

利用資料採礦技術分析臺灣商業健康保險和全民健康保險

Using Data Mining Techniques to Investigate Taiwan Commercial Health Insurance and National Health Insurance

Abstract

本文主要研究目的是將比較不同資料探勘技術，以不同的資料型態與三種不同的分析方式，分別是橫剖面迴歸（Cross-section Regression）、羅吉斯迴歸（Logistic Regression）、時序迴歸（Time Series Regression）來進行分析，來找出影響臺灣商業健康保險購買因素之特徵。本文實證結果中可發現：GDP（國內生產毛額，Gross Domestic Product）成長率和平均收入對於人身意外災害保險之保費支出具顯著正向影響，而平均消費支出對於人身意外災害保險之保費支出具負向影響，顯示所得愈高，保費支出就愈高。另外，全民健保保費及人身意外災害醫療保費支出高的家戶特徵中經濟戶長的性別為一重要變數。保費支出較高的家庭，其經濟戶長多為男性。本研究也以戶內平均年齡作為解釋變數加以分析，結果發現當戶內平均年齡愈低，人身意外災害醫療保險之保費支出就愈高。

Cross-section Regression

變數	係數	t	p
常數項	-1068.85	-0.15	
經濟戶長年齡	-154.25	-2.15	**
戶內人口數	17277.54	17.2	***
就業人口數	-12814.88	-9.71	***
經濟戶長教育程度			
國(初)中(初職)	-9679.28	-2.46	**
高中	-10952.46	-2.67	***
高職	-9669.87	-2.45	**
專科(五專前三學期(初高職))	-12188.01	-2.98	***
大學	-13418.66	-3.24	***
碩士	-15686.47	-3.29	***
博士	-15892.50	-1.94	*
成年人口數	4004.48	3.34	***
所得收入者人口數	5337.34	4.24	***
六十五歲以上人口數	9920.27	8.53	***
戶內平均年齡	875.24	4.23	***

R² = 0.1210
註：
*表示達顯著水準，p<0.1，**表示達顯著水準，p<0.05，***表示達顯著水準，p<0.01。

由左圖發現變數對家戶全民健保保費支出具有顯著影響的變數之因素特徵分別為經濟戶長年齡、戶內人口數、就業人口數、經濟戶長教育程度、成年人口數、所得收入者人口數、六十五歲以上人口數與戶內平均年齡。其中戶內人口數、成年人口數、所得收入者人口數、六十五歲以上人口數及戶內平均年齡具有顯著的正面影響。

變數	係數	t	p
常數項	-5546.06	-1.19	
經濟戶長年齡	204.97	6.68	***
經濟戶長性別			
女	-1307.34	-2.15	**
戶內人口數	1941.91	4.78	***
就業人口數	3035.38	5.7	***
經濟戶長教育程度			
高職	8354.17	2.14	**
專科(五專前三學期(初高職))	9979.56	2.54	**
大學	11120.10	2.82	***
碩士	14885.06	3.69	***
博士	17819.33	3.65	***
所得收入者人口數	3185.19	6.33	***
六十五歲以上人口數	-2910.39	-5.66	***
戶內平均年齡	-73.10	-1.88	*

R² = 0.1405
註：
*表示達顯著水準，p<0.1，**表示達顯著水準，p<0.05，***表示達顯著水準，p<0.01。

由左圖發現變數對家戶投保人身意外災害醫療保險費支出具有顯著影響的變數之因素特徵分別為經濟戶長年齡、經濟戶長性別、戶內人口數、就業人口數、經濟戶長教育程度、所得收入者人口數、六十五歲以上人口數、戶內平均年齡。其中經濟戶長年齡、經濟戶長性別為男性、戶內人口數、就業人口數、經濟戶長教育程度為高職以上到博士、所得收入者人口數具有顯著的正面影響。

Logistic Regression

本研究中人身意外災害醫療保險總樣本數為16528筆，由右圖可得知其中有11802個家庭有購買人身意外災害醫療保險，佔整體比例的71.4%，且平均值有落在信賴區間內。

Effect	N	y ₁ mean	y ₁ Std. Dev.	y ₁ Std. Err.	Cnf.Lmt -95.00	Cnf.Lmt +95.00
Total	11802	28623.33	28011.77	257.8474	28117.90	29128.75

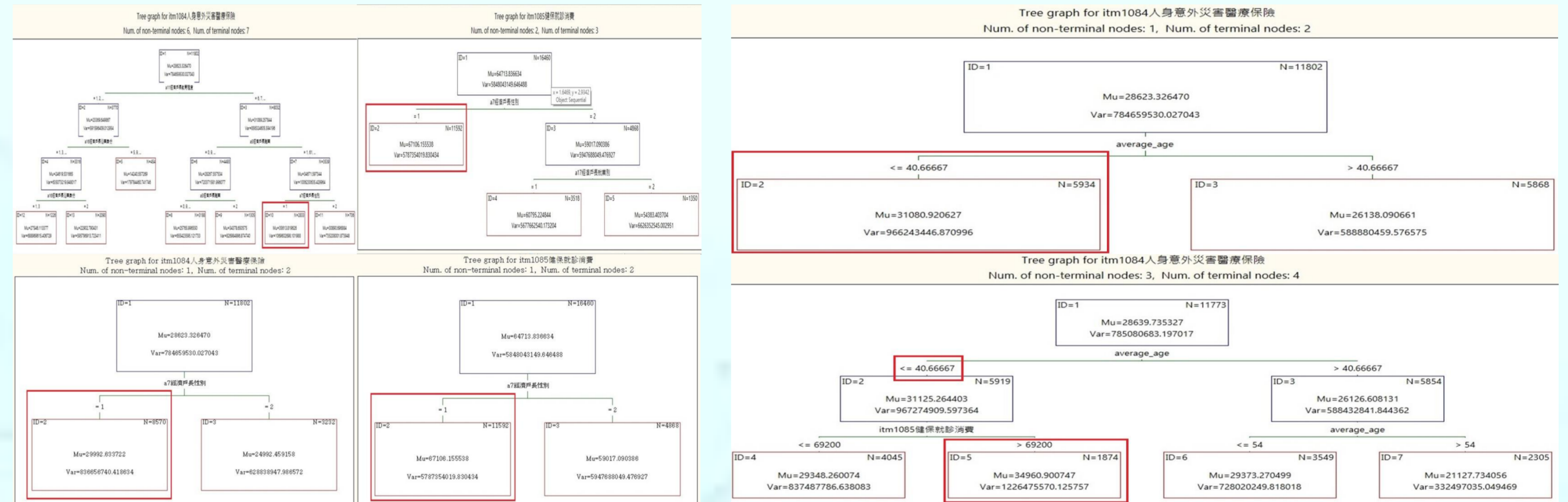
註：
(a)變數定義：
y₁=是否購買人身意外災害醫療保險。

本研究中全民健康保險總樣本數為16528筆，由左圖可得知其中有16460個家庭有購買全民健康保險，佔整體比例的99.58%，且平均值有落在信賴區間內。

Effect	N	y ₂ mean	y ₂ Std. Dev.	y ₂ Std. Err.	Cnf.Lmt -95.00	Cnf.Lmt +95.00
Total	16460	64713.84	76472.50	596.0605	63545.49	65882.18

註：
(a)變數定義：
y₂=是否購買全民健康保險。

Decision Tree



當經濟戶長為男性時，購買保險之需求均比女性為主的家庭高。

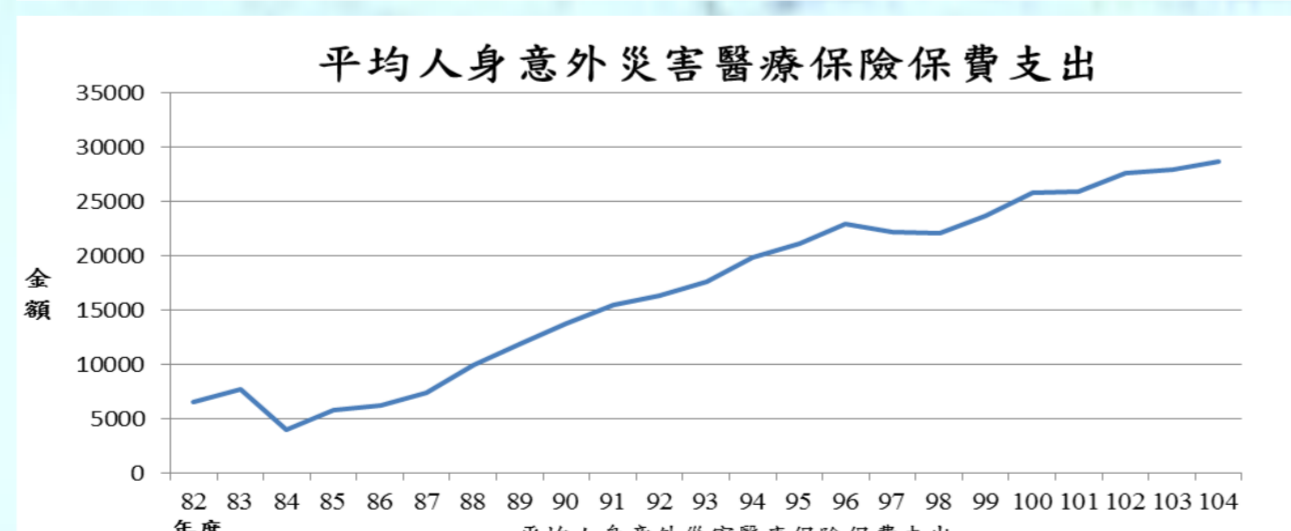
戶內平均年齡小於41歲之家庭，購買人身意外災害醫療保險之需求較高。

Time Series Regression

變數	Mean	Std.Dev	Median	Max.	Min.	Obs.
PAI_M	16965.78	8171.283	17569.85	28623.33	4025.715	23
GR	4.89	3.536544	5.15	10.52	-1.91	23
TI_M	1069967	74105.15	1081698	1163268	836451.6	23
CE_M	675281.2	62266.62	686850	756769.6	500442.3	23

註：
(a)變數解釋：
PAI_M (Mean of Personal Accident Insurance)=人身意外災害醫療保險保費支出總和/人口數=各人身意外災害醫療保險保費支出之算術平均數。
GR(GDP Rate)=各年GDP成長率。
TI_M (Mean of Total Income)=各年總所得收入金額之算術平均數。
CE_M (Mean of Consumption Expenditure)=各年消費支出金額之算術平均數。

本研究透過蒐集民國82到104年之GDP成長率、國民所得收入及消費支出依照年份統整為時序資料，並建立自迴歸模型，藉由實證結果了解這類經濟指標與人身意外災害醫療保險保費支出的影響。



變數	AR(2)	AR(3)	AR(4)	AR(5)
Constant	-20.02374 *** (-3.98)	-20.60682 *** (-4.13)	-19.81782 *** (-3.75)	-12.89221 ** (-2.43)
GR _{t-1}	0.0221057 ** (2.52)	0.0235009 ** (2.18)	0.219057 ** (2.04)	0.0075201 (0.92)
TI_M _{t-2}	3.022275 *** (3.30)	3.029914 *** (3.08)	3.015608 *** (2.63)	2.858112 *** (6.04)
CE_M _{t-2}	-1.635313 ** (-2.43)	-1.600196 ** (-2.07)	-1.643635 * (-1.85)	-1.992281 *** (-6.09)
AIC	-19.57541	-17.97178	-16.41614	-22.68986
BIC	-12.28375	-9.615597	-7.015434	-12.24464

註：
(a) AR(2)、AR(3)、AR(4)、AR(5)代表在ARIMA(p, d, q)中d-q值不變下AR(p)模型的落後期數。
(b) 依P-value判定，***代表1%的顯著水準，**代表5%的顯著水準、*代表10%的顯著水準。
(c) 括號內為檢定統計量之值。
(d) AIC與BIC為選擇模型準則，AIC值愈小，其模型愈好。

由實證結果可看出，GDP成長率與年平均收入對平均人身意外災害醫療保險之保費支出具正向影響，年平均消費支出則具負向影響，故可解釋為GDP成長率與平均收入的提高會促使國民人身意外災害醫療保險保費支出減少，顯示出消費支出會排擠投保支出，因此消費支出愈高，投保保費則愈低。

Conclusions

- 根據本文實證結果，首先利用橫剖面迴歸去找出影響家戶於全民健保和人身意外災害醫療保險保費支出之相關影響特徵，結果發現影響家戶健保就診保費支出較高之主要特徵為：戶內人口數多、成年人口數多、所得收入者人口數多、六十五歲以上人口數多及戶內平均年齡高；而影響家戶人身意外災害醫療保費支出較高之主要特徵為：經濟戶長年齡高、性別為男性、戶內人口數多、就業人口數多、經濟戶長教育程度在高職以上及所得收入者人口數多。
- 其次，在羅吉斯迴歸分析方面上，可看到說臺灣大部分的家庭均有購買全民健康保險，比例將近100%，相較於購買人身意外災害醫療保險之家庭數高；從決策樹模式方面上來看，可明顯看到當經濟戶長為男性、戶內人口數多、就業人口數多、成年人口數多、所得收入者人口數多、戶內平均年齡低時，保費支出較高的特徵就愈明顯。
- 針對民國84年至104年的GDP成長率與家庭收支調查進行時序迴歸分析發現：國民的收入與平日消費支出對商業健康保險保費支出有顯著的影響，當國內經濟成長率提升、所得收入愈高、消費支出愈低，投保商業健康保險的保費支出愈高。
- 觀察時序迴歸的實證結果中之落後期數與模型準則，可發現前兩年的收入與消費支出對投保商業健康保險之保費支出具最顯著的影響。