, 51, 306-315.

[6] Budescu, D. V. (1993). Dominance analysis: A new approach to the problem of relative importance of predictors in multiple regression. *Psychological Bulletin*, 114, 542-551.

[7] Chu, R. (2002). Stated-importance versus derived-importance customer satisfaction measurement. *Journal of Services Marketing, 16*, 285-301.

[8] Duarte Silva, A. P. (2001). Efficient variable screening for multivariate analysis. *Journal of Multivariate Analysis*, *76*, 35-62.

[9] Efroymson, M. A. (1960). Multiple regression analysis. In A. Ralston & H. S. Wilf (Eds.), *Mathematical methods for digital computers* (pp. 191-203). New York: Wiley.

[10] Furnival, G. M., & Wilson, R. W. (1974). Regression by leaps and bounds. *Technometrics*, 16, 499-512.

[11] Garside, M. J. (1971). Some computational procedures for the best subset problem. *Applied Statistics*, 20, 8-15.

[12] Goodnight, J. H. (1979). A tutorial on the SWEEP operator. *The American Statistician*, 33 (3), 149-158.

[13] Grisaffe, D. (1993). Appropriate use of regression in customer satisfaction research. *Quirk's Marketing Research Review*, (February). Available at http://www.quirks.com/articles/article_print.asp?arg_articleid=189 [Accessed 3 March 2006]

[14] Grömping, U. (2015). Variable importance in regression models. *WIREs Computational Statistics*, 7, 137-152.

[15] Hoffman, P. J.(1960). The paramorphic representation of clinical judgment. *Psychological Bulletin*, 57, 116–131.

[16] Gustafsson, A., & Johnson, M. D. (2004). Determining attribute importance in a service satisfaction model. *Journal of Service Research,* 7, 124-141.

[17] LaMotte, L. R., & Hocking, R. R. (1970). Computational efficiency in the selection of regression variables. *Technometrics*, 12, 83-93.

[18] Lindeman, R. H., Mirenda, P. F., Gold, R. Z. (1980). *Introduction to Bivariate and Multivariate Analysis*. Glenview, IL: Scott, Foresman.

[19] Mallows, C. L. (1973). Some comments on C_p. *Technometrics*, 15, 661-675.

[20] Martin, D. (2008). A spreadsheet tool for learning the multiple regression *F*-test, *T*-tests, and multicollinearity. *Journal of Statistics Education*, **16**, doi:10.1080/10691898.2008.11889573

[21] Miller, A. J. (2002). *Subset Selection in Regression* (2nd ed.). London: Chapman and Hall.

[22] Nimon, K. F., & Oswald, F. L. (2013). Understanding the results of multiple linear regression: Beyond standardized regression coefficients. *Organizational Research Methods*, 16, 650-674.

[23] Olejnik, S., Mills, J., & Keselman, H. (2000). Using Wherry's adjusted R^2 and Mallows' Cp for model selection from all possible regressions. *Journal of Experimental Education*, 68, 365-380.

[24] Pratt, J. W. (1987). Dividing the indivisible: Using simple symmetry to partition variance explained. In T. Pukkila & S. Putanen (Eds.), *Proceedings of the Second International Conference in Statistics* (pp. 245–260). Tampere, Finland: University of Tampere

[25] Ryan, M. J., Rayner, R., & Morrison, A. (1999). Diagnosing customer loyalty drivers: Partial least squares vs. regression. *Marketing Research,* 11, 19-26.

[26] Schwartz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, 6, 461-464.

[27] Smith, D. M. (1991). Algorithm AS 268, all possible subset regressions using the QR decomposition. *Applied Statistics*, 40, 502-513.

[28] Thomas D. R., Hughes, E., & Zumbo, B. D. (1998). On variable importance in linear regression. *Social Indicators Research*, 45, 253–275.

[29] Thomas D. R., Kwan, E., Zumbo, B. D. (2018). In defense of Pratt's variable importance axioms: A response to Grömping. *WIREs Computational Statistics* e1433. Available at https://doi.org/10.1002/wics.1433 [Accessed 5 June 2018].

[30] Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society B*, 58, 268-288.

[31] Wherry, R. J. (1931). A new formula for predicting the shrinkage of the coefficient of multiple correlation. *Annals of Mathematical Statistics*, 2, 440-451.