IOSR Journal of Research & Method in Education (IOSR-JRME)
e-ISSN: 2320–7388,p-ISSN: 2320–737X Volume 6, Issue 3 Ver. I (May. - Jun. 2016), PP 01-07
www.iosrjournals.org

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

Chun-Min Hung[1], Chun-Yen Chung[2], Yan-Kuin Su[3], Chin-Chiuan Lin[4*]
[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.

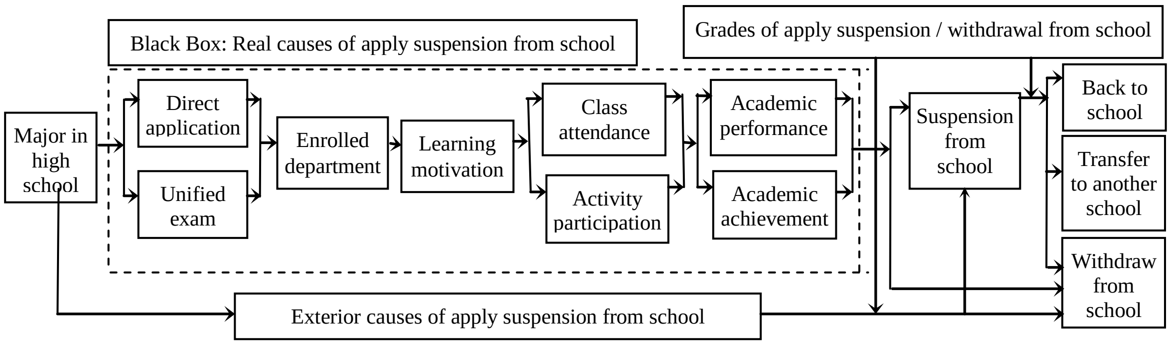

Fig. 1. Decision model for applied suspension or withdrawal from school

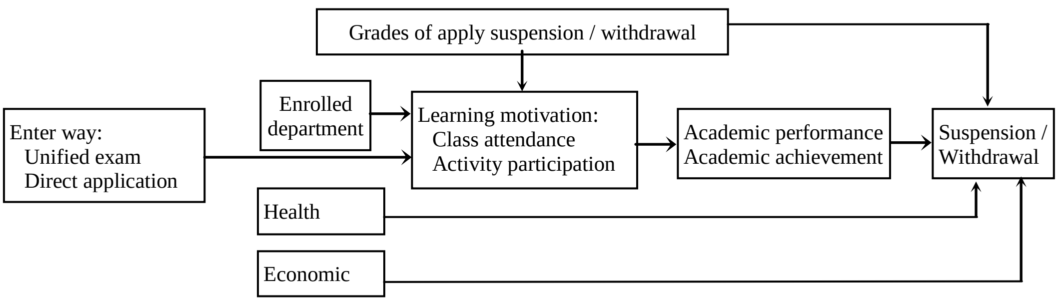

Fig. 2. Significant path coefficients for applied suspension or withdrawal

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 $f(X=i,Y=j,Z=k)$, $i=1,\ldots I,\ j=1,\ldots J,$ *and* $k=1,\ldots,K.$ The Shannon entropy [9] defines the basic information identity (1): $H(X) + H(Y) + H(Z)= I(X,Y,Z) + I(X,Y|Z)$……………………………………..……….……...(1)

where $I(X,Y,Z)= \Sigma_{(i,j,k)} f(i,j,k) \times log\{f(i,j,k) / (f(i)g(j)h(k)\}$, $\{f, g, h\}$ 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:

$I(X,Y,Z)= I(X,Z) + I(Y,Z) + I(X,Y|Z)$…………………………………………………….….....… (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, $I(X,Y|Z)= \Sigma_{(i,j,k)} f(i,j|k) \times log\{f(i|k) / (f(i|k)g(i|k)\}$.

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

$I(X,Y|Z)= Int(X,Y|Z) + I(X,Y||Z)$…………………………………………………………………..(3)

Here, $Int(X,Y,Z)$ 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, 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.

Table I. List Of Codes And Descriptions Of The Variables

Variable	Code	Description
Major in high school	MS, x_1	1: ordinary high schools; 2: vocational high schools.
Enter way	EW, x_2	1: unified exam; 2: direct application.
Living city	LC, x_3	1: Tainan city; 2: other cities.
Gender	GE, x_4	1: Male; 2: Female.
Enrolled department	ED, x_5	1: college of creative media; 2: college of applied human ecology; 3: college of information technology; 4: college of business and management; 5: college of engineering.
Class attendance	CA, x_6	1: number of class absence < 5; 2: others.
Academic performance	AP, x_7	1: average scale > 60; 2: others.
Response	R	1: suspension; 2: withdrawal; 3: others.

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

		Response		
Code	level	(1)	(2)	(3)
MS, x_1	1	131	154	3637
	2	230	280	5166
EW, x_2	1	189	229	4662
	2	172	205	4141
LC, x_3	1	148	166	3304
	2	213	268	5499
GE, x_4	1	212	257	5079
	2	149	177	3724
ED, x_5	1	88	70	1654
	2	37	54	749
	3	52	64	1504
	4	92	109	2076
	5	92	135	2820
CA, x_6	1	16	69	7293
	2	345	365	1510
AP, x_7	1	217	234	8023
	2	144	200	780

A full model of (3) for the case of using seven predictors $\{x_{1,}\ x_2, \ldots, x_7\}$ for the response variable R is:

$I(\{x_1,x_2,x_3,x_4,x_5,x_6,x_7\},R)$

$= 7\ x_i$

$+21\ x_ix_j$

$+35\ x_ix_jx_k$

$+35\ x_ix_jx_kx_l$

$+\ 21\ x_ix_jx_kx_lx_m$

$+\ 7\ x_ix_jx_kx_lx_mx_n$

$+\ 1\ x_ix_jx_kx_lx_mx_nx_o$……..…………………………………………………….……………………..……..(4)

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, …, and x_7. In this case there would be 5,040 (7!) 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 high-way 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 $\{x_1,x_2, \ldots,x_7\}$ for the response variable R is:

$$\begin{aligned} &I(\{x_1,x_2,x_3,x_4,x_5,x_6,x_7\},R) \\ &= I(R,x_1|x_7,x_6,x_5,x_4,x_3,x_2) \\ &+ I(R,x_2|\, x_7,x_6,x_5,x_4,x_3) \\ &+ I(R,x_3|x_7,x_6,x_5,x_4) \\ &+ I(R,x_4|x_7,x_6,x_5) \\ &+ I(R,x_5|x_7,x_6) \\ &+ I(R,x_6|x_7) \\ &+ I(R,x_7) \qquad (5) \end{aligned}$$

Identity (5) is constructed by the rule of selecting the first least significant (7^{th} order) conditional MI (CMI) term, then selecting the least significant 6^{th} order CMI term, and continuing until the last 2^{nd} order CMI term $I(R,x_6|x_7)$. Table III shows the calculation of decomposed of identity (5) for each variables.

Table III. Calculation of decomposed of identity (5)

Var.	$I(R,x_i\|x_j,\ldots)$			$Int(R,x_i,x_j,\ldots)$			$I(R,x_i\|\|x_j,\ldots)$		
	Likelihood ratio	*df*	*Sig.*	Likelihood ratio	*df*	*Sig.*	Likelihood ratio	*df*	*Sig.*
MS, x_1	232.719	320	1.000	230.414	318	1.000	2.305	2	0.316
EW, x_2	254.648	320	0.997	254.050	318	0.997	0.598	2	0.742
LC, x_3	267.606	320	0.985	264.227	318	0.987	3.379	2	0.185
GE, x_4	203.915	320	1.000	203.605	318	1.000	0.310	2	0.856
ED, x_5	442.330	512	0.988	420.270	504	0.997	22.060	8	0.000
CA, x_6	1631.292	320	0.000	423.085	318	0.000	1208.207	2	0.000
AP, x_7	500.579	320	0.000	463.513	318	0.000	37.066	2	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:

$$\begin{aligned} &I(R,\{x_1,x_2,x_3,x_4,x_5,x_6,x_7\}) \\ &= Int(R,x_1,\{x_2,x_3,x_4,x_5,x_6,x_7\}) \\ &+ I(R,x_1||\, x_2,x_3,x_4,x_5,x_6,x_7\}) \\ &+ Int(R,x_2,\{x_3,x_4,x_5,x_6,x_7\}) \\ &+ I(R,x_2||\{x_3,x_4,x_5,x_6,x_7\}) \\ &+ Int(R,x_3,\{x_4,x_5,x_6,x_7\}) \\ &+ I(R,x_3||\{x_4,x_5,x_6,x_7\}) \\ &+ Int(R,x_4,\{x_5,x_6,x_7\}) \\ &+ I(R,x_4||\{x_5,x_6,x_7\}) \\ &+ Int(R,x_5,\{x_6,x_7\}) \\ &+ I(R,x_5||\{x_6,x_7\}) \\ &+ Int(R,x_6,\{x_7\}) \\ &+ I(R,x_6||\{x_7\}) \\ &+ Int(R,x_7) \qquad (6) \end{aligned}$$

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.

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

MI, CMI Terms	$I(R, X_{(t)}\|R^{(t)}\backslash X_{(t)})$			$Int(R, X_{(t)}\|R^{(t)}\backslash X_{(t)})$			$I(R, X_{(t)}\|\|R^{(t)}\backslash X_{(t)})$		
	CMI	*df*	*Sig.*	Interaction	*df*	*Sig.*	Partial Asso.	*df*	*Sig.*
$I(R,x_6\|x_1,x_2,x_3,x_4,x_5,x_7)$	1631.292	320	0.000	423.085	318	0.000	1208.207	2	0.000
$I(R,x_7\|x_1,x_2,x_3,x_4,x_5)$	766.894	160	0.000	187.588	158	0.054	579.306	2	0.000
$I(R,x_5\|x_1,x_2,x_3,x_4)$	189.369	128	0.000	155.139	120	0.017	34.230	8	0.000
$I(R,x_1\|x_2,x_3,x_4)$	17.703	16	0.342	9.289	14	0.812	8.414	2	0.014
$I(R,x_3\|x_2,x_4)$	7.445	8	0.490	5.587	6	0.471	1.858	2	0.395
$I(R,x_4\|x_2)$	1.959	4	0.743	1.444	2	0.486	0.515	2	0.773
$I(R,x_2)$	0.056	2	0.973	-	-	-	-	-	-

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_6x_7, x_5, $(x_6x_7)x_5$, x_2, $(x_6x_7x_5)x_1$, x_3, $(x_6x_7x_5x_1)x_3$, x_4, $(x_6x_7x_5x_1 x_3)x_4$, x_2} as identity (7).

$$
\begin{aligned}
&\text{Logit}(R|\{x_6,x_7,x_5,x_1,x_3,x_4,x_2\}) \\
&= -1.476^* \\
&- 4.464\, x_6^* \\
&-0.181\, x_7 \\
&- 2.232\, x_6x_7^* \\
&+ 0.172\, x_5 \\
&- 0.103\, x_6x_7x_5 \\
&- 0.044\, x_1 \\
&- 1.000\, x_6x_7x_5x_1 \\
&- 0.094\, x_3 \\
&+ 1.064\, x_6x_7x_5x_1x_3 \\
&+ 0.328\, x_4 \\
&+ 0.976\, x_6x_7x_5x_1x_3x_4 \\
&- 0.005\, x_2 \quad \cdots\cdots (7)
\end{aligned}
$$

Identity (7) shows only the main effect of x_6 and interaction of x_6x_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 be suspended or to withdraw when they have more than 5 class absences.

The interaction effect of x_6x_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

Factor		Response		
CA, x_6	AP, x_7	Suspension	Withdrawal	Others
1	1	13	22	7121
	2	3	47	152
2	1	224	212	902
	2	141	153	608

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

[1]. C.-M. Hung, C.-Y. Chung, Y.-K. Su snd C.-C. Lin, Decision model of suspension or withdrawal of college sthuents in Taiwan: Constructing a conceptual model, British Journal of Education, 4(3), 2016, pp. 89-98.

[2]. C.-M. Hung, C.-Y. Chung, Y.-K. Su snd C.-C. Lin, Decision model of suspension or withdrawal of college sthuents in Taiwan: Verification of model, British Journal of Education, 4(4), 2016, pp. 65-74.

[3]. G.-S. Zheng, Construct the generalize associate classification multiple support system to predict the college withdraw, Unpublished Master Thesis, Department of Information Engineering, National Kaohsiung University, Taiwan, 2013. (in Chinese).

[4]. S. Kullback, Information and statisticc, John & Wiley, New Tork, 1959.

[5]. P.E. Cheng, J.W. Liou, M, Liou and J.A.D. Aston, Linear information models: An introduction, Journal of Data Science, 5, 2007, pp. 297-313.

[6]. J.A. Calderhead, New and old proofs of the Pythagorean theorem, The American Mathematical Monthly, *3*(4), 1989, pp.110-113.

[7]. P.E. Cheng, M, Liou, J.A.D. Aston and A.C. Tsai, (2008) Information identities and testing hypotheses: Power analysis for contingency tables. Statistica Sinica, 18, 2008, pp. 535-558.

[8]. P.E. Cheng, M, Liou and J.A.D. Aston, Likelihood ratio tests with three-way tables, Journal of the American Statistical Association, 105, 2010, pp. 740-749.

[9]. C.E. Shannon, A note on the concept of entropy, Bell Systwm Technology Journal, 27, 1948, pp. 379-423.

[10]. S. Kullback and R.A. Leibler, On information and sufficiency, The Annals of Mathematical Statistics 22, 1951, pp. 79-86.

[11]. T.M. Cover and J.A. Thomas, Elements of Information Theory, Wiley, New York, 1991.

[12]. P.E. Cheng, J.W. Liou, M. Liou and J.A.D. Aston, Data information in contingency tables: A fallacy of hierarchical loglinear models, Journal of Data Science, 4, 2006, pp. 387-398.

[13]. W.E. Deming, and F.F. Stephan, On a least squares adjustment of a sampled frequency table when the expected marginal totals are known, Ann. Math. Statist, 11, 1940, pp. 427-444.

[14]. Y.M.M. Bishop, S.E. Fienberg and P.W. Holland, Discrete Multivariate Analysis. Cambridge, MA: MIT Press, 1975.

Appendix Table I. Calculation of decomposed after deleted CA

Factor	$I(R,x_i\|xj,\ldots)$			$Int(R,x_i,x_j,\ldots)$			$I(R,x_i\|\|x,\ldots)$		
	Likelihood ratio	*df*	*Sig.*	Likelihood ratio	*df*	*Sig.*	Likelihood ratio	*df*	*Sig.*
MS, x_1	180.382	160	0.129	173.114	158	0.194	7.268	2	0.026
EW, x_2	202.244	160	0.013	201.402	158	0.011	0.842	2	0.656
LC, x_3	183.970	160	0.094	181.027	158	0.101	2.943	2	0.230
GE, x_4	186.256	160	0.076	162.776	158	0.381	23.480	2	0.000
ED, x_5	319.172	256	0.004	282.743	248	0.064	36.429	8	0.000
AP, x_7	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	$I(R,x_i\|xj,\ldots)$			$Int(R,x_i,x_j,\ldots)$			$I(R,x_i\|\|x,\ldots)$		
	Likelihood ratio	*df*	*Sig.*	Likelihood ratio	*df*	*Sig.*	Likelihood ratio	*df*	*Sig.*
MS, x_1	110.569	80	0.013	99.684	78	0.050	10.885	2	0.004
EW, x_2	107.847	80	0.021	107.747	78	0.014	0.100	2	0.951
LC, x_3	93.008	80	0.152	90.644	78	0.155	2.364	2	0.307
GE, x_4	92.901	80	0.153	84.042	78	0.300	8.859	2	0.012
ED, x_5	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	$I(R,x_i\|xj,\ldots)$			$Int(R,x_i,x_j,\ldots)$			$I(R,x_i\|\|x,\ldots)$		
	Likelihood ratio	*df*	*Sig.*	Likelihood ratio	*df*	*Sig.*	Likelihood ratio	*df*	*Sig.*
MS, x_1	17.703	16	0.342	9.289	14	0.812	8.414	2	0.014
EW, x_2	8.079	16	0.946	7.981	14	0.890	0.098	2	0.952
LC, x_3	13.098	15	0.666	11.160	14	0.673	1.938	2	0.379
GE, x_4	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	$I(R,x_i\|xj,\ldots)$			$Int(R,x_i,x_j,\ldots)$			$I(R,x_i\|\|x,\ldots)$		
	Likelihood ratio	*df*	*Sig.*	Likelihood ratio	*df*	*Sig.*	Likelihood ratio	*df*	*Sig.*
EW, x_2	4.729	8	0.786	4.684	6	0.585	0.045	2	0.977
LC, x_3	7.445	8	0.490	5.587	6	0.471	1.858	2	0.395
GE, x_4	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	$I(R,x_i\|xj,\dots)$			$Int(R,x_i,x_j,\dots)$			$I(R,x_i\|\|x,\dots)$		
	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	$I(R,x_i\|xj,\dots)$			$Int(R,x_i,x_j,\dots)$			$I(R,x_i\|\|x,\dots)$		
	Likelihood ratio	*df*	*Sig.*	Likelihood ratio	*df*	*Sig.*	Likelihood ratio	*df*	*Sig.*
EW, x_2	0.056	2	0.973	-	-	-	-	-	-