

98

信用卡信用度分析

本報告利用 4000 名客戶的信用卡相關資料，包含有無違約以及性別、行業別、工作年資等，建立羅吉斯迴歸模型，並透過另外 1000 名客戶的資料評估模型效能。

羅吉斯迴歸是一種處理依變數屬於類別變數的統計分析方法，我們的研究對象是客戶有無違約，數據上只以 0、1 來表示的二元反應變數。相較於原本的線性迴歸模型，羅吉斯迴歸模型透過下列的變數轉換使得依變數介於 0 到 1 之間，通常會將自變數代入模型後的結果視為事件發生的機率值。

$$\text{簡單線性迴歸模型: } g(x) = v_0 + \sum_{i=1}^r v_i x_i$$

$$\text{羅吉斯迴歸模型: } Y = 1 / (1 + e^{-g(x)})$$

$v_1 \sim v_r$ 在兩個模型內皆可視作各解釋變數的權重。

自變數，也就是「有無違約」以外的項目共14項，但分析開始前要注意，以行業別為例，資料將農工商等各行各業以數字表示，但這些數字本身是沒有任何意義的（例：6 是 3 的 2 倍，但編號 3 和編號 6 的行業之間沒有「倍數」的關聯）至於比例尺度的資料，自變數和依變數之間不一定存在線性關聯。因此必須先將變數轉換成可適應模型的 Dummy Variable。

以下是轉換後的變數與依變數的交叉資料表：

是否申請結餘代償類別型(0.沒有申請、1.有申請) * 違約戶(0.不是、1.是) 交叉表

			違約戶(0.不是、1.是)		總和
			0	1	
是否申請結餘代償類別型(0.沒有申請、1.有申請)	沒有申請	個數	1656	5	1661
			99.7%	0.3%	100.0%
	有申請	個數	2224	115	2339
			95.1%	4.9%	100.0%
總和		個數	3880	120	4000
			97.0%	3.0%	100.0%

<最近一個月>近期是否逾期類別型(0.沒有逾期、1.逾期) * 違約戶(0.不是、1.是) 交叉表

			違約戶(0.不是、1.是)		總和
			0	1	
<最近一個月>近期是否逾期類別型(0.沒有逾期、1.逾期)	沒有逾期	個數	3838	115	3953
			97.1%	2.9%	100.0%
	有逾期	個數	42	5	47
			89.4%	10.6%	100.0%
總和		個數	3880	120	4000
			97.0%	3.0%	100.0%

性別(0.女性 1.男性) * 違約戶(0.不是、1.是) 交叉表

			違約戶(0.不是、1.是)		總和
			0	1	
性別(0.女性 1.男性)	女性	個數	2036 98.0%	41 2.0%	2077 100.0%
	男性	個數	1844 95.9%	79 4.1%	1923 100.0%
總和		個數	3880 97.0%	120 3.0%	4000 100.0%

是否使用扣款帳號類別型(0.沒有使用、1.有使用) * 違約戶(0.不是、1.是) 交叉表

			違約戶(0.不是、1.是)		總和
			0	1	
是否使用扣款帳號類別型(0.沒有使用、1.有使用)	沒有使用	個數	3534 96.7%	119 3.3%	3653 100.0%
	有使用	個數	346 99.7%	1 0.3%	347 100.0%
總和		個數	3880 97.0%	120 3.0%	4000 100.0%

以上四者本身就是二元變數，可以直接套入模型。

工作年資(類別型(0.無工作經驗) * 違約戶(0.不是、1.是) 交叉表

			違約戶(0.不是、1.是)		總和
			0	1	
工作年資(類別型(0.無工作經驗)	0	個數	1598 99.3%	11 0.7%	1609 100.0%
	1	個數	2282 95.4%	109 4.6%	2391 100.0%
總和		個數	3880 97.0%	120 3.0%	4000 100.0%

0：無工作經驗與六年以上。沒有工作經驗的新人對社會事務較不熟悉，就比較沒膽子違約，工作六年以上的人則可能會為了生活安穩、不節外生枝而盡量守約。

1：工作一到五年。有一定程度的社會知識與消費能力，這段期間可能會由於轉換跑道、被資遣等因素導致收入無法趕上原本的消費力，造成違約。

有效卡張數類別型(0.沒有) * 違約戶(0.不是、1.是) 交叉表

			違約戶(0.不是、1.是)		總和
			0	1	
有效卡張數類別型(0.沒有)	0	個數	3268	96	3364
			97.1%	2.9%	100.0%
	1	個數	612	24	636
			96.2%	3.8%	100.0%
總和		個數	3880	120	4000
			97.0%	3.0%	100.0%

0：二張以上。能讓銀行發出第二張卡基本上就代表有一定的信用能力。

1：零或一張。有效卡張數和違約戶比例大致上成負斜率的線性關係。

被查詢總家數(類別型(0.沒有) * 違約戶(0.不是、1.是) 交叉表

			違約戶(0.不是、1.是)		總和
			0	1	
被查詢總家數(類別型(0.沒有)	0	個數	3210	76	3286
			97.7%	2.3%	100.0%
	1	個數	602	25	627
			96.0%	4.0%	100.0%
	2	個數	68	19	87
			78.2%	21.8%	100.0%
總和		個數	3880	120	4000
			97.0%	3.0%	100.0%

0：零到二家。

1：三到五家。

2：六家以上。被查詢家數越多，違約戶比例愈高。且六家以上相較五家以下有著顯著差異。

行業別類別型 * 違約戶(0.不是、1.是) 交叉表

			違約戶(0.不是、1.是)		總和
			0	1	
行業別類別型	0	個數	2645	53	2698
			98.0%	2.0%	100.0%
	1	個數	1235	67	1302
			94.9%	5.1%	100.0%
總和		個數	3880	120	4000
			97.0%	3.0%	100.0%

0：製造業、營造業、批發零售業、運輸倉儲業、住宿餐飲業、通訊傳播業、不動產業、專業服務業、公共行政及國防、社會工作服務業、娛樂休閒服務業（違約率 < 4%）

1：農林漁牧、礦業、金融保險業、支援服務業、教育服務業、其他服務業（> 4%）

過去六個月延遲月數比率 * 違約戶(0.不是、1.是) 交叉表

			違約戶(0.不是、1.是)		總和
			0	1	
過去六個月延遲月數比率	0	個數	3485	62	3547
			98.3%	1.7%	100.0%
	1	個數	375	28	403
			93.1%	6.9%	100.0%
	2	個數	20	30	50
			40.0%	60.0%	100.0%
總和		個數	3880	120	4000
			97.0%	3.0%	100.0%

0：0 ~ 0.2

1：0.2 ~ 0.8

2：0.8 ~ 1

過去六個月循環月數比率 * 違約戶(0.不是、1.是) 交叉表

			違約戶(0.不是、1.是)		總和
			0	1	
過去六個月循環月數比率	0	個數	1604	5	1609
			99.7%	0.3%	100.0%
	1	個數	2276	115	2391
			95.2%	4.8%	100.0%
總和		個數	3880	120	4000
			97.0%	3.0%	100.0%

0：0 ~ 0.5

1：0.5 ~ 1

負債所得比 * 違約戶(0.不是、1.是) 交叉表

			違約戶(0.不是、1.是)		總和
			0	1	
負債所得比	0	個數	3491	88	3579
			97.5%	2.5%	100.0%
	1	個數	389	32	421
			92.4%	7.6%	100.0%
總和		個數	3880	120	4000
			97.0%	3.0%	100.0%

0：0~0.3、0.9以上

1：0.3~0.8

一般情況下負債愈小愈不容易違約。對「負債所得比」和「有效卡張數」作交叉資料表後發現負債比0.9以上的客戶全部都擁有3張以上的有效卡，他們有可能是企業經營者或是其親屬，也就是負債比雖然高，卻有穩定管道取得資金以支付履約的費用。

近一年M2繳款期間曾經有逾期兩個月次數 * 違約戶(0.不是、1.是) 交叉表

			違約戶(0.不是、1.是)		總和
			0	1	
近一年M2繳款期間曾經有逾期兩個月次數	0	個數	3839	90	3929
			97.7%	2.3%	100.0%
	1	個數	41	30	71
			57.7%	42.3%	100.0%
總和		個數	3880	120	4000
			97.0%	3.0%	100.0%

0：零次。

1：一次以上。

教育別類別型 * 違約戶(0.不是、1.是) 交叉表

			違約戶(0.不是、1.是)		總和
			0	1	
教育別類別型	0	個數	430	2	432
			99.5%	0.5%	100.0%
	1	個數	3450	118	3568
			96.7%	3.3%	100.0%
總和		個數	3880	120	4000

		97.0%	3.0%	100.0%
--	--	-------	------	--------

0：研究所、大學。

1：高中、初中、小學、未上學。學歷越高、違約戶比例越低。

緊急聯絡人與正卡關係 * 違約戶(0.不是、1.是) 交叉表

			違約戶(0.不是、1.是)		總和
			0	1	
緊急聯絡人與正卡關係	0	個數	118	0	118
			100.0%	0.0%	100.0%
	1	個數	2875	63	2938
			97.9%	2.1%	100.0%
	2	個數	887	57	944
			94.0%	6.0%	100.0%
總和		個數	3880	120	4000
			97.0%	3.0%	100.0%

0：配偶、父母兄弟姐妹

1：子女、配偶父母

2：朋友、其他

持卡人與緊急聯絡人之關係愈疏離、則違約戶比例愈高。

將上述14項變數透過軟體進行羅吉斯迴歸分析，得到下列結果：

模式摘要

步驟	-2 對數概似	Cox & Snell R 平方	Nagelkerke R 平方
1	342.032 ^a	.168	.711

= Hosmer 和 Lemeshow 檢定 =

步驟	卡方	df	顯著性
1	20.615	8	.008

變數在方程式中

	B 之估計值	S.E.	Wals	df	顯著性	Exp(B)
是否申請結餘代償(1)	-1.970	.539	13.342	1	.000	.139
最近一個月近期是否逾期(1)	-.012	.687	.000	1	.986	.988
性別(1)	-.746	.239	9.723	1	.002	.474
是否使用扣款帳號(1)	1.837	1.064	2.977	1	.084	6.276
工作年資(1)	-1.182	.353	11.210	1	.001	.307
有效卡張數(1)	-2.890	.372	60.449	1	.000	.056
被查詢總家數			52.623	2	.000	
被查詢總家數(1)	-3.016	.420	51.664	1	.000	.049
被查詢總家數(2)	-2.314	.461	25.145	1	.000	.099
行業別(1)	-1.297	.239	29.378	1	.000	.273
過去六個月延遲月數比率	1.761	.195	81.261	1	.000	5.819
過去六個月循環月數比率	1.813	.549	10.925	1	.001	6.129
負債所得比	.832	.279	8.878	1	.003	2.297
近一年M2繳款期間曾經有逾期兩個月次數	2.188	.376	33.811	1	.000	8.921
教育別(1)	-2.794	.784	12.696	1	.000	.061
緊急聯絡人與正卡關係			2.422	2	.298	
緊急聯絡人與正卡關係(1)	-14.829	3010.334	.000	1	.996	.000
緊急聯絡人與正卡關係(2)	-.380	.244	2.422	1	.120	.684
常數	-.596	1.436	.172	1	.678	.551

分類表^a

觀察次數		預測次數		
		違約戶(0.不是、1.是)		百分比修正
		不是	是	
步驟 1	違約戶(0.不是、1.是) 不是	3870	10	99.7
	是	37	83	69.2
	概要百分比			98.8

上列分類表（交叉資料表）是用來評估模型效能的重要參考依據。模型的評估中，事件發生的情況被稱為正例。

精確率（precision）：在被模型判定為正例的結果中，所含真正的正例比率。

$$83 / (83+10) = 83 / 93$$

召回率（recall）：在所有正例中，被模型判定為正例的比率。

$$83 / (83+37) = 83 / 120$$

最後，利用與 4000 筆模型資料獨立的 1000 筆測試資料進行評估。結果如下：

違約戶(0.不是、1.是) * 預測的群組 交叉表

		預測的群組		總和
		不是	是	
違約戶(0.不是、1.是)	不是	968	2	970
	是	16	14	30
總和		984	16	1000

精確率 (precision) : $14 / (14+2) = 7 / 8$

召回率 (recall) : $14 / (14+16) = 7 / 15$

精確率很高，代表這組模型能夠很精準地找出潛在違約戶。

召回率很低，代表和模型發現的潛在違約戶相比，仍有約一倍的潛在違約戶隱藏在其他客戶之中。