©EIJEAS 2016 Volume: 2 Issue: Special Issue, 60-81, Ohio, USA
Electronic International Journal of Education, Arts, and Science
http://www.eijeas.com

Stepwise Regression and All Possible Subsets Regression in Education

Ke Wang
Texas A&M University, kewang2010@tamu.edu,
Zhuo Chen
Texas A&M University, zhuoyue87@tamu.edu

Abstract

Stepwise methods are quite common to be reported in empirically based journal articles (Huberty, 1994). However, many researchers using stepwise methods failed to realize that software packages had been programmed in error. The purpose of this study is to introduce the procedure of stepwise regression and used experiments and Venn diagrams to illustrate the three main problems of stepwise regression: wrong degree of freedom, capitalization on sampling, and error R^2 not optimized. Meanwhile, the study also depicted an alternative method: All-possible-subsets regression and used an experiment to illustrate how to use it in real complex study. Finally, some matters needing attention when using both automatic procedures are discussed.

Keywords: Stepwise Regression, Degree of Freedom, Sampling Error, R^2, All-possible Subsets

Introduction

One of common problems in regression analysis is to make the decision in selecting variables. Researchers usually utilize or manipulate numbers of predictor variables in research and they are trying to create a best-fit model with a relatively smaller subset of predictors. During the process of building the model, researchers could try many combinations based on theories. However, it should be noteworthy that the entire process may take a considerable amount of time.

As one of commonly-used approaches in variable selection, the stepwise method is often reported in empirically based journal articles (Huberty, 1994). As a fact, 260 academic journal papers (peer reviewed) after 1994 can be found in EBSCOhost Online Research Databases by using the key words of "stepwise regression". Although the stepwise is extensively utilized in selecting variables, the step procedure in commonly-sued software packages is programmed with errors. In this case, in spite of the prevalence of stepwise methods in variable selection, the results or outcomes from the stepwise method may not be appropriate.

In the present paper, we firstly argue for the procedure of stepwise regression and three major problems in stepwise procedure with examples and Venn diagrams, then, illustrate the problem caused by collinearity in stepwise. In addition, we also introduce the all-possible-subsets analysis as an alternative of selecting variables and illustrate how to use all-possible-subsets analysis.

What is stepwise regression?

Stepwise regression aims to select a model step by step, adding or deleting one predictor only based on the statistical significance. The result of this process is a single regression model. Stepwise analysis has either forward or backward progression. The forward progression is more commonly encountered than backward analysis. Nowadays, researchers can control details of the process, including the significance level and variable manipulation (e.g., add or remove) in statistical software programs, such as Minitab and Statistical Software (Frost, 2012).

Taking forward stepwise regression as an example, firstly, the stepwise process computes all bivariate r^2 values for all independent variables and dependent variable. Then, it selects the independent variable with the largest r^2. At the second stage, the remaining independent variables will be added in the model separately with the “best” single

independent variable and the stepwise will select the independent variable that yield the largest increase in R^2. At the third stage, the stepwise regression will evaluate the contributions of remaining variables to the model with the best single variable and the second "biggest contribution" variable. Following the same rule, the variable, which can yield the largest increase in R^2 with the two selected predictors, can be chosen as the third predictor.

The forward selection procedure in the commonly-used statistical packages will stop when the increase in R^2 from one step versus the R^2 in the previous step is not statistically significant. It should be noteworthy that, over the course of selecting variables, stepwise process utilizes the equation ($F_{calculated}=[(R_{Larger}^2 - R_{Small}^2)/(k_L-k_S)]/[(1-R_{Larger}^2)/(n-k_L-1)]$) to test the null hypothesis H_0: $R_{Larger}^2=R_{Small}^2$. Specifically, k_L is the number of predictors used to obtain R_{Larger}^2, and k_S is the number of predictors used to obtain $R_{Smaller}^2$. The degrees of freedom are (k_L-k_S)for the numerator, and (n-k_L-1) for the denominator (Thompson, 2006). Therefore, the sample size plays an essential role in calculating $F_{calculated}$ value on the above null hypothesis. In other words, researchers may evaluate the same Delta R^2 with the same number of stepwise procedures but get different $p_{calculated}$ value based on different sample sizes (Thompson, 2006).

Why Stepwise Regression Do Not Work: Three Problems

Criticisms of the stepwise method has been given by increasing numbers of scholars (e.g., Huberty, 1989, 1995; Snyder, 1991, Thompson, 1989, 1995, 2001). The first major problem of the stepwise procedure is the fact that commonly-used statistical program packages choose the wrong degrees of freedom in their statistical tests, providing incorrect $F_{calculated}$ and $p_{calculated}$. The second major problem is that stepwise multiple regression packages tend to capitalize on small amount of sampling error, thus, offering the final model with different selected variables. The third major problem is that stepwise multiple regression packages cannot yield the best predictor set for a given size.

Wrong Degree of Freedom

Commonly-used statistical packages utilize incorrect degrees of freedom to calculate the mean square (MS), F value and p value in stepwise regression. According to Walker (1940), the number of degrees of freedom is equal to the number of observations minus the number of necessary relations obtaining among these observations. In other words, the number of degrees of freedom is equal to the number of original observations minus the number of parameters estimated from them (Walker, 1940).

Based on the definition of degree of freedom, df_{total} is n-1 (n is the number of sample size) in the stepwise regression analysis; $df_{regression}$ is the number of variables that stepwise has entered; and $df_{residual}$ equals df_{total} minus regression degrees of freedom (Thompson, 2006). Therefore, $df_{regression}$ is the number of predictor variables in a given study. However, stepwise regression has special procedures being mentioned in the first section to select variables into the model. In the selection process after the first step, all remaining variables are separately entered into the model and the predictor which yields the largest R^2 is selected. As Thompson (2006) noted, software should charge for 1 degree of freedom for every predictor "tasted", regardless of how many predictors are ultimately retained in the analysis.

Table 1 presents a heuristic example regarding the wrong degree of freedom. Presuming that there are 526 samples, 5 steps of forward stepwise with 50 predictor variables, and an R^2 of 10%. The computer packages use the incorrect degree of freedom 5, and $p_{calculated}$ is 0.022 (statistically significant). However, when the correct degree of freedom should be 50, the $p_{calculted}$ is 0.369 (not statistically significant). Thus, this problem easily causes Type I errors (Cliff, 1987).

Table 1. *Stepwise NHSST for Hypothetical Example*

Analysis/source	*SOS*	*df*	*MS*	$F_{calculated}$	$p_{calculated}$	R^2
Incorrect						
Regression	100.0	**5**	20.00	2.667	**0.0215**	10.00%
Residual	900.0	**520**	7.50			
Total	1000.0	525	1.90			
Correct						
Regression	100.0	**50**	2.00	1.06	**0.3686**	10.00%
Residual	900.0	**475**	1.89			
Total	1000.0	525	1.90			

Note: The original table (Thompson 2006, p. 273) was modified to this table

Capitalization on Sampling Error

Based on procedures of stepwise statistical packages, the packages do not recognize the mistakes caused by sampling errors. Unfortunately, statistical software will select the

predictor that contributes more to the increase of R^2 than other predictors in a given step, even if the difference is very small between two predictors and the difference is caused by the sampling error. Snyder (1991) presented a heuristic example of these dynamics on sampling error.

Any mistakes in the sequence will influence the following subsequent choices because stepwise regression is a linear sequence of selection based on the rules mentioned in the previous section and only one predictor is selected in each step. If one predictor enters into the model because of an infinitesimal advantage being caused by a small amount of sampling error, this may produce an incorrect regression model. For example, as Figure 1 shown, the common area of Y and X_1 is 80 (B+C=80) and is bigger than other X_2 and X_3 variables. Thus, X_1 is the first predictor in the model. However, if the difference of contribution to R^2 between X_1 and X_2 is caused by sampling error, namely, X_2 should be the first predictor in the model. Then, in stepwise regression, X_2 will be the second predictor because X_2 can contribute unique 39 to the overlap area with Y. But if the first predictor is X_2, the second predictor should be X_3 because X_3 can contribute 50 to R^2 and X_2 is 40.

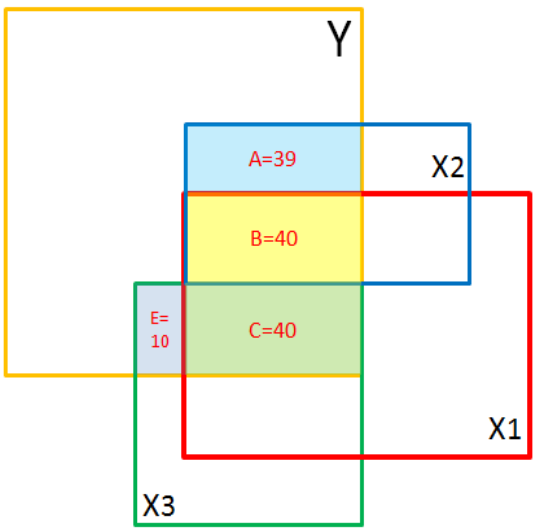

Figure 1. Venn diagram on sampling error.

Therefore, sampling error is another important issue in the process of stepwise regression. Less sampling error tends to be presented in data sets involving (a) larger samples, (b) fewer predictor variables, and (c) larger effect sizes, as reflected in the factors involved in most statistical corrections for positive bias in uncorrected variance-accounted-for effect sizes (Snynder & Lawson, 1993, Thompson, 1990).

R^2 Not Optimized

Some researchers, who did not enough understand working procedures of stepwise, may erroneously believe that the best predictor set of size two or more are included in the results of stepwise. However, the purpose of the stepwise method is to select a relatively smaller subset of predictors that may be almost as effective as the full set of predictors in yielding accurate Y_hat_i scores (Thompson, 2006). Most importantly, the stepwise method selects one predictor in each step, meanwhile, each step needs to consider the variables selected in the model. Nevertheless, the correct selection method should choose the best combination in a given set of predictors of size rather than identified a sequence of variables. In fact, the stepwise incremental selection standard is completely irrelevant in selecting the best combination. In other words, the best combination might not include the single best predictor, when the best single predictor is high multiply collinearity. Just as Figure 2 shown, stepwise regression will select X_1 as the first predictor because X_1 has the largest bivariate r^2 value with Y_i scores ($B+C+D$=80). In the second step, based on the model with X_1, stepwise regression will select X_4 as the second predictor because X_4 can contribute unique the area of 15 to multiple R^2. Then, given the set of predictors of size 2, the result from stepwise regression is the combination of X_1 and X_4. However, the best combination of size 2 in Figure 1 is X_2 and X_3 rather than X_1 and X_4. The stepwise procedure defines an a posteriori order based solely on the relative uniqueness of variables in the sample at hand (Cohen, et al., 2003, p. 161).

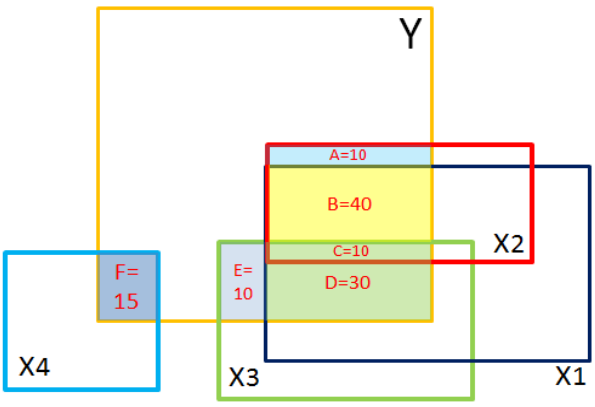

Figure 2. Venn diagram on optimized R^2.

In Figure 2, the X_1 predictor's explanatory power is presented in X_2 and X_3 predictors, namely, there is high collinearity among predictors. Cohen et al. (2003) stated, "when the *IV*s are substantially correlated with each other, the losers in the competition may not make a

sufficiently large unique contribution to be entered at any subsequent step before the problems is terminated by the absence of a variable making a statistically significant addition." (Cohen, et al., 2003, p.116) The sum of the common areas between Y & X_1 and Y& X_4 is 95 (80+15=95) which is smaller than the sum of the common area between Y & X_2 and Y & X_3 (10+30+10+40+10 =100).

Additionally, there is another possible reason that may cause "unoptimized" R^2 in a give subset. In the process of predictor selection, stepwise packages may move previously-entered variable out of the model in the next and all the remaining steps. Taking Figure 3 as an example, the area of A and E are 18 and 20 and X_1 is the first variable in the model. In the second step, X_3 will be selected and enters into the model with X_1, because X_3 contributes unique 20 to R^2. In the third step, when the computer package enters the X_2 into the model with X_1 and X_3, the program will delete the $X1$ variable in the model. In this situation, X_1's explanatory power is presented in X_2 and X_3, thus, the R-square of the model with X_1, X_2 and X_3 is the same to the model with X_2 and X_3. In other words, the model with two variables and the largest R^2 is X_2 and X_3. As Figure 3 shown, X_2 and X_3 contribute 118 to R^2, X_1 and X_2 contribute 98, and X_1 and X_3 contribute 100. Therefore, the model with X_2 and X_3 has a largest R^2 in the predictors of size 2. This supports the statement from Thompson (1995): "The true best set may have considerably higher effect sizes and may even include none of the variables selected by the stepwise algorithm."

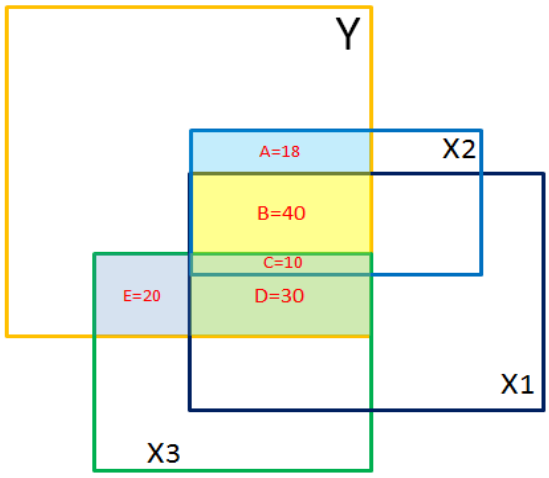

Figure 3. Venn diagram on three predictors.

Experiment on Collinearity

In fact, all used examples in the above present the problem of collinearity. For example, X_1, X_2, and X_3 have relatively high correlation coefficients in Figure 1. Here,

another heuristic example will be used to explain the problem of collinearity in stepwise regression. As to the situation of all predictors are completely uncorrelated, the results of stepwise selection are complete correct because the variables entry order is exactly correspond to the squared correlations of the predictors with *Y*. However, this case is nearly nonexistent in a real study.

The data is from Teaching and Learning International Survey (*TALIS*) on the United States. The present example only selects three independent variables (efficacy in classroom management (*ECM*), efficacy in instruction (*EINS*), efficacy in student engagement (*ESE*)), and one dependent variable (Teacher job satisfaction (*TJS*)).

Table 2. *Descriptive Statistic on Variables*

Variables	*TJS*	*ECM*	*EINS*	*ESE*	Mean	*SD*	Valid N (listwise)
TJS	1.00				12.28	2.01	1845
ECM	.18	1.00			13.06	1.94	1845
EINS	.16	.70	1.00		12.97	1.94	1845
ESE	.25	.64	.75	1.00	11.97	2.23	1845

Based on Table 2, the valid sample size is relatively large with 1845 participants; the difference of mean and *SD* between four variable are small; and the correlation coefficient between predictors are relatively high (from 0.64 to 0.75). This experiment provides some evidence on the statement that "Probably the most serious problems in the use of stepwise regression programs arises when a relatively large number of *IV*s is used."(Cohen, Cohen, West, & Aiken, 2003, p. 161)

First of all, the output from stepwise regression is that the final model only includes the *ESE* predictor, namely, *TJS*_hat = 9.630 +0.222**ESE*, the *p* value of *ESE* coefficient is 0.000, and R^2 is 6.1%. However, when all variables are entered into the model, the model equation with three variables is *TJS*_hat = 9.905 + 0.249**ESE* -0.101**EINS* + 0.069**ECM*. Surprisingly, the *p* values suggest that the model is good fit the data. Moreover, the R^2 (6.45%) is 0.35% more than the stepwise regression. Table 3 shows the detailed difference between two methods.

Table 3. *Results Comparison from Stepwise and Enter Regression*

Source	Stepwise	Enter
DV	*TJS*	*TJS*
IV	*ESE*	*ESE, EINS, ECM*
R^2	6.1%	6.4%
p value of IV	0.00	0.00(*ESE*), 0.01(*EINS*), 0.04(*ECM*)

Less sampling error tends to be presented in data sets involving: (a) larger samples, (b) fewer predictor variables, and (c) larger effect sizes, as reflected in the factors involved in most statistical corrections for positive bias in uncorrected variance-accounted-for effect sizes (Snynder & Lawson, 1993; Thompson, 1990). Thus, the use of stepwise methods in these circumstances may be somewhat less sinful. However, stepwise regression also made a serious mistake on this example. In order to explore the reasons for the difference between stepwise and enter regression on this example, commonality analysis was used to answer the questions: (a) how much explanatory power is unique to the first predictor (*ESE*) and other two predictors (*EINS* and *ESM*)? (b) how much explanatory power is common to both predictors and could be derived from either predictor? Knowing the explanatory power of three predictors can promote better understanding of the differences between the two methods on the example.

Table 4. *Coefficients Required and Unique and Commonality Components of Shared Variance*

Predictors	r^2 or R^2
ESE (1)	6.0527%
EINS (2)	2.4427%
ECM (3)	3.0668%
ESE, EINS (1,2)	6.2338%
ESE, ECM (1,3)	6.1070%
EINS, ECM (2,3)	3.2924%
ESE, EINS, ECM (1,2,3)	6.4465%
Three independent variables	

U1=R^2(123)- R^2(23)=6.4465%-3.2924%=3.1541%

U2= R^2(123)- R^2(13)=6.4465%-6.1070%=0.3395%

U3= R^2(123)- R^2(12)=6.4465%-6.2338%=0.2127%

U12= R^2(13)+ R^2(23)- R^2(3)- R^2(123)=6.1070%+3.2924%-3.0668%-6.4465%= -0.1139%

U13= R^2(12)+ R^2(23)- R^2(2)- R^2(123)=6.2338%+3.2924%-2.4427%-6.4465%=0.637%

U23= R^2(12)+ R^2(13)- R^2(1)- R^2(123)=6.2338%+6.1070%-6.0527%-6.4465%= -0.1584%

U123= R^2(123)+R^2(1)+R^2(2)+R^2(3)-R^2(12)-R^2(13)-R^2(23) =6.4465%+6.0527%+2.4427%+3.0668%-6.2338%-6.1070%-3.2924%=2.3755%

Table 4 shows the calculating procedures of commonality analysis. The results can easily be plugged into a spreadsheet program to produce the output reported in Table 5. Noting that the sum of unique (3.15%) and the common explanatory (2.90%) partitions of *ESE* in the model with R^2 (6.05%) is equal to the correlation coefficient of r^2 of ESE with TJS (0.246^2=0.0605). As Table 5 shown, there are seven non-overlapping partitions of R^2 and the sum of seven partitions is 6.45%. However, two of seven partitions are negative numbers. As area-world statistics, variances theoretically have minimum value of zero (*ESE* and *EINS*, *EINS* and *ECM*). Therefore, there are presumably suppressor effects in the model. By checking β_{EINS} (-0.098) and r_{EINS} (0.156) from the output of the final model (Table 6), *EINS* predictor is judged to be a suppressor variable.

Table 5. *Unique and Common Components of Shared Variance (R^2)*

Predictor Combination	Variable			
Statistic	*ESE*	*EINS*	*ECM*	Partition
ESE	3.1541%			3.1541%
EINS		0.3395%		0.3395%
ECM			0.2127%	0.2127%
ESE, EINS	-0.1139%	-0.1139%		-0.1139%
ESE, ECM	0.637%		0.637%	0.637%
EINS, ECM		-0.1584%	-0.1584%	-0.1584%
ESE, EINS, ECM	2.3755%	2.3755%	2.3755%	2.3755%
Unique	3.1541%	0.3395%	0.2127%	

Common	2.8986%	2.1032%	2.8541%	
Total	6.0527%	2.4427%	3.0668%	6.4465%

In addition, based on the outputs of linear regression, the *p* values of *EINS* in the model with *ESE* and *EINS* is 0.059 and *p* value of *ECM* in the model with *ESE* and *ECM* is 0.302. Therefore, the stepwise regression will exclude two variables based on the rules of stepwise. However, when the two variable of *ECM* and *EINS* are entered into the model with *ESE*, the *p* values of three predictors are statistical significant just because of the effect from the suppressor variable *EINS*.

Table 6. *The Output Table on The Final Model with Enter Method*

Model	*B*	*S.E.*	*Beta*	*S.E.*	*p* value	*r*	r_s
Constant	9.705046	.332622			.000000		
ESE	.249228	.031635	.276724	.035125	.000000	.246023	.968976
EINS	-.100719	.038964	-.097573	.037747	.009816	.156292	.615565
ECM	.068762	.033608	.066438	.032472	.040896	.175122	.689727

On the other hand, three predictors' structure coefficients are separately 0.99, 0.62, and 0.69. Namely, they are significant with high relations with *Y*_hat, even if they have high collinearity. In a word, the stepwise packages make a serious mistake on selecting predictors when variables are collinearity. As this example shown, the stepwise regression equation is *TJS*_hat=9.630+0.222**ESE*, but the real regression equation is *TJS*_hat = 9.905 + 0.249**ESE* -0.101**EINS* + 0.069**ECM*.

Consequently, this example provides the evidence that the collinearity between predictor variables certainly affects the order of entry or deletion of variable, and illustrates possible mistakes in predictor selection in a collinear situation. Meanwhile, this example also illustrates that an independent variable (EINS or ECM) with a high Pearson *r* with the first-entered variables (ESE) may never be entered into the model, even if this variable is the second-best single variable in the independent variable set. In order to have a better understanding of this statement, a simple Venn diagram example is illustrated. Based on Figure 2, the stepwise will select predictor X_1 in the model. Then, the computer package selects X_4 variable. If the remaining variables can contribute the area more than 10, the

"important variable" X_2 and X_3 never be entered into the model because of the high collinearity among X_1, X_2, and X_3, even if the variable of X_2 and X_3 are the best predictors in the real model.

All-possible-subsets Analysis

Huberty (1989) noted, "A user of stepwise analysis may be led to believe that the first q variable entered into the analysis would constitute a good subset (of size q) of the initial set of p variables, or even the best subset of size q." However, the correct method should begin by computing the R^2 for every combination of predictors for one and more variable set size.

What is all-possible-subset regression?

All-possible-subsets regression runs all possible models using a different set of predictors, and displays models that contain one predictor, two predictors, and so on. The output is a number of models and their summary statistics. In addition, the result does not show the best model among the all models.

Most importantly, all-possible-subsets regression goes beyond stepwise regression, testing all possible subsets of the set of predictors. For example, the number of predictor variables is n, then, this program runs 2^n models with all subsets of predictors. For example, n=10, the results show 1024 models with their summary statistics. The entire analysis may sound tedious, but can be done rapidly, accurately, and painlessly by readily available computer software (Thompson, 1991). On the contrary, stepwise regression superficially works reasonably well as an automatic variable selection method, but the result is not guaranteed. Sometimes stepwise enters into a wrong turn and gets a suboptimal model (e.g., the above experiment).

When using an all-possible-regressions procedure, researchers need to select the best model and rank the models. Generally, researchers can draw the line plot of successive R^2 values (the plots of R-squared versus the number of variables) to find which number of variables is diminishing returns. "This plot can inform the researchers' subjective judgment regarding the optimal number of predictors to retain." (Thompson, 2006, p. 277) In addition, researchers can search whether there are one or two models standing out in the output, whether there are some almost-equally-good models. This also can give some hints on finding the optimum model.

However, it is a real danger of letting automated data-mining packages help researchers select a model without theoretical guidance and practical experience. Therefore, a better approach is to select the predictor set based on theory or previous empirical results, or based on the accessibility of the variables in a given set (Thompson, 1991).

Experiment on All-possible-subsets Analysis

In this section, a heuristic example is used to illustrate how to use all-possible-subsets regression to analyze the data. The data has eight variables. Dependent variable is work satisfaction scores (*WSQ*), and independent variables are age, male, married, income, length of service in the current organization in years (length), total work experience in years (experience), total number of job changes (changes), and score on love of money scale (*LOMS*). The sample size is 375. SPSS software (syntax are shown in the appendix) is utilized to run all-possible analysis. All statistical results are recorded into an Excel spreadsheet. Figure 4 was drawn and it presents a line plot of the maximum R^2 values for a given number of *n* predictors. As Figure 4 shown, the plot seems to level off after the predictor variables set size is 5 (R^2 = 85.5% with n=5, versus R^2 =86.2% with n=6). Thus, effect size of 85.5% can be deemed sufficient, based on the theories or previous empirical studies. Then, the focus turns to which five predictors should be retained for future use, namely, age, male, income, change, and *LOWS*. Moreover, through checking the others n=5 and 6 variable set with a slightly less-than-optimal effect size might be preferred for practical reasons, the purpose is to make sure the rationality of the model. The R^2 of model with age, male, income, experience, and *LOMS* variables is 85.3%; the R^2 of model with age, male, married, income, change, and *LOMS* variables is 85.7%; the R^2 of model with age, male, income, experience, length, and *LOMS* variables is 85.5%, the R^2 of model with age, male, income, experience, change, and *LOMS* variables is 86.2%. We found that the variables of male, income, and *LOMS* are included in above five models. The other age, experience, change, married, length variables are distributed in the other four models. Length can be deleted and experience can be added based on theories on job satisfaction. Therefore, the final model with age, male, income, experience, change, and *LOMS* was determined. As Thompson (2006) claimed, "A variant on this all-possible-subsets analysis would instead plot "adjusted R^2" value. This alternative will be most appealing when sample size is relatively small because sampling error variance is likely to be larger." In addition, Derksen & Keselman (1992) stated, "educational and psychological researchers who use automated subset selection procedures should be aware of operation characteristics of these procedures."

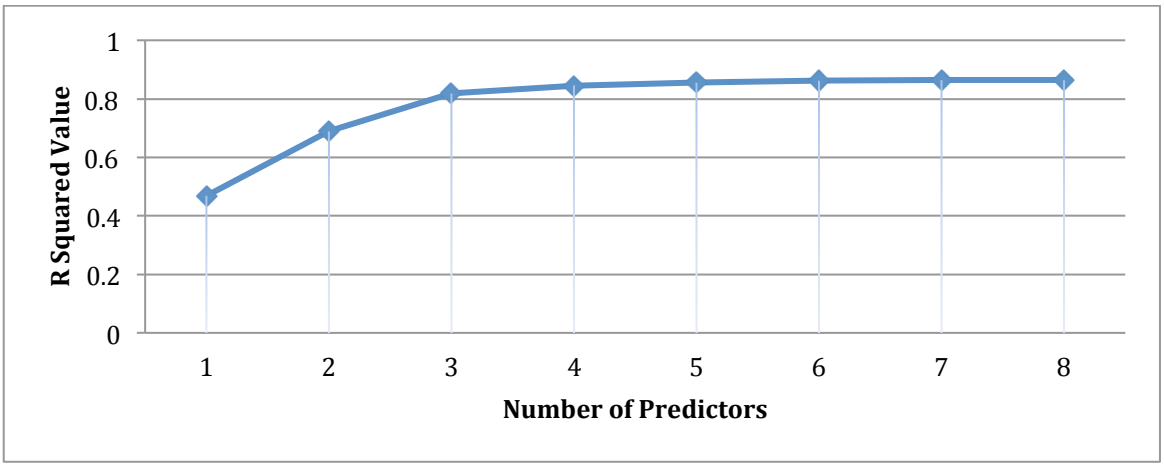

Figure 4. Line plot of successive R square values.

Conclusion

Although many researchers (e.g., Huberty, 1989; Snyder, 1991; Thompson, 1985, 1989, 2001) suggested stepwise method cannot be employed in psychological and behavioral research, many academic papers using stepwise method are being published. The issue is that only researchers being aware of main problems on stepwise method can correctly use stepwise method as a tool to select variables in the process of building models, while many researchers who did not realize the weakness of stepwise are using stepwise in their studies. The main problems of stepwise regression include (1) statistical software used the wrong degrees of freedom to compute *MS*, (2) stepwise method is very sensitive to sampling error and yield non-replicable results, and (3) stepwise cannot identify the best subset of predictors. These possible problems give rise to unreliable results from stepwise. Meanwhile, researchers need to realize that the number of variables, sample size, and multicollinearity in the data before running stepwise regression.

It is noteworthy that stepwise regression could be a valuable tool in the early stages of building a model. However, when the procedure terminates, researchers should check the order in which variables were added and deleted, confirm whether or not the variables that were included or excluded are reasonable, and judge the model with a largest R^2 based on a theoretical perspective. According to Huberty (1989), the order of variables entered into stepwise should not be used to assess relative variable contribution or importance. Likewise, as Cliff (1987, p. 187) stated, "in a sense all the variable are in the equation, even though some of them have (effectively) been given zero weights."

Although there is another alternative method: all-possible-subsets regression. It can assess all possible models and displays all subset results long, research reality is complex and an automated algorithm cannot solve all problems for researchers. Researchers can run both of stepwise and all-possible-subsets regressions to get the additional information that they need. The model that stepwise regression selected and the model with a largest R^2 that all-possible-subset regression showed may not be the best from a practical and theoretical perspective. Therefore, researchers need to use their professional subject area knowledge and common sense to analyze the output of all subset possible regression. As Kerlinger (1986, P. 454) stated, "The research problem and the theory behind the problem (and not stepwise methods) should determine the order of entry of variables in multiple regression analysis." In a word, don't just blindly accept the computer's choice!

Reference

Cliff, N. (1987). *Analyzing multivariate data*. San Diego: Harcourt Brace Jovanovich.

Cohen, J., Cohen, P., West, S. G., & Alken, L. S. (2003). *Applied multiple regression/correlation analysis in the behavioral science* (Third Edition). Mahwas, NJL Erlbaum.

Derksen, S., & Keselman, H. J. (1992). Backward, forward and stepwise automated subset selection algorithms: Frequency of obtaining authentic and noise variables. *British Journal of Mathematical and Statistical Psychology*, *45*(2), 265-282.

Frost, J. (2012). *Regression smackdown: Stepwise versus best subset*. Retrieved from http://blog.minitab.com/blog/adventures-in-statistics/regression-smackdown-stepwise-versus-best-subsets.

Huberty, C. J. (1989). Problems with stepwise methods—Better alternatives. In B. Thompson (Ed.), *Advances in Social Science Methodology* (Vol.1, pp. 43-70). Greenwich, CT: JAI Press.

Huberty, C. J. (1994). *Applied discriminant analysis*. New York: Wiley and Sons.

Kerlinger, F. N. (1986). *Foundations of behavioral research* (Third Edition). New York: Holt, Rinehart and Winston.

Snyder, P. (1991). Three reasons why stepwise regression methods should not be used by researcher. In B. Thompson (Ed.), *Advances in educational research: Substantive findings, methodological developments* (Vol. 1, pp. 99-105). Greenwich, CT: JAI Press.

Thompson, B. (1989). Editorial: Why won't stepwise methods die? *Measurement and Evaluation in Counseling and Development, 21*, 146-148.

Thompson, B. (1995). Stepwise regression and stepwise discriminant analysis need not apply here: A guidelines editorial. *Educational and Psychological Measurement, 55*, 525-534.

Thompson, B. (2001). Significance, effect sizes, stepwise methods, and other issues: Strong arguments move the field. *Journal of Experimental Education, 70*(1), 80-93.

Thompson, B. (2006). *Foundations of behavioral statistics: An insight-based approach*. New York: Guilford Press.

Thompson, B., Smith, Q. W., Miller, L. M., & Thomson, W. A. (1991). *Stepwise methods lead to bad interpretations: Better alternatives*. Paper presented at the annual meeting

of the Southwest Educational Research Association, San Antonio, TX. (ERIC Document Reproduction Service No. ED 327 573)

Walker, H. W. (1940). Degrees of freedom. Journal of Educational Psychology, 31(4), 253-269.

Welge, P. (1990). *The reasons why stepwise regression methods should not be used by researchers*. Paper presented at the annual meeting of the Southwest Educational Research Association, Austin, TX. (ERIC Document Reproduction Service No. ED 316 583)

Appendix

Note: This syntax is from Dr. Bruce Thompson.

```
* INSTRUCTIONS: First, make sure you have a subdirectory/folder
on your "C:" drive named "C:\TEMP", and if you don't then create
one.
* Second, search for "???" within this syntax file and then make
all the noted changes in order to analyze your own data.

*(Q) How can I do an all-subsets regression using SPSS?
    Whereas a stepwise regression yields one final equation,
    the goal of all-subsets regression is to perform all possible
    regressions combination of and then let the user (rather
    than the stepwise regression) choose the "best" equation.
*    So, if one had 5 independent variables, the all-subsets
    regression would perform 5 regressions of each predictor
    on y, and then work up towards one final regression with
    all the predictors.  The output can be any number of things,
    such as the r^2 for each equation, but I would rather use the
    adjusted predicted variables that SPSS can already create.

* (A) by rlevesque@videotron.ca 2001/08/30;
SPSS Dedicated web site:
http://pages.infinit.net/rlevesqu/index.htm.

SET MPRINT=no.

*//////////////////////////////.
DEFINE !combine (n=!TOKENS(1)
    /m=!TOKENS(1)
    /dep=!TOKENS(1)
    /indepv=!CMDEND).
```

```
/* Find all combinations on n items out of m */
/* August 30,2001 rlevesque@videotron.ca */

!DO !thisn=1 !TO !n
NEW FILE.
INPUT PROGRAM.
LOOP i=1 TO !thisn.
END CASE.
END LOOP.
END FILE.
END INPUT PROGRAM.
LIST.
!LET !list=!NULL
!DO !cnt=1 !TO !thisn
    !LET !list=!CONCAT(!list," ","j",!cnt)
!DOEND
COMPUTE n=!thisn.
* Calculate variable names for LOOP of the next WRITE command *.
STRING cntname cntbeg(A8).
COMPUTE cntname=CONCAT('j',LTRIM(STRING(i,F8.0))).
* Calculate first parameter for the LOOP of the next WRITE
  command *.
DO IF i=1.
COMPUTE cntbeg="1".
ELSE.
COMPUTE cntbeg=CONCAT('j',LTRIM(STRING(i-1,F8.0))," + 1").
END IF.
* Calculate second parameter for the LOOP of the next WRITE
  command *.
COMPUTE k=!m - !thisn + i.
FORMATS i k n(F8.0).
```

```
STRING quote(A1) strlist(A255).
COMPUTE quote="".
COMPUTE strlist=!QUOTE(!list).
* Write the syntax file which will store all the combinations in
the list.txt file*.
WRITE OUTFILE "c:\temp\macro.sps"
    /"LOOP "cntname"="cntbeg" TO "k".".
DO IF i=!thisn.
+    WRITE OUTFILE "c:\temp\macro.sps"
    /"WRITE OUTFILE "quote"c:\temp\list.txt"quote "/" strlist "."
+    LOOP cnt=1 TO !thisn.
+        WRITE OUTFILE "c:\temp\macro.sps" /"END LOOP.".
+    END LOOP.
+    WRITE OUTFILE "c:\temp\macro.sps" /"EXECUTE.".
END IF.
EXECUTE.
INCLUDE FILE="c:\temp\macro.sps".

/* Convert data from list.txt to the corresponding sav file */.
DATA LIST FILE='c:\temp\list.txt' LIST /!list.
SAVE OUTFILE=!QUOTE(!CONCAT('c:\temp\list',!thisn,'.sav')).
!DOEND
/* Combine all the sav files */.
GET FILE='c:\temp\list1.sav'.
!DO !nb=2 !TO !n.
ADD FILES FILE=*
    /FILE=!QUOTE(!CONCAT('c:\temp\list',!nb,'.sav.')).
!DOEND
/* Eliminate duplicates  */.
SORT CASES BY ALL.
MATCH FILES FILE=* /BY=ALL /FIRST=first.
SELECT IF first.
```

```
SAVE OUTFILE='c:\temp\all_comb.sav'.
/* Find name of last variables */
!DO !var !IN (!indepv)
!LET !lastone=!var
!DOEND
VECTOR vnames(!m A8).
!LET !cnt=!BLANK(1)
/* Create variables containing the names of the indep variables */
!DO !var !IN (!indepv)
COMPUTE vnames(!LEN(!cnt))=!QUOTE(!var).
!LET !cnt=!CONCAT(!cnt,!BLANK(1))
!DOEND
/* Construct the string containing the list of indep var of each regression */.
STRING dep (A8) indepv(A255).
COMPUTE dep=!QUOTE(!dep).
VECTOR j=j1 TO !CONCAT('j',!n) /ind=vnames1 TO
!CONCAT('vnames',!m).
COMPUTE nvar=NVALID(j1 TO !CONCAT('j',!n)).
LOOP cnt=1 TO nvar.
COMPUTE indepv=CONCAT(RTRIM(indepv)," ",vnames(j(cnt))).
END LOOP.
* Write the syntax file which will run all regressions */.
WRITE OUTFILE='c:\temp\syntax.sps'
    /"!regres dep=" dep "indepv=" indepv ".".
EXECUTE.
!ENDDEFINE.
*//////////////////////////.
DEFINE !regres (dep=!TOKENS(1) /indepv=!CMDEND)
REGRESSION
  /MISSING LISTWISE
  /STATISTICS COEFF OUTS R ANOVA
  /CRITERIA=PIN(.05) POUT(.10)
```

```
  /NOORIGIN
  /DEPENDENT !dep
  /METHOD=ENTER !indepv .
*discriminant
  groups = !dep (1,2) / variables = !indepv /
  analysis = !indepv / method = direct / statistics = table crossvalid .
!ENDDEFINE.
*//////////////////////////.
*************************.
* EXAMPLE OF USE.
*************************.
*???
     For regression, all possible subsets, change "n=" to
the number of predictors.
* Change "m=" to the number of predictors; change "dep="
to the name of the outcome variable within your dataset;
change "indepv=" to the names of the predictor variables
in your dataset.
SET MPRINT=yes.
****** Run the following macro to do the preparatory work.
!combine n=8 m=8 dep=WSQ indepv=AGE Male Married INCOME LENGTH EXP
CHANGE LOMS.
execute.
*???
* NOTE: The command below presumes that you put the data on
a USB stick, and that the USB drive on your computer is "E:\"
but it instead might be "F:\" or "G:\".
  GET
  FILE='C:\Users\kewang\Downloads\LOMS.sav'.
DATASET NAME DataSet1 WINDOW=FRONT .
INCLUDE FILE='c:\temp\syntax.sps'.
execute.
```