The Optimal Cut-off Point for Predictive Classification of Ungrouped Data with Binary Logistic Regression Model

Durongwatana, Supol
Department of Statistics
Faculty of Commerce and Accountancy
Chulalongkorn University
Bangkok 10330, Thailand
E-mail: Supol@acc.chula.ac.th

Introduction

Recently, the binary logistic regression model has been used to predict the qualitative dependent variable into one of the two categories for the interesting characteristic using independent variables as predictors. Almost all of researches, conducted to classify individual subject or object into one of the two categories, have used the cut-off point equal to 0.5 to minimize[1] classification error rate. For instance, a researcher in finance might be interested in using a group of financial distress to predict a chance of corporate acquisitions[2], a medical researcher would like to predict a chance of Down's syndrome in young pregnant women using maternal serum biomarkers[3], cardiologist would like to predict the disappearance of left Atrial Thrombi among cardiovascular-disease patients[4], and a pediatrician would like to predict the outcome of pregnancies of unknown location[5]. Almost all of problems or researches, conducted to predict one of the two categories using a group of independent variables, normally, has been use the level of 0.5 as the cut-off point. This cut-off point has been usually used for several years. The question is how well this point of the value 0.5 will be appropriate point for classifying into one of the two categories, has come across and sparkling the idea for starting this research.

Binary Logistic Regression Model

The binary logistic regression model for predictive classification is shown as follows:

$$Y_i = \tilde{\mathbf{X}}_i^T \tilde{\boldsymbol{\beta}} + \varepsilon_i ; i = 1,2,...,n$$

$$\tilde{\mathbf{X}}_i^T = \begin{pmatrix} 1 & X_{1i} & X_{2i} & \dots & X_{pi} \end{pmatrix}, \tilde{\boldsymbol{\beta}}^T = \begin{pmatrix} \beta_0 & \beta_1 & \beta_2 & \dots & \beta_p \end{pmatrix}$$

where

$$\Pr\left(Y_i = 1 \mid \tilde{\mathbf{X}}_i^T = \tilde{\mathbf{x}}_i^T\right) = \pi_i, \Pr\left(Y_i = 0 \mid \tilde{\mathbf{X}}_i^T = \tilde{\mathbf{x}}_i^T\right) = 1 - \pi_i ; 0 < \pi_i < 1$$

where Y_i is independently Bernoulli distributed and the mean and variance as shown:

$$E\left(Y_i \mid \tilde{\mathbf{X}}_i^T = \tilde{\mathbf{x}}_i^T\right) = \tilde{\mathbf{x}}_i^T \tilde{\boldsymbol{\beta}} \text{ , } Var\left(Y_i \mid \tilde{\mathbf{X}}_i^T = \tilde{\mathbf{x}}_i^T\right) = \pi_i (1 - \pi_i) ; i = 1,2,...,n$$

---

[1] Efron, B. (1983) Estimating the error rate of a prediction rule: improvement on cross-validation, *Journal of American Statistical Association*, 78, pp. 316-331.

[2] Theodossiou, P., Kahya, E., Saidi, R. & Philippatos, G (1996) Financial distress and corporate acquisitions: Further empirical evidence, Journal of Business Finance and Accounting, 23, pp. 699-719.

[3] Hwa, H.L., Ko, T.M., Hsieh, F.J., Yen, M.F., Chou, K.P. & Tony, H.H. (2007) Risk prediction of Down's syndrome in young pregnant women using maternal serum biomarkers: determination of cut-off risk from receiver operating characteristic curve analysis, Journal of Evaluation in Clinical Practice, 13(2), pp.254-258.

[4] Silaruks, S., Thinkhamrop, B., Tantikosum, W., Wongwipaporn, C., Tatsanavivat, P. & Klungboonkrong, V. (2002) Aprognostic model for predicting the disappearance of left Atrial Thrombi among candidates for percutaneous transvenous mitral commissurotomy, *Journal of the American College of Cardiology*, 39(5), pp. 886-891.

[5] Condous, G., Okaro, E., Khalid, A., Timmerman, D., Lu, C., Zhou, Y., Huffet, S. Van & Bourne, T. (2004) The use of a new logistic regression model for predicting the out come of pregnancies of unknown location, *Human Reproduction*, pp. 1-11.

If regression model is

$$Y_i = \widetilde{\mathbf{X}}_i^T \widetilde{\boldsymbol{\beta}} + \varepsilon_i, E\left(Y_i \mid \widetilde{\mathbf{X}}_i^T = \widetilde{\mathbf{x}}_i^T\right) = \widetilde{\mathbf{x}}_i^T \widetilde{\boldsymbol{\beta}}; i = 1,2,\ldots,n$$

where $$E\left(\varepsilon_i \mid \widetilde{\mathbf{X}}_i^T = \widetilde{\mathbf{x}}_i^T\right) = 0; i = 1,2,\ldots,n$$

The logit function is:

$$\text{Logit}\left(E\left[Y_i \mid \widetilde{\mathbf{X}}_i^T = \widetilde{\mathbf{x}}_i^T\right]\right) = \ln\left(\frac{E\left[Y_i \mid \widetilde{\mathbf{X}}_i^T = \widetilde{\mathbf{x}}_i^T\right]}{1 - E\left[Y_i \mid \widetilde{\mathbf{X}}_i^T = \widetilde{\mathbf{x}}_i^T\right]}\right) = \ln\left(\frac{\pi_i}{1-\pi_i}\right)$$

It shows the binary logistic regression model as follows:

$$E\left(Y_i \mid \widetilde{\mathbf{X}}_i^T = \widetilde{\mathbf{x}}_i^T\right) = \frac{\exp\left(\widetilde{\mathbf{x}}_i^T \widetilde{\boldsymbol{\beta}}\right)}{1 + \exp\left(\widetilde{\mathbf{x}}_i^T \widetilde{\boldsymbol{\beta}}\right)}; i = 1,2,\ldots,n$$

The likelihood function is:

$$l\left(\widetilde{\boldsymbol{\beta}}^T \mid Y_i = y_i, \widetilde{\mathbf{X}}_i^T = \widetilde{\mathbf{x}}_i^T; i = 1,2,\ldots,n\right) = \prod_{i=1}^{n} \left[\pi_i\left(\widetilde{\mathbf{x}}_i^T\right)\right]^{y_i} \left[1 - \pi_i\left(\widetilde{\mathbf{x}}_i^T\right)\right]^{1-y_i}$$

Hence, the log likelihood function is:

$$L(\widetilde{\boldsymbol{\beta}}^T) = \sum_{i=1}^{n} y_i (\widetilde{\mathbf{x}}_i^T \widetilde{\boldsymbol{\beta}}) - \sum_{i=1}^{n} \ln\left[1 + \exp(\widetilde{\mathbf{x}}_i^T \widetilde{\boldsymbol{\beta}})\right]; i = 1,2,\ldots,n$$

Maximum Likelihood Estimation

Since we know the distribution of Y, we can formulate the likelihood function and maximize with respect to the parameters. Estimates do not have closed forms, which means we have to numerically estimate using iterate procedure. Once we get estimates, $\widetilde{\mathbf{b}}^T = \begin{pmatrix} b_0 & b_1 & b_2 & \ldots & b_p \end{pmatrix}$ then calculate the estimate of $\pi_i; i = 1,2,\ldots,n$ is shown in the following:

$$\hat{\pi}_i = \frac{\exp\left(\widetilde{\mathbf{x}}_i^T \widetilde{\mathbf{b}}\right)}{1 + \exp\left(\widetilde{\mathbf{x}}_i^T \widetilde{\mathbf{b}}\right)}; i = 1,2,\ldots,n$$

Use of Estimated Binary Logistic Regression Equation

When the binary logistic regression model is estimated by the maximum likelihood estimates for the parameters in the model as mentioned earlier, it can be used as a predictive classification model by the following formula:

- The i[th] individual will be classified as success group (1) if

$$\hat{\pi}_i = \frac{\exp\left(\widetilde{\mathbf{x}}_i^T \widetilde{\mathbf{b}}\right)}{1 + \exp\left(\widetilde{\mathbf{x}}_i^T \widetilde{\mathbf{b}}\right)} \geq c \text{ where } 0 \leq c \leq 1,$$

- The i[th] individuals will be classified as failure group (0) if

$$\hat{\pi}_i = \frac{\exp\left(\tilde{\mathbf{x}}_i^T \tilde{\mathbf{b}}\right)}{1+\exp\left(\tilde{\mathbf{x}}_i^T \tilde{\mathbf{b}}\right)} < c \text{ where } 0 \le c \le 1,$$

where c is the cut-off point or level of probability that is used to categorize an individual as "failure group" or "success group."

Hadjicostas's Theory[6]

This method is derived to come up with the optimum of cut-off point for classifying object or subject into either success group or failure group regarding to the minimum of classification error rate. Assume that the data have been sorted in ascending order according to the value of $\hat{\pi}$, i.e. without loss of generality assume that $\hat{\pi}_1 < \hat{\pi}_2 < \ldots < \hat{\pi}_n$. (Note that the following results assume that, if $\hat{\pi}_i$ equals the cut-off point, then the predicted group is 0. If one alters this convention, the results have to be modified appropriately.) For each $i \in \{1,2,\ldots,n\}$, let $M(i)$ be the maximum $j \in \{1,2,\ldots,n\}$ such that $\hat{\pi}_i = \hat{\pi}_j$. Let also $M(0) = 0$. Obviously, $i \le M(i) \le n$.

For any $c \in [0,1]$, let $p(c)$ (and $N(c)$) be the overall proportion (and the overall number, respectively) of correct classification corresponding to c. Note that $p(c) = \dfrac{N(c)}{n}$.

Also, for $i \in \{0,1,2,\ldots,n\}$, let $A_i = [0,\hat{\pi}_1)$ if $i = 0$; $A_i = [\hat{\pi}_i, \hat{\pi}_{i+1})$ for $\hat{\pi}_i < \hat{\pi}_{i+1}$ and $1 \le i < n$; $A_i = \{\hat{\pi}_i\} = \{\hat{\pi}_{M(i)}\}$ if $\hat{\pi}_i = \hat{\pi}_{i+1}$ and $1 \le i < n$; $A_i = [\hat{\pi}_n, 1]$ if $i = n$. Notice that $\bigcup_{i=0}^{n} A_i = [0,1]$.

Lemma 1: For any $i \in \{0,1,2,\ldots,n\}$ and $c \in A_i$,

$$N(c) = \sum_{j=1}^{M(i)} \left(1 - y_j\right) + \sum_{j=M(i)+1}^{n} y_j \tag{1}$$

where a sum over the empty set is zero by definition.

Theorem 1: Let $a_i = \sum_{k=1}^{M(i)} (-1)^{y_k}$ for $i = 0,1,2,\ldots,n$. Let I_0 be the set of all $j \in \{0,1,2,\ldots,n\}$ such that $a_j = \max_{0 \le i \le n} a_i$ and let C_0 be the set of all $c_0 \in [0,1]$ such that $p(c_0) = \max_{c \in [0,1]} p(c)$. Then $C_0 = \bigcup_{i \in I_0} A_i$

Theorem 2: Assume $w_0, w_1 \ge 0$ and $w_0 + w_1 = 1$. Let $d_i = \sum_{j=1}^{M(i)} (1 - y_i)$ and $f_i = \sum_{j=M(i)+1}^{n} y_i$ for $i = 0,1,2,\ldots,n.$ Let J_0 be the set of all $j \in \{0,1,2,\ldots,n\}$ such that

$$\tilde{w}_0 d_j + \tilde{w}_1 f_j = \max_{0 \le i \le n}\left(\tilde{w}_0 d_i + \tilde{w}_1 f_i\right)$$

where $\tilde{w}_k = \dfrac{w_k}{n_k}$ for $k = 0, 1$. Let also D_0 be the set of all $c_0 \in [0,1]$ such that

$$w_0 p_0(c_0) + w_1 p_1(c_0) = \max_{c \in [0,1]}\left(w_0 p_0(c) + w_1 p_1(c)\right).$$

[6] Hadjicostas,P. (2006) Maximizing proportions of correct classifications in binary logistic regression, *Journal of Applied Statistics*, 33, pp.629-640.

Then $D_0 = \bigcup_{i \in J_0} A_i$.

Corollary 1: Consider the notation and assumptions of Theorem 2. Let J'_0 be the set of all $j \in \{0,1,...,n\}$ such that

$$\frac{\widetilde{w}_0}{\widetilde{w}_0 + \widetilde{w}_1} M(j) - \sum_{m=1}^{M(j)} y_m = \max_{0 \le i \le n} \left(\frac{\widetilde{w}_0}{\widetilde{w}_0 + \widetilde{w}_1} M(j) - \sum_{m=1}^{M(j)} y_m \right),$$

where $\widetilde{w}_k = w_k / n_k$ for $k = 0,1$. then $J_0 = J'_0$ and $D_0 = \cup_{i \in J_0} A_i$.

Measure of Association between Predicted Probabilities and Observe Responses

Consider a pair of observations $(y_i, x_{1i}, ..., x_{pi})$ and $(y_k, x_{1k}, ..., x_{pk})$ where $(1 \le i \ne k \le n)$ with different responses, i.e. assume $y_i \ne y_k$. Let $\hat{\pi}_i$ and $\hat{\pi}_k$ be the corresponding predicted probabilities. If $\hat{\pi}_i = \hat{\pi}_k$, the pair is call *tied*. If either $y_i > y_k$ and $\hat{\pi}_i > \hat{\pi}_k$, or $y_i < y_k$ and $\hat{\pi}_i < \hat{\pi}_k$, the pair is called *concordant*. If the pair is neither tied nor concordant, then it is called *discordant*.

Denote by m_c, m_d, and m_t the number of concordant, discordant, and tied pairs of observations with different y. Since this method does not always give the exact values of m_c, m_d, and m_t, we will not use it.

The following theorem gives some formulas regarding the number of concordant, discordant and tied pairs of observations with different responses. Note that the order of the indices i and j in the double sums of parts (b), (c) and (d) is very important. For a real number r denote by $\lfloor r \rfloor$ the greatest integer less than or equal to r.

Theorem 3: Assume $\hat{\pi}_1 \le \hat{\pi}_2 \le ... \le \hat{\pi}_n$. Then:

(a) $m_c + m_d + m_t = (\sum_{i=1}^{n} y_i)(n - \sum_{i=1}^{n} y_i)$.

(b) $m_t = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} |y_i - y_j| \lfloor 1 + \hat{\pi}_i - \hat{\pi}_j \rfloor$.

(c) $m_c = (1/2) \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} (-1 + (-1)^{\min(y_i - y_j, 0)}) \lfloor \hat{\pi}_i - \hat{\pi}_j \rfloor$.

(d) $m_d = (1/2) \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} (-1 + (-1)^{\max(y_i - y_j, 0)}) \lfloor \hat{\pi}_i - \hat{\pi}_j \rfloor$.

Corollary 2: Assume $\hat{\pi}_1 < \hat{\pi}_2 < ... < \hat{\pi}_n$. Then:

(a) $m_c + m_d = (\sum_{i=1}^{n} y_i)(n - \sum_{i=1}^{n} y_i)$.

(b) $m_c = n - (n-1)/4 - (1/2) \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} (-1)^{\min(y_i - y_j, 0)}$.

(c) $m_d = n - (n-1)/4 - (1/2) \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} (-1)^{\max(y_i - y_j, 0)}$.

Optimal Cut-Off Point Procedure

This method can be shown in the following steps:

- Order the value $\hat{\pi}_i$ in ascending order $\hat{\pi}_1 < \hat{\pi}_2 < ... < \hat{\pi}_n$ where if $\hat{\pi}_i$ is the cut-off point, then classify for the failure group as interesting group $(Y = 0)$.

- For $i \in \{1,2,...,n\}$, find the value $M(i)$. $M(i)$ is $\max j \in \{1,2,...,n\}$ if $\hat{\pi}_i = \hat{\pi}_j$ where $M(0) = 0$; $i \le M(i) \le n$
- For $i = 0,1,2,...,n$, find the value $a_i = \sum_{k=1}^{M(i)} (-1)^{y_k}$ which there are 2 cases as shown:
- $a_{i+1} = a_i + \sum_{k=M(i)+1}^{M(i+1)} (-1)^{y_k}$ when $M(i) < i+1$
- $a_{i+1} = a_i$ when $i+1 \le M(i)$
- Find I_0 which is the set of all $j \in \{0,1,2,...,n\}$ where $a_j = \max_{0\le i\le n} a_i$
- Find C_0 which is the set of all $c_0 \in [0,1]$ where if $p(c_0) = \max_{c\in[0,1]} p(c)$, then $C_0 = \bigcup_{i\in I_0} A_i$

Kaiser-Meyer-Olkin Measure of Sampling Adequacy[7]

This KMO measure is an index for comparing the magnitudes of observed simple correlation coefficients among the independent variables to the magnitudes of their observed partial correlation coefficients. Normally this measure is used for measuring the inter-correlation both simply and partially among independent variables in multivariate analysis named as the factor analysis. Large values for the KMO measure indicate that a factor analysis of the variables is a good idea. The computation of this measure is shown as follows:

$$0 < \text{KMO} = \frac{\sum_{i<j=1}^{p} r_{ij}^2}{\sum_{i<j=1}^{p} r_{ij}^2 + \sum_{i<j}^{p} r_{ij|1,2,...,(j-1),(j+1),...,p}^2} < 1$$

where

$$\mathbf{r}_{p\times p} = \begin{pmatrix} 1 & r_{12} & r_{13} & \cdots & \cdots & r_{1p} \\ r_{21} & 1 & r_{23} & \cdots & \cdots & r_{2p} \\ r_{31} & r_{32} & 1 & \cdots & \cdots & r_{3p} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ r_{p1} & r_{p2} & r_{p3} & \cdots & \cdots & 1 \end{pmatrix}$$

r_{ij} is the sample coefficient of simple correlation between the i^{th} independent variable and the j^{th} independent variable for $i < j = 1,2,..., p$, and $r_{ij|1,2,...,(j-1),(j+1),...,p}$ is the sample coefficient of partial correlation between the i^{th} independent variable and the j^{th} independent variable while keeping the rest of independent variables constant or unchanged for $i < j = 1,2,..., p$. This measure was refined and suggested[8] that if it is between 0.90 and 1.00 then it is considered to have surprisingly high degree of multicollinearity, if it is between 0.80 and 0.89 then it is considered to have excellently high degree of multicollinearity, if it is between 0.70 and 0.79 then it is considered to have medium degree of multicollinearity, if it is between 0.60 and 0.69 then it is considered to have indifferently degree of multicollinearity, if it is between 0.50 and 0.59 then it is considered to have miserably degree of

[7] Cerny, B.A. and Kaiser, H.F. (1977), "A Study of a Measure of Sampling Adequacy for Factor-Analytic Correlation Matrices," *Multivariate Behavioral Research*, 12, 289-325.

[8] Kaiser, H.F., and J. Rice (1974), "Little Jiffy Mark IV," *Educational and Psychological Measurement*, 34, 111-117.

multicollinearity, and finally if it is between 0 and 0.49 then it is considered to have little degree of multicollinearity.

Simulation Design

Each set of data is generated using R program according to the number of independent variables (p), the failure rate of the data set (a), the sample size (n), the degree of inter-correlation among the independent variables (KMO), and the distribution of each independent variable. Each factor is designed as shown in the following:

- Number of independent variables (p) is simulated having $p = 1, 2, 3, 4, 5,$ and 6.
- Failure rate of data set (a) is generated with $a = 0.1, 0.3, 0.5, 0.7,$ and 0.9.
- Sample size (n) of each data set is designed having $n = 20, 40, 60, 80, 100,$ and 120.
- Degree of inter-correlation among the independent variables is designed with $\mathrm{KMO} = 0, 0.33, 0.67,$ and 0.99.
- The distribution of the observations for each independent variable is designated to have its spacing structure with the uniform distribution.

Results of Simulation Study

For each run, the optimal cut-off point is located using the Hadjicostas's theory. . The average from 500 runs of all located cut-off points is computed for each situation. Then, the multiple linear regression equation is fitted using the data from all simulated situations. The results are as follows:

- As the number of independent variables increases, the average of the cut-off points will decrease in the neighborhood of 0.5 while keeping the other factors constant.
- As the sample size increases, the average of the cut-off points will also decrease in the neighborhood of 0.5 while keeping the other factors unchanged.
- As the failure rate of the data set increases, the average of the cut-off points will decrease in the neighborhood of 0.5 while keeping the other factors unchanged.
- As the KMO measure of sample adequacy increases, the average of the cut-off points will decrease in the neighborhood of 0.5 while keeping the other factors constant.

Use of the Fitted Binary Logistic Regression Equation

From the estimated binary logistic regression equation, one can simply apply the equation to find the optimal cut-off point for a selected set of independent variables for predicting the probability of success by computing the KMO degree of sampling adequacy, putting the failure rate of the data set, the sample size, and the number of independent variables into the estimated equation. Then the estimated average value of the cut-off point could be predicted. Finally, one can also use that estimated cut-off point as a criterion for classification of individual subject or object into either success group of failure group.

REFERENCES

Journals

Agresti, A., 1999. On Logic Confidence Intervals for the Odds Ratio with Small Samples. **Biometrics** 55: 597 – 602.

Alishiri, G.H., Bayat, N., Ashtiani, A.F., Tavallaii, S.A., Assari, S., and Moharamzad, Y., 2008. Logistic Regression Models for Predicting Physical and Mental Health-related Quality of Life in Rheumatoid Arthritis Patients. **Mod Rheumatol** 18: 601-608.

Anderson, T.W., 1951b. Classification by Multivariate Analysis. **Psychometrica** 16: 13 -50.

Anderson, T.W., and Bahadur, R.R., 1962. Classification into Two Multivariate Normal Distributions with Different Covariance Matrices. **Annals of Mathematical Statistics** 33: 420 -431.

Cerny, B.A. and Kaiser, H.F., 1977. A Study of a Measure of Sampling Adequacy for Factor-Analytic Correlation Matrices. **Multivariate Behavioral Research.** 12: 289-325.

Kaiser, H.F., and Rice, J., 1974. Little Jiffy Mark IV. **Educational and Psychological Measurement**. 34: 111-117.

Chatrchaiwiwatana, S., and Ratanasiri, A., 2008. Exhaled Carbon Monoxide Level and Smoking Status in Urban Khon Kaen Adults. **Journal of Thai Medical Association** 91(11): 1669-1676.

Condous, G., Okaro, E., Khalid, A., Timmerman, D., Lu, C., Zhou, Y., and Huffel, S.V., 2004. The Use of a New Logistic Regression Model for Predicting the Outcome of Pregnancies of Unknown Location. **Human Reproduction: European Society of Human Reproduction and Embryology**: 1-11.

Duran, E.H., Gürgan, T., Günalp, S., Enginsu, M.E., Yarali, H., and Ayhan, A., 1998. A logistic Regression Model Including DNA Status and Morphology of Spermatozoa for Prediction of Fertilization *In Vitro*. **Human Reproduction** 13(5): 1235-1239.

Efron, B., 1983. Estimating the error rate of a prediction rule: improvement on cross-validation. **Journal of American Statistical Association**. 78: 316-331.

Estrella, A., 1998. A New Measure of Fit for Equations with Dichotomous Dependent Variables. **Journal of Business and Economic Statistics** 16(2): 198 – 205.

Gorsevski, P.V., Gessler, P.E., Foltz, R.B., and Elliot, W.J., 2006. Spatial Prediction of Landslide Hazard Using Logistic Regression and ROC Analysis. **Transactions in GIS** 10(3): 395-415.

Hadjicostas,P., 2006. Maximizing proportions of correct classifications in binary logistic regression. **Journal of Applied Statistics**. 33: 629-640.

Hirji, K.F., Mehta, C.R. and Patel, N.R., 1987. Computing distributions for exact logistic regression. **Journal of the American Statistical Association** 82: 1110-1117.

Hirji, K.F., Tsiatis, A.A. and Mehta, C.R., 1989. Median unbiased estimation for binary data. **The American Statistician** 43: 7-11.

Hwa, H.L., Ko, T.M., Hsieh, F.J., Yen, M.F., Chou, K.P. & Tony, H.H. 2007. Risk prediction of Down's syndrome in young pregnant women using maternal serum biomarkers: determination of cut-off risk from receiver operating characteristic curve Analysis. **Journal of Evaluation in Clinical Practice**. 13(2): 254-258.

Isaranurag, S., Klinman, S., Chompikul, J., Nantamongkolchai, S., and Plubrukan, R., 2006. Implication of Family Protective-risk Index for Screening Cognitive Development of Children Aged 13-15 Years. **Journal of Thai Medical Association** 89(9): 1427-1433.

Kaiser, H.F., and Rice, J., 1974. Little Jiffy Mark IV. **Educational and Psychological Measurement**. 34: 111-117.

Press, S.J., and Wilson, S., 1978. Choosing Between Logistic Regression and Discriminant Analysis. Journal of the American Statistical Association. 73: 699-705.

Sharifi, F., Mousavinasab, N., Mazloomzadeh, S., Jaberi, Y., Saeini, M., Dinmohamadi, M., and Angomshoaa, A. 2008. Cutoff Point of Waist Circumference for the Diagnosis of Metabolic Syndrome in an Iranian Population. **Obesity Research and Clinical Practice** 2: 171-178.

Silaruks, S., Thinkhamrop, B., Tantikosum, W., Wongwipaporn, C., Tatsanavivat, P. and Klungboonkrong, V., 2002. A Prognostic Model for Predicting the Disappearance of Left Atrial Thrombi among Candidates for Percutaneous Transvenous Mitral Commissurotomy. **Journal of the American College of Cardiology**. 39(5): 886-891.

Theodossiou, P., Kahya, E., Saidi, R. & Philippatos, G., 1996. Financial distress and corporate acquisitions: Further empirical evidence. **Journal of Business Finance and Accounting**. 23: 699-719.

Tuan, Nguyen T., Adair, L.S., He, K., and Popkin, B.M., 2008. Optimal Cutoff Values for Overweight: Using Body Mass Index to Predict Incidence of Hypertension in 18- to 65-Year-Old Chinese Adults. **The Journal of Nutrition: Nutritional Epidemiology**, May 2: 1377-1382.

Books

Agresti, A., 2002. **Categorical Data Analysis**. 2nd Edition, New York: Wiley.

Cox, D.R., and Snell, E.J., 1989. **Analysis of Binary Data**. 2nd Edition, London: Chapman Hall.

Dielman, T.E., 2005. **Applied Regression Analysis.** 4th Edition, Boston: Thomson Learning.

Gill, P.E., Murray, W., and Wright, M.H., 1991. **Numerical Linear Algebra and Optimization**. 1, Redwood City, CA: Addison-Wesley.

Hosmer, D., and Lemeshow, S., 1989. **Applied Logistic Regression**. New York: John Wiley & Sons.

Jambu, M., 1989. **Exploratory and Multivariate Data Analysis**. Boston: Academic Press.

Johnson, R.A., and Wichern, D.W., 2002. **Applied Multivariate Methods for Data Analysis**. 5th Edition, New York: Prentice Hall.

Kleinbaum, D.G., 1994. **Logistic Regression: A Self-learning Text**. New York: Springer-Verlag.

Kleinbaum, D.G., Kupper, L.L., Muller, K.E., and Nizan, A., 1998. **Applied Regression Analysis and Other Multivariable Methods**. 3rd Edition, Boston: Duxbury Press.

Mennard, S., 2002. **Applied Logistic Regression Analysis**. 2nd Edition, 16: Sage Publication, Series: Quantitative Applications in the Social Sciences.

Morrison, D.F., 2005. **Multivariate Statistical Methods**. 4th Edition, Boston: Duxbury.

O' Connell, A.A., 2005. **Logistic Regression Models for Ordinal Response Variables**. 146, Thousand Oaks, California: Sage Publication, Series: Quantitative Applications in the Social Science