AN ABSTRACT OF THE THESIS OF

Xiaotong Li for the Master Degree

in Mathematics presented on 11/15/2013

Title:

Model Selection Techniques for Multiple Linear Regression Models

Abstract approved: ________________________________
(Thesis Advisor Signature)

Statistics is an important tool for researchers in almost every field that impacts modern life. Multiple linear regression analysis is one of the most important tools available to these researchers. A difficult, but frequently encountered problem in multiple regression analysis, is model selection. Classical model selection techniques included forward selection, backward elimination, and stepwise regression. Many new techniques have become available with the tremendous advances that have been made in computational power. These techniques include Mallow's Cp, Akaike's Information Criterion (AIC), Sawa's Bayesian Criterion (BIC), Schwartz' Bayesian Criterion (SBC) and many others.

This study focused on the Akaike's Information Criterion, Sawa's Bayesian Criterion and Schwartz' Bayesian Criterion. A simulation of several situations was conducted to try to answer two important questions. First, how good are these techniques? Second, are there any characteristics the researcher can use to determine which technique to use? The results indicated that there are some situations where the answers to these questions are clear cut but in other situations the results are somewhat unpredictable.

Keywords: Multiple Linear Regression, Model Selection Techniques, Akaike's Information Criterion, Sawa's Bayesian Information Criterion, Schwarz' Bayesian Criterion, Simulation.

MODEL SELECTION TECHNIQUES

FOR MULTIPLE LINEAR REGRESSION MODELS

A Thesis

Presented to

The Department of Mathematics and Economics

EMPORIA STATE UNIVERSITY

In Partial Fulfillment

of the Requirements for the Degree

Master of Science

----

by

Xiaotong Li

November 2013

Approved by the Department Chair

Committee Chair

Committee Member

Committee Member

Dean of the Graduate School and Distance Education

ACKNOWLEDGMENTS

I would like to extend my deepest gratitude to my advisor, Dr. Larry Scott, for his patience, encouragement and expert advice. He spent lots of his time helping me revise and improve my thesis.

I also would like to thank my committee members: Dr. Brian Hollenbeck, and Dr. Qiang Shi, for developing my thesis and giving helpful advice.

My thanks also go to my parents and my friends. They supported me to finish the thesis. I hope they receive my gratitude for their support and encouragement.

PREFACE

My thesis contains seven chapters. The first chapter will introduce regression analysis and give the definition of multiple linear regression. In Chapter 2, I introduce Root Mean Square Error (RMSE), Adjusted Coefficient of Determination, Mallows' C_p, Forward Selection, Backward Elimination, Stepwise Regression, Akaike's Information Criterion (AIC), Sawa's Bayesian Information Criterion (BIC) and Schwarz' Bayesian Criterion (SBC). Chapter 3 shows the development of the modern selection techniques such as Maximum R^2 Improvement (MAXR), Minimum R^2 Improvement (MINR), The Corrected Akaike Information Criterion (AIC_c) and Deviance Information Criterion (DIC). Chapter 4 shows that multicollinearity. Chapters 5, 6 and 7 will focus on simulation of several situations and draw the conclusion.

TABLE OF CONTENTS

Page

LIST OF FIGURES

LIST OF TABLES

Page

1. Introduction

The concept of “Regression” was created by Francis Galton in the nineteenth century to study human genetic problems. His work indicated that the height of offspring does not tend toward the size of parents, but rather towards the mean as compared to their parents. It is called “Regression toward the mean” in Statistics. Even Though Mr. Galton just analyzed the biological problem and created the concept of “”Regression”, the next work was extended and given a more general statistical context (Myers, 1989).

Regression analysis (Kutner etc., 2004) is a common and widely used statistical technique for estimating the relationships between a response (dependent) variable and one or more predictor (independent) variables. It has been successfully used to predict and analyze data in industrial production, agricultural production and scientific experiments including biology, economics, engineering, geology, medicine and almost every field. In addition to prediction, regression analysis can also be used for variable screening, model specification and parameters estimation.

There are three general kinds of regression models, simple linear regression, multiple linear regression and nonlinear regression. Compared to simple linear regression which focuses on the relationship between one dependent variable and one independent variable, multiple linear regression attempts to express the relationship between two or more predictor variables and a response variable as a linear equation.

We can use a multiple linear regression model as a mathematical model tool to describe the relationship among variables. Its aims are to provide a more scientific and sophisticated data analysis for the research of uncertain phenomenon and tries to predict and control the related random variable and to find the statistical regularity among variables. Multiple linear regression is an important tool used in modern statistical analysis.

Multiple linear regression is a linear function with the general form of

$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p + \varepsilon$,

where y is a dependent variable, $x_1, x_2, \dots, x_p$are independent variables, $\beta_1, \beta_2, \dots, \beta_p$ are unknown constants which are called partial regression coefficients and ε is called error term. In order to test hypotheses about the parameters of this model, it is necessary to assume that there is at least one predictor variable in this model, and ε is normally distributed with $E(\varepsilon) = 0$ and $Var(\varepsilon) = \sigma^2$. The variance is assumed to be constant for all values of the predictor variables. .

For a set of given data, there are many regression models that can be selected. We always prefer to get the "best" model which will predict the respond variable well to show the relationships between predictor variables and response variable and explain the data in the simplest way. On one side, we know that the more variables in the model, the better will the model fit the data. Many researchers have a tendency to overspecify the model (Freund etc., 2006). On the other side, we don't want to underfit but simple models tend to predict data better because the unnecessary predictor variables will add noise to the model and complicate the relationship among variables in the model. Applied in the specific case, adding unnecessary predictor variables will affect the accuracy of estimation and prediction.

Since multiple linear regression plays a very important role in the modern life, model selection, which is also called variable selection becomes a hot topic in statistics

(Myers, 1989). This paper will review several criteria that can be used to help choose the “best” model. These criteria include:

1. Root Mean Square Error (RMSE)

2. Adjusted Coefficient of Determination

3. Mallows’ C_p

4. Forward Selection

5. Backward Elimination

6. Stepwise Regression

7. Akaike’s Information Criterion (AIC)

8. Sawa’s Bayesian Information Criterion (BIC)

9. Schwarz’ Bayesian Criterion (SBC)

This thesis uses simulations to investigate the effectiveness of some of the above model selection techniques. Some researchers, including Sawa (1978), Mallows (1973), have suggested some situations where specific techniques are thought to work well. These situations will be investigated in this thesis. The paper summarizes long run comparisons of the model selected by each technique to the true underlying model. This paper will use the Statistical Analysis System (SAS) to simulate the data and perform multiple linear regression analysis.

2. Some Techniques for Model Selection

In order to introduce the criteria method, we assume that there are n observations, k predictor variables and p estimated parameters in the multiple linear regression model. Multiple linear regression model is defined as

$$Y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_k x_{ki} + \varepsilon_i \,, \ \ (i = 1, 2, 3, \dots, n; \ n \geq k + 1)$$

and where ε_i is a model error;

The residual sum of square (RSS) is shown as

$$RSS = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \qquad (2.1)$$

$where\ \hat{y}_i\ is\ the\ predicted\ value\ of\ y\ for\ given\ x\ at\ the\ ith\ data\ point$

$and\ \hat{y}_i = b_0 + b_1 x_{1i} + b_2 x_{2i} + \cdots + b_k x_{ki},$

$b_0, b_1, b_2, \ldots, b_k\ are\ the\ estimators\ of\ \beta_0, \beta_1, \beta_2, \ldots, \beta_k;$

The total sum of squares (TSS) is shown as

$$TSS = \sum_{i=1}^{n} (y_i - \bar{y}_i)^2, \qquad (2.2)$$

$where\ \bar{y}_i\ is\ the\ mean\ of\ the\ observed\ values\ y_i$

2.1 Criteria for All Possible Subset Models

Standard criteria are based on calculating for all possible subset models and choose the "best" linear model to fit the data. For any set of k predictor variables, there are 2^k-1 models that can be constructed. In most cases we exclude the null model.

2.1.1 Root Mean Square Error (RMSE)

The root mean square error, which is also called the root mean square deviation (RMSD) is defined as

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{n-p}} \qquad (From\ 2.1)$$

$$= \sqrt{\frac{RSS}{n-p}} \qquad (2.3)$$

It is a quick method for model selection criterion. It is an unbiased estimator. From the formula, we know that if we remove a predictor variable, the value of both RSS and (n-p) will increase. So the value of RMSE may increase or decrease. If we add a predictor variable to this model, the value of RMSE may or may not be reduced. We would generally look at all models with small RMSE. Knowledge of the variables can sometimes help the researcher select a good model.

2.1.2 Adjusted Coefficient of Determination

The adjusted coefficient of determination is defined as

$$R_{adj}^2 = 1 - \frac{\frac{RSS}{n-p}}{\frac{TSS}{n-1}}$$

$$= 1 - \frac{RSS}{n-p} \times \frac{n-1}{TSS}$$

$$= 1 - \frac{(n-1)MSE}{TSS}$$

since $MSE = \frac{RSS}{n-p}$ $\quad (From\ 2.3)$

First of all, we need to know the coefficient of determination R^2.

R^2 is defined as the form of

$$R^2 = \frac{\sum_{i=1}^{n}(\hat{y}_i - \bar{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y}_i)^2}$$

Since

$$\sum_{i=1}^{n}(y_i - \bar{y}_i)^2 = \sum_{i=1}^{n}(\hat{y}_i - \bar{y}_i)^2 + \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 ,$$

then

$$\sum_{i=1}^{n}(\hat{y}_i - \bar{y}_i)^2 = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 - \sum_{i=1}^{n}(y_i - \bar{y}_i)^2 , \qquad (2.4)$$

$$R^2 = \frac{\sum_{i=1}^{n}(\hat{y}_i - \bar{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y}_i)^2}$$

$$= \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 - \sum_{i=1}^{n}(y_i - \bar{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y}_i)^2} \qquad (From\ 2.4)$$

$$= 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y}_i)^2} \qquad (From\ 2.1\ and\ 2.2)$$

$$= 1 - \frac{RSS}{TSS}\ .$$

It is easy to show that $0 \leq R^2 \leq 1$ from the definition. Since R^2 does not take into account the number of parameters p, so the maximum R^2 will occur when all predictor variables are the regression model. There may be a very small change in R^2 when adding variables to the model. The value of R^2 will increase even if a variable is totally unrelated to the response variables. In order to avoid the problem we use adjusted coefficient of determination in multiple linear regression to do model selection.

Since we use the adjusted coefficient of determination to compare which candidate model is good, that means the data set is fixed. So we can know that both TSS and n is fixed. By the formula of R_{adj}^2 , R_{adj}^2 increases if and only if MSE decreases. Then R_{adj}^2 will increase when p decreases. We would select the model with the maximum R_{adj}^2as the choice for the best model provided that this model makes sense.

2.1.3 Maximum R^2 Improvement (MAXR)

Maximum R^2 Improvement (MAXR) (SAS, 1989) is a method which is based on the value of R^2 and looks for the “best” one-variable model, the “best” two-variable model and so forth. MAXR starts by finding the one-variable model which produced the maximum R^2 and then adding another variable, the one that would yield the greatest increase in R^2. When the two-variable model is obtained, MAXR determines if the removal of one variable and replacement of another variable would increase R^2. This is done by comparing each of the variables in the model and each variable not in the model. After that, MAXR helps us pick up the model which produces the largest increase in R^2.

2.1.4 Minimum R^2 Improvement (MINR)

The Minimum R^2 Improvement (MINR) (SAS, 1989) is similar to the MAXR technique. MINR picks the model which produces the smallest increase in R^2. For the given number of variables in the model, MAXR and MINR always show the same result for selecting the "best" model. Since MINR prefers the smallest increase in R^2, MINR has more steps to pick the best model by a given size.

2.1.5 Mallows' Cp Criterion

The Mallows' C_p criterion (Mallows, 1973) is defined as

$$C_p = \frac{RSS}{MSE} + 2p - n$$

We calculate values of C_p for all of the possible models and compare these values and choose the smallest. C_p is an unbiased estimator.

If we consider a multiple regression model, which containing all p-1 predictor variables, then we can get

$$\begin{aligned} C_p &= \frac{RSS}{MSE} + 2p - n \\ &= \frac{RSS}{\frac{RSS}{n-p}} + 2p - n \qquad (From\ 2.3) \\ &= n - p + 2p - n \\ &= p \end{aligned}$$

We prefer to choose the smaller and the value of C_p, which is close to value p as the best model.

2.2 Classical Model Selection Techniques

The above standard criteria are easy to implement when there are a small number of predictor variables in the multiple linear regression models but it is infeasible to use them for a large number of independent variables in the study. In some cases it is desirable to use automatic search methods that are based on computer algorithms to help us select the best model. These methods focus on adding or dropping one or more predictor variables from the model and compare the resulting regression models.

I will introduce some basic ideas about the procedure of these criteria and show how to use it with SAS program in the next part.

2.2.1 Forward Selection

In Forward selection (Weisberg, 1947) procedure, variables are added at each step. It contains only a term in the initial model and tests whether we should add a variable to the model.

- Begin with the simple regression model that has a single predictor variable. We select the predictor variable that has the highest correlation with the response variable.
- Forward selection then computes separate F-ratios for each variable that is not already in the model. The predictor variable with the smallest p-value is the second variable added to the model provided that its p-value is smaller than α.
- After that, we add another variable at each stage until no variable produces a significant p-value. We usually chose a rather large value for α. Typical values range from 0.10 to 0.25.
- This process is terminated when no variable meets the chosen level of significance.

- Traditionally, forward selection was a nice way to do model selection, because the computations were fairly simple. Sometimes this approach works well but it doesn't always produce the best model.

2.2.2 Backward Elimination

The backward elimination (Draper etc., 1981) tries to examine only the "best" regressions containing a certain number of variables. This technique starts with the model containing all the predictor variables. It is an opposite search method comparing with forward selection.

- First of all, we start by using all the predictor variables in the regression model. We compute the partial F-ratios for each variable and remove the one that has the largest p-value provided it is not significant.
- The variables are deleted from the model one by one until all remaining variables have significant F-ratios.
- Generally, it is recommended that the analysts use fairly large value for α for entry into the model and a more traditional value of α to stay in the model (Myers, 1989). Typical values of α for a variable to stay in the model are usually less than 0.1.

Researchers tend to use this method since it will test all the variables in the model so nothing is missed (Draper etc., 1981). But once we decide to drop a variable from this model, this variable will never come back to this model.

2.2.3 Stepwise Regression

Stepwise regression (Draper etc., 1981) is a combination of backward elimination and forward selection. It can be looked at as an improved method for the forward selection procedure. The difference between these two methods is that stepwise regression retests at each stage then adds a variable in previous stages, since some other variable may no longer be needed in the model.

Stepwise regression procedure focuses on checking when and where to enter the new variable which may be nice in the early stage or the later stage. For example, when we have only x_1 in the model and want to enter a new variable say x_2 , stepwise regression will test to see if x_1 still produces a significant partial F-Ratio.

- We begin by using a single predictor variable in the simple regression model as same as forward selection procedure.
- Comparing p-value in the F-test and decide if we should add another variable.
- Before adding another variable, stepwise regression will check the current model and eliminate the variables which product partial F-ratios which lead to p-value larger than α.
- It attempts to add a variable, eliminate a variable, or interchange between an entered variable and a previous variable at each step.

2.3. Modern Model Selection Techniques

These techniques are usually applied to all possible models if the number of predictor variables is small. In cases where p is large, we first reduce the number of predictor variables by using some statistics like R^2 or Cp to choose variables that fit the data well and which model is best. Then we will select one of the modern criteria to help us figure out the "best" among these possible models.

2.3.1 Akaike Information Criterion (AIC)

Akaike Information Criterion (AIC) (Akaike, 1974) is a model selection technique which helps to get the optimal model. Akaike Information Criterion is an asympotically unbiased estimator of the expected relative Kullback-Leibler information quantity or distance (K-L) (Posada D. and Buckley T.R., 2004).

AIC is defined as the form of

$$AIC = -2ln(L) + 2p,$$

where L is the maximized likelihood estimation of σ^2 for the giving model and p is the number of parameters.

For the linear regression models, AIC can be written as

$$AIC = nln\left(\frac{RSS}{n}\right) + 2p,$$

The model with a smallest value of AIC is the techniques choice for the "true" model.

2.3.2. Sawa's Bayesian Information Criterion (BIC)

Sawa's Bayesian Information Criterion (BIC) (Sawa, 1978) is defined as

$$BIC = nln\left(\frac{RSS}{n}\right) + \frac{2(p+2)n\sigma^2}{RSS} + \frac{2(n\sigma^2)^2}{RSS^2}$$

Takamitsu Sawa showed that BIC is a criterion, which is not only an estimation procedure, but also a procedure for model identification in 1978. More precisely, BIC aimed to develop a procedure for identifying the most adequate model from a given set of alternatives rather than estimating unknown parameters involved in a given true model.

The model with a smallest value of BIC is the model selected by this technique.

2.3.3. Schwarz' Bayesian Criterion (SBC)

The Schwarz' Bayesian Criterion (SBC) was developed by Schwarz (1978) and defined as

$$SBC = -2ln(L) + pln(n),$$

For the linear regression models, the SBC can be written as

$$SBC = nln\left(\frac{RSS}{n}\right) + pln(n)$$

Compared with the AIC, SBC will change depending on the sample size n. The model with a lowest SBC value is treated the best to the "true" model.

3. The Development of the Model Selection Technique

3.1 The Corrected Akaike Information Criterion (AIC_c)

AIC_c is derived by AIC and defined as

$$AIC_c = AIC + \frac{2p(p+1)}{n-p-1}$$

$$= -2ln(L) + 2p + \frac{2p(p+1)}{n-p-1}$$

$$= -2ln(L) + \frac{2pn}{n-p-1}$$

For the linear regression models, the AIC_c can be written as

$$AIC_c = ln\left(\frac{RSS}{n}\right) + \frac{2pn}{n-p-1},$$

When we the value of n is much larger than the number of parameter p, we have $n/(n-p-1) \rightarrow 1 \; as \; n \rightarrow$infinity. Then this criterion becomes the same as AIC. So we usually use it when n is smaller than p. The model with a smallest value of AIC_c is the best approximation to the "true" model.

3.2 Deviance Information Criterion (DIC)

Deviance Information Criterion (Spiegelhalter etc., 2006) is an improvement of the Bayesian Estimation. The DIC is defined as

$$DIC = -2\ln(L) + 2\ln f(y)$$

where f(y) is some fully specified standardizing term.

From the references, since $L = p(y|\theta)$ (where θ is the unknown parameter of the model) is the likelihood function of the observation data and $DIC = \overline{D(\theta)} + p_D$, $p_D = \overline{D(\theta)} - D(\bar{\theta})$,

Then we can modify the DIC as

$$DIC = \overline{D(\theta)} + p_D$$

$$= D(\bar{\theta}) + 2p_D$$

and $$DIC = 2\overline{D(\theta)} - D(\bar{\theta})$$

where $\bar{\theta}$ is the Bayesian estimator.

Comparing AIC and SBC which try to get the best model which will be approximation to the "true" model, DIC will pick up the model without basing on any assumption of the "true" model. We prefer the smallest DIC as the best.

3.3 Other Model Selection Techniques

Since the application of the statistics tool of predicting the practical problem is more useful, the development of the model selection becomes more and more important.

Most of these new techniques are automatic search methods and based on the AIC and Bayesian Estimation, such as Takeuchi's Information Criterion (TIC) (Takeuchi, 1976), Focused Information Criterion (FIC) (Claeskens and Hjort, 2003), The Risk Information Criterion (RIC) (Foster etc., 1994) and so on. These techniques are known and used to predict data in Biology, Economics, Engineering, Medicine and any other field.

4. Multicollinearity in Regression Models

In the linear regression models, we will meet the situation that two or more independent variables are correlated. It is called multicollinearity. There are two types of multicollinearity. One is called perfect multicollinearity, and the other one is called approximate multicollinearity or intercorrelated.

If we have $c_1x_{1i} + c_2x_{2i} + \cdots + c_kx_{ki} = 0, \quad i = 1, 2, 3, \ldots, n,$

and $c_i's$ are not all equal to zero.

This type of relationship among the independent variables is called perfect multicollinearity.

If we have $c_1x_{1i} + c_2x_{2i} + \cdots + c_kx_{ki} + \delta_i = 0, \quad i = 1, 2, 3, \ldots, n,$

and $c_i's \; are \; not \; all \; equal \; to \; zero$, δ_i is a random disturbance term.

This type of relationship among the independent variables is called approximate multicollinearity. This means some of the predictor variables are intercorrelated. In the practical problems, the perfect multicollinearity rarely happens.

When one variable is linearly correlated with the others, multicollinearity may cause the researcher to leave out some of the important predict variables in the regression process. As a result, it becomes difficult to estimate the relationship among variables and get the best regression model and may lead an infeasible result. Multicollinearity makes the problems more difficult. Before we test the data, we should check if there exists variables that are correlated.

5. Simulation

In this example, I set up a set of data with totally 10 predict variables which comes from the normal, lognormal, exponential, and uniform distributions and focus on comparing AIC, BIC and SBC. Since k is equal to 10, we have $2^{10} - 1 = 1023$ possible subset models. It is really a large number and we cannot test all of the possible subset models. I tested three of one-variable “true” models, and reported the results of simulation with rep=100 and rep=1000, then tested three of two-variable "true" models, etc. I averaged the SAS output from one-variable true model, two-variable true model, etc. Based on these values, I drew a picture and showed how well AIC, BIC and SBC fit the "true" model by different numbers of predictor variables in the true model. In the other words, the graphs showed that the frequency of models correctly selected for AIC, BIC and SBC by the number of predict variables in the true model.

Data:

```
x1=10+5*rannor(0);          * normal(10,25);
x2=exp(3*rannor(0));        * lognormal;
x3=5+10*ranuni(0);          * uniform;
x4=50+10*rannor(0);         * normal(50,100);
x5=x1+x4+rannor(0);         *normal bimodal;
x6=5+2*x2+3*ranexp(0);      *lognormal and exponential mixture;
x7=0.5*exp(4*rannor(0));    *lognormal;
x8=10+8*ranuni(0);          * uniform;
x9=x2+x8+2*rannor(0);       * lognormal, uniform and normal mix;
x10=20+x7+9*rannor(0);      * lognormal and normal mix;
```

The results are shown in Figure 1, Figure 2, Figure 3, Figure 4, Figure 5, Figure 6, Figure 7 and Figure 8.

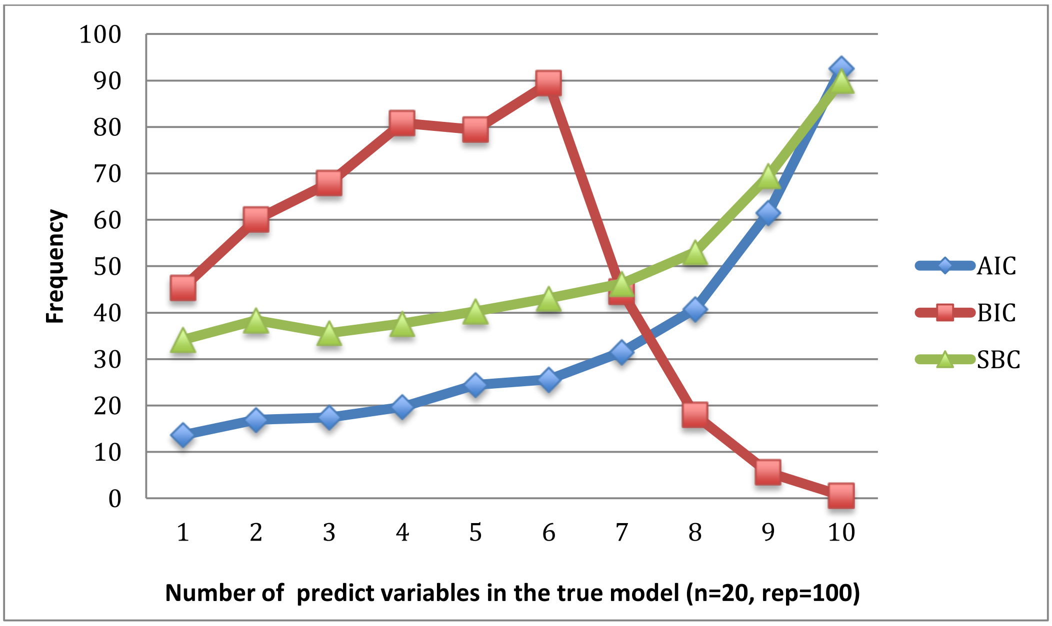

Fig.1. Line plot of frequency of models correctly selected for AIC, BIC and SBC by the number of predict variables in the true model with n=20 and rep=100.

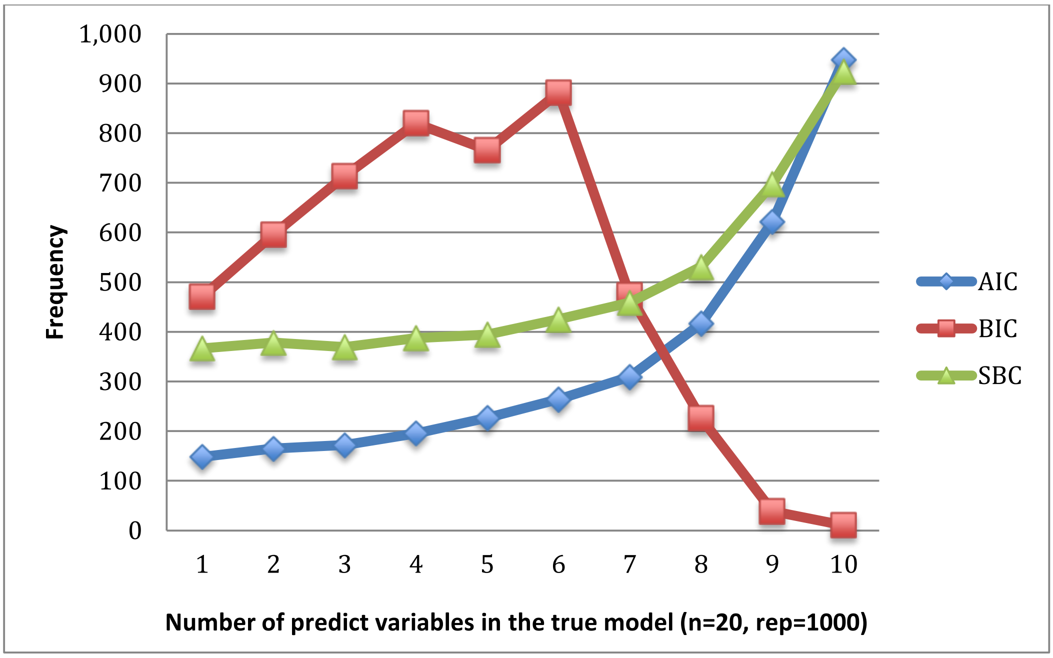

Fig.2. Line plot of frequency of models correctly selected for AIC, BIC and SBC by the number of predict variables in the true model with n=20 and rep=100.

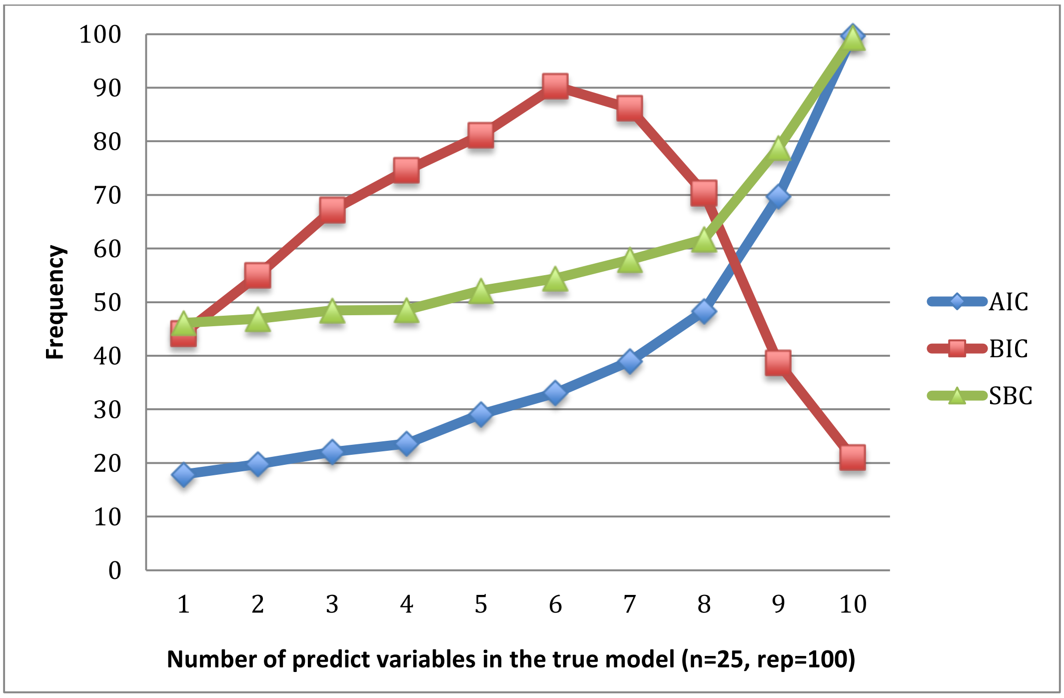

Fig.3. Line plot of frequency of models correctly selected for AIC, BIC and SBC by the number of predict variables in the true model with n=25 and rep=100.

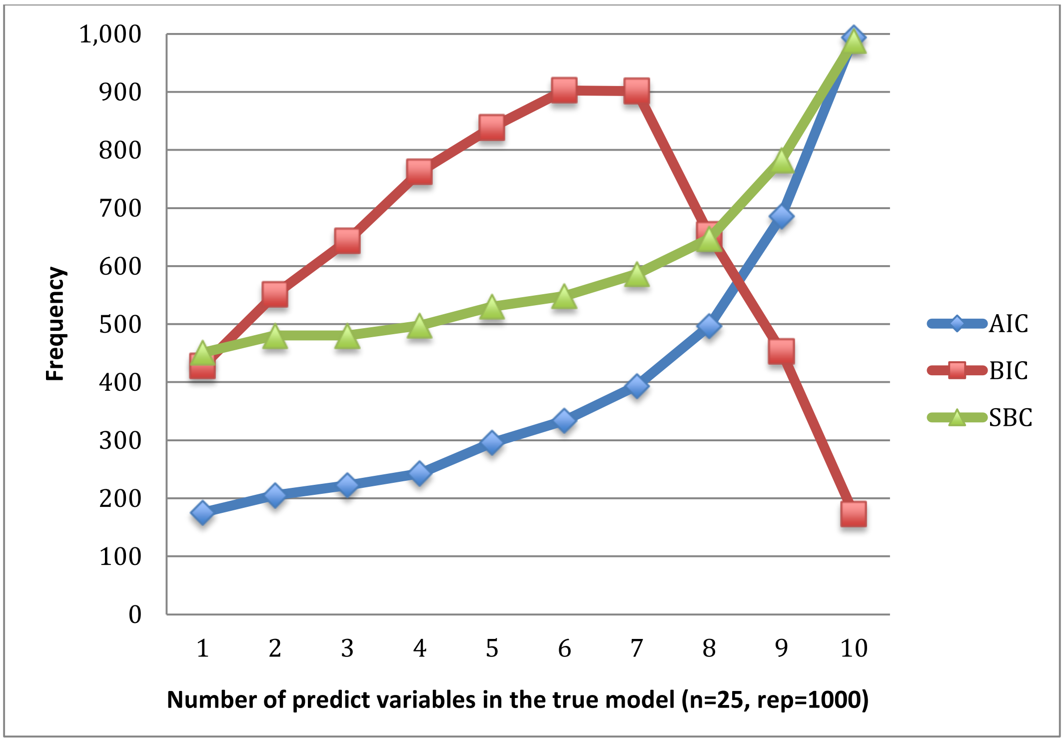

Fig.4. Line plot of frequency of models correctly selected for AIC, BIC and SBC by the number of predict variables in the true model with n=25 and rep=1000.

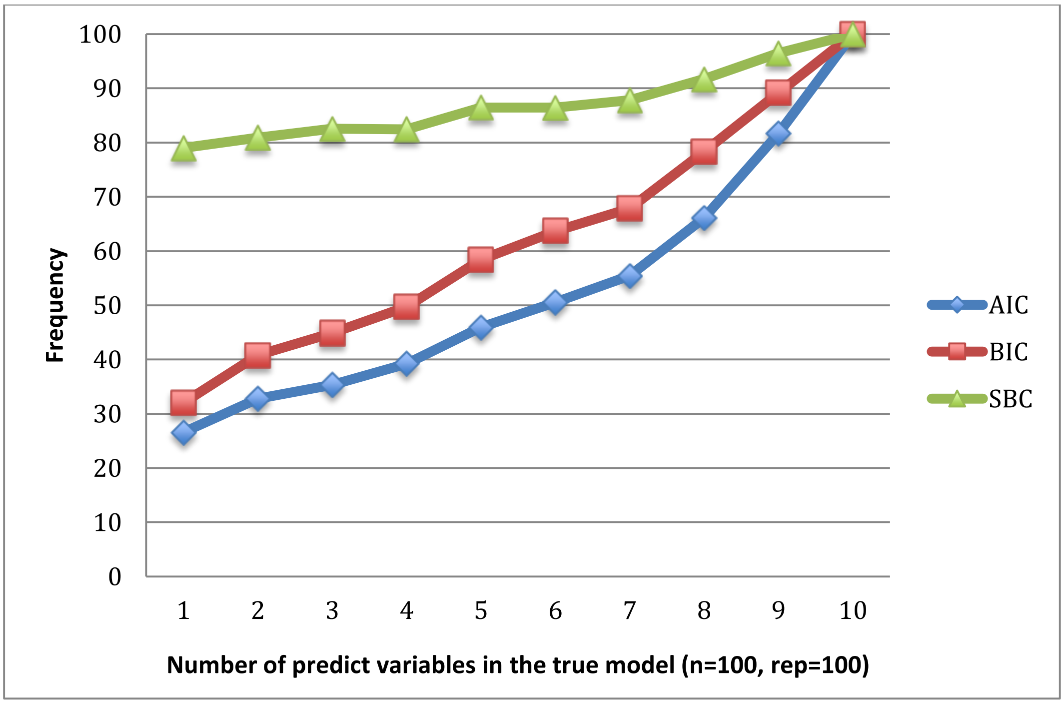

Fig.5. Line plot of frequency of models correctly selected for AIC, BIC and SBC by the number of predict variables in the true model with n=100 and rep=100.

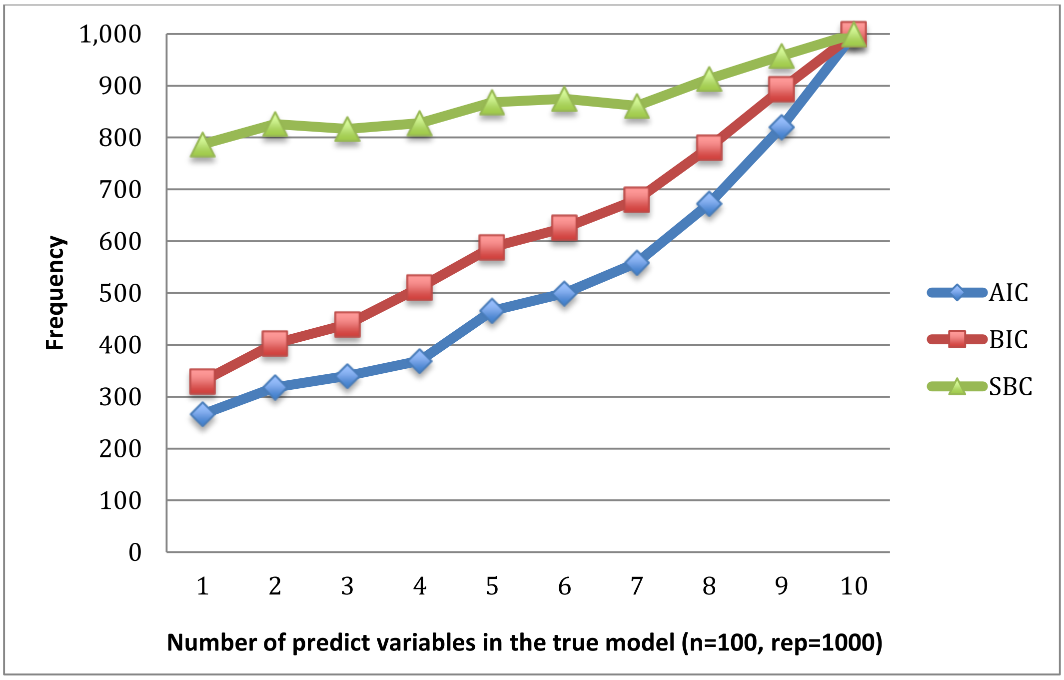

Fig.6. Line plot of frequency of models correctly selected for AIC, BIC and SBC by the number of predict variables in the true model with n=100 and rep=1000.

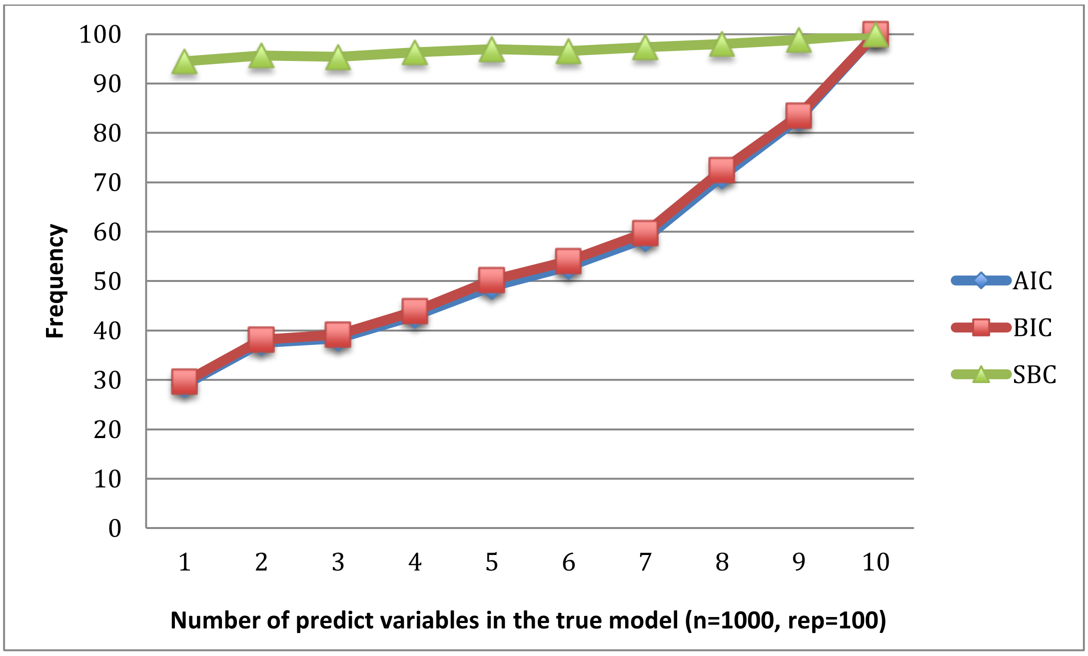

Fig.7. Line plot of frequency of models correctly selected for AIC, BIC and SBC by the number of predict variables in the true model with n=1000 and rep=100.

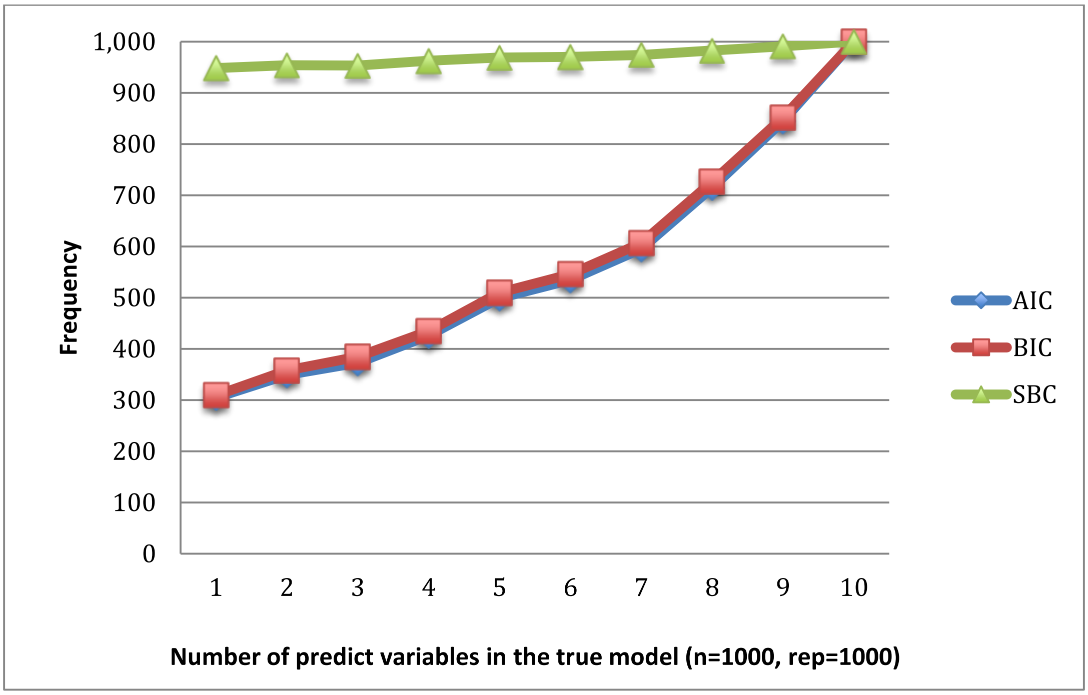

Fig.8. Line plot of frequency of models correctly selected for AIC, BIC and SBC by the number of predict variables in the true model with n=1000 and rep=1000.

Comparing Figure 1 and Figure 2, Figure 3 and Figure 4, Figure 5 and Figure 6, and Figure 7 and Figure 8, respectively, we can see that when rep=100 and rep=1000, we get almost the same result. Since I only recorded three outputs for rep=1000 and got the average number of the output, they are not strict same. But they reflected how well AIC, BIC and SBC performed by different numbers of predictor variables in the true model. Comparing Figure 1, Figure 2, Figure 3, Figure 4 and Figure 5, Figure 6, Figure 7 and Figure 8 for data, which has both the small sample size problems and the small number of predictor variables in the true model, BIC was always the best to fit the true model.

Evaluating Figure 1, Figure 2, Figure 3 and Figure 4, we can see that for the small sample size, BIC increased first and then decreased, which depends on how many estimated parameters in the model. Evaluating Figure 5, Figure 6, Figure 7 and Figure

8, for a large samples size, the SBC was always the best to fit the true model no matter how many predictor variables in the true model. They also showed that when n is large enough, AIC and BIC have the same frequency of models being correctly selected.

From all of the graphs, which are shown in this thesis, the frequency of models being correctly selected by AIC only changed a little by increasing the sample size n.

6. Discussion

Comparing Table 1 and Table 2, both have the same predictor variables (they have 8 predictor variables) in the true model. We can see that BIC did worse for n=20 for the true model, which has 8 predict variables in the true model. When n increases from 20 to 25, BIC did better than before (in Table 1), but still as bad as before (in Table 2). In Table 1 and Table 3, BIC when n is increasing from 20 to 25, the frequency of BIC increases and becomes the best to fit the true model.

Table 1: SAS output for testing true model "y = **33.5** **-9***x1+**3.1***x2+**7.4***x3+**3.5***x4-**4***x5+**3.2***x6+**2.2***x7-**6***x8+**3***rannor(**0**)" (rep=100)

	AIC	BIC	SBC		AIC	BIC	SBC
n=20	38	23	54	n=25	46	97	62
	46	15	69		49	85	66
	41	13	57		55	89	71
	41	19	62		54	95	67
	38	19	56		46	99	68
	41	9	47		48	96	68
	42	13	57		51	84	70
	44	28	53		50	80	72
	49	36	63		48	74	64
	43	12	57		50	83	63
Average	42.3	18.7	57.5	Average	49.7	88.2	67.1

Table 2: SAS output for testing true model "y = **33.5** **-4***x1+**3.2***x3+**2.2***x4-**6***x5-**9***x6-**3.1***x7+**7.4***x8+**3.5***x9+**3***rannor(**0**)"(rep=100)

	AIC	BIC	SBC		AIC	BIC	SBC
n=20	36	2	45	n=25	42	3	34
	29	1	3		46	19	55
	33	7	37		52	54	68
	40	8	46		52	59	69
	35	5	44		41	27	55
	31	0	46		50	42	62
	50	4	62		41	16	47
	30	13	44		45	29	53
	43	13	59		40	12	51
	39	8	53		43	4	45
Average	36.6	6.1	43.9	Average	45.2	26.5	53.9

Table 3: SAS output for testing true model "y = **33.5** +**3.1***x2+**7.4***x3+**3.5***x4-**4***x5+**3.2***x6+**2.2***x7-**6***x8-**9***x10+**3***rannor(**0**)"(rep=100)

	AIC	BIC	SBC		AIC	BIC	SBC
n=20	42	66	62	n=25	56	98	70
	43	25	58		55	100	69
	46	14	58		53	98	65
	46	41	61		36	97	55
	39	17	60		55	79	66
	39	6	52		52	96	60
	40	24	58		41	99	56
	43	54	55		47	99	66
	45	32	56		55	99	71
	50	13	57		51	99	64
Average	43.3	29.2	57.7	Average	50.1	96.4	64.2

By discussing why Table 2 has different results with Table 1 and Table 3, we may test the relationship among numbers of predictor variables in the true model, number of total predictor variables and sample size. We may also consider the number of predictor variables which are multicollinearity.

7. Conclusion

There are several methods for model selection of the multiple linear regression. Also, statisticians are still working on modifying these methods so that a method will fit most of the practical problems. The development of the model selection technique has become more popular. This thesis focused on the most popular modern model selection techniques Akaike's Information Criterion (AIC) (Akaike, 1974), Sawa's Bayesian Information Criterion (BIC) (Sawa, 1978) and Schwarz' Bayesian Criterion (SBC) (Schwarz, 1978) and attempted to give a more general ideas about choosing the best model for the multiple linear regression models.

From the output of SAS, BIC may be the best choice for a small sample size, which also has a small number of predictor variables in the true model. For the large sample size, where n is larger than 100, SBC performed the best for the multiple linear regression models.

Actually, this thesis didn't figure out how small the sample size is, since the different examples showed the different results. It may depend on the ratio of sample size and number of predict variables in the true model. In the future study, I will test different kinds of data and discuss the relationship between the number of estimated parameters and sample size. The ratio of the amount of predictor variables in the true model and sample size or some other factors may give a more reasonable suggestion about which one fits the practical data well. I will also test if the ratio of the number of variables which are correlated and number of predictor variables in the true model will affect the result.

BIBLIOGRAPHY

Akaike, H. (1974), "A New Look at the Statistical model identification IEEE Trans," *Automat*. Control 19, 716-723.

Cleaskens, G., and Hjort, N. L. (2003), "The focused Information Criterion," *Journal of the American Statistical Association,*98, 900-916.

Draper, N. R., and Smith H., *Applied Regression Analysis,* New York, Chichester, Brisbane, Toronto: John Wiley & Sons, Inc., 1981. Print.

Foster, Dean P., and George, Edward I. (1994), "The Risk Inflation Criterion for Multiple Regression," *The Annals of Statistics,* Vol. 22, No. 4, 1947-1975.

Freund Rudolf J., William J. Wilson, and Ping Sa, *Regression Analysis: Statistical Model of A Response Variable*, Burlingto, MA: Academic Press, 2006. Print.

Kunter, Michael H., Christopher J. Nachtsheim, and John Neter, *Applied Linear Regression Models*, New York, NY: McGraw-Hill Itwin, 2004. Print.

Mallows, C. L. (1973), "Some Comments on Cp," *Technometrics*, 15, 661-675.

Myers, Raymond H., *Classical and Modern Regression with Applications*, Boston, MA: PWS-KENT, 1989. Print.

Posada, D. and Buckley, T.R. (2004), "Model Selection and Model Averaging in Phylogenetics: Advantages of Akaike Information Criterion and Bayesian Approaches Over Likelihood Ratio Tests," Biol. 53(5):793-808, 2004, DOI: 10.1080/10635150490522304.

Sawa, T. (1978), "Information Criteria for Discriminating Among Alternative Regression Models," *Econometrica*, 46, 1273-1282.

SAS Institute Inc., *SAS/STAT User's Guide, Version 6, Fourth Edition, Volume 2,* Cary, NC: SAS Institute Inc., 1989. 846 pp.

Schwarz, G. (1978), “Estimating the Dimension of a Model,” *Annals of Statistics,* 6, 461464.

Takeuchi, K. (1976), “Distribution of Information Statistics and Criteria for Adequacy of Models,” *Econometrica,* 57, 307-333.

Weisberg, Sanford (1947), *Applied Linear Regression,* New York, Chichester, Brisbane, Toronto, Singapore: John Willey & Sons, 1980. Print.

Appendix A

Appendix A

SAS program for testing true model y = **33.5** + **3***x2+**3***rannor(**0**) with n=20, rep=100.
*We need change "Do I = **1** to **20**;" to 25, 100, 1000. and the rep from 100 to 1000.

```
Data Design;
Do I = 1 to 20;
x1=10+5*rannor(0); * normal(10,25);
x2=exp(3*rannor(0)); * lognormal;
x3=5+10*ranuni(0); * uniform;
x4=50+10*rannor(0); * normal(50,100);
x5=x1+x4+rannor(0); *normal bimodal;
x6=5+2*x2+3*ranexp(0); *lognormal and exponential mixture;
x7=0.5*exp(4*rannor(0)); *lognormal;
x8=10+8*ranuni(0); * uniform;
x9=x2+x8+2*rannor(0); * lognormal, uniform and normal mix;
x10=20+x7+9*rannor(0); * lognormal and normal mix;
output;End; drop i;run;
Data One;Set Design;
Do Rep = 1 to 100;
y = 33.5 + 3*x2+3*rannor(0); *true model;
OutPut; End;
Proc Sort; By Rep;
*Proc Print;*Var x1-x10 y;
proc reg Outest=Stat NoPrint; By Rep;
model y=x1-x10/Selection=adjRsq aic bic sbc;run;
proc sort; by Rep _aic_; *Proc Print; *Var  Rep  _aic_;
Data Final1;Set Stat;
m = Mod(_n_, 1023); *Proc Print; *Var rep x1-x10 _aic_;
Data AIC; Set Final1;
If m=1;
 If x1 = "." and x2 ne "." and x3 = "." and x4 = "." and x5 = "." and
x6 = "." and x7 = "." and x8 = "." and x9 = "." and x10 = "." Then ax2+1;
Data Final;Set AIC;
If Rep=100;
proc print; Var m Rep  _aic_ ax2;
proc sort Data=Stat; by Rep _bic_; *Proc Print; *Var  Rep  _bic_;
Data Final2;Set Stat;
m = Mod(_n_, 1023); *Proc Print; *Var rep x1-x10 _bic_;
Data BIC; Set Final2;
If m=1;
 If x1 = "." and x2 ne "." and x3 = "." and x4 = "." and x5 = "." and
x6 = "." and x7 = "." and x8 = "." and x9 = "." and x10 = "." Then ax2+1;
Data Final;Set BIC;
If Rep=100;
proc print; Var m Rep  _bic_ ax2;
proc sort Data=Stat; by Rep _sbc_; *Proc Print; *Var  Rep  _sbc_;
Data Final3;Set Stat;
m = Mod(_n_, 1023); *Proc Print; *Var rep x1-x10 _sbc_;
Data SBC; Set Final3;
```

```
If m=1;
 If x1 = "." and x2 ne "." and x3 = "." and x4 = "." and x5 = "." and
x6 = "." and x7 = "." and x8 = "." and x9 = "." and x10 = "." Then ax2+1;
Data Final;Set SBC;
If Rep=100;
proc print; Var m Rep  _sbc_ ax2;
Run;
```

SAS program for testing true model y = **33.5 -2.7***x3+**3***rannor(**0**) with n=20, rep=100.
*We need change “Do I = **1** to **20**;” to 25, 100, 1000. and the rep from 100 to 1000.

```
Data Design;
Do I = 1 to 20;
x1=10+5*rannor(0); * normal(10,25);
x2=exp(3*rannor(0)); * lognormal;
x3=5+10*ranuni(0); * uniform;
x4=50+10*rannor(0); * normal(50,100);
x5=x1+x4+rannor(0); *normal bimodal;
x6=5+2*x2+3*ranexp(0); *lognormal and exponential mixture;
x7=0.5*exp(4*rannor(0)); *lognormal;
x8=10+8*ranuni(0); * uniform;
x9=x2+x8+2*rannor(0); * lognormal, uniform and normal mix;
x10=20+x7+9*rannor(0); * lognormal and normal mix;
output;
End; drop i;run;
Data One;Set Design;
Do Rep = 1 to 100;
y = 33.5 -2.7*x3+3*rannor(0); *true model;
OutPut; End;
Proc Sort; By Rep;
*Proc Print;*Var x1-x10 y;
proc reg Outest=Stat NoPrint; By Rep;
model y=x1-x10/Selection=adjRsq aic bic sbc;
run;
proc sort; by Rep _aic_; *Proc Print; *Var  Rep  _aic_;
Data Final1;Set Stat;
m = Mod(_n_, 1023); *Proc Print; *Var rep x1-x10 _aic_;
Data AIC; Set Final1;
If m=1;
 If x1 = "." and x2 = "." and x3 ne "." and x4 = "." and x5 = "." and
x6 = "." and x7 = "." and x8 = "." and x9 = "." and x10 = "." Then ax3+1;
Data Final;Set AIC;
If Rep=100;
proc print; Var m Rep  _aic_ ax3;
proc sort Data=Stat; by Rep _bic_; *Proc Print; *Var  Rep  _bic_;
Data Final2;Set Stat;
m = Mod(_n_, 1023); *Proc Print; *Var rep x1-x10 _bic_;
Data BIC; Set Final2;
```

```
If m=1;
 If x1 = "." and x2 = "." and x3 ne "." and x4 = "." and x5 = "." and
x6 = "." and x7 = "." and x8 = "." and x9 = "." and x10 = "." Then ax3+1;
Data Final;Set BIC;
If Rep=100;
proc print; Var m Rep  _bic_ ax3;
proc sort Data=Stat; by Rep _sbc_; *Proc Print; *Var  Rep  _sbc_;
Data Final3;Set Stat;
m = Mod(_n_, 1023); *Proc Print; *Var rep x1-x10 _sbc_;
Data SBC; Set Final3;
If m=1;
 If x1 = "." and x2 = "." and x3 ne "." and x4 = "." and x5 = "." and
x6 = "." and x7 = "." and x8 = "." and x9 = "." and x10 = "." Then ax3+1;
Data Final;Set SBC;
If Rep=100;
proc print; Var m Rep  _sbc_ ax3;
Run;
```

SAS program for testing true model y = **33.5 -2.2***x10+**3***rannor(**0**) with n=20, rep=100.

*We need change “Do I = **1** to **20**;” to 25, 100, 1000. and the rep from 100 to 1000.

```
Data Design;
Do I = 1 to 20;
x1=10+5*rannor(0); * normal(10,25);
x2=exp(3*rannor(0)); * lognormal;
x3=5+10*ranuni(0); * uniform;
x4=50+10*rannor(0); * normal(50,100);
x5=x1+x4+rannor(0); *normal bimodal;
x6=5+2*x2+3*ranexp(0); *lognormal and exponential mixture;
x7=0.5*exp(4*rannor(0)); *lognormal;
x8=10+8*ranuni(0); * uniform;
x9=x2+x8+2*rannor(0); * lognormal, uniform and normal mix;
x10=20+x7+9*rannor(0); * lognormal and normal mix;
output; End; drop i;run;
Data One;Set Design;
Do Rep = 1 to 100;
y = 33.5 -2.2*x10+3*rannor(0); *true model;
OutPut;  End;
Proc Sort; By Rep;
*Proc Print;*Var x1-x10 y;
proc reg Outest=Stat NoPrint; By Rep;
model y=x1-x10/Selection=adjRsq aic bic sbc;
run;
proc sort; by Rep _aic_; *Proc Print; *Var  Rep  _aic_;
Data Final1;Set Stat;
m = Mod(_n_, 1023); *Proc Print; *Var rep x1-x10 _aic_;
Data AIC; Set Final1;
If m=1;
```

```
 If x1 = "." and x2 = "." and x3 = "." and x4 = "." and x5 = "." and
x6 = "." and x7 = "." and x8 = "." and x9 = "." and x10 ne "." Then ax10+1;
Data Final;Set AIC;
If Rep=100;
proc print; Var m Rep  _aic_ ax10;
proc sort Data=Stat; by Rep _bic_; *Proc Print; *Var  Rep  _bic_;
Data Final2;Set Stat;
m = Mod(_n_, 1023); *Proc Print; *Var rep x1-x10 _bic_;
Data BIC; Set Final2;
If m=1;
  If x1 = "." and x2 = "." and x3 = "." and x4 = "." and x5 = "." and
x6 = "." and x7 = "." and x8 = "." and x9 = "." and x10 ne "." Then ax10+1;
Data Final;Set BIC;
If Rep=100;
proc print; Var m Rep  _bic_ ax10;
proc sort Data=Stat; by Rep _sbc_; *Proc Print; *Var  Rep  _sbc_;
Data Final3;Set Stat;
m = Mod(_n_, 1023); *Proc Print; *Var rep x1-x10 _sbc_;
Data SBC; Set Final3;
If m=1;
 If x1 = "." and x2 = "." and x3 = "." and x4 = "." and x5 = "." and
x6 = "." and x7 = "." and x8 = "." and x9 = "." and x10 ne "." Then ax10+1;
Data Final;Set SBC;
If Rep=100;
proc print; Var m Rep  _sbc_ ax10;
Run;
```

SAS program for testing true model y = **33.5** -**2.7***x3-**4***x5+**3***rannor(**0**) with n=20, rep=100.

*We need change "Do I = 1 to 20;" to 25, 100, 1000. and the rep from 100 to 1000.

```
Data Design;
Do I = 1 to 20;
x1=10+5*rannor(0); * normal(10,25);
x2=exp(3*rannor(0)); * lognormal;
x3=5+10*ranuni(0); * uniform;
x4=50+10*rannor(0); * normal(50,100);
x5=x1+x4+rannor(0); *normal bimodal;
x6=5+2*x2+3*ranexp(0); *lognormal and exponential mixture;
x7=0.5*exp(4*rannor(0)); *lognormal;
x8=10+8*ranuni(0); * uniform;
x9=x2+x8+2*rannor(0); * lognormal, uniform and normal mix;
x10=20+x7+9*rannor(0); * lognormal and normal mix;
output; End; drop i;run;
Data One;Set Design;
Do Rep = 1 to 100;
y = 33.5 -2.7*x3-4*x5+3*rannor(0); *true model;
OutPut;  End;
```

```
Proc Sort; By Rep;
*Proc Print;*Var x1-x10 y;
proc reg Outest=Stat NoPrint; By Rep;
model y=x1-x10/Selection=adjRsq aic bic sbc;
run;
proc sort; by Rep _aic_; *Proc Print; *Var  Rep  _aic_;
Data Final1;Set Stat;
m = Mod(_n_, 1023); *Proc Print; *Var rep x1-x10 _aic_;
Data AIC; Set Final1;
If m=1;
If x1 = "." and x2 = "." and x3 ne "." and x4 = "." and x5 ne "." and
x6 = "." and x7 = "." and x8 = "." and x9 = "." and x10 = "." Then x35+1;
Data Final;Set AIC;
If Rep=100;
proc print; Var m Rep  _aic_ x35;
proc sort Data=Stat; by Rep _bic_; *Proc Print; *Var  Rep  _bic_;
Data Final2;Set Stat;
m = Mod(_n_, 1023); *Proc Print; *Var rep x1-x10 _bic_;
Data BIC; Set Final2;
If m=1;
 If x1 = "." and x2 = "." and x3 ne "." and x4 = "." and x5 ne "." and
x6 = "." and x7 = "." and x8 = "." and x9 = "." and x10 = "." Then x35+1;
Data Final;Set BIC;
If Rep=100;
proc print; Var m Rep  _bic_ x35;
proc sort Data=Stat; by Rep _sbc_; *Proc Print; *Var  Rep  _sbc_;
Data Final3;Set Stat;
m = Mod(_n_, 1023); *Proc Print; *Var rep x1-x10 _sbc_;
Data SBC; Set Final3;
If m=1;
If x1 = "." and x2 = "." and x3 ne "." and x4 = "." and x5 ne "." and
x6 = "." and x7 = "." and x8 = "." and x9 = "." and x10 = "." Then x35+1;;
Data Final;Set SBC;
If Rep=100;
proc print; Var m Rep  _sbc_ x35;
Run;
```

SAS program for testing true model y = **33.5** **-2.7***x3+**3.2***x9+**3***rannor(**0**) with n=20, rep=100.

*We need change “Do I = **1** to **20**;” to 25, 100, 1000. and the rep from 100 to 1000.

```
Data Design;
Do I = 1 to 20;
x1=10+5*rannor(0); * normal(10,25);
x2=exp(3*rannor(0)); * lognormal;
x3=5+10*ranuni(0); * uniform;
x4=50+10*rannor(0); * normal(50,100);
x5=x1+x4+rannor(0); *normal bimodal;
x6=5+2*x2+3*ranexp(0); *lognormal and exponential mixture;
```

```
x7=0.5*exp(4*rannor(0)); *lognormal;
x8=10+8*ranuni(0); * uniform;
x9=x2+x8+2*rannor(0); * lognormal, uniform and normal mix;
x10=20+x7+9*rannor(0); * lognormal and normal mix;
output; End; drop i;run;
Data One;Set Design;
Do Rep = 1 to 100;
y = 33.5 -2.7*x3+3.2*x9+3*rannor(0); *true model;
OutPut;  End;
Proc Sort; By Rep;
*Proc Print;*Var x1-x10 y;
proc reg Outest=Stat NoPrint; By Rep;
model y=x1-x10/Selection=adjRsq aic bic sbc;
run;
proc sort; by Rep _aic_; *Proc Print; *Var  Rep  _aic_;
Data Final1;Set Stat;
m = Mod(_n_, 1023); *Proc Print; *Var rep x1-x10 _aic_;
Data AIC; Set Final1;
If m=1;
 If x1 = "." and x2 = "." and x3 ne "." and x4 = "." and x5 = "." and
x6 = "." and x7 = "." and x8 = "." and x9 ne "." and x10 = "." Then x39+1;
Data Final;Set AIC;
If Rep=100;
proc print; Var m Rep  _aic_ x39;
proc sort Data=Stat; by Rep _bic_; *Proc Print; *Var  Rep  _bic_;
Data Final2;Set Stat;
m = Mod(_n_, 1023); *Proc Print; *Var rep x1-x10 _bic_;
Data BIC; Set Final2;
If m=1;
 If x1 = "." and x2 = "." and x3 ne "." and x4 = "." and x5 = "." and
x6 = "." and x7 = "." and x8 = "." and x9 ne "." and x10 = "." Then x39+1;
Data Final;Set BIC;
If Rep=100;
proc print; Var m Rep  _bic_ x39;
proc sort Data=Stat; by Rep _sbc_; *Proc Print; *Var  Rep  _sbc_;
Data Final3;Set Stat;
m = Mod(_n_, 1023); *Proc Print; *Var rep x1-x10 _sbc_;
Data SBC; Set Final3;
If m=1;
If x1 = "." and x2 = "." and x3 ne "." and x4 = "." and x5 = "." and
x6 = "." and x7 = "." and x8 = "." and x9 ne "." and x10 = "." Then x39+1;
Data Final;Set SBC;
If Rep=100;
proc print; Var m Rep  _sbc_ x39;
Run;
```

SAS program for testing true model y = **33.5** + **3***x4-**2.2***x5+**3***rannor(**0**) with n=20, rep=100.
*We need change “Do I = **1** to **20**;” to 25, 100, 1000. and the rep from 100 to 1000.

```
Data Design;
Do I = 1 to 20;
x1=10+5*rannor(0); * normal(10,25);
x2=exp(3*rannor(0)); * lognormal;
x3=5+10*ranuni(0); * uniform;
x4=50+10*rannor(0); * normal(50,100);
x5=x1+x4+rannor(0); *normal bimodal;
x6=5+2*x2+3*ranexp(0); *lognormal and exponential mixture;
x7=0.5*exp(4*rannor(0)); *lognormal;
x8=10+8*ranuni(0); * uniform;
x9=x2+x8+2*rannor(0); * lognormal, uniform and normal mix;
x10=20+x7+9*rannor(0); * lognormal and normal mix;
output; End; drop i;run;
Data One;Set Design;
Do Rep = 1 to 100;
y = 33.5 + 3*x4-2.2*x5+3*rannor(0); *true model;
OutPut;  End;
Proc Sort; By Rep;
*Proc Print;*Var x1-x10 y;
proc reg Outest=Stat NoPrint; By Rep;
model y=x1-x10/Selection=adjRsq aic bic sbc;
run;
proc sort; by Rep _aic_; *Proc Print; *Var  Rep  _aic_;
Data Final1;Set Stat;
m = Mod(_n_, 1023); *Proc Print; *Var rep x1-x10 _aic_;
Data AIC; Set Final1;
If m=1;
If x1 = "." and x2 = "." and x3 = "." and x4 ne "." and x5 ne "." and
x6 = "." and x7 = "." and x8 = "." and x9 = "." and x10 = "." Then x45+1;
Data Final;Set AIC;
If Rep=100;
proc print; Var m Rep  _aic_ x45;
proc sort Data=Stat; by Rep _bic_; *Proc Print; *Var  Rep  _bic_;
Data Final2;Set Stat;
m = Mod(_n_, 1023); *Proc Print; *Var rep x1-x10 _bic_;
Data BIC; Set Final2;
If m=1;
If x1 = "." and x2 = "." and x3 = "." and x4 ne "." and x5 ne "." and
x6 = "." and x7 = "." and x8 = "." and x9 = "." and x10 = "." Then x45+1;
Data Final;Set BIC;
If Rep=100;
proc print; Var m Rep  _bic_ x45;
proc sort Data=Stat; by Rep _sbc_; *Proc Print; *Var  Rep  _sbc_;
Data Final3;Set Stat;
m = Mod(_n_, 1023); *Proc Print; *Var rep x1-x10 _sbc_;
Data SBC; Set Final3;
If m=1;
If x1 = "." and x2 = "." and x3 = "." and x4 ne "." and x5 ne "." and
x6 = "." and x7 = "." and x8 = "." and x9 = "." and x10 = "." Then x45+1;
```

```
Data Final;Set SBC;
If Rep=100;
proc print; Var m Rep  _sbc_ x45;
Run;
```

SAS program for testing true model

y = 33.5 + 3*x2-4*x3+5*x5+3*rannor(0) with n=20, rep=100.

*We need change “Do I = 1 to 20;” to 25, 100, 1000. and the rep from 100 to 1000.

```
Data Design;
Do I = 1 to 20;
x1=10+5*rannor(0); * normal(10,25);
x2=exp(3*rannor(0)); * lognormal;
x3=5+10*ranuni(0); * uniform;
x4=50+10*rannor(0); * normal(50,100);
x5=x1+x4+rannor(0); *normal bimodal;
x6=5+2*x2+3*ranexp(0); *lognormal and exponential mixture;
x7=0.5*exp(4*rannor(0)); *lognormal;
x8=10+8*ranuni(0); * uniform;
x9=x2+x8+2*rannor(0); * lognormal, uniform and normal mix;
x10=20+x7+9*rannor(0); * lognormal and normal mix;
output; End; drop i;run;
Data One;Set Design;
Do Rep = 1 to 100;
y = 33.5 + 3*x2-4*x3+5*x5+3*rannor(0); *true model;
OutPut; End;
Proc Sort; By Rep; *Proc Print;*Var x1-x10 y;
proc reg Outest=Stat NoPrint; By Rep;
model y=x1-x10/Selection=adjRsq aic bic sbc;
run;
proc sort; by Rep _aic_; *Proc Print; *Var  Rep  _aic_;
Data Final1;Set Stat;
m = Mod(_n_, 1023); *Proc Print; *Var rep x1-x10 _aic_;
Data AIC; Set Final1;
If m=1;
If x1 = "." and x2 ne "." and x3 ne "." and x4 = "." and x5 ne "." and x6 = "." and x7 =
"." and x8 = "." and x9 = "." and x10 = "." Then x235+1;
Data Final;Set AIC;
If Rep=100;
proc print; Var m Rep  _aic_ x235;
proc sort Data=Stat; by Rep _bic_; *Proc Print; *Var  Rep  _bic_;
Data Final2;Set Stat;
m = Mod(_n_, 1023); *Proc Print; *Var rep x1-x10 _bic_;
Data BIC; Set Final2;
If m=1;
If x1 = "." and x2 ne "." and x3 ne "." and x4 = "." and x5 ne "." and x6 = "." and x7 =
"." and x8 = "." and x9 = "." and x10 = "." Then x235+1;
Data Final;Set BIC;
If Rep=100;
```

```
proc print; Var m Rep  _bic_ x235;
proc sort Data=Stat; by Rep _sbc_; *Proc Print; *Var  Rep  _sbc_;
Data Final3;Set Stat;
m = Mod(_n_, 1023); *Proc Print; *Var rep x1-x10 _sbc_;
Data SBC; Set Final3;
If m=1;
If x1 = "." and x2 ne "." and x3 ne "." and x4 = "." and x5 ne "." and x6 = "." and x7 =
"." and x8 = "." and x9 = "." and x10 = "." Then x235+1;
Data Final;Set SBC;
If Rep=100;
proc print; Var m Rep  _sbc_ x235;
Run;
```

SAS program for testing true model
y = 33.5 + 3*x2-4*x8+5*x9+3*rannor(0) with n=20, rep=100.
*We need change “Do I = 1 to 20;” to 25, 100, 1000. and the rep from 100 to 1000.

```
Data Design;
Do I = 1 to 20;
x1=10+5*rannor(0); * normal(10,25);
x2=exp(3*rannor(0)); * lognormal;
x3=5+10*ranuni(0); * uniform;
x4=50+10*rannor(0); * normal(50,100);
x5=x1+x4+rannor(0); *normal bimodal;
x6=5+2*x2+3*ranexp(0); *lognormal and exponential mixture;
x7=0.5*exp(4*rannor(0)); *lognormal;
x8=10+8*ranuni(0); * uniform;
x9=x2+x8+2*rannor(0); * lognormal, uniform and normal mix;
x10=20+x7+9*rannor(0); * lognormal and normal mix;
output; End; drop i;run;
Data One;Set Design;
Do Rep = 1 to 100;
y = 33.5 + 3*x2-4*x8+5*x9+3*rannor(0); *true model;
OutPut;  End;
Proc Sort; By Rep; *Proc Print;*Var x1-x10 y;
proc reg Outest=Stat NoPrint; By Rep;
model y=x1-x10/Selection=adjRsq aic bic sbc;
run;
proc sort; by Rep _aic_; *Proc Print; *Var  Rep  _aic_;
Data Final1;Set Stat;
m = Mod(_n_, 1023); *Proc Print; *Var rep x1-x10 _aic_;
Data AIC; Set Final1;
If m=1;
If x1 = "." and x2 ne "." and x3 = "." and x4 = "." and x5 = "." and
x6 = "." and x7 = "." and x8 ne "." and x9 ne "." and x10 = "." Then x289+1;
Data Final;Set AIC;
If Rep=100;
proc print; Var m Rep  _aic_ x289;
proc sort Data=Stat; by Rep _bic_; *Proc Print; *Var  Rep  _bic_;
```

```
Data Final2;Set Stat;
m = Mod(_n_, 1023); *Proc Print; *Var rep x1-x10 _bic_;
Data BIC; Set Final2;
If m=1;
If x1 = "." and x2 ne "." and x3 = "." and x4 = "." and x5 = "." and
x6 = "." and x7 = "." and x8 ne "." and x9 ne "." and x10 = "." Then x289+1;
Data Final;Set BIC;
If Rep=100;
proc print; Var m Rep  _bic_ x289;
proc sort Data=Stat; by Rep _sbc_; *Proc Print; *Var  Rep  _sbc_;
Data Final3;Set Stat;
m = Mod(_n_, 1023); *Proc Print; *Var rep x1-x10 _sbc_;
Data SBC; Set Final3;
If m=1;
If x1 = "." and x2 ne "." and x3 = "." and x4 = "." and x5 = "." and
x6 = "." and x7 = "." and x8 ne "." and x9 ne "." and x10 = "." Then x289+1;
Data Final;Set SBC;
If Rep=100;
proc print; Var m Rep  _sbc_ x289;
Run;
```

SAS program for testing true model
y = **33.5** -**4***x3+**2.2***x4+**3.2***x7+**3***rannor(**0**) with n=20, rep=100.
*We need change “Do I = 1 to 20;” to 25, 100, 1000. and the rep from 100 to 1000.

```
Data Design;
Do I = 1 to 20;
x1=10+5*rannor(0); * normal(10,25);
x2=exp(3*rannor(0)); * lognormal;
x3=5+10*ranuni(0); * uniform;
x4=50+10*rannor(0); * normal(50,100);
x5=x1+x4+rannor(0); *normal bimodal;
x6=5+2*x2+3*ranexp(0); *lognormal and exponential mixture;
x7=0.5*exp(4*rannor(0)); *lognormal;
x8=10+8*ranuni(0); * uniform;
x9=x2+x8+2*rannor(0); * lognormal, uniform and normal mix;
x10=20+x7+9*rannor(0); * lognormal and normal mix;
output; End; drop i;run;
Data One;Set Design;
Do Rep = 1 to 100;
y = 33.5 -4*x3+2.2*x4+3.2*x7+3*rannor(0); *true model;
OutPut;  End;
Proc Sort; By Rep; *Proc Print;*Var x1-x10 y;
proc reg Outest=Stat NoPrint; By Rep;
model y=x1-x10/Selection=adjRsq aic bic sbc;
run;
proc sort; by Rep _aic_; *Proc Print; *Var  Rep  _aic_;
Data Final1;Set Stat;
m = Mod(_n_, 1023); *Proc Print; *Var rep x1-x10 _aic_;
```

```
Data AIC; Set Final1;
If m=1;
If x1 = "." and x2 = "." and x3 ne "." and x4 ne "." and x5 = "." and
x6 = "." and x7 ne "." and x8 = "." and x9 = "." and x10 = "." Then x347+1;
Data Final;Set AIC;
If Rep=100;
proc print; Var m Rep  _aic_ x347;
proc sort Data=Stat; by Rep _bic_; *Proc Print; *Var  Rep  _bic_;
Data Final2;Set Stat;
m = Mod(_n_, 1023); *Proc Print; *Var rep x1-x10 _bic_;
Data BIC; Set Final2;
If m=1;
If x1 = "." and x2 = "." and x3 ne "." and x4 ne "." and x5 = "." and
x6 = "." and x7 ne "." and x8 = "." and x9 = "." and x10 = "." Then x347+1;
Data Final;Set BIC;
If Rep=100;
proc print; Var m Rep  _bic_ x347;
proc sort Data=Stat; by Rep _sbc_; *Proc Print; *Var  Rep  _sbc_;
Data Final3;Set Stat;
m = Mod(_n_, 1023); *Proc Print; *Var rep x1-x10 _sbc_;
Data SBC; Set Final3;
If m=1;
If x1 = "." and x2 = "." and x3 ne "." and x4 ne "." and x5 = "." and
x6 = "." and x7 ne "." and x8 = "." and x9 = "." and x10 = "." Then x347+1;
Data Final;Set SBC;
If Rep=100;
proc print; Var m Rep  _sbc_ x347;
Run;
```

SAS program for testing true model

y = **33.5**+**2.2***x3-**6***x5-**9***x6-**3.1***x7+**3***rannor(**0**) with n=20, rep=100.

*We need change “Do I = 1 to 20;” to 25, 100, 1000. and the rep from 100 to 1000.

```
Data Design;
Do I = 1 to 20;
x1=10+5*rannor(0); * normal(10,25);
x2=exp(3*rannor(0)); * lognormal;
x3=5+10*ranuni(0); * uniform;
x4=50+10*rannor(0); * normal(50,100);
x5=x1+x4+rannor(0); *normal bimodal;
x6=5+2*x2+3*ranexp(0); *lognormal and exponential mixture;
x7=0.5*exp(4*rannor(0)); *lognormal;
x8=10+8*ranuni(0); * uniform;
x9=x2+x8+2*rannor(0); * lognormal, uniform and normal mix;
x10=20+x7+9*rannor(0); * lognormal and normal mix;
output; End; drop i;run;
Data One;Set Design;
Do Rep = 1 to 100;
y = 33.5+2.2*x3-6*x5-9*x6-3.1*x7+3*rannor(0); *true model;
```

```
OutPut;  End;
Proc Sort; By Rep; *Proc Print;*Var x1-x10 y;
proc reg Outest=Stat NoPrint; By Rep;
model y=x1-x10/Selection=adjRsq aic bic sbc;
run;
proc sort; by Rep _aic_; *Proc Print; *Var  Rep  _aic_;
Data Final1;Set Stat;
m = Mod(_n_, 1023); *Proc Print; *Var rep x1-x10 _aic_;
Data AIC; Set Final1;
If m=1;
If x1 = "." and x2 = "." and x3 ne "." and x4 = "." and x5 ne "." and
x6 ne "." and x7 ne "." and x8 = "." and x9 = "." and x10 = "." Then x3567+1;
Data Final;Set AIC;
If Rep=100;
proc print; Var m Rep  _aic_  x3567;
proc sort Data=Stat; by Rep _bic_; *Proc Print; *Var  Rep  _bic_;
Data Final2;Set Stat;
m = Mod(_n_, 1023); *Proc Print; *Var rep x1-x10 _bic_;
Data BIC; Set Final2;
If m=1;
If x1 = "." and x2 = "." and x3 ne "." and x4 = "." and x5 ne "." and
x6 ne "." and x7 ne "." and x8 = "." and x9 = "." and x10 = "." Then x3567+1;
Data Final;Set BIC;
If Rep=100;
proc print; Var m Rep  _bic_  x3567;
proc sort Data=Stat; by Rep _sbc_; *Proc Print; *Var  Rep  _sbc_;
Data Final3;Set Stat;
m = Mod(_n_, 1023); *Proc Print; *Var rep x1-x10 _sbc_;
Data SBC; Set Final3;
If m=1;
If x1 = "." and x2 = "." and x3 ne "." and x4 = "." and x5 ne "." and
x6 ne "." and x7 ne "." and x8 = "." and x9 = "." and x10 = "." Then x3567+1;
Data Final;Set SBC;
If Rep=100;
proc print; Var m Rep  _sbc_ x3567;
Run;
```

SAS program for testing true model
y = **33.5** +**3.2***x3-**6***x5-**3.1***x7-**5.7***x10+**3***rannor(**0**) with n=20, rep=100.
*We need change "Do I = **1** to **20**;" to 25, 100, 1000. and the rep from 100 to 1000.

```
Data Design;
Do I = 1 to 20;
x1=10+5*rannor(0); * normal(10,25);
x2=exp(3*rannor(0)); * lognormal;
x3=5+10*ranuni(0); * uniform;
x4=50+10*rannor(0); * normal(50,100);
x5=x1+x4+rannor(0); *normal bimodal;
x6=5+2*x2+3*ranexp(0); *lognormal and exponential mixture;
```

```
x7=0.5*exp(4*rannor(0)); *lognormal;
x8=10+8*ranuni(0); * uniform;
x9=x2+x8+2*rannor(0); * lognormal, uniform and normal mix;
x10=20+x7+9*rannor(0); * lognormal and normal mix;
output; End;
drop i;run;
Data One;Set Design;
Do Rep = 1 to 100;
y = 33.5 +3.2*x3-6*x5-3.1*x7-5.7*x10+3*rannor(0); *true model;
OutPut;  End;
Proc Sort; By Rep; *Proc Print;*Var x1-x10 y;
proc reg Outest=Stat NoPrint; By Rep;
model y=x1-x10/Selection=adjRsq aic bic sbc;
run;
proc sort; by Rep _aic_; *Proc Print; *Var  Rep  _aic_;
Data Final1;Set Stat;
m = Mod(_n_, 1023); *Proc Print; *Var rep x1-x10 _aic_;
Data AIC; Set Final1;
If m=1;
If x1 = "." and x2 = "." and x3 ne "." and x4 = "." and x5 ne "." and
x6 = "." and x7 ne "." and x8 = "." and x9 = "." and x10 ne "." Then x35710+1;
Data Final;Set AIC;
If Rep=100;
proc print; Var m Rep  _aic_ x35710;
proc sort Data=Stat; by Rep _bic_; *Proc Print; *Var  Rep  _bic_;
Data Final2;Set Stat;
m = Mod(_n_, 1023); *Proc Print; *Var rep x1-x10 _bic_;
Data BIC; Set Final2;
If m=1;
If x1 = "." and x2 = "." and x3 ne "." and x4 = "." and x5 ne "." and
x6 = "." and x7 ne "." and x8 = "." and x9 = "." and x10 ne "." Then x35710+1;
Data Final;Set BIC;
If Rep=100;
proc print; Var m Rep  _bic_  x35710;
proc sort Data=Stat; by Rep _sbc_; *Proc Print; *Var  Rep  _sbc_;
Data Final3;Set Stat;
m = Mod(_n_, 1023); *Proc Print; *Var rep x1-x10 _sbc_;
Data SBC; Set Final3;
If m=1;
If x1 = "." and x2 = "." and x3 ne "." and x4 = "." and x5 ne "." and
x6 = "." and x7 ne "." and x8 = "." and x9 = "." and x10 ne "." Then x35710+1;
Data Final;Set SBC;
If Rep=100;
proc print; Var m Rep  _sbc_ x35710;
Run;
```

SAS program for testing true model

y = **33.5**+**3.2***x1 +**3.5***x2-**4***x9-**5.7***x10+**3***rannor(**0**) with n=20, rep=100.

*We need change “Do I = 1 to 20;” to 25, 100, 1000. and the rep from 100 to 1000.

```
Data Design;
Do I = 1 to 20;
x1=10+5*rannor(0); * normal(10,25);
x2=exp(3*rannor(0)); * lognormal;
x3=5+10*ranuni(0); * uniform;
x4=50+10*rannor(0); * normal(50,100);
x5=x1+x4+rannor(0); *normal bimodal;
x6=5+2*x2+3*ranexp(0); *lognormal and exponential mixture;
x7=0.5*exp(4*rannor(0)); *lognormal;
x8=10+8*ranuni(0); * uniform;
x9=x2+x8+2*rannor(0); * lognormal, uniform and normal mix;
x10=20+x7+9*rannor(0); * lognormal and normal mix;
output; End; drop i;run;
Data One;Set Design;
Do Rep = 1 to 100;
y = 33.5+3.2*x1 +3.5*x2-4*x9-5.7*x10+3*rannor(0); *true model;
OutPut;  End;
Proc Sort; By Rep; *Proc Print;*Var x1-x10 y;
proc reg Outest=Stat NoPrint; By Rep;
model y=x1-x10/Selection=adjRsq aic bic sbc;
run;
proc sort; by Rep _aic_; *Proc Print; *Var  Rep  _aic_;
Data Final1;Set Stat;
m = Mod(_n_, 1023); *Proc Print; *Var rep x1-x10 _aic_;
Data AIC; Set Final1;
If m=1;
If x1 ne "." and x2 ne "." and x3 = "." and x4 = "." and x5 = "." and
x6 = "." and x7 = "." and x8 = "." and x9 ne "." and x10 ne "." Then x12910+1;
Data Final;Set AIC;
If Rep=100;
proc print; Var m Rep  _aic_  x12910;
proc sort Data=Stat; by Rep _bic_; *Proc Print; *Var  Rep  _bic_;
Data Final2;Set Stat;
m = Mod(_n_, 1023); *Proc Print; *Var rep x1-x10 _bic_;
Data BIC; Set Final2;
If m=1;
If x1 ne "." and x2 ne "." and x3 = "." and x4 = "." and x5 = "." and
x6 = "." and x7 = "." and x8 = "." and x9 ne "." and x10 ne "." Then x12910+1;
Data Final;Set BIC;
If Rep=100;
proc print; Var m Rep  _bic_  x12910;
proc sort Data=Stat; by Rep _sbc_; *Proc Print; *Var  Rep  _sbc_;
Data Final3;Set Stat;
m = Mod(_n_, 1023); *Proc Print; *Var rep x1-x10 _sbc_;
Data SBC; Set Final3;
If m=1;
If x1 ne "." and x2 ne "." and x3 = "." and x4 = "." and x5 = "." and
x6 = "." and x7 = "." and x8 = "." and x9 ne "." and x10 ne "." Then x12910+1;
Data Final;Set SBC;
```

```
If Rep=100;
proc print; Var m Rep  _sbc_ x12910;
Run;
```

SAS program for testing true model
y = 33.5 + 3*x1-4*x2+5*x3+3.2*x4-2.2*x5+3*rannor(0) with n=20, rep=100.
*We need change “Do I = 1 to 20;” to 25, 100, 1000. and the rep from 100 to 1000.

```
Data Design;
Do I = 1 to 20;
x1=10+5*rannor(0); * normal(10,25);
x2=exp(3*rannor(0)); * lognormal;
x3=5+10*ranuni(0); * uniform;
x4=50+10*rannor(0); * normal(50,100);
x5=x1+x4+rannor(0); *normal bimodal;
x6=5+2*x2+3*ranexp(0); *lognormal and exponential mixture;
x7=0.5*exp(4*rannor(0)); *lognormal;
x8=10+8*ranuni(0); * uniform;
x9=x2+x8+2*rannor(0); * lognormal, uniform and normal mix;
x10=20+x7+9*rannor(0); * lognormal and normal mix;
output; End; drop i;run;
Data One;Set Design;
Do Rep = 1 to 100;
y = 33.5 + 3*x1-4*x2+5*x3+3.2*x4-2.2*x5+3*rannor(0); *true model;
OutPut;  End;
Proc Sort; By Rep; *Proc Print;*Var x1-x10 y;
proc reg Outest=Stat NoPrint; By Rep;
model y=x1-x10/Selection=adjRsq aic bic sbc;
run;
proc sort; by Rep _aic_; *Proc Print; *Var  Rep  _aic_;
Data Final1;Set Stat;
m = Mod(_n_, 1023); *Proc Print; *Var rep x1-x10 _aic_;
Data AIC; Set Final1;
If m=1;
If x1 ne "." and x2 ne "." and x3 ne "." and x4 ne "." and x5 ne "." and x6 = "." and x7
= "." and x8 = "." and x9 = "." and x10 = "." Then x12345+1;
Data Final;Set AIC;
If Rep=100;
proc print; Var m Rep  _aic_ x12345;
proc sort Data=Stat; by Rep _bic_; *Proc Print; *Var  Rep  _bic_;
Data Final2;Set Stat;
m = Mod(_n_, 1023); *Proc Print; *Var rep x1-x10 _bic_;
Data BIC; Set Final2;
If m=1;
If x1 ne "." and x2 ne "." and x3 ne "." and x4 ne "." and x5 ne "." and x6 = "." and x7
= "." and x8 = "." and x9 = "." and x10 = "." Then x12345+1;
Data Final;Set BIC;
If Rep=100;
proc print; Var m Rep  _bic_ x12345;
```

```
proc sort Data=Stat; by Rep _sbc_; *Proc Print; *Var  Rep  _sbc_;
Data Final3;Set Stat;
m = Mod(_n_, 1023); *Proc Print; *Var rep x1-x10 _sbc_;
Data SBC; Set Final3;
If m=1;
If x1 ne "." and x2 ne "." and x3 ne "." and x4 ne "." and x5 ne "." and x6 = "." and x7
= "." and x8 = "." and x9 = "." and x10 = "." Then x12345+1;
Data Final;Set SBC;
If Rep=100;
proc print; Var m Rep  _sbc_ x12345;
Run;
```

SAS program for testing true model
y = 33.5 + 3*x2-4*x3+5*x5+3.2*x8-2.2*x9+3*rannor(0) with n=20, rep=100.
*We need change “Do I = 1 to 20;” to 25, 100, 1000. and the rep from 100 to 1000.

```
Data Design;
Do I = 1 to 20;
x1=10+5*rannor(0); * normal(10,25);
x2=exp(3*rannor(0)); * lognormal;
x3=5+10*ranuni(0); * uniform;
x4=50+10*rannor(0); * normal(50,100);
x5=x1+x4+rannor(0); *normal bimodal;
x6=5+2*x2+3*ranexp(0); *lognormal and exponential mixture;
x7=0.5*exp(4*rannor(0)); *lognormal;
x8=10+8*ranuni(0); * uniform;
x9=x2+x8+2*rannor(0); * lognormal, uniform and normal mix;
x10=20+x7+9*rannor(0); * lognormal and normal mix;
output; End; drop i;run;
Data One;Set Design;
Do Rep = 1 to 100;
y = 33.5 + 3*x2-4*x3+5*x5+3.2*x8-2.2*x9+3*rannor(0); *true model;
OutPut;  End;
Proc Sort; By Rep; *Proc Print;*Var x1-x10 y;
proc reg Outest=Stat NoPrint; By Rep;
model y=x1-x10/Selection=adjRsq aic bic sbc;
run;
proc sort; by Rep _aic_; *Proc Print; *Var  Rep  _aic_;
Data Final1;Set Stat;
m = Mod(_n_, 1023); *Proc Print; *Var rep x1-x10 _aic_;
Data AIC; Set Final1;
If m=1;
If x1 = "." and x2 ne "." and x3 ne "." and x4 = "." and x5 ne "." and x6 = "." and x7 =
"." and x8 ne "." and x9 ne "." and x10 = "." Then x23589+1;
Data Final;Set AIC;
If Rep=100;
proc print; Var m Rep  _aic_ x23589;
proc sort Data=Stat; by Rep _bic_;
*Proc Print; *Var  Rep  _bic_;
```

```
Data Final2;Set Stat;
m = Mod(_n_, 1023); *Proc Print; *Var rep x1-x10 _bic_;
Data BIC; Set Final2;
If m=1;
If x1 = "." and x2 ne "." and x3 ne "." and x4 = "." and x5 ne "." and x6 = "." and x7 =
"." and x8 ne "." and x9 ne "." and x10 = "." Then x23589+1;
Data Final;Set BIC;
If Rep=100;
proc print; Var m Rep  _bic_ x23589;
proc sort Data=Stat; by Rep _sbc_; *Proc Print; *Var  Rep  _sbc_;
Data Final3;Set Stat;
m = Mod(_n_, 1023); *Proc Print; *Var rep x1-x10 _sbc_;
Data SBC; Set Final3;
If m=1;
 If x1 = "." and x2 ne "." and x3 ne "." and x4 = "." and x5 ne "." and x6 = "." and x7 =
"." and x8 ne "." and x9 ne "." and x10 = "." Then x23589+1;
Data Final;Set SBC;
If Rep=100;
proc print; Var m Rep  _sbc_ x23589;
Run;
```

SAS program for testing true model
y = **33.5 -2.7***x3-**4***x5-**1.5***x7+**3.2***x9-**2.2***x10+**3***rannor(**0**) with n=20, rep=100.
*We need change "Do I = 1 to 20;" to 25, 100, 1000. and the rep from 100 to 1000.

```
Data Design;
Do I = 1 to 20;
x1=10+5*rannor(0); * normal(10,25);
x2=exp(3*rannor(0)); * lognormal;
x3=5+10*ranuni(0); * uniform;
x4=50+10*rannor(0); * normal(50,100);
x5=x1+x4+rannor(0); *normal bimodal;
x6=5+2*x2+3*ranexp(0); *lognormal and exponential mixture;
x7=0.5*exp(4*rannor(0)); *lognormal;
x8=10+8*ranuni(0); * uniform;
x9=x2+x8+2*rannor(0); * lognormal, uniform and normal mix;
x10=20+x7+9*rannor(0); * lognormal and normal mix;
output; End; drop i;run;
Data One;Set Design;
Do Rep = 1 to 100;
y = 33.5 -2.7*x3-4*x5-1.5*x7+3.2*x9-2.2*x10+3*rannor(0); *true model;
OutPut;  End;
Proc Sort; By Rep; *Proc Print;*Var x1-x10 y;
proc reg Outest=Stat NoPrint; By Rep;
model y=x1-x10/Selection=adjRsq aic bic sbc;
run;
proc sort; by Rep _aic_; *Proc Print; *Var  Rep  _aic_;
Data Final1;Set Stat;
m = Mod(_n_, 1023); *Proc Print; *Var rep x1-x10 _aic_;
```

```
Data AIC; Set Final1;
If m=1;
 If x1 = "." and x2 = "." and x3 ne "." and x4 = "." and x5 ne "." and x6 = "." and x7 ne
"." and x8 = "." and x9 ne "." and x10 ne "." Then x357910+1;
Data Final;Set AIC;
If Rep=100;
proc print; Var m Rep  _aic_ x357910;
proc sort Data=Stat; by Rep _bic_; *Proc Print; *Var  Rep  _bic_;
Data Final2;Set Stat;
m = Mod(_n_, 1023); *Proc Print; *Var rep x1-x10 _bic_;
Data BIC; Set Final2;
If m=1;
 If x1 = "." and x2 = "." and x3 ne "." and x4 = "." and x5 ne "." and x6 = "." and x7 ne
"." and x8 = "." and x9 ne "." and x10 ne "." Then x357910+1;
Data Final;Set BIC;
If Rep=100;
proc print; Var m Rep  _bic_ x357910;
proc sort Data=Stat; by Rep _sbc_; *Proc Print; *Var  Rep  _sbc_;
Data Final3;Set Stat;
m = Mod(_n_, 1023); *Proc Print; *Var rep x1-x10 _sbc_;
Data SBC; Set Final3;
If m=1;
  If x1 = "." and x2 = "." and x3 ne "." and x4 = "." and x5 ne "." and x6 = "." and x7
ne "." and x8 = "." and x9 ne "." and x10 ne "." Then x357910+1;
Data Final;Set SBC;
If Rep=100;
proc print; Var m Rep  _sbc_ x357910;
Run;
```

SAS program for testing true model
y = **33.5** +**3.1***x1-**5.5***x3+**4***x5+**3.2***x7+**2.2***x8-**6***x9+**3***rannor(**0**) with n=20, rep=100.
*We need change “Do I = **1** to **20**;” to 25, 100, 1000. and the rep from 100 to 1000.

```
Data Design;
Do I = 1 to 20;
x1=10+5*rannor(0); * normal(10,25);
x2=exp(3*rannor(0)); * lognormal;
x3=5+10*ranuni(0); * uniform;
x4=50+10*rannor(0); * normal(50,100);
x5=x1+x4+rannor(0); *normal bimodal;
x6=5+2*x2+3*ranexp(0); *lognormal and exponential mixture;
x7=0.5*exp(4*rannor(0)); *lognormal;
x8=10+8*ranuni(0); * uniform;
x9=x2+x8+2*rannor(0); * lognormal, uniform and normal mix;
x10=20+x7+9*rannor(0); * lognormal and normal mix;
output; End; drop i;run;
Data One;Set Design;
Do Rep = 1 to 100;
```

```
y = 33.5 +3.1*x1-5.5*x3+4*x5+3.2*x7+2.2*x8-6*x9+3*rannor(0);*true model;
OutPut;  End;
Proc Sort; By Rep; *Proc Print;*Var x1-x10 y;
proc reg Outest=Stat NoPrint; By Rep;
model y=x1-x10/Selection=adjRsq aic bic sbc;
run;
proc sort; by Rep _aic_; *Proc Print; *Var  Rep  _aic_;
Data Final1;Set Stat;
m = Mod(_n_, 1023); *Proc Print; *Var rep x1-x10 _aic_;
Data AIC; Set Final1;
If m=1;
 If x1 ne "." and x2 = "." and x3 ne "." and x4 = "." and x5 ne "." and x6 = "." and x7
ne "." and x8 ne "." and x9 ne "." and x10 = "." Then x135789+1;
Data Final;Set AIC;
If Rep=100;
proc print; Var m Rep  _aic_ x135789;
proc sort Data=Stat; by Rep _bic_; *Proc Print; *Var  Rep  _bic_;
Data Final2;Set Stat;
m = Mod(_n_, 1023); *Proc Print; *Var rep x1-x10 _bic_;
Data BIC; Set Final2;
If m=1;
If x1 ne "." and x2 = "." and x3 ne "." and x4 = "." and x5 ne "." and x6 = "." and x7
ne "." and x8 ne "." and x9 ne "." and x10 = "." Then x135789+1;
Data Final;Set BIC;
If Rep=100;
proc print; Var m Rep  _bic_ x135789;
proc sort Data=Stat; by Rep _sbc_; *Proc Print; *Var  Rep  _sbc_;
Data Final3;Set Stat;
m = Mod(_n_, 1023); *Proc Print; *Var rep x1-x10 _sbc_;
Data SBC; Set Final3;
If m=1;
If x1 ne "." and x2 = "." and x3 ne "." and x4 = "." and x5 ne "." and x6 = "." and x7
ne "." and x8 ne "." and x9 ne "." and x10 = "." Then x135789+1;
Data Final;Set SBC;
If Rep=100;
proc print; Var m Rep  _sbc_ x135789;
Run;
```

SAS program for testing true model

y = **33.5 -2.1***x1-**5.5***x3+**4***x6+**3.2***x7-**2.2***x9+**9***x10+**3***rannor(**0**) with n=20, rep=100.

*We need change "Do I = **1** to **20**;" to 25, 100, 1000. and the rep from 100 to 1000.

```
Data Design;
Do I = 1 to 20;
x1=10+5*rannor(0); * normal(10,25);
x2=exp(3*rannor(0)); * lognormal;
x3=5+10*ranuni(0); * uniform;
x4=50+10*rannor(0); * normal(50,100);
```

```
x5=x1+x4+rannor(0); *normal bimodal;
x6=5+2*x2+3*ranexp(0); *lognormal and exponential mixture;
x7=0.5*exp(4*rannor(0)); *lognormal;
x8=10+8*ranuni(0); * uniform;
x9=x2+x8+2*rannor(0); * lognormal, uniform and normal mix;
x10=20+x7+9*rannor(0); * lognormal and normal mix;
output;End;drop i;run;
Data One;Set Design;
Do Rep = 1 to 100;
y = 33.5 -2.1*x1-5.5*x3+4*x6+3.2*x7-2.2*x9+9*x10+3*rannor(0); *true
model;OutPut; End;
Proc Sort; By Rep; *Proc Print;*Var x1-x10 y;
proc reg Outest=Stat NoPrint; By Rep;
model y=x1-x10/Selection=adjRsq aic bic sbc;
run;
proc sort; by Rep _aic_; *Proc Print; *Var  Rep  _aic_;
Data Final1;Set Stat;
m = Mod(_n_, 1023); *Proc Print; *Var rep x1-x10 _aic_;
Data AIC; Set Final1;
If m=1;
 If x1 ne "." and x2 = "." and x3 ne "." and x4 = "." and x5 = "." and x6 ne "." and x7
ne "." and x8 = "." and x9 ne "." and x10 ne "." Then x1367910+1;
Data Final;Set AIC;
If Rep=100;
proc print; Var m Rep  _aic_ x1367910;
proc sort Data=Stat; by Rep _bic_; *Proc Print; *Var  Rep  _bic_;
Data Final2;Set Stat;
m = Mod(_n_, 1023); *Proc Print; *Var rep x1-x10 _bic_;
Data BIC; Set Final2;
If m=1;
 If x1 ne "." and x2 = "." and x3 ne "." and x4 = "." and x5 = "." and x6 ne "." and x7
ne "." and x8 = "." and x9 ne "." and x10 ne "." Then x1367910+1;
Data Final;Set BIC;
If Rep=100;
proc print; Var m Rep  _bic_ x1367910;
proc sort Data=Stat; by Rep _sbc_; *Proc Print; *Var  Rep  _sbc_;
Data Final3;Set Stat;
m = Mod(_n_, 1023); *Proc Print; *Var rep x1-x10 _sbc_;
Data SBC; Set Final3;
If m=1;
 If x1 ne "." and x2 = "." and x3 ne "." and x4 = "." and x5 = "." and x6 ne "." and x7
ne "." and x8 = "." and x9 ne "." and x10 ne "." Then x1367910+1;
Data Final;Set SBC;
If Rep=100;
proc print; Var m Rep  _sbc_ x1367910;
Run;
```

SAS program for testing true model

y = **33.5** + **3***x2-**4***x3+**5***x5+**3.2***x8-**2.2***x9+**9***x10+**3***rannor(**0**) with n=20, rep=100.

*We need change “Do I = 1 to 20;” to 25, 100, 1000. and the rep from 100 to 1000.

```
Data Design;
Do I = 1 to 20;
x1=10+5*rannor(0); * normal(10,25);
x2=exp(3*rannor(0)); * lognormal;
x3=5+10*ranuni(0); * uniform;
x4=50+10*rannor(0); * normal(50,100);
x5=x1+x4+rannor(0); *normal bimodal;
x6=5+2*x2+3*ranexp(0); *lognormal and exponential mixture;
x7=0.5*exp(4*rannor(0)); *lognormal;
x8=10+8*ranuni(0); * uniform;
x9=x2+x8+2*rannor(0); * lognormal, uniform and normal mix;
x10=20+x7+9*rannor(0); * lognormal and normal mix;
output; End; drop i;run;
Data One;Set Design;
Do Rep = 1 to 100;
y = 33.5 + 3*x2-4*x3+5*x5+3.2*x8-2.2*x9+9*x10+3*rannor(0); *true model;
OutPut;  End;
Proc Sort; By Rep; *Proc Print;*Var x1-x10 y;
proc reg Outest=Stat NoPrint; By Rep;
model y=x1-x10/Selection=adjRsq aic bic sbc;
run;
proc sort; by Rep _aic_; *Proc Print; *Var  Rep  _aic_;
Data Final1;Set Stat;
m = Mod(_n_, 1023); *Proc Print; *Var rep x1-x10 _aic_;
Data AIC; Set Final1;
If m=1;
 If x1 = "." and x2 ne "." and x3 ne "." and x4 = "." and x5 ne "." and x6 = "." and x7 =
"." and x8 ne "." and x9 ne "." and x10 ne "." Then x2358910+1;
Data Final;Set AIC;
If Rep=100;
proc print; Var m Rep  _aic_ x2358910;
proc sort Data=Stat; by Rep _bic_; *Proc Print; *Var  Rep  _bic_;
Data Final2;Set Stat;
m = Mod(_n_, 1023); *Proc Print; *Var rep x1-x10 _bic_;
Data BIC; Set Final2;
If m=1;
If x1 = "." and x2 ne "." and x3 ne "." and x4 = "." and x5 ne "." and
x6 = "." and x7 = "." and x8 ne "." and x9 ne "." and x10 ne "." Then x2358910+1;
Data Final;Set BIC;
If Rep=100;
proc print; Var m Rep  _bic_ x2358910;
proc sort Data=Stat; by Rep _sbc_; *Proc Print; *Var  Rep  _sbc_;
Data Final3;Set Stat;
m = Mod(_n_, 1023); *Proc Print; *Var rep x1-x10 _sbc_;
Data SBC; Set Final3;
If m=1;
If x1 = "." and x2 ne "." and x3 ne "." and x4 = "." and x5 ne "." and x6 = "." and x7 =
"." and x8 ne "." and x9 ne "." and x10 ne "." Then x2358910+1;
```

```
Data Final;Set SBC;
If Rep=100;
proc print; Var m Rep  _sbc_ x2358910;
Run;
```

SAS program for testing true model
y = **33.5** +**3.1***x1-**4.4***x2+**5.5***x3+**4***x5+**3.2***x7+**2.2***x8-**6***x9+**3***rannor(**0**) with n=20, rep=100.
*We need change “Do I = **1** to **20**;” to 25, 100, 1000. and the rep from 100 to 1000.

```
Data Design;
Do I = 1 to 20;
x1=10+5*rannor(0); * normal(10,25);
x2=exp(3*rannor(0)); * lognormal;
x3=5+10*ranuni(0); * uniform;
x4=50+10*rannor(0); * normal(50,100);
x5=x1+x4+rannor(0); *normal bimodal;
x6=5+2*x2+3*ranexp(0); *lognormal and exponential mixture;
x7=0.5*exp(4*rannor(0)); *lognormal;
x8=10+8*ranuni(0); * uniform;
x9=x2+x8+2*rannor(0); * lognormal, uniform and normal mix;
x10=20+x7+9*rannor(0); * lognormal and normal mix;
output; End; drop i;run;
Data One;Set Design;
Do Rep = 1 to 100;
y = 33.5 +3.1*x1-4.4*x2+5.5*x3+4*x5+3.2*x7+2.2*x8-6*x9+3*rannor(0); *true
model; OutPut;  End;
Proc Sort; By Rep;  *Proc Print;*Var x1-x10 y;
proc reg Outest=Stat NoPrint; By Rep;
model y=x1-x10/Selection=adjRsq aic bic sbc;
run;
proc sort; by Rep _aic_; *Proc Print; *Var  Rep  _aic_;
Data Final1;Set Stat;
m = Mod(_n_, 1023); *Proc Print; *Var rep x1-x10 _aic_;
Data AIC; Set Final1;
If m=1;
 If x1 ne "." and x2 ne "." and x3 ne "." and x4 = "." and x5 ne "." and x6 = "." and x7
ne "." and x8 ne "." and x9 ne "." and x10 = "." Then x1235789+1;
Data Final;Set AIC;
If Rep=100;
proc print; Var m Rep  _aic_ x1235789;
proc sort Data=Stat; by Rep _bic_; *Proc Print; *Var  Rep  _bic_;
Data Final2;Set Stat;
m = Mod(_n_, 1023); *Proc Print; *Var rep x1-x10 _bic_;
Data BIC; Set Final2;
If m=1;
 If x1 ne "." and x2 ne "." and x3 ne "." and x4 = "." and x5 ne "." and x6 = "." and x7
ne "." and x8 ne "." and x9 ne "." and x10 = "." Then x1235789+1;
Data Final;Set BIC;
```

```
If Rep=100;
proc print; Var m Rep  _bic_ x1235789;
proc sort Data=Stat; by Rep _sbc_; *Proc Print; *Var  Rep  _sbc_;
Data Final3;Set Stat;
m = Mod(_n_, 1023); *Proc Print; *Var rep x1-x10 _sbc_;
Data SBC; Set Final3;
If m=1;
 If x1 ne "." and x2 ne "." and x3 ne "." and x4 = "." and x5 ne "." and x6 = "." and x7
ne "." and x8 ne "." and x9 ne "." and x10 = "." Then x1235789+1;
Data Final;Set SBC;
If Rep=100;
proc print; Var m Rep  _sbc_ x1235789;
Run;
```

SAS program for testing true model

y = **33.5** +**3.1***x1+**7.4***x3+**3.5***x4-**4***x5+**3.2***x6+**2.2***x7-**6***x8+**3***rannor(**0**) with n=20, rep=100.

*We need change "Do I = **1** to **20**;" to 25, 100, 1000. and the rep from 100 to 1000.

```
Data Design;
Do I = 1 to 20;
x1=10+5*rannor(0); * normal(10,25);
x2=exp(3*rannor(0)); * lognormal;
x3=5+10*ranuni(0); * uniform;
x4=50+10*rannor(0); * normal(50,100);
x5=x1+x4+rannor(0); *normal bimodal;
x6=5+2*x2+3*ranexp(0); *lognormal and exponential mixture;
x7=0.5*exp(4*rannor(0)); *lognormal;
x8=10+8*ranuni(0); * uniform;
x9=x2+x8+2*rannor(0); * lognormal, uniform and normal mix;
x10=20+x7+9*rannor(0); * lognormal and normal mix;
output; End; drop i;run;
Data One;Set Design;
Do Rep = 1 to 100;
y = 33.5 +3.1*x1+7.4*x3+3.5*x4-4*x5+3.2*x6+2.2*x7-6*x8+3*rannor(0); *true
model; OutPut;  End;
Proc Sort; By Rep; *Proc Print;*Var x1-x10 y;
proc reg Outest=Stat NoPrint; By Rep;
model y=x1-x10/Selection=adjRsq aic bic sbc;
run;
proc sort; by Rep _aic_; *Proc Print; *Var  Rep  _aic_;
Data Final1;Set Stat;
m = Mod(_n_, 1023); *Proc Print; *Var rep x1-x10 _aic_;
Data AIC; Set Final1;
If m=1;
 If x1 ne "." and x2 = "." and x3 ne "." and x4 ne "." and x5 ne "." and x6 ne "." and x7
ne "." and x8 ne "." and x9 = "." and x10 = "." Then x1345678+1;
Data Final;Set AIC;
If Rep=100;
```

```
proc print; Var m Rep  _aic_ x1345678;
proc sort Data=Stat; by Rep _bic_; *Proc Print; *Var  Rep  _bic_;
Data Final2;Set Stat;
m = Mod(_n_, 1023); *Proc Print; *Var rep x1-x10 _bic_;
Data BIC; Set Final2;
If m=1;
 If x1 ne "." and x2 = "." and x3 ne "." and x4 ne "." and x5 ne "." and x6 ne "." and x7
ne "." and x8 ne "." and x9 = "." and x10 = "." Then x1345678+1;
Data Final;Set BIC;
If Rep=100;
proc print; Var m Rep  _bic_ x1345678;
proc sort Data=Stat; by Rep _sbc_; *Proc Print; *Var  Rep  _sbc_;
Data Final3;Set Stat;
m = Mod(_n_, 1023); *Proc Print; *Var rep x1-x10 _sbc_;
Data SBC; Set Final3;
If m=1;
 If x1 ne "." and x2 = "." and x3 ne "." and x4 ne "." and x5 ne "." and x6 ne "." and x7
ne "." and x8 ne "." and x9 = "." and x10 = "." Then x1345678+1;
Data Final;Set SBC;
If Rep=100;
proc print; Var m Rep  _sbc_ x1345678;
Run;
```

SAS program for testing true model
y = **33.5** +**3.1***x2-**4.4***x3+**5.5***x4+**4***x5+**3.2***x7+**2.2***x8-**6***x10+**3***rannor(**0**) with n=20, rep=100.
*We need change "Do I = **1** to **20**;" to 25, 100, 1000. and the rep from 100 to 1000.

```
Data Design;
Do I = 1 to 20;
x1=10+5*rannor(0); * normal(10,25);
x2=exp(3*rannor(0)); * lognormal;
x3=5+10*ranuni(0); * uniform;
x4=50+10*rannor(0); * normal(50,100);
x5=x1+x4+rannor(0); *normal bimodal;
x6=5+2*x2+3*ranexp(0); *lognormal and exponential mixture;
x7=0.5*exp(4*rannor(0)); *lognormal;
x8=10+8*ranuni(0); * uniform;
x9=x2+x8+2*rannor(0); * lognormal, uniform and normal mix;
x10=20+x7+9*rannor(0); * lognormal and normal mix;
output; End; drop i;run;
Data One;Set Design;
Do Rep = 1 to 100;
y = 33.5 +3.1*x2-4.4*x3+5.5*x4+4*x5+3.2*x7+2.2*x8-6*x10+3*rannor(0); *true
model; OutPut;  End;
Proc Sort; By Rep; *Proc Print;*Var x1-x10 y;
proc reg Outest=Stat NoPrint; By Rep;
model y=x1-x10/Selection=adjRsq aic bic sbc;
run;
```

```
proc sort; by Rep _aic_; *Proc Print; *Var  Rep  _aic_;
Data Final1;Set Stat;
m = Mod(_n_, 1023); *Proc Print; *Var rep x1-x10 _aic_;
Data AIC; Set Final1;
If m=1;
 If x1 = "." and x2 ne "." and x3 ne "." and x4 ne "." and x5 ne "." and x6 = "." and x7
ne "." and x8 ne "." and x9 = "." and x10 ne "." Then x23457810+1;
Data Final;Set AIC;
If Rep=100;
proc print; Var m Rep  _aic_ x23457810;
proc sort Data=Stat; by Rep _bic_; *Proc Print; *Var  Rep  _bic_;
Data Final2;Set Stat;
m = Mod(_n_, 1023); *Proc Print; *Var rep x1-x10 _bic_;
Data BIC; Set Final2;
If m=1;
  If x1 = "." and x2 ne "." and x3 ne "." and x4 ne "." and x5 ne "." and x6 = "." and x7
ne "." and x8 ne "." and x9 = "." and x10 ne "." Then x23457810+1;
Data Final;Set BIC;
If Rep=100;
proc print; Var m Rep  _bic_ x23457810;
proc sort Data=Stat; by Rep _sbc_; *Proc Print; *Var  Rep  _sbc_;
Data Final3;Set Stat;
m = Mod(_n_, 1023); *Proc Print; *Var rep x1-x10 _sbc_;
Data SBC; Set Final3;
If m=1;
 If x1 = "." and x2 ne "." and x3 ne "." and x4 ne "." and x5 ne "." and x6 = "." and x7
ne "." and x8 ne "." and x9 = "." and x10 ne "." Then x23457810+1;
Data Final;Set SBC;
If Rep=100;
proc print; Var m Rep  _sbc_ x23457810;
Run;
```

SAS program for testing true model
y = **33.5** -**9***x1+**3.1***x2+**7.4***x3+**3.5***x4-**4***x5+**3.2***x6+**2.2***x7-**6***x8+**3***rannor(**0**) with n=20, rep=100.
*We need change “Do I = 1 to 20;” to 25, 100, 1000. and the rep from 100 to 1000.

```
Data Design;
Do I = 1 to 20;
x1=10+5*rannor(0); * normal(10,25);
x2=exp(3*rannor(0)); * lognormal;
x3=5+10*ranuni(0); * uniform;
x4=50+10*rannor(0); * normal(50,100);
x5=x1+x4+rannor(0); *normal bimodal;
x6=5+2*x2+3*ranexp(0); *lognormal and exponential mixture;
x7=0.5*exp(4*rannor(0)); *lognormal;
x8=10+8*ranuni(0); * uniform;
x9=x2+x8+2*rannor(0); * lognormal, uniform and normal mix;
x10=20+x7+9*rannor(0); * lognormal and normal mix;
```

```
output; End; drop i;run;
Data One;Set Design;
Do Rep = 1 to 100;
y = 33.5 -9*x1+3.1*x2+7.4*x3+3.5*x4-4*x5+3.2*x6+2.2*x7-6*x8+3*rannor(0);
*true model; OutPut;  End;
Proc Sort; By Rep; *Proc Print;*Var x1-x10 y;
proc reg Outest=Stat NoPrint; By Rep;
model y=x1-x10/Selection=adjRsq aic bic sbc;
run;
proc sort; by Rep _aic_; *Proc Print; *Var  Rep  _aic_;
Data Final1;Set Stat;
m = Mod(_n_, 1023); *Proc Print; *Var rep x1-x10 _aic_;
Data AIC; Set Final1;
If m=1;
 If x1 ne "." and x2 ne "." and x3 ne "." and x4 ne "." and x5 ne "." and x6 ne "." and
x7 ne "." and x8 ne "." and x9 = "." and x10 = "." Then x12345678+1;
Data Final;Set AIC;
If Rep=100;
proc print; Var m Rep  _aic_ x12345678;
proc sort Data=Stat; by Rep _bic_; *Proc Print; *Var  Rep  _bic_;
Data Final2;Set Stat;
m = Mod(_n_, 1023); *Proc Print; *Var rep x1-x10 _bic_;
Data BIC; Set Final2;
If m=1;
 If x1 ne "." and x2 ne "." and x3 ne "." and x4 ne "." and x5 ne "." and x6 ne "." and
x7 ne "." and x8 ne "." and x9 = "." and x10 = "." Then x12345678+1;
Data Final;Set BIC;
If Rep=100;
proc print; Var m Rep  _bic_ x12345678;
proc sort Data=Stat; by Rep _sbc_; *Proc Print; *Var  Rep  _sbc_;
Data Final3;Set Stat;
m = Mod(_n_, 1023); *Proc Print; *Var rep x1-x10 _sbc_;
Data SBC; Set Final3;
If m=1;
 If x1 ne "." and x2 ne "." and x3 ne "." and x4 ne "." and x5 ne "." and x6 ne "." and
x7 ne "." and x8 ne "." and x9 = "." and x10 = "." Then x12345678+1;
Data Final;Set SBC;
If Rep=100;
proc print; Var m Rep  _sbc_ x12345678;
Run;
```

SAS program for testing true model
y = **33.5** -**4***x1+**3.2***x3+**2.2***x4-**6***x5-**9***x6-**3.1***x7+**7.4***x8+**3.5***x9+**3***rannor(**0**) with n=20, rep=100.
*We need change “Do I = **1** to **20**;” to 25, 100, 1000. and the rep from 100 to 1000.

```
Data Design;
Do I = 1 to 20;
x1=10+5*rannor(0); * normal(10,25);
```

```
x2=exp(3*rannor(0)); * lognormal;
x3=5+10*ranuni(0); * uniform;
x4=50+10*rannor(0); * normal(50,100);
x5=x1+x4+rannor(0); *normal bimodal;
x6=5+2*x2+3*ranexp(0); *lognormal and exponential mixture;
x7=0.5*exp(4*rannor(0)); *lognormal;
x8=10+8*ranuni(0); * uniform;
x9=x2+x8+2*rannor(0); * lognormal, uniform and normal mix;
x10=20+x7+9*rannor(0); * lognormal and normal mix;
output; End; drop i;run;
Data One;Set Design;
Do Rep = 1 to 100;
y = 33.5 -4*x1+3.2*x3+2.2*x4-6*x5-9*x6-3.1*x7+7.4*x8+3.5*x9+3*rannor(0);
*true model;
OutPut;  End;
Proc Sort; By Rep; *Proc Print;*Var x1-x10 y;
proc reg Outest=Stat NoPrint; By Rep;
model y=x1-x10/Selection=adjRsq aic bic sbc;
run;
proc sort; by Rep _aic_; *Proc Print; *Var  Rep  _aic_;
Data Final1;Set Stat;
m = Mod(_n_, 1023); *Proc Print; *Var rep x1-x10 _aic_;
Data AIC; Set Final1;
If m=1;
 If x1 ne "." and x2 = "." and x3 ne "." and x4 ne "." and x5 ne "." and x6 ne "." and x7
ne "." and x8 ne "." and x9 ne "." and x10 = "." Then x13456789+1;
Data Final;Set AIC;
If Rep=100;
proc print; Var m Rep  _aic_ x13456789;
proc sort Data=Stat; by Rep _bic_; *Proc Print; *Var  Rep  _bic_;
Data Final2;Set Stat;
m = Mod(_n_, 1023); *Proc Print; *Var rep x1-x10 _bic_;
Data BIC; Set Final2;
If m=1;
If x1 ne "." and x2 = "." and x3 ne "." and x4 ne "." and x5 ne "." and x6 ne "." and x7
ne "." and x8 ne "." and x9 ne "." and x10 = "." Then x13456789+1;
Data Final;Set BIC;
If Rep=100;
proc print; Var m Rep  _bic_  x13456789;
proc sort Data=Stat; by Rep _sbc_; *Proc Print; *Var  Rep  _sbc_;
Data Final3;Set Stat;
m = Mod(_n_, 1023); *Proc Print; *Var rep x1-x10 _sbc_;
Data SBC; Set Final3;
If m=1;
If x1 ne "." and x2 = "." and x3 ne "." and x4 ne "." and x5 ne "." and x6 ne "." and x7
ne "." and x8 ne "." and x9 ne "." and x10 = "." Then x13456789+1;
Data Final;Set SBC;
If Rep=100;
proc print; Var m Rep  _sbc_ x13456789;
Run;
```

SAS program for testing true model
y = **33.5** +**3.1***x2+**7.4***x3+**3.5***x4-**4***x5+**3.2***x6+**2.2***x7-**6***x8-**9***x10+**3***rannor(**0**)
with n=20, rep=100.
*We need change "Do I = 1 to 20;" to 25, 100, 1000. and the rep from 100 to 1000.

```
Data Design;
Do I = 1 to 20;
x1=10+5*rannor(0); * normal(10,25);
x2=exp(3*rannor(0)); * lognormal;
x3=5+10*ranuni(0); * uniform;
x4=50+10*rannor(0); * normal(50,100);
x5=x1+x4+rannor(0); *normal bimodal;
x6=5+2*x2+3*ranexp(0); *lognormal and exponential mixture;
x7=0.5*exp(4*rannor(0)); *lognormal;
x8=10+8*ranuni(0); * uniform;
x9=x2+x8+2*rannor(0); * lognormal, uniform and normal mix;
x10=20+x7+9*rannor(0); * lognormal and normal mix;
output; End; drop i;run;
Data One;Set Design;
Do Rep = 1 to 100;
y = 33.5 +3.1*x2+7.4*x3+3.5*x4-4*x5+3.2*x6+2.2*x7-6*x8-9*x10+3*rannor(0);
*true model;
OutPut;  End;
Proc Sort; By Rep; *Proc Print;*Var x1-x10 y;
proc reg Outest=Stat NoPrint; By Rep;
model y=x1-x10/Selection=adjRsq aic bic sbc;
run;
proc sort; by Rep _aic_; *Proc Print; *Var  Rep  _aic_;
Data Final1;Set Stat;
m = Mod(_n_, 1023); *Proc Print; *Var rep x1-x10 _aic_;
Data AIC; Set Final1;
If m=1;
 If x1 = "." and x2 ne "." and x3 ne "." and x4 ne "." and x5 ne "." and x6 ne "." and x7
ne "." and x8 ne "." and x9 = "." and x10 ne "." Then x234567810+1;
Data Final;Set AIC;
If Rep=100;
proc print; Var m Rep  _aic_ x234567810;
proc sort Data=Stat; by Rep _bic_; *Proc Print; *Var  Rep  _bic_;
Data Final2;Set Stat;
m = Mod(_n_, 1023); *Proc Print; *Var rep x1-x10 _bic_;
Data BIC; Set Final2;
If m=1;
 If x1 = "." and x2 ne "." and x3 ne "." and x4 ne "." and x5 ne "." and x6 ne "." and x7
ne "." and x8 ne "." and x9 = "." and x10 ne "." Then x234567810+1;
Data Final;Set BIC;
If Rep=100;
proc print; Var m Rep  _bic_ x234567810;
proc sort Data=Stat; by Rep _sbc_; *Proc Print; *Var  Rep  _sbc_;
Data Final3;Set Stat;
m = Mod(_n_, 1023); *Proc Print; *Var rep x1-x10 _sbc_;
```

```
Data SBC; Set Final3;
If m=1;
 If x1 = "." and x2 ne "." and x3 ne "." and x4 ne "." and x5 ne "." and x6 ne "." and x7
ne "." and x8 ne "." and x9 = "." and x10 ne "." Then x234567810+1;
Data Final;Set SBC;
If Rep=100;
proc print; Var m Rep  _sbc_ x234567810;
Run;
```

SAS program for testing true model y = **33.5** -**4***x1+**3.5***x2+**3.2***x3+**2.2***x4-**6***x5-**9***x6-**3.1***x7+**7.4***x8-**5.7***x10+**3***rannor(**0**) with n=20, rep=100.
*We need change “Do I = **1** to **20**;” to 25, 100, 1000. and the rep from 100 to 1000.

```
Data Design;
Do I = 1 to 20;
x1=10+5*rannor(0); * normal(10,25);
x2=exp(3*rannor(0)); * lognormal;
x3=5+10*ranuni(0); * uniform;
x4=50+10*rannor(0); * normal(50,100);
x5=x1+x4+rannor(0); *normal bimodal;
x6=5+2*x2+3*ranexp(0); *lognormal and exponential mixture;
x7=0.5*exp(4*rannor(0)); *lognormal;
x8=10+8*ranuni(0); * uniform;
x9=x2+x8+2*rannor(0); * lognormal, uniform and normal mix;
x10=20+x7+9*rannor(0); * lognormal and normal mix;
output; End; drop i;run;
Data One;Set Design;
Do Rep = 1 to 100;
y = 33.5 -4*x1+3.5*x2+3.2*x3+2.2*x4-6*x5-9*x6-3.1*x7+7.4*x8-
5.7*x10+3*rannor(0); *true model;
OutPut;  End;
Proc Sort; By Rep; *Proc Print;*Var x1-x10 y;
proc reg Outest=Stat NoPrint; By Rep;
model y=x1-x10/Selection=adjRsq aic bic sbc;
run;
proc sort; by Rep _aic_; *Proc Print; *Var  Rep  _aic_;
Data Final1;Set Stat;
m = Mod(_n_, 1023); *Proc Print; *Var rep x1-x10 _aic_;
Data AIC; Set Final1;
If m=1;
 If x1 ne "." and x2 ne "." and x3 ne "." and x4 ne "." and x5 ne "." and x6 ne "." and
x7 ne "." and x8 ne "." and x9 = "." and x10 ne "." Then x1234567810+1;
Data Final;Set AIC;
If Rep=100;
proc print; Var m Rep  _aic_ x1234567810;
proc sort Data=Stat; by Rep _bic_; *Proc Print; *Var  Rep  _bic_;
Data Final2;Set Stat;
m = Mod(_n_, 1023); *Proc Print; *Var rep x1-x10 _bic_;
Data BIC; Set Final2;
```

```
If m=1;
If x1 ne "." and x2 ne "." and x3 ne "." and x4 ne "." and x5 ne "." and x6 ne "." and
x7 ne "." and x8 ne "." and x9 = "." and x10 ne "." Then x1234567810+1;
Data Final;Set BIC;
If Rep=100;
proc print; Var m Rep  _bic_  x1234567810;
proc sort Data=Stat; by Rep _sbc_; *Proc Print; *Var  Rep  _sbc_;
Data Final3;Set Stat;
m = Mod(_n_, 1023); *Proc Print; *Var rep x1-x10 _sbc_;
Data SBC; Set Final3;
If m=1;
If x1 ne "." and x2 ne "." and x3 ne "." and x4 ne "." and x5 ne "." and x6 ne "." and
x7 ne "." and x8 ne "." and x9 = "." and x10 ne "." Then x1234567810+1;
Data Final;Set SBC;
If Rep=100;
proc print; Var m Rep  _sbc_ x1234567810;
Run;
```

SAS program for testing true model y = **33.5**+**3.2***x1 +**3.5***x2+**2.2***x3-**6***x4-**9***x6-**3.1***x7+**7.4***x8-**4***x9-**5.7***x10+**3***rannor(**0**) with n=20, rep=100.
*We need change "Do I = 1 to 20;" to 25, 100, 1000. and the rep from 100 to 1000.

```
Data Design;
Do I = 1 to 20;
x1=10+5*rannor(0); * normal(10,25);
x2=exp(3*rannor(0)); * lognormal;
x3=5+10*ranuni(0); * uniform;
x4=50+10*rannor(0); * normal(50,100);
x5=x1+x4+rannor(0); *normal bimodal;
x6=5+2*x2+3*ranexp(0); *lognormal and exponential mixture;
x7=0.5*exp(4*rannor(0)); *lognormal;
x8=10+8*ranuni(0); * uniform;
x9=x2+x8+2*rannor(0); * lognormal, uniform and normal mix;
x10=20+x7+9*rannor(0); * lognormal and normal mix;
output; End; drop i;run;
Data One;Set Design;
Do Rep = 1 to 100;
y = 33.5+3.2*x1 +3.5*x2+2.2*x3-6*x4-9*x6-3.1*x7+7.4*x8-4*x9-
5.7*x10+3*rannor(0); *true model;
OutPut;  End;
Proc Sort; By Rep; *Proc Print;*Var x1-x10 y;
proc reg Outest=Stat NoPrint; By Rep;
model y=x1-x10/Selection=adjRsq aic bic sbc;
run;
proc sort; by Rep _aic_; *Proc Print; *Var  Rep  _aic_;
Data Final1;Set Stat;
m = Mod(_n_, 1023); *Proc Print; *Var rep x1-x10 _aic_;
Data AIC; Set Final1;
If m=1;
```

```
If x1 ne "." and x2 ne "." and x3 ne "." and x4 ne "." and x5 = "." and x6 ne "." and x7
ne "." and x8 ne "." and x9 ne "." and x10 ne "." Then x1234678910+1;
Data Final;Set AIC;
If Rep=100;
proc print; Var m Rep  _aic_  x1234678910;
proc sort Data=Stat; by Rep _bic_; *Proc Print; *Var  Rep  _bic_;
Data Final2;Set Stat;
m = Mod(_n_, 1023); *Proc Print; *Var rep x1-x10 _bic_;
Data BIC; Set Final2;
If m=1;
If x1 ne "." and x2 ne "." and x3 ne "." and x4 ne "." and x5 = "." and x6 ne "." and x7
ne "." and x8 ne "." and x9 ne "." and x10 ne "." Then x1234678910+1;
Data Final;Set BIC;
If Rep=100;
proc print; Var m Rep  _bic_  x1234678910;
proc sort Data=Stat; by Rep _sbc_; *Proc Print; *Var  Rep  _sbc_;
Data Final3;Set Stat;
m = Mod(_n_, 1023); *Proc Print; *Var rep x1-x10 _sbc_;
Data SBC; Set Final3;
If m=1;
If x1 ne "." and x2 ne "." and x3 ne "." and x4 ne "." and x5 = "." and x6 ne "." and x7
ne "." and x8 ne "." and x9 ne "." and x10 ne "." Then x1234678910+1;
Data Final;Set SBC;
If Rep=100;
proc print; Var m Rep  _sbc_ x1234678910;
Run;
```

SAS program for testing true model y = **33.5** +**3.5***x2+**3.2***x3+**2.2***x4-**6***x5-**9***x6-**3.1***x7+**7.4***x8-**4***x9-**5.7***x10+**3***rannor(**0**) with n=20, rep=100.
*We need change "Do I = 1 to 20;" to 25, 100, 1000. and the rep from 100 to 1000.

```
Data Design;
Do I = 1 to 20;
x1=10+5*rannor(0); * normal(10,25);
x2=exp(3*rannor(0)); * lognormal;
x3=5+10*ranuni(0); * uniform;
x4=50+10*rannor(0); * normal(50,100);
x5=x1+x4+rannor(0); *normal bimodal;
x6=5+2*x2+3*ranexp(0); *lognormal and exponential mixture;
x7=0.5*exp(4*rannor(0)); *lognormal;
x8=10+8*ranuni(0); * uniform;
x9=x2+x8+2*rannor(0); * lognormal, uniform and normal mix;
x10=20+x7+9*rannor(0); * lognormal and normal mix;
output; End; drop i;run;
Data One;Set Design;
Do Rep = 1 to 100;
y = 33.5 +3.5*x2+3.2*x3+2.2*x4-6*x5-9*x6-3.1*x7+7.4*x8-4*x9-
5.7*x10+3*rannor(0); *true model;
OutPut;  End;
```

```
Proc Sort; By Rep; *Proc Print;*Var x1-x10 y;
proc reg Outest=Stat NoPrint; By Rep;
model y=x1-x10/Selection=adjRsq aic bic sbc;
run;
proc sort; by Rep _aic_; *Proc Print; *Var  Rep  _aic_;
Data Final1;Set Stat;
m = Mod(_n_, 1023); *Proc Print; *Var rep x1-x10 _aic_;
Data AIC; Set Final1;
If m=1;
If x1 = "." and x2 ne "." and x3 ne "." and x4 ne "." and x5 ne "." and x6 ne "." and x7
ne "." and x8 ne "." and x9 ne "." and x10 ne "." Then x2345678910+1;
Data Final;Set AIC;
If Rep=100;
proc print; Var m Rep  _aic_ x2345678910;
proc sort Data=Stat; by Rep _bic_; *Proc Print; *Var  Rep  _bic_;
Data Final2;Set Stat;
m = Mod(_n_, 1023); *Proc Print; *Var rep x1-x10 _bic_;
Data BIC; Set Final2;
If m=1;
If x1 = "." and x2 ne "." and x3 ne "." and x4 ne "." and x5 ne "." and x6 ne "." and x7
ne "." and x8 ne "." and x9 ne "." and x10 ne "." Then x2345678910+1;
Data Final;Set BIC;
If Rep=100;
proc print; Var m Rep  _bic_  x2345678910;
proc sort Data=Stat; by Rep _sbc_; *Proc Print; *Var  Rep  _sbc_;
Data Final3;Set Stat;
m = Mod(_n_, 1023); *Proc Print; *Var rep x1-x10 _sbc_;
Data SBC; Set Final3;
If m=1;
If x1 = "." and x2 ne "." and x3 ne "." and x4 ne "." and x5 ne "." and x6 ne "." and x7
ne "." and x8 ne "." and x9 ne "." and x10 ne "." Then x2345678910+1;
Data Final;Set SBC;
If Rep=100;
proc print; Var m Rep  _sbc_ x2345678910;
Run;
```

SAS program for testing true model y = **33.5**+**3.2***x1 +**3.5***x2+**2.2***x3-**6***x4+**7.7***x5-**9***x6-**3.1***x7+**7.4***x8-**4***x9-**5.7***x10+**3***rannor(**0**) with n=20, rep=100.
*We need change "Do I = **1** to **20**;" to 25, 100, 1000. and the rep from 100 to 1000.

```
Data Design;
Do I = 1 to 20;
x1=10+5*rannor(0); * normal(10,25);
x2=exp(3*rannor(0)); * lognormal;
x3=5+10*ranuni(0); * uniform;
x4=50+10*rannor(0); * normal(50,100);
x5=x1+x4+rannor(0); *normal bimodal;
x6=5+2*x2+3*ranexp(0); *lognormal and exponential mixture;
x7=0.5*exp(4*rannor(0)); *lognormal;
```

```
x8=10+8*ranuni(0); * uniform;
x9=x2+x8+2*rannor(0); * lognormal, uniform and normal mix;
x10=20+x7+9*rannor(0); * lognormal and normal mix;
output; End; drop i;run;
Data One;Set Design;
Do Rep = 1 to 100;
y = 33.5+3.2*x1 +3.5*x2+2.2*x3-6*x4+7.7*x5-9*x6-3.1*x7+7.4*x8-4*x9-
5.7*x10+3*rannor(0); *true model; OutPut;  End;
Proc Sort; By Rep; *Proc Print;*Var x1-x10 y;
proc reg Outest=Stat NoPrint; By Rep;
model y=x1-x10/Selection=adjRsq aic bic sbc;
run;
proc sort; by Rep _aic_; *Proc Print; *Var  Rep  _aic_;
Data Final1;Set Stat;
m = Mod(_n_, 1023); *Proc Print; *Var rep x1-x10 _aic_;
Data AIC; Set Final1;
If m=1;
If x1 ne "." and x2 ne "." and x3 ne "." and x4 ne "." and x5 ne "." and x6 ne "." and
x7 ne "." and x8 ne "." and x9 ne "." and x10 ne "." Then x12345678910+1;
Data Final;Set AIC;
If Rep=100;
proc print; Var m Rep  _aic_  x12345678910;
proc sort Data=Stat; by Rep _bic_; *Proc Print; *Var  Rep  _bic_;
Data Final2;Set Stat;
m = Mod(_n_, 1023); *Proc Print; *Var rep x1-x10 _bic_;
Data BIC; Set Final2;
If m=1;
If x1 ne "." and x2 ne "." and x3 ne "." and x4 ne "." and x5 ne "." and x6 ne "." and
x7 ne "." and x8 ne "." and x9 ne "." and x10 ne "." Then x12345678910+1;
Data Final;Set BIC;
If Rep=100;
proc print; Var m Rep  _bic_  x12345678910;
proc sort Data=Stat; by Rep _sbc_; *Proc Print; *Var  Rep  _sbc_;
Data Final3;Set Stat;
m = Mod(_n_, 1023); *Proc Print; *Var rep x1-x10 _sbc_;
Data SBC; Set Final3;
If m=1;
If x1 ne "." and x2 ne "." and x3 ne "." and x4 ne "." and x5 ne "." and x6 ne "." and
x7 ne "." and x8 ne "." and x9 ne "." and x10 ne "." Then x12345678910+1;
Data Final;Set SBC;
If Rep=100;
proc print; Var m Rep  _sbc_ x12345678910;
Run;
```

Appendix B

Appendix B

Table 4. SAS output for the program with n=20, rep=100 and rep=1000

No. of predict variables	predict variables in the true model	n	Rep=100			Rep=1000		
			AIC	BIC	SBC	AIC	BIC	SBC
1	x2	20	6	41	26	131	441	339
			13	52	39	134	448	347
			9	42	31	124	464	353
			17	44	31			
			14	44	32			
			15	45	35			
			10	37	30			
			15	43	30			
			13	41	27			
			9	45	30			
1	x3	20	15	44	37	191	503	404
			14	44	36	158	471	404
			19	55	41	144	476	355
			10	41	31			
			5	42	28			
			22	52	45			
			15	46	34			
			14	51	43			
			14	49	36			
			14	43	31			
1	x10	20	13	44	33	164	492	367
			19	49	42	145	472	379
			15	48	31	145	473	354
			15	45	36			
			14	40	28			
			13	54	43			
			12	45	37			
			15	44	34			
			15	48	33			
			15	43	35			
	Average	20	13.63	45.37	34.17	148.44	471.11	366.89
2	x35	20	17	61	40	193	654	429
			19	58	41	166	622	388
			15	68	38	149	617	380

			18	68	41			
			18	67	44			
			20	64	43			
			16	56	37			
			20	63	38			
			18	60	40			
			17	56	40			
2	x39	20	17	60	40	186	589	383
			18	66	38	151	585	382
			14	63	39	155	586	360
			18	63	41			
			20	65	49			
			20	62	40			
			13	63	31			
			10	49	28			
			25	67	43			
			14	51	31			
2	x45	20	13	59	37	150	549	346
			13	48	35	161	570	362
			18	58	30	173	591	374
			15	56	36			
			23	64	42			
			17	57	40			
			15	54	33			
			14	61	36			
			14	55	37			
			19	60	43			
	Average	20	16.93	60.07	38.37	164.89	595.89	378.22
3	x235	20	16	66	37	188	706	368
			17	65	31	178	731	382
			15	65	31	164	724	364
			14	78	30			
			15	75	43			
			24	66	41			
			15	70	38			
			22	70	42			
			16	69	35			
			18	63	37			
3	x289	20	18	64	31	175	730	390
			17	69	31	159	697	356
			21	71	31	170	693	379

			20	77	45			
			17	62	32			
			15	62	27			
			14	66	29			
			15	68	36			
			18	70	33			
			20	69	35			
3	x347	20	20	72	45	167	716	350
			13	71	35	180	716	369
			18	72	35	167	713	366
			16	63	32			
			17	71	34			
			16	70	28			
			16	37	43			
			19	71	36			
			19	73	43			
			22	72	42			
	Average	20	17.43	67.90	35.60	172.00	714.00	369.33
4	x12910	20	16	76	24	190	813	388
			11	80	37	168	797	329
			22	79	37	181	789	352
			12	74	26			
			25	83	44			
			14	80	24			
			15	74	35			
			35	78	35			
			17	82	35			
			16	83	42			
4	x35710	20	17	86	42	226	848	418
			21	88	39	226	829	472
			30	92	55	199	841	392
			19	82	44	201	839	409
			18	78	41	219	832	406
			16	86	37			
			23	82	40			
			26	82	45			
			20	82	33			
			24	80	46			
4	x3567	20	17	81	32	191	812	372
			20	81	37	205	795	400
			24	83	45	170	838	347

			19	81	41	183	812	378
			13	75	36	184	815	373
			21	82	33			
			14	74	33			
			33	77	35			
			16	78	32			
			17	87	44			
	Average		19.70	80.87	37.63	195.62	820.00	387.38
5	x12345	20	29	63	35	230	569	366
			29	77	42	181	346	283
			30	64	44	208	605	369
			12	50	24	234	726	419
			25	60	41	216	594	346
			20	47	35			
			23	57	35			
			18	57	34			
			19	77	31			
			30	76	39			
5	x23589	20	24	86	41	237	882	419
			27	87	43	238	772	431
			24	94	42	252	911	432
			25	79	41			
			24	74	43			
			27	78	41			
			26	81	43			
			21	82	41			
			22	90	40			
			24	89	46			
5	x357910	20	27	94	45	244	927	424
			27	93	46	220	887	392
			24	93	43	215	908	406
			26	95	48	247	914	441
			26	91	46	226	916	401
			26	91	43			
			29	94	43			
			24	85	34			
			21	92	37			
			24	86	43			
	Average		24.43	79.40	40.30	226.77	765.92	394.54
6	135789	20	19	87	37	258	916	412

			25	81	38	238	777	421
			16	83	33	285	744	419
			26	91	39	256	944	413
			26	90	48	252	874	430
			27	81	42			
			24	86	45			
			25	73	41			
			27	91	43			
			25	88	44			
6	x1367910	20	23	98	38	252	969	425
			25	97	39	255	962	420
			25	95	47	272	971	433
			19	97	39			
			34	98	49			
			34	94	43			
			23	96	38			
			23	94	38			
			25	96	47			
			33	98	48			
6	x2358910	20	34	83	48	287	786	422
			24	74	46	267	830	443
			29	92	43	277	928	443
			31	96	49	270	876	423
			26	70	44			
			18	74	42			
			20	92	38			
			22	96	47			
			27	96	43			
			33	95	57			
	Average		25.60	89.40	43.10	264.08	881.42	425.33
7	x1235789	20	37	47	57	315	470	471
			33	33	51	313	168	439
			34	30	49	304	473	454
			32	39	50	283	194	391
			24	61	40	301	694	466
			28	15	37			
			28	37	50			
			30	60	47			
			31	46	45			
			27	49	41			
7	x1345678	20	30	58	42	318	433	458

			21	27	35	333	489	479
			35	52	45	294	781	441
			31	43	43	286	484	452
			29	44	49	345	605	495
			37	36	46			
			38	66	52			
			31	49	52			
			35	37	46			
			34	46	51			
7	x23457810	20	38	61	52	299	568	449
			26	58	42	341	401	487
			29	28	42	312	357	469
			31	49	46	287	523	470
			35	49	48			
			34	36	49			
			27	67	41			
			30	55	44			
			35	48	53			
			35	10	43			
	Average		31.50	44.53	46.27	309.36	474.29	458.64
8	x12345678	20	38	23	54	410	155	552
			46	15	69	451	234	571
			41	13	57	465	339	606
			41	19	62			
			38	19	56			
			41	9	47			
			42	13	57			
			44	28	53			
			49	36	63			
			43	12	57			
8	x13456789	20	36	2	45	411	96	510
			29	1	3	257	9	266
			33	7	37	410	94	517
			40	8	46			
			35	5	44			
			31	0	46			
			50	4	62			
			30	13	44			
			43	13	59			
			39	8	53			
8	x234567810	20	42	66	62	471	413	603

			43	25	58	453	444	584
			46	14	58	422	250	569
			46	41	61			
			39	17	60			
			39	6	52			
			40	24	58			
			43	54	55			
			45	32	56			
			50	13	57			
	Average		40.7 3	18.0 0	53.0 3	416.6 7	226.0 0	530.8 9
9	x1234567810	20	52	0	61	583	19	660
			53	1	52	599	38	689
			59	2	65	582	6	617
			58	0	68			
			66	2	64			
			63	2	65			
			62	2	66			
			57	4	65			
			49	1	54			
			57	1	69			
9	x1234678910	20	64	11	76	626	73	702
			66	6	76	622	43	706
			64	6	74	622	46	711
			64	9	73	635	42	715
			62	5	71			
			68	7	77			
			67	9	77			
			62	10	72			
			55	19	64			
			61	7	72			
9	x2345678910	20	65	9	74	651	39	725
			57	7	67	659	38	725
			67	3	73	634	42	726
			69	9	77			
			66	8	79			
			59	4	61			
			69	5	79			
			55	7	60			
			66	6	79			
			63	6	73			
	Average		61.5	5.60	69.4	621.3	38.60	697.6

			0		3	0		0
10	x12345678910	20	97	0	96	937	2	909
			96	1	95	1000	28	1000
			86	0	80	906	2	865
			95	0	93			
			92	0	90			
			85	0	82			
			100	1	100			
			78	0	67			
			99	2	99			
			98	2	98			
	Average	20	92.60	0.60	90.00	947.67	10.67	924.67

Table 5. SAS output for the program with n=25, rep=100 and rep=1000

No. of predict variables	predict variables in the true model		Rep=100			Rep=1000		
		n	AIC	BIC	SBC	AIC	BIC	SBC
1	x2	25	12	38	37	160	401	423
			16	52	53	179	421	437
			13	32	34	160	424	444
			23	49	53			
			21	43	45			
			15	37	40			
			13	36	39			
			12	39	38			
			14	45	43			
			16	39	41			
1	x3	25	26	52	55	182	423	454
			21	41	45	206	467	493
			15	44	44	189	440	458
			22	50	49			
			20	43	46			
			20	49	48			
			25	47	50			
			24	53	54			
			21	50	51			
			21	46	48			
1	x10	25	13	43	49	161	439	460
			19	46	48	170	426	446

			11	47	53	171	409	442
			16	44	43			
			21	43	47			
			16	44	48			
			17	45	47			
			20	43	45			
			15	43	45			
			18	41	45			
	Average	25	17.87	44.13	46.10	175.33	427.78	450.78
2	x35	25	23	49	43	204	586	519
			26	64	58	231	572	516
			24	54	45	206	575	501
			23	62	54			
			20	58	50			
			15	52	43			
			20	55	48			
			24	60	50			
			14	44	35			
			17	51	45			
2	x39	25	17	51	44	209	523	452
			20	49	50	216	551	484
			20	58	48	189	535	453
			16	52	44			
			24	60	52			
			25	52	47			
			19	61	49			
			22	57	51			
			20	57	47			
			17	56	47			
2	x45	25	17	50	42	199	529	463
			22	57	53	182	553	471
			18	52	47	209	538	464
			20	56	42			
			15	48	39			
			17	57	45			
			15	55	48			
			18	61	45			
			27	58	49			
			19	54	47			
	Average	25	19.80	55.00	46.90	205.00	551.33	480.33
3	x235	25	22	75	49	228	667	501
			23	63	47	219	645	482

			13	69	47	202	612	443
			20	58	40			
			25	65	50			
			16	66	48			
			27	69	50			
			24	69	47			
			24	74	51			
			22	67	47			
3	x289	25	23	67	51	212	615	450
			18	68	52	225	645	481
			24	72	45	202	630	465
			21	64	52			
			22	67	46			
			22	64	46			
			23	69	47			
			23	62	49			
			20	62	48			
			23	62	40			
3	x347	25	26	69	51	231	652	493
			22	66	54	236	667	499
			16	60	41	247	660	510
			19	69	50			
			23	73	52			
			26	71	49			
			28	77	55			
			26	73	57			
			20	65	51			
			21	60	41			
	Average	25	22.07	67.17	48.43	222.44	643.67	480.44
4	x12910	25	20	82	45	213	732	439
			13	68	44	226	730	455
			29	81	53	226	735	481
			25	71	51			
			18	64	46			
			18	64	45			
			14	76	38			
			15	75	43			
			22	75	47			
			20	75	50			
4	x35710	25	28	80	51	303	790	539
			29	75	56	288	810	570
			19	71	50	271	797	546

			29	83	61	273	792	558
			30	80	52			
			26	83	51			
			32	76	52			
			28	80	64			
			35	81	62			
			31	77	53			
4	x3567	25	22	74	48	218	752	460
			28	73	53	222	741	474
			22	72	43	239	759	484
			26	71	46	192	746	468
			25	74	42			
			23	76	51			
			26	70	43			
			16	69	37			
			16	74	45			
			24	68	35			
	Average	25	23.63	74.60	48.57	242.82	762.18	497.64
5	x12345	25	30	70	48	282	767	505
			28	82	56	258	812	481
			27	77	53	299	816	531
			31	79	54	263	699	466
			22	71	44			
			31	65	52			
			26	81	49			
			28	63	53			
			30	70	46			
			31	80	52			
5	x23589	25	33	84	49	286	877	544
			21	83	48	320	892	569
			25	84	56	315	858	553
			31	88	55			
			25	85	51			
			28	72	50			
			29	85	57			
			24	84	50			
			31	88	57			
			32	86	54			
5	x357910	25	30	89	58	309	886	549
			22	81	45	294	895	534
			38	86	57	332	886	570
			40	89	56			

			28	79	46			
			30	85	55			
			34	84	56			
			33	89	53			
			28	90	55			
			27	85	49			
	Average	25	29.10	81.13	52.13	295.80	838.80	530.20
6	135789	25	30	91	53	322	859	515
			29	86	54	337	889	530
			39	93	63	326	875	536
			36	95	55	326	925	539
			26	91	43			
			31	94	57			
			31	88	51			
			35	79	47			
			41	92	63			
			35	85	51			
6	x1367910	25	33	96	60	357	922	561
			26	89	46	337	926	564
			40	88	51	343	942	593
			29	92	53			
			35	95	54			
			30	96	60			
			30	86	56			
			41	92	59			
			31	91	54			
			34	91	52			
6	x2358910	25	30	89	52	324	921	531
			36	95	59	344	945	601
			34	93	58	326	795	515
			38	87	57	332	929	549
			29	89	51			
			30	93	58			
			33	71	45			
			34	93	58			
			38	92	54			
			28	96	58			
	Average	25	33.07	90.27	54.40	334.00	902.55	548.55
7	x1235789	25	43	91	65	420	868	588
			37	87	51	369	896	576
			37	89	57	375	844	577
			35	67	53			

			32	90	57			
			36	89	51			
			39	95	57			
			46	86	60			
			37	76	52			
			41	80	57			
7	x1345678	25	36	98	61	437	950	629
			40	69	60	373	923	562
			44	84	64	429	813	616
			38	60	53	389	835	585
			43	96	60			
			32	94	55			
			40	71	64			
			37	74	49			
			43	63	61			
			46	93	59			
7	x23457810	25	43	99	64	381	972	563
			41	87	65	370	942	579
			40	88	64	393	972	589
			39	92	51			
			42	92	60			
			41	98	61			
			36	99	58			
			34	95	55			
			35	92	61			
			36	91	52			
	Average	25	38.97	86.17	57.90	393.60	901.50	586.40
8	x12345678	25	46	97	62	526	808	703
			49	85	66	513	702	685
			55	89	71	527	647	689
			54	95	67	503	917	663
			46	99	68			
			48	96	68			
			51	84	70			
			50	80	72			
			48	74	64			
			50	83	63			
8	x13456789	25	42	3	34	449	205	557
			46	19	55	486	430	633
			52	54	68	424	95	499
			52	59	69			
			41	27	55			

			50	42	62			
			41	16	47			
			45	29	53			
			40	12	51			
			43	4	45			
8	x234567810	25	56	98	70	481	976	661
			55	100	69	520	806	671
			53	98	65	536	960	711
			36	97	55			
			55	79	66			
			52	96	60			
			41	99	56			
			47	99	66			
			55	99	71			
			51	99	64			
	Average	25	48.33	70.37	61.73	496.50	654.60	647.20
9	x1234567810	25	62	9	66	676	161	737
			66	21	82	678	213	781
			74	16	77	645	88	716
			69	12	77			
			60	7	67			
			68	24	75			
			75	16	81			
			73	15	81			
			70	22	78			
			61	15	69			
9	x1234678910	25	73	82	86	693	556	798
			70	78	82	723	635	824
			74	34	84	691	921	806
			69	36	79			
			72	90	78			
			72	50	83			
			70	53	83			
			66	50	80			
			69	77	81			
			66	64	76			
9	x2345678910	25	70	30	75	687	449	803
			73	65	82	690	852	819
			73	46	84	707	503	811
			73	31	80	667	155	731
			63	32	69			
			70	28	79			

			77	26	85			
			70	31	81			
			71	50	81			
			74	49	83			
	Average	25	69.77	38.63	78.80	685.70	453.30	782.60
10	x12345678910	25	98	4	97	986	75	974
			100	11	100	999	255	998
			100	17	99	998	188	992
			100	21	100			
			100	28	100			
			100	33	100			
			99	28	99			
			100	27	100			
			100	21	99			
			100	20	100			
10	Average	25	99.70	21.00	99.40	994.33	172.67	988.00

Table 6. SAS output for the program with n=100, rep=100 and rep=1000

No. of Predict variabl es	predict variables in the true model		Rep=100			Rep=1000		
		n	AIC	BIC	SBC	AIC	BIC	SBC
1	ax2	100	20	23	69	252	315	783
			18	23	71	243	297	769
			30	35	87	253	312	753
			27	33	72			
			17	23	74			
			30	35	78			
			35	40	74			
			24	29	77			
			26	30	80			
			19	27	81			
1	ax3	100	28	36	78	280	336	822
			29	38	81	291	363	820
			23	28	74	287	337	794
			29	35	85			
			26	30	75			
			32	39	85			
			34	40	86			
			26	34	84			
			35	39	84			

			24	31	76			
1	ax10	100	27	30	77	272	338	789
			28	36	79	271	349	785
			29	34	86	247	321	766
			31	35	84			
			22	27	71			
			26	29	85			
			30	31	78			
			30	32	80			
			21	30	81			
			21	28	78			
	Average	100	26.57	32.00	79.00	266.22	329.78	786.78
2	x35	100	33	43	89	356	431	858
			34	39	80	333	412	821
			34	43	81	362	465	828
			29	33	77			
			36	41	79			
			31	41	85			
			41	52	87			
			38	48	79			
			30	37	79			
			43	50	88			
2	x39	100	31	41	78	298	375	811
			41	45	85	286	366	799
			44	46	80	325	406	822
			31	38	75			
			36	45	77			
			33	40	79			
			32	42	83			
			37	46	84			
			35	39	78			
			21	34	83			
2	x45	100	35	39	82	311	399	844
			24	32	79	298	385	835
			29	33	79	296	388	816
			29	37	80			
			30	40	81			
			24	31	76			
			27	42	80			
			31	46	76			
			39	47	87			
			26	34	82			

	Average	100	32.80	40.80	80.93	318.33	403.00	826.00
3	x235	100	29	37	84	326	419	833
			34	46	81	337	439	829
			31	45	86	324	422	825
			33	46	78			
			35	46	85			
			38	52	87			
			34	47	79			
			29	40	84			
			35	46	84			
			30	37	84			
3	x289	100	32	42	79	310	398	766
			33	44	84	350	455	794
			35	43	80	351	431	803
			34	43	85	321	440	821
			32	47	82	346	456	814
			42	50	78			
			44	47	78			
			31	36	86			
			32	41	74			
			35	43	75			
3	x347	100	38	46	87	352	451	838
			43	49	86	362	462	831
			37	42	84	362	462	832
			44	56	80			
			41	47	85			
			44	55	90			
			29	35	76			
			45	51	83			
			28	42	89			
			33	45	83			
	Average	100	35.33	44.87	82.53	340.09	439.55	816.91
4	x12910	100	33	37	80	226	735	793
			35	47	80	348	443	818
			33	48	79	344	458	813
			31	39	82	380	487	826
			34	49	78	359	478	814
			38	42	89			
			38	46	75			
			37	45	78			
			47	55	81			
			40	50	77			

4	x35710	100	45	56	90	414	521	836
			45	56	84	447	550	871
			36	47	81	418	542	863
			36	51	77			
			52	63	90			
			44	55	87			
			51	58	87			
			39	50	87			
			44	52	88			
			37	48	84			
4	x3567	100	30	42	76	369	492	826
			42	53	81	392	485	838
			40	53	88	366	464	807
			30	40	76	369	474	830
			41	53	87			
			44	52	82			
			46	56	86			
			33	50	80			
			38	50	81			
			38	49	83			
	Average	100	39.23	49.73	82.47	369.33	510.75	827.92
5	x12345	100	41	53	82	445	582	860
			46	60	89	438	571	865
			49	60	85	435	567	872
			38	56	87	465	576	866
			53	64	91			
			39	50	84			
			51	65	85			
			49	59	89			
			52	60	87			
			37	48	86			
5	x23589	100	43	52	88	466	595	875
			49	58	91	490	593	851
			47	58	82	490	595	880
			39	59	90			
			48	60	86			
			43	55	83			
			43	63	83			
			48	60	81			
			46	62	91			
			40	61	80			
5	x357910	100	54	66	88	472	602	864

			56	64	91	462	590	873
			39	52	81	495	615	876
			63	71	94			
			44	52	83			
			52	62	93			
			43	51	83			
			40	57	93			
			46	61	88			
			40	53	80			
	Average	100	45.93	58.40	86.47	465.80	588.60	868.20
6	135789	100	44	60	88	492	623	858
			54	66	83	492	607	876
			51	67	93	468	610	869
			52	59	85			
			51	67	92			
			51	65	85			
			60	69	85			
			46	57	84			
			47	67	89			
			48	62	82			
6	x1367910	100	57	67	85	523	638	884
			53	61	87	503	637	888
			50	65	86	524	648	898
			55	64	88			
			52	60	86			
			49	63	83			
			50	65	87			
			44	58	87			
			51	67	89			
			50	63	88			
6	x2358910	100	54	65	84	492	617	863
			50	65	86	515	641	885
			48	60	86	486	611	854
			41	61	84			
			50	63	88			
			49	66	86			
			55	71	94			
			51	63	81			
			48	58	87			
			56	66	85			
	Average	100	50.57	63.67	86.43	499.44	625.78	875.00
7	x1235789	100	46	59	83	556	687	883

			56	74	93	566	685	897
			56	68	85	553	672	873
			57	68	90			
			54	68	90			
			61	73	87			
			54	64	89			
			49	65	90			
			66	73	93			
			56	71	89			
7	x1345678	100	56	71	87	577	694	694
			49	64	89	597	697	881
			64	77	90	519	646	857
			61	73	89	553	675	883
			55	68	82			
			59	68	86			
			50	65	85			
			51	65	90			
			51	67	84			
			52	62	85			
7	x23457810	100	57	68	87	543	664	876
			53	71	86	552	680	876
			58	67	88	568	698	894
			60	72	93			
			58	66	89			
			57	70	89			
			50	63	83			
			50	60	85			
			68	73	91			
			49	63	88			
	Average	100	55.43	67.87	87.83	558.40	679.80	861.40
8	x12345678	100	61	84	96	685	781	907
			69	80	94	688	788	918
			64	77	93	679	772	900
			66	80	93			
			61	77	88			
			66	77	93			
			74	77	89			
			73	84	94			
			66	78	88			
			70	77	90			
8	x13456789	100	66	79	94	662	784	916
			68	79	94	661	789	924

			67	78	94	671	789	922
			62	70	91			
			72	84	92			
			65	79	91			
			63	81	91			
			67	79	93			
			72	84	95			
			71	82	91			
8	x234567810	100	58	73	89	650	758	897
			70	76	92	658	776	912
			59	73	90	695	790	923
			68	81	92			
			62	73	92			
			60	69	86			
			69	79	90			
			62	73	88			
			73	87	96			
			59	79	92			
	Average	100	66.10	78.30	91.70	672.11	780.78	913.22
9	x1234567810	100	81	90	96	821	895	958
			74	84	94	812	894	960
			84	93	98	825	890	953
			80	91	96			
			79	85	97			
			80	86	92			
			83	89	97			
			85	91	96			
			70	81	90			
			76	84	94			
9	x1234678910	100	81	91	98	820	895	960
			84	94	98	816	887	950
			87	91	97	831	905	961
			81	87	95			
			87	91	100			
			83	93	98			
			87	92	97			
			84	89	97			
			83	90	98			
			82	87	97			
9	x234567891	100	80	89	96	816	891	970

	0							
			88	94	99	830	901	965
			80	87	97	808	883	943
			82	89	98			
			80	87	97			
			81	91	97			
			87	95	98			
			80	86	95			
			81	91	98			
			81	88	94			
	Average	100	81.70	89.20	96.47	819.89	893.44	957.78
10	x12345678910	100	100	100	100	1000	1000	1000
			100	100	100	1000	1000	1000
			100	100	100	1000	1000	1000
			100	100	100			
			100	100	100			
			100	100	100			
			100	100	100			
			100	100	100			
			100	100	100			
			100	100	100			
10	Average	100	100.00	100.00	100.00	1000.00	1000.00	1000.00

Table 7. SAS output for the program with n=1000, rep=100 and rep=1000

No. of predict variables	predict variables in the true model		Rep=100			Rep=1000		
		n	AIC	BIC	SBC	AIC	BIC	SBC
1	x2	1000	27	28	96	276	284	954
			23	23	93	284	285	942
			27	27	94	265	271	953
			26	26	93			
			28	28	94			
			23	24	92			
			33	34	93			
			26	28	96			
			31	31	98			
			24	25	91			
1	x3	1000	35	36	95	333	337	947
			31	32	97	339	344	954

			38	38	96	345	350	957
			31	32	94			
			29	29	96			
			27	27	92			
			28	28	98			
			24	25	93			
			35	35	94			
			31	31	96			
1	x10	1000	35	35	97	291	297	953
			34	34	94	311	317	947
			27	28	98	295	298	931
			25	25	92			
			29	29	92			
			35	36	89			
			28	29	95			
			29	31	96			
			25	26	93			
			27	28	98			
	Average	1000	29.03	29.60	94.50	304.33	309.22	948.67
2	x35	1000	37	37	98	349	360	964
			33	36	95	372	380	959
			39	39	98	384	392	964
			47	49	98			
			33	33	96			
			41	41	96			
			41	40	95			
			38	39	98			
			43	44	98			
			34	35	95			
2	x39	1000	31	31	95	300	308	940
			33	34	97	339	351	941
			32	32	95	346	354	944
			41	41	97			
			35	37	96			
			37	37	92			
			43	44	99			
			34	35	97			
			37	39	97			
			37	37	93			
2	x45	1000	42	42	95	367	379	954
			40	41	96	348	355	958
			35	35	96	335	340	960

			32	32	96			
			40	40	98			
			32	32	93			
			37	39	92			
			34	36	93			
			44	45	93			
			43	43	93			
	Average	1000	37.50	38.17	95.67	348.89	357.67	953.78
3	x235	1000	38	39	96	365	371	955
			44	45	93	342	371	952
			38	39	94	353	364	942
			40	41	95			
			37	38	96			
			38	39	94			
			42	43	95			
			24	25	97			
			35	36	93			
			36	36	96			
3	x289	1000	35	35	90	376	383	944
			38	39	96	364	372	950
			36	38	96	365	375	953
			36	38	95			
			40	40	93			
			43	43	97			
			33	33	95			
			33	34	96			
			37	38	95			
			33	33	95			
3	x347	1000	41	42	95	410	423	959
			45	46	98	385	399	971
			35	35	98	392	402	953
			40	40	95			
			50	51	98			
			40	40	96			
			35	35	98			
			41	43	95			
			47	48	97			
			43	43	95			
	Average	1000	38.43	39.17	95.40	372.44	384.44	953.22
4	x12910	1000	38	40	92	399	406	954
			41	45	99	405	416	962
			37	37	97	380	388	952

			33	33	99	402	409	957
			30	30	97			
			41	41	94			
			37	37	93			
			45	46	96			
			37	39	99			
			39	40	94			
4	x35710	1000	47	48	100	487	492	985
			46	47	96	469	479	967
			61	61	98	461	471	965
			40	42	95	474	487	970
			54	54	99			
			54	55	96			
			51	52	97			
			54	54	99			
			38	39	98			
			45	47	96			
4	x3567	1000	48	48	98	398	401	966
			46	46	93	423	434	954
			39	41	96	409	415	963
			39	39	95	401	414	958
			42	43	97			
			44	45	93			
			37	38	97			
			40	41	94			
			39	41	96			
			48	49	97			
	Average	1000	43.00	43.93	96.33	425.67	434.33	962.75
5	x12345	1000	51	53	98	475	487	965
			43	43	97	470	478	967
			57	59	99	487	499	980
			47	49	98			
			52	52	98			
			51	53	95			
			41	44	97			
			47	49	97			
			51	52	98			
			42	43	95			
5	x23589	1000	39	42	97	491	498	965
			46	50	99	496	508	964
			49	51	98	520	528	975
			51	52	97			

			49	51	98			
			52	52	96			
			44	44	92			
			47	48	95			
			44	45	97			
			49	46	98			
5	x357910	1000	57	58	98	524	535	967
			56	56	100	536	546	967
			45	49	96	494	506	969
			46	50	97			
			60	60	98			
			52	52	95			
			47	49	97			
			55	56	99			
			47	47	95			
			49	50	95			
	Average	1000	48.87	50.17	96.97	499.22	509.44	968.78
6	135789	1000	58	59	97	533	548	962
			56	56	95	509	523	972
			52	53	94	515	528	967
			52	54	96			
			54	54	98			
			64	65	96			
			42	42	93			
			45	45	97			
			51	53	96			
			54	56	99			
6	x1367910	1000	50	52	98	536	544	974
			54	56	96	529	538	973
			49	51	95	578	589	970
			49	50	97			
			59	62	96			
			48	49	96			
			61	62	98			
			43	43	96			
			61	61	98			
			56	57	94			
6	x2358910	1000	55	57	96	547	556	969
			56	57	97	534	547	974
			55	55	99	529	542	970
			53	55	94			
			56	56	98			

			54	56	98			
			45	45	98			
			56	56	98			
			47	49	95			
			54	56	100			
	Average	1000	52.97	54.07	96.60	534.44	546.11	970.11
7	x1235789	1000	61	62	95	611	623	984
			64	66	99	559	571	972
			61	62	100	600	608	966
			55	59	98			
			54	56	99			
			61	62	99			
			57	58	93			
			60	60	99			
			52	52	95			
			62	66	97			
7	x1345678	1000	64	64	96	602	612	972
			59	59	99	612	625	982
			53	54	97	609	621	972
			68	69	97			
			52	52	97			
			59	61	95			
			69	70	99			
			58	58	98			
			58	59	99			
			53	55	98			
7	x23457810	1000	56	58	98	577	590	970
			58	58	96	605	619	972
			59	61	97	580	589	978
			59	60	100			
			54	54	99			
			56	57	97			
			54	55	98			
			63	65	98			
			60	61	95			
			60	60	94			
	Average	1000	58.63	59.77	97.37	595.00	606.44	974.22
8	x12345678	1000	70	74	99	712	717	980
			72	73	97	694	708	982
			74	74	99	727	740	987
			72	74	97			
			75	75	97			

			70	73	99			
			73	75	97			
			69	70	99			
			66	66	98			
			69	71	97			
8	x13456789	1000	74	76	97	723	736	983
			65	65	95	692	708	981
			63	64	97	741	753	985
			73	73	99			
			80	81	99			
			68	69	98			
			75	75	100			
			64	65	99			
			76	77	97			
			67	68	100			
8	x234567810	1000	69	70	98	728	738	987
			76	79	100	697	708	976
			67	68	98	722	729	983
			79	80	99			
			79	80	97			
			71	74	97			
			70	71	98			
			68	68	98			
			76	76	96			
			68	70	99			
	Average	1000	71.27	72.47	98.00	715.11	726.33	982.67
9	x1234567810	1000	84	84	98	832	840	994
			80	80	100	832	846	989
			77	77	97	845	852	993
			80	80	98			
			84	84	100			
			78	79	99			
			85	86	98			
			83	83	98			
			83	84	100			
			81	81	97			
9	x1234678910	1000	84	84	99	837	840	991
			78	79	98	856	858	993
			87	88	99	860	869	985

			84	85	100			
			84	85	100			
			83	83	100			
			87	87	99			
			82	82	100			
			83	83	99			
			84	86	99			
9	x23456789 10	1000	82	84	100	862	865	992
			76	78	100	827	839	989
			84	84	97	854	858	994
			87	87	100			
			90	90	98			
			85	85	100			
			85	85	99			
			82	83	98			
			83	86	98			
			84	84	98			
	Average	1000	82.97	83.53	98.87	845.00	851.89	991.11
10	x12345678 910	1000	100	100	100	1000	1000	1000
			100	100	100	1000	1000	1000
			100	100	100	1000	1000	1000
			100	100	100			
			100	100	100			
			100	100	100			
			100	100	100			
			100	100	100			
			100	100	100			
			100	100	100			
10	Average	1000	100.00	100.00	100.00	1000.00	1000.00	1000.00

Table 8. The average of SAS output of AIC BIC and SBC with rep=100 and rep=1000.

				Rep=100		Rep=1000	
	No. of predict variable	AIC	BIC	SBC	AIC	BIC	SBC
n=20	1	13.63	45.37	34.17	148.44	471.11	366.89
	2	16.93	60.07	38.37	164.89	595.89	378.22
	3	17.43	67.90	35.60	172.00	714.00	369.33

	4	19.70	80.87	37.63	195.62	820.00	387.38
	5	24.43	79.40	40.30	226.77	765.92	394.54
	6	25.60	89.40	43.10	264.08	881.42	425.33
	7	31.50	44.53	46.27	309.36	474.29	458.64
	8	40.73	18.00	53.03	416.67	226.00	530.89
	9	61.50	5.60	69.43	621.30	38.60	697.60
	10	92.60	0.60	90.00	947.67	10.67	924.67
n=25	1	17.87	44.13	46.10	175.33	427.78	450.78
	2	19.80	55.00	46.90	205.00	551.33	480.33
	3	22.07	67.17	48.43	222.44	643.67	480.44
	4	23.63	74.60	48.57	242.82	762.18	497.64
	5	29.10	81.13	52.13	295.80	838.80	530.20
	6	33.07	90.27	54.40	334.00	902.55	548.55
	7	38.97	86.17	57.90	393.60	901.50	586.40
	8	48.33	70.37	61.73	496.50	654.60	647.20
	9	69.77	38.63	78.80	685.70	453.30	782.60
	10	99.70	21.00	99.40	994.33	172.67	988.00
n=100	1	26.57	32.00	79.00	266.22	329.78	786.78
	2	32.80	40.80	80.93	318.33	403.00	826.00
	3	35.33	44.87	82.53	340.09	439.55	816.91
	4	39.23	49.73	82.47	369.33	510.75	827.92
	5	45.93	58.40	86.47	465.80	588.60	868.20
	6	50.57	63.67	86.43	499.44	625.78	875.00
	7	55.43	67.87	87.83	558.40	679.80	861.40
	8	66.10	78.30	91.70	672.11	780.78	913.22
	9	81.70	89.20	96.47	819.89	893.44	957.78
	10	100.00	100.00	100.00	1000.00	1000.00	1000.00
n=1000	1	29.03	29.60	94.50	304.33	309.22	948.67
	2	37.50	38.17	95.67	348.89	357.67	953.78
	3	38.43	39.17	95.40	372.44	384.44	953.22
	4	43.00	43.93	96.33	425.67	434.33	962.75
	5	48.87	50.17	96.97	499.22	509.44	968.78
	6	52.97	54.07	96.60	534.44	546.11	970.11
	7	58.63	59.77	97.37	595.00	606.44	974.22
	8	71.27	72.47	98.00	715.11	726.33	982.67
	9	82.97	83.53	98.87	845.00	851.89	991.11
	10	100.00	100.00	100.00	1000.00	1000.00	1000.00

I, Xiaotong Li, hereby submit this thesis/report to Emporia State University as partial fulfillment of the requirements for an advanced degree. I agree that the Library of the University may make it available to use in accordance with its regulations governing materials of this type. I further agree that quoting, photocopying, digitizing or other reproduction of this document is allowed for private study, scholarship (including teaching) and research purposes of a nonprofit nature. No copying which involves potential financial gain will be allowed without written permission of the author. I also agree to permit the Graduate School at Emporia State University to digitize and place this thesis in the ESU institutional repository.

Signature of Author

Date

Model Selection Techniques
for Multiple Linear Regression Models
Title of Thesis

Signature of Graduate School Staff

Date Received