模糊統計導論

吳柏林 著

方法與應用

methods and applications

模糊統計理論是一種定量化處理人類語言、思維的一個新興學問。
模糊邏輯並非如字面上那樣不精確，而是面對生活上各種不確定性，
以更合理的規定去管理控制，得到更有效率、
更合乎人性與智慧的結果。
因此，如何將統計方法論延伸到模糊理論的結構體系之中，
是本書的目的。

Introduction to Fuzzy Statistics

如何預估台北市選情？如何為台灣茶葉模糊分類？
模糊[illegible]歸與景氣循環的關係為何？
[illegible]間數列與股價指數預測又有什麼關聯？

模糊統計導論

方法與應用

吳　柏　林　著

國立政治大學應用數學系教授

五南圖書出版公司 印行

國家圖書館出版品預行編目資料

模糊統計導論 : 方法與應用 / 吳柏林著. --
初版. -- 臺北市 : 五南, 2005[民 94]
面 ; 公分
參考書目:面
ISBN 957-11-4102-X(平裝)

1. 模糊理論 2. 數理統計

319.4 94017782

1H36

模糊統計導論：方法與應用

作　者　吳柏林

出版者　五南圖書出版股份有限公司
發行人　楊榮川
地　　址：台北市大安區 106
和平東路二段 339 號 4 樓
電　　話：(02)27055066（代表號）
傳　　真：(02)27066100
劃　　撥：0106895-3
網　　址：http://www.wunan.com.tw
電子郵件：wunan@wunan.com.tw
顧　問　財團法人資訊工業策進會科技法律中心

版　刷　2005 年 10 月　初版一刷

定　價　550 元

序言

統計是用來分析、處理自然科學及社會科學資訊的工具。幫助人們在複雜的自然或社會現象中，藉由樣本資料所提供的訊息，經歸納分析、推論檢定、決策、預測等過程，使我們對現實狀況更了解，更能明確地處理現實世界的問題。傳統統計學的目地主要針對各類資訊，擬定一套估計檢定的測度方法，其過程包括：(1)設定合適的理論或模式 (2)收集樣本資料，實驗設計、抽樣或模擬(3)資料分析與研判 (4)估計與檢定 (5)決策或預測。

近年來由於智能科技發展一日千里，研究方法亦不斷的更新。傳統統計分析工具已漸感到不敷應用。一個主要的原因是: 如何更有效處理分析日益複雜，鉅量的網路情報資料。雖然資料採礦的興起，解決了不少資料分析的問題。但是對於如何處理非實數樣本資料，比如區間資料，多值資料型式之模糊樣本，應用架構在實變函數與機率論之傳統統計方法，實在已無法有效的分析與掌控。尤其是我們在決策過程中所遇到的不確定性問題比我們想像的更爲複雜。情報資訊除了隨機性外，還包括不完全的信息，部分已知的知識，或者對環境模糊的描述等。

事實上，我們所獲得的信息來自測量與感知，而感知信息中的不確定因素，主要是我們的語言對某些概念表達模糊所引起的。顯然要做出比較好的判斷，我們必須盡量將所能得到的信息都考慮在內。這包括用自然語言描述的行爲、意義等之屬性信息。因此我們需要用機率將模糊概念數學模式化，其實這也展示了不確定性的另一種型式。模糊理論是一種定量化處理人類語言，思維的一個新興學門。模糊邏輯並非如字面上意思那樣的馬虎，不精確。而是面對生活上各種的不確定性，以更合理的規則去分析去管理控制，以期得到更有效率，更合乎人性與智慧的結果。模糊統計並不模糊，它是處理不確定事件的新技術，帶領我們從古典的統計估計與檢定研究計算，進入一個需要軟計算，穩健性的高科技的 e 世代。

在傳統的統計推論方法中，爲了了解未知母體參數值，我們常藉由一些評估準則，找出適當的統計量來對母體參數進行估計。平均數是了解母體集中趨勢最重要的母體參數之一，我們常以其不偏估計量，亦即樣本平均數來估計。然而，在日常生活中，母體平均數常爲帶有模糊、不確定性的語意變數，或爲一可能區間，傳統的估計量評估準則及估計方法便無法適用於此種情形。

本書基於以軟計算方法，配合模糊集合理論，定義出模糊樣本均數，模糊樣本眾數及模糊中位數，並給訂很多相關之性質。同時，針對模糊參數之估計量，我們提出適當可行估計法的評判準則。對於古典的統計檢定必須陳列明確的假設。當我們想檢定兩母體平均數是否有差異時，虛無假設是“兩個平均數相等”。然而，有時我們想要知道的只是兩平均值是否模糊相等，此時傳統的檢定方法並不適用於這種包含不確定性的模糊假設檢定。因此本書提出基於模糊樣本之統計檢定方法，針對模糊均數相等，模糊屬於與卡方齊一性檢定作一進一步探討。

爲了將傳統統計方法延伸到模糊集合與系統的實務應用之中，本書將詳細介紹: 模糊問卷調查，模糊聚類分析，模糊回歸分析，模糊無母數統計，模糊時間數列分析與預測。我們舉了很多社會科學的應用實例，尤其是台灣生活化例子如：模糊問卷北市選情預估，樂觀量表，風景區滿意度調查，台灣茶葉模糊分類，模糊回歸與景氣循環，模糊時間數列與股價指數預測等等。期望藉以拋磚引玉，開創21世紀模糊統計與應用的嶄新領域。

吳柏林于台北

秋季， 2005

我不知道我在這世界上的地位；但是在神遊科學的世界裡，
我常把自己比喻成一位充滿好奇心的小孩，
在繽紛的海灘邊，撿拾一些更美麗的貝殼。
而浩瀚的知識大海則不斷地拍打在我的腳丫上，
激起無數的美麗浪花。

目錄

第 1 章 緒言

人類的思維主要是來自於對自然現象和社會現象的認知意識，而人類的知識語言也會因本身的主觀意識、時間、環境和研判事情的角度不同而具備模糊性。模糊理論的產生即是參考人類思維方式對環境所用的模糊測度與分類原理，給予較穩健的描述方式，以處理多元複雜的曖昧和不確定現象。因此，人類思維有兩類，一爲形式化思維(formal thinking)，另一爲模糊思維(fuzzy thinking)；前者是有邏輯性和順序性的思考，而後者則是全體性和綜合性的思考。當面臨決策判斷而進行思考時，基於形式化思維的二元邏輯，常很難表示出人類思考的多元邏輯特性。

當有人說他今天感到很快樂時，究竟他對於快樂的認知爲何呢？什麼樣的測量標準可以稱得上快樂呢？或是這樣的感覺持續多久的時間以上才能算是快樂呢？然而，這樣的問題，每個人的回答皆因其主觀性而有不同，即使回答者爲同一人，也會因爲所處的環境、或是外在條件的不同，而可能出現與之前相異的答案。諸如此類很多的論點和問題，都不是能夠用絕對的二元邏輯所可以界定的。原因則皆來自於人類思維的模糊性。但人類卻常常被要求做出絕對的判斷或選擇，以人性的觀點來看，這是十分不合理的。

模糊理論的概念，主要強調個人喜好程度不需非常清晰或數值精確，因此對人類而言，模糊模式比直接指定單一物體一個值，較合適於評估物體間的多元或相關特性。

對不確定性的事物作決策，是相當重要的人類活動。如果這種不確定性僅僅是由於事物的隨機所引起的，模糊統計分析發展爲這類決策活動提供了不錯的理論依據。是事實上，我們在決策過程中所遇到的不確定性問題，往往不只是由於事物的隨機所引起，這種不確定性還可能是：不完全的信息、部份已知的知識、對環境模糊的描述等，這類信息來自於測量與感知中的不確定因

素，主要是我們的語言及人類思維對某些概念表達模糊所引起。這些不明確性經常比我們想像的要複雜許多。

顯然地說，如果要對人類思維的模糊性做出比較好的判斷，我們必須盡量將所得到的信息都考慮在內，特別是屬性問題。由於屬性問題本身的不確定性與模糊性，若我們利用此假性的精確值來做因果分析與計量度量，可能造成判定偏差及決策誤導，甚至會擴大預測結果與實際狀態之間的差異。因此對於這些在思考認知不易表達完善的屬性問題，藉由軟計算方法與模糊統計分析可更明確表達出來。

雖然古典集合在數理科學上建立一套相當有系統邏輯的。但是，若將此集合關係應用於描述某些實務現象時，常發現不合理的情形。因爲某些現象並不一定存在「非此及彼」的關係。例如，進行某一教學單元後，將班級的學生劃分成「精熟」和「不精熟」兩類，這樣的劃分很明顯的有不合理之處，因爲學生的精熟度並非是二元的現象，而是有各種不同精熟程度連續性之特性。自Zadeh(1965)提出模糊理論以來，此思維可解釋許多實務現象。模糊理論將元素和集合之間的關係，以介於[0,1]之間的隸屬度(membership)描述。

由於傳統集合中二元邏輯與人類思維模式出入頗大，若能引用隸屬度函數，應能得到較合理的解釋。例如：人們認爲身高 200 公分絕對屬於高，則其隸屬度函數值自然屬於 1，而身高 180 公分或 178 公分的隸屬度函數值則約等於 0.8，此表示身高 180 公分或 178 公分屬於高的程度有 0.8 之多，再根據隸屬度函數的定義，我們可描繪出模糊集合中高的隸屬度函數。又如果某人認爲 40 歲絕對屬於中年，則其隸屬度函數值自然屬於 1，而 39 歲或 41 歲的隸屬度函數值則約等於 0.9，此表示 39 歲或 41 歲屬於中年的程度有 0.9 之多。

根據隸屬度函數的定義，我們可繪出模糊集合中年的隸屬度函數。與傳統集合的特徵函數比較，隸屬度函數似乎是將特徵函數平滑化了。不僅如此，隸屬度函數讓每個年齡層都擁有一個介於 0 到 1 之間的值，來代表屬於高或中年的程度。相較於傳統集合的特徵函數，在描述模糊的概念時，利用模糊集合

的隸屬度函數來解釋，是更適當的。

模糊理論是以模糊邏輯爲基礎，它將傳統數學之二元邏輯做延伸，不再是只有對錯或是非二分法。對於元素與集合的關係，古典集合論中元素是否屬於集合 A，必須十分明確不容模糊。即 $X \in A$ 或 $X \notin A$ 二者必居其一，且只能居其一。這種邏輯正是所謂的二元邏輯。然而人類的思維，因來自於對自然現象和社會現象的主觀意識影響，其知識語言也會因本身的主觀意識、時間、環境和研判事情的角度不同而具模糊性。對和錯之間還有不完全對，一點對或不完全錯等，是非之間還有有些是，有些非等地帶，正所謂的灰色地帶與模糊觀念。要瞭解模糊的意義亦可從模糊的相反詞明確來做反向思考。

有些學者認爲模糊理論既是研究不確定的現象，應與機率論類似。然而機率論是研究隨機性問題，隨機性雖不確定，但那是因爲條件不充分引起的，事件的發生是隨機的，事件之後卻是確定的，例如：丟一公正骰子，出現 1,2,3,4,5,6 點之機率均爲 1/6，當丟完一次之後，出現多少就是多少。而模糊理論的事件本身卻是模糊的不明確的，例如：回答家中經濟屬於不錯、小康或中等等，這些均不屬於隨機，而是事件本身的不完整性與不明確性。Zedah (1999) 更建議引用感覺測度 (perception measure) 和軟計算 (soft computing system) 共同應用做爲模糊函數估計量，這種應用模糊概念將屬性關係數學模式化的方法，我們統稱爲軟計算方法。希望此研究方向提供未來計量研究方法的一個有用的工具。

模糊概念並不只侷限在研究人類的思維與情感而已。在以往嚴謹精確的原則要求下，許多技術層面所衍生出的灰色地帶，都必須耗費相當大的心力爲複雜的系統寫下嚴密的定義與敘述，灰色地帶中的每一個細微末節，都必須完全地考慮到，盡全力使得其中的模糊變得明確，但若稍有一遺漏，則全盤皆墨，一切又得從頭做起。而模糊理論卻提供一種新的思維模式，只需要明瞭各種屬性的狀況，利用軟計算方法建立大略性的處理模式，即可處理系統中灰色地帶的問題。所以我們應該要了解到：灰色或是模糊不清的事件是層出不窮的，也是無法完全避免的，也因此，才讓我們體認到研究模糊理論的重要性。

隸屬度函數是模糊理論的基礎，它是從傳統集合中的特徵函數(characteristic function)所衍生出來的，用以表達元素對模糊集合的隸屬度(membership grade)，其範圍介於 0 到 1 之間。對於元素和集合的關係，古典集合將元素和集合之間的關係以特徵函數來說明，亦即 $I(x)=1,\ 若 x\in A;\ I(x)=0,\ 若\ x\notin A.$ 但是 Zadeh(1965)在模糊集合論中則提到，若一個元素屬於某一個集合的程度越大，則其隸屬度值越接近於 1，反之則越接近 0。

隸屬度函數是模糊理論最基本的概念，它不僅可以描述模糊集合的性質，更可以對模糊集合進行量化，並且利用精確的數學方法，來分析和處理模糊性資訊。然而，要建立一個足以表達模糊概念的隸屬度函數，並不是一件容易的事。其原因在於隸屬度函數脫離不了個人的主觀意識，故沒有通用的定理或公式，通常是根據經驗或統計來加以確定，很難像客觀事物一樣有很強的說服力。因此，隸屬度函數的建立經常是具有爭議性的，也沒有一種隸屬度函數是可以被廣泛接受而使用。

近年來，由於科技知識水平的提高與智慧科技多元發展，造就了現今財金，經濟與教育與心理研究環境的多變與複雜化。以往的社會科學研究多利用傳統的統計分析方法，如今卻因爲時代的不斷進步，而漸漸不符合現今多變環境的複雜性，以致於場感到研究方法之缺乏與不適用。如何以較爲進步而精確的方法來分析目前瞬息萬變的大環境是非常必要的。故書提出應用模糊理論的概念，將人類的喜好程度及各種屬性關係，轉換成各種便於計算的效用函數，進而適當的建立假設的數學模式。這些參考人類思維方式而建構出來的各種模糊統計分析，將可廣泛的應用於處理分析各種多元複雜的不確定現象。

第 2 章 隸屬度函數與軟計算方法

隸屬度函數通常可分為離散型(discrete)與連續型(continuous)兩類。離散型隸屬度函數是直接給定有限模糊集合內每個元素的隸屬度，並以向量的形式表現出來，而連續型隸屬度函數則有幾種常用的函數形式（S-函數、Z-函數、π-函數、三角形函數、梯形函數、高斯(指數)函數）來描述模糊集合。函數定義的表現，可以是無限模糊集合的元素及其隸屬度之間的關係，也可以是有限模糊集合的元素及其隸屬度之間的關係。

2.1 隸屬度函數與模糊數

過去我們利用傳統集合定義具有模糊性質的語言變數時，常會造成許多不合理的現象。例如，最近相當熱門的「景氣」一詞，當我們考慮景氣指標 0 到 100 的範圍時，若定義 30 到 50 為「景氣好」，則根據傳統集合的定義，可繪出「景氣」的特徵函數，如圖 2.1 所示。

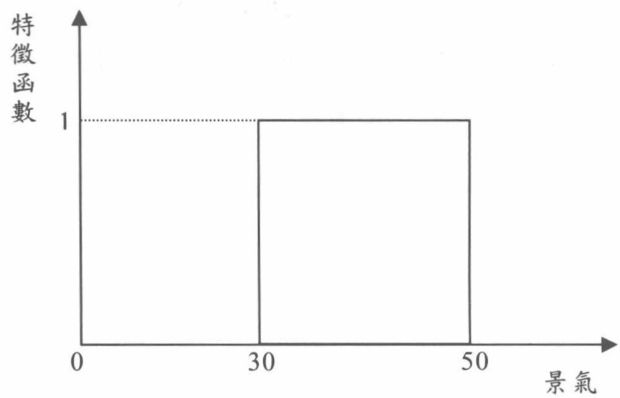

圖 2.1 傳統集合:景氣，的特徵函數

在圖 2.1 中顯示當某月景氣指標介於 30 到 50 之間，則屬於景氣好，其特徵值為 1；反之，便不屬於景氣好，特徵值為 0。但若我們假設有 A、B、C 三個月，景氣指標各為 29、31、49，其中 B、C 兩月指標值相差 18，且都屬於景氣好；但 A、B 兩月指標值雖只差 2，但 A 月卻不屬於景氣好，這是相當不合理的。故並不是全部都合理，須看特性決定。

對於這種傳統集合的二分法與人類思維模式格格不入的問題，利用隸屬度函數則能獲得較為合理的答案。如果某人認為 40 絕對屬於景氣好，則其隸屬度函數值自然屬於 1，而 39 或 41 的隸屬度函數值則約等於 0.9，此表示 39 或 41 屬於景氣好的程度有 0.9 之多。根據隸屬度函數的定義，我們可繪出模糊集合景氣指標的隸屬度函數。與傳統集合的特徵函數比較，隸屬度函數似乎

是將特徵函數平滑化了。不僅如此，隸屬度函數讓每個景氣指標都擁有一個介於 0 到 1 之間的值，來代表屬於景氣好的程度。相較於傳統集合的特徵函數，在描述模糊的概念時，利用模糊集合的隸屬度函數來解釋，是更爲恰當的。

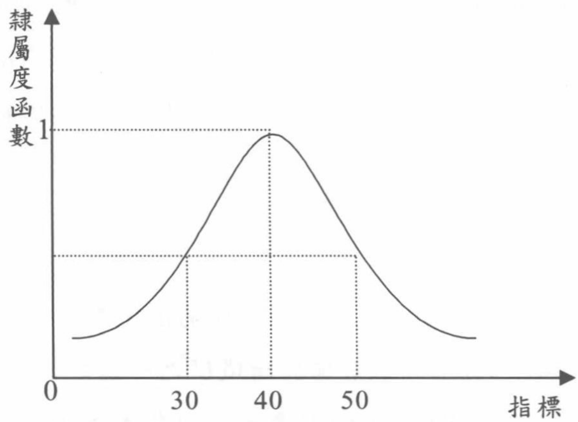

圖 2.2 模糊集合:景氣，的隸屬度函數

隸屬度函數是模糊理論最基本的概念，它不僅可以描述模糊集合的性質，更可以對模糊集合進行量化，並且利用精確的數學方法，來分析和處理人類模糊性的資訊。然而，要建立一個足以表達模糊概念的隸屬度函數，並不是一件容易的事。其原因在於隸屬度函數仍舊脫離不了個人的主觀意識，故沒有通用的定理或公式。一般而言，解決的辦法是根據經驗法則，或是利用以往的統計資料來輔助加以確定，很難像客觀事物一樣有很強的說服力。因此，隸屬度函數的建立經常最具有爭議性的，也沒有一種隸屬度函數是可以被廣泛接受而使用。

隸屬度函數可分爲離散型(discrete type)與連續型(continuous type)兩種。離散型的隸屬度函數是直接給予有限模糊集合內每個元素的隸屬度，並以向量的形式表現出來；而連續型隸屬度函數則有幾種常用的函數形式(S-函數、Z-函數、π-函數、三角形函數、梯形函數、高斯(鐘形)函數)來描述模糊集合。函數定義的表現，可以是無限模糊集合的元素及其隸屬度之間的關係，也可以是有限模糊集合的元素及其隸屬度之間的

關係。

傳統的統計方法透過一般的抽樣調查往往只能得到單一的數值資料、或是固定尺度的選擇，但如此並不足以能夠完整地反應人類個體的想法。若能讓受訪者根據自己的意識，利用隸屬度函數或區間值表達心中對於問項真正屬意的程度，則可更完整地傳達人類真實的思維。在考慮具有模糊特性的問項時，資料本身便具有不確定性與模糊性，所以我們先定義模糊數如下。

定義 2.1 模糊數

設U為一論域，令$\{A_1, A_2, ..., A_n\}$為論域U的因子集。u為一對應到$[0,1]$間的實數函數，即$u:U \to [0,1]$。假若佈於論域U之一述句X其相對於因子集的隸屬度函數以$\{\mu_1(X), \mu_2(X), ..., \mu_n(X)\}$表示，則在離散(discrete)的情形下，述句$X$的模糊數可表示成：

$$\mu_U(X) = \frac{\mu_1(X)}{A_1} + \frac{\mu_2(X)}{A_2} + ... + \frac{\mu_n(X)}{A_n}$$

其中+是或的意思，$\frac{u_i(X)}{A_i}$表示述句X隸屬於因子集A_i的程度。當U為連續時，述句X的模糊數可表示成：$\mu(X) = \int_{x \in X} \frac{u_i(X)}{A_i}$。

例 2.1 一天運動時間的模糊數表示

假設X為國中學生一天運動幾個小時，以模糊數表示為$\mu_\Omega(X)$，論域Ω可視為整數論域，即是運動時數。設$\Omega = \{0,1,2,3,4,5\}$，X:一天運動時間模糊數的隸屬度函數為

$$\{\mu_0(X) = 0.25, \mu_1(X) = 0.4, \mu_2(X) = 0.2, \mu_3(X) = 0.1, \mu_4(X) = 0, \mu_5 = 0\}.$$

則X:一天運動幾個小時的模糊數可表示為

$$\mu_U(X) = \frac{0.3}{0} + \frac{0.4}{1} + \frac{0.2}{2} + \frac{0.1}{3} + \frac{0}{4} + \frac{0}{5}.$$

例 2.2 居民對河川惡臭感覺的模糊數表示

假設 X＝高雄愛河沿岸居民對河川惡臭的感覺。以模糊數表示為 $\mu_U(X)$。假設論域 U={1=很嚴重, 2=嚴重, 3=普通, 4=輕度, 5=無影響}. 若 X＝高雄愛河沿岸居民對河川惡臭的感覺隸屬度函數為

$\{\mu_0(X)=0.25, \mu_1(X)=0.4, \mu_2(X)=0.2, \mu_3(X)=0.1, \mu_4(X)=0.05, \mu_5=0\}$;

亦可以模糊數表示為

$$\mu_U(X)=\frac{0.25}{1}+\frac{0.6}{2}+\frac{0.1}{3}+\frac{0.05}{4}+\frac{0}{5}.$$

例 2.3 投資報酬率的模糊數表示

假設 X 公司投資報酬率以模糊數表示為 $\mu_U(X)$，論域 U 可視為實數論域，即是投資報酬率。設 $U=\{2\%,5\%,10\%,20\%\}$，且令 X 公司投資報酬率模糊數的隸屬度函數為

$$\{\mu_2(X)=0.2, \mu_5(X)=0.4, \mu_{10}(X)=0.3, \mu_{20}(X)=0.1,\}$$

則 X 公司投資報酬率的模糊數可表示為

$$\mu_U(X)=\frac{0.2}{2\%}+\frac{0.4}{5\%}+\frac{0.3}{10\%}+\frac{0.1}{20\%}$$

在計算模糊樣本其相對於語言變數的隸屬度函數時，經常利用到三角形隸屬度函數及其他函數來進行計算的工作，其定義及計算方式如下：

定義 2.2 三角形隸屬度函數之計算

令 $\{X_i\}_{i=1}^n$ 為一模糊樣本，U 為其論域。給定 U 的一個次序分割集合，$\{P_j\}_{j=1}^r$，且其相對於語言變數為 $\{L_j\}_{j=1}^r$。設 m_j 為其分割集合 P_j 的中間值，若 X_i 介於 m_j 與 m_{j+1} 之間，則其屬於語言變數 L_j 的隸屬度為