INTRODUCTION

Background

The central and local government have developed various programs in order to improve the welfare of society in all Indonesian regions. Centralization of development has affected large gaps among regions in and out of Java, western and eastern part of Indonesia, and between urban and rural area. There are more than 70.611 villages in Indonesia, 32.379 of them are categorized as backward region with 62% of these villages are located in the eastern regions of Indonesia (KNPDT, 2004).

The Government of Indonesia (GoI) needs to accelerate development in backward regions to overcome the problems. The fundamental purpose of accelerating development in backward regions is to empower backward society to fulfill their basic needs so they can do the activities that play a crucial role in balancing with the other societies in Indonesia. Therefore, since 2004 the government has realized the importance of developing the State Ministry of Acceleration of Development in Backward Regions (KNPDT, 2004).

The functions of State Ministry of Acceleration of Development in Backward Regions (KNPDT) are (1) to formulate national policy in the field of development in backward regions sector, (2) to implement the policy (3) to organize stated-owned properties or assets, (4) to supervise the implementation of its duty, (5) to submit the evaluation report, suggestion and consideration in its assignment and function to the President (KNPDT, 2004). In order to reach these objectives, the GoI must know which are the prior concerns so that the development will be effective and efficient.

The GoI has used 33 explanatory variables to determine backward region status. These variables are possibly correlated one to another. Therefore, it is very important to simplify factors that most influence backward region status for further analysis. Ordinal logistic regression was implemented to find the most influential factor. In addition, biplot was used to present graphically information of relationships between explanatory variables and observations.

Objective

The objectives of this research were to

1. Determine factors that strongly influence the backward region status and give recommendation to the GoI for making good policy and appropriate allocation of assets or fund based on these factors.
2. Present graphical information of relationship between explanatory and observation variables. It is also interesting to compare the condition of western and eastern part of Indonesia, taking into mind that most of backward regions are located in the eastern regions of Indonesia.

LITERATURE REVIEW

Backward Regions

Backward regions are regencies in Indonesia that are relatively undeveloped compared with other regions in the country (KNPDT, 2004).

Poverty

Poverty is a deprivation of common necessities that determine the quality of life, including food, clothing, shelter and safe drinking water, and may also include the deprivation of opportunities to learn, to obtain better employment to escape poverty, and or to enjoy the respect of fellow citizens. The World Bank defines extreme poverty as living on less than US$ 1 per day, and moderate poverty as less than US$ 2 a day (Wikipedia, 2008).

Ordinal Logistic Regression

Logistic regression extends categorical data analysis to data sets with binary response and one or more continuous factor (Freeman 1987). Ordinal logistic regression perform logistic regression on an ordinal response variable. One way to use category ordering forms logit of cumulative probabilities for ordinal response Y with c categories, $\underline{x}$ are explanatory variables. The cumulative probability for each category can be formulated as

$$P(Y \le j \mid \underline{x}) = F_j(\underline{x}) = p_1(\underline{x}) + \ldots + p_j(\underline{x}) \quad \ldots\ldots\ldots(1)$$

where $p_j(\underline{x})$ is the response probability of the j^{th} category of an explanatory variable $\underline{x}$. Cumulative logits for each category j are defined as

$$L_j(x) = \ln\left[\frac{F_j(x)}{1-F_j(x)}\right]; \text{ where } j = 1,2,\ldots,c-1 \ldots\ldots(2)$$

A model that simultaneously uses all cumulative logits can be written as

$$\hat{L}_j(\underline{x}) = \hat{a}_j + \underline{\hat{b}}'\underline{x} \quad \text{.....(3)}$$

Each cumulative logit has its own intercept. The $\hat{a}_j$ are increasing in j, since $P(Y \le j \mid \underline{x})$ increases in j when x is fixed, and the logit is an increasing function of this probability (Agresti, 2002). $\hat{a}_j$ and $\underline{\hat{b}}'$ are the maximum likelihood estimators for each a_j and $\underline{b}'$. These estimators represent the change in logits cumulative for each j category, if the other explanatory variables do not influence $\hat{L}_j(\underline{x})$. The interpretation of the $\underline{\hat{b}}'$ is the change in logit cumulative for each j category, in other hand, odds ratio will change equal to exp ($\underline{\hat{b}}'$) for each change of explanatory variables $\underline{x}$ (Agresti, 2002).

The estimate value for $P(Y \le j \mid \underline{x})$ can be derived with inverse transformation of logit cumulative function, the result will be shown below.

$$P(Y \le j \mid \underline{x}) = \left[\frac{\exp(\hat{a}_j + \underline{\hat{b}}'\underline{x})}{1 + \exp(\hat{a}_j + \underline{\hat{b}}'\underline{x})} \right] \quad \text{.....(4)}$$

where $j=1, 2, \ldots, c-1$

or

$$P(Y \le j \mid \underline{x}) = \left[\frac{1}{1 + \exp(-\hat{a}_j - \underline{\hat{b}}'\underline{x})} \right] \quad \text{.....(5)}$$

so that

$$P(Y \le j \mid \underline{x}) = \left[\frac{1}{1 + \exp(-\hat{L}_j(\underline{x}))} \right] \quad \text{.....(6)}$$

Testing the Model Significance

Likelihood ratio test of the overall model is used to assess parameter b_i with hypothesis :

$H_0 : b_1 = \ldots = b_p = 0$

H_1 : at least there is one $b_i \neq 0; i = 1,2,\ldots, p$

where i is the number of explanatory variables.

The likelihood-ratio test uses G statistic, which is G = -2 ln(L_0/L_k) where L_0 is likelihood function without variables and L_k is likelihood function with variables (Hosmer & Lemeshow 2000). If H_0 is true, the G statistic will follow chi-square distribution with p degree of freedom and H_0 will be rejected if value of $G > X^2_{(p,a)}$ or p-value < a .

A Wald test is used to test the statistical significance of each coefficient b_i in the model. Hypothesis are

$H_0 : b_i = 0$

$H_1 : b_i \neq 0; i = 1,\ldots, p$

where i is the number of explanatory variables.

A Wald test calculates a *W* statistic, which is formulated as

$$W_{\hat{b}_i} = \frac{\hat{b}_i}{SE(\hat{b}_i)} \quad \text{.....(7)}$$

Reject null hypothesis if |W| > $Z_{a/2}$ or p-value < a (Hosmer & Lemeshow, 2000).

Assumption of Logistic Regression

Logistic regression is popular in part because it enables the researcher to overcome many of the restrictive assumptions of OLS (Ordinary Least Square) regression:

1. Logistic regression does not assume a linear relationship between the dependent and the independent variables. It can handle nonlinear effects even when exponential and polynomial terms are not explicitly added as additional independents because the logit link function on the left-hand side of the logistic regression equation is non-linear.
2. The dependent variable does not need to be normally distributed (but does assume that its distribution is within the range of the exponential family distributions, such as normal, Poisson, binomial, gamma). Solutions may be more stable if the predictors have a multivariate normal distribution.
3. The dependent variable does not need to be homoscedastic for each level of the independents; that is, there is no homogeneity of variance assumption: variances does not need to be the same within categories.
4. Normally distributed error terms are not assumed.
5. Logistic regression does not require that the independents be an interval scale variable.

However, other assumptions still apply:

1. The data doesn't have any outliers. As in OLS regression, outliers can affect results significantly. The researcher should analyze standardized residuals for outliers and consider removing them or modeling them separately. One way for detecting multivariate outliers is with mahalanobis distance. Mahalanobis distance is the leverage times (n - 1), where n is the sample size. As a rule of thumb, the maximum Mahalanobis distance should not exceed the critical chi-square value with degrees of freedom equal to the number of predictors and a = 0.001, or else outliers may be a problem in the data.

2. Between explanatory variables there should be no multicollinearity: to the extent that one independent is a linear function of another independent, the problem of multicollinearity will occur in logistic regression, as it does in OLS regression. As the correlation among each other increase, the standard errors of the logit (effect) coefficients will become inflated. Multicollinearity does not change the estimates of the coefficients, only their reliability. High standard errors flag possible multicollinearity (www.chass.ncsu.edu).

Biplot Analysis

Biplot similarity provides plots of the n observations, but simultaneously they give plots of positions of the p variables in two dimensions. Furthermore, superimposing the two types of plots provides additional information about relationships between variables and observations not available in either individual plot (Jolliffe, 2002).

The plots are based on the singular value decomposition (SVD). This state that the (n x p) matrices **X** on observations on p variables measured about their sample means can be written

$$\mathbf{X} = \mathbf{ULA}'$$

where **U, A** are *(nxr),(pxr)* matrices respectively, each with orthonormal columns, **L** is an *(rxr)* diagonal matrix with elements $t_1^{1/2} \geq t_1^{1/2} \geq \ldots \geq t_r^{1/2}$, and r is the rank of X.

To include the information on the variables in this plot, we consider the pair of eigenvectors. These eigenvectors are the coefficient vectors for the first two sample principal components. Consequently, each row of matrix positions a variable in the graph, and the magnitudes of the coefficients (the coordinates of the variable) show the weightings that the variable has in each principal component. The positions of the variables in the plot are indicated by a vector.

MATERIAL AND METHODS

Source of Data

The data used in this study were collected from the KNPDT. These data were derived from data Potensi Desa (Podes) 2005 and Survei Sosial Ekonomi nasional (Susenas) 2006 conducted by Central Bureau of Statistics (CBS). The data consists of five categories as response variable and 33 explanatory variables which can be seen in Appendix 1.

Method

The methods used in this research were:

1. Data preparation.
   This step consist of selecting regencies with backward region status namely fairly backward, backward, very backward and the most backward regions.
2. Early data description.
3. The assumption of a logistic regression examination.
4. Data analysis.
   Analyze selected data with ordinal logistic regression. This analysis is conducted for each sub criteria of determining backward region status.
5. Determine the prior factors that influence backward region status.
6. Significant variables were further analyzed through biplot and then explain the relationship of these variables based on globally and part of regions (west and east).

The Software used in this research are Microsoft Excel 2007, Minitab 14, SPSS 13 and SAS 9.1.

RESULTS AND DISCUSSION

Early Description

According to the data released by KNPDT, there are 434 regencies in Indonesia. KNPDT has determined five categories of region index and status based on six major criteria, such as (1) economic, (2) human resources, (3) infrastructures, (4) regional finance, (5) accessibility, and (6) characteristic of region. Each criteria has indicators which are relevant to measure the criteria score. Then the GoI calculated region score with giving weight for each criteria based on their experiences and then multiply it with standardized data.

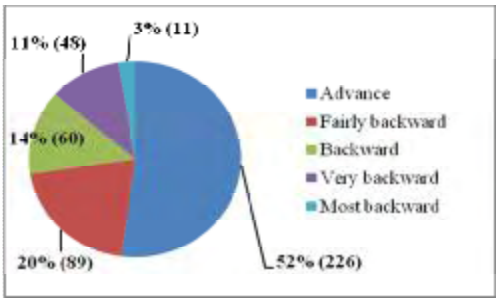

Figure 1. The number and percentage of regency with each status

Regencies with advance status were not used in this analysis because this research