INTRODUCTION

Background
Credit risk is one of the eight risks that banks must consider. It is important to make a measurable, documented, and developable credit risk system. Logistic regression, discriminant analysis, and artificial neural network are some methods that are used in credit risk model. They are useful to predict whether a new applicant will become a good or bad debtor if he or she receives a loan.

Multicollinearity is a common problem in credit risk modeling. Usually, the solution for this problem is using variable selection method (forward, backward, and stepwise). But this solution may cause missing information about the response variable if the deleted predictor variable is an important one. Ridge regression is another statistical procedure for dealing with the problem of multicollinearity (Ravinshanker & Dey 2001). With logistic ridge regression, the multicollinearity is expected to be handled without deleting any variables and there will be no missing information from the data that has been collected.

Bank of Indonesia noted that the growth of credit of national banks in January 2010 was 10%. Until the end of August 2010, the credit of banking industry grew and reached 20.3% (Purnomo 2010a, 2010b). This may conduce on a greater risk that has not been faced by banks before. Hence, it is important to build a more accurate credit scoring model to decide whether a new applicant is credible enough to get a loan.

Objectives
The objectives of this research are:
1. To build a credit risk model using logistic regression with variable selection and logistic ridge regression.
2. To determine the optimal probability cutpoint.
3. To compare the classification rate and the c statistic of logistic regression with variable selection and logistic ridge regression.

LITERATURE REVIEW

Credit Risk Model
Banks loan to individuals, first by asking to fill out a loan application. The customer is asked to submit several documents that the bank needs in order to evaluate the loan request. There are six aspects of the loan application to determine whether a new applicant is creditworthy or not, The Six Basic Cs of Lending are namely character, capacity, cash, collateral, condition, and control. Character is the data about the personality. Capacity is the capacity to borrow money. Cash is related to the borrower income and balance in saving account. Collateral is the adequacy of the borrower to provide adequate support for the loan. Age and degree of specialization of the borrower's assets are the example of collateral. Condition is the prospect of business associated with economics conditions. Correctly prepared loan document is the example of control.

The basic theory of credit scoring is that the bank can identify the financial, economic, and motivational factors that separate the good debtors from the bad ones by observing a large group of people who have borrowed in the past. Credit scoring systems are usually based on discriminant models or related techniques such as logit or probit models or neural networks. If the applicant’s score exceeds a critical cutpoint level, he or she is more likely to be approved for credit. Among the most important variables used in evaluating consumer’s loan are age, marital status, number of dependents, home ownership, telephone ownership, type of occupation, and length of employment in a current job.

The Cramer Statistic
The chi-square test of independence is used to conclude whether there is an association between two categorical variables. When the number of rows and columns of the contingency table are unequal, Cramer coefficient is the measure of the strength of this association. The value is between 0 and 1. The Cramer coefficient is defined as:

$$C = \sqrt{\frac{X^2}{n(t-1)}}$$

Where $X^2$ is the chi-square statistic, $n$ is the total sample size, and $t$ is either the number of