Factors Affecting College Students Applied Suspended or Withdraw From School in Taiwan

Chun-Min Hung1, Chun-Yen Chung2, Yan-Kuin Su3, Chin-Chuan Lin4*

1(Department Of Information Management, Kun Shan University, Taiwan)
2(Department Of Information Engineering, Kun Shan University, Taiwan)
3(President, Kun Shan University, Taiwan)
4(Department Of Business Administration, Kun Shan University, Taiwan)

Abstract: The present study investigates factors that might affect college students who are suspended or apply to withdraw from school in Taiwan based on the institute research database of KS University. The logistic regression method which, associated by information geometry, was conducted to analyze the data. This modified logistic regression method can efficiently reduce the number of factors and recognize significant factors. Results showed that the class attendance and the interaction of class attendance and academic performance were the factors that significantly affect college students who are suspended or apply to withdraw in Taiwan.

Keywords: Factors; Suspension; Withdrawal; Logistic regression model; Institutional research

I. Introduction

Problems stemming from students’ suspension or withdrawal have become a very import issue for private colleges in Taiwan. To better understand the causes of suspension and withdrawal, Hung et al. [1] proposed a conceptual model (Fig. 1) intended to reveal the real causes of suspension or withdrawal of college students in Taiwan. Further, Hung et al. [2] verified the model based on empirical data and indicated that the decision model (Fig. 2) can best reflect the current situation in Taiwan.

However, the results obtained by Hung et al. [2] only indicated the significance coefficients of path between the factors rather than recognizing the impact of factors. Therefore, the present study evaluates the impact of factors.
II. Literature Review

There are few studies employing quantitative methods to investigate possible causes of college students being suspended. Zheng [3] reported that the major causes of suspension were personal interests, aptitudes, and self-expectations; the environmental causes included life stress, crisis situations, and family factors. Hung et al. [1] employed path analysis to verify the decision model they proposed based on empirical data, and results showed that the path analysis was valid to quantitative analysis of the model [1]. However, Hung et al. [2] did not analyze the effect of each individual factor.

To obtain the multivariate multinomial distribution of a contingency table, an information identity is defined as a decomposition of the log-likelihood as a sum of mutually orthogonal terms of relative entropy [4,5]. The mutual information (MI) identity has been developed based on the invariant Pythagorean laws [6] for relative entropy for testing two-way independence and three-way conditional independence, as geometric counterparts to the classical Pearson chi-square tests [7,8]. An extension to multi-way tables can be carried out by analogy to examine associations among variables through testing low- and high-order association effects. Due to the close connection between the contingency table and logistic regression (LR) analyses, information identities can be applied to provide a geometric approach for model selection. This geometrically supported LR analysis method (linear information model, LIM) can efficiently reduce the number of variables to reduce calculation time.

III. Methodology

In this section, we introduce the theory and the MI identities used in the following LIM analysis method, which characterize the geometry of association between categorical variables.

Let \((X,Y,Z)\) denote a three-way \(I \times J \times K\) contingency table with the joint probability density function \(f_{X,Y,Z}(i,j,k)\), \(i=1,...,I, j=1,...,J, k=1,...,K\). The Shannon entropy [9] defines the basic information identity:

\[
I(X,Y,Z) = H(X) + H(Y) + H(Z) - H(X,Y,Z) \tag{1}
\]

where \(H(X) = -\sum_{i} f(i) \log f(i)\) is the marginal probability density function (p.d.f.), \(\{f, g, h\}\) being the marginal probability density function (p.d.f.), is the MI of \((X,Y,Z)\) [10,11].

The MI defines the minimum divergence from the joint p.d.f. to the product space of marginal p.d.f., i.e., the parameter space of the null hypothesis of independence [7,12]. Taking \(Z\) as the conditioning variable, the MI can be expressed as the sum of three orthogonal components:

\[
I(X,Y,Z) = I(X,Z) + I(Y,Z) + I(X,Y|Z) \tag{2}
\]

where the variables may be exchanged to yield three information-equivalent forms. The conditional mutual information (CMI) \(I(X,Y|Z)\) of equation (2) defines the expected log-likelihood ratio (deviance) from the data distribution to the parameter space of conditional independence between \(X\) and \(Y\) across the levels of \(Z\), that is,

\[
I(X,Y|Z) = -\sum_{i,j,k} f(i,j,k) \log \frac{f(i,j,k)}{f(i)g(j)h(k)}
\]

This can be further decomposed as the sum of two orthogonal components:

\[
I(X,Y|Z) = \text{Int}(X,Y|Z) + I(X,Y||Z) \tag{3}
\]

Here, \(\text{Int}(X,Y,Z)\) is the three-way interaction between the three variables, which is defined as the fitted projection table from the raw data table by the classical iterated proportional fit [13,14]. And, \(I(X,Y||Z)\), obtained as the difference of the other two terms, defines the uniform association between \(X\) and \(Y\) across the levels of \(Z\), also termed the partial association between \(X\) and \(Y\) given \(Z\).

IV. Practical data Analysis

An institutional research (IR) was conducted to evaluate the status of college students in KS University being suspended or withdraw from school in 2012. It was of interest to examine how factors affected the prediction of suspension and withdrawal in an LR model.

Based on data likelihood decomposition, an information approach supported by geometry theory for selecting the main and interaction effects of the predicting variables is introduced to the LR analysis. Therefore, the present study employed the LR method in association with geometry theory deployed to analyze the IR data.

4.1. Data And Codes

In the IR database of KS University, a nominal variable is used to define the status of the result (response) variable (suspension=1, withdrawal=2, or others=3), denoted by \(R = 1, 2, \text{ or } 3\). Seven predictors, each coded as “1 to 2 or 1 to 5” are used, thus, the data consists of a multivariate contingency table of seven variables, having 320 cells and total counts 9,598. Table I lists the codes of seven prediction variables and result variable, and their description.
Factors Affecting College Students Applied Suspended Or Withdraw From School In Taiwan

Table I. List Of Codes And Descriptions Of The Variables

<table>
<thead>
<tr>
<th>Variable</th>
<th>Code</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Major in high school</td>
<td>MS, 1</td>
<td>1: ordinary high schools; 2: vocational high schools.</td>
</tr>
<tr>
<td>Enter way</td>
<td>EW, 2</td>
<td>1: unified exam; 2: direct application.</td>
</tr>
<tr>
<td>Living city</td>
<td>LC, 3</td>
<td>1: Tainan city; 2: other cities.</td>
</tr>
<tr>
<td>Gender</td>
<td>GE, 4</td>
<td>1: Male; 2: Female.</td>
</tr>
<tr>
<td>Class attendance</td>
<td>CA, 6</td>
<td>1: number of class absence &lt; 5; 2: others.</td>
</tr>
<tr>
<td>Academic performance</td>
<td>AP, 7</td>
<td>1: average scale &gt; 60; 2: others.</td>
</tr>
</tbody>
</table>

4.2. Classical LR Analysis

Table II shows the association between R (response variable) and each individual prediction variable.

Table II. Association between R and each individual factor

<table>
<thead>
<tr>
<th>Code</th>
<th>Description</th>
<th>Level</th>
<th>Response</th>
</tr>
</thead>
<tbody>
<tr>
<td>MS, 1</td>
<td>Major in high school</td>
<td>1</td>
<td>131</td>
</tr>
<tr>
<td>EW, 2</td>
<td>Enter way</td>
<td>2</td>
<td>230</td>
</tr>
<tr>
<td>LC, 3</td>
<td>Living city</td>
<td>1</td>
<td>189</td>
</tr>
<tr>
<td>GE, 4</td>
<td>Gender</td>
<td>2</td>
<td>172</td>
</tr>
<tr>
<td>ED, 5</td>
<td>Enrolled department</td>
<td>1</td>
<td>148</td>
</tr>
<tr>
<td>CA, 6</td>
<td>Class attendance</td>
<td>2</td>
<td>213</td>
</tr>
<tr>
<td>AP, 7</td>
<td>Academic performance</td>
<td>1</td>
<td>212</td>
</tr>
<tr>
<td>AP, 7</td>
<td>Academic performance</td>
<td>2</td>
<td>149</td>
</tr>
<tr>
<td>ED, 5</td>
<td>Enrolled department</td>
<td>1</td>
<td>88</td>
</tr>
<tr>
<td>GE, 4</td>
<td>Gender</td>
<td>2</td>
<td>37</td>
</tr>
<tr>
<td>GE, 4</td>
<td>Gender</td>
<td>3</td>
<td>52</td>
</tr>
<tr>
<td>CA, 6</td>
<td>Class attendance</td>
<td>5</td>
<td>92</td>
</tr>
<tr>
<td>AP, 7</td>
<td>Academic performance</td>
<td>1</td>
<td>354</td>
</tr>
<tr>
<td>AP, 7</td>
<td>Academic performance</td>
<td>2</td>
<td>144</td>
</tr>
</tbody>
</table>

A full model of (3) for the case of using seven predictors \{x_1, x_2, ..., x_7\} for the response variable R is:

\[
R(\{x_1, x_2, x_3, x_4, x_5, x_6, x_7\}, R) = 7 x_1 + 21 x_2 + 35 x_3 x_2 + 35 x_4 x_2 x_1 + 21 x_5 x_3 x_4 a_1 + 7 x_6 x_4 x_3 a_1 a_2 + 1 x_7 x_4 x_3 x_2 a_1 a_2 a_3 + \cdots + 780
\]

The full model of classical LR analysis for equation (4) would include 7 main effects, 21 two-order interactions, 35 three-order interactions, 35 four-order interactions, 21 five-order interactions, 7 six-order interactions, and 1 seven-order interaction. It would take a very long time to obtain the full model results, furthermore, the full model result would be too complex to interpret.

If we select the factors one by one, this still too complex to calculate the relationship between the seven prediction variables and the response variable because any variable of the seven predict variables can be coded as \(x_1, x_2, ..., x_7\). In this case there would be 5,040 (!!) combinations that need to be calculated. Thus, how to select variables efficiently for the LR model is clearly the major problem.
Thus, classical LR analysis methods usually assume that the high-order interactions are insignificant. However, this might miss some significant high-order interactions and lead to incorrect interpretation.

Therefore, how to reduce items without missing significant interactions is very important, especially for significant high-order interactions.

4.3. Lim Analysis
A basic approach is to eliminate redundant predictors and to test LR models that can be interpreted using the least number of significant interaction terms. A straightforward extension of the MI identities in (4) to highway tables will be examined for the IR data analysis of this study. An extension of the first equation of (4) to the case of using seven predictors \{x_1, x_2, ..., x_7\} for the response variable R is:

\[
I(R(x_1, x_2, x_3, x_4, x_5, x_6, x_7), R) = I(R(x_1), R(x_2)) + I(R(x_2), R(x_1)) + I(R(x_3), R(x_1)) + I(R(x_4), R(x_1)) + I(R(x_5), R(x_1)) + I(R(x_6), R(x_1)) + I(R(x_7), R(x_1)) \]

(5)

Identity (5) is constructed by the rule of selecting the first least significant (7th order) conditional MI (CMI) term, then selecting the least significant 6th order CMI term, and continuing until the last 2nd order CMI term \(I(R, x_{j-1}|x_j)\). Table III shows the calculation of decomposed of identity (5) for each variables.

| Var. | \(I(R, x_{j-1}|x_j)\) | \(Int(R, x_{j-1}|x_j)\) | \(I(R, x_{j-1}|x_j)\) |
|------|-------------------|-------------------|-------------------|
|      | Likelihood ratio  | df                | Sig.              | Likelihood ratio  | df                | Sig.              |
| MS, x_1 | 232.719          | 320               | 1.000             | 230.414          | 318               | 1.000             | 2.305          | 2.000             | 0.316             |
| EW, x_1 | 254.648          | 320               | 0.997             | 254.050          | 318               | 0.997             | 0.598          | 2.000             | 0.742             |
| LC, x_1 | 267.606          | 320               | 0.985             | 264.277          | 318               | 0.987             | 3.379          | 2.000             | 0.185             |
| GE, x_1 | 203.915          | 320               | 1.000             | 203.605          | 318               | 1.000             | 0.310          | 2.000             | 0.856             |
| ED, x_1 | 442.330          | 512               | 0.988             | 420.270          | 504               | 0.997             | 22.060         | 8.000             | 0.000             |
| CA, x_6 | 1631.292         | 320               | 0.000             | 423.085          | 318               | 0.000             | 1208.207       | 2.000             | 0.000             |
| AP, x_1 | 500.579          | 320               | 0.000             | 463.513          | 318               | 0.000             | 37.066         | 2.000             | 0.000             |

The efficiency way to select an initial sequence of variables is according to the significance level of the main and interaction term based on MI and CMI. Therefore, the identity (5) is updated as:

\[
I(R, \{x_1, x_2, x_3, x_4, x_5, x_6, x_7\}) = Int(R, \{x_1, x_2, x_3, x_4, x_5, x_6, x_7\})
\]

(6)

4.4. Determine Initial Sequence Of Factors
Table III showed that the CA is most significant factor. Therefore, the CA factor is first entered into the LR model. Then repeating the calculation procedure (Appendix Table I-VI), the factor AP is entered into the LR model secondly, followed by ED, MS, LC, GE and EW. When the entry sequence of variables is determined, the following problem is the selection of a proper LR model.
Factors Affecting College Students Applied Suspended Or Withdraw From School In Taiwan

### Table IV. Sequential decomposed CMI components of identity (5)

<table>
<thead>
<tr>
<th>ML CMI Terms</th>
<th>( \text{CMI} )</th>
<th>( \text{df} )</th>
<th>( \text{Sig.} )</th>
<th>Interaction</th>
<th>( \text{df} )</th>
<th>( \text{Sig.} )</th>
<th>Partial Asso.</th>
<th>( \text{df} )</th>
<th>( \text{Sig.} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( R, x_0 )</td>
<td>1631.292</td>
<td>320</td>
<td>0.000</td>
<td>423.085</td>
<td>318</td>
<td>0.000</td>
<td>1208.207</td>
<td>2</td>
<td>0.000</td>
</tr>
<tr>
<td>( R, x_0 )</td>
<td>1631.292</td>
<td>320</td>
<td>0.000</td>
<td>423.085</td>
<td>318</td>
<td>0.000</td>
<td>1208.207</td>
<td>2</td>
<td>0.000</td>
</tr>
<tr>
<td>( R, x_0 )</td>
<td>766.894</td>
<td>160</td>
<td>0.000</td>
<td>187.588</td>
<td>158</td>
<td>0.054</td>
<td>579.306</td>
<td>2</td>
<td>0.000</td>
</tr>
<tr>
<td>( R, x_0 )</td>
<td>189.369</td>
<td>128</td>
<td>0.000</td>
<td>155.139</td>
<td>120</td>
<td>0.017</td>
<td>34.230</td>
<td>8</td>
<td>0.000</td>
</tr>
<tr>
<td>( R, x_0 )</td>
<td>17.703</td>
<td>16</td>
<td>0.342</td>
<td>9.289</td>
<td>14</td>
<td>0.812</td>
<td>8.414</td>
<td>2</td>
<td>0.014</td>
</tr>
<tr>
<td>( R, x_0 )</td>
<td>7.445</td>
<td>8</td>
<td>0.490</td>
<td>5.587</td>
<td>6</td>
<td>0.471</td>
<td>1.858</td>
<td>2</td>
<td>0.395</td>
</tr>
<tr>
<td>( R, x_0 )</td>
<td>1.959</td>
<td>4</td>
<td>0.743</td>
<td>1.444</td>
<td>2</td>
<td>0.486</td>
<td>0.515</td>
<td>2</td>
<td>0.773</td>
</tr>
<tr>
<td>( R, x_0 )</td>
<td>0.056</td>
<td>2</td>
<td>0.973</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

### 4.5. Selection Of A Proper LR Model

An LR model is constructed from identity (5) using the hierarchical set of variable parameters \( \{ x_6, x_7, x_8, x_9, x_{10}, x_{11}, x_{12}, x_{13}, x_{14}, x_{15}, x_{16}, x_{17}, x_{18}, x_{19}, x_{20}, x_{21}, x_{22}, x_{23}, x_{24}, x_{25}, x_{26} \} \) as identity (7).

\[
\text{Logit}(R) = -1.476^* x_0 + 4.464 x_6^* + 0.181 x_7 - 2.323 x_8^* + 0.172 x_9 + 0.103 x_9 x_7 - 0.044 x_1 - 1.000 x_9 x_7 x_1 - 0.094 x_1^2 + 1.064 x_9 x_7 x_1 x_2 + 0.328 x_7 + 0.976 x_9 x_7 x_1 x_3 x_2 + 0.005 x_2 \ 
\]

Identity (7) shows only the main effect of \( x_0 \) and interaction of \( x_9 x_7 \) had reached the statistical significant level \( p<0.01 \). In contrast, the other variables and their interactions did not reach the level of statistical significance.

### V. Results

Class attendance \( (x_6) \) significantly affects the students’ being suspended or applying for withdrawal. The negative sign indicates that a student with \( x_6=2 \) was more likely to be suspended or to withdraw than \( x_6=1 \). In other words, the student will be more likely to suspend or to withdraw when they have more than 5 class absences.

The interaction effect of \( x_0 x_7 \) significantly affects the student’s application for temporary suspension or permanent withdrawal. Table V shows the association between CA and AP with response. Table V indicated that with students who have fewer than 5 class absences and an average score greater 60, only about 0.36% applied for suspension or withdrawal. Conversely, the students are more likely to apply for temporary suspension or withdrawal when they have more than 5 class absences.

### Table V. Association Between CA And AP With R

<table>
<thead>
<tr>
<th>Factor</th>
<th>Response</th>
</tr>
</thead>
<tbody>
<tr>
<td>CA, x_0</td>
<td>AP, x_7</td>
</tr>
<tr>
<td>Suspended</td>
<td>Withdrawal</td>
</tr>
<tr>
<td>1</td>
<td>13</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>1</td>
<td>224</td>
</tr>
<tr>
<td>2</td>
<td>141</td>
</tr>
</tbody>
</table>

### VI. Conclusions

Although there are many factors that might affect the students applied suspension or withdrawal. However, the present study found the class absence is the most significant factor the affect the students applied suspension or withdrawal. Furthermore, the interaction of class absence and academic performance also significant factor the affect the students applied suspension or withdrawal.
References


Appendix Table I. Calculation of decomposed after deleted CA

| Factor | \( R(x_k|x_i,...) \) Likelihood ratio | \( \text{Int}(R(x_k|x_i,...)) \) Likelihood ratio | \( R(x_k|x_i,...) \) Likelihood ratio |
|--------|-----------------------------------|-----------------------------------|-----------------------------------|
|        | df | Sig. | df | Sig. | df | Sig. |
| MS, \( x_i \) | 140.382 | 160 | 0.129 | 173.114 | 158 | 0.194 | 7.268 | 2 | 0.026 |
| EW, \( x_i \) | 202.244 | 160 | 0.013 | 201.402 | 158 | 0.011 | 0.842 | 2 | 0.656 |
| LC, \( x_i \) | 183.970 | 160 | 0.094 | 181.027 | 158 | 0.101 | 2.943 | 2 | 0.230 |
| GE, \( x_i \) | 186.256 | 160 | 0.076 | 162.776 | 158 | 0.381 | 23.480 | 2 | 0.000 |
| ED, \( x_i \) | 319.172 | 256 | 0.004 | 282.743 | 248 | 0.064 | 36.429 | 2 | 0.000 |
| AP, \( x_i \) | 766.894 | 160 | 0.000 | 187.588 | 158 | 0.054 | 579.306 | 2 | 0.000 |

Appendix Table II. Calculation of decomposed after deleted CA and AP

| Factor | \( R(x_k|x_i,...) \) Likelihood ratio | \( \text{Int}(R(x_k|x_i,...)) \) Likelihood ratio | \( R(x_k|x_i,...) \) Likelihood ratio |
|--------|-----------------------------------|-----------------------------------|-----------------------------------|
|        | df | Sig. | df | Sig. | df | Sig. |
| MS, \( x_i \) | 110.569 | 80 | 0.015 | 99.684 | 78 | 0.050 | 10.885 | 2 | 0.004 |
| EW, \( x_i \) | 107.847 | 80 | 0.021 | 107.747 | 78 | 0.014 | 0.100 | 2 | 0.951 |
| LC, \( x_i \) | 90.008 | 80 | 0.152 | 90.644 | 78 | 0.155 | 2.364 | 2 | 0.307 |
| GE, \( x_i \) | 92.901 | 80 | 0.153 | 84.042 | 78 | 0.300 | 8.859 | 2 | 0.012 |
| ED, \( x_i \) | 189.369 | 128 | 0.000 | 155.139 | 120 | 0.017 | 34.23 | 8 | 0.000 |

Appendix Table III. Calculation of decomposed after deleted CA, AP and ED

| Factor | \( R(x_k|x_i,...) \) Likelihood ratio | \( \text{Int}(R(x_k|x_i,...)) \) Likelihood ratio | \( R(x_k|x_i,...) \) Likelihood ratio |
|--------|-----------------------------------|-----------------------------------|-----------------------------------|
|        | df | Sig. | df | Sig. | df | Sig. |
| MS, \( x_i \) | 17.703 | 16 | 0.342 | 9.289 | 14 | 0.812 | 8.414 | 2 | 0.014 |
| EW, \( x_i \) | 8.079 | 16 | 0.946 | 7.981 | 14 | 0.890 | 0.098 | 2 | 0.952 |
| LC, \( x_i \) | 13.098 | 15 | 0.666 | 11.160 | 14 | 0.673 | 1.938 | 2 | 0.379 |
| GE, \( x_i \) | 7.234 | 15 | 0.968 | 7.175 | 14 | 0.928 | 0.059 | 2 | 0.971 |

Appendix Table IV. Calculation of decomposed after deleted CA, AP, ED and MS

| Factor | \( R(x_k|x_i,...) \) Likelihood ratio | \( \text{Int}(R(x_k|x_i,...)) \) Likelihood ratio | \( R(x_k|x_i,...) \) Likelihood ratio |
|--------|-----------------------------------|-----------------------------------|-----------------------------------|
|        | df | Sig. | df | Sig. | df | Sig. |
| EW, \( x_i \) | 4.729 | 8 | 0.786 | 4.684 | 6 | 0.585 | 0.045 | 2 | 0.977 |
| LC, \( x_i \) | 7.445 | 8 | 0.490 | 5.587 | 6 | 0.471 | 1.858 | 2 | 0.395 |
| GE, \( x_i \) | 6.006 | 8 | 0.647 | 5.454 | 6 | 0.487 | 0.552 | 2 | 0.759 |
Appendix Table V. Calculation of decomposed after deleted CA, AP, ED, MS and LC

| Factor | $I(R|x|y|...)$ | $\text{Int}(R,x,y|...)$ | $I(R|y|...)$ |
|--------|----------------|------------------------|-------------|
|        | Likelihood ratio | df | Sig. | Likelihood ratio | df | Sig. | Likelihood ratio | df | Sig. |
| EW, $x_2$ | 1.490 | 4 | 0.828 | 1.444 | 2 | 0.486 | 0.046 | 2 | 0.977 |
| GE, $x_4$ | 1.959 | 4 | 0.743 | 1.444 | 2 | 0.486 | 0.515 | 2 | 0.773 |

Appendix Table VI. Calculation of decomposed after deleted CA, AP, ED, MS, LC and GE

| Factor | $I(R|x|y|...)$ | $\text{Int}(R,x,y|...)$ | $I(R|y|...)$ |
|--------|----------------|------------------------|-------------|
|        | Likelihood ratio | df | Sig. | Likelihood ratio | df | Sig. | Likelihood ratio | df | Sig. |
| EW, $x_2$ | 0.056 | 2 | 0.973 | - | - | - | - | - | - |