

國立臺灣大學國際企業研究所

碩士論文

Graduate Institute of International Business

College of Management

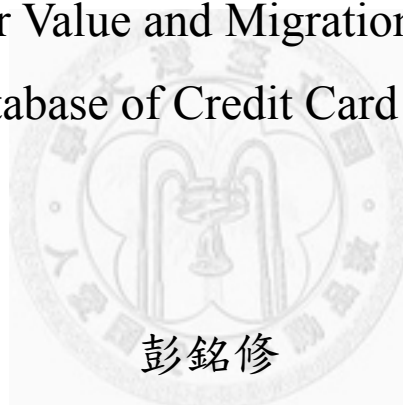
National Taiwan University

Master Thesis

顧客價值與遷徙路徑分析—以信用卡資料庫為例

Customer Value and Migration Analysis

– Take Database of Credit Card as example



彭銘修

Ming-Hsiu Peng

指導教授：任立中 博士

Advisor: Lichung Jen, Ph.D.

中華民國 101 年 6 月

June, 2012

誌謝

寫謝詞的此刻，再次體認到自己是一個極其幸運的人，一直都是。

這篇論文的完成，首先要感謝任立中教授，老師對我們的指導不僅在論文上，同時也教導我們許多待人處事的道理，言猶在耳、銘記在心，謝謝老師。

謝謝韋杰學長、前堯學長、功新學長，沒有你們的幫助就不會有這篇論文。謝謝修維、契漢、韋菁、城美、紋綺，我的同門師兄弟姐妹們。謝謝可愛的研究所同學們，這段時間互相打氣、一起擠在研究室裡面寫著論文、一起愁眉苦臉、一起歡笑，是你們為兩年的時間增添許多色彩，謝謝你們。謝謝偶爾出現的大學同學跟高中同學和國中同學，雖然你們對我的論文沒有貢獻，但看到你們就能帶走許多壓力，也因此更有動力，謝謝你們。謝謝黎立。

最重要的是，謝謝我的父母，如果沒有你們的支持，我將什麼都不是，我愛你們。

彭銘修 謹誌

民國一〇一年八月

中文摘要

現今的行銷模式必須考慮到消費者的「動態性」和「異質性」，本研究以國內某銀行的信用卡顧客為研究對象，嘗試以資料庫行銷的理論與統計方法達到此目的，並分析其顧客價值。首先，結合 RFM 模型和層級貝氏模型、最大概似估計法等估計方式，以購買區間、活躍性、穩定性等指標分析顧客價值。其次，利用馬可夫鏈理論建立移轉機率矩陣模型，以預測顧客的未來消費之狀態，達到平均 54% 的擊中率。並設定 6 條路徑，以逐步迴歸法分析價值遷徙路徑與顧客特質之間的關聯性。本研究的結果能幫助企業了解其顧客價值，並提供模型估計顧客未來之狀態，以利銀行規畫未來行銷策略，並建立長期顧客關係。

關鍵字：資料庫行銷、馬可夫鏈理論、層級貝氏模型



ABSTRACT

The main purpose of this considering consumer dynamic and consumer heterogeneity by theory of database marketing and statistical method, is to recognize the customer value which is based on a domestic bank's database of credit card customers. In the first part of this research, we analyze customer value by indexes such as inter-purchase time, reliability index, activity index conducted by RFM model, Hierarchical Bayesian model and maximum likelihood estimates. Second, we build individual customer's Transition Probability Matrix of Markov Chain to predict customer purchase state, and the average hit rate reaches 54%. Further, we set 6 meaningful paths and find out if there is any correlation within the transition paths and demographic variables. The contribution of this study could help enterprises verify customer value, identify customer's transition paths, develop follow-up marketing strategies and build up long-term customer relationship.

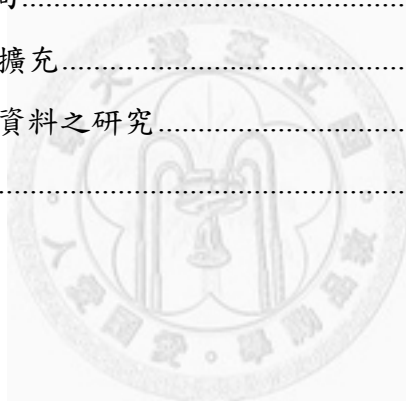
Key words: Database Marketing, Markov Chain, Hierarchical Bayesian Model

目錄

口試委員會審定書	#
誌謝	i
中文摘要	ii
ABSTRACT	iii
目錄	iv
圖次	vii
表次	viii
第一章 緒論.....	1
第一節 研究背景與研究動機.....	1
第二節 研究問題與範圍.....	2
一、 顧客價值初步分析	2
二、 建構層級貝氏 Probit 模式之移轉機率矩陣	2
三、 顧客價值遷徙路徑分析	2
第三節 研究流程.....	3
第二章 文獻探討	4
第一節 顧客關係管理.....	4
一、 顧客關係管理的定義	6
二、 顧客關係管理的功能	7
三、 顧客關係管理的執行方法	8
第二節 資料庫行銷.....	11
一、 資料庫行銷的定義	12
二、 資料庫行銷的功能	13
三、 資料庫行銷之應用	14
四、 資料庫行銷於成長策略之應用	16
五、 資料庫行銷的執行	17
第三節 RFM 模型	18

一、	RFM 模型的定義	18
二、	RFM 模型的應用	19
第三章	研究方法	20
第一節	購買期間 (Inter-purchase Time) 之估計方法.....	20
一、	最大概似估計法.....	21
二、	加權最大概似估計法.....	21
三、	層級貝氏統計模型.....	22
第二節	馬可夫鏈理論.....	23
一、	馬可夫鏈之前提假設與定義.....	23
二、	移轉機率矩陣.....	24
第三節	層級貝氏 Probit 模型.....	25
一、	篩選資料的方法.....	25
二、	HB Probit 模型設定	27
第四章	實證研究	29
第一節	資料分析.....	30
一、	交易紀錄資料整理.....	30
二、	人口統計變數資料整理.....	32
第二節	購買區間分析.....	38
一、	最大概似估計量.....	39
二、	加權最大概似估計量.....	39
三、	層級貝氏估計量.....	40
第三節	顧客穩定性及活躍性分析.....	42
一、	活躍性分析.....	42
二、	穩定性分析.....	45
第四節	層級貝氏 Probit 模型估計與預測力分析.....	48
一、	資料整理.....	48
二、	模型估計.....	49
三、	預測力分析.....	50
第五節	顧客落點及路徑分析.....	52

一、	最終落點分析	52
二、	顧客價值遷徙路徑分析	54
第五章	結論與建議	59
第一節	研究結果與行銷意涵	59
一、	刷卡次數與頻率	59
二、	購買區間與穩定性、活躍性分析	59
三、	層級貝氏 Probit 模型預測力	60
四、	最終落點與顧客價值遷徙路徑	60
第二節	研究限制	61
一、	交易紀錄資料上的限制	61
二、	其他資料上的限制	61
第三節	未來研究方向	62
一、	資料庫之擴充	62
二、	交易明細資料之研究	62
REFERENCE	63



圖次

圖 1.1	研究流程架構圖	3
圖 2.1	資料庫行銷組織關係圖	17
圖 3.1	HB 條件狀態選擇模型之投入資料篩選	26
圖 4.1	顧客消費金額與頻率分布圖	30
圖 4.2	人口變數結構特性圖	34
圖 4.3	最大概似估計量直方圖	39
圖 4.4	加權最大概似估計量直方圖	40
圖 4.5	層級貝氏估計量直方圖	41
圖 4.6	預測力比較圖	51
圖 4.7	路徑 1 先增額後增次	54
圖 4.8	路徑 2 先增次後增額	55
圖 4.9	路徑 3 同時增次增額	55
圖 4.10	路徑 4 先減次後減額	55
圖 4.11	路徑 5 先減額後減次	55
圖 4.12	路徑 6 同時減次減額	56

表次

表 2.1	顧客關係管理定義整理	6
表 2.2	資料庫行銷的定義	12
表 2.3	資料庫行銷於成長策略之應用	16
表 4.1	消費金額與消費頻率之敘述統計量	31
表 4.2	人口統計變數水準合併表	32
表 4.3	人口結構特性	33
表 4.4	平均消費金額之變異數分析	35
表 4.5	平均消費次數之變異數分析	36
表 4.6	購買區間估計量之比較	41
表 4.7	顧客活躍性指標範例	42
表 4.8	活躍性分析	43
表 4.9	活躍性指標列聯表卡方檢定	44
表 4.10	顧客穩定性指標範例	46
表 4.11	穩定性分析	46
表 4.12	穩定性指標列聯表卡方檢定	47
表 4.13	狀態定義與資料個數	48
表 4.14	移轉機率矩陣 (以顧客 100089 為例).....	49
表 4.15	狀態移轉預測範例	50
表 4.16	最終落點個數統計	52
表 4.17	最終落點區隔特性分析	52
表 4.18	最終落點列聯表卡方檢定	53
表 4.19	各路徑之人數	56
表 4.20	上升路徑之逐步迴歸分析	57
表 4.21	下降路徑之逐步迴歸分析	58

第一章 緒論

第一節 研究背景與研究動機

行銷模式隨著時代演進、經濟的發展，逐漸發展出截然不同的面貌，現今的企業不得不考慮到消費者的「動態性」和「異質性」，過去生產導向和銷售導向的行銷模式已不再適用於今日的消費者，行銷人員面臨的將是更嚴峻的挑戰。在今天，行銷的思維逐漸轉變為顧客導向，在設計產品與服務時，以顧客的角度出發並以解決顧客的問題為目的。但消費者常是多變的、不專情的，要怎麼做好顧客關係管理並留住顧客，成為一項企業不得不去思考的重大議題。

在過去，企業了解的顧客的方式僅限於前線人員的經驗或是行銷人員的問卷發放、訪談等研究方式，但是隨著資訊科技的發展，資料庫行銷提供企業更有效的方法，可以幫助企業了解顧客、並落實顧客關係管理，最終達到一對一行銷的目的。

台灣信用卡市場自 1974 年發行第一張信用卡後，發展迄今已將近 10 年的歷史，現今信用卡已成為人們日常生活的一部分 (賴柏志、孫銘誼，2012)。但國內發卡銀行眾多，競爭愈趨激烈，儘管銀行爭相推出不同的行銷活動，希望能刺激持卡人消費，但各家信用卡發卡銀行常面臨顧客辦了卡卻不使用的狀況，主要的原因還是在於銀行對顧客了解的度不夠，以至於行銷活動的成效不彰，也無法與顧客建立長期的關係。

本研究嘗試以國內某發卡銀行的資料庫為資料來源，結合 RFM 模型和層級貝氏模式等估計方式，捕捉消費者的異質性；並建立移轉機率矩陣模型以估計顧客的動態性，希望能藉由上述兩個過程，了解該銀行之顧客價值，並提供模型估計顧客未來之狀態，以利發卡銀行規畫未來行銷策略，並建立良好顧客關係。

第二節 研究問題與範圍

一、 顧客價值初步分析

整理與顧客價值有關之資訊，包括刷卡金額、刷卡頻率、顧客購買區間 (Inter-purchase Time) 並進行顧客價值分析，計算活躍性指標、穩定性指標，並根據不同的人口統計變數區隔來探討其在顧客價值上的差異，試圖利用統計方法找出不同顧客價值的消費者之特性。

二、 建構層級貝氏 Probit 模式之移轉機率矩陣

利用馬可夫鏈理論建立一層級貝氏 Probit 模式之移轉機率矩陣，藉此預估顧客未來刷卡的金額及頻率，並檢驗模型的預測力。

三、 顧客價值遷徙路徑分析

利用先前建構之層級貝氏 Probit 模式之移轉機率矩陣，分別針對顧客的消費狀態做最終落點分析與路徑分析。透過最終落點分析，可以了解顧客長期而言最終的消費狀態為何；透過路徑分析，找出顧客最有可能遵循的路徑，並根據人口統計變數去探討遵循不同路徑的顧客群之間，是否有顯著的差異性，這些性質將可以幫助發卡銀行在審核發卡對象時就能先預估其未來的顧客價值。

第三節 研究流程

本研究的流程架構圖如下圖表示：

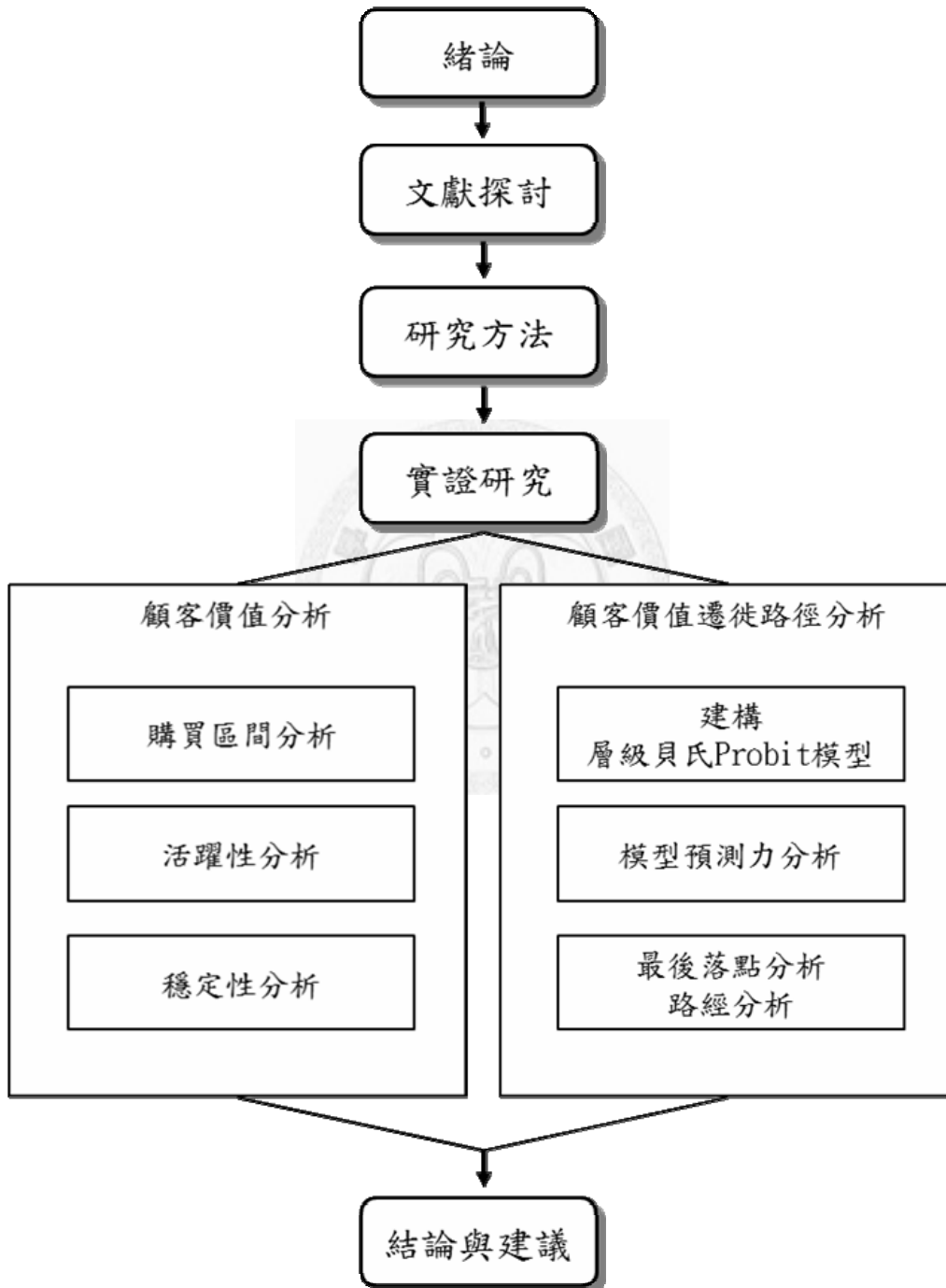


圖 1.1 研究流程架構圖

第二章 文獻探討

第一節 顧客關係管理

學者 Theodore Levitt (2000) 曾對於顧客關係有這樣的比喻：「銷售不過只是求愛時期的結束，然後才是婚姻關係的開始。婚姻生活的幸福與否端看雙方對彼此關係的經營績效如何」。「交易行為」僅僅是整個行銷過程的開始，而不是結束，完成一筆交易並不是企業最終目的，試著去維持、增進與顧客間的長期關係，並進而創造更多的顧客價值，才是企業在行銷上追求的重點。

在過去，行銷概念以生產導向 (Production&Manufacturing Orientation) 為主，企業與顧客間只有短暫的交易關係，買賣雙方除了交易進行以外沒有太多的互動，買賣雙方只是進行了一個所謂的「產品與貨幣交換過程」，雙方僅以短期利益為考量的重點，因此，每次的交易都視為是獨立的。

但隨著時代演進，以生產導向的行銷概念逐漸失去其主流地位，隨後在 1950 至 1980 年代興起的大眾行銷 (Mass Marketing) 也失去了它的魅力，其中主要原因即是『邊際效用遞減』的經濟學概念，過度的生產使得產品的邊際效用越來越低，對於消費者而言即表示產品已失去吸引力。由於科技及生產技術的進步、經濟的發展，市場上的供需結構產生劇烈的變化，發生所謂『過多商品追求有限顧客』的情況。當多數的消費者處於衣食無缺的狀態時，他們將會變得挑剔、變得難以被同質化的商品及服務滿足，企業間為了尋求更多的獲利，紛紛開始致力於差異化的產品及服務的提供。

而後，行銷的思維因而轉變成顧客導向為主，企業主愈趨重視消費者的需求，因此針對消費者的需求而產生的市場區隔概念因應而生，但長期而言，當每個市場區隔都充斥著競爭者的時候，企業成功的因素將取決於顧客關係與顧客的忠誠度，更進而衍生出「關係行銷」的概念。

根據研究顯示，當顧客對於品牌的忠誠度提升，其重複購買比例將是其它消費者的 12 倍，此現象不僅反應在奢侈品如凱迪拉克轎車，也反應於便宜如披薩這類的產品。另一研究資料更顯示，僅僅 5% 的顧客忠誠度成長，將會提高企業的利潤高達 25% 至 85%。麥肯全球關係行銷(McCann Relationship Marketing Worldwide)營運長 Pamela Maphis Larrick 曾指出「留住舊客戶為企業創造利潤的不二法門」。以金融業來為例，減少有價值顧客 5% 的流失量，就可增加 75% 的顧客貢獻度；此外，吸引一位新顧客，比留住一位舊客多花 5 倍成本；100 位滿意的顧客可衍生出 25 位新的顧客。

然而，許多調查顯示現代的消費者愈來愈不專情，對於行銷經理人而言，要令消費者度對品牌忠貞不二，是一件愈趨艱鉅的任務。6%-85% 的顧客表示，他們對於之前使用過的品牌之產品其實還算滿意；有 68% 的顧客表示他們選擇另一個品牌事實上「並沒有什麼特別的原因」。在美國的汽車市場，90% 是顧客滿意度的標準，然而重複購買同一品牌比例往往只剩下一半（45%）。根據以上數據，可以了解顧客關係管理的重要性。以下將依序討論顧客關係管理的定義、功能及應用。

一、 顧客關係管理的定義

顧客關係管理系統此一名詞在實務界被廣為運用，但學術界較少對其作定義。又實務界的文獻雖有定義，但由於各系統供應商所建置的功能項目不盡相同，因此定義並不統一。因此，本研究將一些國內外文獻中較具代表性的顧客關係管理之定義整理於表 2.1。

表 2.1 顧客關係管理定義整理

學者	定義
Pepper & Rogers & Dorf (1999)	「顧客關係管理」具有與一對一行銷相同的意義，皆為一種傾透過聽顧客需求而獲得顧客的一種方式，需要透過與顧客建立起一種學習性關係才能達成。
Davids Meryl (1999)	「顧客關係管理」包含了關係行銷、終身價值行銷、忠誠度行銷和一對一行銷。企業整合這些策略能夠創造出與顧客之間長期並且互利的關係，並發展忠誠度和利潤。
McKinsey Company, Inc. (1999)	顧客關係管理的定義是：透過分析顧客資料，將不同價值的顧客群區隔出來後，了解其需求與特性，並提供不同的產品和通路滿足不同區隔的需求。甚至能夠為其量身訂做產品與服務，持續耕耘顧客關係並與之溝通，強化顧客貢獻價值，是一套結合服務、行銷與專業的支援系統。
Kalakota & Robinson (1999)	運用整合性銷售、行銷、與服務下發展出的一致性行為被稱為顧客關係管理。在企業流程與科技的整合下，找出顧客的真正需求，同時改進產品與服務，致力於提昇顧客忠誠度與滿意度。
American Database Marketing Association (1999)	「顧客關係管理」是一種協助企業與顧客建立關係，使雙方都互利的管理模式。
P. Kotler (2000)	「顧客關係管理」是藉由良好的顧客價值與滿意度，建立並維持長久的互利關係。
R.S. Swift (2001)	「顧客關係管理」是一種企業透過溝通而了解、影響顧客行為，作為增加顧客、防止流失、提高忠誠度和獲利的手段。
IBM Inc. (2012)	客戶關係管理(CRM)指的是一個包含科技應用、流程及組織變革，並以顧客為中心的企業策略。CRM 緊密結合了組織的銷售、行銷與顧客服務活動，以實現獲利、營收及顧客滿意的最佳化。

二、 顧客關係管理的功能

(1) 學者陳文華 (1999) 認為顧客關係管理能為企業帶來許多益處，其主要分為 7 點如下：

- I. 增加收益：針對較有價值的顧客進行交叉銷售，提高銷售之績效。
- II. 增加獲利率：增加和顧客之間的互動，提高開發潛在顧客的機會，並滲透現有客戶，提高其獲利率。
- III. 降低成本：企業可以避免將資源投注在沒有價值的顧客身上。
- IV. 提高市場的專注性：透過與顧客之間的互動與了解，可以提供行銷人員於制定行銷決策時參考，進而提高對市場的專注性。
- V. 減少新銷售活動執行的循環次數：企業可研究如何與顧客進行有效的溝通，增進彼此雙方的了解。可減少行銷活動執行的次數，降低嘗試錯誤的頻率。
- VI. 增加小型目標行銷的次數：因為對顧客的互動與了解，因此能將顧客區隔成更多不同的市場，而針對不同的市場企業可提供不同的產品或運用不同的行銷方法，如此一來，銷售成功的機率也會提高。
- VII. 增加知識：行銷人員可藉由資料庫儲存大量資料的能力，了解所有銷售活動的執行結果及其成敗的原因，更可以藉其分析處理的能力，增加對於消費者行為面的一些知識。

(2) Swift (2001) 提到企業在進行顧客關係管理之後，將為企業帶來競爭優勢如下：

- I. 降低開發新顧客的成本：更有效地透過連絡與追蹤現有顧客，以維持穩定的交易量，因此不需要過度仰賴開發新顧客。
- II. 降低銷售成本：由於現有顧客通常是較為活躍並反應較熱絡的顧客，除了對通路與經銷商有更多的認識，建立關係也使得通路對顧客更有影響力。因此，藉由顧客關係管理能提高行銷的效率與減少銷售的成

本。

- III. 提高顧客利潤：更多的顧客和更高佔有率、更多的追蹤銷售、更高的顧客滿意度等等，皆能帶來更多交叉銷售與延伸銷售的能力。
- IV. 提高顧客忠誠度：持續提供正確的服務，讓顧客買得更多，留得越久，自然能增加顧客的忠誠度與終身價值。
- V. 有效地獲利評估：透過系統了解哪些顧客具有價值貢獻，哪些顧客應該透過行銷策略提升貢獻，哪些顧客根本不具有貢獻能力，哪些顧客又能在未來帶來商機等等。

簡而言之，顧客關係管理系統使得企業能夠辨別不同顧客對企業的重要性，並以最有效慮的方式，將資源分配給適合的顧客，減少行銷成本，並提高個別顧客的獲利率與價值貢獻度。除了增進銷售額以外，企業可以利用顧客關係管理系統快速處理顧客的問題，也更容易了解顧客的潛在需求，有助於企業在未來產品的開發與服務的加強，更幫助行銷人員在制定行銷策略上時有所依據。而良好的顧客關係管理必須透過資料庫的建置與資料庫行銷實現之，下一節將針對資料庫行銷做進一步的討論。

三、顧客關係管理的執行方法

(1) 陳文華(2000)提出顧客關係管理的執行步驟應分為以下 7 個步驟：

- I. 決定顧客關係管理的目標：企業首先要訂出一個施行顧客關係管理所欲達成的目標，且此目標為一量化數字，較容易掌握與衡量，例如增加獲利率、提昇顧客再購率等等明確的目標。
- II. 了解「現行行銷模式改變時可能之障礙」：未來的行銷將以顧客導向為主，講求能在適當的時點、透過適當的通路、針對適當的顧客，提供適當的產品，這樣的行銷方式比傳統行銷、目標行銷方式更能滿足

個別的顧客需求。在實行顧客關係管理系統前，必須先了解可能存在之障礙。

- III. 調整組織及作業程序之規劃：在企業考慮調整外部行銷活動的同時，假如企業內組織的結構和作業不能隨之調整。將會影響外部行銷的效果。
- IV. 利用資訊技術分出顧客群：利用資料挖掘、各種統計分析等方法，不同於以往用傳統人口統計變數的方式，而是用一個全新、多屬性區分標準的分群方式。這樣的分析結果對於下一個階段「銷售活動之規劃」會有幫助。
- V. 規劃銷售活動：在分出數個顧客群之後，由於各群的顧客各有各的特性，我們必須針對各群的特性規劃出適合於各群的銷售活動。
- VI. 執行銷售活動計畫：規劃銷售活動後，根據經過調整的組織和流程，配合新的銷售活動加以執行。
- VII. 監督、控制、反饋：在執行之後，必須監督和控制銷售活動的成效，紀錄結果，並反饋給決策階層，作為下次目標制訂和調整的依據。

(2) Peppers and Rogers(2001)提出以「IDIC」(Identify, Differentiate, Interact, Customize) 流程，創造良好顧客關係的四大階段，整理如下：

- I. 找出 (**Identify**) 顧客：除非知道顧客是誰，否則便無法與他們發展一對一關係。因此，顧客關係管理第一步便是找出個體、區隔顧客群。
- II. 依據價值及需求區隔 (**Differentiate**) 顧客：企業希望差異化地對待不同的顧客，意味著企業能夠指出他們個別的特徵，根據顧客價值的不同，列舉以下三種類型：
 - a. 最有價值的顧客 (**Most valuable customers, MVCs**)：最有價值的顧客是目前業務的核心，企業最應該關心的如何留住此類顧客，因為他們不會希望將這些顧客拱手讓給競爭對手。

- b. 最具成長潛力的顧客 (**Most Growable Customers, MGCs**) :是指那些「當企業發展前瞻性的策略以增加與顧客往的業務量，他們便會具有更重大的價值」的人。
- c. 毫無價值的顧客: (**Below Zero, BZs**):是指「無論企業再怎麼努力，都無法從他們那裡獲取足夠的利潤以彌補他們付出去的成本」的人，企業如無法發展從他們身上賺取足夠利潤，就應放棄，或者鼓勵其投向競爭者。

由上述分類，讓企業方便安排銷售及行銷的優先次序；了解不同顧客的需求，並依據顧客的價值與需求提供最適合的產品及服務，是維持顧客忠誠度最好的方式。

- III. 與顧客互動 (**Interact**) :互動、對話、資訊交流，這些都是產生關係的重點所在，為了與顧客達到最大的互動，所以要有成本效率地互動，顧客與企業的對話、關係的背景資料越豐富、越複雜，而且充滿愉悅，那麼就意謂競爭對手越來越難從企業手中奪取客戶，另一思考是企業都希望驅使更多的與顧客互動，流向最具成本效率的通路，像是網際網路。
- IV. 提供顧客客製化 (**Customize**) 服務：要與顧客發展關係，企業必須能夠修改自己的行為或流程，以滿足個別顧客的需求，依據現在對某特定顧客的了解，以適合、差異的方式對待不同的顧客，是任何一段關係的「決定性的階段」。

第二節 資料庫行銷

行銷模式隨著時代演進、科技的進步，逐漸發展出截然不同的面貌，特別是資訊科技突飛猛進之後，資料庫行銷儼然成為行銷領域之顯學。過去的行銷人員為了了解消費者行為及偏好，常常需要利用問卷發放、訪談等方式取得所需資料，不但費時費力，且容易產生抽樣誤差。不同於傳統行銷模式，在電腦及網路普及的今天，資料庫行銷不但可以記錄顧客的人口變數（例如：性別、年齡、居住地、經濟狀況、教育程度、生活型態等），更能夠詳實地留下顧客的動態資料（包括：購買品項、消費時間、消費通路、付款方式、配合何種促銷活動、滿意度，網路消費行為甚至可以記錄顧客在網頁上瀏覽的頁面及時間）。行銷人員得以利用資料庫整合、產生、並分析所需的資料，以協助行銷策略的規劃。

時至今日，資料庫行銷日漸受到企業的重視，主要的原因是：資料庫行銷可以有效幫助企業實施顧客關係管理。建立資料庫之後，企業與顧客之間不再只是簡單的買賣關係，透過資料庫行銷，企業得以更加了解消費者，打造更深入的顧客關係。對於企業而言，可以利用資料庫行銷研擬更有效率的行銷策略；以消費者角度來看，能夠減少接觸多餘的資訊，並提高購買產品的機率（Denise, Geoffrey, Dawn & Linda, 1997）。

科技技術的進步以及網路環境的成熟，是促使資料庫行銷蓬勃發展的重要條件，現代的行銷人員利用電腦儲存並處理數量龐大的顧客資料與互動記錄，可以快速的了解顧客，與傳統的行銷研究方法相比，資料庫在行銷研究方面可以為企業省下更多成本。

除了科技因素，統計方法的進步配合電腦卓越的演算能力，許多過去由於過於繁複的計算而無法實施的統計方法得以實現，使得行銷人員可以進行更透徹的、更多元的資料分析，也是促使資料庫行銷大為風行的原因之一。

以下將依序介紹資料庫行銷之定義、資料庫行銷的優點，以及資料庫行銷在實務上的應用。

一、 資料庫行銷的定義

表 2.2 資料庫行銷的定義

學者	定義
Shani and Chalasani (1992)	將現在和過去顧客的資料蒐集起來，並建立一個資料庫，用以改善行銷的績效。
Cespedes, Frank and Jeff Smith (1993)	利用顧客資料，以改善企業對行銷資源的分配與應用，以針對特定對象傳遞更有價值的資訊，並維持與顧客之間良好的關係。
Arthur Hughes (1994)	「資料庫行銷」是利用電腦科技，管理一套現有顧客與潛在顧客相關資訊的資料庫系統，企業可以透過這些資料，提供顧客更好的產品或服務，並與顧客建立良好的長期關係。
劉穎壽(1995)	資料庫行銷是指應用顧客資料庫系統從事行銷活動之行銷方式，企業藉由蒐集有關顧客、潛在顧客的各種人口統計變數、興趣、偏好、購買行為及生活型態資料來建立企業本身之顧客資料庫，並尋找相對較有價值之顧客，追蹤並評估促銷活動之效果，所有資料庫行銷的活動目的是希望能建立與顧客之良好關係並建立品牌知名度，使其願意再度購買相關產品及服務。
Frederick Newell(1997)	是一套中央資料庫系統，用來儲存所有企業與顧客的相關資訊，並利用這些資料規劃個人化的溝通，此資料庫更具備加強顧客終身價值的功能。
任立中(1998)	資料庫行銷的重心可分為二：其一為互動；另一為異質性。與顧客的互動方能正確的與消費者溝通；在資訊密集的競爭市場中，行銷策略的成功與否全賴其對顧客異質性的調適。
Kotler(2000)	「資料庫行銷」是一建立並維持顧客關係的過程，並利用顧客資料庫與其他資料庫，來達到接觸顧客並與之交易的目的。

綜合各個學者對於資料庫行銷的定義，可以整理出三要素：資料庫系統、蒐集顧客資訊、資訊的分析與應用。資料庫行銷的特色在於，透過三要素的整合，可以了解每位顧客的異質性，透過一對一行銷，達到接觸顧客並與之交易的目的，長期而言更能幫助企業建立良好的顧客關係。

二、 資料庫行銷的功能

在顧客管理關係裡面，資料庫行銷扮演著分析工具的角色，所以，了解資料庫行銷對於企業行銷決策上具備什麼樣的功能，將是本節討論的重點。根據林慧晶（1997）的研究，把資料庫行銷的功能分為四大部分：

（1） 顧客價值分析

資料庫行銷最主要的功能在於針對顧客的價值進行分析。在過去，企業雖然可以掌握每日的銷售金額，但卻難以評估個別顧客與銷售狀況的連結性；而資料庫行銷可以蒐集個別顧客的消費行為及其他資訊，利用統計方法進行分析，企業因此可以得知每位顧客對於企業而言的價值及重要性。

（2） 計算顧客終身價值

根據 Arthur Hughes（1994）的定義：「所謂的顧客價值（Customer Lifetime Value）是指在未來一定時間之內，企業或廠商可以從個別顧客獲得之利潤的淨現值。」透過資料庫行銷，企業可以根據資料庫對於每位顧客所記錄的消費金額及消費頻率等，計算出該顧客可能貢獻於廠商的顧客終身價值。藉由顧客終身價值的計算，企業不但可以預測未來營收，還可以判斷出顧客價值的高低，而分配不同的資源於不同價值之顧客身上，以提高行銷的效益。

（3） 進行向上銷售（Up-selling）與交叉銷售（Cross-selling）

根據 Bob Stone（1995）所言：「向上銷售是指企業針對顧客目前所購買的產品項目，推測其往後可能需要的產品項目，而進行推銷。交叉銷售則是指針對目前顧客所購買的品項，進行相關產品的推銷。」因此，透過資料庫行銷對於消費者購買之產品項目的記錄，企業很容易達到向上銷售和交叉銷售的目的。

(4) 行銷決策支援系統

根據 Little(1979)所言：「所謂的行銷決策支援系統(MDSS, Marketing Decision Support System)是指將顧客的購買記錄透過模型分析，再利用模型分析出來的結果配合專家知識，使決策者能做出有利的決策。」因此，顧客的購買記錄以及模型的分析是行銷決策支援系統的兩大重要因素，而資料庫行銷結合兩者，不但可以幫助企業管理其顧客，更重要的是資料庫行銷具備了行銷決策支援系統的功能。

三、 資料庫行銷之應用

Denise(1997)等人曾提出發展資料庫行銷的指南，其內容如下(引自林宗龍，民90)：

(1) 企業需求分析 (Corporate Needs Analysis)

這是第一步也可能是最重要的一步，它將會決定資料庫在使用上的方向、功能及適切性。另一個需要被討論的議題則是：誰會來使用這個資料庫系統，以及它應該包含什麼資料，同時必須決定要以什麼方式來發展資料庫，是自行發展抑或是外包，不同的方式將會影響行銷程序的維持、更新及執行。

(2) 收集資料 (Compiling Data)

- I. 內部資料：內部資料的項目包括顧客姓名、地址、聯絡電話、主要人口統計變數、過去交易的歷史資料例如 RFM (最近購買、購買頻率、購買金額) 以及付款歷史記錄等。在 B2B 資料庫中，還會包含 SIC 碼、員工人數、購買行為的偏好，以及購買的決策者之姓名、職稱、及相關資訊。
- II. 外部資料：外部資料包括以彙編完成的資料(如：總體的人口統計資料)、行為資料(如購買型態)以及模型資料(如預測其未來交易的模式)。

(3) 初步分析 (Initial Analysis)

透過不同知識、資源以及創造力所構成的資料庫系統，必須將大量的資料轉變成有用的資訊才有意義。有二種方法可以達成這項任務，一是利益分析 (Profitability Analysis)：計算每位顧客對公司的貢獻。另一項是趨勢分析 (Trends Analysis)：以利益分析做基礎，藉此分辨不同獲利群所呈現的特徵和屬性。這些分析將幫助行銷人員更有效率地對高貢獻顧客設計行銷活動。

(4) 定義市場 (Defining the Market)

透過資料庫所建構的模型有助於分析現有及潛在顧客，並尋找市場機會。利益分析模型可以區隔出最有價值的顧客，而後廠商可以決定是否要減少或停止對於低獲利顧客群的行銷活動。

(5) 發展行銷計畫

發展行銷計畫的最佳資訊來源可能是過去的績效，不論是企業本身成功和失敗的經驗，或其它企業的經驗，都可能提供有價值的建議，如何能辨識出哪一種行銷計畫能產生什麼樣的績效，才是最重要的課題。其中有許多該做 (Do's) 或不該做 (Don'ts) 的準則，這些準則有助於引導成果，而這些都可以從歷史的經驗學習。

(6) 追蹤結果及趨勢

資料庫行銷相較傳統的行銷，最大優勢在於創造了回饋途徑 (Feedback Loop)，另一優勢則在於可藉由歷史交易資料找出顧客，求出適切的銷售型態。最後則是可以有效追蹤所有成功的直效行銷案例，以供未來作為行銷決策上的參考。

四、 資料庫行銷於成長策略之應用

Joseph、Lackman、Peace and Tatar (1999) 在研究中指出，零售業者若能夠善用資料庫，可以更深入了解現有顧客以及潛在顧客，並找出更精確的市場區隔方法以鎖定目標客群，更能有效的判斷其行銷策略的成功機率。他們針對成長策略，解釋選擇不同的成長策略上，資料庫行銷所能發揮的功用為何，如表 2.2 所述：

表 2.3 資料庫行銷於成長策略之應用

		產品構面	
		舊產品	新產品
市場構面	舊市場	<p>市場滲透策略</p> <p>更精確地挑選目標客群，在維持現有廣告支出下，降低銷售成本，並找出最適合之產品組合，以符合多數消費者需求。</p>	<p>產品發展策略</p> <p>分辨出現有顧客對於現有產品所認知的偏好與缺點為何，並作為開發新產品參考的資料。</p>
	新市場	<p>市場發展策略</p> <p>從現有市場蒐集資訊，精簡為幾個關鍵因素，作為事件導向行銷之依據。</p>	<p>多角化策略</p> <p>利用更有效的市場區隔方式，挑選出潛在顧客。並運用在既有產品與市場所得之資訊，作為樣本外預測之基礎。</p>

資料來源：Joseph、Lackman、Peace and Tatar (1999)

五、 資料庫行銷的執行

任立中（2010）曾說明人是資料庫行銷能夠成功的關鍵因素，這群人由一組跨領域的專家所組成，包括電腦專家、統計專家、策略專家與管理銷售專家。這四種專家透過彼此溝通，進行跨領域的合作，以執行資料庫行銷之功能。

管理銷售將其銷售狀況與策略專家進行討論，策略專家將提出策略作法並與統計專家說明，統計專家則根據策略上的需求設法找出資料分析時所需要的模型，再進而跟電腦專家溝通討論，使其進行電腦相關之統計分析系統的建置。其後，統計專家以電腦專家所規劃的系統進行統計分析，並將分析結果說明給策略專家，策略專家依照分析結果進行策略規劃並透過銷售專家執行。這一連串跨領域、跨功能的團隊合作即是資料庫行銷的關鍵成功因素，其組織關係圖如下：

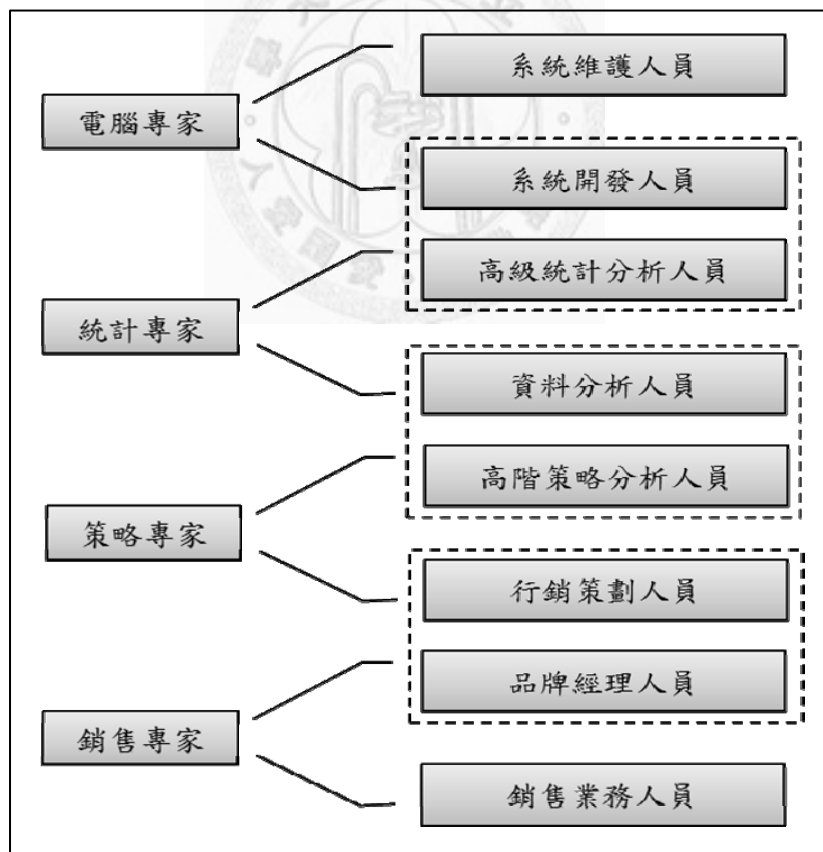


圖 2.1 資料庫行銷組織關係圖

第三節 RFM 模型

一、 RFM 模型的定義

(1) 最近購買日

Hughes (1994) 定義最近購買日為「顧客最近一次購買距離分析時點的天數」。一般而言，若某顧客最近購買日較小，即是顧客最近一次購買的日期距離現在不遠，企業認為該顧客再度消費的機會較高，因此會將此顧客認定為較重要之顧客。然而，以最近購買日來評估顧客之重要程度時，還必續考慮企業產品的特性，耐久品或消耗品皆會有不同的考量，不能單純以最近購買日的大小決定顧客價值高低。

(2) 購買頻率

Hughes (1994) 定義購買頻率為「顧客在一段期間內購買的次數」。一般的情況之下，若顧客購買次數較多，企業對該顧客的忠誠度和顧客價值評價會較高。

(3) 購買金額

根據 Hughes (1994) 對購買金額的定義為「在一段期間之內，購客購買該企業產品之總金額」。一般情況下，若顧客購買金額較高時，對企業而言之顧客價值較高。

二、 RFM 模型的應用

RFM 模型主要應用於顧客價值的分析，兩位學者 Hughes (1994) 和 Stone (1995) 針對 RFM 模型的應用提出不同做法，主要差異在於對不同指標之權重和重要性的看法不一致。

(1) Arthur Hughes 的建構原則

Hughes (1994) 認為 R、F、M 三項指標對於衡量顧客價值的重要性是一致的，在權重上並無分別。於實務上的做法是將資料庫中所記錄的最近購買期間、購買頻率與購買金額，分別從大到小、多到少依序排列，各構面分別將顧客區分為 5 等分，前 20% 的顧客可得 5 分，次之的 20% 得 4 分，依此類推分別給予 3 分、2 分和 1 分，總共可細分為 125 個區隔。

(2) Bob Stone 的建構原則

因 Stone (1995) 採用信用卡消費資料建構 RFM 時，為了符合該產業之特性，所以主觀給予購買頻率 (F) 最高權重、最近購買期間 (R) 次之，購買金額 (M) 最低。

第三章 研究方法

第一節 購買期間 (Inter-purchase Time) 之估計方法

任立中與王仕茹 (1998) 提出以顧客購買期間 (Inter-purchase Time) 結合顧客購買金額，進行顧客價值分析，與過去的 RFM 系統有以下幾點差異：

- (1) 過去 Hughes 和 Stone 所建構的 RFM 模型是一個相對性指標，僅根據消費者的消費金額、消費頻率給予相對性的評分，而顧客購買期間則為一絕對性指標，可以呈現每位顧客個別的行為模式。
- (2) 傳統的 RFM 模型建立於顧客為同質性的前提下，從個別顧客角度出發，計算每位顧客的顧客價值分數，並未考慮顧客異質性可能衍生的情況。而以「層級貝氏模型」所估計的購買期間考量了顧客間的異質性，能夠充分運用隱藏於顧客間的訊息。
- (3) Hughes 和 Stone 所建構的 RFM 模型僅以過去的消費紀錄計算顧客價值，為一落後指標，但利用「層級貝氏模型」所估計的購買期間可以用於預測顧客未來可能的消費行為，配合消費金額和其他模型，更能預測顧客下一期之狀態，可作為顧客價值分析的領先指標。

一、 最大概似估計法

由於購買區間為一連續性事件，時間單位不可能有負值，並考量到平均數與標準差應是不同的行為，因此本研究採取對數常態分配配適之，即假設對購買區間取對數後會服從常態分配。所以我們假設第 i ($i=1, \dots, N$) 位顧客第 j ($j=1, \dots, n_i$) 次的購買區間 t_{ij} 服從期望值為 μ_i ，且變異數為 σ_i^2 的對數常態分配，即 $t_{ij} \sim \ln N(\mu_i, \sigma_i^2)$ ，其中 $t_{ij} \in (0, \infty)$ ， $\mu_i > 0$ ，則 $\ln t_{ij} \sim N(\mu_i, \sigma_i^2)$ ：

$$f(\ln t_{ij} | \mu_i, \sigma_i^2) = \frac{1}{\sqrt{2\pi\sigma_i^2}} e^{-\frac{(\ln t_{ij} - \mu_i)^2}{2\sigma_i^2}} \quad \text{式(1.1)}$$

根據式(3.1)可推得第 i 位顧客的購買期間之最大概似函數：

$$L(\mu_i, \sigma_i^2 | \ln t_{i1}, \dots, \ln t_{in_i}) = \prod_{j=1}^{n_i} f(\ln t_{ij} | \mu_i, \sigma_i^2) \quad \text{式(1.2)}$$

將最大概似函數取對數，並分別對 μ_i 及 σ_i^2 微分，令其一階微分值等於零，整理後可得最大概似估計量：

$$\hat{\mu}_i = \frac{\sum_{j=1}^{n_i} \ln t_{ij}}{n_i} \quad \text{式(1.3)}$$

二、 加權最大概似估計法

加權最大概似估計法的概念是以最大概似估計法為基礎，根據距離目前時點的遠近而給予不同的權重，越接近現在則給予越高的權重，可以用數學式表之，如式(3.4)：

$$\hat{\mu}_i = \frac{\sum_{j=1}^{n_i} j \ln t_{ij}}{\sum_{j=1}^{n_i} j} \quad \text{式(1.4)}$$

三、 層級貝氏統計模型

傳統的統計視機率分配參數 (Parameters) 為固定常數，且過去的統計模式是以抽樣理論 (Sampling Theory) 為基礎，其估計量是在僅考慮樣本資料的情況下所求得。然而，貝氏統計將機率分配的參數視為一隨機變數，並服從某特定機率分配。貝氏統計以貝氏理論 (Bayes' Theory) 為基礎，結合事前情報及樣本資料之後，產生事後機率分配，並依此求得參數的貝氏估計量。貝氏估計量的優點在於，當樣本資料較少的情況下，由於具有充分的先驗知識，可以縮小因樣本資料過少而形成的估計誤差。

層級貝氏模型同樣利用貝氏理論的條件機率為基礎，運用於樣本母體參數的估計上，建構出如階梯般的層級概念，不但包含貝氏統計模式之優點，甚至比一般貝氏統計模式更為精確。一般貝氏統計模式只有一個關於先驗分配的知識，而層級貝氏則具有兩層(或兩層以上)的先驗知識之假設，因此可以提高估計量的準確度。

根據馮淑群 (1998) 對於直銷公司的交易資料進行模型預測和比較，證實利用層級貝氏模型在預測最終消費者的購買行為上，較負二項模型更為準確。王仕茹 (1999) 將層級貝氏模型應用於市場區隔分析，發現預設效度明顯優於傳統市場區隔方式。柳慧琴 (1997) 的研究顯示：就算顧客的交易紀錄較少，貝氏層級模型的預測結果仍是最為準確。

第二節 馬可夫鏈理論

馬可夫鏈的理論基礎為馬可夫過程 (Markov Process) ，可用來預測一個隨機變化的動態系統。現今消費者的消費行為越來越難以掌握，可將其視為一隨機變化的動態系統，因此本研究將採用馬可夫鏈理論針對消費者的消費行為建立移轉機率矩陣，並做出預測與分析。

一、馬可夫鏈之前提假設與定義

馬可夫鏈 (Markov Chain Theory) 由俄國數學家 A. A. Markov (1856-1922) 於 1997 年提出：滿足已知「現在」的條件下，「將來」與「現在」呈現獨立的特性稱為馬可夫性質，而具有這種性質的隨機過程被稱作馬可夫過程。

馬可夫鏈根據其狀態空間 (State Space) 可分為有限型馬可夫鏈 (Finite Markov Chain) 和無限型馬可夫鏈，是指其狀態之個數為有限或無限而言；另外依照時間的特性可分為離散型馬可夫鏈和連續型馬可夫鏈。本文研究的範疇在於離散時間的隨機過程，因此以下主要討論離散型馬可夫鏈。

(1) 馬可夫鏈具有以下基本假設：

- I. 狀態間的轉換機率符合馬可夫性質，即只與目前的狀態有關，與過去的狀態無關。
- II. 狀態空間必須為可數集合 (Countable Set) 。
- III. 轉換機率固定，即 $p_{i|j}$ 不隨時間改變。

其中 $p_{i|j}$ 為狀態轉換機率 (State Transition Probability) ，表示由狀態 i 轉移至狀態 j 的機率， $p_{i|j}$ 不得為負值，且 $\sum_i p_{i|j} = 1$ 。

(2) 馬可夫鏈之定義

當隨機過程 $\{X_t, t \geq 0\}$ 中，第 $t+1$ 期的狀態與第 t 期的狀態有關，且與之前的狀

態無關，則此數列稱為馬可夫鏈，可以利用式子表示：

$$P(X_{t+1} = j | X_0 = i_0, X_1 = i_1, \dots, X_t = i) = P(X_{t+1} = j | X_t = i) = p_{ij} \quad \text{式(1.5)}$$

馬可夫鏈至少需有兩個狀態，當馬可夫鏈具有兩個狀態時，稱為二狀態馬可夫鏈，具有三個狀態時稱為三狀態之馬可夫鏈，以此類推。

二、 移轉機率矩陣

現在定義狀態空間 (State Space) 共存在 S 種狀態，可建立一描述狀態之間移轉可能性的移轉機率矩陣 P 如下：

$$P \equiv (t) \begin{matrix} & \begin{matrix} & \begin{matrix} (t+1) \\ & & & & \end{matrix} \\ & \begin{matrix} p_{1|1} & p_{2|1} & \cdots & p_{s|1} \\ p_{1|2} & p_{2|2} & \cdots & p_{s|2} \\ \vdots & \vdots & \ddots & \vdots \\ p_{1|s} & p_{2|s} & \cdots & p_{s|s} \end{matrix} \end{matrix} \end{matrix} \quad \text{式(1.6)}$$

矩陣 P 稱為一階移轉機率矩陣 (One-Step Transition Probability Matrix)，矩陣內的元素 p_{ij} 定義為在已知現在 (第 t 期) 狀態是 j 的情況下，下期 (第 $t+1$ 期) 狀態為 i 的條件機率，其中 $\sum_{i=1}^S p_{ij} = 1$ 。又矩陣 P 自乘 m 次之後可擴充為 m 階移轉機率矩陣

(m -Step Transition Probability Matrix)：

$$P^m = \underbrace{P \cdot P \cdot \dots \cdot P}_{(\text{自乘}m\text{次})} \equiv (t) \begin{matrix} & \begin{matrix} & \begin{matrix} (t+m) \\ & & & & \end{matrix} \\ & \begin{matrix} p^{(m)}_{1|1} & p^{(m)}_{2|1} & \cdots & p^{(m)}_{s|1} \\ p^{(m)}_{1|2} & p^{(m)}_{2|2} & \cdots & p^{(m)}_{s|2} \\ \vdots & \vdots & \ddots & \vdots \\ p^{(m)}_{1|s} & p^{(m)}_{2|s} & \cdots & p^{(m)}_{s|s} \end{matrix} \end{matrix} \end{matrix} \quad \text{式(1.7)}$$

矩陣 P^m 內的元素 $p^{(m)}_{ij}$ 表示在已知本期 (第 t 期) 狀態是 j 的情況下， m 期之後 (第 $t+m$ 期) 狀態為 i 的條件機率。使用此方法估計顧客的移轉機率矩陣，即可根據目前的起始狀態，預測他們未來各期的消費行為狀態。

第三節 層級貝氏 Probit 模型

任立中和陳靜怡 (2007) 曾利用層級貝氏 Probit 模型和層級貝氏 Logit 模型估計消費者手機通話狀態的移轉機率矩陣。利用這兩種模型估計移轉機率矩陣不但可估計個人層次的行為模式，同時加入人口統計變數，因此也考慮了群體層次的行為差異；另外，即使個別顧客缺少在某狀態的紀錄，也可透過總體資訊估計之。本研究採用其層級貝氏 Probit 模型的方式來建構信用卡顧客消費狀態的移轉機率矩陣。

一、 篩選資料的方法

Probit 模型是一種探討分類性變數行為的統計方法，可以應用於估計移轉機率矩陣。本文的做法參考任立中和陳靜怡 (2007) 的研究，主要概念為：根據樣本給定本期狀態為 j 之條件下，以層級貝氏 Probit 模型直接估計下期狀態 k 的條件發生機率， $k=1,2,\dots,17$ 。

但此模型不適合用來處理組數太多的次數資料，例如本文之連續狀態組數就高達 17^2 組，所以必須設法減少組數。如圖三-一所示，建立第 j 個購買模型之前 ($j=1,2,\dots,17$)，顧客購買紀錄的投入必須加以篩選，圖 3.1(a) 是顧客 i 總共 L 期的購買紀錄，我們利用在「本期狀態為 j 」的條件下所選擇的資料建立第 j 個模型，如圖 3.1(b) 所示。例如，若欲建立顧客 i 的第 3 個 HB 條件狀態選擇模型，則投入資料必須是本期狀態為 3 ($j=3$) 的下期狀態選擇 (k)，如第 2 期與第 5 期這兩筆資料可用於配適第 3 個模型 ($j=3$)。

然而，若某顧客在狀態 j 沒有購買紀錄，將無法建立對應的個人化條件狀態選擇模型，在估計第 j 個模型時，顧客 i 僅有 L_{ij} 筆下期狀態選擇資料可用於配適該模型，且 $L_{ij} < L$ ；而僅有 N_j 位顧客具有符合條件的資料可以推論個人化移轉機率矩陣的第 j 列移轉機率向量，有 $(N - N_j)$ 位顧客則因缺乏資料，而須靠總體資訊估計之。

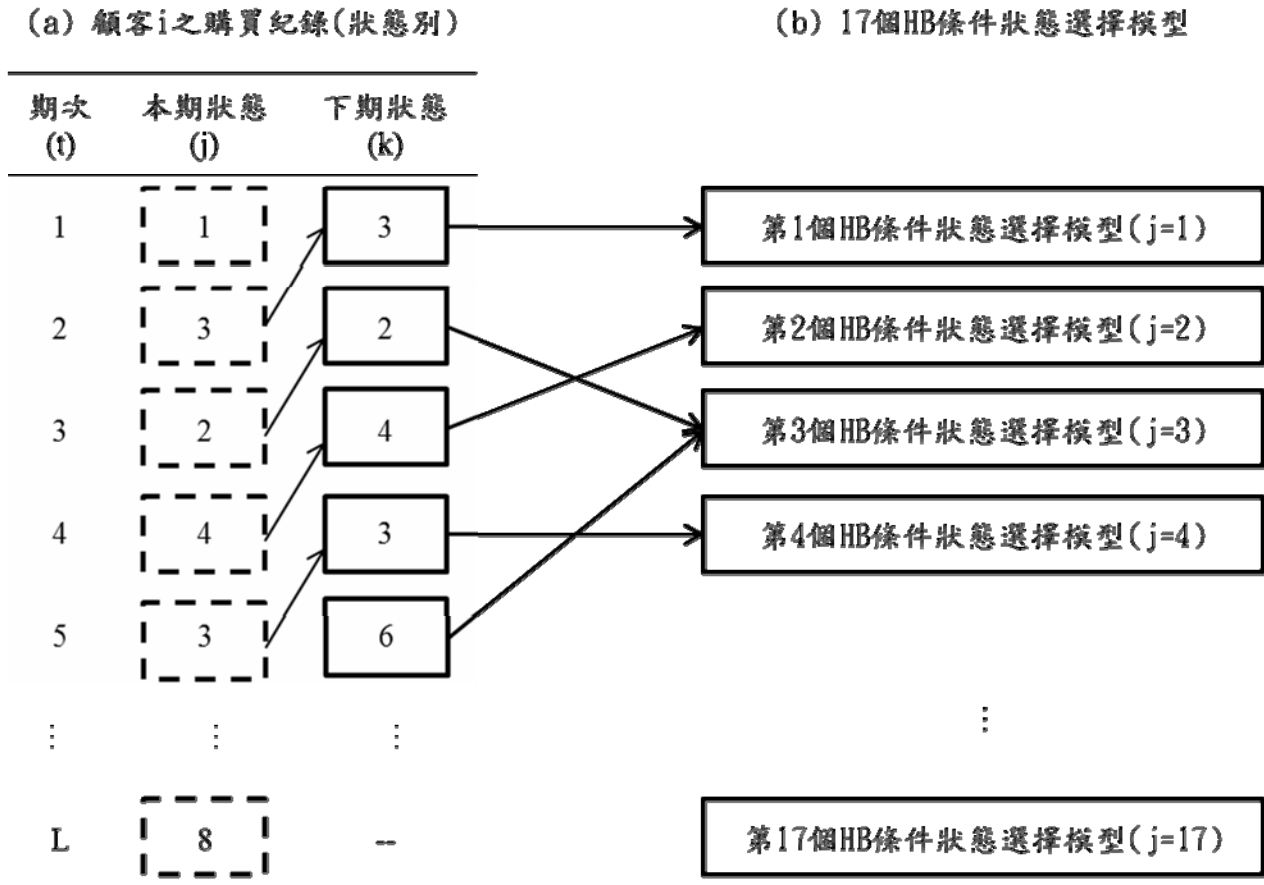


圖 3.1 HB 條件狀態選擇模型之投入資料篩選

二、 HB Probit 模型設定

HB Probit 模型主要的假設為：每一狀態具有一相對效用，且狀態選擇是根據狀態相對效用比較後的結果，並假設所有狀態的相對效用為無法觀測的連續型潛伏資料 (Continuous Latent Data) 向量，且服從多變量常態分配 (Multivariate Normal Distribution；MN)。

(1) 個別顧客層次之模型設定

以第 j 個 ($j=1,2,\dots,S$) HB Probit 模型為例，並且假設經過篩選後，顧客 i 有 $L_{i|j}$ 筆購買紀錄可投入模型配適。以下為個別顧客層次的模型設定：

$$y_{im|j}^* = \mu_{im|j} + \varepsilon_{im|j} \quad m=1,2,\dots,S-1$$

$$\begin{cases} y_{im|j}^* \geq \max(0, y_{im|j}^*), & \text{if } y_{it|j} = k \\ y_{im|j}^* < 0, & \text{if } y_{it|j} = S \end{cases} \quad k=1,2,\dots,S-1, t=1,2,\dots,L_{i|j} \quad \text{式(1.8)}$$

其中，

$y_{it|j}$ 表示顧客 i 在第 t 期的條件狀態選擇，有 S 種選擇；

$y_{im|j}^*$ 表示顧客 i 在第 t 期的狀態選擇為 m 的潛伏相對效用(狀態 S 為比較基礎)，且相對效用的大小排序與觀察值 $y_{it|j}$ 相互對應；

$\mu_{im|j}$ 表示在顧客 i 的認知下，選擇狀態 m 可得之平均相對效用；

$\varepsilon_{im|j} = [\varepsilon_{it1|j} \ \varepsilon_{it2|j} \ \dots \ \varepsilon_{it(S-1)|j}]'$ 為一 $(S-1) \times 1$ 的矩陣，表示相對效用之誤差項，且服從一期望值為 0 向量，變異數矩陣為 \sum_j 的多變量常態分配，並限制 $\text{Var}(\varepsilon_{it(S-1)|j}) = \sigma_{S-1|j}^2 = 1$ 。

式(3.7)說明，相對效用的大小排序與狀態選擇的觀察值彼此之間具有對應關係，當顧客 i 選擇基礎狀態 S 時，其他 $(S-1)$ 個狀態的相對效用皆會小於 0；當顧客 i 選擇非基礎狀態，即 $y_{it|j} = k$ 時，其相對效用會大於 0 且大於其他狀態的相對效用值。

(2) 群體層次之模型設定

由於層級貝式 Probit 模式設定的顧客特質向量 Z_{ij} (表示顧客 i 的特質向量，因人而異) 將人口統計變數納入考量，所以可進一步探討相對效用受個人特質所影響的程度。群體層次之模型設定如下：

$$\mu_{ij} = \Theta'_j z_{ij} + \delta_{ij} \quad i=1,2,\dots,N_j \quad \text{式(1.9)}$$

其中，

μ_{ij} 表示平均相對效用向量，為一 $(S-1) \times 1$ 的向量；

z_{ij} 表示顧客 i 的特質向量，為一 $(P \times 1)$ 向量，包括人口變數如性別、教育程度等共 P 項變數，並根據資料篩選後得到可投入第 j 個 Probit 模型的顧客有 N_j 位；

Θ_j 為回歸係數矩陣，目的在於量平均相對效用如何因顧客特質而異，為一 $P \times (S-1)$ 矩陣；

δ_{ij} 為誤差項向量，為 $(S-1) \times 1$ 向量，假設服從多變量常態分配 $MN(0, \Lambda_j)$ 。

根據群體層次模式所估計之參數，有助於我們預測欠缺購買紀錄的 $(N - N_j)$ 位顧客的第 j 列移轉機率矩陣，概念如下：首先求出具有購買紀錄之顧客的平均相對效用，並以其群體層次所估計的參數，預測缺乏購買紀錄且與前者具有相同特質之顧客的「平均」條件狀態選擇機率，再進而求得第 j 列移轉機率矩陣。

第四章 實證研究

本研究以國內某銀行之信用卡部門為研究對象，自其信用卡的資料庫隨機抽取 10,600 名顧客，共有 349,824 筆交易紀錄，抽樣時間為 2010 年 2 月到 2012 年 2 月共 25 個月的交易資料，及顧客之人口統計變數資料。

分析的主體以個體為主，而非以卡片為單位。另外為了做預測擊中率的比較，本研究將前 20 個月的消費資料投入模型作為估計用的樣本，而後 5 個月的消費紀錄則作為保留樣本 (Handout Sample)，用以比較模型效度。

交易紀錄包含：交易日、交易金額(以新臺幣為單位)、交易明細，其中交易明細由於資料未經分類，內容紊亂而不具分析意義，願不納入研究範圍；人口統計變數包含：卡號、年收入、婚姻狀況、教育程度、居住地(以縣市為單位)、卡片級別。



第一節 資料分析

一、 交易紀錄資料整理

本研究由資料庫隨機抽樣 10,600 位顧客，為期 25 個月，總共 349,824 筆的交易紀錄資料。為了使樣本起始點相同，以利於移轉機率矩陣的估計，所以我們以月為單位，一個月為一期，排除第一期(2010 年 2 月)無交易紀錄的顧客；另外，為了利用顧客購買區間分析顧客價值，所以每位顧客至少需要兩筆以上的購買區間資料，所以排除交易紀錄低於 3 次的資料。篩選過後之人數為 2,312 位顧客，為時 25 期，共 119,228 筆交易紀錄資料，可透過圖 4.1 來了解其分布情況：

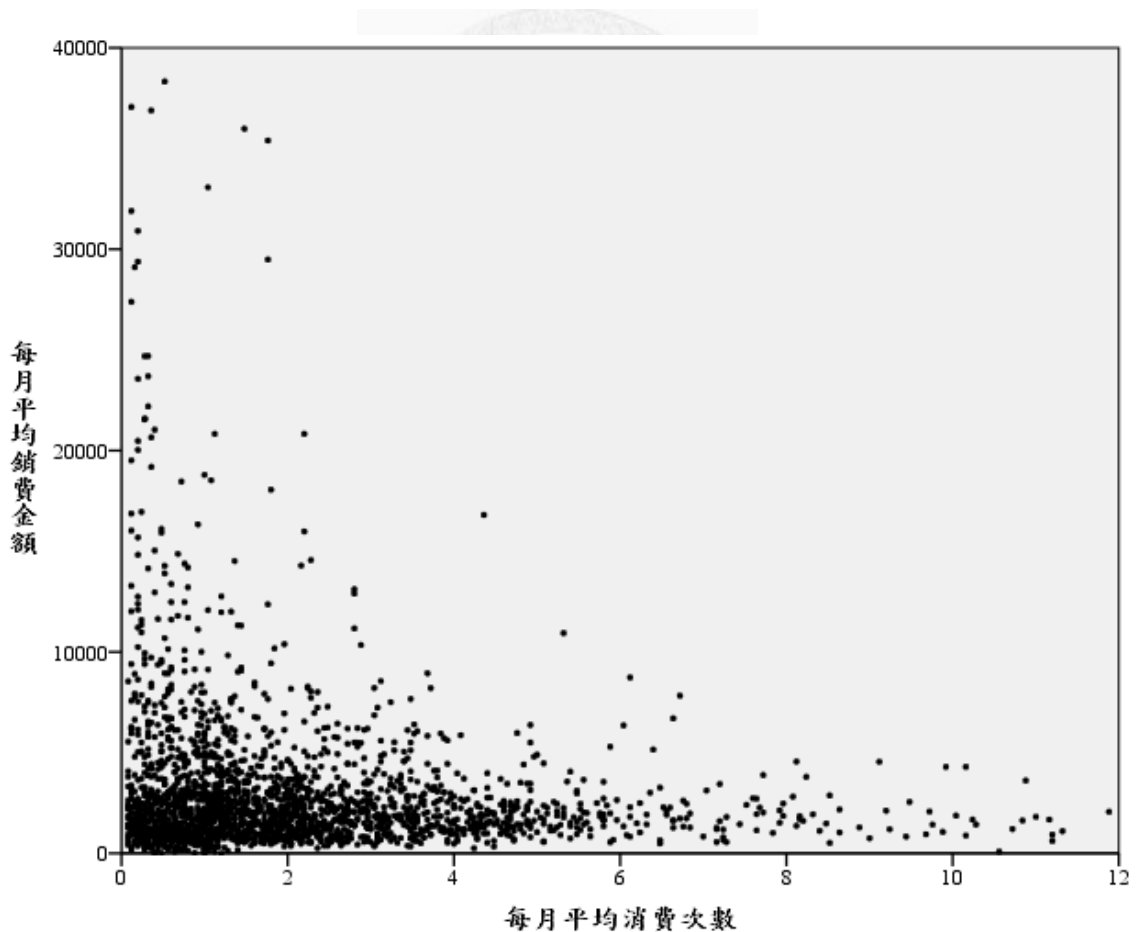


圖 4.1 顧客消費金額與頻率分布圖

敘述統計量如表 4.1：

表 4.1 消費金額與消費頻率之敘述統計量

	消費金額(元)	消費頻率(次)
平均數	3124.31	2.10
中間值	2000.50	1.48
眾數	3000.00	1.00
標準差	4353.37	1.97
最小值	3	0.08
最大值	75676	18.88

二、 人口統計變數資料整理

人口統計變數包含：卡號、年收入、婚姻狀況、教育程度、居住地(以縣市為單位)、卡片級別。為方便研究分析，將部分變數之水準合併，並重新定義之，如下表所示：

表 4.2 人口統計變數水準合併表

變數	水準	原始水準
教育程度	大學以上	大學、碩士、博士
	其他	高中職、專科、其他
卡片等級	白金卡以上	Platinum、Excellent、Infinity
	普卡	Classic、Gold
居住地	北部	台北、基隆、桃園、新竹
	中部	苗栗、台中、彰化、雲林、南投
	南部	嘉義、台南、高雄、屏東
	東部	宜蘭、花蓮、台東、澎湖、金門
收入	低	0~250000
	中低	250001~400000
	中高	400001~600000
	高	>600000

將人口統計變數的各個區隔之交易紀錄資料整理後，計算出各區隔的人數、比例、平均每月消費金額及平均每月消費次數，如表 4.3 所示。

表 4.3 人口結構特性

人口統計變數		人數	百分比	月平均消費次數	月平均消費金額
性別	男性	988	42.68%	2.22	3172.94
	女性	1324	57.32%	2.02	3088.02
婚姻狀況	已婚	1093	47.26%	1.95	2812.51
	未婚	883	38.18%	2.20	3318.21
	其他	336	14.55%	2.32	3266.85
教育程度	大學以上	677	29.41%	2.19	3353.75
	其他	1635	70.59%	2.07	3029.30
卡片級別	普卡	462	20.08%	1.88	3506.11
	白金卡以上	1850	79.92%	2.16	3028.96
居住地	北部	1174	50.60%	2.20	3260.51
	中部	374	16.22%	2.11	3755.67
	南部	679	29.28%	1.92	2606.52
	東部	85	3.90%	2.15	2601.25
年收入	低	808	34.89%	2.23	3025.46
	中低	442	19.26%	1.98	2555.51
	中高	552	23.76%	2.08	3064.45
	高	510	22.09%	2.03	3838.65

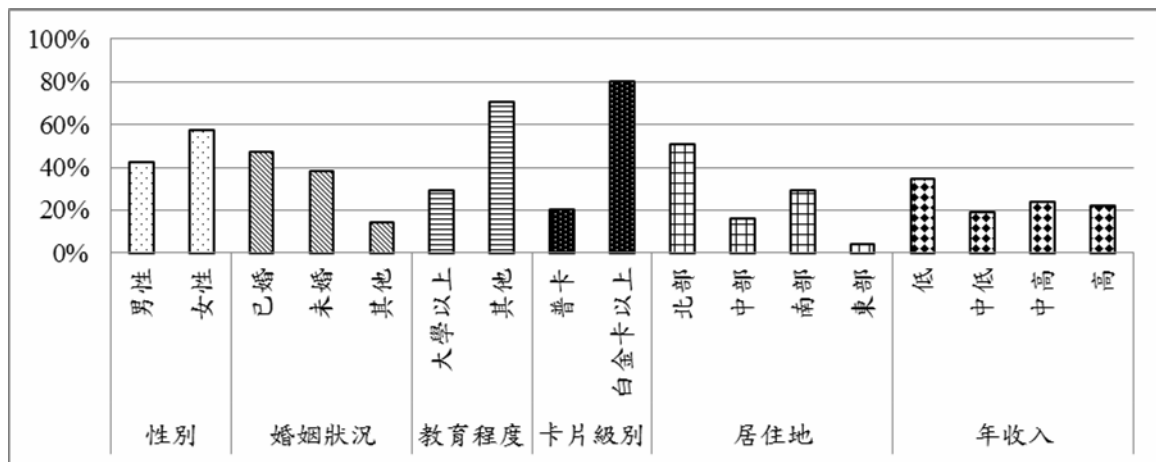


圖 4.2 人口變數結構特性圖

從圖 4.2 可以看出大學以上學歷的顧客比例較低，僅占 29.41%；卡片級別則是白金卡等級以上的居多，佔總人數的 79.92%；從居住地的資料可以看出北部人數比例最高，超過半數，而東部的顧客人數比例最低；婚姻狀況未婚和已婚的人數比例差距不大，其他則表示可能已離婚或喪偶等狀況，佔了 14.55%。

接著，為了分析各區隔在消費次數與消費金額的行為特性是否有差異，本文利用變異數分析做檢定 (表 4.4、表 4.5)。

表 4.4 平均消費金額之變異數分析

人口統計變數		平均消費金額	標準差	F	P-值	臨界值
性別	男性	3172.94	4542.05	0.2152	0.6427	3.8455
	女性	3088.02	4208.44			
婚姻狀況	已婚	2812.51	3313.59	3.5515	0.0288*	2.9996
	未婚	3318.21	5211.89			
	其他	3266.85	3497.64			
教育程度	大學以上	3353.75	4714.34	2.6612	0.1030	3.8455
	其他	3029.30	4192.66			
卡片級別	普卡	3506.11	4867.82	4.4477	0.0351*	3.8455
	白金卡以上	3028.96	4211.15			
居住地	北部	3260.51	3909.97	6.6647	0.0002*	2.6088
	中部	3755.67	6607.31			
	南部	2606.52	3594.48			
	東部	2601.25	2301.71			
年收入	低	3025.46	3980.60	7.3258	0.0001*	2.6088
	中低	2555.51	2879.95			
	中高	3064.45	3454.05			
	高	3838.65	6325.50			

表 4.5 平均消費次數之變異數分析

人口統計變數		平均消費次數	標準差	F	P-值	臨界值
性別	男性	2.22	2.18	5.8424	0.0157*	3.8455
	女性	2.02	1.79			
婚姻狀況	已婚	1.95	1.83	6.2622	0.0019*	2.9996
	未婚	2.20	1.99			
	其他	2.32	2.31			
教育程度	大學以上	2.19	2.00	1.8989	0.1683	3.8455
	其他	2.07	1.96			
卡片級別	普卡	1.88	1.77	7.1220	0.0077*	3.8455
	白金卡以上	2.16	2.01			
居住地	北部	2.20	2.03	2.8172	0.0378*	2.6088
	中部	2.11	1.97			
	南部	1.92	1.86			
	東部	2.15	1.86			
年收入	低	2.23	1.99	2.0292	0.1077	2.6088
	中低	1.98	1.81			
	中高	2.08	2.04			
	高	2.03	1.98			

由以上檢定的結果，我們可以得到以下幾點結論：

1.以性別而言，在平均消費金額上並無顯著差異，但男性持卡人在消費次數上顯著高於女性持卡人。

2.以婚姻狀況而言，未婚的持卡人平均消費金額顯著低於已婚持卡人及其他持卡人；而在刷卡次數方面，已婚的持卡人顯著低於未婚持卡人及其他持卡人。未婚持卡人頻率較高但刷卡金額低，而已婚持卡人則呈現相反的情況，即刷卡金額高但頻率較低，顯示其具有行為上的差異。

3.以教育程度而言，不論是平均消費次數或是平均消費金額，都不具有顯著差異。

4.以卡片級別而言，普卡持卡人的平均消費金額顯著高於白金卡以上等級之持卡人；而白金卡以上等級之持卡人的平均消費次數顯著高於普卡持卡人。

5.以居住地而言，中部地區持卡人的平均消費金額顯著高於其他地區的持卡人，而北部地區持卡人的平均消費金額次高，且顯著高於南部及東部地區的持卡人；在平均消費次數方面，則得到北部持卡人之消費頻率顯著高於南部持卡人之消費頻率。

6.以年收入而言，高收入持卡人的平均消費金額顯著高於其他持卡人，而中低收入持卡人的平均消費金額則顯著低於其他持卡人；在平均消費次數方面則無顯著差異。

根據以上結果，即是一個利用資料庫結合統計初步了解顧客的例子，不同的區隔具有不同的行為模式。我們可以利用這些結果快速(粗略)篩選出可能對企業具有較高價值的顧客，例如我們想找出高消費次數及高消費金額的顧客，可以挑選男性、北部且高收入的顧客，這些人可能為高價值顧客的機率相對較高。資料庫行銷的最終目的是達到一對一行銷，所以分析不會僅止於此，下一節會針對顧客價值的穩定性及活躍性做更進一步的分析。

第二節 購買區間分析

購買區間 (Inter-purchase Time) 指的是一顧客兩次購買之間所間隔的時間，單位可能是分、時、日、周或月，本研究以天為單位計算顧客購買區間，若同一天內刷卡兩次以上皆計為同一筆刷卡紀錄。分析顧客的購買區間可以幫助我們更了解顧客價值，我們可以透過估計購買區間而得知某顧客購買的頻率、週期、甚至可以用來預估顧客未來的行為。

而利用購買區間的最大概似估計量 (MLE)、加權最大概似估計量 (WMLE)、層級貝氏估計量，可以計算出顧客的「活躍性指標」和「穩定性指標」。因此我們可以透過「活躍性指標」知道哪些顧客刷卡的行為越來越活躍，也可以了解我們有可能失去哪些顧客；而「穩定性指標」越高則表示該顧客的個人行為變異性小，反之則顯示個人行為的變異性較大。

在這一節將依序整理最大概似估計量、加權最大概似估計量以及層級貝氏估計量的計算結果，接著依據這些結果計算活躍性指標和穩定性指標，並利用統計學上的檢定方法判斷各區隔的顧客在刷卡行為上是否存在差異性。

一、 最大概似估計量

根據最大概似估計量計算結果，可以得知平均而言，顧客約 19.45 天會有刷卡的行為產生。而根據圖 4.3 可以看出顧客的購買區間絕大部分小於 30 天，佔了約 86% 的比重，顯示這些顧客每個月至少會有一次刷卡的紀錄，且在未來一個月內很有可能再度刷卡。另一方面，也可以看到在圖的右端仍部份顧客的購買區間相當大，購買區間大於 300 以上顯示此顧客可能將近一年才會使用一次信用卡，表示失去此顧客的可能性很高。

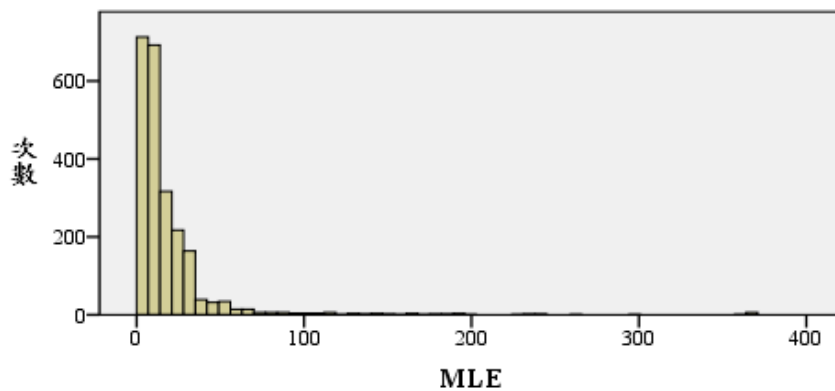


圖 4.3 最大概似估計量直方圖

二、 加權最大概似估計量

加權最大概似估計量 (WMLE) 最主要的特色在於，越接近現在的購買區間會給予較重的加權分數，也就是說加權最大概似估計量比起最大概似估計量更重視最近一次購買紀錄的重要性。當某顧客的加權最大概似估計量大於最大概似估計量時，表示該顧客較不活躍 (Inactive)；反之，當某顧客的加權最大概似估計量小於最大概似估計量時，表示該顧客越來越活躍 (Active)。根據計算結果，加權最大概似估計量的平均值為 20.73 天，大於最大概似估計量的平均值 19.45 天，整體而言顧客的狀態較不活躍。

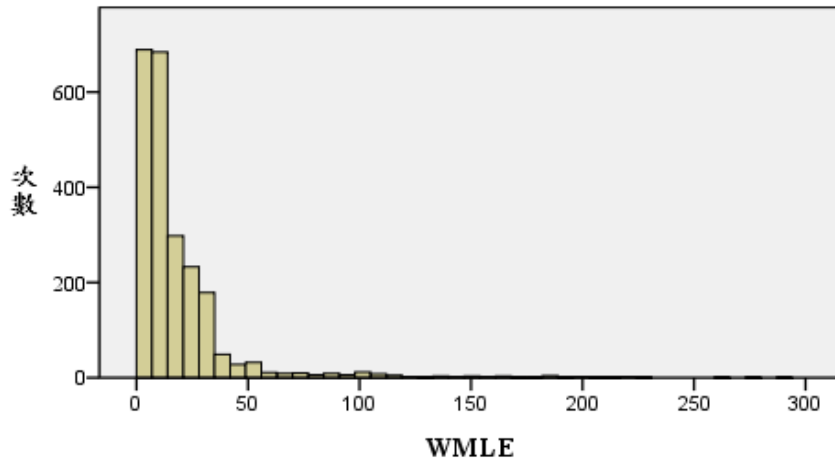


圖 4.4 加權最大概似估計量直方圖

三、 層級貝氏估計量

由於加權最大概似估計量與最大概似估計量在估計時僅考慮到個別顧客的購買紀錄，因此，當某顧客的購買次數太少的時候，傳統的估計方法可能產生嚴重的偏誤。由於層級貝氏估計量投入人口統計變數作為第二層的變數，所以此估計量不但能反映個別顧客層次的行為，也可以考慮到無法被觀察到的顧客異質性，且因為其將全體顧客購買紀錄納入估計，可以幫助推估購買紀錄資料較少之顧客的購買區間，以降低誤差。

本研究利用統計軟體 R 的 Bayesm 套件計算，得到購買區間的層級貝氏估計量，並以吉布斯抽樣方法 (Gibbs Sampling) 隨機抽取且重複模擬 1000 次，並以其平均值得到估計值。計算結果所有顧客的購買區間平均值為 14.87 天，標準差為 18.68，皆低於最大概似估計量與加權最大概似估計量的估計值，以表 4.6 整理後可以看出三者之間的差異。

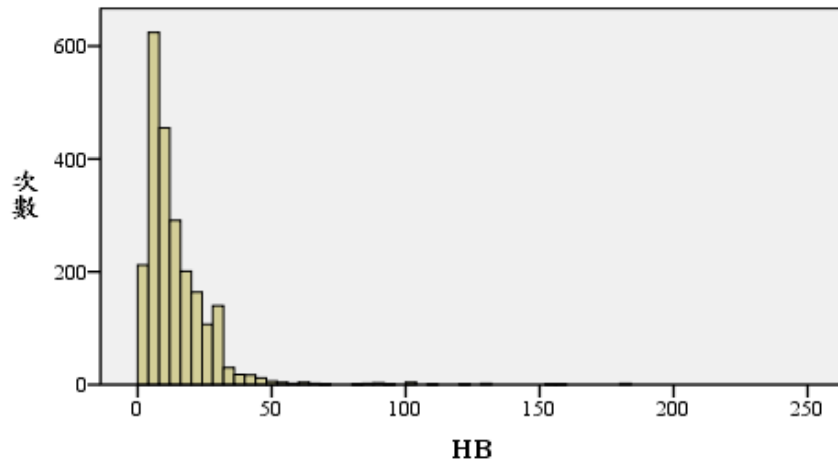


圖 4.5 層級貝氏估計量直方圖

表 4.6 購買區間估計量之比較

	最大概似估計量	加權最大概似估計量	層級貝氏估計量
平均數	19.49	20.73	14.87
標準差	35.53	37.39	18.68
第一四分位數	6.09	6.28	6.23
中位數	10.55	10.91	10.57
第三四分位數	21.31	22.64	18.98

第三節 顧客穩定性及活躍性分析

一、 活躍性分析

本研究採用之「顧客活躍性指標」的概念為最大概似估計值減去加權最大概似估計值的變動百分比，以公式表示如下：

$$\left(\frac{MLE - WMLE}{MLE}\right) \times 100\% \quad \text{式(2.1)}$$

當顧客活躍性指標小於 0 時，表示該顧客不活躍，表示其購買區間越來越長，刷卡的頻率越來越低；反之，當顧客活躍性指標為正且數值越大時，表示其活躍性越高，即該顧客購買區間有縮短的傾向，刷卡的頻率越來越高。我們挑選 4 位顧客的購買區間紀錄作為例子，整理成表格如表 4.7，可以清楚看出購買區間改變的模式，前兩名屬於活躍的顧客，即購買區間縮減；而後兩名屬於不活躍的顧客，購買區間有延長的趨勢。

表 4.7 顧客活躍性指標範例

顧客編號	活躍指標	第 1 期	第 2 期	第 3 期	第 4 期	第 5 期	第 6 期	第 7 期	第 8 期	第 9 期	第 10 期	第 11 期	第 12 期	第 13 期	第 14 期	第 15 期	第 16 期	第 17 期	第 18 期
2115	36.9%	44	48	50	41	23	29	33	28	271	51	28	5	1	16	6	1	3	2
4416	32.6%	56	1	28	13	56	86	25	3	4	19	14	2	7	10	4	1	6	4
9358	-83.1%	2	2	23	34	27	16	15	33	28	31	77	181	186					
3110	-63.6%	53	3	33	57	156	116	274											

整體顧客的活躍性分析如表 4.8，可以看出不活躍的顧客佔了 61.68%的比重，且平均數為-5.52%。為了進一步了解活躍與不活躍的顧客之間，在個人變數的區隔之下，是否有顯著的差異，我們利用列聯表作卡方檢定，如表 4.9 顯示，各人口統計變數之間並無顯著差異。

表 4.8 活躍性分析

活躍性指標	人數	比例
活躍	886	38.32%
不活躍	1426	61.68%
平均數		-5.52%
標準差		18.17%

表 4.9 活躍性指標列聯表卡方檢定

人口統計變數		個數	活躍	不活躍	總和	P-value
性別	女性	觀察個數	495	829	1324	0.284
		期望個數	507.4	816.6	1324	
	男性	觀察個數	391	597	988	
		期望個數	378.6	609.4	988	
婚姻狀況	已婚	觀察個數	431	662	1093	0.353
		期望個數	418.9	674.1	1093	
	未婚	觀察個數	322	561	883	
		期望個數	338.4	544.6	883	
	其他	觀察個數	133	203	336	
		期望個數	128.8	207.2	336	
教育	大專以上	觀察個數	263	414	677	0.738
		期望個數	259.4	417.6	677	
	其他	觀察個數	623	1012	1635	
		期望個數	626.6	1008.4	1635	
卡片	白金卡	觀察個數	710	1140	1850	0.911
		期望個數	709	1141	1850	
	普卡	觀察個數	176	286	462	
		期望個數	177	285	462	
居住地	中	觀察個數	155	219	374	0.278
		期望個數	143.3	230.7	374	
	北	觀察個數	431	743	1174	
		期望個數	449.9	724.1	1174	
	東	觀察個數	37	48	85	
		期望個數	32.6	52.4	85	
	南	觀察個數	263	416	679	
		期望個數	260.2	418.8	679	
收入	中低收入	觀察個數	183	259	442	0.475
		期望個數	169.4	272.6	442	
	中高收入	觀察個數	210	342	552	
		期望個數	211.5	340.5	552	
	低收入	觀察個數	298	510	808	
		期望個數	309.6	498.4	808	
	高收入	觀察個數	195	315	510	
		期望個數	195.4	314.6	510	

二、 穩定性分析

穩定性指標可以幫助我們了解顧客在購買區間的行為上變異性的大小，本研究採用的計算公式如下：

$$\frac{|HB_i - IM_i|}{|HB_i - GM_j|} \quad \text{式(2.2)}$$

其中，

HB_i 表示顧客 i 的層級貝氏估計值；

IM_i 表示顧客 i 的個人平均值 (Individual Mean)，在本研究為最大概似估計值；

GM_j 表示顧客 i 所屬之區隔 j 的群體平均值 (Group Mean)，此區隔依據人口統計變數分割，且由層級貝氏估計量所得到各人口變數的迴歸係數去估計群體平均值 (例如男性、未婚、大專以上、持白金卡、住北部且高收入的顧客，購買區間的群體平均值皆為 10.33)。

當穩定性指標小於 1 時，表示該顧客穩定，其購買區間受到個人資訊影響較大，個人行為變異性小；反之，當穩定性指標大於 1 時，表示該顧客的購買區間受到群體資訊影響較大，個人行為變異大，顧客較不穩定。我們隨機挑選 4 位顧客，列出前 20 筆購買區間紀錄和穩定性指標，如表 4.10，可以幫助我們了解穩定的顧客和不穩定的顧客在購買區間的行為模式。

根據表 4.11，整體而言有 93.03%的顧客屬於穩定，穩定性指標的平均值為 0.54，變異數為 0.11。接著仿照活躍性的做法，利用列聯表卡方檢定，嘗試找出各人口統計變數的區隔之間，是否存在顯著的差異。檢定結果如表 4.12，各個區隔彼此之間在穩定性的行為上並無顯著差異。

表 4.10 顧客穩定性指標範例

顧客編號	穩定性指標	第 1 期	第 2 期	第 3 期	第 4 期	第 5 期	第 6 期	第 7 期	第 8 期	第 9 期	第 10 期	第 11 期	第 12 期	第 13 期	第 14 期	第 15 期	第 16 期	第 17 期	第 18 期
5139	0.00	59	61	62	60	61	62	59	61	58	61	61							
2140	0.04	7	10	12	16	14	13	5	10	19	14	11	6	13	18	11	19	14	19
1062	1.02	10	127	5	20	90	183	157											
3919	1.19	7	5	16	28	1	3	6	87	10	21	78	9	87	8	56	1		

表 4.11 穩定性分析

穩定性指標	人數	比例
穩定	2151	93.03%
不穩定	162	6.97%
平均數		0.54
標準差		0.11

表 4.12 穩定性指標列聯表卡方檢定

人口統計變數		個數	活躍	不活躍	總和	P-value
性別	女性	觀察個數	85	1239	1324	0.234
		期望個數	92.2	1231.8	1324	
	男性	觀察個數	76	912	988	
		期望個數	68.8	919.2	988	
婚姻狀況	已婚	觀察個數	76	1017	1093	0.537
		期望個數	76.1	1016.9	1093	
	未婚	觀察個數	66	817	883	
		期望個數	61.5	821.5	883	
	其他	觀察個數	19	317	336	
		期望個數	23.4	312.6	336	
教育	大專以上	觀察個數	51	626	677	0.489
		期望個數	47.1	629.9	677	
	其他	觀察個數	110	1525	1635	
		期望個數	113.9	1521.1	1635	
卡片	白金卡	觀察個數	126	1724	1850	0.563
		期望個數	128.8	1721.2	1850	
	普卡	觀察個數	35	427	462	
		期望個數	32.2	429.8	462	
居住地	中	觀察個數	24	350	374	0.926
		期望個數	26.0	348.0	374	
	北	觀察個數	85	1089	1174	
		期望個數	81.8	1092.2	1174	
	東	觀察個數	5	80	85	
		期望個數	5.9	79.1	85	
	南	觀察個數	47	632	679	
		期望個數	47.3	631.7	679	
收入	中低收入	觀察個數	26	416	442	0.420
		期望個數	30.8	411.2	442	
	中高收入	觀察個數	43	509	552	
		期望個數	38.4	513.6	552	
	低收入	觀察個數	51	757	808	
		期望個數	56.3	751.7	808	
	高收入	觀察個數	41	469	510	
		期望個數	35.5	474.5	510	

第四節 層級貝氏 Probit 模型估計與預測力分析

一、資料整理

本研究利用 RFM 模型與層級貝氏 Probit 模型，用來預估個別顧客每一期的消費狀態。利用 Probit 模型首先要將所要分析的狀況分為有限個狀態 (States)，本研究以個別顧客為基礎，採用 RFM 模型中的購買頻率 (Frequency) 和購買金額 (Monetary Amount)。另外要注意的是，本研究中購買頻率表示刷卡的頻率，而購買金額則表示刷卡的金額。

我們將各狀態的定義與資料個數整理如表 4.13，其中狀態 1 表示完全沒有消費；狀態 2 低頻率低消費，價值最低；而狀態 17 頻率最高、金額最高，整體價值最高。

表 4.13 狀態定義與資料個數

		月平均消費次數				
		0	1	2~3	4~5	>5
月平均消費金額	>3000	16,837 (狀態 1)	2,973 (狀態 5)	2,909 (狀態 9)	1,383 (狀態 13)	1,919 (狀態 17)
	1501~3000		3,781 (狀態 4)	3,376 (狀態 8)	1,794 (狀態 12)	2,792 (狀態 16)
	801~1500		2,446 (狀態 3)	3,218 (狀態 7)	1,912 (狀態 11)	3,160 (狀態 15)
	0~800		3,464 (狀態 2)	3,123 (狀態 6)	1,273 (狀態 10)	2,041 (狀態 14)

二、 模型估計

本研究利用 Guass 軟體，投入 20 期的交易紀錄資料與人口統計變數資料，並以馬可夫鏈蒙地卡羅 (Markov Chain Monte Carlo; MCMC) 方法中的吉布斯抽樣 (Gibbs Sampling) 模擬 1000 次後取其平均，求出平均相對效用的估計值，以及各迴歸係數的估計值。

原始資料某些顧客可能缺乏某幾個狀態的購買紀錄，造成無法計算其平均相對效用，稱為樣本外 (Out Sample) 顧客，我們可以透過樣本內顧客之迴歸係數的估計值，連結人口統計變數而估計樣本外顧客的平均相對效用。

接著我們將平均相對效用以 CDF 的概念轉換成機率值，得出每位顧客在給定 $t-1$ 期的狀態下各個狀態的移轉機率，由於共有 17 個狀態，所以每位顧客會得到 17 個 1×17 的移轉機率向量。將其依照 $t-1$ 期的狀態排列而組成該顧客的移轉機率矩陣，如表 4.14 所示。整體而言，以狀態 1 的發生機率較高，主要原因應是總體顧客在狀態 1 的資料數較多，以至於估計結果受到總體資訊的影響。

表 4.14 移轉機率矩陣 (以顧客 100089 為例)

State	1	2	3	4	5	6	...	17
1	0.1632	0.0676	0.0498	0.0397	0.0328	0.0560	...	0.0959
2	0.2089	0.1146	0.0330	0.0554	0.0814	0.1008	...	0.1178
3	0.1274	0.0419	0.0920	0.0767	0.0393	0.0711	...	0.0841
4	0.0598	0.0619	0.0615	0.0609	0.0600	0.0611	...	0.0311
5	0.0993	0.0696	0.0433	0.0469	0.0881	0.0555	...	0.0617
6	0.1736	0.0778	0.0390	0.0260	0.0288	0.0946	...	0.1196
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	0.1055
17	0.0799	0.0709	0.0647	0.0477	0.0469	0.0473	...	0.0929

三、 預測力分析

本研究引用馬可夫鏈理論中的 Chapman-Kolmogorov Equation 定理，作為移轉機率矩陣預測的基礎。本研究將 25 期的資料分成兩部分，前 20 期作為估計之用，後 5 期則用來驗證模型的預測力。

令 P 表示某顧客由前 20 期資料所估計出來之完整的移轉機率矩陣， P 為一 17×17 的矩陣，則 P 表示該顧客在給定 $t-1$ 期之狀態下， t 期會轉移到任何狀態之機率，於本研究的涵義，表示可以用第 20 ($t-1$) 期的狀態，去預測下一期，也就是第 21 (t) 期的狀態。

又 P^2 代表該顧客在在給定 $t-1$ 期之狀態下， $t+1$ 期會轉移到任何狀態之機率，於本研究的涵義，表示可以用第 20 ($t-1$) 期的狀態，去預測下二期，也就是第 22 ($t+1$) 期的狀態。以此類推，可以預測第 23 期到第 25 期的狀態。以顧客 100089 為例，如表 4.15，狀態 15 之機率最高，對應第 21 期狀態表示預測正確。

表 4.15 狀態移轉預測範例

顧客編號	第 21 期狀態	第 20 期狀態	狀態 1	狀態 2	...	狀態 15	...	狀態 17
100089	15	11	0.0567	0.0281	...	0.1033	...	0.0820

本研究另外引用其他兩種較為直觀的預測方法，以利預測命中率的比較：

1. 隨機機率法：以隨機猜測的方式預測，因此在給定 $t-1$ 期的狀態下，預測

命中的機率為 $\frac{1}{17}$ 。

2. 前期推測法：在已知 $t-1$ 的狀態下，預測下一期也會發生同樣的狀態。

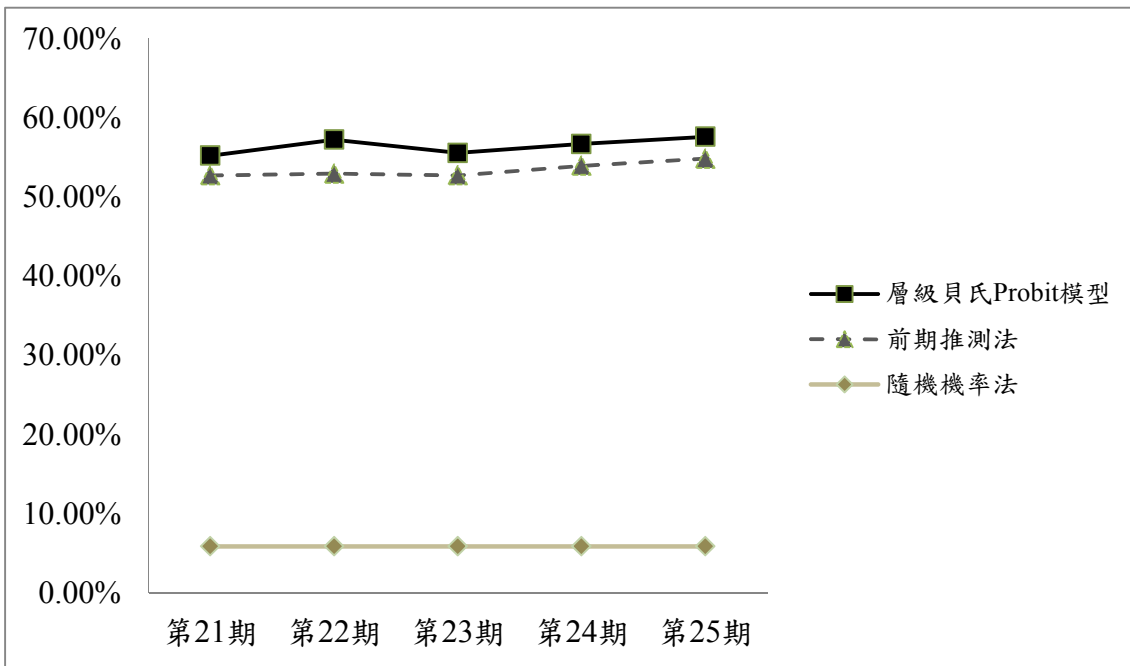


圖 4.6 預測力比較圖

由圖 4.6 可以看出，隨機機率法顯然是效度最差的預測方法；而層級貝氏 Probit 模型則略優於前期推測法，但前期推測法僅能預測下一期的狀態，層級貝氏 Probit 模型則可以利用馬可夫鏈的定理，推測下二期以至於往後數期的狀態，相對而言，比前期推測法更能及早掌握顧客的狀態，提早設計行銷活動應對，且層級貝氏 Probit 模型可以投入更多變數以增加模型的預測能力，這些都是前期推測法無法達到的優點。

下一節將以本節所估計出來的移轉機率矩陣做延伸，更進一步地探討顧客落點與價值路徑的分析。

第五節 顧客落點及路徑分析

一、 最終落點分析

在這一節中，我們首先利用馬可夫鏈理論之狀態機率極限的概念預測每個顧客之最終落點，我們將移轉機率矩陣 P 做 n 期相乘，即 $\lim_{n \rightarrow \infty} P^{(n)}$ ，使原本的移轉機率矩陣收斂成 1×17 之矩陣，表示某顧客在 $t-1$ 期不論發生何種狀態，藉由此收斂矩陣可計算出無數期後該顧客落入任一狀態的機率，機率值最大的狀態即為其最終落點。

最終落點的個數計算結果如下表，接著利用列聯表卡方檢定分析各區隔是否具有顯著特性，除了投入人口統計變數以外，本研究加入穩定性指標和活躍性指標，試圖找出最終落點在此指標的區隔下的差異。由於狀態 2、狀態 4 和狀態 6 個數太少，因此捨棄不討論。檢定結果顯示，穩定性、教育、婚姻狀況、卡片等級、居住地和收入狀況等區隔具有顯著差異，如表 4.17 顯示：

表 4.16 最終落點個數統計

最終落點	狀態 1	狀態 2	狀態 4	狀態 6	狀態 17
層級貝氏 Probit 模型	2223	3	4	4	78

表 4.16 最終落點區隔特性分析

區隔	狀態 1	狀態 17
穩定性	92.76%為穩定	100%顧客皆為穩定
教育	71.35%為其他教育程度	48.72%為大專以上教育程度
婚姻狀況	48.72%為已婚	47.44%為未婚
卡片	79.44%為白金卡以上	97.44%為白金卡以上
居住地	49.66%居住於北部 30.32%居住於南部	78.21%居住於北部
收入	34.37%為低收入 24.02%為中高收入	50.00%為低收入 25.64%為高收入

表 4.18 最終落點列聯表卡方檢定

區隔變數		個數	狀態 1	狀態 17	總和	P-value	區隔變數		個數	狀態 1	狀態 17	總和	P-value
穩定性	不穩定	觀察	161	0	161	0.01*	居住地	中	觀察	362	12	374	0.00*
		期望	155.5	5.5	161				期望	361.3	12.7	374	
	穩定	觀察	2062	78	2140			北	觀察	1104	61	1165	
		期望	2067.5	72.5	2140				期望	1125.5	39.5	1165	
活躍性	不活躍	觀察	1366	52	1418	0.41		東	觀察	83	2	85	
		期望	1369.9	48.1	1418				期望	82.1	2.9	85	
	活躍	觀察	857	26	883			南	觀察	674	3	677	
		期望	853.1	29.9	883				期望	654.1	22.9	677	
性別	女性	觀察	1277	40	1317	0.30	收入	中低	觀察	438	1	439	0.00*
		期望	1272.4	44.6	1317				期望	424.1	14.9	439	
	男性	觀察	946	38	984			中高	觀察	534	18	552	
		期望	950.6	33.4	984				期望	533.3	18.7	552	
教育	大專以上	觀察	637	38	675	0.00*		低	觀察	764	39	803	
		期望	652.1	22.9	675				期望	775.8	27.2	803	
	其他	觀察	1586	40	1626			高	觀察	487	20	507	
		期望	1570.9	55.1	1626				期望	489.8	17.2	507	
婚姻狀況	已婚	觀察	1083	10	1093	0.00*	卡片	白金卡	觀察	1766	76	1842	0.00*
		期望	1055.9	37.1	1093				期望	1779.6	62.4	1842	
	未婚	觀察	840	37	877			普卡	觀察	457	2	459	
		期望	847.3	29.7	877				期望	443.4	15.6	459	
	其他	觀察	300	31	331								
		期望	319.8	11.2	331								

二、 顧客價值遷徙路徑分析

我們可以利用移轉機率矩陣進行顧客價值遷徙路徑分析，分析結果可以幫助企業在發卡的同時就能預測顧客未來價值的發展藍圖，進一步針對不同的遷徙路徑提供不同的顧客管理方法。建構分析的步驟依序為：定義具有意義的路徑、計算路徑發生之機率、顧客特質分析，以下將依據各步驟做詳細的說明。

(1) 定義具有意義的路徑

本研究引用任立中和陳靜怡 (2007) 的研究，設定上升及下降兩大類價值遷徙路徑，每一類並根據消費次數與金額的增減而細分為三種，總共六種路徑。上升類路徑的顧客將由低價值狀態逐步增加頻率或次數，最終達到最高的顧客價值；反之，下降路徑的顧客由高價值狀態逐步降低頻率或次數，最終達到最低的顧客價值。

以下將六種顧客價值遷徙路徑以圖示之，其中上色方塊及箭頭為顧客價值遷徙方向，◎表示停留一期，停留的目的在於保證個路徑具有相同遷徙次數，方便比較移轉機率。如圖四—七為上升路徑中，先提高消費金額再提高消費頻率的路徑；而圖四—九則表示圖時增加消費金額和提高消費頻率的上升路徑；圖四—十則是下降路徑中，先降低消費頻率，再減少消費金額的路徑。

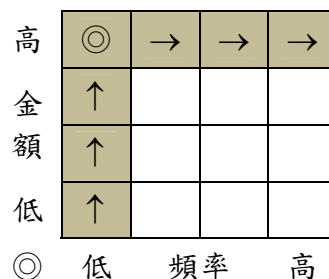


圖 4.7 路徑 1 先增額後增次

高				↑
金				↑
額				↑
低	→	→	→	◎

◎ 低 頻率 高

圖 4.8 路徑 2 先增次後增額

高				◎
金			◎	
額		◎		
低	◎			

◎ 低 頻率 高

圖 4.9 路徑 3 同時增次增額

高	◎	←	←	←
金	↓			
額	↓			
低	↓			

◎ 低 頻率 高

圖 4.10 路徑 4 先減次後減額

高				↓
金				↓
額				↓
低	←	←	←	◎

◎ 低 頻率 高

圖 4.11 路徑 5 先減額後減次

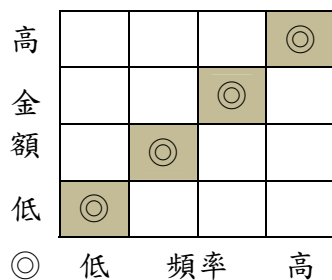


圖 4.12 路徑 6 同時減次減額

(2) 計算顧客遵循不同價值遷徙路徑之可能性

透過個人化移轉機率矩陣計算其遵循不同價值遷徙路徑的可能性，假設該顧客狀態移轉之步驟為 $j \rightarrow k \rightarrow m$ ，則顧客遵循該路徑的機率計算公式如下：

$$p(j, k, m) = p(j) \times p(k|j) \times p(m|k) \quad \text{式(2.3)}$$

由於路徑發生機率是經過數次狀態移轉機率的乘積，其機率值逼近 0 不易比較，因此本研究首先對於路徑發生機率取自然對數之後，再進一步分析：

$$\ln p(j, k, m) = \ln p(j) + \ln p(k|j) + \ln p(m|k) \quad \text{式(2.4)}$$

計算結果各路徑的顧客人數統計如表 4.19，可以看出路徑 5，也就是先減額後減次路徑的人數最多，顯示當顧客開始減少消費金額時，未來可能進一步減少消費次數的可能性也會提升；而路徑 3，同時增額與增次路徑人數最少。

表 4.19 各路徑之人數

	路徑 1	路徑 2	路徑 3	路徑 4	路徑 5	路徑 6
人數	297	221	149	311	980	354

(3) 顧客特質分析

最後，我們以逐步迴歸 (Stepwise Regression) 模型進行分析，探討顧客價值遷徙路徑與顧客特質之間的關係，統計結果如下表：

表 4.20 上升路徑之逐步迴歸分析

上升	路徑 1			路徑 2			路徑 3		
	迴歸係數		P-Value	迴歸係數		P-Value	迴歸係數		P-Value
顧客特質	截距	-19.618	.000	截距	-18.021	.000	截距	-18.097	.000
	男性	.212	.001	男性	.555	.000	北部	-.230	.001
	北部	-.304	.000	南部	.377	.000	南部	.288	.001
	收入	.132	.014	次數	.025	.009	白金卡	-.211	.005
	次數	-.016	.017	已婚	-.431	.002			
<i>F</i> -Statistics	11.879			16.400			11.701		
<i>P</i> -Value	.000			.000			.000		
<i>adj R</i> ²	.128			.219			.224		

以路徑 1 而言，男性且收入越高越有可能遵循此路徑；而住北部及首期刷卡次數越高越不會遵循此路徑。

以路徑 2 而言，男性、住南部、首期刷卡次數越高越會遵循此路徑；而已婚的顧客則遵循此路徑的可能性較低。

以路徑 3 而言，住南部的顧客較可能遵循此路徑；而住北部且持白金卡的顧客遵循此路徑的可能性較低。

表 4.21 下降路徑之逐步迴歸分析

下降	路徑 4			路徑 5			路徑 6		
	迴歸係數		P-Value	迴歸係數		P-Value	迴歸係數		P-Value
顧客特質	截距	-18.356	.000	截距	-15.629	.000	截距	-17.562	.000
	金額	.083	.001	男性	.529	.000	白金卡	-.574	.000
	南部	-.390	.000	大學	.297	.000	已婚	-.567	.000
	中部	-.271	.019	白金卡	-.216	.000	男性	.259	.000
	大學	.156	.011	收入	-.142	.000	南部	.345	.000
	白金卡	-.161	.046	次數	.011	.000	大學	.407	.000
				已婚	-.298	.000			
				南部	-.500	.000			
			中部	-.460	.000				
F-Statistics	7.563			56.974			49.959		
P-Value	.000			.000			.000		
adj R ²	.113			.340			.454		

以路徑 4 而言，首期消費金額越高、教育程度在大學以上的顧客較容易遵循此路徑；住南部或中部、持白金卡以上等級卡片的顧客遵循路徑 4 的可能性較低。

以路徑 5 而言，男性、教育程度在大學以上、首期刷卡次數越高的顧客較可能遵循此路徑；持白金卡以上等級卡片、收入越高、已婚、住中南部的顧客遵循此路徑的可能性較低。

以路徑 6 而言，男性、住南部、教育程度在大學以上較可能遵循此路徑；持白金卡等級以上、已婚的顧客遵循此路徑的可能性較低。

第五章 結論與建議

第一節 研究結果與行銷意涵

一、 刷卡次數與頻率

本研究整理顧客交易資料與顧客基本資料後，以變異數分析檢定各人口變數區隔在刷卡金額及刷卡次數的行為是否具有顯著差異，結果如下：以刷卡的消費金額而言，已婚、普卡、中部、高收入的持卡人的平均消費金額顯著高於其他區隔的持卡人；而以刷卡的次數而言，男性、未婚、持白金卡以上等級之卡片、居住北部的持卡人平均消費次數高於其他區隔的持卡人。

二、 購買區間與穩定性、活躍性分析

本研究在估計購買區間的最大概似估計量、加權最大概似估計量及層級貝氏估計量後，進一步計算各個顧客的穩定性及活躍性。以整體資料庫而言，大部分的顧客屬於穩定狀態，約佔了 93.03% 的比例；而活躍的顧客僅佔了 38.32%，顯示整體資料庫的顧客為不活躍的狀態，有購買區間增加，刷卡頻率減少的傾向，也間接說明整體顧客價值在下降的狀況。以行銷的角度來看，由於預期整體顧客價值將有下降的可能，行銷人員可以及早著手研究，利用資料庫挑選出不活躍的顧客，深入了解顧客趨於不活躍狀態的原因。

而進一步利用列聯表卡方檢定，試圖找出個人口變數區隔之間的顧客，在活躍性及穩定性上的差異，檢定結果顯示不論是穩定性或活躍性，都不具顯著差異，此結果說明人口統計變量在穩定性及活躍性的行為上解釋效果薄弱。

三、 層級貝氏 Probit 模型預測力

本研究利用馬可夫鏈理論對每位顧客建立移轉機率模型，用以預測各個顧客未來的消費狀態，並以最後 5 期的消費紀錄作為擊中率分析的依據，結果顯示模型預測力平均高達 56.48%，模型預測力高於直觀的前提推測法及隨機預測法。

四、 最終落點與顧客價值遷徙路徑

本研究利用先前所建構的移轉機率矩陣做最終落點分析，結果顯示絕大多數的顧客最終狀態為狀態 1 或狀態 17，分別表示無價值的狀態與最高價值的狀態。其中最終落點為狀態 17 的顧客多為未婚、持白金卡以上等級卡片、居住於北部，有半數為低收入，有 25% 為高收入的持卡人；最終落點為狀態 1 的顧客其多為已婚，居住於北部或南部，高收入的顧客比例較低。

顧客價值遷徙路徑分析的研究結果可以幫助發卡銀行了解不同顧客最有可能的價值遷徙路徑，整理結果如下：男性且收入越高越有可能遵循此先增額後增次的之路徑；男性、住南部、首期刷卡次數越高越會遵循先增次後增額路徑；住南部的顧客較可能遵循此同時增次稱額之路徑。在下降路徑方面，首期消費金額越高、教育程度在大學以上的顧客較容易遵循先減次後減額的路徑；男性、教育程度在大學以上、首期刷卡次數越高的顧客較可能遵循先減額後減次路徑；男性、住南部、教育程度在大學以上較可能遵循同時減額及減次路徑。

發卡銀行可藉此推算個別顧客未來的顧客價值，並擬訂不同的行銷策略。舉例而言，假設某顧客遵循上升路徑的可能性較高，則可以主動建議該顧客將額度升等，以刺激其消費；反之，假設某顧客遵循下降路徑的可能性較高，則在行銷策略上可以著重於各種優惠或紅利積點的活動，提高顧客使用信用卡消費的意願。

第二節 研究限制

一、 交易紀錄資料上的限制

在交易紀錄資料方面，主要有缺失值過多、資料期數較少兩大限制。缺失值的來源是因為許多顧客在某些月份沒有消費，使得狀態 1 的筆數較高，在建構移轉機率矩陣時，總體資訊的估計容易受到其影響，降低預測效果。另外，由於本研究只能取得 25 期的資料，並以前 20 期作為估計之用，後 5 期作為預測力的檢驗，若能增加資料期數，或許能提高模型之準確度。

二、 其他資料上的限制

在人口統計變量資料方面，主要有變數過少或區隔不具代表性的限制，如果可以加入更多變數，例如：年齡、職業等或許可以解釋更多現象。另外，信用卡具有附卡的機制，而其附卡的交易紀錄資料可能導致無法反映主卡持卡人實際刷卡的情況。

除了人口統計變數之外，不同的信用卡本身所具備的優惠與增值服務（例如紅利集點、哩程累積、停車優惠、電影優惠等等），對於信用卡持卡人之使用行為可能有更大的影響力，這也是本研究在資料上的限制。

第三節 未來研究方向

一、 資料庫之擴充

正如研究限制所述，未來若能取得樣本數更大的資料庫，將抽樣時間拉長、抽樣的期數增加，或許能提高模型的準確度。另外，若能投入不同變數，例如年齡、職業、持卡人使用信用卡之優惠的行為紀錄資料（例如紅利集點之點數、哩程累積數、電影優惠使用次數等等）相信能夠強化模型的解釋能力，並更精確地區隔顧客。

二、 交易明細資料之研究

本研究所採用的信用卡資料庫，雖記錄了交易明細資料，但因為資料紊亂未經分類，而缺乏分析價值。若發卡銀行未來有意強化資料庫行銷，可以從此處下手，將交易明細資料做全面性的分門別類，區分出不同的屬性，例如同樣是餐廳的消費紀錄，但速食店與高級餐廳代表著不同的消費行為，同時也具備了不同行銷意涵，若能對此做深入的研究，必能充分發揮資料庫行銷的優點，以達到一對一行銷的最終目的。

REFERENCE

一、 中文部分

- [1] 王仕茹 (1999) ,「整合層級貝氏聯合區隔與定位分析模式：來源國效應評價、品牌權益衡量與新產品設計之應用」, 國立台灣大學國際企業學研究所博士論文。
- [2] 任立中、王仕茹 (1999) ,「通路結構中市場佔有率之競爭分析：不完全資訊下之貝氏統計模式」, 中華決策科學學會 1999 年會, 創造 21 世紀決策管理新典範研討會論文。
- [3] 任立中、陳靜怡 (2007) ,「顧客價值遷移路徑分析：馬可夫鏈模型」, 台大管理論叢第 17 卷第二期。
- [4] 林宗龍 (1994) ,「產量彈性與改進活動之關係—台灣汽車廠之實證研究」, 國立台灣大學未出版碩士論文。
- [5] 柳慧琴 (1997) ,「資料庫行銷之顧客價值分析模式」, 國立台灣大學國際企業學研究所未出版碩士論文。
- [6] 陳文華 (1999) , 應用資料倉儲系統建立 CRM, 《資訊與電腦》第 226 期 p122-127。
- [7] 陳文華 (2000) , 顧客關係管理成功的關鍵—高品質的顧客資料, 《能力雜誌》第 278 期 p132-135。
- [8] 陳文華 (2000) , 運用資料倉儲技術於顧客關係管理, 《能力雜誌》第 527 期 p.132-138。
- [9] 馮淑群 (1998) ,「傳銷公司顧客價值分析與產品組合策略之研究」, 國立台灣大學國際企業學研究所未出版碩士論文。
- [10] 劉穎壽 (1994) ,「資料庫行銷—顧客資料庫的建立其應用之研究」, 國立政治大學企業管理研究所碩士論文。
- [11] 賴柏志、孫銘誼 (2011) ,「台灣信用卡市場的發展趨勢與現況分析」, 《金融聯合徵信雜誌》民國一百年六月號。
- [12] IBM Inc (2012) ,「客戶關係管理白皮書」。
- [13] R.S. Swift (2001) ,「深化顧客關係管理」, 遠擎管理顧問股份有限公司。

二、 英文部分

- [1] Allenby, Greg M., Neeraj Arora and James L. Ginter (1995), "Incorporating Prior Knowledge into the Analysis of Conjoint Studies," *Journal of Marketing Research*, 32, 152-162.
- [2] Allenby, Greg M. and Peter E. Rossi (1999), "Marketing Models of Consumer Heterogeneity," *Journal of Econometrics*, 89, 57-78.
- [3] Allenby, Greg M., Peter E. Rossi (2000), "Statistics and Marketing," *Journal of the American Statistical Association*, Vol.95, pp.635-638.
- [4] Allenby, Greg M., R. P. Leone, and Lichung Jen (1999), "A Dynamic Model of Purchase Timing with Application to Direct Marketing," *Journal of the American Statistical Association*, 94, 365-374.
- [5] Arthur Hughes (1995). *Strategic Database Marketing*, McGraw-Hill Inc.
- [6] Cespedes, Frank V. and H. Jeff Smith (1993), "Database Marketing: New Rules for Policy and Practice," *Sloan Management Review* (Summer), 7-22
- [7] Denise, Geoffrey, Dawn, Linda, "Understanding consumer database marketing," *Journal of Consumer Marketing*, Vol 14, Number 1, 1997
- [8] Davids, Meryl (1999), "How to Avoid the 10 Biggest Mistake in CRM," *Journal of Business Strategy*, November ,pp.22-26.
- [9] David Shani & Sujana Chalasani,, (1992) *Exploiting Niches Using Relationship Marketing*, *The Journal of Services Marketing*, Vol. 6 No. 4, Fall 1992, pp.43-51
- [10] Hughes, Arthur M., *Strategic Database Marketing*, Chicago : Probus Publishing. 1994
- [11] John D.C. Little(1979),"Decision Support Systems For Marketing Managers", *Journal of Marketing*, Vol 43, Summer 1979,p9-26
- [12] Joseph, Anthony P., Conway Lackman, A. Graham Peace and Gerald Tatar (1999), "Leveraging Customer Database for Strategic Marketing Advantage in The Retail Industry," *Journal of Database Marketing*, Vol.7, No.1, pp.53-59.
- [13] Kalakota, R., & Robinson , M.(1999) "e-Business:Roadmap for Success, 1st ed.", Mary T.O'Brien, U.S.A.
- [14] Kotler, Philip, (2000), "Kotler on marketing: how to create, win, and dominatemarkets", Simon & Schuster Inc.

- [15] Newell, Frederick (2003), 「 Why CRM Doesn't Work? —How To Win By Letting Customers Manage the Relationship. 」, Princeton N.J: Bloomberg Press.
- [16] Pepper, D., M. Rogers & B. Dorf (1999), “Is your company ready for one to one marketing?”, Harvard Business Review, Jan/Feb:151-160.
- [17] Pfeifer, Philip E. and R. L. Carraway. "Modeling Customer Relationships as Markov Chains." Journal of Interactive Marketing, Vol.14, No.2, 2000, p.43.
- [18] Rossi, Peter E., Greg M. Allenby and Robert McCulloch (2005) Bayesian Statistics and Marketing, John Wiley & Sons.
- [19] Stone Bob, (1995), Successful Direct Marketing Methods, pp37-59, Lincolnwood, IL: NTC Business Books.

