

國立臺灣大學管理學院國際企業學研究所



碩士論文

Graduate Institute of International Business

College of Management

National Taiwan University

Master Thesis

以信用卡資料庫建構信貸預測模型

Constructing Credit Loan Forecasting Model

By Using Credit Card Database

何俊億

Chun-Yi Ho

指導教授：任立中 博士

Advisor: Li-Chung Jen, Ph.D.

中華民國 102 年 6 月

June, 2013

# 口試委員會審定書



## 誌 謝



碩一結束時決定休學先服兵役，回到學校重拾課業難免生疏，但抱著當初考上國企所的那份熱情，繼續把所剩的學分修畢，感謝 99 級很多同學在我這一年的協助，其中一起先去服役的同梯同學宗暉、永寰，謝謝你們一起為碩二這年的課業互相打氣。時光飛逝，晃眼間碩士班的日子也告一段落，撰寫碩士論文的這半年，感謝任立中教授的殷勤指導，每一句話語如沐春風，讓我獲益匪淺，特別是每次與老師接觸的機會中，學習許多作人處事的道理。此外，特別感謝邵功新學長每周辛勤的指導，帶領著學生撰寫論文的方向。當然還有任大師桃李門的同學宜靜、俐婷、郁婷和書琴，能和你們一起學習、努力是最棒的回憶。也謝謝小潔這些日子以來的鼓勵與陪伴。最後，我想將這份論文和與畢業證書獻給我的家人，沒有你們的支持，就沒有今天的我，我愛你們，謝謝。

何俊億 謹誌

民國一〇二年六月

## 中文摘要



企業與顧客建立關係後，企業遂能依據資料庫記錄顧客相關的行為，透過資料庫行銷的方式與顧客互動，挖掘顧客潛在的購物需求，提高企業的獲利空間。而本研究以國內某銀行信用卡資料庫為研究範圍，資料類型包含顧客的人口統計變數與刷卡交易記錄等，試圖從顧客過去的交易情況進行價值分析，並先假設影響顧客出現資金需求的變數是否對於顧客辦理信用貸款具顯著影響力。

因此，本研究的目的是在於找出顧客出現資金需求的原因，並配合研究主題與模型規劃的適切性，首先將每一位信用卡使用者視為可能出現資金需求的潛在顧客，實證上的作法為以 2012 年 5 月為基準，挑選出實際辦理信貸與否的兩組人員如基本組與信貸組，根據信用卡顧客的交易行為模式與假設的顯著變數，建構機率預測模型判讀顧客辦理信用貸款的可能性。對於銀行業者而言，實質的意義在於當顧客出現資金需求的訊號時，能搶先其他同業以正確的時間點進行行銷活動的推廣，增加目標客群來本行辦理信貸的機會。在建構機率預測模型的選擇上，本研究採用 Logistic 和 Probit 迴歸模式進行配適，並比較兩者的預測能力高低。

關鍵字：資料庫行銷、層級貝式統計、Logistic 迴歸、Probit 迴歸

## Abstract



The main purpose of this study is to construct credit loan forecasting model based on the database of customer transaction history, it could help banks to execute timely marketing approach to those who are highly demanding of credit loan. Through interaction with customers, we can find out the target customers to launch marketing mix to improve profit margins. And the types of data contain customer demographic variables and credit card transaction records. We analyze customer value through two indices by past transactions and assume different customer transaction behavioral variables to fit the model to see which variable does really have a significant influence of credit loan applying.

Therefore, the purpose of this study is to identify the potential reasons why customer would apply credit loan. In the choice of constructing probability forecasting model, we use logistic and probit model and thus compare the credit loan hit rate of both.

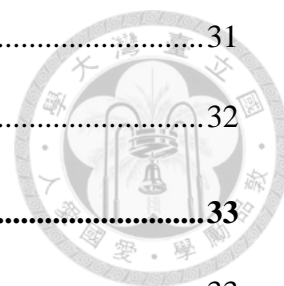
**Key words:** Database Marketing, Hierarchical Bayesian, Logistic Model, Probit Model

# 目 錄



口試委員會審定書 .....	I
誌 謝.....	II
中文摘要.....	III
Abstract.....	IV
圖 次.....	VII
表 次.....	VIII
<b>第一章 緒論.....</b>	<b>1</b>
第一節 研究背景與動機.....	1
第二節 研究目的.....	3
第三節 研究問題與範圍.....	4
第四節 論文架構.....	5
<b>第二章 文獻探討 .....</b>	<b>6</b>
第一節 行銷觀念的演進.....	6
第二節 顧客關係管理.....	11
第三節 資料庫行銷.....	17
<b>第三章 研究方法 .....</b>	<b>22</b>
第一節 最大概似估計法.....	22
第二節 加權最大概似估計法.....	24
第三節 層級貝氏統計模式.....	26

第四節	Logistic 迴歸模式.....	31
第五節	Probit 迴歸模式.....	32
<b>第四章</b>	<b>研究結果.....</b>	<b>33</b>
第一節	資料簡介.....	33
第二節	變數假設.....	40
第三節	顧客交易行為分析.....	43
第四節	顧客活躍性及穩定性分析.....	47
第五節	迴歸模型配適與預測.....	51
<b>第五章</b>	<b>結論與建議.....</b>	<b>66</b>
第一節	研究結論與行銷意涵.....	66
第二節	研究限制.....	69
第三節	後續研究建議.....	70
<b>參考文獻.....</b>		<b>71</b>
一、中文部分.....		71
二、英文部分.....		72



# 圖 次



圖 1-1 論文架構.....	5
圖 2-1 顧客關係管理架構.....	12
圖 4-1 基本組顧客每月平均消費金額與頻率散佈圖.....	34
圖 4-2 信貸組顧客每月平均消費金額與頻率散佈圖.....	35
圖 4-3 基本組與信貸組—平均購買期間 MLE.....	44
圖 4-4 基本組與信貸組—平均購買期間 WMLE.....	44
圖 4-5 基本組與信貸組—平均購買期間 HBE.....	45
圖 4-6 模型推力績效比較.....	65



# 表 次

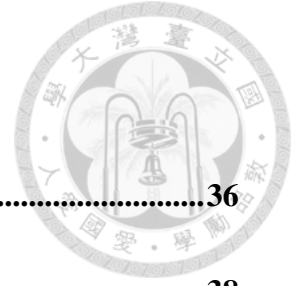


表 4-1 基本組與信貸組每月消費金額、頻率敘述統計量 .....	36
表 4-2 人口統計變數水準定義 .....	38
表 4-3 顧客人口統計結構分布 .....	39
表 4-4 最大概似估計法與層級貝氏估計法之比較 .....	46
表 4-5 顧客活躍性指標範例 .....	47
表 4-6 基本組與信貸組活躍性分析 .....	48
表 4-7 顧客活躍性指標範例 .....	50
表 4-8 基本組與信貸組活躍性分析 .....	50
表 4-9 人口統計變數定義 .....	52
表 4-10 交易資料變數定義 .....	53
表 4-11 Logistic 迴歸式係數與顯著性 .....	57
表 4-12 Logistic 模式擊中率 .....	58
表 4-13 Logistic 模式樣本外預測擊中率 .....	59
表 4-14 Probit 迴歸式係數與顯著性 .....	61
表 4-15 Probit 模式擊中率 .....	62
表 4-16 Probit 模式樣本外預測擊中率 .....	62
表 4-17 Logistic 和 Probit 模式推力圖比較 .....	64

表 5-1 顯著變數重要性列表 .....67



# 第一章 緒論



## 第一節 研究背景與動機

以往的商品銷售模式，買賣雙方的關係僅止於交易完成之時間點，隨著近年行銷概念的發展，已經從大眾行銷演化成關係行銷，針對分眾族群給予不同的行銷方案。然而，隨著時間的推演，消費者的購物不僅只是購買商品而已，亦有其他附加價值的考量，因此企業不得不針對消費者的「動態性」

(Dynamic) 與「異質性」(Heterogeneity) 適時修正行銷策略，而又隨著資料儲存的成本降低，企業可以維護顧客交易的行為記錄，又發展至資料庫行銷的階段，對此不僅能針對顧客進行消費者購買行為之分析，制定相對應的行銷策略，提高企業的獲利性。

談到台灣信用卡的業務自 1974 年推行以來，改變大眾的支付習慣，有了先購物後付款的概念，而銀行給予信用卡客戶的額度根據辦卡人的年收入、信用情況而有高低之分，而信用卡除了簽帳的功能之外，當持卡人出現資金需求時還可向銀行辦理信用貸款。對於台灣的銀行而言，信用卡是經營消費金融市場重要的業務，也是主要的營收來源，包含客戶簽帳金額、預借現金、循環信用餘額等。如今面臨國內太多銀行業務重疊且趨於飽和的今日，競爭的激烈程度不在話下。

因此，銀行除積極推高發卡量之外，若能有效找出顧客潛在的商機，必能增加獲利性。而各家銀行擁有顧客龐大的交易資料，依資料庫行銷的概念，如果可以透過分析顧客交易記錄的方式，找出符合銀行想要鎖定的族群，給予適時的訊息傳達，例如針對出現資金需求的客戶給予適時的廣告宣傳，對銀行而

言，確實可以對特定族群行銷，減少相關成本，增加目標族群來本行辦理信用貸款的擊中率。

由 80/20 法則，當運用在企業獲利的觀點上，公司 80% 的利潤由 20% 的顧客所創造，因此當行銷預算有所限制下，如何找出價值最高的前 20% 的顧客，就成了影響企業獲利高低的關鍵因素。因此，本研究針對銀行提供的信用卡資料庫，進行相關顧客交易行為的分析，並給予一些影響信貸的變數假設，建構辦理信用貸款的機率預測模型，期待透過模型的建立，提升銀行在推展信用貸款的業務時，能夠有方向地針對目標族群，給予適時的廣告或宣傳，達到顧客來行申辦信貸的可能性。

## 第二節 研究目的

本研究的目的是在於以資料庫行銷的概念，針對信用卡資料庫，從客戶過去的交易記錄中，找出可能出現資金需求的潛在客群，主要核心概念在於從辦理信用貸款的客戶中，分析其過去交易記錄的模式，比較與尚未來行辦理信貸客戶的差異，建構信貸的機率預測模型，如此一來，能夠將顧客的購買行為，投入該模型中試算，找出出現資金需求訊號的族群，實質的意義是，對銀行來說可以針對該族群進行行銷活動的設計與推廣，增加客戶來本行辦理信貸的可能性。

如同前述的動機與目的，本研究試圖從實證研究的角度切入，達到以下幾個分析目標：

1. 針對資料庫中的顧客購買行為，分析每月消費金額與頻率，以不同估計法估算其平均購買期間，建構活躍性與穩定性指標，分析顧客過去的交易習慣。
2. 給予假設，試圖找出影響顧客辦理信用貸款的潛在變數，加入本研究所採用的兩個模型中，計算擊中率，並將樣本外的資料帶入模型中，觀察模型的預測力高低。
3. 比較兩模型在預測力上的差異，給予銀行人員鎖定潛在信貸客群的方向，依不同的行銷預算可針對顧客進行資訊傳遞行銷活動推廣。

## 第三節 研究問題與範圍



銀行提供的顧客資料可分為三大類型，如下所示。

### 一、人口統計資料

記錄來行信用卡客戶的基本資料，包含性別、年齡、居住地、婚姻狀況、職業別等。

### 二、交易行為資料

記錄信用卡客戶過去半年期的交易明細資料，包含交易日期、交易明細及交易金額，並有一些其餘相關資料，如顧客與銀行往來年資、來電客服數、持有信用卡張數等資料。

### 三、個人信用資料

由聯合徵信中心（JCIC）提供的顧客資料，包含顧客過去六個月往來銀行、每月簽帳金額、預借現金餘額、動用循環信用餘額以及該月應繳金額等相關資料。

## 第四節 論文架構

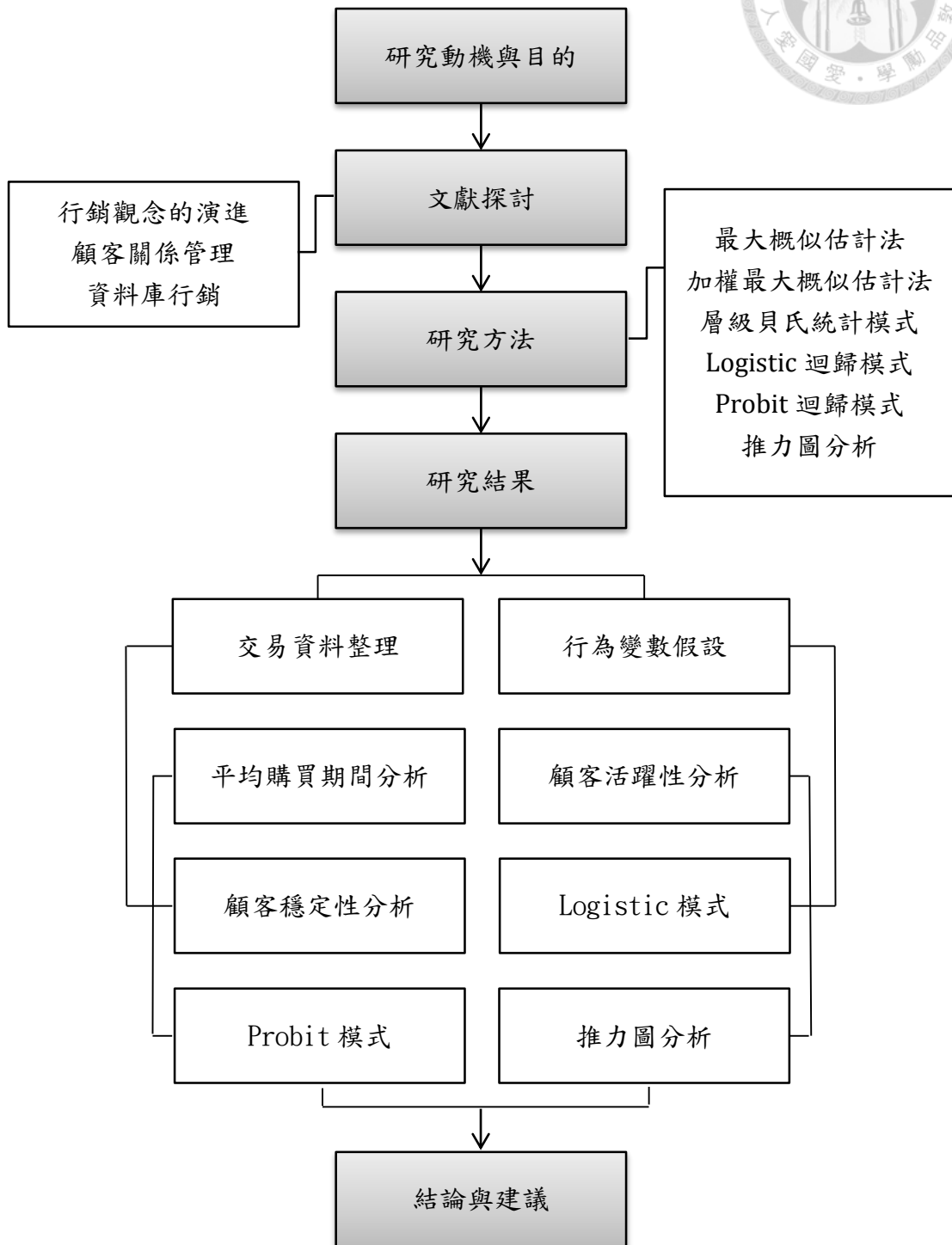


圖 1-1 論文架構

## 第二章 文獻探討



### 第一節 行銷觀念的演進

#### 一、市集交易時期

傳統上買賣雙方的交易，無論是以物易物或者以貨幣為交易媒介的型態，行銷的發展演進，最早以生產與銷售導向為主，企業將製成後的產品透過各種手法向消費者傳達產品價值，進而產生購買行為，這種最基本的銷售模式可認為行銷的鼻祖，亦即買賣雙方產生財貨與價值交換的過程。後來，這種交易模式出現了「中介商」的角色（Webster，1992），賣方會透過任何可以吸引或促使消費者產生購買行為的方式，盡可能的包裝產品價值。

行銷大師 Philip Kotler 對行銷的定義為，行銷是一種管理性的過程，個人或群體可以經由此過程，透過創造及交換產品的價值，滿足雙方的需求與慾望。因此可得知，行銷即是買賣雙方之間勞務交換的過程，當此過程愈趨複雜時，行銷活動就需要更多的溝通方式與消費者接觸互動，行銷管理（Marketing Management）便成了一門學科。

美國行銷學會於 1985 年對行銷管理的定義，表示行銷是一種分析、規劃與執行的一連串過程，藉由制定創新、創意的服務與配套的訂價、促銷活動，進而創造滿足個人和組織目標的交換活動。而行銷的重要性漸漸地被企業所接受，遂而發展出不同觀點的理論。





## 二、大量行銷時期 (Mass Marketing)

工業革命時期，企業投入大量的資本於工廠與機器設備，機械式的標準化生產取代傳統勞力，使得生產效率倍增，此時對於消費者的前提假設是同質性高的，因此產量愈多就能滿足愈多的消費者、提高銷售額，此時的行銷手法因應大量生產的原因，強調大量行銷策略。此外，鐵路與公路網絡趨於成熟，代表著商品通路配送的效率達到高峰，也大幅降低大量產品運輸的成本。

學者榮泰生 (1992) 表示，大量行銷是一種典型的生產導向策略，以相同的行銷組合 (4Ps) 面對市場上每一位消費者作銷售訴求，在此階段，行銷人員擬定行銷組合的依據大多以人口統計變數 (demographic variables) 為主，找出一種能夠涵蓋到最多消費者且獲利最大的 STP (Segmenting Targeting Positioning)，進行生產、研發及銷售等相關的企業活動。

## 三、個體行銷時期 (Micromarketing)

二十世紀初期，科技的快速進步，人們的生活方式更為多元，市場上消費者的需求與慾望出現更多不同的聲音，市場分割愈來愈細。在大量行銷時期，企業以規模經濟的產出模式創造競爭優勢，但隨著消費者的需求的改變，使得企業逐漸無法以一套單一的行銷手法滿足這群異質的顧客。因此 Hapoienu (1990) 提出個體行銷 (Micromarketing) 的概念，企業若想將商品銷售給需求異質的顧客，勢必要與消費者進行互動接觸，挖掘顧客需求，進而提供滿足需求的產品。

透過企業與消費者接觸的機會，包括問卷調查、電話訪問或樣本測試等活動來了解顧客真正的需求，並針對不同分群的顧客提供產品或服務的行銷組合

策略。Kotler (1995) 說明，個體行銷所強調的是在異質性的假設基礎上，專注於理解顧客的需求與慾望，進行行銷策略的擬定。



根據以上對於個體行銷的說明，我們可有兩種方式來區隔市場。第一種為 Top-down 的觀點，由企業（供給面，supply）的立場來區隔市場，即以銷售者的主觀認定消費者特性的區分，或依據成本因素做市場區分，完成市場劃分後，再將同一區隔的消費者視為同質，接著擬定行銷策略。然而，被劃分後屬於同一群的消費者是否需求一致，銷售者無法準確的得知，因此 George S. Day (1981) 提出另一種反向的市場區隔觀念—Bottom-up，此種觀點以顧客（需求面，demand）的立場來觀察消費者需求與異質性，再將相同需求的顧客歸類為同一群，由下而上分割市場，達到顧及顧客異質需求的市場區隔方法。

#### 四、直效行銷（Direct Marketing）

Kotler (1994) 指出，以往直效行銷的定義不斷改變，最初的意義指的是不透過中介商的配銷系統，直接將產品或服務從生產者移轉到消費者手中的一種行銷方式。依此定義，銷售人員將產品直接販售給顧客，或是生產者兼營銷售據點，都可視為直效行銷的一環，而後演變成的型錄行銷或郵購行銷皆屬於直效行銷的範疇。

美國的直效行銷協會（Direct Marketing Association; DMA）(1994) 提出，直效行銷是一種互動式的行銷系統，利用一種或多種的廣告媒體，對消費者產生產品的價值傳達，藉以獲得顧客反應的訊息。此種反應通常指顧客接收廣告訊息後下訂單訂購的行為，故直接行銷也可稱為直接訂購行銷（Direct-order Marketing）(Rapp & Collins, 1987)。

洪順慶（1999）則認為，直效行銷透過各種媒體，如電視、網路與 DM 等方式直接和消費者接觸，其特色如下所示：



（一）廣告與行動合一：直效行銷將廣告與行動結合在一起，誘發一個即時的反應。

（二）特定性：意指使用媒體傳達訊息給篩選的特定族群，此非一般性的廣告，通常針對不特定的個人。

（三）回饋：直效行銷所使用的工具可使企業或銷售員直接客觀衡量廣告效果，也可藉由顧客基本資料追蹤反應來源。近年來由於資訊科技進步，也使得直效行銷的回饋能力更加強化。

## 五、關係行銷（Relationship Marketing）

在過去的大量行銷年代，企業以量為前提生產，少與消費者進行產品上的溝通，獲利空間來自於生產成本的減少與大量的產品銷售，企業與顧客之間並不存在買賣關係的維護，雖然到了個體行銷時期，企業開始與顧客互動，但是僅止於了解的過程，尚不存在關係維護的概念。

後來產生一對一行銷（One by one Marketing）的概念，建議企業在過去忽略顧客關係的觀念下，進可能修正行銷手法，利用提供顧客個人化服務的行銷方式，創造競爭優勢，這可算是關係行銷概念的起頭。Jackson（1985）認為，關係行銷是將行銷定位為強化並維持企業與個別客戶的關係，也有學者認為關係行銷的目的在於，藉由過去與顧客發展出來的信任關係來贏得產品銷售的支持。

Berry & Parasuraman (1991) 則認為，關係行銷關心如何吸引、發展並保留與顧客的良好關係。此外，國內學者任立中 (2005) 則提出，關係行銷在 B2C 的互動模式中，企業對消費者應該發展出一對一、個人化的行銷策略。因此我們可以知道，關係行銷的發展目的是為了建立、維持與顧客關係的相關行銷活動，至於關係的維護，倚靠資料庫系統的建立，可記錄顧客的基本資料與交易明細，透過分析消費者行為模式，企業可以找出顧客偏好、消費習慣，進而維持關係，創造未來的收益來源。

## 第二節 顧客關係管理



### 一、顧客關係管理的定義

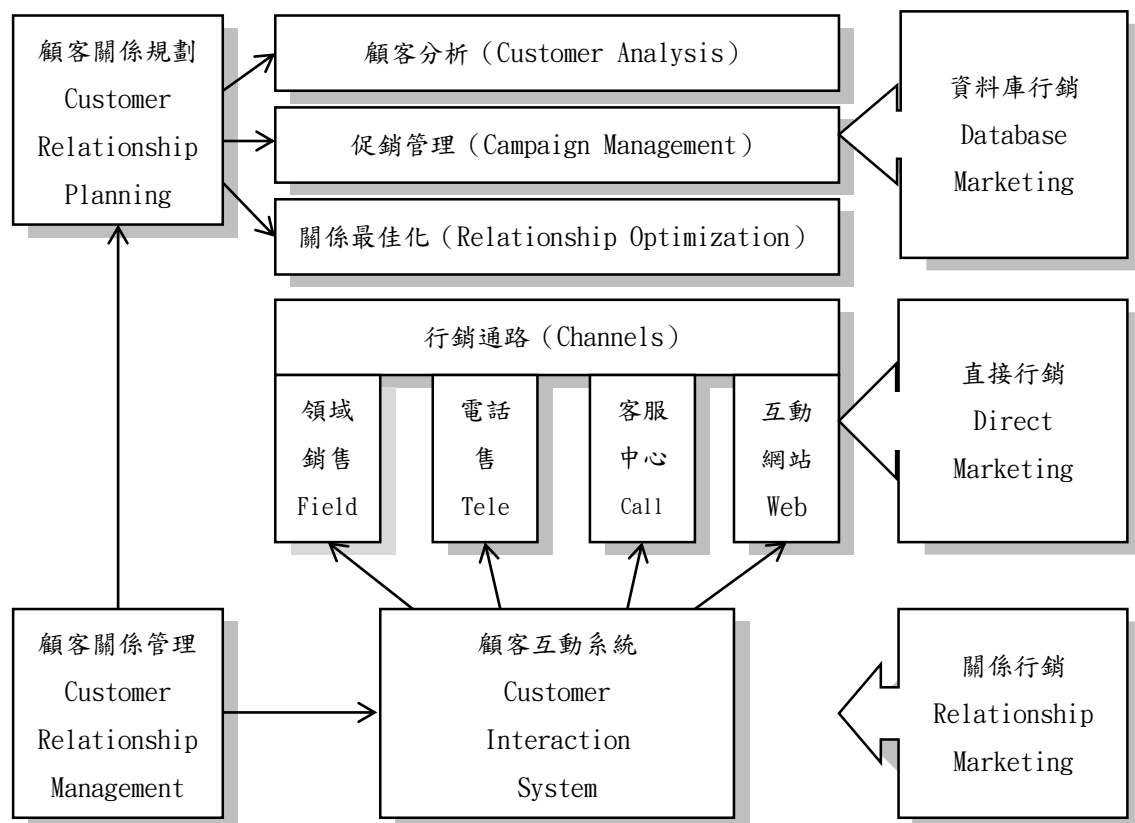
行銷概念的演進從生產銷售導向到顧客關係導向，企業愈來愈重視消費者需求，如何使與顧客互動、了解需求、刺激需求乃至於提升忠誠度，是企業長期獲利的必要條件，學術上顧客關係管理（Customer Relationship Management, CRM）的定義諸多，Pepper & Rogers（1999）說明，顧客關係管理是為了與客戶建立關係後，藉此獲取顧客的資訊，並進而需求。

陳文華（1999）對顧客關係管理的定義如下：指出目標為提供顧客優良的服務品質，能夠更有效率獲取、開發並保留企業最重要的資產－顧客。顧客關係管理是多種資訊的結合，是一個資料蒐集、分析、獲取資訊的管道，並且不斷重複、持續地找出更有效率的方式行銷企業的產品或服務。

根據 Meryl Davids（1999）指出，顧客關係管理即所謂的關係行銷、終身價值行銷、忠誠度行銷和一對一行銷。這些策略能使企業創造出與客戶之間的長期互利關係，並發展忠誠關係和利潤。R.S. Swift（2001）說明，顧客關係管理是一種企業透過和顧客溝通、互動，進而了解、影響顧客的行為，作為增加顧客、提高忠誠度和獲利的手段。IBM Inc.（2012）則提出，顧客關係管理（CRM）是一個包含科技應用、流程及組織變革，並以顧客為核心的企業行銷策略。CRM 結合了銷售、行銷與顧客服務活動，目的在於實現營收、獲利及顧客滿意度的最佳化。

## 二、顧客關係管理的架構

由於資訊科技的快速發展，1980 年代發展出來的接觸管理進一步帶出了顧客關係管理概念，無論是直接行銷（Direct Marketing）、關係行銷（Relationship Marketing）或者資料庫行銷（Database Marketing）都可算是管理的一部份，根據 ABC 遠擎管理顧問公司之研究部門總經理 Brian Spengler（1993）所提出的顧客關係管理架構，將上述三種行銷模式整合起來，可由圖 2-1 中進一步了解顧客關係管理和三種行銷模式的關係。



資料來源：電子化企業－經理人報告（1999）

圖 2-1 顧客關係管理架構

### 三、顧客關係管理的功能



了解顧客關係管理的定義之後，我們可以得知 CRM 對於企業的重要性不可言喻，特別是需要龐大的整合程序，包含公司內部的五管部門整合以及外部供應商，都需要同步導入 CRM 的概念與流程。以下我們將針對企業實施顧客關係管理後，具體獲得的利益內容進行詳細說明。

國內學者陳文華（1999）指出，企業實施顧客關係管理的益處可分為以下幾點：（引自廖怡芳，民 93）

- （一）增加收益：對於高價值的顧客進行銷售管理，提高購買商品或服務的績效。
- （二）增加獲利率：藉由了解潛在顧客需求，實施特別的行銷方案增加和顧客互動性。
- （三）降低成本：對於不同族群實施分眾行銷的策略，將資源有效運用在對的顧客身上，避免資源無謂了浪費。
- （四）提高市場專注性：企業藉由與顧客之間的互動與了解，可以提供行銷人員制定行銷決策的參考，以便提高對市場的專注性。
- （五）減少新銷售活動執行的循環次數：企業得以研究如何與顧客進行有效溝通，增進彼此的了解程度，減少行銷活動執行次數，如此一來便可降低試誤的頻率。

(六) 增加小型目標行銷的次數：導因於對顧客的互動與了解，將顧客區隔成各個具有代表性的市場，可針對不同的市場提供不同的產品或行銷手法，成功銷售的機會也會較高。



(七) 增加知識：企業的行銷人員可藉由資料庫的大量訊息，了解設計的行銷活動其執行結果與成效，增加對於消費者行為的知識。

國外學者 Swift (2001) 說明，企業實施顧客關係管理能帶來競爭優勢，主因如下所示：

(一) 正確的顧客：管理顧客整個生命週期，藉由增加荷包占有率 (Share of Wallet) 實現顧客潛能。

(二) 正確的建議：根據不同的顧客給予客製化建議，並提供有效的產品或服務給潛在客戶。

(三) 正確的通路：協調每位客戶之間的溝通，並學習透過合適的通路溝通，擷取和分析通路資訊。

(四) 正確的時機：能在對時間找到對的顧客進行有效溝通，亦即有能力以即時性的行銷方式和顧客互動。



#### 四、顧客關係管理的執行方法



Peppers, Rogers, and Dorf (1999) 提出創造良好顧客關係的四個階段，以「IDIC」流程說明：

- (一) 確認 (Identify)：顧客關係管理的第一步是找出有價值的顧客，掌握顧客是誰方能發展一對一關係。
- (二) 區隔 (Differentiate)：重視顧客之間的差異性，進一步分析顧客價值且區隔不同族群，可分為以下三種類型：
  1. 最有價值顧客 (Most Valuable Customers, MVCs)：最有價值顧客是企業核心的獲利來源，更應該關心如何留住這群顧客。
  2. 最具成長潛力顧客 (Most Growable Customers, MGCs)：指的是當企業未來發展的策略或擴大業務範圍時能夠鎖定那些潛在的客群。
  3. 毫無價值顧客 (Below Zeros, BXs)：指的是企業現今或未來無論再怎麼努力，始終都無法從顧客身上賺取收益的群體，此時就應直接放棄，避免成本的無謂浪費。
- (三) 互動 (Interact)：促進與顧客之間的互動，改善溝通的成本效益，並可透過自動化程序降低成本，也使企業能更深入的與客戶對話，增加對顧客的理解。

(四) 客製化 (Customize)：配合不同顧客的專屬性需求，可提供不同的服務、調整的產品內容以回應之，進而提高顧客滿意度。

因此透過以上「IDIC」的說明，可做為企業建構顧客關係管理的基礎方法，透過這種結構式的標準流程，能讓企業依據本身競爭優勢擬定未來策略藍圖，增加獲利空間。

陳文華 (1999) 指出顧客關係管理可以透過不斷重複、改善的過程，達到最好的行銷效果，於此可將此循環過程分為三大階段，如下所示：

(一) 評估 (Assess)：此階段整和企業內、外部的資源，並針對目標客群進行相關的消費者偏好分析，在此階段一開始，必須先釐清顧客類型、顧客屬性以及顧客價值，並且計算可能的顧客反應。

(二) 規劃 (Plan)：由行銷人員針對分群的顧客進行不同行銷活動的設計，目的在於能夠準確地打入市場，吸引顧客上門來消費。

(三) 執行 (Execute)：確認企業鎖定的目標客群之後，確實執行不同市場的行銷活動，並適時針對顧客訴求改善產品或服務的流程，達到較佳的顧客關係管理。

### 第三節 資料庫行銷



#### 一、資料庫行銷的定義

有別於傳統上的行銷活動以交易為核心基礎，資料庫行銷是以顧客為核心發展的行銷策略。Shaw & Stone (1990) 定義資料庫行銷是「互動式行銷的一種溝通方式，相關它能針對特定族群的顧客藉由媒介進行互動，持續與顧客接觸刺激其需求，並將顧客資料、交易行為等相關資訊儲存於資料庫中，以便未來能與顧客保持密切接觸，助於顧客關係的維護」。

Shani & Chalasani (1992) 認為，資料庫行銷收集過去、現在以及潛在客群的資料，建立資料庫以改善行銷策略擬定的績效，而資料庫的內容包含人口統計資料、顧客交易偏好、顧客購買行為以及顧客生活型態資料等。Cespedes & Smith (1993) 認為，資料庫行銷是利用顧客的資料來改善企業行銷資源的配置與應用，針對特定族群給予有用的資訊傳遞，維持顧客與企業之間的良好關係。Hughes (1994) 認為，以電腦科技來管理一套記錄現有顧客及潛在顧客相關訊息的資料庫系統，使企業得以利用這些資料提供較佳的產品或服務，與顧客建立互動式的長期關係。

國內學者任立中 (1998) 則認為，資料庫行銷的重心可分為二，其一為互動，其二為異質性。首先與顧客互動才能建立溝通的橋梁，接著針對異質性的顧客給予不同行銷策略的調適，方能發揮資料庫行銷之概念。

## 二、資料庫行銷的功能



美國資料庫行銷中心（National Center for Database Marketing）認為資料庫行銷應有下列功能（葉德昌，1994 譯）：

- （一）確認出最容易打動的顧客及潛在顧客。
- （二）與常客建立長期、高品質的良好關係。
- （三）建立先期模型方能：
  - 於適當時機以適切的方式將訊息傳遞給顧客
  - 有效贏得顧客歡心
  - 讓行銷支出更具效益
  - 建立品牌忠誠度
  - 增加獲利性

Stone & Woodcock（1995）指出，從資料庫行銷接觸個別顧客的特性，使其能夠達到下列行銷功能：（引自李宗龍，民 90）

- （一）更精確的瞄準產品的行銷與風險。
- （二）確保顧客的忠誠度，避免競爭的風險。
- （三）確保最有可能購買新產品與服務的客戶。

(四) 提高銷售效率。

(五) 為傳統的銷售方式提供低成本的新方案。

(六) 促使行銷功能的結果更容易量化。

(七) 改善產品管理、銷售通路、廣告與促銷活動，提供更好的連結。

(八) 能夠在任何時點下，提供客戶相關的資訊，因此能改進顧客服務品。

(九) 協同行銷過程中影響顧客的各種因素，達成完全的關係行銷。

根據林慧晶（1997）的研究，資料庫行銷的功能可分為以下四個部份：

（引自蔡欣靜，民 93 年）

(一) 顧客價值分析：資料庫行銷最主要的功能在於進行顧客價值的分析，傳統上企業可以清楚知道每日銷售額，但卻很難能將個別顧客與銷售情況作連結，因此透過資料庫行銷的分析，企業可以很容易地針對顧客進行價值分析，並對不同價值顧客採取不同的行銷策略。

(二) 計算顧客終身價值：根據 Arthur Hughes（1994）的說明，顧客終身價值（Customer Lifetime Value）是指在未來一段時間內，企業可以從每一位顧客身上獲得的利潤淨現值。藉由資料庫行銷，企業可依據資料庫中儲存的顧客交易記錄，計算出個別顧客對企業可能貢獻的終身價值，實質上來說，企業可以預測未來營收情況，亦可針對高價值顧客給予較多的行銷預算。



(三) 進行向上銷售 (Up-selling) 與交叉銷售 (Cross-selling)：根據 Bob Stone (1995) 所言，向上銷售是指企業可以針對顧客目前購買的產品項目，推測往後可能會需要的品項；交叉銷售則是指針對顧客目前購買的產品項目，進行相關產品的銷售服務。因此，針對資料庫中顧客購買，可以達到向上銷售和交叉銷售的目的。

(四) 行銷決策支援系統：Little (1979) 認為，所謂的行銷決策支援系統 (Marketing Decision Support System, MDSS) 指的是顧客的購買記錄透過模型分析，接者分析出的結果專家知識，能企業能作出有利的決策。因此，顧客資料加上模型分析是資料庫行銷最重要的兩大元素，其功能不僅幫助企業管理顧客，更能作為企業未來決策的支援系統。

### 三、資料庫行銷的執行方法

任立中（1994）曾說明資料庫行銷的關鍵在於不同專家組成的團隊，包括電腦專家、統計專家、策略專家與銷售專家，倚靠四種專家彼此溝通，進行跨功能的合作，方能完整執行資料庫行銷的內涵。

四種專家的功能環環相扣，首先銷售專家將商品或服務銷售情形與策略專家討論，經由策略專家擬定出行銷策略後，與統計專家說明，並針對行銷策略的需求找出分析資料所需的統計模式，再與電腦專家討論溝通，進行電腦系統與統計分析相關的規劃。接著進行反向的溝通，統計專家以電腦專家規劃的分析系統進行資料數據分析，並將結果呈現給策略專家，策略專家便能依據分析結果進行行銷策略的規劃與修正，並將執行方案交給銷售專家，進行市場上的銷售活動。因此，這一連串的跨功能團隊運作模式，是資料庫行銷的關建成功因素。



## 第三章 研究方法



### 第一節 最大概似估計法

顧客購買期間發生的事件為連續型變數，且時間不存在負數，符合指數型分配的值域介於零到正無窮大之連續型分配的特性，此處我們假設顧客平均購買期間服從一個參數為  $\lambda_i$  的指數分配，而期望值與變異數皆為  $\lambda_i$ 。

因此，假設總共有  $N$  名顧客，第  $i$  名顧客 ( $i=1,2,\dots,N$ ) 在某一期間發生  $n_i$  次購買事件，且自第一次購買行為發生後，每次購買期間為  $t_{ij}$ ，其中  $j=1,2,\dots,n_i$ ，故第  $i$  名顧客的購買區間可由下式的機率密度函數表示：

$$f(t_{ij} | \lambda_i) = \lambda_i^{-1} \cdot \exp\left(\frac{-t_{ij}}{\lambda_i}\right), \quad i=1,2,\dots,N, \quad j=1,2,\dots,n_i$$

根據上式之機率密度函數，我們可以求出平均購買期間  $\lambda_i$  的最大概似估計函數，推導方式如下所述：

$$\ell(t_{ij} | \lambda_i) = \prod_{j=1}^{n_i} \lambda_i^{-1} \cdot \exp\left(\frac{-t_{ij}}{\lambda_i}\right) = \lambda_i^{-n_i} \cdot \exp\left(\frac{-\sum_{j=1}^{n_i} t_{ij}}{\lambda_i}\right)$$

接者對最大概似函數取對數，得到：

$$\ln(\ell(t_{ij} | \lambda_i)) = -n_i \ln(\lambda_i) - \frac{\sum_{j=1}^{n_i} t_{ij}}{\lambda_i}$$



此時再對  $\lambda_i$  取微分，並令其一階微分為 0，便可得到平均購買期間的最大概似估計量。



$$\frac{\partial \ln(\ell(t_{ij} | \lambda_i))}{\partial \lambda_i} = \frac{-n_i}{\lambda_i} + \lambda_i^{-2} \left( \sum_{j=1}^{n_i} t_{ij} \right) = 0$$

$$\Rightarrow -n_i \lambda_i + \sum_{j=1}^{n_i} t_{ij} = 0$$

$$\Rightarrow \hat{\lambda}_i = \frac{\sum_{j=1}^{n_i} t_{ij}}{n_i} \quad \text{其中 } i=1, 2, \dots, N$$

平均購買期間的最大概似估計量，就其公式上而言代表顧客所有購買期間的總平均數，在計算此估計量方式具有簡單、方便等優點，但卻可能未充分使用資料庫中之其他資訊，因此 Jen & Wang (1998) 曾指出，利用傳統的最大概似估計法所估出的平均購買期間，無法運用整體樣本之聯合資訊 (pool information) 來推估個別顧客之真實估計值，此時對於顧客購買行為之代表性稍嫌不足，因此實際估計效果並非是最準確的。

## 第二節 加權最大概似估計法



除上述介紹的最大概似估計法，加權最大概似估計法（Weighted Maximum Likelihood Estimation, WMLE）是針對顧客交易時點的遠近，設定不同的權重比例，因此，我們可將平均購買期間  $\lambda_i$  的最大概似函數改寫為下式：

$$\ell(t_{ij} | \lambda_i) = \prod_{j=1}^{n_i} \left[ \lambda_i^{-1} \cdot \exp\left(\frac{-t_{ij}}{\lambda_i}\right) \right]^{\sum_j} = \lambda_i^{-1} \cdot \exp\left(\frac{-\sum_{j=1}^{n_i} t_{ij}}{\lambda_i \sum_{j=1}^{n_i} j}\right)$$

接者對此概似函數取對數，得到：

$$\ln(\ell(t_{ij} | \lambda_i)) = -\ln(\lambda_i) - \frac{\sum_{j=1}^{n_i} t_{ij}}{\lambda_i \sum_{j=1}^{n_i} j}$$

此時再對  $\lambda_i$  取微分，並令其一階微分為 0，便可得到平均購買期間的加權最大概似估計量。

$$\frac{\partial \ln(\ell(t_{ij} | \lambda_i))}{\partial \lambda_i} = \frac{-1}{\lambda_i} + \lambda_i^{-2} \left( \frac{\sum_{j=1}^{n_i} j t_{ij}}{\sum_{j=1}^{n_i} j} \right) = 0$$

$$\Rightarrow \hat{\lambda}_i = \frac{\sum_{j=1}^{n_i} jt_{ij}}{\sum_{j=1}^{n_i} j} \quad \text{其中 } i = 1, 2, \dots, N$$



此處的加權最大概似估計量將顧客不同交易時間遠近的因素納入模式中  
運算，在解釋上對於顧客購買行為的預測較為合理，但仍具有前述 Jen &  
Wang 所說的缺點：未將跨樣本間的聯合資訊納入考慮。

### 第三節 層級貝氏統計模式



#### 一、貝氏統計量

假設顧客之平均購買期間服從參數為  $\lambda_i$  之指數分配 (Exponential Distribution)，且每位顧客之參數  $\lambda_i$  皆不相同。令第  $i$  位顧客在  $T$  時間內的購買次數為  $n_i$ ，則其購買期間為  $(t_{i1}, t_{i2}, \dots, t_{in_i})$ 。因此，第  $i$  位顧客之概似函數如下式所示：

$$\ell(t_{ij} | \lambda_i) = \prod_{j=1}^{n_i} \lambda_i^{-1} e^{-\frac{t_{ij}}{\lambda_i}} = \lambda_i^{-n_i} e^{-\frac{\sum t_{ij}}{\lambda_i}}$$

考慮到每位顧客的異質性之下，其平均購買期間的參數  $\lambda_i$  會服從參數為  $\alpha$  和  $\theta_i$  的反伽瑪分配 (Inverse Gamma Distribution)，如下式所示，且期望值為  $\frac{1}{(\alpha-1)\theta_i}$ ；變異數為  $\frac{1}{\theta_i^2(\alpha-1)^2(\alpha-2)}$ ，其中  $\alpha$  和  $\theta_i > 0$ 。

$$\pi(\lambda_i; \alpha, \theta_i) = \frac{1}{\Gamma(\alpha)\theta_i^\alpha \lambda_i^{\alpha+1}} e^{-\frac{\sum t_{ij}}{\lambda_i}}$$

此時令  $\theta_i = \exp(X_i'\beta)$ ，其中  $X_i$  為 0 或 1，故可將具有異質性的伽瑪分配表示為下式：

$$g(\lambda_i; \alpha, \exp(X_i'\beta)) = \frac{1}{\Gamma(\alpha)[\exp(X_i'\beta)]^\alpha \lambda_i^{\alpha+1}} e^{-\frac{1}{[\exp(X_i'\beta)]\lambda_i}}$$

## 二、導出貝式估計值



利用貝氏定理 (Bayesian Theorem) 可推導出貝氏估計量  $\lambda_i$ 、 $\beta_k$  及  $\alpha$  之後驗機率分配，以下分別詳述之：

(一) 推導  $\lambda_i$ ：

在已知  $t_{ij}$ ， $X_i$  和參數  $\alpha$ 、 $\beta$  的條件之下， $\lambda_i$  的後驗分配如下所示：

$$\begin{aligned}\lambda_i | t_{ij}, n_i, \alpha, X_i, \beta &\propto \left[ \prod_{j=1}^{n_i} \lambda_i^{-1} \cdot e^{-\frac{t_{ij}}{\lambda_i}} \right] \cdot \left[ \lambda_i^{-(\alpha+1)} \cdot e^{-\frac{1}{[\exp(X_i' \beta)] \lambda_i}} \right] \\ &\propto \lambda_i^{-(n_i + \alpha + 1)} e^{-\frac{\{\sum t_{ij} + [\exp(X_i' \beta)]\}}{\lambda_i}}\end{aligned}$$

因此， $\lambda_i$  可以從一個參數為  $A$ 、 $B$  的反向伽瑪分配導出，即

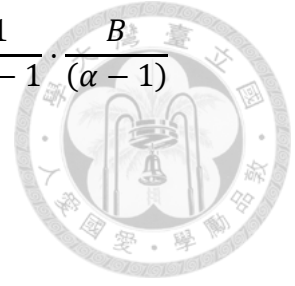
$\lambda_i \sim IG(A, B)$ 。其中：

$$A = n_i + \alpha$$

$$B = \left[ \sum_{i=1}^{n_i} t_{ij} + [\exp(X_i' \beta)]^{-1} \right]^{-1}, i = 1, 2, \dots, N$$

此時可導出  $\lambda_i$  的估計值，即為層級貝氏估計量 (Hierarchical Bayesian Estimator)，如下式所示：

$$\begin{aligned}\lambda_i &= w_1 \cdot \frac{\sum t_{ij}}{n_i} + w_2 \cdot \frac{B}{(\alpha-1)} = \frac{n_i}{n_i + \alpha - 1} \cdot \frac{\sum t_{ij}}{n_i} + \frac{\alpha - 1}{n_i + \alpha - 1} \cdot \frac{B}{(\alpha-1)} \\ &= \frac{\sum t_{ij}}{n_i + \alpha - 1} + \frac{B}{n_i + \alpha - 1} = \frac{\sum t_{ij} + B}{n_i + \alpha - 1}\end{aligned}$$



(二) 導出  $\beta_k$

在已知其他參數的條件之下， $\beta_k$  的後驗機率分配為：

$$[\beta_k | \lambda_i, \alpha, X_{ik}] \propto [\lambda_i | \alpha, \beta_k, X_{ik}] \cdot [\beta_k]$$

為了求出  $\beta_k$ ，必須先令  $e^{\beta_k} = \phi_k$ ，並導出  $\phi_k$  的後驗機率分配。其中，

$X_{ik} (k=1, 2, \dots, m)$  為虛擬變數 (Dummy Variable)。因為：

$$e^{X_i \beta} = e^{\sum_{k=1}^m X_{ik} \beta_k} = \prod_{k=1}^m [e^{\beta_k}]^{X_{ik}} = \prod_{k=1}^m \phi_k^{X_{ik}}$$

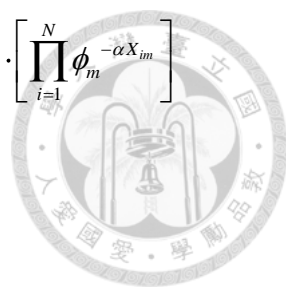
因此，上式可改寫成：

$$\pi(\lambda_i; \alpha, \phi_k, X_{ik}) = \frac{1}{\Gamma(\alpha) \left[ \prod_{k=1}^m \phi_k^{X_{ik}} \right]^\alpha} \cdot \frac{\left[ \prod_{k=1}^m \phi_k^{X_{ik}} \right]^{-1}}{\lambda_i^{\alpha+1}} \cdot e^{-\frac{\left[ \prod_{k=1}^m \phi_k^{X_{ik}} \right]^{-1}}{\lambda_i}}$$

其中獨立變數  $X_{ik}$  為虛擬變數，且定義如下：

$$\sum_{i=1}^N X_{ik} = N_k, \quad D_i^{(j)} = \prod_{\substack{k=1 \\ k \neq j}}^m \phi_k^{X_{ik}}$$

因此，以  $\phi_k$  來表示  $\beta_k$  的概似函數如下所示：



$$\begin{aligned}
\ell(\phi_k | X_{ik}, \alpha, \lambda_i) &\propto \left[ \prod_{i=1}^N \phi_1^{-\alpha X_{i1}} \right] \left[ \prod_{i=1}^N \phi_2^{-\alpha X_{i2}} \right] \dots \left[ \prod_{i=1}^N \phi_k^{-\alpha X_{ik}} \right] \dots \left[ \prod_{i=1}^N \phi_m^{-\alpha X_{im}} \right] \\
&\times \exp \left[ -\sum_{i=1}^N (\phi_1^{-X_{i1}} \phi_2^{-X_{i2}} \dots \phi_m^{-X_{im}} \lambda_i^{-1}) \right] \\
&\propto \phi_k^{-\alpha \sum_{i=1}^N X_{ik}} \times \exp \left[ -\phi_k^{-1} \sum_{i \in N_k} (\lambda_i D_i^{(k)})^{-1} + \sum_{i \notin N_k} (\lambda_i D_i^{(k)})^{-1} \right] \\
&\propto \phi_k^{-\alpha N_k} \exp \left[ -\phi_k^{-1} \sum_{i \in N_k} (\lambda_i D_i^{(k)})^{-1} \right]
\end{aligned}$$

假設  $\phi_k$  的先驗機率分配為反伽瑪分配  $(a_{k0}, b_{k0})$  時，即：

$$\phi_k \sim IG \left[ \alpha N_k + \alpha_{k0} \left( b_{k0}^{-1} + \sum_{i \in N_k} \frac{\lambda_i}{D_i^{(k)}} \right)^{-1} \right]^{-1}$$

則在已知其他參數的條件下， $\phi_k$  的後驗機率分配為：

$$\pi(\phi_k | X, \alpha, \lambda_i, \alpha_{k0}, b_{k0}) \propto \phi_k^{-\alpha N_k - \alpha_{k0} - 1} \times \exp \left[ -\phi_k^{-1} \left( \sum_{i \in N_k} \lambda_i D_i^{-(k)} + b_{k0}^{-1} \right) \right]$$

最後，將  $\phi_k$  轉變為  $\beta_k$ ，即  $\beta_k = \log(\phi_k)$ ，即可求出  $\beta_k$ 。

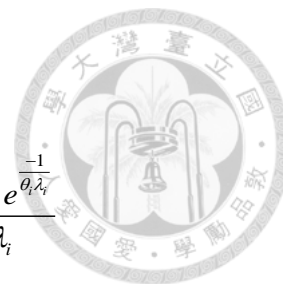
### (三) 導出 $\alpha$

在已知其他參數的條件之下， $\alpha$  的後驗機率分配為：

$$(\alpha | \lambda_i, \beta_k, X_{ik}) \propto (\lambda_i | \alpha, \theta_i)(\alpha)$$

根據反伽瑪分配，推出  $\alpha$  之概似函數如下式：

$$\ell(\alpha | \lambda_i, \theta_i) = \prod_{i=1}^N \frac{1}{\Gamma(\alpha) \theta_i^\alpha \lambda_i^{\alpha+1}} e^{-\frac{1}{\theta_i \lambda_i}} = \prod_{i=1}^N \frac{(1/\theta_i \lambda_i)^\alpha e^{-\frac{1}{\theta_i \lambda_i}}}{\Gamma(\alpha) \lambda_i}$$



令  $\alpha^* = \alpha - 1$ ，若  $\alpha^*$  的先驗機率分配為間斷行均勻分配 (Discrete Uniform Distribution)，則在已知  $\lambda_i$  和  $\theta_i$  之下， $\alpha^*$  的後驗機率分配為：

$$\prod_{i=1}^N \Pr \left[ \alpha^* = \alpha^* \mid \frac{1}{\theta_i \lambda_i} \right]$$

其中， $\Pr \left[ \alpha^* = \alpha^* \mid \frac{1}{\theta_i \lambda_i} \right]$  是參數為  $\frac{1}{\theta_i \lambda_i}$  的卜瓦松分配。



## 第四節 Logistic 迴歸模式



Logistic 迴歸模式類似線性迴歸模式，線性迴歸分析是描述一個反應變數 (Dependent variable) 與一個或多個解釋變數 (Independent variable) ( $X_1, X_2 \dots X_n$ ) 之間的關係，目的在於探討每一個解釋變數對反應變數的影響情形，而 Logistic 探討的反應變數是二元型 (Binary)，例如是與否、成功與失敗，因此可令  $\pi(x)$  表示某件事情發生的機率，且  $x$  為影響機率的因素，假若  $\pi(x)$  與  $x$  滿足下列式子：

$$\pi(x) = \frac{e^{f(x)}}{1 + e^{f(x)}} , 0 \leq \pi(x) \leq 1$$

則失敗的機率為：

$$1 - \pi(x) = \frac{1}{1 + e^{f(x)}}$$

而兩者的比值稱為勝算比 (Odds ratio)：

$$\frac{\pi(x)}{1 - \pi(x)} = e^{f(x)}$$

所謂的勝算比為二元事件發生的機率強度比，取對數後可得 Logistic 迴歸模式，如下所示。

$$\ln \frac{\pi(x)}{1 - \pi(x)} = f(x) = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_n X_{in}$$

## 第五節 Probit 迴歸模式



在統計模式中，Probit 是類似迴歸的一種，和前一節所述的 Logistic 迴歸一樣，主要都是處理依變數  $Y$  為二元的資料，模型設定如下所示：

$$\Pr(Y=1 | X) = \psi(X'\beta)$$

其中，Pr 為機率值 (Probability)， $\psi$  為累積分配函數 (Cumulative Distribution Function)，參數  $\beta$  為最大概似估計值 (Maximum Likelihood Estimator)，此外，Probit 模型設定如下：

$$Y^* = X'\beta + \varepsilon, \text{ 其中 } \varepsilon \sim N(0,1)$$

而  $Y$  則可被視為潛藏變數是否為正值的指標：

$$Y = 1_{(Y^* > 0)} = \begin{cases} 1, & \text{if } Y^* > 0 \\ 0, & \text{o/w} \end{cases}$$

結合上述公式，我們亦可寫成：

$$\Pr(Y = 1 | X) = \Pr(Y^* > 0) = \Pr(-\varepsilon < X'\beta) = \Phi(X'\beta)$$

類似於 Logistic 迴歸模式，本研究將會多採取 Probit 迴歸模式來進行分析，比較兩者間預測力的高低差異。

## 第四章 研究結果



### 第一節 資料簡介

本研究以國內某銀行（以下稱為本行）提供之信用卡資料庫為研究對象，樣本來源依據臺灣 2012 年內政部戶政司公佈之人口年齡分配中關於人口結構的資訊，本行依五個直轄市之人口性別和年齡比例，以配額抽樣原則抽取出 20,000 名顧客做為基本組，本行也提供行內信用卡持卡人於 2012 年曾經申辦信用貸款者（包括向行內、行外申辦者）共 12,000 名顧客，做為信貸組。

接著定義顧客的債務狀況，假設以 2012 年 5 月為基準，共計有 RR、RT、TR、TT 四種債務類型，R 和 T 字母的意義分別為：R 代表顧客截至本月尚有債務未償清，可能是信用卡循環信用餘額或預借現金餘額或信用貸款餘額；T 則代表顧客截至本月與銀行往來並無債務積欠。至於字母順序的意義為：第一個字母為截至本月顧客於本行債務狀況，第二個字母為截至本月顧客於其它銀行的債務狀況。因此舉例來說，某一位顧客於 5 月的債務標記為 RT，代表該名顧客截至五月底，於本行有未償清債務，於他行無債務積欠，依此類推。

因此，本研究以 2012 年 5 月為軸，從基本組與信貸組各自挑出 5 月份債務類型為 TT 者，其中再從信貸組 5 月份 TT 者中特別挑選 2012 年 6 月或 7 月於本行或他行申辦信用貸款者，這樣的目的是在於，觀察兩組 TT 類型的顧客截至 5 月底，在相同債務類型的比較基礎下（皆未積欠債務），找出信貸組顧客為何在 6 月或 7 月份會出現資金需求、向銀行申辦信用貸款的原因，即為本研究的最大目的，符合此篩選條件的基本組顧客人數 5,611 人，貸款組人數 3,309 人。接著我們先依照不同的資料類型如以下部分做說明，包含交易記錄資料、人口統計變數資料、及所有銀行往來記錄。

## 一、交易記錄資料整理



由本研究篩選出來之基本組與信貸組，以 2012 年 5 月為時間點，計算過去 6 個月的交易記錄資料，資料型態包含顧客消費日期、消費明細、消費產業別。此外，為方便後續的顧客購買區間價值分析，因此刪除交易記錄低於 4 筆的顧客，此時基本組人數為 5,407 名，信貸組為 3,150 名，共計 6 個月 412,351 筆交易記錄，兩族群每月平均消費金額及頻率散佈圖，如圖 4-1 及圖 4-2 所示。

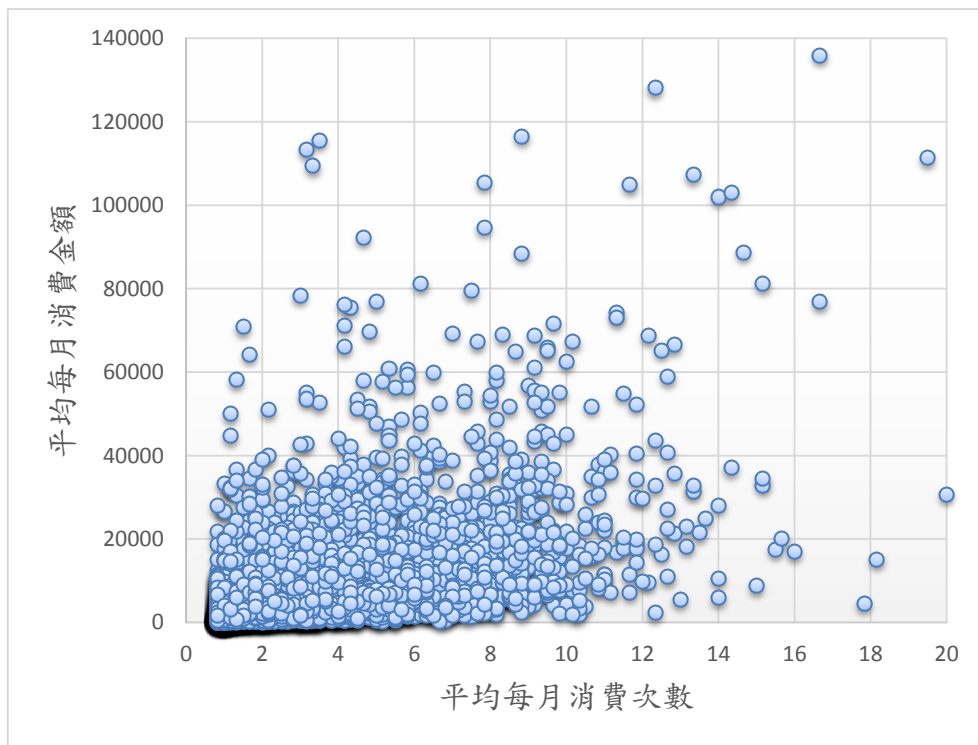


圖 4-1 基本組顧客每月平均消費金額與頻率散佈圖

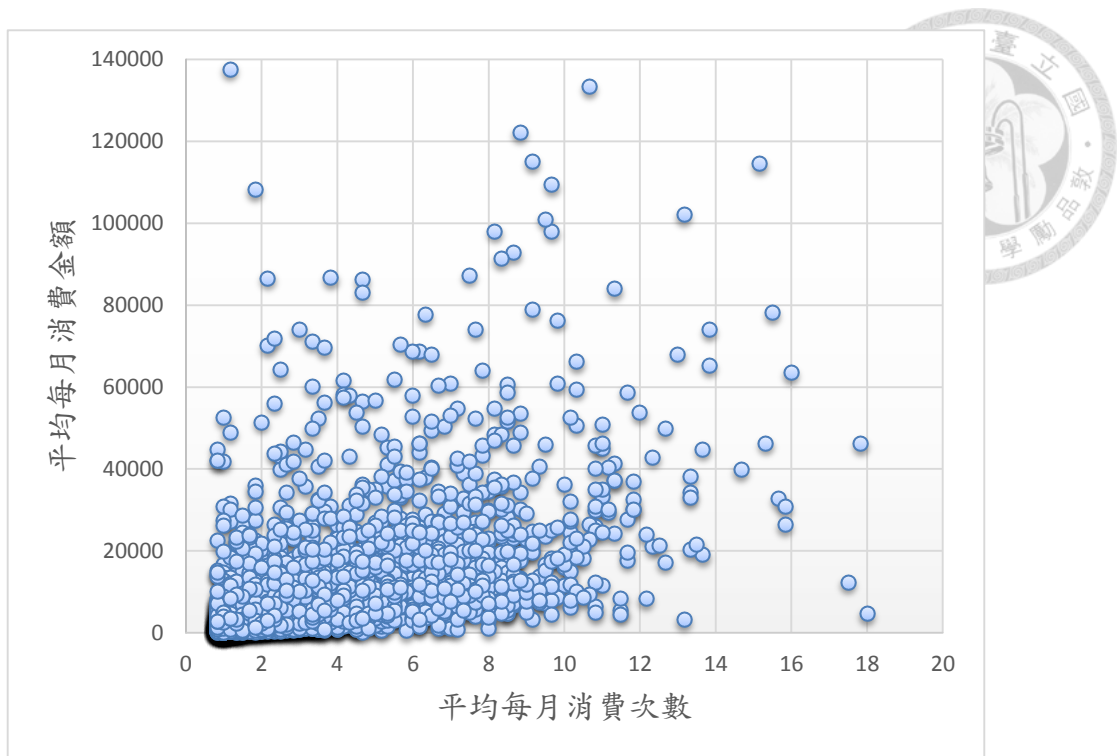


圖 4-2 信貸組顧客每月平均消費金額與頻率散佈圖

兩組顧客的基本敘述統計量如表 4-1 整理，顯示在信貸組當中，整體族群每月消費金額的每個分位數均高於基本組，其平均值也較基本組高出近 3,000 元，且消費金額的極端程度較高，最大值出現超過百萬的龐大刷卡金額；在每月消費頻率方面，兩組顧客並無太大差異。

表 4-1 基本組與信貸組每月消費金額、頻率敘述統計量

單位：次／元	基本組		信貸組	
	每月消費頻率	每月消費金額	每月消費頻率	每月消費金額
最小值	0.83	34.17	0.83	19.17
第一四分位	1.50	2,438.58	1.67	2,995.54
中位數	2.67	5,199.83	3.00	6,277.92
第三四分位	4.50	10,665.50	5.17	12,623.58
最大值	20.00	158,112.80	18.00	1,444,242.67
平均值	3.42	8,848.26	3.73	11,772.12
標準差	2.49	11,524.19	2.64	36,664.36


## 二、人口統計資料整理

顧客的人口統計資料包含性別、年齡、婚姻狀況、教育程度、居住地區及職業別，為方便研究分析，將資料中部分定義的原始水準合併為新的水準，如表 4-2 所示。

接著，依據挑選出來的兩組人員，計算合併後人口統計變數下各水準的人數分布，如表 4-3 所示，依照所設定的樣本挑選，基本組與信貸組各人口統計變數的不同水準做以下說明。

- (一) 性別：基本組與信貸組兩族群的男女比例相近，接近一比一。
- (二) 年齡：信貸組顧客有約四成的人數年齡落在 31-40 歲之間，基本組的人數則是平均落在各年齡層。
- (三) 婚姻狀況：兩個族群的單身、已婚比例相近，接近一比一。
- (四) 教育程度：信貸組顧客有接近六成的學士以上學歷。
- (五) 居住地區：信貸組顧客在各區域的人數比例上相等。
- (六) 職業別：兩組群在各水準的比例並無太大差異，大致也符合信用卡客戶的職業別比例。

表 4-2 人口統計變數水準定義



變數	更新水準	原始水準
性別	男	男
	女	女
年齡	21-30	21-30
	31-40	31-40
	41-50	41-50
	51 以上	51-60、61-70
婚姻狀況	單身	單身
	已婚	已婚
教育程度	學士以上	學士、研究所以上
	高中以下	初中、專科、高中
居住地區	北部	大台北地區
	中部	桃竹苗、中彰投地區
	南部及其它	雲嘉南、高屏、東部離島地區
職業別	自由業／服務業	自由業、服務業
	白領族	白領從業人員
	藍領族	藍領從業人員
	軍公教	軍人、公務人員、教職
	學生／退休／其它	學生、無業、退休



表 4-3 顧客人口統計結構分布

		基本組		信貸組	
		人數	百分比	人數	百分比
性別	男	2714	50.2%	1771	56.2%
	女	2693	49.8%	1379	43.8%
年齡	21-30	1277	23.6%	535	17.0%
	31-40	1328	24.6%	1307	41.5%
	41-50	1153	21.3%	889	28.2%
	51 以上	1649	30.5%	419	13.3%
婚姻狀況	單身	2559	47.3%	1711	54.3%
	已婚	2848	52.7%	1439	45.7%
教育程度	學士以上	2367	43.8%	1759	55.8%
	高中以下	3040	56.2%	1391	44.2%
居住地區	北部	2556	47.3%	1098	34.9%
	中部	1056	19.5%	1122	35.6%
	南部及其它	1795	33.2%	930	29.5%
職業別	自由業／服務業	765	14.1%	376	11.9%
	白領族	2308	42.7%	1448	46.0%
	藍領族	1140	21.1%	818	26.0%
	軍公教	399	7.4%	233	7.4%
	學生／退休／其它	795	14.7%	275	8.7%

資料來源：本研究整理

## 第二節 變數假設




本研究的目的是在於找出顧客出現資金需求的原因，並配合研究主題與模型規劃的適切性，將每一位信用卡使用者視為可能出現資金需求的潛在顧客，目標亦即以 2012 年 5 月為基準，試算每一位信用卡持卡人辦理信用貸款的可能性，對於銀行業者而言，實質的意義在於當顧客出現資金需求的訊號時，能搶先其他同業以正確的時間點進行行銷活動的推廣，增加來本行辦理信貸的機會。

根據信用卡資料庫的內容，本研究嘗試在尚未建構預測模型之前，以直覺的想法大膽假設以下交易行為變數和人口統計變數，並說明會如何影響信貸與否的情況，在往後的章節將會以下述假設的變數投入模型中，觀察真實情形是否對顧客辦理信用貸款具有顯著影響。

### 一、交易行為變數假設

- (一) 顧客持有信用卡數：信用卡數多代表顧客對銀行依賴程度或者互動程度愈高，雖然並非每一張信用卡都是活躍的，但可以判斷顧客與銀行往來的密切性，也可反應刷卡消費的習慣，將可能使顧客對於信用貸款取得資金管道的認同感比一般民眾高。
- (二) 顧客與本行往來年資：代表顧客使用本來第一張信用卡至今的時間長度，愈長代表顧客對於使用銀行服務有一定的熟悉度，能夠較容易掌控自己刷卡消費的習慣。

- 
- (三) 顧客過去半年曾欠款次數：顧客過去有未償清債務或者曾借款者，代表可能再次出現資金需求。
- (四) 本行刷卡金額之荷包佔有率：若顧客持本行信用卡交易金額占比愈高，雖不能說明其是否和辦理信貸有直接關係，但可推估顧客對本行信用卡忠誠較高，當出現資金需求時或許會第一順位向本行辦理信貸。
- (五) 本行來電客服數：會打客服專線的顧客代表其對於銀行的服務有一定的信任與連結性。
- (六) 顧客消費金額波動性：根據過去半年每月消費的變異程度作分析，變異程度愈高的顧客代表其使用信用卡交易的金額不穩定。
- (七) 顧客消費金額平均：平均消費金額愈高者代表其對於使用信用卡支付的依賴程度較高。
- (八) 顧客消費金額變異程度：此為消費金額波動性（標準差）除以消費金額平均之值。
- (九) 顧客簽帳金額占欠款總餘額：簽帳金額占欠款總餘額高者，代表其相對使用循環利息較低。
- (十) 顧客循環信用占欠款總餘額：代表顧客每月循環信用餘額占總欠款的金額，比例愈高愈可能出現資金短缺。
- (十一) 顧客每月平均還款比率：代表每月顧客償還簽帳金額的比例。

## 二、人口統計變數假設



- (一) 顧客性別：平均而言，男性的可支配所得較女性為高，但消費支出不一定比女性來得高。
- (二) 顧客年齡：隨顧客年齡的增長，可能有更多的支出必須負擔。
- (三) 顧客婚姻狀況：已婚者較單身可能有較多必需性生活支出，
- (四) 顧客教育程度：教育程度低者由於踏入社會較早，收入可能不高，較容易出現資金短缺。
- (五) 顧客居住地區：北部地區較其他地區物價高，且因居住地區的不同可能有資訊接收與認知的差異，可能會導致顧客對於信用貸款的觀念落差。
- (六) 顧客職業別：軍公教人員收入穩定，可將未來所得提前支配，又或者較能接受分期付款等概念，推測其較不容易出現資金需求，不同職業的薪資收入相異。

### 第三節 顧客交易行為分析



信用卡持卡人交易的購買期間 (Inter-purchase Time) 指的是兩次交易所間隔的時間，本研究以天為單位計算，藉由分析購買期間可以得知每次交易的間隔時間變化，對於每位客戶來說，其購買期間可能有逐漸縮短或者拉長的趨勢，不同的變化情形代表顧客的行為資訊，亦可分析顧客對銀行的價值貢獻程度。

樣本處理方式為合併同一天內的兩筆以上交易記錄為一筆，並剔除過去半年內少於四次交易次數的客戶，本研究將採用文獻回顧中說明的最大概似估計法 (MLE)、加權最大概似估計法 (WMLE) 及層級貝式估計量 (Hierarchical Bayesian Estimation) 三種方法來估計篩選顧客的平均購買期間，如此一來可分析顧客截至目前的交易狀態，還能利用這三種估計量之間的關係來建構顧客的活躍性與穩定性指標。

#### 一、最大概似估計法 (MLE) 與加權最大概似估計法 (WMLE)

計算平均購買期間的估計，最大概似估計量的概念為，舉例來說某一位客戶共有五次的購買期間，間隔期間分別為 40 天、30 天、20 天及 10 天，因此以 MLE 所估算的平均購買期間為  $(50 + 40 + 30 + 20 + 10)/5 = 30$  天。

除了最大概似估計法外，加權最大概似估計法的特色在於加入權重的因素，距離越近的購買區間給予較高的權重比例，如上述的例子，此時利用 WMLE 所估計的加權平均購買期間為  $(50 \times 1 + 40 \times 2 + 30 \times 3 + 20 \times 4 + 10 \times 5)/15 = 23.33$  天。

接著分別將基本組與信貸組的 MLE 和 WMLE 整理成圖 4-3 和圖 4-4，兩種估計法計算的平均購買期間均為右偏分配，就 MLE 而言，基本組與信貸組分別約 12.68 天與 11.99 天即會產生刷卡交易行為，而以 WMLE 而言，基本組與信貸組分別約 12.77 天與 12.05 天即會產生刷卡交易行為，且絕大多數的顧客的購買區間小於 30 天，基本組與信貸組分別占了 95.3% 與 95.7% 的比例，代表這群顧客每月至少會刷卡一次。

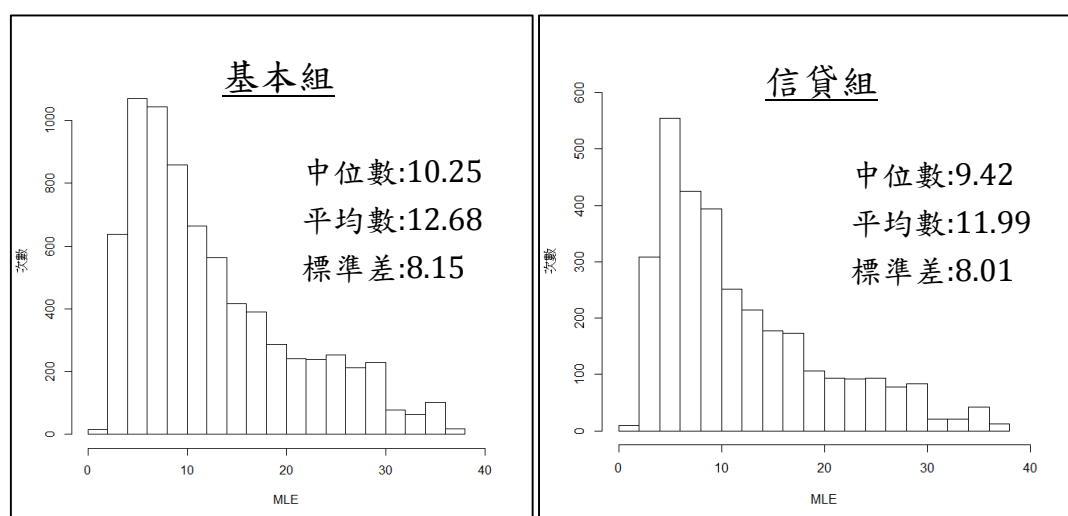


圖 4-3 基本組與信貸組—平均購買期間 MLE

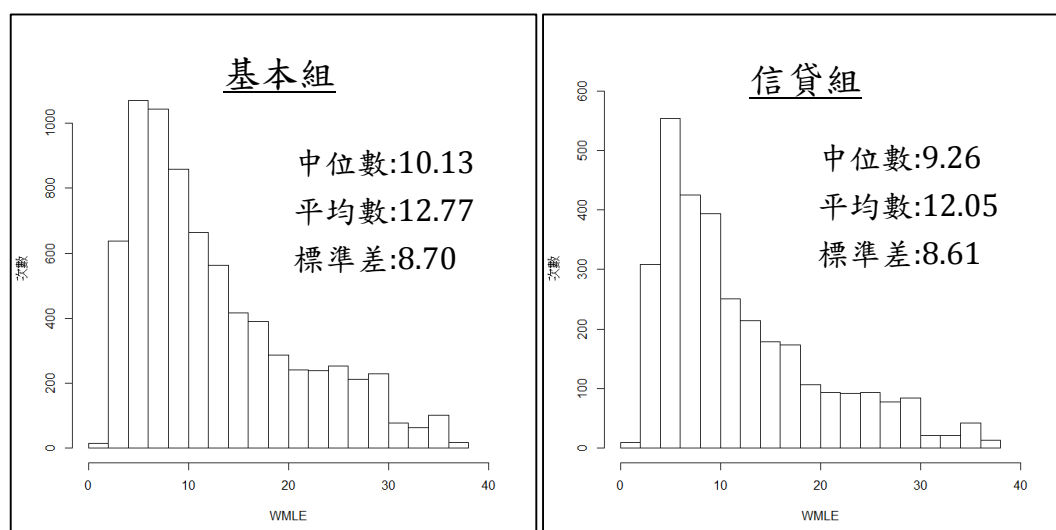


圖 4-4 基本組與信貸組—平均購買期間 WMLE

## 二、層級貝氏估計法 (Hierarchical Bayesian Estimation)

由於 MLE 與 WMLE 在估計時僅使用顧客個人的資訊，當個人交易次數的樣本過少時，無法真正代表顧客的真實購買情況，將產生嚴重偏誤，因此利用層級貝氏估計法，除考量個人樣本的資訊外，並加入整體顧客的交易記錄的資訊，可以修正個別顧客的異質性情況，修正購買次數較少的顧客，降低產生偏誤的可能性，因此層級貝氏估計法所計算的平均購買期間，如圖 4-5 所示。

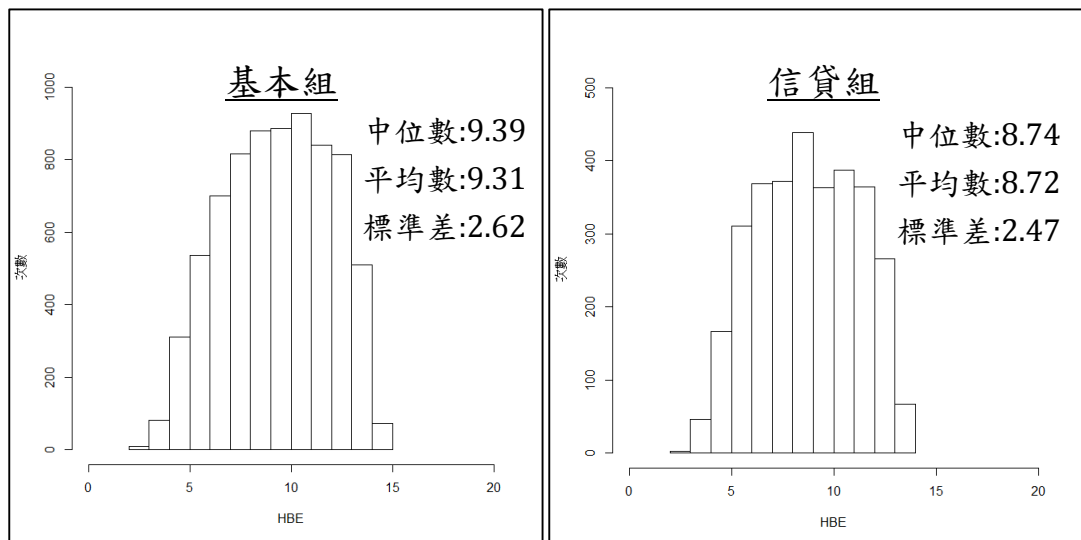


圖 4-5 基本組與信貸組—平均購買期間 HBE

貝氏層級估計法相較 MLE 與 WMLE 增加考率整個母體參數的修正機制，在預測顧客購買期間上擁有修正小樣本誤差之功用，如今透過隨機抽樣 6 名顧客資料，比較 HBE 與 MLE 的差異性，如表 4-4 所示，從資料中可以得知，當顧客交易次數愈高時，由 HBE 所推估的平均購買期間愈小，且與 MLE 的平均購買期間差距愈小。因此，當顧客在交易次數上有較多樣本時，顧客的事後機率分配受本身購買行為（即 MLE）影響較大，而受所有樣本資訊建構的事前機率分配影響較小。相反地，當顧客在交易次數上樣本較少時，顧客的事後機率

分配受本身購買行為（即 MLE）影響較小，而受所有樣本資訊建構的事前機率分配影響較大。以上的說明即為層級貝氏模式的修正機制。



表 4-4 最大概似估計法與層級貝氏估計法之比較

顧客編號	交易次數	MLE	HB	兩者差距
3820	5	32.4	13.42	18.68
7790	8	21.88	12.45	9.43
14719	10	16.9	11.42	5.48
13596	16	11.19	9.8	1.39
15970	22	8.05	8.35	-0.3
16542	49	3.73	5.17	-1.44



## 第四節 顧客活躍性及穩定性分析



### 一、顧客活躍性分析

本研究嘗試以第二節定義的 MLE 與 WMLE 做組合，建構顧客活躍性指標 (Customer Activity Index, CAI)，能夠更進一步針對顧客購買區間的變化情況，了解顧客在整個時間內的交易活躍程度，活躍性指標的計算方式如下所示：

$$CAI = \left( \frac{MLE - WMLE}{MLE} \right) \times 100\%$$

計算出來的 CAI 值域介於  $\{-100\%, 100\%$  之間，當活躍性指標大於 0 且數值愈大時，代表該名顧客活躍程度較高 (Active)，購買期間有逐漸縮短之趨勢；相反地，當活躍性指標小於 0 時，代表該名顧客活躍程度較低 (Inactive)，購買期間呈現逐漸拉長之趨勢。因此，我們從資料中挑 4 名顧客的活躍性指標，列出 10 次交易的購買期間間隔，幫助了解顧客活躍性的義涵，如表 4-5 所示。

表 4-5 顧客活躍性指標範例

顧客 編號	CAI	購買期間									
		1	2	3	4	5	6	7	8	9	10
113895	41.81%	56	21	42	9	14	9	10	4	3	9
12303	33.79%	75	28	5	7	14	10	7	11	7	16
10118	34.49%	6	2	10	2	8	7	11	18	35	52
17341	60.19%	1	1	4	1	9	17	2	13	9	99

比較兩族群的活躍性指標，由表 4-6 可看出基本組和信貸組的顧客約各有一半的人數活躍、一半人數不活躍（因活躍性指標之中位數落於 0% 附近），而整體而言，在不同分位數之下，信貸組的活躍程度稍較基本組高，顯示當信貸組的顧客愈接近出現資金需求的時間點時，消費的購買期間會逐漸縮短。

表 4-6 基本組與信貸組活躍性分析

	基本組	信貸組
最小值	-74.34%	-62.51%
第一四分位	-7.66%	-7.23%
中位數	0.90%	0.96%
第三四分位	8.04%	8.41%
最大值	59.88%	64.45%

## 二、顧客穩定性分析



考慮個別樣本的變異性，顧客穩定性指標（Customer Reliability Index, CRI）的建立有助於了解顧客在購買區間上的行為變異大小，計算方式如下所示：

$$CRI = \frac{IM_i - HB_i}{IM_i - GM_i}$$

其中， $HB_i$ （Hierarchical Bayesian）：第  $i$  位顧客以層級貝氏模型所得的平均購買期間估計值。 $IM_i$ （Individual Mean）：第  $i$  位顧客個人行為資訊所得之平均購買期間估計值，即為最大概似估計值（MLE）。 $GM_i$ （Group Mean）：表示第  $i$  位顧客所屬之特定區隔  $j$  內，由層級貝氏估計量所得之群體平均購買期間估計值。

計算出來的穩定性指標數值介於  $\{0, 100\%$  之間，當 CRI 數值愈小時，表示該名顧客的購買期間受到個人資訊影響較大，變異性小，顧客較為穩定；相反地，當 CRI 數值愈大時，表示該名顧客的購買期間受到群體資訊影響較大，變異性大，顧客較不穩定。因此，我們從資料中挑 4 名顧客的穩定性指標，列出 10 次交易的購買期間間隔，幫助了解顧客活躍性的義涵，如表 4-7 所示。

表 4-7 顧客活躍性指標範例

顧客 編號	CAI	購買期間									
		1	2	3	4	5	6	7	8	9	10
3339	14.53%	1	3	2	5	1	1	2	1	3	10
3430	30.36%	5	8	11	2	2	7	4	6	11	7
4037	58.89%	22	20	2	5	26	9	21	7	14	7
15907	80.05%	6	116	32	3	25					

比較兩族群的穩定性指標，由表 4-8 可看出基本組和信貸組顧客的穩定性指標在不同分位數之下並無太大差異，顯示顧客消費的購買期間穩定性不會對於是否產生資金需求有一定的影響力。

表 4-8 基本組與信貸組活躍性分析

	基本組	信貸組
最小值	14.32%	16.54%
第一四分位	42.62%	40.84%
中位數	55.63%	54.32%
第三四分位	69.03%	68.15%
最大值	80.05%	81.06%

## 第五節 迴歸模型配適與預測



### 一、變數定義

本研究主要目的在於觀察顧客在不同交易型為資料下，是否出現資金需求，因此依變數 (Dependent Variable) 的選擇為辦理信用貸款與否，為二元 (Binary) 型態，接著加入不同的自變數 (Independent Variable) 後，以 Logistic 迴歸模式和 Probit 模式產生迴歸式，計算樣本擊中率 (Hit Ratie)。

首先，原始的人口統計變數為類別變數 (Category Variable)，為使其能夠放進迴歸模式中，於是將資料調整為虛擬變數 (Dummy Variable)，如表 4-9 所示。而交易行為資料的自變數選擇如同本章第二節的變數假設，整理如表 4-10 所示。

表 4-9 人口統計變數定義

變數 名稱	水準	Coding	變數 代號	虛擬變數			
				(1)	(2)	(3)	(4)
性別	男	1	Gen(1)	1	0		
	女	2	Gen	0	0		
年齡	21-30	1	Age(1)	1	0	0	
	31-40	2	Age(2)	0	1	0	
	41-50	3	Age(3)	0	0	1	
	51 以上	4	Age	0	0	0	
婚姻 狀況	單身	1	Mar(1)	1	0		
	已婚	2	Mar	0	0		
教育 程度	學士以上	1	Edu(1)	1	0		
	高中以下	2	Edu	0	0		
居住 地區	北部	1	Area(1)	1	0	0	0
	中部	2	Area(2)	0	1	0	0
	南部及 其它	3	Area(3)	0	0	1	0
		4	Area	0	0	0	0
職業別	自由服務業	1	Occu(1)	1	0	0	0
	白領族	2	Occu(2)	0	1	0	0
	藍領族	3	Occu(3)	0	0	1	0
	軍公教	4	Occu(4)	0	0	0	1
	學生退休其它	5	Occu	0	0	0	0

表 4-10 交易資料變數定義

變數名稱及代號	說明	單位
持有信用卡數 (Card_Num)	顧客持有本行及他行總信用卡數	張
與本行往來年資 (Relation_Yr)	顧客持有本行第一張信用卡至今時間長短	年
過去半年曾欠款月數 (R_Num)	過去半年每個月分債務狀況， 即行內或行外出現 R 的次數	次
荷包占有率 (Share of Wallet)	顧客持本行與他行信用卡消費占比	百分比
本行來電客服數 (Phone_srv)	顧客過去半年來電客服數	次
平均每月刷卡金額 (Avg_mth_spt)	顧客於本行平均每月消費金額	元
平均每月刷卡次數 (Avg_mth_frq)	顧客於本行平均每月消費次數	次
顧客消費金額波動性 (JCIC_Std)	顧客交易金額的變異大小	無
顧客消費金額平均 (JCIC_Avg)	顧客於行內行外交易金額的平均	元
顧客消費金額變異程度 (JCIC_CV)	顧客交易金額的變異程度	百分比
前期較平均消費變動 (JCIC_Last_to_avg)	顧客前期交易金額對於 平均交易金額的變動程度	百分比

前期較前二期消費變動 (JCIC_05_to_04)	顧客前期交易金額對於 前二交易金額的變動程度	百分比
本行荷包占有率 (CT_SOW)	顧客於本行與他行消費金額占比	百分比
平均簽帳金額占餘額 (SA_RB)	每月簽帳金額占總欠款的比率	百分比
循環信用金額占餘額 (LA_RB)	每月循環信用餘額占總欠款的比率	百分比
顧客每月平均還款比率 (Pay_ratio)	每月顧客償還簽帳金額的比率	百分比





## 二、交易行為變數相關性分析

將人口統計變數及交易行為變數放入迴規模式前，必須先針對所有變數進行相關性分析，為避免產生共線性的問題，降低迴歸式的解釋力，當兩變數存在高度相關時（本研究設定相關係數大於 0.7 即為高度相關），則不宜同時放入迴歸式中。根據本研究針對所有變數進行相關性分析，發先以下兩組變數出現高度相關，分別為：

（一）每月消費次數（Avg\_amt\_frq）和穩定性指標（CRI）：選擇 CRI 放入迴歸式中。

（二）顧客消費金額波動性（JCIC\_Std）和顧客消費金額平均（JCIC\_Avg）：選擇 JCIC\_Avg 放入迴歸式中。


### 三、Logistic 模式



本研究根據先前整理之基本組顧客 5,407 名和信貸組顧客 3,150 名的資料中，採用隨機抽樣的方式各抽取 1,500 名，兩組共 3,000 名樣本，將所有設定好的人口統計變數和交易行為相關變數放入 Logistic 模型中試算，結果如表 4-11 所示。採用的顯著水準為  $\alpha=0.05$ ，當表中的變數之  $p\text{-value} < 0.05$  時，即可判定該變數顯著，代表對於辦理信用貸款具有影響力。

舉例而言，Gen (1) 為顯著變數且係數估計值為 0.219425，代表顧客若為男性，則相較於女性，男性出現資金需求而去辦理信貸的勝算比 (Odds ratio) 為  $\exp(0.219425)=1.245$ ，顯示單就性別來說，男性辦理信貸的機率為 55.46% (由  $p/(1-p)=1.245$  可得)；而 Card\_Num 亦為顯著變數且係數估計值為 0.160203，代表顧客每增加持有一張信用卡，辦理信用貸款的勝算比為  $\exp(0.160203)=1.174$ ，顯示多持有一張信用卡的顧客去辦理信貸的機率為 54.00% (由  $p/(1-p)=1.174$  可得)。此外，若顯著變數的係數估計值為負，代表該變數對於辦理信貸與否具有反向關係，例如來電客服數愈多，辦理信貸的可能性愈低。

表 4-11 Logistic 迴歸式係數與顯著性



	係數估計值	p-value
Gen(1)	0.219425	<b>0.000</b>
Age		<b>0.000</b>
Age(1)	0.636671	<b>0.000</b>
Age(2)	1.202203	<b>0.000</b>
Age(3)	1.057307	<b>0.000</b>
Mar(1)	0.081629	<b>0.000</b>
Edu(1)	0.341534	<b>0.000</b>
Area		<b>0.000</b>
Area(1)	-0.71621	<b>0.000</b>
Area(2)	1.007075	<b>0.000</b>
Occu		<b>0.000</b>
Occu(1)	0.333032	<b>0.011</b>
Occu(2)	0.372704	<b>0.001</b>
Occu(3)	0.782998	<b>0.000</b>
Occu(4)	0.63166	<b>0.000</b>
Card_num	0.160203	<b>0.000</b>
Relation_Yr	-0.05467	<b>0.000</b>
R_num	0.087362	<b>0.001</b>
Phone_num	-2.89273	<b>0.000</b>
Avg_mth_spt	0.000025	<b>0.000</b>
JCIC_AVG	0.000000	0.530
JCIC_CV	-0.52989	<b>0.000</b>
JCIC_Last_to_avg	0.035914	0.541
JCIC_05_to_04	-0.00027	0.786
CAI	-0.01744	0.935
CRI	0.738482	<b>0.001</b>
CT_SOW	-0.04367	0.479
SA_RB	57.4513	<b>0.001</b>
LA_RB	57.2397	<b>0.001</b>
Avg_pay_ratio	-2.20656	0.126
常數	57.70588	<b>0.001</b>

接續我們可以根據投入 Logistic 迴歸模式所估計出的預測機率值，設定顧客的預測機率大於 50% 則有可能辦理信貸，並比較該名顧客真實的信貸申辦情形，究竟為有申辦或無申辦，進行擊中率分析如表 4-12 所示。結果顯示，Logistic 迴歸模式的整體擊中率為  $(1,068+1,113)/3,000 = 72.7\%$ ，比起一開始各投入 1,500 名顧客時，以隨機預測法可得到的 50% 擊中率來的好。

表 4-12 Logistic 模式擊中率

單位：人數			預測次數		
			信貸申辦		百分比修正
			無申辦	有申辦	
觀察次數	信貸申辦	無申辦	1,068	432	71.2%
		有申辦	387	1,113	74.2%
	擊中率				<b>72.7%</b>

此外，為使該模型可以應用於隨機顧客的機率預測，並了解其真實擊中情形為何，本研究接續以樣本外預測的方式，從基本組與信貸組各隨機抽樣 500 名樣本，依據 Logistic 迴歸模式所估計的變數係數，將 1,000 名樣本外顧客的觀察值代入，再藉由 Logistic 迴歸公式的轉換，計算信貸貸款的預測機率，當預測機率大於 0.5 時，視為辦理信貸者，反之當預測機率小於 0.5 時，視顧客為不會辦理信用貸款者。

因此藉由整理樣本外 1,000 名顧客真實信貸觀察與預測的信貸結果，如表 4-13 所示，整體的信貸預測擊中率為 72.2%，稍低於樣本內的擊率率，但表現尚可接受。對於銀行端而言，行銷的義涵在於，不管真實顧客信貸狀況為何，當篩選顧客進行信貸資訊的推廣時，可以鎖定預測結果為申辦信貸


者，命中率將高達  $388/(161+388)=70.4\%$ ，而預測結果為無信貸但真實情況卻辦理信貸者（117 人），可視為銀行端對於信貸宣傳的「漏網之魚」，亦即不給予任何訊息，這群人依然會出現資金需求前來辦理信貸。



表 4-13 Logistic 模式樣本外預測擊中率

單位：人數			預測次數		
			信貸申辦		百分比修正
			無申辦	有申辦	
觀察次數	信貸申辦	無申辦	339	161	67.8%
		有申辦	117	383	76.6%
		擊中率			<b>72.2%</b>


#### 四、Probit 模式



除 Logistic 迴歸模型外，我們建立另一種模式來比較預測結果，將原本隨機抽樣的基本組與信貸組同一群樣本納入 Probit 中，並同樣設定顯著水準  $\alpha=0.05$ ，模型結果如表 4-14 所示，顯著的變數與 Logistic 模式中差異不大，其中平均每月還款金額 (Avg\_pay\_ratio) 此時轉為顯著。當變數顯著時，代表其對於辦理信貸與否具有影響力，係數估計值的正負符號則代表影響的相關方向。

相同地，我們也能針對 Probit 模式進行擊中率分析，結果如表 4-15 所示，顯示整體擊中率高達 70.2%，和 Logistic 模式相差並不大。

表 4-14 Probit 迴歸式係數與顯著性



	係數估計值	p-value
Gen(1)	0.12489	<b>0.001</b>
Age		<b>0.000</b>
Age(1)	0.36706	<b>0.000</b>
Age(2)	0.716337	<b>0.000</b>
Age(3)	0.63277	<b>0.000</b>
Mar(1)	0.04498	0.337
Edu(1)	0.2065897	<b>0.000</b>
Area		<b>0.000</b>
Area(1)	-0.4338735	<b>0.000</b>
Area(2)	0.5928723	<b>0.000</b>
Occu		<b>0.000</b>
Occu(1)	0.1978833	<b>0.010</b>
Occu(2)	0.2191341	<b>0.000</b>
Occu(3)	0.4660015	<b>0.000</b>
Occu(4)	0.3617609	<b>0.000</b>
Card_num	0.0939858	<b>0.000</b>
Relation_Yr	-0.0323789	<b>0.000</b>
R_num	0.0536713	<b>0.000</b>
Phone_num	-1.224394	<b>0.000</b>
Avg_mth_spt	0.0000114	<b>0.000</b>
JCIC_AVG	0.000000	0.547
JCIC_CV	-0.3195744	<b>0.000</b>
JCIC_Last_to_avg	0.0204857	0.559
JCIC_05_to_04	-0.0001331	0.770
CAI	-0.0148068	0.907
CRI	0.3657143	<b>0.005</b>
CT_SOW	-0.0102113	0.779
SA_RB	30.13759	<b>0.000</b>
LA_RB	29.91436	<b>0.000</b>
Avg_pay_ratio	-1.743382	<b>0.044</b>
常數	30.80054	<b>0.000</b>

表 4-15 Probit 模式擊中率

單位：人數			預測次數		
			信貸申辦		百分比修正
			無申辦	有申辦	
觀察次數	信貸申辦	無申辦	1,075	425	71.7%
		有申辦	469	1,031	68.73%
	擊中率				<b>70.2%</b>

接著，樣本外預測的樣本選擇和 Logistic 模式相同，將各變數帶入模型中試算進行樣本外擊中率分析，結果如表 4-16 所示，擊中率達 71.9%。

表 4-16 Probit 模式樣本外預測擊中率

單位：人數			預測次數		
			信貸申辦		百分比修正
			無申辦	有申辦	
觀察次數	信貸申辦	無申辦	337	163	67.4%
		有申辦	136	364	72.8%
	擊中率				<b>70.1%</b>





## 五、模型預測力評估

本研究以 2012 年 5 月為基準，比較基本組與信貸組過去六個月的交易記錄資料，用以預測下兩個月（6 月或 7 月）是否會辦理信用貸款，透過兩模型的建立，試評估不同模式的預測力效果。

- (一) 隨機機率法：在不參考任何資訊的情況下，根據隨機猜測方式的擊中率為 50%。
- (二) Logistic 模式：樣本外擊中率為 72.2%。
- (三) Probit 模式：樣本外擊中率為 70.1%。

接著，本研究的模型評估方式以推力圖（Lift Chart）呈現，推力可以指出不同模式在預測信貸的績效上比起隨機機率法提升多少，依據基本組與信貸組各選取 500 名顧客樣本的模型外預測結果，我們將這 1,000 名顧客不同變數的觀察值分別投入 Logistic 和 Probit 模型後產生每位顧客的機率預測值，再將這些機率值由大到小排序，並切割為十個等分區塊，每個區塊 100 名顧客，計算每名顧客辦理信貸與否的真實觀察結果，記錄在每一區塊的信貸人數中，即可計算每一區塊的回應率，結果如表 4-17 所示。

此外，推力的計算方式為， $\text{推力} = (\text{每一區塊回應率}) / (\text{整體回應率})$ 。當推力值大於 1，代表比平均績效好，當推力值小於 1，代表比平均績效差，故由兩模型的樣本外預測結果顯示，機率預測值的前 6 個區塊推力值均大於一，代表模型在前 60% 的顧客信貸預測績效較佳。

表 4-17 Logistic 和 Probit 模式推力圖比較

區塊	顧客 人數	信貸人數		回應率		推力		累積推力	
		L	P	L	P	L	P	L	P
1	100	85	80	85%	80%	1.70	1.60	1.70	1.60
2	100	74	75	74%	75%	1.48	1.50	3.18	3.10
3	100	72	68	72%	68%	1.44	1.36	4.62	4.46
4	100	69	65	69%	65%	1.38	1.30	6.00	5.76
5	100	56	56	56%	56%	1.12	1.12	7.12	6.88
6	100	51	48	51%	48%	1.02	0.96	8.14	7.84
7	100	42	45	42%	45%	0.84	0.90	8.98	8.74
8	100	32	36	32%	36%	0.64	0.72	9.62	9.46
9	100	18	25	18%	25%	0.36	0.50	9.98	9.96
10	100	1	2	1%	2%	0.02	0.04	10.00	10.00
總和	1,000	500	500	50%	50%				

接著我們可以將兩模式的累積推力以推力圖表示，三種模式的比較如圖 4-6 所示，其中 45 度角的斜對角線為隨機機率法預測，導因於樣本篩選個數相等，而 Logistic 和 Probit 模型的各區塊預測績效相等，均可同時搭配使用於信貸機率的預測。

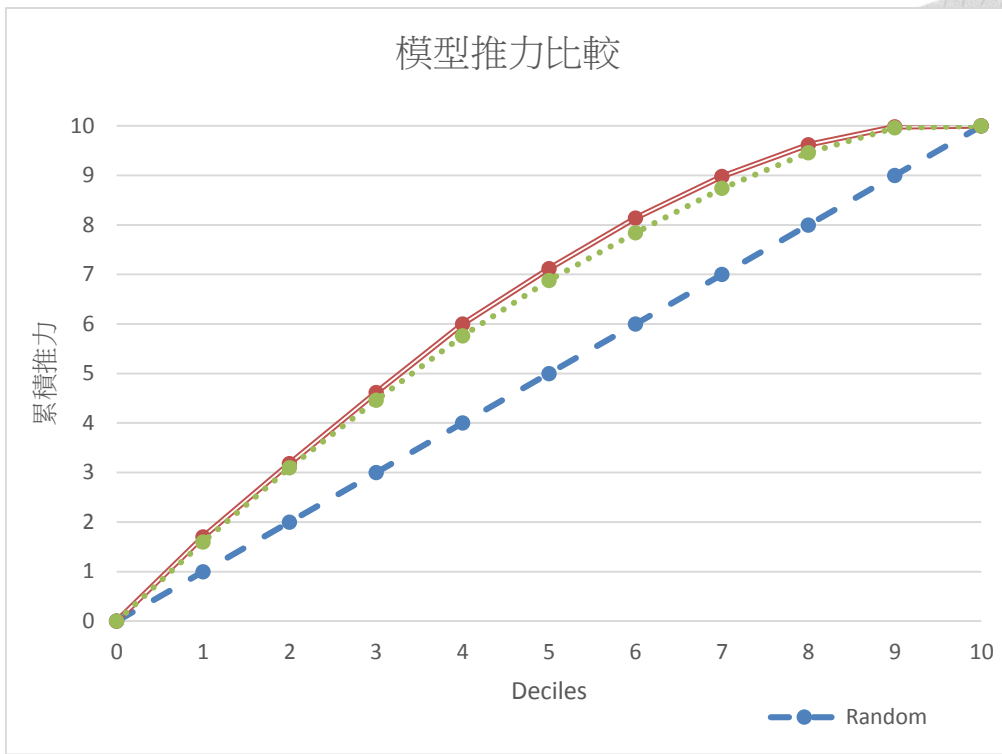


圖 4-6 模型推力績效比較

## 第五章 結論與建議



### 第一節 研究結論與行銷意涵

本研究的目的是在於推估潛在資金需求的信用卡顧客名單，藉由資料庫分析的概念，建立兩種模式預測顧客辦理信貸的機率，使得銀行端的行銷人員可以更準確的知道該向哪些顧客通提供即借款服務的資訊或者利率優惠方案，並且在不同的行銷預算之下，鎖定不同範圍的顧客，達到資料庫行銷的目的。

根據第四章的實證分析，透過對樣本資料的了解、變數的假設乃至於迴歸模型的建立，最終得到的結論如以下幾點說明。

#### 一、顧客交易行為分析

由顧客活躍性與穩定性指標的建立，試圖了解基本組與信貸組顧客在平均購買期間的變化情況，發現整體而言，信貸組的顧客稍較基本組的顧客活躍，而購買期間的穩定性則無太大差異，兩組顧客每月至少消費一次的比率皆高達 95% 左右。

#### 二、顯著變數整理

依據第四章 Logistic 和 Probit 兩模式的分析結果顯示，大部分的顯著如同一開始對於信貸與否的變數假設中，亦即，影響信貸相關顯著的變數整理如表 5-1 所示，並標明其方向。

表 5-1 顯著變數重要性列表

編號	變數名稱	與信貸機率關係
1	持有信用卡數 (Card_Num)	(+)
2	與本行往來年資 (Relation_Yr)	(-)
3	過去半年曾欠款月數 (R_Num)	(+)
4	本行來電客服數 (Phone_srv)	(-)
5	平均每月刷卡金額 (Avg_mth_spt)	(+)
6	顧客消費金額變異程度 (JCIC_CV)	(-)
7	穩定性指標 (CRI)	(+)
8	平均簽帳金額占餘額 (SA_RB)	(+)
9	循環信用金額占餘額 (LA_RB)	(+)

### 三、模型預測

本研究的兩種迴歸模型預測結果相仿，僅相差一個變數的顯著性，代表那些於兩模型中皆顯著變數具有一定的信貸預測解釋力，這些顯著的變數可以視為銀行內衡量顧客是否辦理信貸的關鍵性指標，可以提供行銷人員在維護資料庫並且做決策時的一個依據。

此外，模型對於樣本外預測的擊中率高達 70% 以上，在判定係數方面，Logistic 模型的 Cox & Snell R-square=0.280 和 Nagelkerke R-square=0.373，Probit 模型的 Pseudo R-square = 0.245，均高於一般假設的 0.15，顯示迴歸模式具有一定解釋力。

綜合以上的結論，對於銀行端業者來說，可將兩種迴歸模式應用於實際的信貸預測，當試算出顧客不同的預測機率時，可以依照成本預算的考量，劃分鎖定信貸活動宣傳的臨界機率值，例如當預算充足時，可將顧客信貸的預測機率以 0.4 為界，反之當預算緊縮時，將預測機率較高的顧客視為優先鎖定的潛在客群，如機率 0.7 為界。

## 第二節 研究限制



本研究的所有資料來自於銀行所提供，受限於非銀行內部企業活動研究之故，目前處理的資料僅為銀行信用卡資料庫的部份資訊，因資料的豐富性與變數的多寡決定模型預測的能力呈現，研究限制主要來自於顧客資料的完整性，可分為以下幾種類型說明。

### 一、交易記錄資料限制

本研究僅針對顧客過去半年的交易紀錄做分析，受限於資料月份較少的限制，僅觀察到顧客短期的消費行為模式，不容易觀察到顧客長期真實的交易習慣，容易產生預測上的偏誤，降低預測結果。

### 二、交易明細的產業別

雖有顧客過去交易紀錄的金額與明細，但缺乏產業別的明確的定義，如能提供產業別的劃分，如此一來可看出顧客於不同產業別消費金額多寡與占比，比較是否哪些產業別的消费支出或變異高，將是潛在的資金需求者。

### 三、其他類型限制

在進行變數假設時，尚與許多潛在變數可做討論，例如顧客持有房貸或車貸餘額、信用額度和信貸利率敏感度等，相信這些變數對於顧客是否出現資金需求具有一定的影響力。

### 第三節 後續研究建議



#### 一、資料庫的完整性

如同研究限制所述，未來如能向銀行取得更多的資料數量，將能提高模型預測的準確度與豐富性，且能針對不同變數的表現作顧客分群，便能夠有效地運用模型的配適提高信貸預測擊中率。

#### 二、銀行顧客的問卷發放

做為學術研究之目的，本研究主要用途在於預測顧客是否出現資金需求，因此在樣本的選擇上為顧客辦理信用貸款者（不分向行內、行外申辦），對於提供資料的銀行來說，更在乎顧客出現資金需求後，是否會第一優先考慮來本行申辦，但又關係到銀行經營的與品牌知名度，未來如能透過行內問卷的發放，分析顧客出現資金需求時，第一選擇辦理信用貸款的銀行，目的能和銀行業者相結合，更符合實務上的應用。

#### 三、模型配適的選擇

本研究以 Logistic 和 Probit 兩種模式建構預測模式，在資料採礦（Data Mining）的技術上可採用別種諸如類神經網路的概念，找出最佳的預測模型。此外，隨著時間的變遷，例如先前出現的雙卡風暴和金融海嘯，皆會影響顧客交易消費的意願與金額，此時沒有一個模型是可以完美預測、不斷被使用的，必須針對選擇的模型或者投入變數做出適當的調整，才能確保預測的目的與準確性是否一致，達到資料庫行銷的最終目的—從顧客中挖掘潛在的獲利模式。

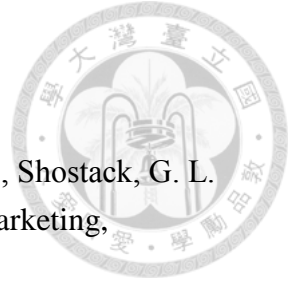


# 參考文獻



## 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未出版博士論文。
2. 李章偉 (2001), 「資料庫行銷之顧客價值分析-以 3C 流通業為例」，國立台灣大學國際企業學研究所未出版碩士論文。
3. 李宗龍 (2001), 「信用卡資料庫行銷之顧客價值分析與促銷成效之研究」，國立台灣大學國際企業學研究所未出版碩士論文。
4. 林慧晶 (1997), 「資料庫行銷之顧客價值分析與行銷策略應用」，國立台灣大學國際企業學系研究所未出版碩士論文。
5. 陳文華 (1999), 「應用資料倉儲系統建立 CRM」，資訊與電腦雜誌，第 226 期，頁 112-127。
6. 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，頁 132-138。
7. 葉德昌 (1994), 「資料庫行銷」，中國商銀月刊，第 13 卷，第 1 期，1994 年 2 月，pp.36-38。
8. 廖怡芳 (2004), 「資料庫行銷之應用-以百貨業為例」，國立台灣大學國際企業學研究所未出版碩士論文。
9. 遠擎管理顧問 (2001), 「顧客關係管理深度解析：執行以客戶為中心的企業轉型策略」，遠擎管理顧問公司。



## 二、英文部分

1. Berry, L. L. (1983) , “Relationship Marketing,” in Berry, L. L., Shostack, G. L. and Upah, G. D. (Eds) , Emerging Perspectives of Services Marketing, American Marketing Association, Chicago, IL, pp. 25-28.
2. Bitner, M. J., (1995) , “Building service relationships: It"s all about promises”, Academy of Marketing Science, Journal, Greenvale.
3. Christy, Richard, Gordon Oliver and Joe Penn (1996) , “Relationship Marketing in Consumer Markets,” Journal of Marketing Management, 12 ( Spring ) ,pp.175-187.
4. Copulsky, J. R. and M. J. Wolf (1990) , “Relationship Marketing: Positioning for the Future,” The Journal of Business Strategy, 11 ( 7-8 ) , pp.16-20.
5. Fletcher, K., Wright, G. and Desai, C. (1996) , “The Role of Organizational Factors in The Adoption an Sophistication of Database Marketing in The UK Financial Services Industry”, Journal of Direct Marketing, Vol. 10, Winter, pp 10~21
6. Hughes, A. M., “Strategic Database Marketing,” Chicago: Probus Publishing, 1994.
7. Jen, L. and Wang, S. J. (1998) , “Incorporating Heterogeneity in Customer Valuation : An Empirical Study of Health Care Direct Marketing in Taiwan”, International Journal of Operation and Quantitative Management, Vol.4, No.3, pp.217~228.
8. Kotler, Philip (2000) , Kotler on Marketing: How to Create, Win, and Dominate Markets, New York: Simon & Schuster Inc.
9. Newell, F. (1997) The New Rules of Marketing: How to Use One-on-One Relationship Marketing to be the Leader in Your Industry. New York: McGraw-Hill.
10. Peppers, D. and Rogers, M. (1999) , The One to One Manager: an Executive's Guide to Customer Relationship Management, 1st edition, New York: Sagalyn Literary Agency.
11. Peppers, D. and Dorf, B. (1999) , “Is Your Company Ready for One-to-One

- Marketing?” Harvard Business Review, 77 (January-February) , 151-160.
12. Shaw, R. and Stone, M. (1988) , Database marketing, Gower, English, P3.
13. Shani, D. and Chalasani s. (1992) , “Exploiting Niches Using Relationship Marketing”, Journal of Consumer Marketing, Vol.3, pp. 33-42.
14. Sheth, J. N., & Parvatiyar, A. (1995) , “The evolution of relationship marketing”. International Business Review, 4:4, 397-418.
15. Stone, Bob (1995) , “Successful Direct Marketing Methods, Lincolnwood”, IL : NTC Business Books, pp.37-59.
16. Swift, R.S. (2000) Accelerating Customer Relationships. Prentice Hall.

