

餐飲感官評估

目錄

自序

第一章 感官評估簡介

第二章 人類的感覺機制

第三章 品評試驗室、測試產品及品評小組的管理

第四章 影響感官判定的因子

第五章 測量反應

第六章 總體差異測試

第七章 屬性差異測試

第八章 感覺閾值和敏感度測試

第九章 品評員的選擇及訓練

第十章 描述性分析技巧

第十一章 喜好性測試：消費者測試及內部品評組接受度測試

第十二章 技術選擇及報告結果指南

第十三章 基本統計方法

第十四章 進階統計方法

第十五章 統計表

自序

餐飲乃感官評估中最受矚目之業種，常在大眾場舉行各種評鑑、比賽，甚至各電視台都有美食節目，在名廚、藝人之加持下，蔚然成為顯學，但始終以一二人之口味嗜好，判定美食與否。消費大眾耳濡目染，習以為常，人云

亦云，積非成是。

感官評估學應用於食品工業已數十年，結合於市調，使其信度和效度大為增加，但鮮用於餐飲。業界仍以名廚、名店、名劇、名人推崇其餐飲，忘卻餐飲之本質。

餐飲管理涉及諸多學科，但始終對感官品質眾說紛紜，本書嘗試以感官評估學為骨架，應用於餐飲實例，闡述餐飲之感官評估。拋磚引玉，不值識者之一哂。

第一章 感官評估簡介

1 Introduction

本章詳細說明餐飲品質的概念，為什麼要進行感官評估測試，並追述感官評估的發展歷史。接著介紹現代感官分析的基本方法，這些方法以品評員為評估工具，因此其可變性很大且可能有偏見的傾向，但是它們是用來檢測需要被檢測品項的唯一儀器，因此差異性必須降到最小，並儘量利用心理及心理物理學(psychophysics)的最佳現有技術來控制偏見。此外，簡介感官知覺鏈。最後，提供一套示範操作，以說明在 7 個實際步驟中如何使用這些技術。

1.1 餐飲的品質

近年來，餐飲有走向國際化的大趨勢，不但外國料理大舉入據台灣；台灣餐飲也走入東南亞、歐美、中東，甚至混搭各國料理特色，形成跨國料理。另一方面，道地的異國料理移植到台灣，有時也會有適應不良的情形出現，關鍵因素是消費者對品質的認知。品質是人們選擇餐飲的重要依據，但品質很難正確地定義，因它既代表餐飲的優良程度，也包含了餐飲所有的特性，這包括餐飲的重要特性及可接受性(acceptability)。

餐飲品質包括主觀方面及客觀方面。用餐環境、餐飲的外觀、質地結構及風味具有較大的主觀特性，而餐飲的營養及微生物品質則可藉由化學分析方法、檢測總菌數或使用其他特定方法客觀地檢測出來。

主觀特性包括用餐環境(用餐環境、品牌等)和餐飲的本質。其中，餐飲的本質包括：(1)**食物外觀**：餐飲的份量、形狀、顏色、結構、透明度或濁度、色調或色澤、完整性或破損程度，甚至形象包裝(盤飾等)。消費者在選擇、判斷餐飲品質時，或根據品質選擇餐飲時，會將這些因素作為品質指標。例如：擺飾整齊意謂食材清潔。水果的顏色代表其成熟度，顏色也代表濃度（如果汁、茶葉或咖啡）、烹煮程度、新鮮度或腐敗程度。消費者往往希望某種食物應呈現某種顏色，否則會認為是品質不良。同樣，在大小方面亦是如此，例如：消費者常選擇較大的雞蛋或較大的水產。食物結構對烘焙食品很重要。例如：整個麵包中氣室應該均勻分佈，而不是只在頂部有一個大的孔洞。濁度為飲料的重要指標，例如：柳橙汁因含有果肉呈混濁狀，而白葡萄汁應該是透明無沈澱，如果有沈澱物的話就

表示品質不良。同樣的牛排餐，在飯店的用餐環境也比夜市被認為高品質，甚至提高數倍價格。(2)**食物質地(口感)**：是指可藉由手指、舌頭、上顎或牙齒感覺到的餐飲特性。餐飲有不同的質地結構，例如薄脆餅乾或洋芋片、脆芹菜、硬糖、嫩肉排、巧克力夾心餅乾及牛奶霜淇淋。質地結構也是品質的指標。許多因素使餐飲質地結構在儲藏過程中改變。如果水果或蔬菜在貯藏過程中失去水分，則萎縮或使膨脹壓降低，一個原本青脆的蘋果會變得如同皮革般的堅硬，令人難以接受。麵包貯存後會變硬、失去新鮮度。如果冰箱溫度上下波動，冰淇淋等產品會溶化再冷凍，會因乳糖沉澱，冰晶變大，而導致產品呈砂砂狀。**口感(mouthfeel)**常用於描述口腔對餐飲感覺的質地結構特性。質地結構的主觀測試是對餐飲流變學特性的間接測試。例如：感官評估小組可藉由餐飲的濃度及口感來評估其黏度。不過，黏度也可由黏度計直接檢測出來。(3)**風味**：風味包含味道及氣味，有較強的主觀性。感冒常常使人覺得食物沒有味道，但這並非是由於味蕾而是由於嗅覺受到影響。味道是藉由舌尖、兩側及舌根的味蕾來偵測的，而香味是藉由鼻腔上部的嗅覺上皮偵測。對於任何食物來講，若要具有香味，就必須具有易揮發性物質，而易揮發性物質能夠以很低的量被偵測出來。香味是品質的重要指標。一般來講，變質的食物先產生異味然後外觀才開始變化，久放的肉類可藉由氣味很容易分辨出來。目前的研究，餐飲的基本味道有五種—鹹、甜、酸、苦、鮮(umami)。其他如脂肪味、鈣味等基本味正在研究中，其複雜性很難全面描述。「辣」則不是基本味，只是痛覺受器對食物中的刺激性成分(如辣椒素<capsaicin>)產生反應而引起舌頭、口腔、鼻腔黏膜、皮膚和三叉神經的一種疼痛的感覺，基本上屬於痛覺。

感官評估的定義是要求利用人的五感(視覺、聽覺、味覺、嗅覺、包含觸覺在內的皮膚感覺)或體感(深部感覺、平衡感覺、內臟感覺)來評估品質，而此種品質的評估作業稱為「感官評估技術」。可是，以往常使用「官能檢查」或「感官品評」(evaluation 翻為品評，乃前人之誤，以訛傳訛，已數十年。品評是判定詞，評估是估計詞)。

美國食品技師協會(Institute of Food Technologists)對感官評估(sensory evaluation)的定義是以科學方法藉著人的視、嗅、嚐、觸、聽等五種感覺，來測量食物或其他使用物品之性質的一門科學(IFT, 1975)。

1.2 感官評估測試的發展史

1.2.1 原始的感官分析

很早以來，人們就利用自身的感覺器官進行許多評估和判別工作，像香腸顏色和味道的評估、酒類品質優劣的劃分、茶的香

味比較等。因此，人類使用感官評估(sensory evaluation)已經很久了。評估(evaluation)可能接近事實，但不等於事實。擔任這種評估和判別工作的人員，是一些在某些方面具有敏銳的感覺和長年經驗累積的專家。在許多場合下，其評估結果具有絕對的權威性。當幾位專家的意見分歧時，往往採用少數服從多數的簡單方法決定最後的評估結果。這種方式被稱為**原始的感官分析**。

貿易的興起逐漸地啟動較正式的感官評估測試，買家希望以測試整批貨物的少量樣品來代表全部產品，賣家開始以貨品品質的評估為基礎來訂出價格。隨著時間的演進，有些酒、茶、咖啡、奶油、魚和肉類的分級方法被開發出來，其中某些方法到今日仍然被使用。

專業的品評員(專家)和顧問以分級(grading)的方法促進了 1900 年早期食品、餐飲和化妝品工業的萌芽，文獻上使用“感官評估”此一名詞的意思是指對感官屬性的假設性客觀測試，但實際上，測試經常是主觀的，因評估人數太少且解釋趨於偏見。

顯而易見，這種依據經驗和權威選擇品評人員進行感官分析的做法有很多弊端：第一、由專家擔任品評員(或稱為評估員)，只能是少數幾人，而且不易召集；第二、不同的人具有不同的感覺敏感性、嗜好和品評標準，所以，許多品評員對同一物品進行品評，往往得不到一致的結果；第三、人的感官的狀態(如出現疲勞或生病)和環境條件的變化經常影響感官分析的結果；第四、每個人具有的感情傾向和利益衝突，有時使得評估結果出現偏差，甚至做假；第五、專家對物品的評估標準，與消費大眾的實際看法可能有差異。

1.2.2 現代的感官分析

由於上述問題的存在，使得原始感官分析方法的可信度令人懷疑。為了克服這種不足，在感官分析試驗中逐漸引入了生理學、心理學和統計學方面的研究成果，盡量避免原始感官分析中存在的各種缺陷，從而發展成為今天的**現代感官分析**，此外，電腦技術的發展也影響和推動感官分析學的發展。在品評員的選擇、試驗環境的佈置、試驗方案的設定、結果的處理等方面，不再依靠經驗和權威，而是依靠**科學**。系統性“感官評估”分析的歷史，是奠基於提供合格食物給美軍的戰時工作和在 Scandinavia 發展的三角試驗法。感官評估測試方法主要是由位於美國 Davis 的加州大學食品科學系所開發。

1.2.2.1 引入統計學方法

英國著名的推測統計學家 R . A . Fisher 在 1935 年所著的《實驗計劃法》一章中，記述了一個與感官分析有關的試驗。這是首次將統計學方法應用在感官分析中的例子。當時，英國有一位妃女自稱可以分辨出奶茶中的紅茶和牛奶是哪一種先加。為此，

R . A . Fisher 設計了一個方案驗證她的說法。它沖了 8 杯奶茶，其中 4 杯是先加紅茶後加牛奶，另外 4 杯順序相反。然後隨機遞給那位妃女，並預先告訴她加入順序不同的奶茶各是 4 杯，要求她分出各自相同的 2 組。試驗結果表明，這位妃女時並不具備自稱的那種分辨能力，因為在調配 70 次的試驗中，她僅分對 1 次，正確率為 1.4% ，所以，即使分對了，也可以認為是偶然所致。這種試驗方法現在稱為類別檢驗。但是，Fisher 舉此例的目的是為了闡明歸零假設與有意性檢驗等統計學上的重要概念。真正把統計學的方法應用於感官分析的首推 S . Keber。他在 1936 年，首次採用二點試驗法，感官檢查了肉的嫩度。

1941 年，美國某工廠老闆 Whisky 在產品的出場檢查中，試行了 1：2 識別法。這是首次將感官分析方法應用於品質管理的實例。此後，感官分析方法逐漸進入工業生產的各領域，得到廣泛應用和飛速發展。

感官分析作為一種以人的感覺為測定手段或測定對象的方法，誤差是不可避免的。但是，引入統計學的方法後，可以合理有效地糾正誤差帶來的影響，並且使感官分析法成為一種有說服力的科學測定方法。

1.2.2.2 引入心理學的方法

心理學具有悠久的歷史，發展至今至少有二千多年了。但在很長的一段時間內，心理學一直作為哲學的一部分使用。它成為一門獨立的學科是在十九世紀後期。在感官分析中，引入了許多心理學的內容。當然，感官分析與心理學的研究目的迥然不同，但心理學的許多測定技術可以直接應用於感官分析。心理學測定技術的發展，是近百年的事情，特別是最近數十年間，隨著統計學和電腦的進步，有了十分顯著的發展。

心理學測定方法從歷史上可以分為三種：精神物理方法、實驗美學方法和測驗法。

（一）精神物理方法

德國 G . T . Fechner(1801~1887)是現代實驗心理學的鼻祖，他在 1860 年所著的《精神物理學要論》中，描述了精神物理的測定方法。這種方法即是測定刺激與反應之間數學關係的方法。其中典型的方法有：調整法、權限法、恆常法。這種方法是用於測定精神物理中的常數。常數包括刺激閾、辨別閾和等價刺激三種。在感官分析領域中，經常接觸到刺激閾、辨別閾和等價刺激問題，也經常使用權限法和恆常法。

（二）實驗美學方法

Fechner 最初的實驗美學方法，是研究人類的眼睛在受到刺激時，什麼形狀能夠引起最美好感覺的問題。其最偉大的研究成果之一，是發現了矩形的黃金分割法。且沿用至今。Fechner 的貢獻在於提供用物理尺度衡量具有美感物質的方法，

為感覺數值化邁出了第一步。

作為實驗美學的方法，有選擇法、產出法和使用法。後來，經人們不斷改進，出現了順序法、成對比較法、評比尺度法等。而且，評定對象也從最初的美學研究對象，擴大到人的愛好、態度和興趣等，特別是成為感官分析中重要方法之一。

（三）測驗法

科學心理學的另一流派是研究人類個體之間的差異。著名學者達爾文的堂兄 F. 戈爾多(1822~1911)受達爾文進化論的影響，首創了優生學。它從優生學角度，設計了許多試驗方法，對個體差進行了系統研究，這是測驗法的開始。此後在許多領域中逐漸得到發展和應用。如智商測定、人格檢查以及各種能力測試等。由此發展而來的試驗法的思維方式和技術，在感官分析中，可以用於測定品評員的能力，檢查試驗的可靠性等。

1.2.2.3 引入生理學的方法

人類對於外界適宜刺激均有愉快或不悅的感覺。在產生感覺的同時。脈搏、呼吸、血壓、腦波；心電圖、眼球等身體各器官都有某些變動。

1874 年，W. Bund 出版了《生理學的心理學綱要》，創建了使用生理學實驗方法的心理學體系。1879 年，在萊比錫大學設立了世界上第一個生理學試驗室，開始對人體各器官的變化與外界刺激之間關係的研究。

隨著研究的不斷深入和製造業的發展，近代出現了可以測定刺激與感覺之間相互關係的儀器（如測謊器），即感覺可以通過儀器直接顯示出來。

最近，對味覺和嗅覺的電生理學研究方法的進步十分引人注目。例如：用味覺物質刺激動物的舌頭時，連接在味覺神經上的傳感器，可以將刺激結果轉換成電信號紀錄下來。有數據表明，通過動物末梢神經傳出的反應，與人類的感官分析結果在許多方面是一致的。

可以說，人類的感覺器官對應於不同刺激，具有不同的生理變化。把這些生理變化通過電訊紀錄下來，可以防止某些品評員為了某種目的的有時產生的撒謊現象。如同測量瞳孔大小的方法確定樓道內的照明度是否合適；用測定眼球電位來品評汽車前燈是否晃眼等。目前，這些方法尚在研究階段，要進入實用階段還需要摸索很長時間，但毫無疑問，在感官分析中利用生理學的方法，是感官分析的一個發展方向。

1.2.2.4 電腦技術

自從 1946 年世界上誕生第一台電腦 ENIAC 以來，至今僅有四十多年的歷史，但電腦在自身結構和應用範圍上的發展速度十分驚人。電腦的發展經歷了電子管、晶體管、中小規模積體電路和大規模積體電路四個階段。積體電路的應用和發展，使得電腦的運算速度大幅提高。在體積縮小，價格降低的同時，功能卻加強和擴充，從最初的純科學計算發展到今天的智能模擬、系統模擬和設計。電腦已在研發、生產、事務處理等各個領域發揮愈來愈大的作用。

在感官分析領域中，電腦技術的應用和發展也是不容忽視的，它包括二個方向：

（一）利用電腦處理分析結果。這首先需要編輯程序，不同方法編成不同程序。編程序對於不了解電腦語言的人來說很困難，但對資訊管理員來說，編輯一個用於處理數據的程式是輕而易舉的工作。感官分析試驗的管理者者可以隨時調用任何一個編好的程式。每次試驗後，只需將每個品評員的姓名、結果等有關情報輸入電腦，電腦就能自動將零散數據予以分類、排列、計算，並得出結論。然後，根據管理者的要求，印出有關試驗的分析結果報告單或其它需要了解的內容。

（二）在感官試驗室中使用。由管理者控制一台電腦，在每個品評員的座位前連結安裝一個電腦終端，形成一個小型電腦網路。管理者通過電腦提示給品評員有關試驗的各項內容和要求，品評員則通過終端把分析結果通知組織者，同時也可以向管理者提出問題。試驗結束後，記算機寫上就可以輸出試驗結果。這些結果同時被貯存在電腦的硬碟中，可供管理者隨時瞭解每次試驗的結果，或每個品評員以往的工作成績。

科學家近來已將感官評估測試發展為一科正式的、有架構的和經系統編輯的方法學，並繼續發展出新的及更新的現存方法，感官評估技術的現況已被記錄在專業期刊：化學感官(Chemical Senses)、感官評估研究期刊(Journal of Sensory Studies)、組織研究期刊(Journal of Texture Studies)以及 Pangborn 評論會（三年一次）和國際感官韻律團體（每兩年一次）的公報中，後兩者經常發表專論在食品品質和偏好期刊(Journal of Food Quality and Preference)和 Weurman 專題討論會公報(Proceedings of the Weurman Symposia)（每三年一次，但以書的方式出版）等期刊。發表在食品技師協會（Institute of Food Technologists）的感官評估文章經常被刊在 IFT 的食品科學期刊(Journal of Food Science)或食品技術(Food Technology)中。由於現在代感官評估學的發展始於歐美，充滿西方人的見解與認知，在台灣的推廣往往有格格不入之感，亞洲各因素的考慮近十幾年才逐漸加入。

已開發的方法創造了經濟利益，感官評估測試可以建立商品的價值及其可接受性，感官評估測試評估各種替代方式來代替適用金錢價值的品項。感官評估技巧主要使用在品質管制、新產品開發與研究，它們不但可應用在餐飲和加工食品的特性驗證及評估，也可應用在其它領域如環境的氣味、個人衛生用品、疾病診斷、純化學品的測試等。感官評估測試的主要功能是導入正確和

可信的測試方法，以提供數據而做出正確的決策。

1.3 以人的感覺作為儀器

可靠的感官評估分析是以感官評估分析師的技巧為基礎，使四項因子最適化，我們都知道這些重要因子，因為都適用於各種方法。

1. 定義問題：要評估什麼問題必須明確地定義，重要的是在“嚴格的”科學下，對感官和感情更要明確定義。
2. 測試設計：設計不但必須沒有遺留主觀的空間和考慮偏差的來源，而且必須將為獲得正確結果的測試數量減到最小量。
3. 儀器化：受測者必須經選擇並加以訓練以達到具有再現性的判定。分析師必須與受測者們一起工作，直到受測者知道在既定條件下本身的敏感度和偏差。
4. 解釋結果：使用統計學，分析師選擇正確零假設(null hypothesis)和正確的替代假設(alternative hypothesis)，並下定結果有保證的結論。

品評員如同測試儀器，因而（1）隨著時間的變化有很大的差異、（2）各品評員之間有很大的差異和（3）有很高的偏差傾向。適當地考慮這些缺點，需要（1）重複檢測、（2）提供足夠的受測者（通常 20-50 人）使判定具有代表性和（3）感官評估分析師尊重管理品評小組意見的許多規則和陷阱（參見第 4 章）。受測者對 2-10 個以上因子的敏感性是天生有差異的，且在進行一個專題時中途不應該換手，受測者必須以其敏感性被選擇並給予一再的訓練（參見第 9 章），直到他們完全了解即將進行的工作。感官評估測試的活動充滿著不可信的結果，因為許多品評員不了解該測試所使用的問題和（或）術語，不知道產品的風味或組織參數，或對測試技巧或所用數值的表達覺得不適。

為了這些原因和其他因素，感官評估分析師主動參與用來檢測品評員回應的尺度（scales,或譯為評量）之開發是很重要的。好的尺度需要好好學習，必須徹底加以了解，以管理有問題的感官變數的物理和化學因子為基礎，和需要許多參考點及品評小組的徹底訓練。期望一位有經驗的品評員擁有發展出一致性的正確與精準的尺度所需的知識和技巧是不合理的，唯有透過在尺度開發時有知識的感官評估專家的直接參與才能獲得描述性的分析，例如：使他們在六個月內所做的評估工作將和今天所做的一樣有再現性。

1.4 感官知覺鏈(The Chain of Sensory Perception)

當感官評估分析師研究既定的物理刺激和其回應之間的關係時，所得結果通常被認為是一個步驟(one-step)的過程，事實上這過程至少有 3 個步驟，刺激(stimulus)觸動感覺器官後轉換成傳達到大腦的神經信號，因為在記憶中存在著之前的經驗，經大腦解釋、組織和整合所接到的感覺(sensation)轉為知覺(Perception)，最後以受體的知覺為基礎產生回應(response)，如圖 1-1。

在處理人們對相同的刺激卻時常得到不同的回應的事實時，感官評估專家需要了解兩個人之間的判斷不同，可能是因其器官的敏感度不同，他們收到的感覺不同或其知覺的心理處理不同所造成，例如因為缺乏特殊的氣味、味道等知識或因為缺乏以語言和數字來表達感覺的訓練。透過訓練和運用參考樣品，感官評估專家可以塑造心理過程，使得受測試者在接到既定刺激時可顯示相同的回應。

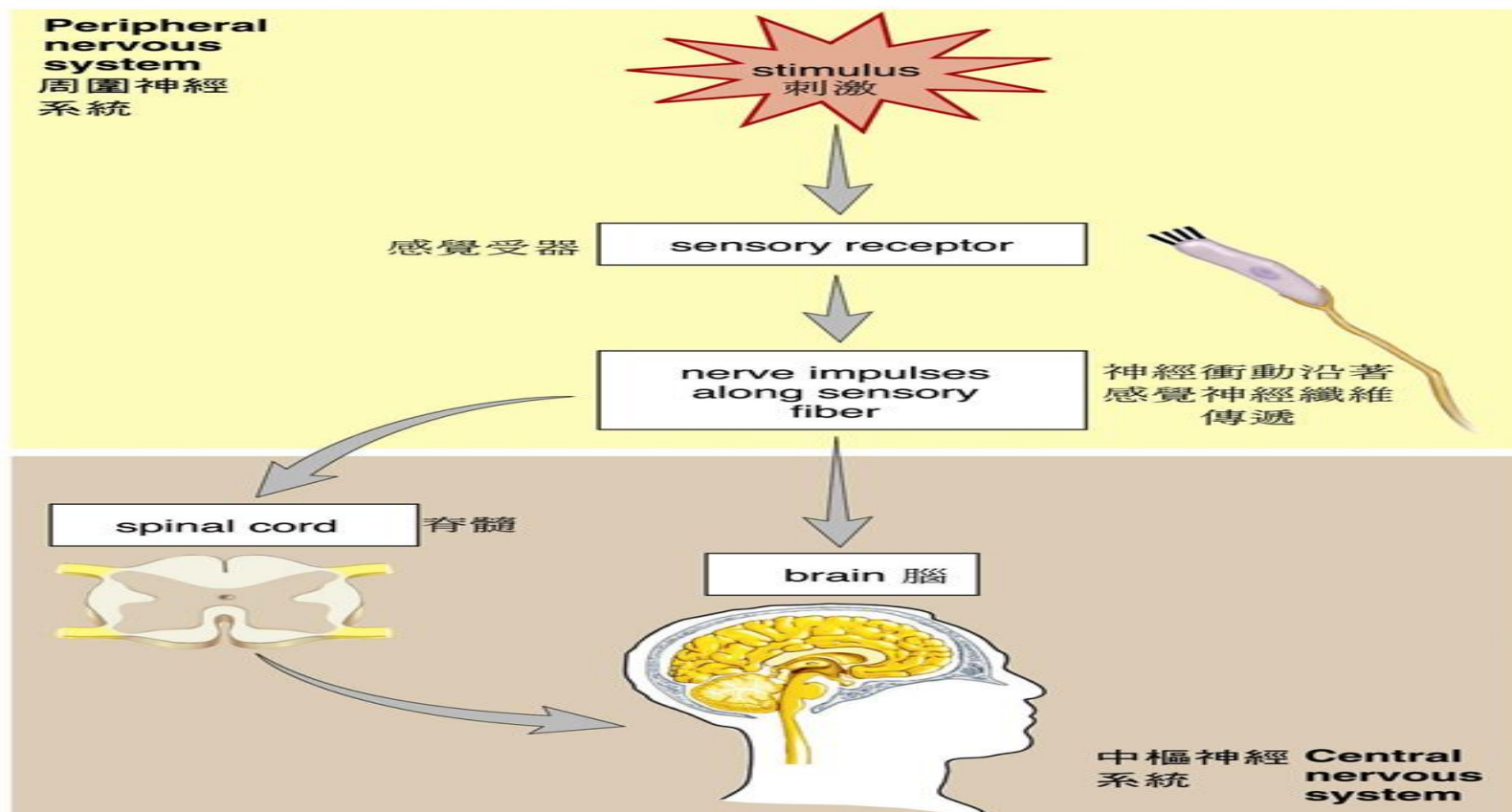


圖 1.1 感官知覺鏈

1.5 進行感官評估研究的實務

開發出最好產品的組織大多是由感官評估專家，而非特定測試服務的提供者所組成的。唯有透過全程參與，才能使他(她)在整項研究計畫生涯中的每一刻都了解什麼測試才是必需且適當的。感官專家(如統計學家)在開發研究計畫中必須扮演積極的角色，和參與實驗設計開發的其他團體(如研發、品管等人員)共同合作，最終才能在所引發的問題中幫上忙。Erhardt 將感官分析區分為以下七項實務：

1. 決定計畫(項目)目標:定義計劃領導人的需求(計畫目標)是完成一項正確測試的重要關鍵。樣品是當作改良產品被提交，還是當作降低成本或取代原料的樣品被呈現，或是被定位為一項能與競爭對手匹敵的產品?它是被期待和其他樣品相似還是不同?有優先順序還是平分秋色?具備單一特性還是多元屬性?如果這關鍵的一步沒有徹底執行，感官評估師不可能做出恰當的測試且正確地解讀資訊。
2. 決定測試主題:一旦計畫的目標清楚地豎立，感官評估師及計畫領導人便能決定測試主題:整體差異、屬性差異、相對偏好、可接受性等。避免在單一測試中企圖回答太多問題。好的方法是要能為感官評估師及計畫領導人在測試前便寫下計畫目標、測試主題，以及簡要地說明測試結果將會做怎樣的用途。
3. 篩選樣品:在計畫目標和測試主題的討論期間，感官評估師必須審視測試樣品的所有感官特性。這使感官評估師能夠透過測試方法考慮到所有樣品所引起的感官偏差。例如:目視方法(顏色、厚度、光澤)可能影響整體差異反應，如以三角測試法檢測蔗糖與阿斯巴甜的甜度差異。在此種情況下，屬性測試將會較合適。此外，樣品篩選也可提供評分表某些可用的詞彙資訊。
4. 設計測試:在決定了計畫目標、測試主題以及篩選樣品後，感官評估師可著手進行設計測試。這包括了測試技術的選擇(參見第六~八章，第十到十二章及第十四章)；品評人員的挑選與訓練(參見第九章)；設計伴隨而來的評分表(投票、問卷)；將樣品的準備介紹程序標準化(參見第三章)；及決定資訊該如何分析(參見第十三章及第十四章)。每一步驟都須小心，並堅持實驗的統計設計原則，以確保對受測物做最靈敏的評估。
5. 執行測試:即使是由技術人員執行測試，感官評估師仍負責確認測試設計所有要素都符合要求。
6. 分析資訊:由於分析資訊的程序是在設計測試時便已決定，所需的專業及統計程式將可在調查完成後隨即被用來展開資訊分析。資訊不但應本著主要論述的意義(測試主題)進行分析，還需考慮其他測試差異，如呈現順序、一天的次數、不同的天數，以及受測者的差異，如年齡、性別、地理區域等。
7. 解釋及回報結果:計畫目標及測試主題的最初簡潔敘述，可使感官評估師能回顧(review)測試結果，並做出有根據的後續行動建議。

後者應以書面簡潔清楚地總結資訊、確認樣品、陳述受測者人數及資格(參見第十六章)。

本書的主要目的是為了幫助感官評估者在受過訓練或有經驗的品評員的協助下，發展在分析感官評估測試上所需的方法學、受測者組織、設備及測試管制。除此之外，第十一章討論了消費者測試的組織，亦即使用經驗不足的消費者(非分析人員)進行大規模的評估，構成足以代表產品市場大量人口的消費及反應。

感官評估的角色是提供有效且可信賴的資料給研發部、生產部及行銷部，以便經理人根據產品特性做出穩當的商業決策。任何感官評估程序的終極目標皆是找出最符合成本效益且有效的方法以來獲得感官資訊。若是可以，公司內部實驗及描述性方法也能取代昂貴又耗時的消費者測試。更省成本的做法還可在儀器、物理、化學的分析上，儘可能地連結相似的感官特性。在某些例子裡還可利用便宜又迅速的儀器分析取代部分例行的感官評估。

第二章 人類的感覺機制

2 Mechanism of Human Sensation

2.1 簡介

這一章複習了(1)本書所討論的各種感官屬性，如：不同產品的外觀、氣味、風味和感覺，以及(2)人們用來察覺這些屬性的機制，如：視覺、嗅覺、味覺、觸(動)覺。礙於版面的限制，本章只能對此主題做簡要的說明。感官評估專家應更進一步參考其他出版物，以求在感官評估上建立一本好的書籍和期刊資料庫。感官測試不是一科嚴謹的科學，因此實驗的設計需要以隱藏在感官屬性的物理和化學因子的整體知識為基礎。原則上，感官測試的結果可以有很多種解釋，而我們對人類感官及產品屬性的每一新認識都將減少可能的誤解機會。

2.2 感官屬性

感官評估同時兼具有感知(sensation)與知覺(perception)的作用，感知是生理上的反應，而知覺的形成則主要是以心理為主，且感知是知覺的基礎，知覺具有的個別差異又比感知部分受到許多個體種種身心狀態的影響。

不同的刺激由不同的感受(知)器接收，再經由不同的神經系統傳導。不論哪種系統，除了刺激的質(quality)(品質)，通常還會有量(quantity)(大小)、空間(距離)和時間(持久性)等訊息種類。體內神經衝動的電位都很微弱，通常在 10 毫伏特左右。食品品項的屬性通常以下順序被察覺：

- 外觀
- 氣味/香味/香氣
- 硬度及質地
- 風味(香味、化學感受、味覺)

然而在被察覺的過程中，大多的屬性是互相重疊的。這句話的意思即是：對幾乎是同步接收所有感官印象的測試對象而言，它接收到的只是一團混亂，而沒有經過訓練的品評員無法對任一項感官做出獨立的評估。感官屬性乃是在被察覺且被接受的條件下才能存在的特殊性質，本節提供了一些典型的範例。

在本書中所提及的風味，是指食物在口中經由化學感受而察覺到的綜合印象，也就是說它不包含外觀和結構。而香氣一詞則指的是在口中由食物產生的揮發性成分，嗅覺系統經由鼻孔後端察覺而得知。

2.2.1 外觀

如同每一位消費者所知，產品的外觀及包裝通常是唯一被用來決定是否購買與消費的屬性。因此，人們傾向於從瑣碎的線索中做出大膽廣泛的推論，測試對象也會在品評小室(booth)內做出同樣的舉動。所以品評員必須小心翼翼的看待測試樣品的每一項外觀，並試圖消除或模糊樣品多彩顏色或不透光容器的偽裝等等。

一般外觀特性列舉如下，在附件 11.1A 有藉助於評量而描述外觀的例子。

顏色 融合了物理及心理兩者所構成的現象：視覺系統對光的波長為

400—500nm(藍色)、500—600nm(綠色和黃色)、以及 600—800 nm(紅色)的感知，通常以 Munsell 表色彩中的彩度、明暗度、色度來表示。上色均勻與否、乾淨或骯髒也都是很重要的。食物腐敗通常伴隨而來的便是顏色的轉變。Clydesdale、McDougall、Lawless 和 Heymann 在關於外觀及色澤的感官評估方面已提出很好的描述。

形狀和大小 長度、厚度、寬度、顆粒大小、幾何形狀（方形、圓形等）、部位的分佈，如：蔬菜、餅皮、經調製食物等、大小及形狀，這些都是缺陷的指標（Kramer 和 Twigg 1973、Gatchalian 1981）。

表面質地 表面是暗晦或明亮，粗糙或平滑？表面是溼的或乾的，軟的還是硬的，脆的或堅固的？

清澈度 透明液體或固體的矇矓(Siebert、Stenroos 和 Reid 1981)或昏暗(McDougall 1988)，可見粒子的存在與否？

充碳酸氣 碳酸飲料倒出時的冒泡程度。一般是用 Zahm-Nagel 設備來測量，可用下述各項來判斷：

充碳酸氣(體積)	充碳酸氣(%，重量百分比)	冒泡程度	範例
1.5 或 1.5 以下	0.27 或更少	沒有	蒸餾飲料

1.5-2.0	0.27-0.36	輕微	水果飲料
2.0-3.0	0.36-0.54	中等	啤酒、西打
3.0-4.0	0.54-0.72	高	清涼飲料、香檳

2.2.2 氣味/香味/香氣

當產品的揮發物質進入鼻腔透過嗅覺系統被察覺，稱為此產品的氣味。氣味的討論是在揮發物質透過鼻子被聞、嗅到的階段（自發的或其他）。香味指的是食物發出的氣味，而香氣則是香水或化妝品發出的氣味。前面提過，香味是指口中的物品所含的揮發物質被嗅覺系統所察覺而得知。

揮發物質從物品逸散的程度受溫度及組成元素的本質影響。物質的氣體壓力以下列的公式，隨溫度呈現指數性增加：

$$\log p = -0.05223a/T + b$$

p 是氣體的壓力，單位 mmHg， T 是絕對溫度 ($T = t^{\circ}\text{C} + 273.1$)， a 和 b 則是物質常數，可在參考書籍 (Howard 1996) 中找到。揮發性也受表面狀況影響：在加溫的狀況下，柔軟、多孔、潮濕的表面較堅硬、光滑、乾燥的表面容易揮發。

許多的氣味只會在酵素反應在新鮮切口部份發生時釋放（如洋蔥的氣味），氣味分子只能藉著大氣、水氣或工業氣體傳遞，而被察覺的氣味密度是由其在空氣中所佔比例，及嗅覺受器接觸的空氣量多寡而決定。

感官評估專家持續地嘗試香氣種類，使其成為可識別詞彙（參見第十章）。但此刻在描述氣味的術語上仍無任何國際標準術語。本領域非常廣大；已知的氣味大約有一萬七千種，而一位優秀的品香師可分辨出一百五十至兩百種氣味品質。有許多詞彙被歸於單一化合物（麝香草酚＝類香料、綠色的、類橡皮），也有單一詞彙被用來形容多種化合物（檸檬＝ α －松油精、 β －松油精、 α －檸檬油精、檸檬醛、香茅醛、迦羅木醇、 α －香瑯油等）。

2.2.3 黏稠與質地

第三項要被討論的屬性也是利用嘴巴這個受器來察覺，但卻不是味覺及其他化學感受。質地則是在評估個人用品或家用產品時，以

身體的皮膚和肌肉而非嘴巴所察覺的屬性。

依慣例，可參考以下所列舉的：

- 黏度(viscosity)(對均勻的牛頓流體而言)
- 黏稠(consistency)(對非牛頓或不均勻的流體學說及半固體而言)
- 質地(texture)(對固體或半固體而言)

“黏度(viscosity)”意謂流體在某力(force)(如重力)下的流動速率，這可被精確量測，從水或啤酒的約 1cP (centipoises，厘泊)到啤酒的膠狀產品的

1000s cP。液狀濃湯、醬汁、汁、果醬、果凍和化妝品的黏稠(consistency)原則上可用感官評估來測量，但實際上可能要借助黏度計幫忙。質地(texture)更複雜，如質地研究期刊(Journal of Texture Studies)所示。質地可被定義為產品組織或內部組成的感官呈現。

• 對應力(stress)的反應，可用手、指、舌、頷或唇的動覺感(kinesthetic sense)量測其機械性質(如硬度< hardness/firmness>、附著性< adhesiveness>、黏合性、膠著性< gumminess>、彈性< springiness/resilience>、黏度< viscosity>)。

• 觸覺可用手、唇或舌的肌肉觸覺神經量測其幾何學微粒(geometrical particle)(粒狀、砂礫感、結晶、片狀)或水感(濕潤、油性、水水的、乾)。

表 2.1 列出食品、護膚產品和織品的一般機械、幾何學和濕潤性。要注意的是涵蓋如此廣泛多樣的產品，質地性(textural property)是以動覺或觸知(tactile-wise)量測而衍生自質地名詞的同一一般類別。其他食品質地名詞列餘地 11 章、附錄 11.2C、11.2D 和 11.3。

2.2.4 風味(Flavor)

風味被當作食物、飲料和調料的一種屬性，已被定義為來自感覺末端的刺激的知覺總和，該末端與消化道和呼吸道的入口群集在一起，但實際感官分析的目的，作者偏好依照 Caul (1957)且限制其名詞於“一食品在嘴內經化學感產生的感覺(the impressions perceived via the chemical senses from a product in the mouth)”。依照此定義，風味包括：

香味(aromatics), 意即經由鼻後部從嘴中一產品釋出的揮發物質引起的嗅覺知覺。

味道(tastes), 意即經由鼻後部從嘴中可溶性物質引起的味覺知覺(鹹、甜、酸、苦和鮮)。

化學感覺因子因刺激口鼻孔的軟膜神經末端而產生(收斂、調味品熱、冷、刺痛、金屬味、鮮味)。

2.2.5 聲響(Noise)

食品咀嚼時或結構處理所產生的聲響是微小但不可忽略的感官屬性。通常是測量由食物或構造物所產生的聲音的音高(pitch)、響度(loundness)和持久(persistence)。聲音的音高和響度歸因於整體的感官印象(sensory impression)。某些脆裂食品(酥脆、易脆、脆的)的音高差異提供感官輸入,我們用來評估新鮮度或走味。Vickers and Bourne 以示波(oscilloscopic)測量允許以酥脆(crispy)描述的產品和以易碎(crunchy)描述的產品間的尖銳差異。這些動覺的(kinesthetical)差異對應依產品的硬度、濃厚和脆裂的可測量差異。

2.2.6 餘味(Aftertaste)

餘味是刺激離開口腔後所殘留的味覺、嗅覺或觸覺。通常餐飲皆有易殘留或遲滯的物質,如甘味、鹹味、苦味、澀味和鮮味等。餘味的呈現有其時間序列性,時間愈近,愈會形成口感的層次感。若時間愈遠,則形成回味現象。

2.3 感覺閾值(Threshold)

感覺閾值是指從剛能引起感覺到剛好不能引起感覺刺激強度的一個範圍。敏感性是感覺器官對刺激的感受、識別和分辨能力。敏感性亦稱感受性。把剛剛能引起感覺最小的刺激量稱為絕對閾值的下限,又稱刺激閾或察覺閾。低於下限的刺激稱為閾下刺激。剛剛導致感覺消失的最大刺激量,稱為絕對感覺閾值的上限,高於上限的刺激稱為閾上刺激。

剛剛能引起差別感覺刺激的最小變化量,叫做差別(異)感覺閾值或差別(異)閾。差別閾值會隨原來刺激量的變化而變化,即差別閾與刺激量的比值是個常數,如以 K 代表常數,則

$$K = \Delta I / I$$

其中 ΔI 代表差別閾值, I 代表刺激量(刺激強度)。常數 K 又稱為韋伯分數。此公式就是韋伯定律。

費希納(G. H. Fechner)發現感覺的大小和刺激強度的對數成正比

$$S = K \log R,$$

其中 S 為感覺, R 為刺激強度, K 為常數。此規律被稱為費希納定律。

後來的許多實驗證明，韋伯定律和費希納定律只適合於中等強度的刺激，當刺激強度接近絕對閾值時，韋伯比值(常數)則大於中等強度刺激的比值。表 2-1 為常見味道與香氣的化合物識別閾值。

表 2-1 常見味道與香氣的化合物識別閾值。

表 3-4 常見味道與香氣之化合物的識別閾值

屬性		化合物	識別閾值(g/L)
味道	鹹味	食鹽(氯化鈉 NaCl)	1.19
	甜味	砂糖(蔗糖 $C_{12}H_{22}O_{11}$)	5.76
	酸味	檸檬酸(citric acid $C_6H_8O_7$)	0.43
	苦味	奎寧(quinine $C_{20}H_{24}N_2O_2$)	0.195
	鮮味	麩胺酸鈉(monosodium glutamate $C_5H_8NNaO_4$)	0.595
香氣	新鮮檸檬味	檸檬醛(citral)	0.01
	玫瑰味	香葉醇(geraniol)	0.01
	青草味	葉醇(順-3-己烯-1-醇 $cis\text{-}3\text{-hexen-1-ol}$)	0.05
	杏仁味	苯甲醛(benzaldehyde)	0.05
	香蕉味	丁酸乙酯(ethyl butanoate)	0.005
	酸敗或起司味	丁酸(butyric acid)	0.01
	丁香味	丁香酚(eugenol)	0.005
	馬鈴薯泥味或烤肉味	3-甲硫基丙醛(methional)	0.01

2.4 感覺機制

2.4.1 視覺

人類在認識世界，獲取知識的過程中，90% 的訊息是靠視覺提供的。在感官分析中，視覺檢查佔有重要的位置，幾乎所有產品的檢查都離不開視覺檢查。在市場銷售的產品能否得到消費者的歡迎，往往取絕於“第一印象”，即視覺印象。

一、視覺的生理特點

視覺的產生依賴於視覺的適宜刺激和視覺的生理機制。視覺的適宜刺激為波長 380~780nm 的電磁波。這部分電磁波又叫光波，屬可見光部分，它僅占全部電磁波的 1/70。可見光分為兩類：一類是由發光體直接發射出來，如太陽光，燈光等；另一類是光源照射到物體表面，由反光體把光反射出來。我們平常所見的光多數是反射光。在完全缺乏光源的環境下，就不會產生視覺。視覺的生理機制是光源光或反射光刺激於晶狀體，光線經過晶狀體的折射，在視網膜上形成物象。物象刺激視網膜上的感光細胞，使細胞產生的神經衝動沿視神經傳入大腦皮層的視覺中樞，最後產生視覺。

二、視覺的敏感性

在不同的光照條件下，眼睛對被觀察物的感受性(即敏感性)是不同的。在明亮光線的作用下，人眼可以看清物體的外型和細小的地方，並能分辨出不同的顏色。所以，感官分析中的視覺檢查應在相同的光照條件下進行，特別是同一次試驗過程中的樣品檢查。

三、顏色對食品的感官評估有以下作用

1. 顏色本身即為食品品質的表徵，顏色對食品的挑選有很大的影響。
2. 人們依經驗對食物本有顏色或習慣上的認知會成為判斷食品特性與新鮮度之依據。
3. 食物的顏色會影響人的食慾。
4. 人們對顏色有許多既定的偏好，有些食物較受歡迎是因為顏色較令人愉悅，但有些食物卻因顏色不被喜愛而失去吸引力。

四、視覺的感官檢查

感官檢查順序中首先由視覺判斷物體的外觀，確定物體的外形、色澤。通過視覺檢查還可知產品的品質，如腌臘肉的脂肪變黃，則說明脂肪已氧化酸敗。麵包和糕點的烘烤也可通過視覺控制烘烤時間和溫度。隨著科學技術的發展，有些外觀指標可以由儀器測定而獲得控制。如香腸的顏色就可以用儀器測定，而何種顏色的香腸可增加人的食慾，能受到人們的喜愛，這是儀器所不能代替的。視覺檢查在生產過程中佔有很重要的地位。

2.4.2 聽覺

一、聽覺的生理特點

頻率在 16~20000Hz 的聲波是人耳的適宜刺激。聲波是物體振動所產生的一種縱波。聲波的振幅是決定聲音強弱的特性。聲音的強弱是一個客觀物理量，可用聲壓，聲壓級等度量，一般以分貝(dB)表示。響度是由聲波的物理強度(振幅)決定的。外界的聲波經外耳傳入鼓膜，引起鼓膜振動，而刺激耳朵內聽覺感受器，使聽覺感受器產生神經衝動，最後傳入大腦皮層的聽覺中樞形成聽覺。聲波必須借助於氣體、液體或固體的媒介物才能傳播。頻率是指聲波每秒振動的次數。頻率不同，給人的音感也不一樣，是決定音高的主要因素之一，同時，聲音強度對音高也有一定影響。但是，響度和強度並不完全相當，強度相同而頻率不同的聲音，人感覺到的響度可能不一樣。不同頻率和振幅的純音(正弦波)混合可以產生一切聲音。音樂聲、噪聲、音語聲等都是不同純音的混合。因此，每種聲音都有自己的波形。外界的聲波經外耳傳入鼓膜，引起鼓膜振動，而刺激耳朵內聽覺感受器，使聽覺感受器產生神經衝動，最後傳入大腦皮層的聽覺中樞形成聽覺。

二、聽覺的敏感性

聽覺的敏感性是指人的聽力，即對聲波的音調和響度的感受能力。正常聽覺的成人可聽到的聲音在 16~20000 Hz。18~28 歲是聽力的最佳年齡。人耳對一個聲音的強度或頻率的微小變化是很敏感的。實驗表明，純音強度差別閾隨刺激強度的增加而降低。1000Hz 的純音強度為 20 dB 時，差別閾為 1.5 dB；在 40 dB 時，差別閾降至 0.7 dB。

三、聲音對食品的感官評估有以下作用

用餐時播放不同的音樂，可以使用餐的感覺產生不同的變化，不同形式的音樂背景可以讓用餐體驗帶入不同的氛圍與情境。

三、聽覺的感官檢查

利用聽覺進行感官檢查的應用範圍十分廣泛。對於同一種物品，在外來的敲擊下或內部自身的原因，應該發出相同的聲音。但當其中的一些成分、結構發生變化後，會導致原有的聲音發生一些變化。據此，可以檢查許多產品的品質。如敲打罐頭，用聽覺檢查其品質，生產中稱為打檢。從機器發出的聲音來判斷是否出現異常。另外容器有無裂縫等，也可用聽覺判斷。

2.4.3 嗅覺

一、嗅覺的生理特點

嗅覺是辨別各種氣味的感覺。嗅覺的感受器位於鼻腔最上端的嗅上皮肉，其中嗅細胞是嗅覺刺激的感受器，接受有氣味的分子一

種濃度很低的氣味，必須用力吸氣，才能使氣體分子到達嗅區，產生嗅感。嗅覺的適宜刺激物必須具有揮發性和可溶性的特點，否則不易刺激鼻黏膜，無法引起嗅覺。嗅覺敏銳者並非對所有氣味都敏銳，因不同氣味而異。如長期從事評酒工作的人，其嗅覺對酒香的變化非常敏感，但對其他氣味就不一定敏感。人的身體狀況可以影響嗅覺器官。

嗅細胞容易產生疲勞，而且當嗅球等中樞系統由於氣味的刺激陷入負反饋狀態時，感覺受到抑制，氣味感消失，這便是對氣味產生了適應性。

嗅覺的個體差異很大，有嗅覺敏銳者和嗅覺遲鈍者。嗅覺敏銳者並非對所有氣味都敏銳，因不同氣味而異。如長期從事評酒工作的人，其嗅覺對酒香的變化非常敏感，但對其他氣味就不一定敏感。人的身體狀況可以影響嗅覺器官。如人在感冒時，品嚐咖啡的香味，顯然不如平常那樣芳香撲鼻。當身體疲勞或營養不良時，都會引起嗅覺功能降低，女性在月經期、妊娠期及更年期都會發生嗅覺缺失或過敏的現象。

二、氣味

氣味的種類非常多，有人認為，在 200 種有機化合物中，40 種都有氣味，而且各不相同。人僅能分辨出 5000 餘種氣味。借助分析儀器可以準確區分各種氣體。對氣味的分類以 Amoore 的分類方法最為著名。他發現：樟腦味、麝香味、花香味、薄荷香味、乙醚味、刺激味和腐臭味這七個詞彙的應用頻度最高，因此，這七種氣味被認為是基本的氣味。任何一種氣味的產生，都是由七種基本氣味中幾種氣味混合的結果。各種氣味之間存在掩蔽現象，所以，有時為了去除某種討厭的或難聞的氣味，就用其他強烈氣味加以掩蔽。或使某種氣味和其他氣味混合後性質發生改變，成為令人喜歡的氣味。

三、嗅覺的感官檢查

在食品的風味化學研究中，通常由色譜和質譜將風味各組分定性和定量，但整個過程中的提取、捕集、濃縮等都必須伴隨著器官的嗅覺檢查，才可保證試驗過程中風味組分無損失。在生產、檢驗和鑑定方面，嗅覺起著十分重要的作用。有許多方面的分析是無法用儀器和理化分析方法所代替。如在食品的風味化學研究中，通常由色譜和質譜將風味各組分定性和定量，但整個過程中的提取、捕集、濃縮等都必須伴隨著器官的嗅覺檢查，才可保證試驗過程中風味組分無損失。另外，調香和酒的調配等也需要用嗅覺來評判，才可最後投入生產。食品很多氣味的描述語是很主觀的，穩定性比風味差。嗅覺的敏感度可達 ppm，甚至 ppb。由於嗅細胞有易疲勞的特點，所以，對產品氣味的檢查或對比，數量和時間應儘可能縮短。

2.4.4 味覺

一、味覺的生理特點

人類的味覺感受器是覆蓋在舌面上的味蕾。這些味蕾分布在不同的乳頭—蕈狀乳頭、輪廓乳頭、葉狀乳頭上。蕈狀乳頭主要分布在舌尖和舌側部位。輪廓乳頭是V字型，分布在舌根部位，在每個乳頭的乳突體內，有幾千個味蕾。葉狀乳頭主要位於靠近舌兩側後區。味蕾中含有味細胞，可溶性物質最終刺激於味細胞，使其呈興奮狀態，由味覺神經立即傳入中樞，進入大腦皮質，產生味覺。輪廓乳頭對苦味最敏感，葉狀乳頭內的味蕾對酸味最敏感。圖 2-1 為四種基本味道接受器在舌頭上的分佈。

從刺激味覺感受器到出現味覺，一般需 1.5~4.0 毫秒。其中鹹味的感覺最快，苦味的感覺最慢。水溶性好的物質味覺產生快消失也快，水溶性較差的物質味覺產生較慢，但維持時間較長。最能刺激味覺的溫度在 10~40℃ 之間，其中 30℃ 時味覺最為敏感。如甜味在 50℃ 以上時，感覺明顯地遲鈍。隨著年齡的增長，味覺逐漸衰退。

味覺的強度和出現味覺的時間與味刺激物(呈味物質)的水溶性有關。完全不溶於水的物質實際上是無味的。只有溶解在水中的物質才能刺激味覺神經，產生味覺。因此，呈味物質與舌表面接觸後，先在舌表面溶解，而後才產生味覺。這樣，味覺產生的時間和維持的時間因呈味物質的水溶性不同而有差異。水溶性好的物質味覺產生快消失也快，水溶性較差的物質味覺產生較慢，但維持時間較長。蔗糖和糖精就屬於不同的二類。

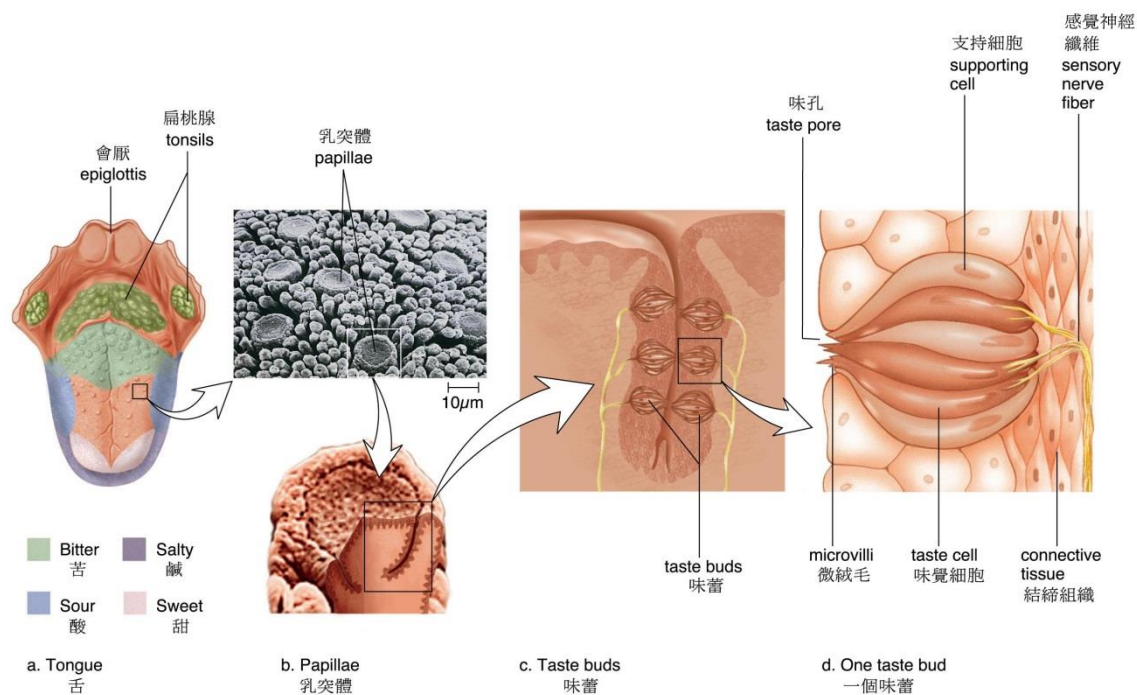


圖 2-1 四種基本味道接受器在舌頭上的分佈

不同年齡的人對呈味物質的敏感性不同。隨著年齡的增長，味覺逐漸衰退。50 歲左右味覺敏感性明顯衰退。甜味約減少 1/2，苦味約減少 1/3，鹹味約減少 1/4，但酸味減少不明顯。

二、五種基本味

關於味的分類方法，各國有一些差異，中國是“酸、甜、苦、辣、鹹”，歐洲是“甜、酸、鹹、苦、金屬性、鹹性”等。但一般認為，味覺與顏色三原色相似，具有四原味，即甜、酸、鹹、苦，是四種基本味覺。這是德國海寧提出的一種假設。他認為，所有的味覺

都是四種原味組合而成。以四種原味各為一個頂點，構成味的四面體。但是在食品的調味等方面，亞洲民族往往也把鮮味當作一個獨立的味覺考慮。目前已證明上述的味覺分區並不正確，而是只要有味蕾的地方都可以感受到甜、鹹、鮮、苦、酸等5種基本味覺。

三、味覺的感官檢查

評估人員的身體狀況，精神狀態，味覺嗜好，樣品的溫度等都對味覺器官的敏感性有一定的影響。味覺檢查主要用來評估，分析食品的品質。評估人員的身體狀況，精神狀態，味覺嗜好，樣品的溫度等都對味覺器官的敏感性有一定的影響，因此，在進行味覺檢查時應施予特別的注意。另外，味覺檢查常用來鑑定中草藥材。

2.4.5 觸覺

一、觸覺的生理特點

皮膚的感覺稱為觸覺。它是辨別物體的機械特性和溫度的感覺。皮膚受到機械刺激尚未引起變形時的感覺為觸覺。若刺激強度增加，使皮膚變形，此時的感覺為壓覺。二者是相互聯繫，故又稱為觸覺壓。觸覺的感受器在有毛的皮膚中就是毛髮感受器，在無毛髮的皮膚中主要是麥斯納小體。壓覺感受器是巴西尼環層小體。10~60℃的溫度刺激為適宜刺激，均能產生溫度覺。劃分溫、冷的分界是皮膚的生理零度及皮膚表面的溫度。通常年輕人較喜歡稍硬而脆的食物，或者像牛排、肉乾等耐咀嚼的食物，不太喜歡受潮的餅乾與過於軟爛的食物。相對地，嬰兒或沒牙齒的老人則不得不接受較軟的食物，否則其牙齦會因無法處理較硬的食材而不舒服。劃分溫、冷的分界是皮膚的生理零度及皮膚表面的溫度。它不是溫度計上的零度。生理零度(T_0)，是實際溫度覺上的中性溫度。高於生理零度的溫度刺激引起溫覺，低於生理零度刺激產生冷覺。人體各部位的生理零度是各不相同的，生理零度也不是固定不變的，它隨皮膚對外界溫度的適應而改變。在分辨物品表面的冷、熱程度時，氣溫和檢查場所的環境溫度，檢查者的體溫等，都能給溫度覺產生感受誤差。

辣椒素刺激皮膚產生痛覺的綜合呈現。就「辣」這個感覺而言，應該是屬於化學成分造成疼痛的觸覺，刺激局部血液循環加速而腫脹，感到灼熱。也因腫脹而造成味覺細胞的靈敏度提高，而加重了食物的味道或品嚐出正常狀態下感覺低微的風味。辣味按其刺激性不同分為火辣味和辛辣味兩類。火辣味在口腔中能引起一種燒灼感，如紅辣椒和胡椒的辣味。辛辣味具有沖鼻刺激感，除了作用於口腔粘膜外，還有一定的揮發性，能刺激嗅覺器官，例如薑、蔥、蒜、芥子等的辛辣味。

二、觸覺的敏感性

觸壓覺的感受器在皮膚內的分布不均勻，所以不同部位有不同敏感性。四肢皮膚比軀幹部敏感，手指尖的敏感性強。此外，不同

皮膚區感受兩點之間最小距離的能力也有所不同。舌尖具有最大的敏感性，能分辨兩個相隔 1.1mm 的刺激。手指掌面能分辨 2.2mm，背部正中只能分辨兩個相隔 6~7mm 的刺激。

皮膚分布著冷點與溫點，其整體是冷點多於溫點，兩者之比為 4：1~10：1，所以皮膚對冷敏感而對熱相對不敏感。面部皮膚對熱和冷有最大敏感性，每平方厘米平均有冷點 4.8~5.2 個，溫點 1.7 個。腿部皮膚每平方厘米平均有冷點 4.8~5.2 個，溫點 0.4 個。一般軀幹部皮膚對冷的敏感性比四肢皮膚大。

三、觸覺的感官檢查

觸覺檢查是用人的手、皮膚表面接觸物體時所產生的感覺來分辨、判斷產品品質特性的一種感官檢查。觸覺檢查主要用於檢查產品表面的粗糙度，光滑度、軟、硬、柔性、彈性、塑性、熱、冷、潮濕等感覺。人體自身的皮膚(手指、手掌)是否光滑，對分辨物品表面粗糙、光滑、細緻程度也有影響。如果皮膚表面有傷口、炎症、裂痕時，觸摸覺的誤差更大。這些是感官檢查中應注意的。

第三章 品評試驗室、測試產品及品評小組的管理

3 Controls for Test Room, Products and Panel

假如感官評估的結果目的是要在調查研究過程中測量一個真實產品的差異，許多變數就必須受到控制。我們常將這些變數分成以下三大類：

A 測試的管理：試驗室的環境、品評小室(或稱為評估小室)或是一個圓桌的使用、光線、室內照明、準備區及進出區。(器)

B 產品的管理：使用的設備、樣品的審視、製備、數量、編碼及呈現方式。(物)

C 小組的管理：品評員(或稱為評估員)評估待測樣品的程序。(人)

3.1 測試的管理

感官評估場所(實體設施)的組成區域包括品評小室、預備室、討論室、電腦室、會議室、廚房或小型工廠、儲藏室、接待室，其

中某些區域可大可小，與其他部門共用，但單獨設置最好，如預備室(單獨設置最好)、討論室、數據處理室、貯藏室、會議室、小型工廠或廚房(單獨設置最好)，特別是人口稠密的地區如台灣。

實體設施的設計乃是為了使實驗品的偏差最小化、並使其靈敏度最大化，以及消除並非來自產品本身的變數。小組測試之成本所以昂貴是來自品評員所耗費的時間成本。是否大幅減少干擾因素是很容易被證明的。小組的出席率及動機是普遍性的問題，而管理階層必須藉由關心及增強測試區域的努力來清楚展現其小組測試的價值。測試區域需要集中設置、容易到達(報到)且沒有人群聚集及任何混亂，舒適、安全、溫度控管、無氣味及噪音。

3.1.1 試驗室設計的發展

藉由對應了國家級及國際的標準，試驗室的設計已趨成熟，在 ISO9000 標準下，加速推動了感官評估設施的一致化，例如：各個品評小室獨立的排氣設計。

早期的試驗室允許放置 6 到 10 個實驗品，且由一個實驗台組成，或是所有樣品擺放在同一張會議桌。為了防止實驗品因互相影響而導致偏差及干擾，帶入了品評小室的概念。同時期，Arthur D. Little 主張品評員應互動來達成一致性，因此需要一張配置圓轉盤的圓桌，使得放置其上的參考材料可以標準化用詞及刻度值。

當前的思維常常結合以下兩個因素：(1.) 品評小室區：也就是主要進行不同實驗及描述性實驗的空間。(2.) 圓桌區：用來訓練及其他描述性任務。樣品準備區域安置在附近以方便取用，但必須與試驗室隔開。一定規模以上的設置也必須考量辦公區域、樣品存放區域及資料處理區域。

3.1.2 地點

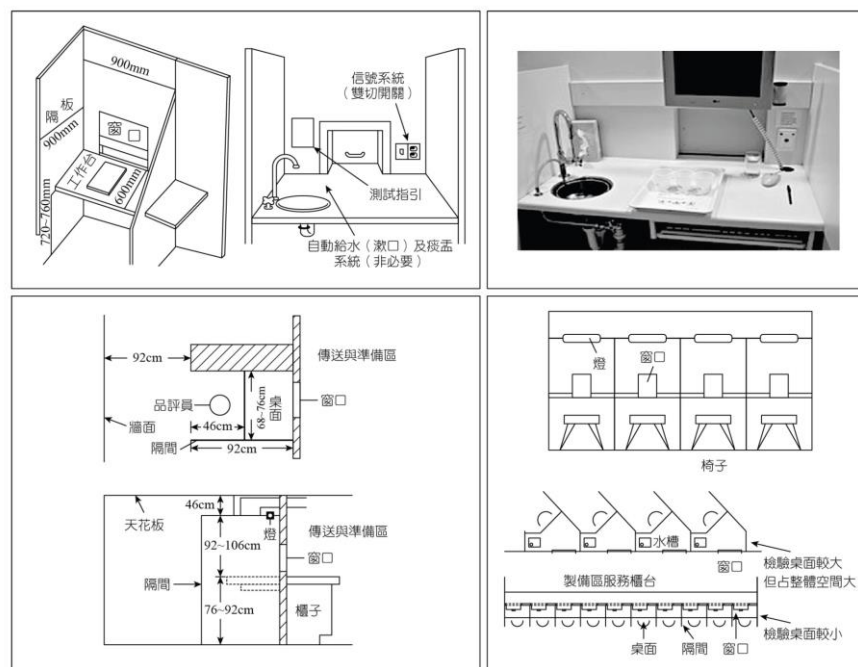
品評地點需要設置在所有人都容易到達的地點。設置在所有品評員前往午餐或晨間休息的路上可能是一個良好的地點。假如品評員是由戶外進入，則品評室需要設置在靠近入口的地方。品評地點也應該與擁擠的區域保持適當的距離，以避免噪音以及不必要的社交。品評地點須遠離噪音及氣味，比如說機房、裝卸區、產品線、休息區或廚房。

3.1.3 試驗室設計

1. 品評小室

如果以感官評估試驗的使用頻率低，不一定要設置品評室，通常可使用臨時性的佈置，將實驗桌分割成一些小隔間，每個小隔間的桌面上都擺放尚待檢測的樣品，這樣可以防止品評員避免受到干擾單獨隔開的品評小室可以避免品評員交談的機會，也可採用含枚方式防止品評員出聲。習慣上一項樣品的準備應可供應 6 到 8 個品評小室。品評小室可以是並排、L 型或是 3 到 4 個品評小室，其中兩組圍繞著供應區而互相面對。L 型放置所形成的工作三角概念，是廚房中最有效率的設計，使技術員供應樣品的時間及距離最小化。6 到 8 個品評小室及 18 到 24 名品評員所構成的單位每年可處理三百到四百批的中等型試驗。至於更大量的測試需求，則建議一個樣品準備區域複合多個測試單位。另外也需考慮放置技術員的監控器(螢幕)以及可處理任何自動化資料處理系統的中央處理區域。

圖 3.1 顯示了一個典型的品評小室。桌子高度則應安裝成與樣品準備桌面等高。空間須容許放置螢幕及鍵盤，如果需要的話。隔板需至桌面往上延伸接近十八吋以隔絕各個品評小室視覺及聽覺上的干擾。在通風及清掃的方便考量下，隔板也可以從地板延伸至天花板已達到完全的隱私，不然則從牆面延伸，包圍評估者的軀幹及頭部即可。在大多數情況下後者為優先的，畢竟幽閉恐懼症是一個永遠的問題，鑒於品評員很快地就學會看對方的肩膀或是大聲地唸出其對樣品品質的評論，其實對一位有交談習慣的品評員，品評小室也不見得有用。建議品評小室彼此的間隔至少四英尺以上，連接個品評小室的通道得以讓品評員容易進入各個品評小室。



■ 圖 4-4 品評小室的規格與設備

圖 3.1 品評小室

特定品評小室設施—通常配置一個小的不鏽鋼水槽及水龍頭以方便使用來沖洗。對於產品像是漱口水、牙膏及家用物品等，這樣的配置是必要的，但如果是一些容易阻塞水管的固體食物則不建議此配置。假如沒有辦法取得無異味的水，則可能需要裝置濾水系統。

訊號系統時常也會被納入品評小室中，使得品評員的督導可以即時知道其是否已準備好，或是樣品是否有問題。通常這個使用開關的形式，燈號從品評小室連接到樣品準備區，有時也會延伸至外部的燈號控制區，顯示各個品評小室的樣品是否可使

用。

若各個品評小室配置一個直接電腦輸入系統，則各品評小室至少需要三十二吋寬，以容納所有輸入裝置（鍵盤、滑鼠、顯示器）。

樣品盤可以手提至各個品評小室，前提是樣品須為無氣味且在二十至三十分鐘內不會變質。假如樣品不符合這些前提，樣品準備區必須設置在品評小室的正後方，且樣品通入口的大小必須使樣品托盤能夠通過。常用的樣品通入口有三種(見圖 3.2)。垂直或水平滑門設計的樣品通入口是最節省空間的，而掀蓋式及旋轉式的設計則是用來防止氣味擴散，以及樣品準備區狀況對品評人員的影響。

用於建造品評小室及其周遭環境的材質必須是防氣味沾附及容易清潔的，最常用的材質為不鏽鋼及樹脂合成板。

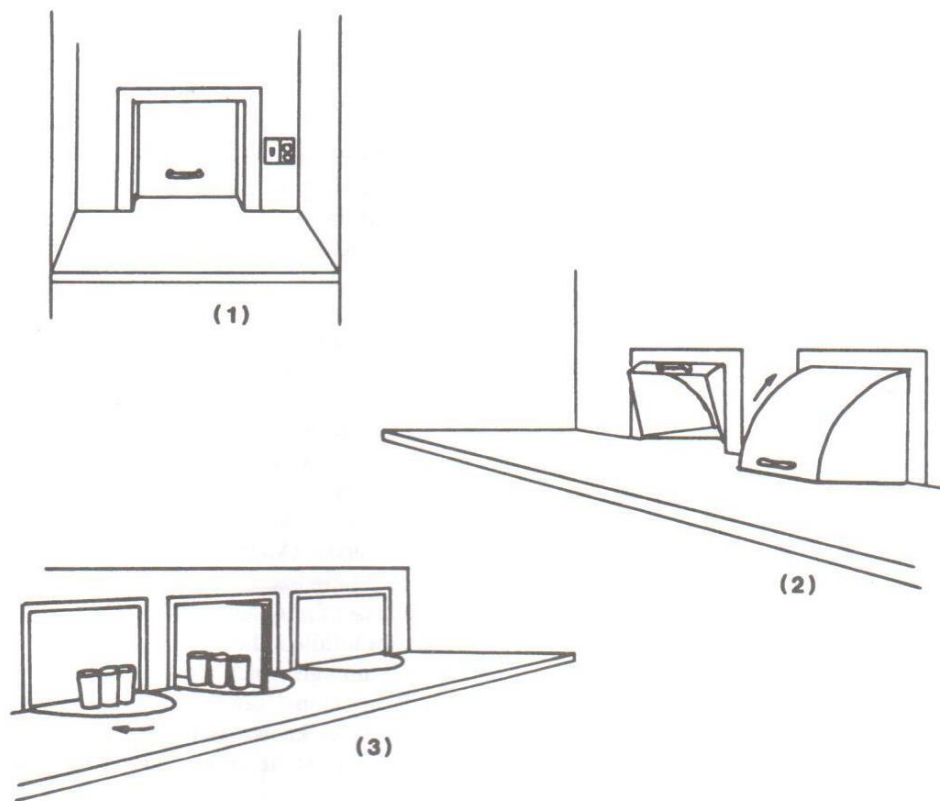


圖 3.2 三種傳遞給品評者的傳送口(1) 拉門 (2) 麵包盒開口式 (3) 旋轉門

3.1.3.2 描述性評估及訓練區域

此功能最簡易的方式可在品評管理員的辦公室中的桌子上進行品評人員的標準教育。若產品需要時常進行描述性分析，或需要比較大的訓練及教育空間，則建議準備下列的設施：

- 具有多張桌子的會議室，且桌子可根據訓練團體的大小而更動擺放位置。
- 視聽設備，包括電子白板及有記號樣品的功能。
- 參考樣品的獨立準備設施可用來展示。根據不同的形式，這有可能包含保存空間(冷凍庫、冷藏室、控溫室或蠟封箱以避免氣味逸散)及用來讓參考樣品展示的等待空間。

3.1.3.3 準備區域

準備區域是一個實驗室空間，其空間必須讓所有可能及可預見的樣品組合可在最高速率下準備所要的樣品。任何一個品評小室及描述性分析必須有獨立的樣品準備區，使技術員得以在每次研究過程最佳地準備、呈現及清潔。除了特殊樣品所指定的特殊設備外，一般而言，樣品準備區需要具備以下幾種設備：

- 一個具有多個樣品通入口的實驗桌，以用來同時供應許多品評小室。
- 廚具、烤箱等樣品準備設備。
- 冷凍庫、冷藏室等用來儲存樣品的設備。
- 用來放置玻璃碗、樣品盤等的櫥櫃。
- 洗碗機、烘碗機、垃圾處理器、廚餘桶、水槽等。
- 提供給品評人員的獎勵品的儲存冰箱。
- 大型垃圾桶，以用來丟棄使用過的樣品

須考慮公司及當地的回收政策，樣品的垃圾處理要有相對應的區域設計。

3.1.3.4 辦公室

辦公室須位於看得見品評小室的區域，因為每次的品評都必須要有管理人員在場。若資料紀錄、儲存空間及電腦設備及其他硬體都位於此空間，可能有利於品評管理員的時間管理最佳化。手機及印表機等設備應保持相當距離以避免干擾辦公人員。

3.1.3.5 品評小室進出區

若是大型的設施，建議分開入口及出口，以避免多餘的資訊交換。出口區一般會設置一張桌子，使品評人員可以閱讀今日的

樣品資訊及領取今日參與的獎勵點心。假如部分的品評人員並非公司內部員工，則品評小室的進出口必須設置相對大小的等候室，內置舒服的沙發、外套懸掛區及獨立的洗手間。

3.1.3.6 儲存空間

此空間須考慮可容納：

- 製備前的樣品、製備後的樣品、等待呈現樣品的空間
- 參考樣品的管理或標準，需保持在適當的溫度及濕度條件。
- 大量的可丟棄式容器及器具
- 清潔用品
- 使用前及使用後的記分紙

圖 3.3、圖 3.4 顯示典型的中型及大型的實驗室佈局，可看到許多的設備圍繞在品評小室的周圍。

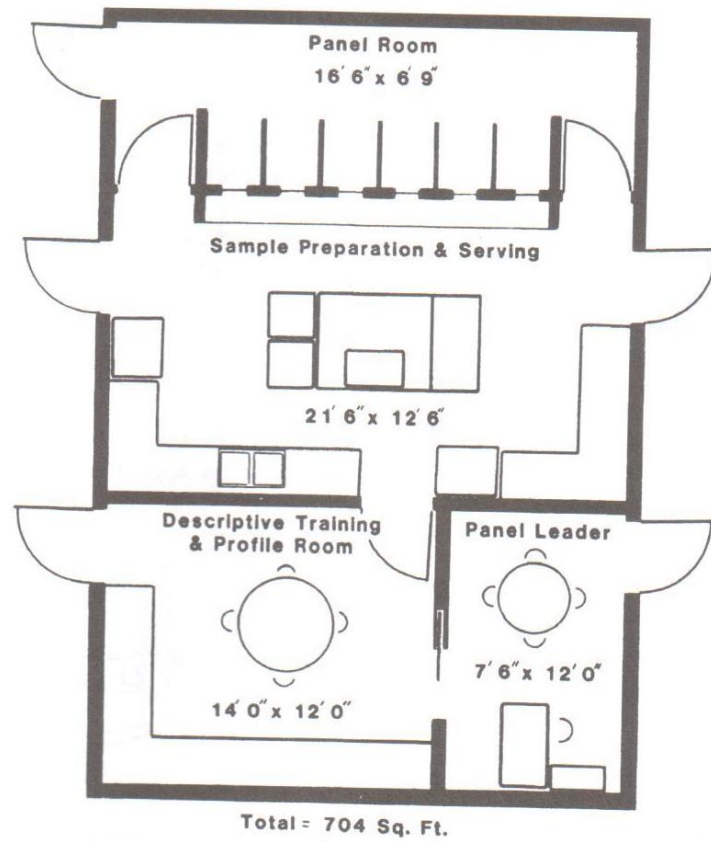


圖 3.3 中型品評區域的佈局，一年約可容納三百至四百個測試

無任何顏色可以防止分心。品評小室的照明需選用平均光源、無陰影且至少 70 至 80 燭光，也就是一般辦公室所需要的照明。若產品的外觀是品評的重點之一，則使用變阻器調節燈光至 100 燭光。使用白燈泡可以搭配較多的燈源變化，並且可以使用彩光燈(見下)，但是會產生較多的熱源，需考量散熱。螢光燈產生較少熱源，而且可以選擇不同種類的白光(如：冷白光、暖白光等，見圖 3.5)

彩光燈—常見於品評小室所適用的彩光為綠色、紅色及使用特殊濾光片所得到的低密度藍光。這些光源用於掩飾樣品外觀的差異度以獲得味道或觸感的品評結果。

許多彩光燈泡會發射過多的白光，使得降低色彩的差異性沒有效率。劇院所使用的彩色濾光板則相當有效率，且可以安置在框板中，過濾鑲嵌式的光源。另一個替代選擇是只發射單一波長的低壓鈉燈。不管是彩色濾光片或是彩光燈都可以過濾特定色彩，但光密度的差異則不可被消除。判斷標準是在黑白螢幕上仍可以偵測灰階的程度。

不正常的燈源或多或少會影響品評人員對樣品的印象，一個替代做法是不要同時呈現所有樣品讓品評人員比較外觀，比如說連續分次上不同樣品並且讓品評人員有參考標準樣品。

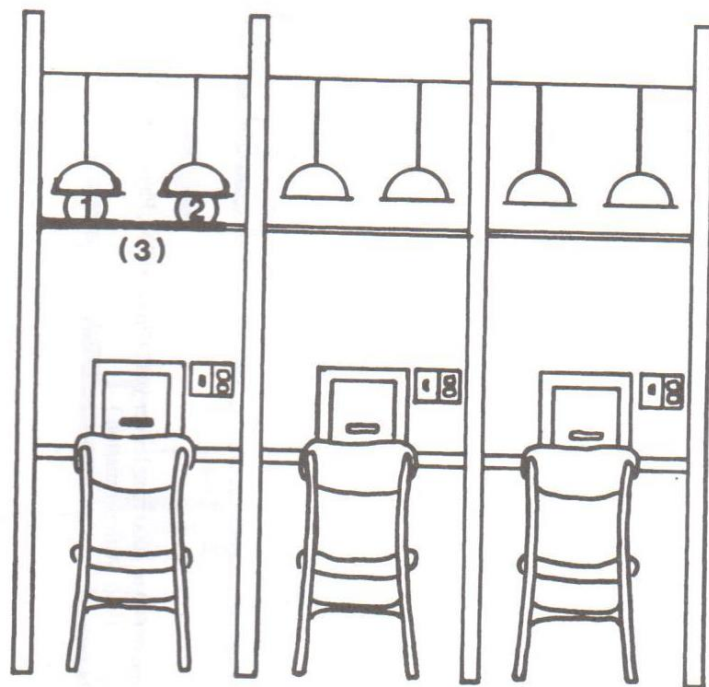


圖 3.5 品評小室的燈光設計(1) 白光 (2) 螢光 (3) 濾光片支架

3.1.4.2 空氣循環、溫度、濕度

品評小室的空間溫度必須在華氏 72°至 75°之間，相對濕度 45%至 55%。若品評項目為布料的觸感、無紡紙或護膚產品等，則需要較緊密的濕度控制。如： $50\pm2\%$ ， $65\pm2\%$ 。再循環的空氣須通過活性碳過濾箱以去除所有能感知的氣味。過濾箱需放置在品評小室之外，且容易更換的地方。建議兩到三個月更換一次。頻繁的檢查過濾箱是必要的，以避免過濾器無效且變成一個氣味的來

源。實驗室中需要輕微的加壓，以避免氣味從外面或是樣品準備室擴散進去。若是有可能品評氣味重的樣品，如：香腸、起司，則各個品評小室需要有獨立的抽風器。

3.1.4.3 建造材料

品評小室的建造材料必須相對應齊所有可能品評樣品而考慮。

避免氣味—紙、布料、地毯、透氣磚等容易產生或留下氣味且從容易藏汙納垢的建造材料需盡量避免。建造材料必須是光滑、易清潔且不會吸收的材質，所以不會留下上一個實驗的氣味。符合以上特性的最佳材料為不鏽鋼、壓克力板及鐵氟龍。乙烯層壓板則適用於天花板、地板及牆壁。

色彩—選擇米白色搭配一些樣式，這樣中性、不唐突的色彩組合搭配可以避免干擾品評人員。尤其是檯面的選擇是至關重要的，必須選擇不會混淆樣品且干擾品評的顏色。一張白紙或是包覆黑色布料的檯面都可以突出樣品的外觀差異，使得外觀品評及觸感品評可以同時進行。

配管系統—若樣品阻塞在品評實驗室的水管中，則會干擾品評的進行且產生氣味。確保開放給所有品評小室使用的水管及水槽都是可以沖洗且清潔的。若某些品評項目需要品評人員吐唾（如：牙膏、漱口水），則須考慮分離式水槽，也就是說此水槽使用可彎曲的軟管連接入水孔及排水孔。當此水槽沒有在使用時可另外放置，並使用蓋子關閉水管。

3.2 產品管理

3.2.1 一般設備

當執行一項品評實驗，產品研究人員及感官分析師其實是在找尋有效的變因，如配方的改變、製程的變動、包裝更換、保存條件等。品評分析的一項主要責任就是在控制早期的操作、準備、呈現各項產品。這些管理在確保無效的變因不會被導入，及找出任何可能影響產品感官的變因。

準備區域必須設置在連接測試區的位置，然而，空氣循環系統必須使測試區有正壓，以確保空氣從測試區流動至準備區域，並且在準備區域設置空氣循環器及補充的抽風機。

3.2.2 樣品準備

3.2.2.1 供應及設備

除了必須的主要設備外，一個受管控的樣品準備區需要有足夠的以下設備：

- 刻度尺：以測量產品及材料。
- 玻璃器皿：以測量及放置產品。
- 計時器：以監控準備過程
- 不鏽鋼設備：以攪拌及儲存產品。

3.2.2.2 材料

用在準備及呈現樣品的材料必須小心挑選，以免帶來任何偏差或引進新的變因。大多數的塑膠刀具、儲存容器及保鮮膜並不適用於餐飲及個人護理用品等的樣品製備。可能從塑膠中擴散，或是沾黏在塑膠中的揮發物質都會改變樣品的氣味及口味。

3.2.2.3 準備程序

控制樣品的準備需要詳細規範及監控流程，以下幾點須特別注意：

- 產品的數量、重量。必須使用準確的儀器測量其重量或容量。
- 任何新增材料的數量及重量
- 準備的過程、時間管理(計時器)、溫度管理(溫度計)、反應速率、準備設備的大小及型號
- 定義樣品呈現至感官評估前最大及最小的等待時間

3.2.3 樣品呈現

3.2.3.1 容器、樣品大小及其他參數

呈現產品所使用的容器及設備必須要經過小心地挑選，以減少帶入變數及偏差的風險。以下幾點須注意：

- 供應容器：再次提醒需使用玻璃或上釉瓷器，而非塑膠，除非是測試過的。
- 供應大小：提供任何測試主體時須特別注意準確的產品總量。當提供最小的操作量時，必須小心地訓練技術員以提供正確的總量。特定的設備可能有助於感官評估階段提供有效地總量。樣品的量：一般，液體 16-30cc；固體，28-30 g。樣

品數：一次以不超過 6 個為原則。

- 供應載體(carrier):所謂的**載體**是指品嚐食物時伴隨所品嚐的食物一起被吞嚥的其他任何食物。對大部分的差異性測試，測試中的產品往往是其本身而無任何添加物。產品如咖啡、茶、花生、蔬菜、肉等在測試時並沒有提供一般消費者在使用這些產品時常用的佐料或添加物。如：牛奶、麵包、奶油、香料等。相反地，在消費者測試中(接受度、偏好測試)，產品則需按照一般使用的方式所呈現。例如：茶或咖啡依需求加牛奶、糖或檸檬；花生搭配麵包或餅乾；蔬菜和肉搭配根據消費者偏好的香料。產品如添加物或佐料等一般搭配其他產品食用的，則須搭配有一致性，且不會剝奪產品本身特色的主食，像是麵糊、糖果軟餡、甜牛奶。也有學者將飲料中的水是為載體。

供應溫度：在所有樣品被分配到其盛裝的容器後，必須在測試前再一次測量溫度是否在合適的範圍內。大部分的感官評估實驗室都會建立包括確認供應溫度的標準製備流程，以確保樣品是在正常的溫度下被測試。標準流程的建立大大地減少了監控每個測試樣品的需求。樣品提供的溫度需要考慮：(1)同次試驗樣品的溫度要維持一致、(2)試驗樣品平時食用的溫度應是該食品最好吃的溫度，也是容易檢出品質間差異的溫度、(3)試驗樣品容易保持的溫度、(4)不會使樣品產生變性的溫度、(5) 不易產生感覺疲勞的溫度。通常樣品的溫度要求：油 40-60℃、熱湯 60-70 ℃、冰淇淋 0 ℃。

- 樣品之均一度
- 受評之方式：樣品停留時間(5 秒)、是否咀嚼、吞嚥？
- 樣品的資料：材料、目的、避免品評員猜疑(不能經過準備室到品評小室)。

3.2.3.2 順序、編碼、及樣品數量

順序、編碼及樣品數量是任何測試中都必須要監控的範圍。為了平衡樣品供應的順序、每個樣品的順序必須平均地分配。舉例來說，以下是三個樣品 ABC 所可能被供應的順序：

ABC—ACB—BCA—BAC—CBA—CAB

這類的測試必須使用六個樣品，以確保六個可能供應的順序組合都被涵蓋。供應的順序也必須是隨機的，可以抽取亂數

卡或是使用亂數程式來達到這樣的效果。每個樣品所指定的編碼也有可能是造成偏差的來源，比如，某實驗對象潛意識地選擇編碼 A 的樣品。因此，應避免單數或兩位數字母的編碼。另外，字母代表某地區、公司或任何產品、樣品等都不應該被使用。大多數的感官測試依靠三位數的亂數表來做編碼。編碼在樣品及評分表上的標示也不該太過顯眼，適當且清楚的標明即可，以避免造成任何可能的偏差。

在一個感官測試中，提供樣品的數量多寡會造成品評人員感官及心理上的疲乏，基本上，若是餅乾類的產品，上限是八到十個，啤酒類則六到八個為上限。若產品帶有強烈的味道，像是煙燻或嗆辣的肉品、苦澀味或油膩的口感等，每個測試可能只能夠提供一到兩個。另一方面，視覺測試考量到心理疲乏的因素，每次最多只能提供二十至三十組。

3.2.4 產品樣品

執行一項感官測試必須決定需要多少產品，以及此項產品的歷史。掌握實驗主體的事前資料以及控制樣品的細節是對於測試的過程及結果解析至關重要的。感官實驗室應該保有一本紀錄簿，以保存其測試樣品的相關資訊：

- 產品來源：生產地及生產時間。樣品履歷對於實驗階段(實驗室編號)，及量產階段(機器編號)的測試產品都極為重要。
- 測試需求：執行此次品評的所有測試，以及可能地二次測試，需要多少數量的產品。所有此樣品來源的產品都必須是同一來源(生產地、生產線、生產資訊。假如產品並非一致的，則須考量混合、拆封不同批的包裝。
- 保存：此產品必須如何保存，及過去曾經歷哪些保存條件。假如一項測試是要比較兩項產品配方或製程上的改變，則保存期限、保存及運送溫度、濕度、包裝方式的不同都不可能測量此項改變，因為其有可能造成過大的偏差而遮蓋了此變異。

3.3 品評人員的管控

品評人員與環境、產品、測試步驟的互動都有可能是造成變異或偏差的來源。管控及規則的建立可以減少這些可能影響結果的無關變異。

3.3.1 小組訓練及培訓

品評員無庸置疑地需要妥善的培訓，包括操作樣品、使用評分表及品評欲確認的資訊等。品評員的培訓方式在第九章有詳細的解釋。品評人員起碼需要被訓練為：在沒有任何額外的指導下，也熟知如何進行任何一項基本的品評測試。他們必須熟知：

- 測試步驟，例如進行一次測試的數量、拿取方式（湯匙、杯子、吸允、喝）、接觸此產品的時間長度（啜飲、吐出、輕嗅、咬、嚼），以及產品的後處理（吞下、吐出、留置皮膚上或取下）。這些常識所有品評人員都必須熟知且遵守。
- 評分表的設計，其中包括了此次測試的指導、問題、術語及判斷表達的規模。品評人員都應該詳知評分表的內容。
- 品評或判斷的方式（差異、描述、偏好、接受）也都必須在培訓中被品評人員清楚地瞭解。

Kelly 和 Heymann(1989)創立了攝入及吐出食物中的無顯著差異測試。例如：在加了鹽的菜豆中，攝入測試的確提供了較窄的信賴區間。

3.3.2 產品及時間

假如品評人員還沒有經過良好的培訓，則安排此產品在其通常被取用的時間下做測試是比較好的方式。不建議在早上安排測試酒精類產品或是有強烈風味的產品，或是在午餐及下午茶後安排產品測試也可能造成結果偏差，而應該盡量避免。為了改善評審的一致性，對於品評人員的皮膚及嘴巴的事前處理也是必要的。

3.3.3 品評人員及環境

如同此章的第二節所討論的，為了避免偏差的前提下，被品評人員所感知的環境都必須設計及管控。然而，某些環境控制如有色的燈光、高濕度及封閉的測試環境也可能使品評人員焦慮或分心，除非品評人員已經接受了多次的特殊環境訓練。

再一次強調，提供品評人員充足的培訓及時間去適應特殊的測試環境及條件是必要的，並提供他們充足的資訊使他們瞭解如何妥善的因應各個可能的變異。

第四章 影響感官判定的因子

4 Factors Influencing Sensory Verdicts

4.1 簡介

一次好的感官評估，必須將感官評估者視為一個測量設備。然而事實上，品評員之中的差異，隨著時間是極有可能成為偏差的。為了使這些變數及偏差最小化，實驗者必須瞭解可能影響感官認知的生理及心理因子。

感受真實世界並不是一個被動的過程，而是一個主動及選擇性的方式。觀察者在一個複雜情況下只能察覺那些他已經準備好去觀察、去聯想及對其有意義的元素，其餘的因子則就算面對著觀察者發生，他也會直接忽略。我們必須將觀察者先放在一個瞭解的框架中，讓他在品評前先充分了解我們希望他感受的所有產品特性。要達成這個條件，必須在訓練過程中(見第九章)教導他們繞過一些在樣品呈現上、問卷調查中，以及操作方式裡固有的陷阱。

4.2 生理因子

4.2.1 適應性

適應性是指感受器在同一刺激物的持續作用下，敏感性發生變化的現象，亦即在連續的刺激或相似的刺激下，敏感度的降低或改變。在感官評估中，這個因子是我們極不想要的閾值及強度的變異來源。

在下列的交叉適應性的實驗中，在條件 B 下的觀察者很有可能察覺到測試刺激中的樣品較不甜，因為在調整刺激中的蔗糖降低了他對甜度的敏感度，而用於條件 A 中的水因為不含甜味而不會造成敏感疲乏（或造成在甜味測試中的適應性）。

調適刺激

測試刺激

條件 A	水	阿斯巴甜
條件 B	蔗糖	阿斯巴甜

條件 A	水	奎寧
條件 B	蔗糖	奎寧

在此，交叉增強的現象則可能發生。在條件 B 下，觀察者在測試刺激的樣品嚐到較多的苦味，因為在調適刺激中嘗到的苦味強化了他對於奎寧的敏感度。

4.2.2 增強或抑制

增強或抑制的現象包含同時的混合刺激之中的互動。

增強(Enhancement)——一項物質的呈現增加了第二項物質的感受強度，是指當兩個刺激物同時或連續存在於同一感受器時，一般把一個刺激的存在比另一個刺激強的現象，又稱為**對比現象**，所產生的反應叫**對比效應**。同時施予兩個刺激時稱為**同時對比**，先後連續施予兩個刺激時，稱為**連續對比**(或稱為**先後對比**)。在品嚐評比幾種食品時，品嚐每一種食品前都要徹底漱口，以避免對比效應帶來的影響。

綜效(Synergy)——一項物質的呈現增加了兩項物質混合後的感受強度，是兩種或多種刺激的綜合效應，它導致感覺水準超過預期的每種刺激各自效應的疊加。又稱為**協同效應**或**相乘效果**。

抑制(Suppression)——一項物質的呈現降低了第二項物質的感受強度，是指因一種刺激的存在，會使另一種刺激強度減弱的現象。又稱為**拮抗效應**或**相抵效應**。當兩個強度相差較大的聲音同時傳到雙耳時，我們只能感覺到其中的一個聲音，這一現象稱為**掩蔽現象**。

舉例(見下列圖例)：

1. 混合物的總感受強度

情況	現象名稱
$MIX < A+B$ (各項單獨)	混合抑制
$MIX > A+B$ (各項單獨)	綜效

2. 可分析混合物的元素

情況	現象名稱
$A' < A$	混合抑制
$A' > A$	增加

圖例：MIX= 混合物感受強度

A= 元素 A 在未混合下的感受強度

A' = 元素 A 在混合後的感受強度

4.3 心理因子

4.3.1 期待錯誤(Expectation Error)

樣品的資訊可能引發實驗前的預期。人們通常會從樣品發現他們想要發現的痕跡。在實驗中，譬如典型的閾值實驗，其中包含了一系列的濃度上升樣品，實驗者(藉由自我暗示)將在試驗前便有了預期的品評，且在其回答中報告。而當品評員聽說工廠間有過剩產品時，則有找出當天過期產品的傾向。啤酒品評員若事先知道啤酒花濃度，則其偵測苦味的結果將會有所偏差。預期錯誤可以破壞任何實驗結果，所以實驗前一定要盡量避免提供實驗者任何詳細的樣品資訊。樣品必須編號並且隨機呈現給任一品評員。受過完整訓練且有良好動機的品評員是否因為任何意外的認知而受到影響，這點頗受爭議，但是實務上，品評員不知道該如何調整其自我暗示導致的預期程度，所以基本上還是避免讓他們知道樣品的來源比較好。

4.3.2 習慣錯誤

人類一直以來被形容為習慣的生物，而這個形容在品評世界裡是真實的，並且導向了失誤，也就是習慣失誤。這個失誤起因於對於一系列緩慢增加或減少刺激的實驗給予相同回應的傾向，例如每天的品質控制。品評員傾向於重複同樣的給分，而因此忽略任何正在進行的趨勢或接受有瑕疵的樣品。習慣性是常見的，且必須藉由提供大量多樣化的樣品來對抗此失誤。

4.3.3 刺激錯誤

這類的失誤起因於與樣品無相關的附加因子影響了觀察者的判斷，例如用餐環境、餐飲的擺盤、樣品的包裝。假如此附加因子有變異，則觀察者傾向於發現變異，就算此變異根本不存在。例如：人們廣泛知道螺旋蓋包裝的紅酒在市面上較為便宜，則品評測試中同樣的酒使用螺旋蓋包裝呈現的得分低於使用軟木塞瓶蓋。緊急召集的品評測試極可能會找出廠內已知的瑕疵。較晚提供的樣品容易得到較有味道的得分，因為品評人員知道管理者會為避免味覺疲勞會先提供較淡味道的樣品。要解決這類失誤的方法很明顯：避開不相關的刺激因子、規律地進行品評測試、頻繁且以不規律順序來提供樣品。

4.3.4 邏輯錯誤

邏輯錯誤出現於當有多個特徵同時出現在品評人員心中時。像是知道深色啤酒較有味道、深色的美乃滋較不新鮮，導致品評人員無視於其感官且改變了判斷。邏輯錯誤必須最小化，我們可以藉由保持樣品的一致性，或者透過使用濾光眼鏡、彩光等的幫助。某些邏輯錯誤無法被遮蔽但可以透過一些方法避免。例如：苦味較濃的啤酒通常會得到啤酒花濃度較高的得分，品評人員領導可以有時在低啤酒花濃度低的樣品中加奎寧以製造高苦味，來破壞這類的邏輯錯誤。

4.3.5 光環效應(Halo Effect)

當一個樣品需要被評估多項特性時，各項特性的評分則可能互相影響。同時評分許多不同的風味特性以及產品接受度可能產生與分別評分時有不同的結果。例如在一項柳橙汁的消費者測試中，品評人員被要求不只要評分總體喜好度，另外還要針對各項產品特性給分。若此柳橙汁的總體喜好度評分高時，則其他特性如甜度、酸度、新鮮度、風味強度、口感等的評分則有偏向高分的趨勢。反之亦然，當此柳橙汁的總體喜好度低時，則其他產品特性則得分也相對偏低。此效應解藥就是，當有任何一單獨特性

是重要的時候，就得分開不同次的樣品來做品評。

4.3.6 樣品呈現的順序

樣品呈現的順序可能導致至少以下五種的偏差：

對比效應：若在呈現完一個品質較好的樣品之後，再呈現的是一個品質較差的樣品，則品質較差的樣品會得到比它在單獨評比時還低的評分。反之亦然，品質較好的樣品在品質較差的樣品後，通常會獲得更高的評分。呈現的另一個例子是：若居住在明尼蘇達州，而華氏溫度達到 42 度，則當地的氣象會報導為遭受熱浪。相反地，若是在邁阿密，而華氏溫度達到 42 度，氣反而會報導為嚴重的寒冷期。

群體效應：一個品質好的樣品與一群品質差的樣品一同呈現時，會得到較低的得分。此效應與對比效應剛好相反。

中間取向錯誤：在中間呈現的樣品比起最後呈現的樣品容易被接受，在著名的三角測試中，在中間呈現有異的樣品比起在後面呈現更容易被偵測到。

軌跡效應：品評人員潛意識就會使用他們能夠得到的所有線索來執行品評實驗，這包括他們也能夠意識到樣品呈現的順序與品質是否有關連性。

時間錯誤／部位偏差：在一系列的測試時，品評人員的態度也隨著時間有微妙的改變。從對第一個樣品期待或是飢餓，到最後疲累、冷淡。通常，第一個樣品容易特別被喜愛(或拒絕)。一個短期品評(淺啜並給評)容易在第一個樣品產生偏差，而長期品評(一週家居佈置)則在最後的樣品產生偏差。第一組及最後一組的判別偏差更是巨大。

所有上述的偏差都必須在品評實驗中被最小化，而我們可以透過亂數化、平衡化供應各個樣品的順序來達成。平衡化指的是所有可能的組合數量都必須被平均的呈現。各個樣品在時間順序上也必須被平均分配：第一組、第二組....到第 N 組。假如有很多種的樣品需呈現，則可以使用平衡過的不完全區組設計。(見第七章)

亂數化則指的是各個樣品組合出現的順序必須由機率法來決定。在實務上，我們可以透過抽籤的方式來達成，或是藉由亂數表的輔助。(見表 T1 及第三章的產品控制)

4.3.7 同儕效應

一個品評人員對產品的反應有可能被另一個品評人員所影響。因為這個原因，各個品評人員被分開在各個品評小室裡，以避免其判斷不會因為看到了其他品評人員的臉部反應而受到影響。在品評過程中也是全面禁止品評人員出聲，另外品評區域也必須要注意噪音及其他容易造成分心的事物干擾。

4.3.8 動機缺乏

品評人員對於察覺各個樣品微小的差別、找尋對於樣品印象的適切表達用語、或者是提供具有一致性的給分上所做的努力程度，對品評結果有絕對的重要性。這是品評人員管理的工作，創造一個正確的氛圍使得所有品評人員感到舒服並且做得好。一個有興趣的品評人員總是更有效率，且在一個完整了解、充分定義的測試環境中，有動機性是最佳的。品評人員的興趣可以透過提供他們各自的品評報告結果來維持。應該讓品評人員感到其品評行為是一項重要的活動，而這可以透過維持控制及有效率的測試來巧妙地達成。

4.3.9 善變及膽怯

有些人傾向使用極端的字眼來描述，從而在品評結果上施加過度的描述。而另外一些人則傾向堅持在等級中間給分，並且最小化各個樣品間的差異。為了得到有意義的結果，對於任何一個新來的品評人員，領導每天都必須監督其品評結果，並且就品評小組已評分過的樣品給予適當的指導。假如有需要的話，使用多種樣品來解說。

4.4 身體疲勞狀態

品評人員在以下狀況下不該參加品評：(1) 假如是味覺測試且他們感冒或發燒，假如是觸感測試且他們處於皮膚神經上的失調。(2) 假如他們處於牙齒保健狀況差或牙齦發炎。(3) 假如他們處於情緒低落、工作壓力大的狀況下(相反地，有時品評試驗反而是忙亂一天的綠洲)。吸菸者也可以是一個好的品評人員，前提是在品評前的三十到六十分鐘內他們不能夠吸菸。濃咖啡會麻痺味覺一小時，而任何品評都不應該安排在餐後兩個小時內。最佳的品評時間(對於日班人員)為早上十點到午餐前。普遍來說最好的時間是基於每個人的生理時間，也就是在一個人最清醒，且心理強度最高峰的時候。馬修(1986)探討了多種健康及營養不協調

是如何影響品評結果的準確度，相反地，品評結果也能夠用來偵查一個人的健康狀況。

第五章 測量反應

5 Measuring Responses

5.1 簡介

這個章節描述了許多測量品評反應的方法，旨在介紹各個方法的原則，並討論其利與弊。在一個極度簡化的世界，我們可以視品評員為一個測量機器，且將其設置為 0 到 100 的量度，期能夠針對多個樣品及待測量的特性設置數個基準點。不幸的是，真實世界中品評試驗並非如此簡單，所以我們需要多種測量方式。其複雜度就像是多所主要大學內的心理系中又設置了許多心理物理實驗室一樣。部分的觀察因子將列在本章節中。

當我們要求品評員使用數字(或標示)來代表其感官印象時，他們可能以下列四種方式來進行：

- 1 **表名資料**(Nominal data)(資料亦譯為數據)：品評人員將被檢視的項目放置在兩組或是多組不同的名稱裡，但遵守某種順序，或是擁有數量關係。例如：甜味、酸味。
- 2 **序列資料**(Ordinal data)：品評人員將被檢視的項目放置在兩組或是多組的序列中。例如：不同辣度(輕微、適中、強烈)的宮保雞丁。
- 3 **區間資料**(Interval data)：品評人員將被檢視的項目放置在數個連續性群組內。例如：三、四、五、六、七、八。
- 4 **比例資料**(Ratio data)：品評人員使用數字測量刺激相對於之前呈現的表名資料包含最少的資訊，而序列資料(數據)承載了較多的資訊且可用非參數型的統計試驗分析。區間及比例資料則更佳，因為他們可以被所有非參數型及參數型的方法分析。

我們偏好使用比例資料，因為它可以避免固定量度的間隔所造成的失真。然而實務中，最容易收集的是區間資料，而他們似乎可以提供相同的結果。(見第六章)

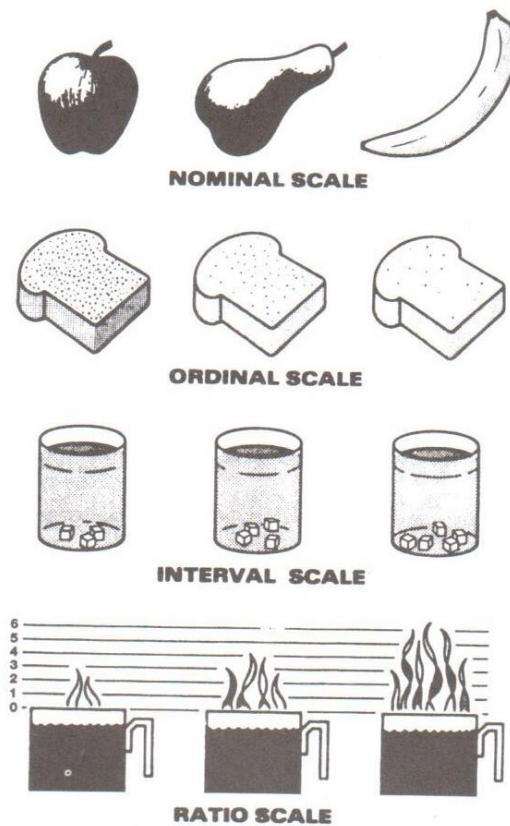


圖 5.1 測量標準的圖像說明。各種食物的名稱（蘋果、梨子、香蕉）提供了表名資料；而序列資料的例子中，三個黑麥麵包依照其所含有的香菜籽進行排序；在區間資料中，三杯飲料則依照連續性的方糖數量進行區隔；在最後一個例子中，使用氣相層析(GC)來測量三杯咖啡中揮發的兩種揮發物，而第一杯的揮發物被設定為第二杯的三分之二，以及第三杯的二分之一。注意圖說中顯示

了物理或化學量度。但品評人員的感官感覺可能不同，舉例來說：方糖從三塊增加到四塊感覺到的甜度增加可能大於從四塊增加到五塊。(引自 Cardello, A.V, and Maller, O. , Psychophysical bases for the assessment of food quality, in Objective Methods in Food Quality Assessment, Kapsalis, J.G., Ed., CRC Press, Boca Raton, FL, 1987.)

為了增加其複雜性，最常使用來測量品評反應的方法為：

- 分類法(Classification)：將要測量的項目使用表名資料做分類。如：使用烘焙程度分類各咖啡。
- 分級法(Grading)：商務上使用的老字號方法，將其從他類產品中脫穎而出。如：USDA 認證的肉品。
- 排序法(Ranking)：使用密度或者某一特性的濃度進行排序，來提供樣品(通常使用 3 組)，使用序列方式來定義範圍。
- 量度法(Scaling，或譯為量尺法)：品評員依據其訓練時所學得，使用一定範圍的數字(通常為 0 到 10)來判定樣品。類別量度(category scaling)產生序列資料，有時產生區間資料。線性量度(category scales)則產生區間資料。幅度預測量度(Magnitude Estimation Scaling)，雖然是被設計來產生比例資料，實務上似乎會產生區間／比例資料。

另一個更進一步的方法是 根據閾值所設定的氣味單位，將會在第八章討論。對於從眾多方法中挑選品評方法，並用來訓練其品評小組，品評小組的領導者必須了解兩個主要的變數來源：(1)各個品評人員對於偵測刺激的差異。(2)品評人員表達其感覺的差異(見第一章)。

每個人感覺(perception)的實際差異在感官資料中可變性很大，所以在感官分析過程中我們學習接納它，而心理學家則試著測量它。感官閾值(threshold)因人而異。在不同的閾值探討中，發現對啤酒中所添加的物質，在 20 位受過專業訓練的品評人員中，有 2 位的感官閾值比起平均值高於 4 倍，而有另外兩位則高於 5 倍。對兩百位未經過專業訓練但健康的正常人中，探討水中純化合物溶液，發現屏除異常值後，感官敏感性的最高值及最低值的差異竟達到一千倍。這代表了一個 4 或 7 人小組的判定有可能與大眾的感覺變異很大，因此這本書傾向建議品評小組規模至少在二十或三十人以上，甚至越多越好。一個規模過小的品評小組所判定的結果，除了代表他們自己外並沒有任何代表性。

第二個變異的來源：品評人員表達既定感覺的差異，也很可能造成巨大失真。幸運地是這可透過適當的訓練以及審慎地挑選提供給品評小組的用語來降低。提供給品評小組的感官表達用語的課程，只能藉由假設品評人員在進行感官評估時處於完全的茫然，且他們感覺得到正在研究中的產品特性，但是他們心中卻並不清楚此品評機制希望他們評量什麼。

為了選擇測量反應的方式，一般來說感官分析選擇最簡單的方式來測量各個樣品之間的差異，因此可以減少訓練品評小組的時間。有時使用越複雜的方式，則品評人員選擇的詞彙越趨複雜，事前訓練及事後的評量時間也就越長。例如：有些樣品的差異也許在計畫階段就沒有被考慮到，則在簡易的品評方式中也就可以忽略不計。因為當品評人員累積的訓練使其越趨專業，整體的培訓時間可能會越來越少，也就越來越不需要針對個別的測試進行事前訓練。

5.2 心理物理理論(Psychophysical Theory)

心理物理學在實驗心理學的領域著重在研究感官刺激及人類反應的關係，也就是在進一步理解人類感官系統是如何運作。大學中的心理物理學家持續地在改善測量人類反應的方法，而在感官分析領域，我們必須了解此技術且確實地與之實驗合作。本章將會提供應用於感官分析中的心理物理學的綱要，若對進一步的細節有興趣，可以參考列於本章節一的參考列表。

心理物理學主要的重點是在探討心理物理功能的形式，介於刺激 C 及其所產生的感官 R 的關係，進一步用數學方程式來表示 $R=f(C)$ (見圖五之二)。

氣味強度

丙酸濃度。 ($\log$ ppm)

乾淨空氣

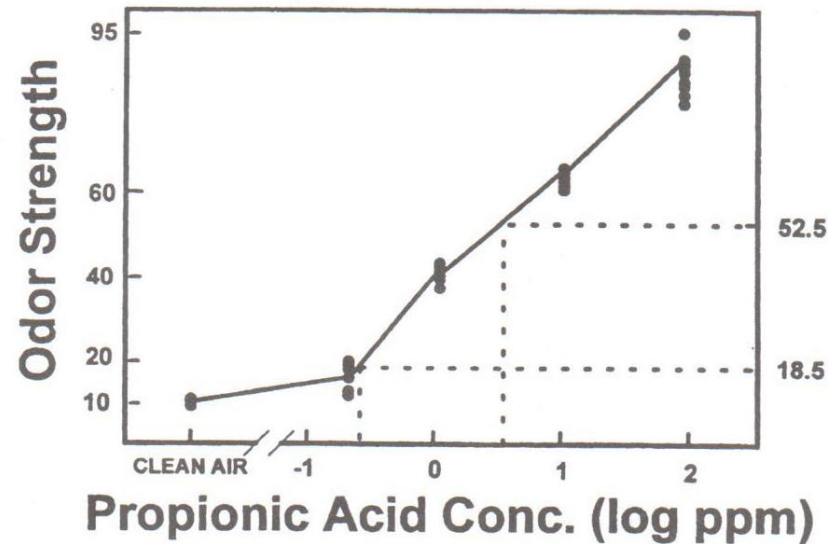


圖 5.2 心理物理功能的例子。氣味強度被評為 0 至 99。0 等於無異味或鼻腔刺激的感覺（引自 Kendal-Reed et al., Human responses to propionic acid, in Chemical Senses 23. Oxford University Press. 1998, 71—82.）。

當刺激(stimulus)是眾所皆知的（外加濃度）或容易測量的（一個高峰，或是一個拉伸(Instron) 讀數），便是感覺(sensation)導致了困難。我們必須問品評人員的問題像是：「判斷這氣味的範圍，從 0~99 的量度」、「感覺是兩倍強或是三倍強？」、「哪些溶液有最強的奎寧味道？」，但沒有人能準確地回答這些問題。我們正在運用各種實驗技術，例如：與第二組做對比，較佳的已知感覺(如找尋音調的響度，這被稱為**跨模態匹配**<cross-modality matching>)，或當受測者的內耳正在運作時，直接電測鼓索（味覺神經）所產生的神經衝動。

在過去的一個世紀，兩種形式的心理功能正被廣泛使用：費希納定律(Fechner' s Law)和 Stevens 定律。雖然兩者都不是完美的，卻分別(前提是在有限的正確使用下)提供了比起簡單直覺更好的實驗設計指導。最近，酶化學領域已知的米氏方程(Michaelis-Menten

equation)，或從它衍生的 Beidler 方程式，都已被用於模擬劑量-反應(dose-response)關係（見下面 C 節）。

5.2.1 費希納定律(Fechner' s Law)

如費希納選作測量人類感覺的強度：最小可覺差（Just Noticeable Difference, JND）；見圖 5.3。例如：他認為 8 JNDs 所感覺的感覺是 4 JNDs 的兩倍強度。當 JNDs 剛剛通過差異測試時，費希納從萊比錫大學的 Ernst Weber 得知此事，在 18 世紀中葉。Weber 發現差異閾值與其初始絕對刺激強度成比例地增加，他們使用下列公式描述：

$$\frac{\Delta C}{C} = k \text{ (韋伯定律) (5.1)}$$

絕對閾值

刺激強度 (C)

(任意單位)

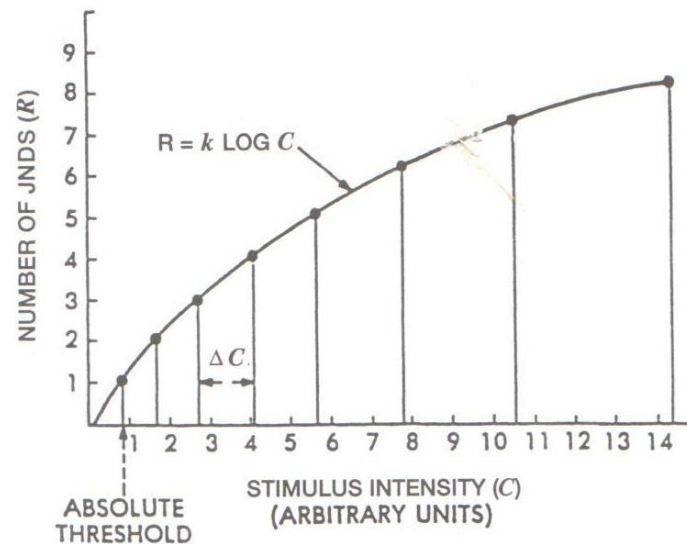


圖 5.3 通過總結 JNDs 的方法導出的費希納定律。

其中 C 是刺激的絕對強度，例如：濃度。 ΔC 是 1 JND 必需刺激強度的變化，而 k 是一個常數，通常是介於 0 和 1 之間。韋伯定律表示：增加的風味可被感測出的量取決於該風味原本存在的量。如果 k 常數已經確定，就可以計算出多少額外風味量是必要的。費希納定律的實際衍生：

$$R = k \log C \text{ (費希納定律) (5.2)}$$

此定律是複雜的，並且取決於許多假設，其中的一些可能不會成立 (Norwich 和 Wong, 1997)。我們可從常見的類別量度中發現費希納定律的支持。當小組成員(品評員, panelist)針對一些樣品給分，而分數沿著一個維度上變化（例如：甜味）。當量度從 0 到 9，並將結果繪製成類似於圖 5.3 的對數曲線。費希納理論的一個明確成果是聲音分貝 (Decibel) 強度的對數量度，分貝量度 (Decibel

scale)。

5.2.2 蒂文斯定律(Stevens' Law)

在費希納之後在哈佛大學工作了將近一個世紀的史蒂文斯(Stevens)指出，如果方程式(5.2)是正確的，100 分貝的聲調應該比起 50 分貝只大聲兩倍。隨後，藉由幅度預測量度(Magnitude Estimation Scaling)的幫助下，他發現了 100 分貝的聲調是 50 分貝的 40 倍響。Steven 主張感覺幅度的成長是刺激強度的冪函數，可以在數學上表示為：

$$R = kC^n \text{ (Stevens 的冪定律<power law>)} \quad (5.3)$$

其中， k 是一個常數，取決於我們選擇來測量 R 和 C 的單位，而 n 是冪函數的指數，即 n 是感覺強度的生長速率的測量，作為刺激強度的函數。圖 5.4 顯示了冪函數，其中 $n = 0.5$ 、 1.0 及 1.5 ，而表 5.1 列出了在各種感官屬性中典型的指數。視覺長度的指數為 1.0 ，也就是說之間有簡單的比例關係，因此在分類品評強度時常使用線性量度。

縱坐標：接收強度(R)

橫坐標：物理密度(C)

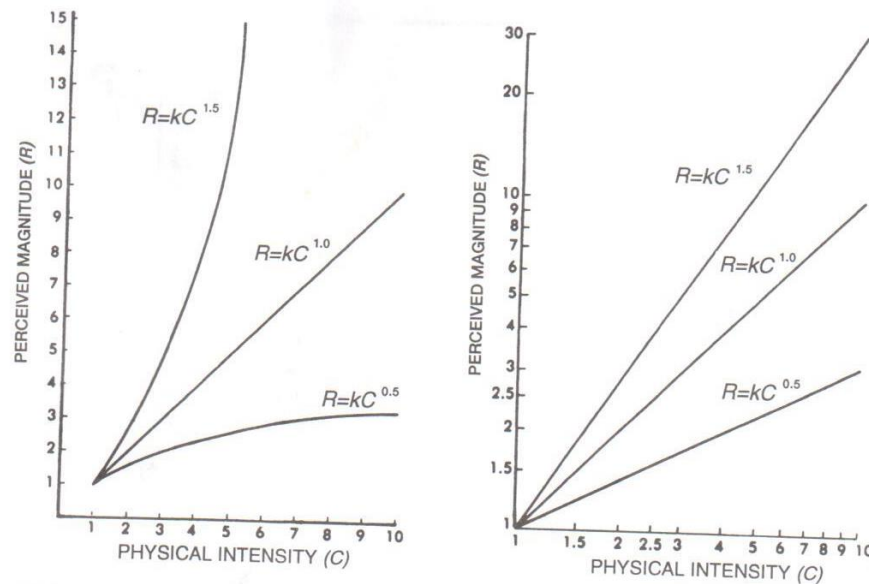


圖 5.4 冪函數的繪圖。當 $K=1$, $n=0.5$ 、 1.0 及 1.5 的的冪函數(左圖)，及其對數函數(右圖)

當 n 大於 1.0 ，則接收的感覺強度的增加比刺激還快；一個極端的例子是電擊（見表 5.1）。相反地，當 n 小於 1.0 時，對許多氣味而言，感覺強度增加的速度慢於刺激，而其曲線結果表面上類似於圖 5.3。

Stevens 建議在接收感覺的測量中，只有比例量度是有效，但幅度預測量度可能已經廣泛使用於心理物理學實驗室。然而，許多作者（Cardello 和 Maller, 1987）已經指出，對食品及香料的感官評價，存在著嚴重的缺陷。指數隨著刺激的範圍和所使用的模組而變化，並且更糟的是，指數在調查者和受試個體間的差異很大，因為受測者對數字的個人習慣使用(idiosyncratic use)不同。

5.2.3 Beidler 模型

對數函數及冪函數僅僅是可以套用在感官評估數據中的數學方程式，兩者之間並沒有什麼生理學上的邏輯。McBride (1987) 提出

了下面公式，該公式是 Beidler (1954, 1974) 從動物實驗和米氏方程式中衍生出用在生物系統中酶－基物關係的動力學，同樣可以套用在人類味覺的反應上。McBride 建議我們從受測者是使用數字或是比例的問題脫身，並應該簡單假設人類的心理反應與神經生理反應具有比例關係：

$$\frac{R}{R_{MAX}} = \frac{C}{k+C} \quad (\text{Beidler 方程式}) \quad (5.4)$$

表 5.1 在各種感官特性中之冪函數典型的指數

感官特性	指數	刺激
苦味	0.65	奎寧，啜飲
	0.32	奎寧，連飲
亮度	0.33	5° 域
寒冷	1.0	手臂上
持續時間	1.1	白色訊號噪音
觸電	3.5	電流通過手指
硬度	0.8	壓縮橡膠
重度	1.45	舉重
明度（視覺）	1.20	灰色紙
響度	0.67	1000 Hz 的聲音
鹹味	1.4	氯化鈉，啜飲
	0.78	氯化鈉，連飲
氣味	0.55	咖啡
	0.60	庚烷
酸味	1.00	氯化氫，啜飲

甜味	1.33	蔗糖，啜飲
觸覺粗糙度	1.5	金剛砂布
熱燙感	1.0	皮膚上的輻射熱
震動	0.95	手指上 60Hz
	0.6	手指上 250Hz
黏性	0.42	攪拌液體
可視面積	0.7	平方預測
視覺長度	1.00	線型預測
暖度	1.6	金屬放手臂上

(改編自 Adapted from Cardello, A.V. and Maller. O., Psychophysical bases for the assessment of food quality, in Objective Methods in food Quality assessment, Kapsalis, J.G., Ed., CRC Press, Boca Raton. FL, 1987. With permission.)

該方程表明，反應， R 除以最大反應 R_{max} ，對刺激 C （莫耳濃度）的函數（見圖 5.5）顯示出了 S 形關係。而常數 k ，是在一半反應時的濃度。Beidler (1974) 稱它為調和常數(association constant)，或結合常數(binding constant)，並指出它可以被看作是與該刺激分子與其接收者結合的方法。Beidler 模型適合中高領域的感官印象，例如對甜食或飲料的甜味。不同於 Fechner 和 Stevens 模型，它假定反應有一個上限 R_{max} ，這是無論刺激濃度多高都不會被超過。它被視作是可讓所有接收者飽和的濃度。

McBride 使用一些糖、鹽、檸檬酸和咖啡因的例子來表示，Beidler 方程式提供了描述人類味覺反應的一個好方法，而 JND 通過兩個心理物理學方法，可累積和分類等級。Beidler 方程式的應用允許我們測量人類味覺反應中迄今無法得到的參數， R 和 k 。因此，相較於費希納法則和史蒂文斯規律，Beidler 方程式的應用提供了測量在人類味覺反應中無法獲得的參數，也就是心理物理運作的方式。

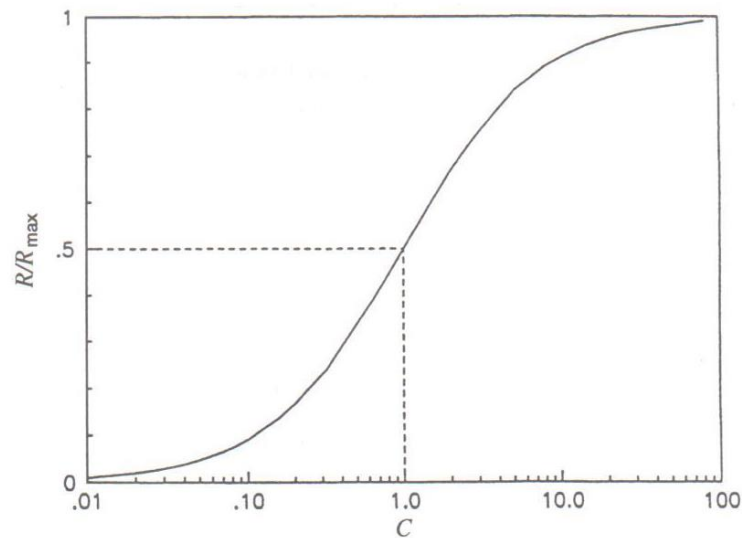


圖 5.5 味覺反應， $R/R_{\max}$ 和刺激濃度， C ，之間的 S 形關係。就像 Beidler 方程式裡所指定， k 可以為了方便設定為 1。曲線的反折點(最大斜率)發生在 $R=0.5 R_{\max}$ ，且 $C = K$ 時。

其他試圖為品評員的判定過程建立模型的技術(經常在心理物理學中被使用)，也應用於感官評價，特別是在閾值和差別測試。這些其他心理物理模型包括 Thurston-Ura 模型及信號偵測模型(Signal Detection model)。

5.3 分類法(Classification)

在分類測試中，受測者被要求選擇一個或多個屬性來描述刺激(Stimulus)。例如：在飲料試驗中，受測者標記能最佳地描述樣品

的術語(複選)：

_____ 甜味 _____ 酸味 _____ 檸檬味
_____ 混合 _____ 厚實 _____ 清爽
_____ 果肉 _____ 自然 _____ 回味

不嘗試將用語標準化，並且結果以每個用語的標記總數報告。這樣的數據是表名類別：沒有使用數字，並且沒有增加或減少數據顯示的序列。例如：一批蘋果可以以主要顏色分類（紅、綠、黃）。

適當地選擇用詞來正確地描述刺激是必要的。如果小組成員沒有受過訓練，也就是和一般消費者一樣的情況下，必須使用一般常用語而非術語。混亂的根源在於受測者經常錯誤地勾選個人常用語。下面的注意事項描述了分類中選擇合適的單詞/術語/分類是關鍵的第一步。選擇最佳的術語是重要的，不僅在分類試驗是必要的，在所有的量測技術中，都應該使用正確的術語來描述及定義感覺到的屬性。

感官屬性的選擇和其相對應的定義應與產品的真實化學及物理性能密切相關。若能貼切一產品被感知的真實流變或化學性質，便能使數據更易於解釋，且對決策更有幫助。

注意事項：

- 一，如果一個產品有明顯的瑕疵，如腐敗或酸敗，而用來描述這種瑕疵的術語還沒有被列入名單，小組成員（尤其是未經訓練者）會使用列表中的另一個術語，以表達此不在名單上的感覺。
2. 如果提供給小組成員的清單沒有提及一些能描述各產品之間真正的差異特性，或者忽略產品的一些重要特性，小組成員會再次地使用列表中的另一個用詞來代替他們所感覺到的。
3. 因此，如果希望產生的結果是有用的，對於分類中術語的選擇（及之後第六節所討論各種形式的評等）必須根據實際產品的特性。這必須在試驗前，先經由訓練有素的品評人員做樣品的預檢，以確保所有適當的屬性被列出。若使用先前的研究或之前的列表，需要再次確認各個屬性，因為目前產品的部分重要屬性可能被忽略，或者是舊列表中有許多術語是無關於目前的測試，從而引起小組成員的混淆。

以下是分類工作或隨後的相關的評等工作中常用的術語列表：

1. 保健品。例如：光滑的、油膩的、柔順的、柔軟的、顆粒感、蠟質的、油質的、收斂毛孔的、緊繃的、乾燥的、潮濕的、滑膩的。

需要注意的是各個屬性之間並沒有參雜類似的參數關係（濕潤/乾燥，光滑/粒狀等）。

2. 香料（受測者可能會被要求在複合的香氣中定義是哪些香料或草本）：牛至、羅勒、百里香、鼠尾草、迷迭香、馬鬱蘭、丁香、肉桂、肉荳蔻、肉荳蔻乾皮、小荳蔻。

5.4 分級法(Grading)

分級是商業上常用的專家評價法，通常有 4 至 5 個步驟：「挑選」、「特別」、「普通」及「拒絕」。進行這類品評分級法的常見商品有咖啡、茶、香料、奶油、魚和肉。品評等級最常涉及分級者的整體感覺過程。

分級者被要求對許多混合物之中的單一特性進行評分，或該混合物和這些內含屬性的平衡，或是負面特徵的組合效果，或根據一些書面或物理標準來比較分級產品。

分級法在商業中提供了一個相當有效且系統性的方式，避免消費者為了低品質的產品付出過於高昂的代價，同時允許生產者得以支付提供高品質產品而產生的額外成本。然而，分級法一個相當大的缺點，即其與可測量的物理或化學屬性的統計關聯是困難且不可能的。

因此，許多歷史悠久的分級量度正在被本書中所介紹的方法取代。良好的分級方法仍在使用的例子是測量魚類新鮮度的 Torry 分級法（Sanders 和 Smith, 1976），以及奶油（USDA, 1977）和肉類（USDA，無日期）的美國農業部(USDA) 分級量度。

5.5 排序法(Ranking)

在排序法中，受測者接受 3 個*或更多個樣本，將強度按照順序排序，或按照指定的屬性程度排序。例如：將 4 個優格樣品依照感官酸度做排序，或 5 個早餐穀類的樣品按照偏好程度排序。關於排序方式的描述和他們的統計處理方式會在第七章中做完整說明。

對於每一個受測者(subject)，排序第一的樣品被標示“1”，排在第二位的標示“2”等等。每個樣品得到的數字相加，所得數字及代表其樣品的總排序。排序的順序無法在強度方法裡有意義的被使用，但他們是經得起如 X2 測試（參見第十三章）和 Friedman 測

試（參見第七章）的顯著性檢驗。

排序測試法快速且品評人員需要相對較少的訓練，儘管不應該被忽略的是受測者必須徹底熟悉要測量的屬性。排序測試有廣泛的應用，但若是三個以上的樣品組，他們無法如使用量度般來作區分。

5.6 量度法(Scaling)

量度法使用數字或文字來表達感知屬性（甜度、硬度、光滑度）的強度、或這種屬性的反應（如太軟、恰到好處、太用力）。如果使用文字，分析人員可以分配數字值（如極度喜愛= 9，極度厭惡= 1），使得數據可以被作統計處理。如此處所述及，量度法正在世界各地被深入的研究（ISO, 1999; Munoz 和 Civille, 1998），而其建議的結果應該被看作是初步的資料。

量度法的技術有效性和可靠性高度依賴於：

- 量度技術的選擇足以包含參數強度的全部範圍，並且量度之間具有足夠距離以區分樣品之間所有微小差異的強度
- 品評人員已經或尚未被教導一特性感覺如何與待測量度的特性建立關聯性。
- 在所有的樣品及所有品評時段，品評人員已經或尚未被訓練如何使用同樣的方式來鑑別。（見第 9 章關於專門小組成員的訓練）。

與差異性測試相比，量度法提供了更多資訊，因此也是記錄感覺強度的一個更有用的方法。就像排序法一樣，其結果的準確度很大程度上取決於小組成員是否熟悉此樣品的特性範圍。在這方面，有以下三個不同的理念（引自 Munoz 和 Civille）：

- 通用量度(Universal scaling)：此量度下小組成員認知的參考點是他們已經歷此樣品的最高強度（例如：在 15 點制的香味頻譜(spectrum)的量度中，肉桂口味在 Big Red 口香糖中為 13 點。）
- 產品特定量度(Product specific scaling)：小組成員考慮的最高參考點在其經驗中的所選產品類別裡（例如：香草印象在典型的 15 點制的香草餅乾類別中被定為 10 點。）
- 屬性特定量度(Attribute specific scaling)：小組成員設定的最高參考點是此特殊屬性在所有產品上表現的經驗（例如：一個特定的牙膏被評定在所有薄荷芳香的產品中 13 點）。

一個常見的問題是小組成員傾向於只使用量度中間的部分。例如：在蘋果酒的蘋果風味判斷中，強度的量度從 0 到 9，受測者將避免使用數字 0、1 或 2，因

為它們往往假設蘋果酒中不會有蘋果風味強度如此低的樣本。同樣地，數字 7、8 和 9 也很容易不被使用，因為品評人員也預期非常高強度的樣本可能性低。其影響的結果是量度法的準確度被扭曲：例如：蘋果風味顯著的蘋果酒可能有 6.8 的強度，卻可能被品評人員評出略高於平均的 6.2 度。

雖然從任何情況得到的屬性量度各不相同（例如：測試者的經驗，屬性的熟悉程度），它通常假設如下：

- 類別量度（ISO 術語：評級<rating>）通常得到序列（或區間）數據；
- 線性量度（ISO 術語：評分<scoring>）產生區間數據；
- 幅度預測量度（通常稱為比例量度<ratio scaling>）有時但非總是，產出比例數據。

5.6.1 類別量度(Category Scaling)

類別（或分區）量度測定的方法，受測者被要求在一個有限的，通常是數字量度，分配一個值（類別）的來評定特定刺激的強度。類別量度的數據通常被認為是量化性數據(ordinal level data)。它們通常無法測量一個樣品多於另一個樣品強度多少。在 7 點硬度測試中，一個 6 點硬度的產品不一定比 3 點硬度等級的產物還硬兩倍。而 3 與 6 的硬度差距可能也與 6 和 9 之間不相同。雖然類別量度的創造是為了鼓勵小組成員使用所有的區間，然而小組成員仍可能傾向平均的使用各個類別，或者為了避免太過極端而不使用前後兩個區間的值。下面是描述性分析有用的證明：

編號	類別	表名類別量度 I	表名類別量度 II
0	0	無	完全沒有
1	1	閾值	剛好察覺
2	2.5	很輕微	很溫和
3	5	輕微	溫和
4	7.5	輕微-適中	鮮明溫和
5	10	中等	鮮明
6	12.5	中等-強	鮮明-強
7	15	強	強

一般情況下，即使表名類別量度轉換為數字。在上述列表中使用的數字即是典型的被轉換的表名類別量度。在風味檔案和質地檔案的描述性分析方法使用了數值型的類別對應表名量度：

數值	對應表名量度
0	無
)(	閾值(剛好察覺)
1/2	非常輕微
1	輕微
1 1/2	輕微-適中
2	中等
2 1/2	中強
3	強

除非感官知覺的範圍是在一個非常小的範圍內，或待測試樣本的數目少（小於5），品評領導人應考慮使用至少10到15點的類別範圍。對於類別範圍的數據可以使用 χ^2 -檢驗來比較在樣品在每個類別中得到的反應分析。或者，如果此類別可以合理假設各量度的間隔相等，其數據可以應用於參數技術如 t-檢定、變異分析和回歸等。Riskey（1986）討論了相當詳細的類別量度的使用及濫用。量度的建設實際步驟在第九章及第十章有詳細的討論。附錄 11.1 包含了已證明實用的量度點的詞目廣泛選擇。附錄 11.2 提供了為超市中販賣的完成加工品所測試的四種基本口味、選擇香味強度，味道，質地特色等的 0~15 基準點的規模為，如 Hellmann' s 美乃滋。

5.6.2 線性量度(Line Scaling)

具有線性或線的量度，小組成員針對一個給定的刺激評定其在一個水平線上對應感知強度。最常用的長度為15公分和6吋，定位點從兩端 1/2 吋或 1.25 公分處做標記（見圖 5.6）。使用兩個以上定位點傾向於降低線性量度到一類別量度，這可能理想或不理想。通常情況下，量度的左端對應於“無”或刺激量為零，而量度的右端代表大量或非常強烈的刺激水平(Anderson,1970; Stone 及 Sidel，

1992)。在某些情況下，量度是雙極性的，亦即：相反類型的刺激被用於定位點的端點。

小組成員在線性量度上做標記來代表所測樣品的感知強度。從線量度得到的標記將手動轉換為數字，藉由用一把透明尺覆蓋在線性量度上，測量其在量度上的位置，或者可使用掛接到電腦的數字化轉換器。數字化轉換器的標誌為數字的位置，根據預先設定的程序，將數據回饋到電腦以供分析。

5.6.3 幅度預測量度(Magnitude Estimation Scaling)

幅度預測 (Moskowitz, 1977; ISO, 1994) 或任意數字匹配是基於 Stevens 定律的量度技術 (見節 IIB)。受測者接收的第一個樣本被分配一個自由選擇的數字。(數字可以由實驗者分配，在這種情況下，它被稱為一個標準值(modulus)，或可以由小組成員分配) 受測者接著被要求依照與第一份樣品的比例分配後續所有樣品的評分。如果第二份樣品比起第一份還多出 3 倍強度，所分配的評級應該為分配給第一個刺激的三倍。



圖 5.6 典型的線性量度。

品評人員依照要求按其感覺評分。例如：

1. 具有標準值(modulus)：將第一個餅乾的脆度評定為 25，並且依比例評定剩下的其他樣品。假設某樣品的脆度是第一個餅乾的一半脆度，則它應被分配 12.5 的脆值。

第一個樣品 25

樣品 549 _____

樣品 306 _____

2. 不具有標準值：測試第一個餅乾的脆度並為其評定任何分數，之後依比例評定剩下的其他樣品。

樣品 928 _____ (第一個樣品)

樣品 549 _____

樣品 306 _____

測量的結果如 Moskowitz (在上述引文中) 和 ISO (在上述引文中) 所述。此評量的替代方法是由 Butler 等 (1987) 和 Lawless (1989) 所提供。

5.6.3.1 幅度預測 vs. 分類量度(Magnitude estimation vs. category scaling)

Pangborn (1984) 比較了幅度預測量度法與類別量度法的優點和缺點，通過幅度預測量度法 (ME) 所得到的數據擁有比例性，像是技術測量法的標準格式 (長度，重量，體積等)。幅度預測量度法 (ME) 的問題是受測者傾向避免使用兩端點的值，以便有未來極端刺激出現時仍留有餘地。另外相信幅度預測量度法 (ME) 的人也認為類別量度法 (CS) 必須花費過多時間和精力以準備量度標準，並且訓練受測者使用它們。而偏好使用類別量度法 (CS) 的人則認為：幅度預測量度法 (ME) 無法在香味強度測試中提供穩定和可重複的值。在實務中，幅度預測量度法 (ME) 仍需要妥善地訓練其品評人員，假如以任何設備使用方法，許多品評員評估”微不足道(nickles and dimes)”使用全數或，或偏好一系列的 10s 或 5s，如 15、20、25 等等，且他們無法使用數字表達如“六倍強”或“1.3 倍弱”等。在許多應用中 (Powers 等，1981; Giovanni 及 Pangborn, 1983; Pearce 等，1986; Lawless 和 Malone，1986)，幅度預測量度法 (ME) 沒有比起類別量度法 (CS) 提供更多的鑑別度。此外，幅度預測量度法 (ME) 也不適合用於喜好程度的測試中 (Pangborn 等人，1989)。然而，在少數定位點 (20 以內) 的學術應用中，幅度預測量度法 (ME) 確實提供了更多的鑑別度和區分，比如像是水中的蔗糖、稀釋劑中的香精、或是調整一項屬性如一維系統的研究。

5.6.3.2 幅度匹配(Magnitude matching) (交叉形態匹配<cross-modality matching>)

在該技術中，受測者比較一項屬性(1)的濃度，例如酸溶液中的酸度，與另一個屬性(2)，強度如 1000 赫茲的音調響度。如果兩個強度由函數表示

$$R1 = k1C1' \text{ 或 } \log R1 = \log k1 + n1 \log C1$$

$$\text{且 } R2 = k2C2^2 \text{ 或 } \log R2 = \log k2 + n2 \log C2$$

藉由比較它們，我們得到：

$$\log k1 + n1 \log C1 = \log k2 + n2 \log C2$$

$$\log C1 = n2/n1 \log C2 \text{ 加上一個常數}$$

換言之，藉由其描述酸度的方式我們得到了一個功率函數，並且此函數的指數等於其原本指數的比例 (Cardello 和 Mailer 1987; Lawless 和 Heymann, 1998; Marks 等, 1988)。這種方法的優點是，沒有數字再被分配，所以它得以繞過受測者使用不同數字的傾向，如上所述。

5.7 品評方法分型

品評方法分為兩種類型：分析型和消費者型，見表 5-2。分析型包括敏感度、差異性和描述性等測試，消費者型則指喜好性/接受性測試。

表 5.2 感官評估方法分類

敏感度	分析型			消費者型	
	差異性 (整體/屬性)	描述性	特性強度時 間法	喜好性/接受性	
	單項比較法	風味描述法		單項比較法	
	多項比較法	定量描述法		多項比較法	
	Scheffe 比較法	質地描述法		排序法	
	二三點比較法	稀釋描述法		評分法	
	三角測試法	標準差離描 述法			
	排序法				
	評分法				
	比例法				

第六章 總體差異測試

6 Overall Differende Tests

6.1 引言

試驗分析型包括敏感度、差異性和描述性等測試。**差異測試**(Difference tests)是分析型測試的一種，其目的是找群樣品間的差異。第六章和第七章包含了各種差異測試的例子，如同食譜般的詳細描述。其基本理論呈現在第五章測量反應中，且在第十三章中闡明了基本的統計方法。對於選擇特定測試的準則將在“範圍與應用(Scope and Application)”中找到，或在第十五章指導選擇技術中以結論的形式呈現。

差異測試可以有數百種不同的方式能合理地運用，但在實務中此處所描述的方法已經有悠久的使用歷史和經驗。兩組差異測試的特徵如下：

- **總體差異測試**(Overall difference tests) (又稱為**異同測試**, 本章)：目的在探討樣本之間是否存在總體感官差異？如三角測試法(Triangle Test)和二三點測試法(Duo-Trio Test)，其目的是為了測試受測者是否可以在樣品間檢測出差異。總體差異測試常用的方法包括配對比較法(paired comparison test)、二三點測試法(duo-trio test)、三角測試法和參考比較法(difference from control method)。
- **屬性差異測試**(Attribute difference tests) (又稱為**區別測試**, 第七章)：目的是探討各樣品之間的特定屬性 X 有什麼不同？受測者被要求集中注意於某一屬性(或某幾個屬性)，並且忽略其他屬性。例如：「請按甜度排列這些樣品」。實例如配對比較法(Paired comparison test)、n 選項比選法(Alternative Forced Choice test, n-AFC test)，以及各種多重比較的檢驗法。可以測試出所選定屬性的感受強度，可用第五章中描述的任何方法來測量，例如：排列、線性刻度(line scaling)或幅度預測(magnitude estimation, ME)等。

2-選項比選法(2-AFC test)和 3-選項比選法(3-AFC test)常用來測定閾值。喜好性測試(Affective test) (偏好測試，或消費者測試)也是一種屬性差異測試。

相似性測試(Similarity test)曾被視為一個不同的主題，因此提供了單獨的表格。但在本書中採用現代統一的方法，其中單張表就涵蓋了從差異到相似測試的所有情況。

6.2 差異測試及相似性測試的統一方法(The Unified Approach to Difference and Similarity Testing)

差異測試(Discrimination test)可用於解決各種實務目的。在某些情況下，研究人員感興趣的是呈現兩個樣品的差異。此外，有時研究員想要確定兩個樣品有足夠的相似性到能夠互相替代。又有時，某部分研究人員希望樣品表現出差異性，但同時參與同一研究的其他研究人員則希望發現其相似性。所有這些情況透過選定測試敏感度各參數(α 、 β 和 P_d)的適當值，可以使用統一的方式處理。而如何選擇這些值則取決於測試的特定目標。(α 誤差又稱為第一類誤差。是指”無差異”為真實時被判斷為有差異之錯誤機率。 β 誤差又稱為第二類誤差。是指”無差異”不成立時被判斷為無差異之錯誤機率。通常 α 愈大， β 就愈小。)

Microsoft Excel 已經開發出電子表格的應用，可幫助研究人員決定 α ， β ，和 P_d 值，並使其可在所需的測試靈敏度和可利用資源之間達到最佳妥協。

「測試敏感度分析儀 (Test Sensitivity Analyzer)」透過品評員數量 n 、正確反應 x 的數目、最大可容許區分值 P_d 的比例，及在 α -風險和 β -風險下所影響的各種情況的不同組合，使研究人員能夠快速操作各種情況。統一方法也適用於配對比較測試，如 2-選項比選法(2-AFC test)。

在差異的基本三角測試法中，僅僅是為了發現兩個樣品之間是否存在可察覺的差異。統計分析是在只有 α -風險（假設可察覺差異存在，但事實上並沒有的機率）的默示前提下進行。品評員的數量由檢視 α -風險表並考慮到顧及資源而求得，如品評員的數量、可供測試樣品的品質和其可用數量等。 β -風險（假設不存在可察覺差異，但事實上存在的機率），以及比例區分值 P_d ，則被品評小組忽略，或者假定是不重要的。其結果是為了求得測試差異且保持所需要的資源在合理範圍內，研究者可能選擇使用極小的 α -風險值，並且接受任意大的 β -風險值和 P_d 值。

在感官分析員的相似性測試中，是為了要確定兩個樣品是否有足夠的相似性到可以互相替換使用。此情況常見的例子像是為了降低成本而改變配方，或驗證供應商的更換等。為了測試相似性，研究人員通過選擇極小的 β -風險值以確保只有很小的機會真的存在差異，而允許 α -風險非常大，以保持其需要資源在合理範圍內。

在某些情況下，平衡差異存在的 β -風險及差異不存在的 α -風險可能是重要的。而在這個例子中，研究人員決定了三個參數 α 、 β 和 P_d 以達到實驗所需的靈敏度以及其擁有的資源條件（見實例 6.4）。

統計學顯著結果的經驗法則為：

- α 風險處於 10-5% (0.10-0.05) 代表有中等的證據顯示，存在顯而易見的差異；
- α -風險 5-1% (0.05~0.01) 代表有力的證據顯示，存在顯而易見的差異；
- α -風險 1-0.1% (0.01-0.001) 代表有非常有力的證據顯示，存在顯而易見的差異；
- α -風險低於 0.1% (<0.001) 代表極端有力的證據顯示，存在顯而易見的差異。

對 β -風險而言，使用與上述相同的標準來評定（將「不明顯的」替代成「明顯的」），差異是不明顯評估的證據強度。

區分值的最大可容忍比例， P ，分為三個範圍：

- $P_d < 25\%$ 代表小的值；
- $25\% < P_d < 35\%$ 為中等大小的值；和
- $P_d > 35\%$ 為大的值。

6.3 三角測試法(Triangle Test)

三角測試法的章節，是這本書中第一個相當複雜且包括很多細節的方法，因為：(1) 所有的感官分析師應該知道；(2) 與許多方法相像；且(3) 因而可省略隨後的方法。統一方法的應用在實例 6.3 和 6.4 中描述。

A. 適用範圍及應用

當測試目的是為了確定兩個產品之間是否存在感官差異，使用這種方法。當測試在不同處理方法下產品所產生的變化，且其不能簡單地歸納到一個或兩個屬性的情況時此方法特別有效。雖然在統計上它比起配對比較法和二三點測試法(Duo-trio test)更有效率，但三角試驗仍有其限制：產品特性涉及感官疲勞，轉移或適應性，且品評人員對測試三個樣品感到混淆等情況時。這個方法在以下某些情況下有效：

1. 為了判斷產品的差異是否因為成分調整、工序、包裝或儲存條件等而產生。
2. 為了判斷總體差異是否存在，是否沒有特定屬性（多個）可以被識別出而已受影響。

3. 為了選擇和監督評估小組成員判斷差異的能力

B. 測試原則

提供品評人員三個編碼的樣本，並告知他們其中兩個樣品是相同的，一個是不同的（或有變異的）。要求品評人員進行測試（感覺，檢驗），從左至右每一件產品並選出變異樣本。參閱表 T8 *計算正確答案的數量。

C. 測試人員數目

通常三角測試法需要 20 至 40 名受測者，但是當所測試的差異很大且容易被發現時，也可以用少至 12 人的測試結果。在另一方面，相似性測試需要 50 至 100 個受測者。作為最低要求，所有受測者應該熟悉三角測試法（形式、任務、品評過程），且熟悉其測試的產品，尤其是因為風味記憶在三角測試法中佔了很重要的部分。

建議在測試前舉行事前的訓練指導，讓受測者熟悉測試程序和產品特性。必須小心提供足夠的信息以便指導和激勵，但同時不讓受測者因特定訊息而對測試產生偏差。

*請參閱本書的最後一章“統計表”表格編號 TI 到 T14。

三角測試法

名稱_____ 日期_____

樣本類型_____

說明

從左到右測試樣本，其中兩個是相同的，判斷哪個不同。

如果沒有差異是明顯的，你必須猜測。

樣品組別

哪一個不同？

評論

圖 6.1 三組三角測試的評分表範例

D. 測試程序

測試控制（詳見第三章介紹）應包括分區試驗區，每個品評員在其中可以獨立工作。照明的控制可能是必要的，以減少任何顏色變量。在最佳條件下準備並呈現待研究的樣品，也就是說，樣品必須看起來開胃且被良好地呈現。

如果可能的話，同時提供所有樣品，然而，若樣品是體積大，或有餘味，或在外觀上有細微的差別，則在測試不會失效的前提下可以依照順序提供。

準備數量相等的六種可能的組合（ABB、BAA、AAB、BBA、ABA 和 BAB），並隨機呈現給受測者。要求受測者從左至右的順序測試（味覺、感覺、嗅覺等）樣品。正在進行中的測試可以重複測試。

如圖 6.1 所示，評分表可以提供給多於一組的樣品。然而這只能在感覺疲勞最小化時提供。在選擇有異樣品的同時，不要問偏好、接受度、差異程度，或任何有關差異性的問題。這是因為受測者可能因為回答這些額外的問題，而在選擇有異樣品時發生偏差。對這類問題的研究可以通過額外的測試來獲得。請參見第十二章偏好和接受度測試，和第七章，對大小或屬性的的差異測試。

E. 結果的分析和解釋

統計正確答案的數目（正確識別有異樣品）和全部回答的數目。檢索 T8 表確定是否正確回答的數目等於或大於此表所指示的數目。不能計算“沒有差別”的回答為有效回答。若受測者找不出差異，應指導受測者猜測答案。

F. 例 6.1：三角差異測試法 - 新麥芽啤酒

受測的啤酒“B”是用新的一批麥芽釀造，且感官分析師想知道它是否與目前生產的啤酒“A”有感受上的差異。5%的錯誤風險是可以接受的，且有 12 位訓練有素的品評人員可使用。準備 18 杯啤酒“A”和 18 杯啤酒“B”，這是為了兩套隨機分佈的 12 個組合 ABB，BAA，AAB，BBA，ABA 和 BAB。

8 名受測者正確識別出有異樣品。校閱表 T8 後得出的結論是，這兩種啤酒在 5% 的顯著性上是不同的。

G. 例 6.2：三角差異測試的詳細例子 - 巧克力條包裝：錫箔 VS. 紙

問題/狀況：巧克力公司的包裝部總監希望測試目前用於包裝的紙包裝與新的錫箔包裝材料是否存在有效的差異。初步觀察表明，紙包裝的巧克力條 3 個月後表現出質地較硬，而錫箔紙包裝的巧克力條仍保持柔軟。總監認為，如果他能在 3 個月時呈現樣品間存在顯著差異，他便可以證明變換錫箔包裝對產品是有利的。

計畫目的：確定三個月的架上樣品是否因包裝的改變而引起風味和(或)質地(texture)的總體差異。

測試目的：測試人們是否可以透過品嚐經 3 個月儲存的兩組產品包裝條件來區分是否有差異。

測試設計：安排 30 至 36 位受測者進行的三角差異測試。該測試將在正常白光照明下進行，以允許外觀差異得以考慮在內。安排六組樣品，以確保順序完全隨機。差異顯著值定義在 5% 的 α 風險，也就是，此測試錯誤地得出有差異的結論的機率是 5%。

審查樣品：檢查樣品初始（包裝前），以確保樣品間沒有明顯感官差異。3 個月後評估測試樣品，以確保沒有肉眼可見的感官特性導致此測試無效。

進行測試：按照表 11 將兩組 54 個樣品做三位數的隨機編碼，將樣品的包裝拆掉，將巧克力條的兩端切除，並且切成適合一口咬下的大小以方便測試。將紙包裝（P）的樣品與錫箔紙包裝（F）的樣品分開。對於每一個品評員，準備標誌著他/她的號碼，並根據圖 6.2 的工作表提供 P 或 F 的樣品。記錄每個品評員各編碼的評分（見圖 6.3）。

分析結果：30 組品評員的測試，其中 17 組正確識別有異樣本。表 T8 表示這項差異是顯著的，在 α -風險為 1%（機率 $p < 0.01$ ）的情況下。

測試報告—報告全文應包含計畫目的、試驗目的、測試設計等如前所述，可以包含工作表和記錄表的範例。任何給予受測者的信息或建議（例如，關於樣品的起源）都必須報告。統計結果（17/30 正確）必須和 α -風險（達到 5% 的目標）一併提到。最後的報告結論是：「紙包裝和錫箔包裝的巧克力條之間存在顯著差異，錫箔包裝確實產生感受差異，其中大約有十個評論稱錫箔包裝的樣品質地較軟。」

日期 XX 年 X 月 X 日

工作表

測試代碼 587 FF03

將此工作表貼在樣品托盤準備區域。評分表提前編碼。供應容器提前標籤。

樣品類型：巧克力條

測試類型：三角測試

樣品標識

代碼

PKG 4736（紙包裝）

P

PKG 3987（錫箔包裝）

F

編碼供應容器如下：

品評員 #

呈現順序

1, 7, 13, 19, 25, 31

P-F-F

2, 8, 14, 20, 26, 32

F-P-F

3, 9, 15, 21, 27, 33

F-F-P

4, 10, 16, 22, 28, 34

F-P-P

5, 11, 17, 23, 29, 35

P-F-P

6, 12, 18, 24, 30, 36

P-P-F

1. 將品評人員的編碼放置在樣品托盤上。
2. 按照呈現順序選擇“P”或“F”的樣品，並將其置於托盤從左到右。
3. 將選擇的編碼記錄在評分表上。
4. 提供樣品。
5. 回收品評人員填寫的評分表，並紀錄呈現的順序，以及是否回答正確（c）或不正確（i）。

圖 6.2 三角測試的工作表。例 6.2：鋁箔包裝與紙包裝的巧克力條。

H. 例 6.3：使用三角測試法測試相似度。使用 α ， β ，Pd 決定品評小組大小- 混和糖漿

問題/狀況：一個專門製作混和糖漿的製造商發現他的玉米糖漿供應商提高了該成分的價格。此製造商的研究小組已經找到另外一家高

品質的玉米糖漿替代供應商，且其價格比較能夠接受。感官分析師被要求測試兩組混和糖漿是否等效，一組是使用目前供應商的產品，另一組則使用成本較低的替代供應商的玉米糖漿。

計畫目的：決定是否使用成本較低的替代供應商所提供的玉米糖漿，其所製作的混和糖漿是否與現行的混和糖漿沒有任何風味改變。

三角試驗

測試代碼：

品嚐者編號_____名稱：_____

日期：_____

樣品類型：_____

說明

從左至右測試樣品，其中兩個樣品是相同的，而一個是有異的。選擇出異樣樣品。並且透過在旁放置 X 標示其為不同的樣品。

樣品	有異樣品	備註
_____	☐	_____
_____	☐	_____
_____	☐	_____

如果您希望為您的選擇做評論，或者如果您想對產品特性進行評論，您可以在備註欄紀錄。

圖 6.3 三角測試的評分表。例 6.2：鋁箔與紙包裝的巧克力條。品評人員在三個樣品的其中之一標示 X，但可以在備註欄裡寫下多行評論。

測試目標：為了測試替代供應商的玉米糖漿所製成的混和糖漿與現有供應商的混和糖漿是否相似。

品評員的數量：由 α ， β ，和 Pd 決定 - 感官分析師和計畫總監查閱表 T7，注意到要最大地避免得到錯誤結論，譬如在眾多選擇假設中設定 β 為 0.1%（即 $\beta = 0.001$ ），而能夠檢測出樣品之間差異的真實人口比例至少為 20%（即 p 值 = 0.20），然後保持適當的 α -風險 0.10，他們發現需要至少有 260 位品評員評估。於是他們決定妥協在 $\alpha = 0.20$ ， $\beta = 0.01$ ，和 Pd = 30% 的假設，也就是需要 64 位

品評員。

測試設計：感官分析師根據既定的測試協議表進行 66 組回應規模的混合糖漿的角測試。品評小室裡使用紅色濾光片，以掩蓋顏色差異。12 位品評人員預計連續執行五次測試，而 6 位成員則安排在第六次，也就是最後一次的測試。圖 6.4 顯示了一項典型測試的分析師工作表。

分析結果：66 個回答當中，有 21 個正確選擇了有異樣品。參照表 T8，對應於 $n = 66$ 和 $\alpha = 0.20$ 的假設，所需的正確反應最小數目是 26。因此，只有 21 個正確回答的情況下，可以得出的結論是兩個糖漿之間的任何感官差異足夠小，可以忽略不計，也就是說，兩個樣品足夠相似，可以互相替換使用。

結果解釋：該分析師通知計畫總監測試結果是 66 個回應中有 21 個正確的選擇，表示在 99% 的信賴區間中，能夠感受差異的人口比例小於 30%，且可能要低得多。因此可接受替代供應商的產品。

Pd 的信賴限制 - 假如需要的話，分析人員可以計算感受得出差異的人口比例 Pd 的信賴限制。計算方法如下，其中 c = 正確反應數目， n = 評估的總數目。

日期 98 年 11 月 5 日 工作表號 35-0032-01

將此工作表貼在樣品托盤準備區域。評分表提前編碼。供應容器提前標籤。

樣品類型：混和糖漿

測試類型：三角相似性測試(Triangle similarity test)

樣品鑑定：

使用代碼

2 個 A 的組別

2 個 B 的組別

A: 實驗室代碼 47-3651 587 246 413
 B: 實驗室代碼 026 (控制) 894 365 751

樣品容器編碼如下：

品評人員 #	編碼	代表模式*
1	587 246 894	AAB
2	413 365 751	ABB
3	751 413 365	BAB
4	246 587 894	AAB
5	751 365 413	BBA
6	587 894 246	ABA
7	413 751 365	ABB
8	246 894 587	ABA
9	894 587 246	BAA
10	365 751 413	BBA
11	894 246 587	BAA
12	365 413 751	BAB

*每個模式被重複兩次，以允許每個代碼都出現在每個位置上。

圖 6.4 相似度的三角測試測試工作表。例 6.3：混合糖漿。

$$Pc \text{ (正確比例)} = c/n$$

$$Pd \text{ (鑑別出差異度的比例)} = 1.5 Pc - 0.5$$

$$Sd \text{ (Pd 的標準差)} = 1.5 - \sqrt{Pc(1-Pc)/n}$$

信賴區間的上限 = $Pd + Z_{\beta}Sd$

信賴區間的下限 = $Pd - Z_{\alpha}Sd$

Z_{α} 和 Z_{β} 是標準正態分配的臨界值。常用的單向信賴區間 z 值包括：

信賴程度	Z
75%	0.674
80%	0.842
85%	1.036
90%	1.282
95%	1.645
99%	2.326

對於範例中的數據，鑑別差異度在信賴區間 99% 下的單向上限的計算公式為：

$$P_{\max} = Pd + Z_{\beta}Sd = [1.5(21/66) - 0.5] + (2.326)(1.5)\sqrt{(21/66)(1 - (21/66))/66} = [-0.023] + 2.326(1.5)(0.05733) \\ = 0.177 \text{ or } 18\%$$

而較低的 80% 信賴度下，單向限制落在

$$P_{\min} = Pd - Z_{\alpha}Sd = [-0.023] - 0.842(1.5)(0.05733) = -0.095 \text{ (i.e., 0.0, it cannot be negative)}$$

或者，用白話表達，感官分析師 99% 肯定可以鑑別樣品差異的人口真實比例不超過 18%，並且可能低至 0%。*

例 6.4：平衡 α ， β 和 Pd 。設定碳酸飲料成分的有效期限

問題/狀況：碳酸飲料的製造商希望在其生產的飲料瓶蓋上印上建議有效期限。已知的是，在冷藏（2°C）條件下，瓶裝樣品可保存一年以上且無任何風味的改變，而在較高溫度下，風味保質期則較短。將樣品儲存在環境溫度較高（30°C）的條件下，分別保存 6、8 和 12 個月，然後進行差異測試。

計畫目的：在瓶蓋上印上建議有效期限。

測試目的 - 確定產品在溫度較高的保存條件下多久會與冷藏的產品產生風味的差異。

品評員的數量、 α ， β 和 Pd 的選擇：生產商希望看到最長的建議有效期限，因此決定願意承擔 5% 研究結果顯示差異但事實卻不存在的機率（即 $\alpha = 0.05$ ）。而另一方面，品管經理合理地希望有把握，直到到期日前客戶無法嚐出過期的味道，所以他同意接受在 90% 的把握（即 $\beta = 0.10$ ）下，不超過 30% 人口的比例（即， $Pd = 30\%$ ）可檢測出差異。根據表 T1、 $\beta = 0.10$ 和 $Pd = 30\%$ 的假設下，發現需要 53 位品評人員。然而，只有 30 位品評成員可以配合此測試的所需時間。因此，他們三人重新協商此測試的敏感度參數，使得在可調配資源下可以獲得最大的保障。再次根據表 T1，它們決定折衷 $Pd = 30\%$ ， $\beta = 0.20$ ， $\alpha = 0.10$ 的敏感度條件下，進行感官分析測試。

測試設計：分析師準備並進行使用 30 位品評人員規模的三角測試。

分析結果：正確的回應數量如下：6 個月時 11 個；8 個月時 13 個；在 12 個月時 15 個。核對表 T8 後，分析師的結論是，在 6 個月時沒有證明差異存在。而在 8 個月之後，產品差異大。

*統一方法與相似性表：請注意，在本版第三版中使用的統一的方法並不包括在第二版中的相似性表。以本範例說明，表 T8 僅僅顯示了相似性的證據存在。為了了解此相似性的可靠狀況如何（亦即： Pd 不超過 18%，並且可能低至 0%），分析師需要計算信賴區間的界限。見第 13 章. II. C，信賴區間的推導。

表 T8 表明了若使用 $\alpha = 0.20$ 的話，差異存在的證明會較高。而在 12 個月，只要 $\alpha = 0.05$ ，便顯示差異存在。

解讀：該小組決定將有效期限設定為 8 個月，以充分保證在此月分內消費者不會嚐出過期的味道。為了額外檢查此結論，使用 80% 單向信賴區間的界線計算每個測試。結果發現，它們可以有 80% 的把握，在 8 個月內不超過 16% 的消費者可以發現，而在 6 個月則不超過 26%，但在 12 個月可能多達 37%。因此為了安全保證產品的品質 $Pd = 30\%$ 的假設，有效期限為 8 個月。*

6.4 二三點測試法(Duo-trio Test)

A. 適用範圍及應用

與三角測試法相比，二三點測試法在統計學裡算是低效的，因為通過猜測獲得正確的結果的可能性是在二分之一。但另一方面，此測試較為簡單且容易理解。與配對比較法相比，它藉由提供參考樣品而避免了造成差異混亂的優點，但缺點是必須測試三個樣品，而非兩個。

當測試目標是為了確定兩個樣品之間是否存在差異，使用這項方法。而在以下的情況，二三點測試法是特別有用的：

1. 為了確定產品差異是否因成分的變化、或是加工、包裝、儲存所造成？
2. 為了確定在沒有特定屬性可以被識別為已受影響的情況下，總體的差異是否存在？

二三點測試法通常具有超過 15 位，最好是大於 30 位的品評人員可以進行測試。兩種形式的測試存在：恆定參考模式(the constant reference mode)，通常是從常規生產中拿出相同的樣品當作始終的參考；與平衡參考模式(the balanced reference mode)，其中兩個樣品隨機作為基準地被比較。當品評人員熟知一個產品，或是他們已經妥善的受過相關訓練時，使用恆定參考模式。若是品評人員對於產品不熟悉，或者並未經過訓練，則使用平衡參考模式，

如果產品有明顯的餘味，則比起配對比較法，二三點測試法較不適合進行此實驗。（見第 7 章 II）。

B. 測試原理

提供每個受測者已標示的參考樣品，後面跟著兩個編碼的樣品，其中一個與參考樣品一樣。詢問對象哪個編碼樣品跟參考樣品是一樣的。計算正確回答的數量，並根據表 T10 做闡釋。

C. 品評人員

甄選、培訓，並根據第三節 C 指導品評人員。一般來說，最少需要 16 位，而小於 28 位受測者的規模下， β 誤差是高的。如果是 32 位、40 位，或更多的品評員下，則可以大幅改善誤差。

*使用 6 個月的結果計算信賴界線的例子如下：

$$Pd = (1.5(11/30) - 0.5) + 0.84(1.5) \sqrt{(11/30)(1 - (11/30))}/30 = 0.16$$

二三點測試法

測試編號

品評員編號 _____ 名稱： _____ 日期： _____

樣品的類型： _____

說明：由左到右測試樣品。最左手邊的樣品是參考樣品。測試兩個樣品中哪一個與參考樣品一樣，並在其旁放置 X 標示。

如果在兩個未知樣本之間分辨不出明顯差異，您必須猜測。

參考代碼 _____ 代碼 _____

評論： _____

圖 6.5 二三點測試法的評分表。

D. 測試程序

對於測試控制和產品控制，請參見前一節。如果可能的話同時呈現所有樣品，否則就按照順序。準備所有可能組合的數量（見範例），並隨機分配給各個品評員。圖 6.5 是評分表的例子（在恆定參考模式和或參考模式裡格式都相同）。在評分表上可以多留一點空間給品評員做評論，但不要提出任何額外的問題（例如差異的程度或品評員的偏好），因為品評員對其回答可能會導致其反應的偏差。計算正確回答的數量和所有回答的總數，並參考表 T10。不能將“無差異”的回答計算進去，若是品評員有懷疑仍必須猜測出一個答案。以下三個例子皆採用統一方法：

E. 範例 6.5：平衡參考模式 - 香水面紙盒

問題/狀況：產品開發部的香味化學家希望知道面紙盒中香味遞送的兩種方法，香味直接灑在面紙上，或是香味灑在面紙盒的內部，是否會產生在香味感受上品質或數量的差異。

計畫目的：為了確定香味傳遞的方法是否會因為在存放一段時間後，因為噴灑方式的不同而產生差異。

測試目的：為了確定兩項樣品在儲存三個月後是否會再感知上出現差異。

測試設計：當此刺激是複雜的，二三點測試比起三角測試法或是屬性差異測試，需要更少的重複嗅聞。這減少了相互比較三個樣品所引起的氣味適應或難度的潛在混亂。這項測試由 40 名在氣味評估領域有相關經驗的受測者進行。由香味化學家準備樣品，他使用在同一天灑在同一面紙和同樣面紙盒內部的同樣香味，然後在相同的條件下儲存 3 個月。測試樣品取自面紙盒中心的 50% 面紙，且每張面紙樣品都在評估之前被放置在密封的玻璃罐內 1 小時。

二三點測試法

測試編號：230S

品評員編號 21 名稱：_____日期：_____

樣品類型：在玻璃罐內的面紙

說明

1. 請從左邊開始依序聞各樣品。將蓋子打開並且簡單地、短淺地嗅一下。
2. 最左手邊的樣品是參考樣品。請確定兩個編碼樣本哪一個與參考樣品一樣。
3. 放置 X 在相應的樣本旁以標明其為相同樣品。如果兩個樣品間沒有明顯的感覺差別，您必須猜測。

參考 代碼_____代碼_____

評論：_____

圖 6.6 二三點測試法的評分表。範例 6.5：平衡參考模式。

這個測試設計防止了某些香味遷移到面紙盒的頂部空間，以及利用上述的密閉容器減少在品評小室空間累積香味。兩個樣品的每一個被當作評估一半的參考（20）。圖 6.6 是所用的評分表。

分析結果：40 名受測者中只有 21 位選擇了正確的樣品和參考樣品相同。根據表 T10 中， α 風險 5% 下需要 26 個正確反應。此外，為了確認樣品的呈現位置是否會影響正確選擇而做了審查，結果表明，正確反應的分佈在兩個位置間是均衡的（10 個和 11 個）。這表示

對這兩種香味的品質和數量的分析結果上，沒有或很少的附加偏差。

結果解釋：感官分析師告知香味化學家的兩種香味灑佈的方式在實驗設置的儲存時間下，經由二三點測試法未能在面紙上檢測出任何相位差異。

測試的敏感度：對於未來可能計劃這種類型的研究，須注意選擇 40 名品評人員規模的二三點測試法，其產生的測試靈敏度參數為：

辨識差異的人口比例	辨識機率	
	$(Pd) (1-\beta) @ \alpha=0.05$	$(1-\beta) @ \alpha=0.10$
10%	0.13	0.21
15%	0.21	0.32
20%	0.32	0.44
25%	0.44	0.57
30%	0.57	0.69
35%	0.70	0.80
40%	0.81	0.88
45%	0.89	0.94
50%	0.95	0.97

例如，擁有 40 位品評者和設定 $\alpha = 0.05$ 的測試所產生的結果為，在 44% 的機率 ($1-\beta = 0.44$) 下有 25% 的人口可以感受出差異 ($Pd = 25\%$)。擁有越多品評者的測試，會增加在任何既定 Pd 值(given Pd)的機率。若在 α 值較大的條件下進行測試，也會增加在任何既定 Pd 值的機率。

F. 範例 6.6：恆定參考模式 - 新的酒瓶生產線

問題/狀況：釀酒商有兩個酒瓶供應商，“A”是已經正常供應多年的供應商，“B”則是提供了在酒瓶保存期上微弱優勢的新供應商。他們想要知道兩種酒瓶之間是否可以檢測出任何差異，且為在引入不必要的變化的風險及為啤酒帶來更長的保存期之間找到平衡是很重要的。

計畫目的：要確定在貨架上存放一定時間的啤酒是否因更改包裝而產生任何明顯的差異，因為這通常是消費者買回啤酒的時機。

測試目的：為了確定兩種包裝的啤酒在室溫下經歷 8 個星期的貨架貯存後，是否可以偵測出任何感覺差異。

品評員的數量：釀酒商從過去的經驗知道，在不超過 $P_d = 30\%$ 的情況下他的品評員可檢測出差異，則在市場上就沒有什麼有意義的風險。他更在意的是引入可延長貨架壽命的“B”酒瓶是否會為他帶來任何不想要的改變。因此，他決定設置 β -風險為 0.05 和 α 風險為 0.10。參照表 T9 中的部分為 $P_d = 30\%$ ， $\beta = 0.05$ 和 $\alpha = 0.10$ 時，他發現此測試需要 96 位品評員。

測試設計：在二三點測試法中的恆定參考模式是合適的，因為該公司裝在酒瓶“A”的啤酒對於測試員來說口味非常熟悉。在啤酒商的三大實驗基地各進行一項單獨測試，且每個測試安排了 32 位品評員，以“A”作為參考樣品；64 杯啤酒“A”和 32 杯啤酒“B”，共設置 16 組 AAB 組合和 16 組 ABA 組合，最左手邊的樣品都是作為受檢者的參考。

分析結果：18，20，和 19 位受測者正確偵測出相應參考樣品的樣品。根據表 T10，在 10% 的顯著性水準下需要 21 個正確的回答。

注意：在許多情況下是允許組合兩個或多個測試。而本例中，酒瓶樣品是來自同一批生產的，且受測者都來自同一品評小組，所以組合是允許的。在總共 $3 \times 32 = 96$ 的回答中，有 $18 + 20 + 19 = 57$ 正確回答。從表 Tb 得知，在 10% 的顯著性水準下 55 個正確回答是臨界值，在 5% 的顯著性水準下 57 個正確回答是臨界值。

結果解釋：在 5% 的顯著性水準且結合三個測試的基礎上，認定兩個酒瓶間存在差異。接下來，必須研究品評小組成員是否留下任何描述差異的評論。如果沒有找到，必須提交樣品進行描述性測試。如果該差異最終的評論是不好不壞，則可能需要進行消費者偏好測試。

G. 例 6.7：二三點相似性測試 - 改變綜合咖啡配方

問題/狀況：一個咖啡製造商發現，其綜合咖啡豆中重要組成中的一種咖啡豆品種將在未來 2 年內供不應求。其研究小組已經制定了三種新的品種以供替代，他們覺得風味與當前的品種在混合後應該是一樣的。該研究小組已要求進行感官評估分析，以測試這些新的混合物是否等同於目前的產品。

計畫目的：要確定哪些三種混合物可以最好地取代目前的混合物。

測試目的：為了測試當前混合物和各種共混物之間的相似性。

日期 99 年 3 月 4 日 實驗室編號 3 工作表測試編號 2803-30

將此工作表貼在樣品托盤準備區域。評分表提前編碼。供應容器提前標籤。

樣品類型：杯裝咖啡

測試類型：二三點相似性測試(Duo-trio similarity test) (平衡參考模式)

樣品 A = 控制組 樣品 B =混合 62-A 樣品 C=混合 223B 樣品 D =混合 211

使用編碼：

	B 對 A		C 對 A		D 對 A	
	2 組 A 的	2 組 B 的	2 組 A 的	2 組 C 的	2 組 A 的	2 組 D 的
樣品 A	317 543	986	866 581	541	121 225	965
樣品 B	314	393 737				
樣品 C			674	373 158		
樣品 D					221	499 134

容器上的編碼規則如下：

品評人員 編號	模式	編碼順序	模式	編碼順序	模式	編碼順序
37	ABA	R-314-543	AAC	R-581-674	DAD	R-965-134
38	BBA	R-737-986	ACA	R-674-866	AAD	R-225-221
39	BAB	R-986-393	CCA	R-158-541	ADA	R-221-121
40	AAB	R-317-314	CAC	R-541-373	DDA	R-499-965
41	ABA	R-314-317	AAC	R-866-674	DAD	R-965-499

42	BBA	R-393-986	ACA	R-674-581	AAD	R-121-221
43	BAB	R-986-737	CCA	R-373-541	ADA	R-221-225
44	AAB	R-543-314	CAC	R-541-158	DDA	R-134-965
45	ABA	R-314-543	AAC	R-581-674	DAD	R-965-134
46	BBA	R-737-986	ACA	R-674-866	AAD	R-225-221
47	BAB	R-986-393	CCA	R-158-541	ADA	R-221-121
48	AAB	R-317-314	CAC	R-541-373	DDA	R-499-965

圖 6.7 二三點相似性測試的工作表。例 6.7：更換綜合咖啡品種。

測試設計：初步測試顯示差異很小且並沒有相關的特定屬性。因此，使用相似性的二三點測試法是合適的。為了減少遺漏可察覺差異的風險，感官分析師提出使用 60 位專門品評員進行每組測試（比起差異檢測一般使用 36 位品評員來說算是增加了人員）。他使用了試算表中的敏感度測試分析器(test-sensitivity analyzer)*(見第 13 章 III.E.)，她確定有 60 位品評員的二三點測試法有 90%的（即 $\beta = 0.10$ ）的機率可檢測出 $P_d = 25\%$ 的差異，伴隨 α 風險大約 0.25。分析師接受相對大的 α 風險，因為她更關心的是不正確地批准了與對照組不同的混合物，且只擁有 60 位品評人員可供測試。對於每一種混合咖啡，感官分析師計劃間隔 1 週以上進行 60 位品評員的測試。由於該產品的準備至賞味時間是影響風味的一個關鍵因素，受測者必須在樣品準備後的 10 分鐘內到達。使用感官實驗室的 12 間品評小室，準備棕色調的濾光片，感官分析師安排 12 位不同的品評人員至每間測試品評小室。每次 12 位的安排使得每個樣品以及參考樣品的準備時間和順序得以平衡。圖 6.7 顯示了分析師的工作表。

		測試編號
		2803-03
二三點測試		
品評人員編號	名稱：	日期：
樣品類型： <u>現煮杯裝咖啡</u>		
說明：從左到右測試樣本，其中最左邊的為參考樣品，判斷哪個與參考樣品相		

同並在其旁標示 X。

如果沒有差異是明顯的，您必須猜測哪一個不同？

參考樣品 編號_____ 編號_____

評論： _____

圖 6.8 為二三點測試的相似性測試。例 6.7：更換綜合咖啡豆品種。

樣品都沒有放奶精和糖。咖啡壺保持在 175°F，且倒入加熱 (130°F) 過的陶瓷杯，陶瓷杯已經按照工作表編碼並且按順序放置好。預先準備評分表（見圖 6.8）以節省時間，並且在受測者已經坐在品評小室上後再將樣品倒至杯中。

分析結果— 3 個測試的正確反應數是

(12 位受測者的)小室號碼	混合 B	混合 C	混合 D
1	3	6	8
2	4	5	8
3	5	7	5
4	7	7	7
5	5	5	7
總數	24	30	35

從試算表中的敏感度測試分析器得知，在選擇用於測試的 α -風險 (0.25) 的情況下若要證明顯著差異存在，需要 33 個正確反應，因此得出結論，60 位品評人員測試規模下，等於或少於 32 個正確反應是有足夠相似性的證據。*

*使用敏感度測試分析器時，不要選擇 α 、 β 和 Pd 的值形成需要的“正確反應數量”小於直接預期的機率（即三角試驗的 $n/3$ ，

二三點測試的 $n / 2$ 等)。事實上，正確反應數低於直接預期的機率，可能是由於外來因素影響而造成選擇的偏差。

從敏感度測試分析器的輸出

輸出				輸入			
總測試數	正確回答數	猜測正確機率	區分差異比例	在 Pd 假設下的正確回答機率	第一類錯誤	第二類錯誤	功率
n	x	Po	Pd	P_{max}	α -risk	β -risk	$1 - \beta$
60	33	0.50	0.25	0.625	0.2595	0.0923	0.9077

解讀：

33	或更多的正確反應數量證明在顯著性 $\alpha = 0.26$ 的水準下存在差異。
32	或更少的正確反應數顯示在 91% 的把握且不超過 25% 的品評人員可檢測出差異，意即，在 $Pd = 25\%$ 及 $\beta = 0.09$ 的水準下樣品間具有相似性。

因此可以得出結論，經過測試後，混合物 B 和 C 與控制組有足夠的相似性，但是得到 35 個正確答案的混合物 D，則不具有足夠相似性。混合物 D（基於二三點測試法）在 90% 的信賴區間單方向上限是

$$\begin{aligned}
 P_{\max(90\%)} &= [2(x/n) - 1] + Z_{\beta} \sqrt{[4(x/n)(1 - (x/n))]}n \\
 &= [2(35/60) - 1] + 1.282 \sqrt{[4(35/60)(1 - (35/60))]} / 60 \\
 &= [0.1667] + 1.282(0.1273) \\
 &= 0.33, \text{ or } 33\%
 \end{aligned}$$

感官分析師在 90% 的信心下做出以下結論：能夠測試出混合物 D 和參考樣品有差異的真實人口比例高於 33%。因此，比起預定的臨界極限 (Pd) 25% 還要多出 8%。

另外感官分析師擔心，混合物 B 的測試結果中，60 位受測者中只有 24 個正確的選擇，而在二三點測試中，60 個反應的直接預測正確數 (PD = 0) 為 $n / 2 = 30$ 。少於直接預測數量的正確反應可能代表了在混合物 B 中的測試，有外來因素嚴重地影響品評人員的選擇而導致偏差，例如樣品貼錯標籤、準備不足，或樣品處理不當等。感官分析測試的假設正確反應的真實機率至少為 50% ($H_0: p \geq 0.5$)。小於 50% (即 $P < 0.5$) 的結果，在單向信賴程度 (one-tailed confidence level) 設定為 95% 下限，使用正常二項式近似法 (normal approximation) (即， $\alpha = 0.05$ ，下尾曲線) 來檢驗。試驗統計學為：

$$\begin{aligned} z &= [(x/n) - p_0] / \sqrt{p_0(1 - p_0)/n} \\ &= [(24/60 - 0.50)] / \sqrt{0.50(1 - 0.50)/60} \\ &= [-0.10] / (0.06455) \\ &= -1.55 \end{aligned}$$

利用表 T3 (查閱 $\Pr[z < -1.55] = \Pr[z > 1.55]$)，感官分析師發現檢驗統計不大於 -1.55 的機率是 $(0.5 - 0.4394) = 0.0606$ 。此機率大於 $\alpha = 0.05$ ，因此分析師的結論是，沒有足夠的證據來拒絕在 95% 的水準上的假設錯誤。24 個正確回應不足以證明有任何積極的外來因素影響結果。

6.5 五取二測試法 (Two-out-of-Five Test)

A. 適用範圍及應用

這項方法在統計學上是非常有效率的，因為比起三角測試法的預期機率三分之一，從五個樣品中直接猜測出兩個正確答案的機率是十分之一。出於和之前同樣的原因，感官疲勞和記憶效應嚴重地影響此測試結果，所以它主要用在視覺、聽覺，和觸覺的應用上，而非風味測試 (flavor testing)。

當測試目的是為了確定兩個樣品之間是否存在差異，特別是當只有少數品評者（例如 10 位）可用時，可使用此測試方法。

與三角測試法相同，五取二測試法在某些情況下有效：

1. 要確定產品是否從成分、加工、包裝，或儲存的條件變化而造成差異，
2. 在沒有特定屬性可以被識別出的情況下，想確定總體差異是否存在，
- 3 在感官疲勞效果小的測試情況下，選擇和監督有區別各特定差異能力的品評員。

B. 測試原理

呈現給品評人員 5 個編碼樣品，並告知受測者其中兩個樣品屬於相同類別，而另外三個則屬另一種類別。請受測者從左至右測試（觸摸、審視、檢查）每個產品並選擇出與其它三個不同的兩種樣品。計算正確回答的數量，並參考表 T14 解釋。

C. 測試受測者

按照前述章節所述甄選、培訓人員。一般安排 10 至 20 名受測者，僅僅在差異很大且容易發現時才可以只使用 5 或 6 位人員。另外，只能使用經過培訓的人員。

D. 測試程序

對於測試控制和產品控制，請參見前面章節。如果可能的話同時提供樣品；不然，若樣品體積較大，或在外觀上略有不同，可以在不會破壞實驗的情況下按順序提供。如果受測者的數量不是 20 位，則從以下組合隨意選擇，選取同等數量的 3 A 和 3 B 的組合：

AAABB	ABABA	BBBAA	BABAB
AABAB	BAABA	BBABA	ABBAB
ABAAB	ABBAA	BABBA	BAABB
BAAAB	BABAA	AEBBA	ABABB
AABBA	BBAAA	BBAAB	AABBB

五取二測試(Two-Out-Of-Five Test)

姓名：_____日期：_____

樣品類型：_____

說明

1. 由左到右檢查樣品。其中兩個是同一種類型，其他三個為另一種類型。

2. 放置 X 在對應的框中以指示出相同的兩個樣品。

測試 1	測試 2	測試 3
左_____	_____	_____
_____	_____	_____
_____	_____	_____
_____	_____	_____
右_____	_____	_____
評論		
左_____	_____	_____
_____	_____	_____
_____	_____	_____
_____	_____	_____
右_____	_____	_____

圖 6.9 五取二測試法的評分表。

五取二測試法的評分表實例提供在圖 6.9。計算正確回答數和總回答數，並參考表 T14。不要計算“無差異”的回答，若品評人員不確定，仍必須猜測答案。

E. 實例 6.8：比較海苔的脆度

問題/狀況：一個海苔製造商使用某種油脂替換現有的油脂。

計畫目的：要確定油脂是否需要修改。

測試目的：為了獲得兩種油脂之間脆度的相對差異量度。

測試設計：由於感官疲勞在本測試中不是一個很大的影響因素，因此選擇五取二測試法進行最有效的評估差異。12 位品評員將能夠檢測出非常小的差異。從先前 20 種組合的表中隨機選擇 12 種組合。詢問品評人員：「哪兩個樣品脆度相同?不同於其他三個？」

進行測試：從隨機三位數列表分配編碼（見表 TI）。使用如圖 6.10 的評分表。

五取二測試

測試編碼

姓名：

日期：

樣品類型：_____

差異類型：_____

說明

1. 檢視以下列出的樣品。其中兩個的脆度是同一種類型，另外三個則屬於另一種類型。

2. 放置 X 在對應的框中以指示出相同的兩個樣品。

編碼	X	評論
_____	☐	_____
_____	☐	_____
_____	☐	_____
_____	☐	_____



圖 6.10 五取二測試表。例 6.8：比較海苔的脆度。

分析結果：在 12 名品評員中，有 9 人能夠正確的分辨出布料樣品。參考表 T14 顯示，在脆度差異檢測有 $\alpha = 0.001$ 顯著水準。

結果解釋：海苔製造商被告知，兩種海苔之間的脆度差異是容易被檢測到的。

F. 例 6.9：美乃滋替代品的視覺差異

問題/狀況：美乃滋中有一個美乃滋替代品顯著地節省了生產成本。但卻減少了產品的表面光澤度。

計畫目的：行銷組希望在測試消費者對外觀的接受度之前，先確定是否兩個成分之間存在視覺上可檢測出的差異。

測試目的：要確定是否兩種美乃滋之間在外觀上有統計學顯著的差異。

測試設計/樣本呈現：使用 10 名受檢者（已經篩選過色盲和視力受損者）。在白熾燈源下，放置 2 毫升的產品在白色背景的手錶玻璃上。

預先測試樣品，以確保表面不會在最長的測試時間中，例如 30 分鐘內，有外觀上的改變（乾硬、出水、軟化等）

日期 99 年 3 月 5 日

工作表

測試代碼 TO-AF88

將此工作表貼在樣品托盤準備區域。評分表提前編碼。供應容器提前編碼。

樣品類型：美乃滋視覺檢視

測試的類型：五取二測試

樣品標示

代碼

PX-231 6（對照品）

A

PX-2602（新美乃滋）

B

在每位品評員前呈現樣品如下：

評審編號

樣品順序

1

AABBB

2	ABBAB
3	BAABB
4	BABBA
5	BBABA
6	BBAAA
7	BAABA
8	ABBAA
9	ABAAB
10	AABAB

圖 6.11 五取二測試的工作表。例 69：美乃滋中的美乃滋外觀檢視。

進行試驗：根據工作表上所示，從左至右在一直線上平面排列樣品（見圖 6.12）；使用一個類似圖 6.11 的評分表。要求受測者「辨別出兩個相同外觀的樣品，不同於第三個樣品」

分析結果：五位品評人員正確將樣品分類。根據表 T14，這對應 1% 的顯著差異。

6.6 差異性對比法 (Same/Different Test, Simple Different Test)

A. 適用範圍及應用

當測試目標是為了確定兩種產品之間是否存在差異時，可以使用此方法，特別是當樣品不適合多次展示時，也就是說，不能使用三角測試或者是二三點測試法。這種情況的例子是，樣品存在強烈的氣味且揮之不去、需要被施用於皮膚上的半臉測試，或樣品存在複雜的刺激且容易干擾品評員。

如同其它總體差異測試中，差異性對比法在以下情況下有效：

1. 要確定產品是否因成分、加工、包裝、或儲存條件改變而造成變化
2. 在沒有特定屬性可以被識別出的情況下，想確定總體差異是否存在。

這個測試需要花費比較多時間，因為它是透過比較來自不同對（A / B 和 B / A）與相同配對（A / A 和 B / B）中的結果來確認產品差異是否存在。呈現相同配對只是感官分析師為了要確認“安慰劑效應”的影響效果。

B. 測試原理

提供給每個品評員兩個樣品，並詢問樣品是相同的或不同的。其中一半品評員呈現兩個不同的樣品；另一半則呈現一對相同的樣品（相同的樣品呈現兩次）。分析比較相同樣品組得到「不同」的回覆數，以及不同樣品組得到的「不同」的回覆數，採用 χ^2 檢驗分析結果。

C. 測試人員

一般情況下，四個樣品組合（A / A，B / B，A / B，B / A）約需呈現 20 至 50 組以確定差異是否存在？需要多達 200 個不同的品評人員，或者安排 100 位受測者並且進行兩次單獨測試。然而若因為刺激的複雜性而選定執行差異性對比法，那麼品評者只能夠進行一次單獨測試而不能重複。可使用已訓練或未經訓練的受試員，但是同次測試不應該同時包含兩種品評員（專業與非專業）。

D. 測試程序

對於測試控制和產品控制，請參見前面章節。如果可能的話同時提供樣品，不然則妥善地按順序提供。準備四組數量相等的配對樣品，如果每位受測者只評價一對樣品，則隨機呈現樣品給受測者。如果此測試中每位受測者需要評估多於一對（一組相同和一個不同，或是所有四種組合），則每位受測者的測試記錄必須保存。例 6.10 是此測試的典型工作表和評分表。

E. 分析和結果解釋：見例 6.10。

F. 例 6.10：更換烤肉醬的加工爐具

問題/狀況：一個調味品製造商為了使其工廠現代化，必須更換用來處理烤肉醬的舊爐具。工廠經理想知道，在新的爐具生產的產品口味是否和舊爐具所生產的一樣。

計畫目的：要確定新的爐具可以投入工廠使用以代替舊爐具。

測試目的：為了確定在不同的爐具生產的兩種烤肉醬產品，是否可從味道來區分。

測試設計：產品有刺激性且有餘味，會造成測試時交叉污染。因此，使用差異性對比法，搭配溫和的載體，如白麵包，是一個合適的測試選擇。60 組樣品，其中 30 組相同，30 組不同，安排 60 位受測者測試。每位受測者只評估相同（A / A 或 B / B）或不同的（A /

B 或 B / A) 單套樣品。測試的工作表和評分表顯示在圖 6.12 和 6.13 中。該測試安排在紅燈下的品評小室進行，以掩蓋任何顏色差異。

樣品呈現：由五個具有經驗的品評員進行初步測試，以確定該樣品是單獨呈現或是搭配載體(如白麵包)，比較容易測試出差異。載體在不引入任何外來因素下，用來更佳地比較樣品差異。初步測試也是以產品合適的量（以重量或體積）相對於麵包量（以大小）來決定樣品搭配，對測試很有幫助。

進行測試：就在每位品評員測試前，將儲存在寒冷且密閉容器中的預定量烤肉醬填抹到預切好的麵包片上。依照每個品評人員的工作表將樣品在放置已編碼的樣品盤上。

結果分析：在下面的表格中，每列指示哪個是被測試的樣品，而對應的該行則表示他們是如何被品評人員測試：

品評人員接收

相同組

相異組

AA 或 BB

AB 或 BA

總數量

品評結果：

相同 17 9 26

不同 13 21 34

總計 30 30 60

χ^2 檢驗（見章節 13.11I.D.6）用來比較安慰劑效應（17/13）與治療效果（9/21）。 χ^2 統計量的計算公式為：

$$\chi^2 = \sum \frac{(O - E)^2}{E}$$

E

其中 O 是所觀察到的數，E 是在四個組合中的相同/相同，相同/相異，相異/相同，和相異/相異的預期數。例如：對於相同/相同組別：

$$E = (26 \times 30)/60 = 13, \text{ 也就是說}$$

$$\chi^2 = \frac{(17-13)^2}{13} + \frac{(9-13)^2}{13} + \frac{(13-17)^2}{17} + \frac{(21-17)^2}{17} = 4.34$$

日期 1999 年 2 月 26 日 工作表

測試代碼 84-46F09

將此工作表貼在製備樣品的托盤區域。評分表提前編碼。供應容器提前標籤。

樣本類型：白麵包片上的烤肉醬

測試類型：差異性對比法(Same/Different test，異同測試<difference test>)

樣品標示

編碼

5-117-36 (舊爐具)

36

5-117-39 (新爐具)

39

以 3 位數隨機亂碼將盛裝容器進行編碼，並將之分成兩批，一批盛裝樣品 36，其他盛裝樣品 39。

準備品評托盤，將樣品按下列順序從左至右排列：

品評員編碼

樣品順序

1-15

36-36

16-30

36-39

31-45

39-36

46-60

39-39

圖 6.12 差異性對比法的工作表。例 6.10：更換加工烤肉醬的爐具。

差異性對比法

測試編號

84-4639

品評員號碼：_____名稱：_____日期：_____

樣品類型：白麵包片上的烤肉醬

說明

1. 從左至右品嚐兩個樣品。
2. 決定樣品是相同或不同。
3. 將您的回應標示在下方空格。

請注意，某些組別有兩個相同的樣品。

_____產品是一樣的

_____產品是不同的

評論：_____

圖 6.13 差異性對比法的評分表。例 6.10：更換製造烤肉醬的爐具。

結果大於表 T5 的值 ($DF = 1$ ，概率 = 0.05， $\chi^2 = 3.84$)，即表示存在著顯著差異。

結果解釋：結果表明，在兩種不同的爐具中製造的烤肉醬存在著顯著差異。感官分析師因此告訴廠長：設備供應商的說法是不正確的，兩種產品之間的差異可被檢測出。分析師表示，如果新爐具的投資是廠內重要的成本或效率項目，則新舊爐具所生產出來的兩種烤肉醬，應在消費者之間進行偏好測試。如果經過測試發現消費者偏愛新爐具所生產出來的醬品，則工廠仍可繼續新爐具的投資過程。

注意：如果例 6.10 中安排 30 位品評員進行測試，而不是 60 位，且此 30 位分別在各節測試中接受一組相同和一組不同的樣品，其結果可能與上述相同，但 χ^2 有可能是不恰當的，需使用麥式考驗 (McNemar test) 進行檢驗 (Conover, 1980)。為了執行麥式考驗，分析師必須追蹤每位品評員的回應，並且使用以下格式計算它們：

品評員收到 A / B 或 B / A，並回答：

相同

不同

品評員收到 A/A 或 B/B	相同	$a = 2$	$b = 15$
並回答：	不同	$c = 7$	$d = 6$

統計檢驗為

$$\text{麥式考驗的 } T(\text{McNemar's } T) = (b - c)^2 / (b + c)$$

當 $(b + c) \geq 20$ ，如果 T 大於表 T5 中一個自由度的 χ^2 的臨界值，則無差異的假設不成立。當 $(b + c) < 20$ ，使用二項式方程式檢驗（見 Conover, loc. cit.）。本例中：

$$\text{麥式考驗的 } T = (15 - 7)^2 / (15 + 7) = 2.91$$

小於 $\chi^2_{1, 0.05} = 3.84$ 。因此，我們不能得出樣品有差異的結論。

如果我們把從 30 位品評員的配對數據，當作它們是從 60 位個別的品評員所得到的個別數據，我們也得到前一節“結果解釋”所述的數據。標準 χ^2 分析(standard χ^2 -analysis)將會導致不正確地結論：樣品間在統計顯著上存在著差異。

6.7 A 非 A 檢驗法(“A” — “Not A” Test)

A. 適用範圍及應用

使用此方法（ISO 1985）的時機是，當測試目標是為了確定兩種產品之間是否存在差異時可以使用此方法，特別是當樣品不適合多重展示時，也就是說，不能使用三角測試或者是二三點測試法。這種情況的例子是，樣品存在強烈的氣味且揮之不去、需要被施用於皮膚上的半臉測試、在外觀上略有不同的產品和樣品、或是樣品存在複雜的刺激且容易干擾品評員。A 非 A 檢驗法(“A” — “Not

A” test)優先於差異性對比法的使用時機是，當兩種產品中的一個具有重要意義的標準，或參考樣品是品評員熟悉的產品，或者現行樣品相對所有其他受測樣品是必不可少的項目等。

如同其他測試總體差異的方法，A 非 A 檢驗法在以下情況下有效：

1. 為了判斷產品的差異是否因為成分調整、加工、包裝、或儲存條件等而產生。
2. 為了判斷總體差異是否存在，是否沒有特定屬性可以被識別出已受影響。

此方法也可以用來篩選品評人員的能力，例如測試受試者（或一組受試者）是否可以從其他甜味劑中辨別出特定的甜味劑。它也可以使用訊號偵測法(Signal detection measure)來界定感官閾值(Macmillan and Creelman, 1991)。

B. 測試原理

使品評人員熟悉 A 非 A 檢驗法。提供每個小組成員多組樣品，其中一些產品為“A”，而另一些產品“非 A”；受試者判斷是否此樣品是“A”或“非 A”，並且透過比較受試者的正確答案及 χ^2 檢驗法來確定受試者的辨別能力。

C. 測試人員

訓練 10 至 50 位品評人員熟悉 A 非 A 檢驗法。每次品評測試使用約 20~50 位品評人員。每個品評人員可能會收到一個樣品（“A”或“非 A”）、兩個樣品（一個“A”和一個“非 A”），或者每位品評員可以在同一系列中測試多達十組樣品。允許的測試樣品數量取決於此樣品的物理特性或其造成品評人員的感官疲勞程度。

注意：在此方法中，不建議使用品評人員熟悉的產品做為“非 A”的樣品，這是因為受試者在本測試中缺乏參考樣品，容易因胡亂猜測而導致結果偏差。

D. 測試程序

為了測試控制和產品控制，請參見先前章節。每次提供一組樣品與評分表。以亂碼編碼所有樣品並以亂碼呈現給受試者，使品評員不至於以任何系列測得“A”或“非 A”樣品的模式。直到受試者完成測試前，不要透露樣品的內容物資訊。

注意：在此測試的標準版本過程中，須遵守以下規定：

1. 直到測試開始時才提供樣品“A”和“非 A”給受試者。

2. 每組樣品中只有一個“非 A”樣品。

3. 每組測試提供相同數量的“A”和“非 A”。

這些規定可因為任何特定的測試而調整，但是需在測試開始前告知品評員。如規則 2，當有多個“非 A”樣品存在時，需在測試前告知品評員。

E. 結果分析和解釋

要分析四個不同的樣品及反應的組合是有點複雜，我們可以透過例 6.11 進一步解說。

日期 1999 年 1 月 15 日 工作表 測試代碼 612A83

將此工作表貼在樣品托盤準備區域。評分表提前編碼。供應容器提前標籤。

樣品的類型：甜味飲料(Sweetened beverage)

測試類型：A 非 A 檢驗

樣品內容	編碼

含 0.1%甜味劑的飲料 (“A”)	A

含 5%蔗糖的飲料 (“非 A”)	B

將 200 個 6 盎司容量的杯子進行 3 位數字隨機編碼，並且分成兩批各 100 杯。

使用 100 杯盛裝樣本“A”，使用另外的 100 杯盛裝樣品“非 A”。準備樣品托盤，將樣品依如下順序從左至右排列：

品評員	樣品順序
1-5	A A B B A B A B B A
6-10	B A B A A B A A B B
11-15	A B A B B A B B A A
16-20	B B A A B A B A A B

圖 6.14 A 非 A 檢驗的工作表。例 6.11：新型甜味劑與蔗糖進行比較。

F. 例 6.11：比較新型甜味劑與蔗糖。

問題/狀況：產品開發的化學家正在研究替代當前配方 5% 蔗糖的甜味飲料，初步味覺測試已經得知新的甜味劑的 0.1% 相當於 5% 的蔗糖，但也發現，如果同時呈現一個以上的樣品，識別力會受到甜度和其他口感因素的影響。化學家想知道這兩種飲料是否存在味覺上的差異。

計畫目的：確定 0.1% 的甜味劑是否能夠代替甜味飲料中的 5% 蔗糖。

測試目的：為了直接比較兩種甜味劑，同時減少味覺疲勞和轉換的影響。

測試設計：A 非 A 檢驗法允許受試者在樣品之間進行間接比較，且可讓品評員對新甜味劑的預期味道建立一個清楚的認知。反覆提供給受試者 0.1% 的甜味劑溶液，其標示為“A”；相反地，5% 的蔗糖溶液標示為“非 A”。在一節 20 分鐘的測試中，20 名受試者各自測試 10 個樣品並且做出判斷。受試者被要求品嚐每個樣品一次，記錄回應（“A”或“非 A”），在品嚐下一個樣品之前用白開水漱口，並等待 1 分鐘。圖 6.14 顯示了測試工作表，圖 6.15 為評分表。

A 非 A 檢驗

測試代碼

測試者編號：_____ 名稱：_____ 日期：_____

樣品類型：甜味飲料

說明

1. 在測試之前的等待時間中，了解樣品“A”及樣品“非 A”的味道。
2. 從左至右品嚐樣品。在測試每個樣品後，記錄下您的反應，並用清水沖洗掉殘留的味道，並等待一分鐘再測試下一個樣品。

注意：您收到的“A”和“非 A 樣品的數量大致相等。

樣品	編碼	此樣品為	樣品	編碼	此樣品為
1	—	“A” “非 A”	6	—	“A” “非 A”

2	—	“A”	“非 A”	7	—	“A”	“非 A”
3	—	“A”	“非 A”	8	—	“A”	“非 A”
4	—	“A”	“非 A”	9	—	“A”	“非 A”
5	—	“A”	“非 A”	10	—	“A”	“非 A”

評論：

圖 6.15 A 非 A 檢驗的紀錄表。例 6.11：新型甜味劑與蔗糖比較。

結果分析：在下面的表格中，行(columns)顯示了受試者接收的樣品類型，而列(rows)則顯示受試者的回應：

	受試者接受的樣品		
	A	非 A	合計
受試者的回應：A	60	35	95
非 A	40	65	105
合計	100	100	200

如第六節所示，經 χ^2 統計量的計算公式

$$\chi^2 = \frac{(60-47.5)^2}{47.5} + \frac{(35-47.5)^2}{47.5} + \frac{(40-52.5)^2}{52.5} + \frac{(65-52.5)^2}{52.5} = 12.53$$

大於於表 T5 的值(df= 1, α -風險 = 0.05, $\chi^2 = 3.84$)，也就是說，顯著存在差異。

注意： χ^2 分析只是為了檢驗是否因為每個受訪者進行了多次評估而造成不完全恰當的結果。然而，目前沒有計算方便的替代方法是可用。且從這個檢驗得到的顯著值應該視為近似值。

結果解釋：測試結果顯示，0.1%甜味劑溶液顯著不同於5%的蔗糖溶液。感官分析師通知化學家：該特定甜味劑很可能造成飲料在風味上出現可檢測的變化。下一個有邏輯的步驟可能是進行描述性分析以求得變異類型。

有人可能會問：什麼是不顯著的差異？這將是以下這種情況：

60 50

40 50

此時 χ^2 等於 2.02，該值低於 3.84。見 ISO 1985 有更多相關的例子。

6.8 參考比較法(DIFFERENCE-FROM-CONTROL TEST)

A. 適用範圍及應用

當計劃或測試的目的是雙重時，使用這種測試，即：(1) 確定是否一個以上(含)的樣品或對照品間存在著差異，和 (2) 估計任何此差異的大小。通常一個樣品被指定為“對照品(control)”、“參考品(reference)”或“標準品(standard)”，而所有其它樣品則相對於該對照品來說是不同的。

參考比較法是用於當樣品間的差異是可檢測且不同大小的差異影響了測試的目的及往後決策的情況，如品質保證、品質管制和儲存研究等，與對照品的差異之相對規模大小對下定管理決策非常重要。另外當不能使用二三點測試法及三角測試法的情況下，如肉類、沙拉或烘焙食品等存在正常異質性(heterogeneity)的產品，使用參考比較法(Difference-from-control test)進行測試也是合適的。

參考比較法也適用於雙樣品測試(two-sample test)：當產品造成感官疲勞或轉移效果時，重複樣品測試是不恰當的。參考比較法本質上是簡單的差異測試，外加差異大小的評估。

B. 測試原理

提供每位受試者一組對照樣品加上一個或多個測試樣品。要求受試者評價各樣品和其對照品之間的差異大小，並提供其相關衡量尺度。告知受試者某些測試樣品可能與其對照品完全相同。通過比較它們與盲測所獲得的差異來計算差異。

C. 品評員測試

一般來說，提供 20 至 50 組樣品和盲測來確定受試者的鑑別程度是必要的。如果因為產品是較為複雜且容易引起感官疲勞而選擇了進行參考比較法，那麼每位受試者只能進行一輪測試。受試者可使用已訓練或未經訓練者，但不應該在同次測試中同時使用兩者。所有受試者都應熟悉測試方式、尺度的意義，以及測試同時會進行盲測的需求。

*使用盲測所得到的效果相當於獲得安慰劑效應的措施。盲測的意思是提供完全一樣的樣品，然後詢問受試者樣品之間是否存在差異。

D. 測試程序

對於測試控制和產品控制，請參見前面章節。如果可能的話，同時提供各樣品而已編碼的對照品先評估。準備一個已編碼的對照樣品給每位品評人員，再加上另外已編碼的測試樣品。如果測試的目的是讓所有品評人員最終測試所有的樣品，但是這並不能在一節測試中完成，則必須保留樣品及測試人員的記錄，以確保剩餘的樣品在隨後的測試中提供無虞。

使用的衡量尺度可以是第 5 章中討論的任何尺度。例如：

<u>口頭描述尺度(Verbal category scale)</u>	<u>數值等級尺度(Numerical category scale)</u>
沒有差異	0 =無差異
極輕度差異(Very slight difference)	1
輕/中度差異(Slight/moderate difference)	2
中度差異(Moderate difference)	3
中度/大的差異	4
大的差異(Large difference)	5
非常大的差異	6
	7
	8
	9 =非常大的差異

(當計算用口頭描述尺度的結果時，將每個一判定值，轉換為相對的數值，例如，大的差異= 5)

E. 結果分析和解釋

計算每一樣品和盲測與對照品的平均值(mean difference-from-control)，並以變異數分析（或配對測試<paired t test>，如果只有一個樣品與對照品相比較的話）評估其結果，如實例中所示。

F. 例 6.12：茯苓膏 -增加粘度

問題/狀況：一間茯苓膏公司計劃增加其茯苓膏基質的粘度。新的兩種樣品是使用儀器控制使其有較厚的質地。樣品 F 的特性是需要更多的力量使其流動，而樣品 N 是最初容易流動，但具有較高的總粘度。該產品的研究人員想知道這兩種產品與對照品存在怎麼樣的差異。

項目目標：為了確定樣品 F 或樣品 N 是否總體上最接近目前的產品。

測試目的：測量兩個樣品與原本的茯苓膏感官上總體的差異。

測試設計：放置已預先稱重好的每種產品的量在一個已經編碼的鉢中。每個樣品需含有相同的量。本次測試採用 42 位品評人員，並要求之後的 3 天進行連續測試。每天受試者測試一組樣品，其組合可以是

- 對照品與樣品 F
- 對照品與樣品 N
- 對照品與盲測

日期 1998 年 10 月 2 日 工作表 編號 13-625

將此工作表貼在樣品托盤準備區域。評分表提前編碼。供應容器提前標籤。

樣本類型：茯苓膏

測試類型：參考比較法(Difference from Control test)

樣品	說明	樣品編碼
C	對照組	C
F	實驗組 10A3	500 以下的隨機號碼
N	實驗組 2-6X (高黏度)	超過 500 隨機號碼

樣品呈現順序如下：

品評人員編號	第 1 天	第 2 天	第 3 天
1-7	C-F	C-N	C- C

8-14	C-N	C-F	C- C
15-21	C-F	C-C	C- N
22-28	C-N	C-C	C- F
29-35	C-C	C-N	C- F
36-42	C-C	C-F	C-N

小時	品評人員編號
9:00	1, 8, 15, 22, 29, 36
9:45	2, 9, 16, 23, 30, 37
10:30	3, 10, 17, 24, 31, 38
11:15	4, 11, 18, 25, 32, 39
1:00	5, 12, 19, 26, 33, 40
1:45	6, 13, 20, 27, 34, 41
2:30	7, 14, 21, 28, 35, 42

圖 6.16 參考比較法的工作表。 例 6.12：茯苓膏。

工作表如圖 6.16。所有受試者第一個接受的是標記對照組的樣品，接著測試實驗組樣品。每個品評小室需通風良好，以減少氣味堆積；光線充足，以允許受試者接收視覺線索。

進行測試：每次測試前 15 分鐘開始採樣品。使用三位數的編碼標籤兩組樣品。使用容易去除的標記筆，(見圖 6.17)。

結果分析：由表 6.1 的結果，以及適用於隨機完全區級設計(Randomized complete block design)程序的變異數分析 (ANOVA, Analysis of variance) 來分析數據。42 名品評人員是測試設計中的“區級(blocks)”。三個樣品是“處理(treatments)” (或，更適當地，是三個級別的處理)(請參見第 13 章，為變異數分析的一般性討論及塊狀設計)。

表 6.2 總結了測試的統計結果。總變異被“劃分”為三個獨立變異來源，也就是小組成員之間的差 (即，區級效應)、樣品之間的差異

(即相關處理效果)以及除此兩組原因之外的不明變異性(即實驗誤差)。

參考比較測試(DIFFERENCE-FROM-CONTROL TEST)

姓名：

日期：

樣品類型：_____測試樣品編碼_____

說明：

1. 您接到的兩個樣品，對照樣品標記為C；測試樣品標記有3位數字。
2. 用湯匙拿取對照樣品並食用。
3. 使用下面的尺度標示測試樣品相對於對照樣品的感覺差異大小。

_____ 0 =沒有差別

_____ 1 =

_____ 2 =

_____ 3 =

_____ 4 =

_____ 5 =

_____ 6 =

_____ 7 =

_____ 8 =

_____ 9 =

_____ 10 =極差

請記住，有時兩個樣品可能是同樣的。

評論：_____

圖 6.17 參考對比法的工作表 例 6.12：茯苓膏。

樣品的統計自由度(F -statistic)顯示為高度顯著 (表 T6) ; $F_{2,82} = 127.0$, $P < 0.0001$ 。 統計自由度是一比率：樣品的均方除以均方誤差。適當程度的自由度是那些與分子和分母相關聯的 F 統計量的均方(分別為 2 和 82)。用於與一對照品多重比較的 Dunnett 檢驗法，被用於樣品平均值，和顯示兩個測試樣品和盲控(blind control)顯著不同。它也可以得出結論：產品 N($P < 0.05$)比起產品 F，與對照品更具顯著差異，基於最小檢驗差異法(LSD, least significant difference)的多重比較($LSD = 0.4$)。

表 6.1 實例 6.12 的結果：

參考比較法－茯苓膏

品評員	盲測	產品 F	產品 N	品評員	盲測	產品 F	產品 N
1	1	4	5	22	3	6	7
2	4	6	6	23	3	5	6
3	1	4	6	24	4	6	6
4	4	8	7	25	0	3	3
5	2	4	3	26	2	5	1
6	1	4	5	27	2	5	5
7	3	3	6	28	2	6	4
8	0	2	4	29	3	5	6
9	6	8	9	30	1	4	7
10	7	7	9	31	4	6	7
11	0	1	2	32	1	4	5
12	1	5	6	33	3	5	5
13	4	5	7	34	1	4	4

14	1	6	5	35	4	6	5
15	4	7	6	36	2	3	6
16	2	2	5	37	3	4	6
17	2	6	7	38	0	4	4
18	4	5	7	39	4	8	7
19	0	3	4	40	0	5	6
20	5	4	5	41	1	5	5
21	2	3	3	42	3	4	4

表 6.2 例 6.12 的變異數分析表
參考比較法－茯苓膏

變異來源	自由度	平方和	均方	F	P
總計	125	545.78			
品評員	41	247.11	6.03	6.8	0.0001
樣品	2	225.78	112.89	127.00	0.0001
誤差	82	72.89	0.89		

Dunnett 多重比較檢驗法的樣品均值

樣品	盲測	產品 F
平均反應	2.4a	4.8b
樣品	盲測	產品 N

平均反應 2.4a 5.5b

注：在同一排中，沒有相同的字母的平均都有 95%的信賴度顯著不同。Dunnett $d_{0.05} = 0.46$ 。產品 N 比起產品 F 相對於控制組更加顯著的不同。 $(LSD_{0.05} = 0.4)$ 。

結果解釋：兩個樣品檢測出顯著差異，得出結論：兩組配方與對照組有足夠的差異，值得進行粘度、厚度、皮膚冷熱感、用後感覺等屬性差異測試或描述性測試。

G. 例 6.13：風味花生零食

問題/狀況：一個大型零食加工廠的品質管理經理需要監控一條製造花生點心的產品線的感官變化，並且設置生產規格。每個不同風味（蜂蜜、辛辣、燒烤等）的批次之間固有的差異已經使用三角測試法、二三點測試法或差異性對比法排除。在大多數上述的總體差異測試中，如果受試者可以檢測出同一批產品的變異，那麼這將嚴重降低了測試出不同批次間差異的機會。所以我們需要的是一個屏除同批次間變異的測試。

計畫目的：制定適用於監測生產花生零食批次之間變化的測試方法。最終設置 QA / QC 等感官指標。

測試目的：為了衡量著名產地的花生零食在批次間的細微差異。

測試設計：將最近的控制批次樣品（正常生產）從倉庫中拉出。分別標記不同生產線的樣品為對照組 A 和對照組 B。這些樣品代表批次內的變化。另外從大量生產的產品線中拉出不同批次的花生作為測試樣品並標記為測試樣品。由參考比較法的設計概念建立三組測試：

- 對照組 A 對對照組 A（盲測）
- 對照組 A 對對照組 B（批次內側是）
- 對照組 A 對測試組（批次間測試）

安排五十位品評員在 3 天的時間內參加 3 項測試。隨機供應每位受試者樣品。在所有的測試對中，首先提供對照組 A 作為參考樣品，接著提供相對比照的樣品，並且要求受試者使用 0 至 10 的規模來比較兩者之間差異。根據 Aust 等（1985）的分析方法，將盲測結果及批次內測試所得差異從批次之間的結果減去，以確定差異具有統計顯著意義。

呈現樣品：從生產、檢測、行銷和研發的單位中找熟悉產品特性的人預先篩選樣品的風味、質地及外觀，以確定每個樣本是足以代表

批次內和批次間的變化的產品。根據感官分析組的決定，只有整顆花生將被採樣和測試。

進行測試：計算 15 顆花生並將其放置於已標示的樣品杯中。第一個呈現的對照組 A 被標記為“對照組”；其他所有樣品都具有三位數的代碼：

第 1 組：對照組 A 對 對照組 A

標示：“對照組”與[三位代碼]

第 2 組：對照組 A 對 對照組 B

標示：“對照組”與[三位代碼]

第 3 組：對照組 A 對 測試樣品

標示：“對照組”與[三位代碼]

評分表顯示於圖 6.18。

參考比較法測試(DIFFERENCE-FROM-CONTROL TEST)

姓名： 日期：1998 年 8 月 7 日 測試 #11 03 6B

樣品類型：風味花生零食

說明：

1. 首先品嚐標記為“對照組”的樣品。
2. 接著品嚐標有三位數代碼的樣品。
3. 使用下列刻度評估兩個樣品之間的總體感覺差異。
4. 標記刻度，以指示總體差異的大小。

	刻度	標記
無差別	0	_____

	1	_____
	2	_____
	3	_____
	4	_____
	5	_____
	6	_____
	7	_____
	8	_____
	9	_____
極端不同	1 0	_____

注意有時兩項樣品可能是重複的。(盲測)

評論： _____

圖 6.18 參考比較表的評分表。例 6.13：風味花生零食。

結果分析：評估表（見表 6.3）上的數據是根據 Aust 等描述的方法分析。（出處同上）。此過程檢驗測試樣品是否與不同的兩個對照樣品平均值具有顯著性差異。無效假設和替代假設為：

$$H_0: \mu_T = (\mu_{CA} + \mu_{CB})/2 \text{ vs. } H_a: \mu > (\mu_{CA} + \mu_{CB})/2$$

用於檢驗這個假設的誤差項，被稱為“純誤差均方(pure error mean square)”（在表 6.4 分析 1.13），藉由總結所有品評人員對於兩個控制樣本之間的平方差，再除以小組成員數的兩倍。使用變異數分析(ANOVA)所得的結果根據表 6.4 顯示 F-檢驗 (F1, 24 = MS 測

試與誤差 / MS 純誤差= 326.54)，代表測試樣品和對照樣品之間的差異是非常顯著的。

表 6.3 結果來自實例 6.13：

參考比較法－風味花生零食

品評員	對照組 A	對照組 B	測試組
1	2	1	6
2	0	3	7
3	1	2	5
4	1	3	7
5	0	3	6
6	2	2	6
7	3	1	6
8	2	3	6
9	2	2	6
10	3	4	6
11	1	2	7
12	0	1	7
13	3	1	4
14	0	2	8
15	0	0	6
16	0	1	8
17	1	1	7
18	3	4	6

19	1	1	9
20	0	3	6
21	0	1	7
22	1	2	6
23	2	1	4
24	1	1	6

表 6.4 根據奧斯特等人（1985）的參考比較法所做的變異數分析(ANOVA)。例 6.13 的數據：風味花生零食

來源	自由度	平方和	平方平均	F	P
合計	71	456.61			
測試與參考	1	367.36	367.36	326.54	<0.0001
純誤差	24	27.00	1.13		
剩餘	46	62.25			

解釋：分析師的結論是，即使在控制樣品中仍有變異存在的情況下，測試樣品仍然顯著不同於對照組的平均值。他建議，作為下一步，批次間進行消費者偏好或接受度測試以確認是否有差異。這種確定使公司能夠進一步知道品評人員所感知的差異是否對消費者是有意義的。參考比較法搭配消費者測試的進一步研究，允許品質管理單位建立現實的規範。

6.9 連續性測試(Sequential Test)

A. 適用範圍及應用

連續性測試是一種減少評估次數而得出結論的一種方法，例如：一個受訓者加入品評小組的接受 vs 拒絕，或一批產物的運輸 vs 銷毀。不像本章中前述的測試，本測試是在固定的 α 和受試者數目 n 中，減少第二型誤差 (β)，在連續性測試中， α 和 β 的值已經預

先決定，並且 n 取決於每個感官測試的結果。此外，由於 α 和 β 是預先確定，連續性測試提供一個直接的方法來同時測試兩個樣本之間具有差異或相似性（見第 II 節）。

連續性測試是非常實用和有效的，因為他們考慮到從最初的幾個評估得出的證據而導出結論的可能性(possibility)可能是相當充足（對於固定的 α 和 β 值）。任何進一步的測試是浪費時間和金錢。事實上，連續性測試可以減少所需評估數目多達 50%。

連續性測試也可用於存在差異的測試(existence-of-difference tests)，其中有一答案正確而有一答案不正確，例如：三角測試法(the Triangle tests)、五取二測試法(Two-out-of-five tests)、二三點測試法(Duo-trio tests)。

B. 測試原理

藉由適合已選擇方法的程序，進行一系列評估(evaluations)，並在如圖 6.20 的表上輸入每個完成測試的結果，其中三個區域被識別：接受區域，拒絕區域和繼續測試區域。在圖 6.20 中，試驗次數繪製在水平 (x) 軸，而正確回應總數被繪製在垂直 (y) 軸。如果第一個測試的結果為正確，輸入 $(X, Y) = (1, 1)$ ；如果不正確，為 $(X, Y) = (1, 0)$ 。對於每個後續試驗， x 增加 1，正確答案則 y 增加 1，不正確答案輸入 0。繼續試驗直到一個點觸及或超過決策區域的範圍。指示的區域（即接受或拒絕）即為結論。

C. 結果的分析和解釋：參數測試(PARAMETERS OF THE TEST)

這裡使用的連續測試的版本是 ISO (1983) 的。該測試本身是由 Wald (1947) 和 Rao (1950) 提出的另一種替代試驗。這兩個試驗皆由 Bradley (1953) 提供決定評估的預期數量的計算方法，以及用於選擇相關參數，如實例 6.14 和 6.15 等清楚的解釋。

D. 例 6.14：兩個訓練人員加入品評小組的接受 VS. 拒絕

計畫目的：藉由一系列測試樣品的差異靈敏度的基礎上，選擇接受或拒絕他們的學員。

測試目的：為了確定每名訓練人員的正確回答長期比例， p ，是否足夠大，而合格加入品評小組。

測試設計：同時間只提供一組樣本，使用三角測試的形式。測試之間的時間保持足夠長的時間，以避免過度疲勞。由於每個三角測試完成時，結果將被輸入到圖 6.20。連續進行測試，直到訓練人員被接受或拒絕。

d，正確的測試數量（累計）

受試者 A

接受

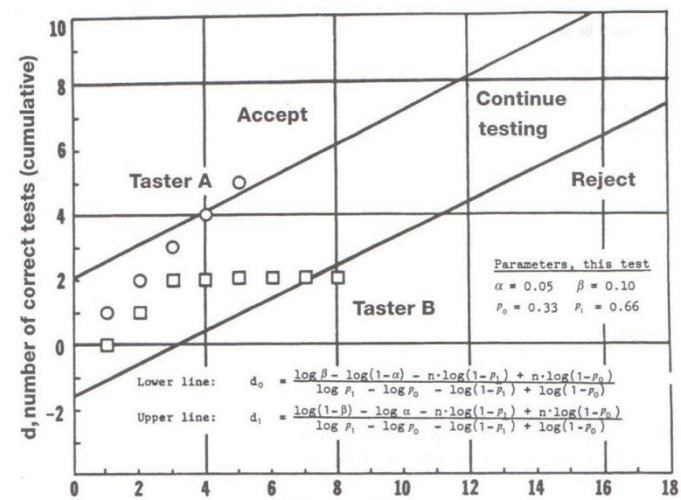
受試者 B

繼續測試

拒絕

參數，這個測試

測試



n，測試次數（累計）

圖 6.19 使用連續性三角測試甄選品評人員的例子

α 是實驗結果證明差異但事實上並不存在的機率

β 為實驗結果證明無差異但事實上存在的機率

p_0 是當樣品是相同時的預期正確決策比例

p_1 是在全部測試有一半的機會偵測出不同樣品（非猜測）的預期正確決策比例

分析結果：測試參數：四個參數的值由品評組長分配如下：

- α 是選擇不可接受訓練人員的可能性
- β 則是拒絕應當接受的訓練人員的可能性
- p_0 是最大不可接受的能力（衡量正確答案的比例）
- p_1 是最低可接受的能力（衡量正確答案的比例）

在圖 6.19 中可以看出，決定接受區域等的線條方程式取決於 α ， β ， p_0 和 p_1 。在本實例中，訓練人員 A 在所有測試皆做出正確回答，因而在連續五個三角測試後被接受。訓練人員 B 在第一個三角測試中錯誤回答，接著兩三個三角測試中正確，但後來未能在每個後續三角測試中做出正確回答，因而在第八次後被拒絕。

四個參數的各種數值可能被用。當 p_0 接近 p_1 ，所需試驗的次數增加。有幾種方法用來減少所需的試驗的平均數。首先，在我們的三角測試例子中，偵測差異的最小可接受機率可以被設定得更高，例如：從例子中的 50% 增加至 67%，這將使得 $p_1 = 0.78$ [從 $p_1 = 0.67 + (1 - 0.67)(1/3)$]。*其次，如果有許多訓練人員， α 及 β 可被分配較大的值（例如， $\alpha > 0.05$ 和/或 $\beta > 0.10$ ）。

E. 例 6.15：連續的二三點測試法 (SEQUENTIAL DUO-TRIO TESTS) - 再加熱風味牛肉餅

計畫目的：一個設於陸軍食品工程站的常規品管小組檢測出冷藏 5 天的牛肉餅再加熱後有重複加熱的味道。該計畫負責人瞭解“軍人飽食後才能前進”的道理，因此希望設置牛肉餅可冷藏天數一個實際的最大值。

測試目的：為了確定對於儲存 1 天、3 天、5 天的樣品，是否能夠被檢測出與新鮮牛肉餅對照組的差異。

測試設計：初步測試表明，在二三點測試中，放置 5 天的牛肉餅表現出強烈的再加熱味，而放置 1 天的餅則沒有，因此連續的測試設計是合適的，且需要很少試驗即可知道針對這兩種樣品的決策。

三組樣品對 (sample pairs)（對照組對 1 天；對照組對 3 天；對照組對 5 天）都進行單獨的二三點測試，其中對照組的樣品被作為為每

一個測試的參考樣品。當每個受試者完成一次測試，結果被添加到先前的回應中，並累計結果繪製（見後面）。測試連續進行，直到儲存樣品被決定是相似或不同於對照組。

分析結果：得到顯示於表 6.5 的結果。這裡的 α 是測試結果顯示有差異但是事實上沒有差異存在的機率； β 是測試結果顯示無差異但是事實上有差異的機率。

-
- 樣品保存一天
 - ▲ 樣品保存三天
 - 樣品保存五天

正確答覆數

試驗數

樣品有差異

樣品相似

繼續測試

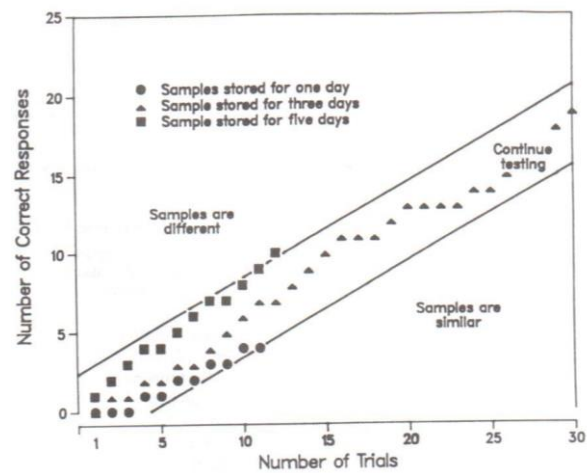


圖 6.20 例 6.15 中重新加熱牛肉餅的連續二三點測試結果繪製。

表 6.5 從例 6.15 中重新加熱牛肉餅的連續二三點測試結果

品評人員	試驗 A		試驗 B		試驗 C	
	對照組 vs. 1 天		對照組 vs. 3 天		對照組 vs. 5 天	
1	I	0	I	0	C	I
2	I	0	C	I	C	2
3	I	0	I	I	C	3
4	C	I	C	2	C	4
5	I	I	I	2	I	4

6	C	2	C	3	C	5
7	I	2	I	3	C	6
8	C	3	C	4	C	7
9	I	3	C	5	I	7
10	C	4	C	6	C	8
11	I	4	C	7	C	9
12			I	7	C	10
13			C	8		
14			C	9		
15			C	10		
16			C	11		
17			I	11		
18			I	11		
19			C	12		
20			C	13		
21			I	13		
22			I	13		
23			I	13		
24			C	14		
25			I	14		
26			C	15		
27			C	16		
28			C	17		

29	C	18
30	C	19

注意：第 1 列：I=不正確；C=正確；第二列：累計正確。

感官分析人員和計畫負責人決定設置 $\alpha = 0.10$ ， $\beta = 0.10$ ，及二三點測試的無效假設 p 值(the null hypothesis p-value)， $P_0 = 0.50$ 。此外，它們決定了可以區分新鮮和儲存樣品的群體(population)最大比例不應該超過 40%。因此， P_1 的值是

$$P_1 = (0.40)(1.0) + (0.60)(0.50) = 0.70.$$

(來源： P_1 = 可區分機率 x 可區分且給出正確回應的機率 + 無法區分機率 x 無法區分而給出正確答案的機率)

而接受、拒絕、繼續測試區域的邊界所構成的兩條線的方程式為

$$d_0 = -2.59 + 0.60n$$

$$d_1 = 2.59 + 0.60n$$

這些線隨著每三組儲存樣品的二三點測試結果累計繪製在在圖 6.21 (見表 6.5)。儲存 1 天的樣品測試結果類似於對照組。保存 5 天的樣品則顯著不同於對照組。而儲存 3 天的樣品在累計 30 次試驗後仍然無法決定類似或不同於對照組。

結果解釋：該計畫負責人得到儲存 1 天和 5 天的樣品結果，並且也被告知儲存 3 天的樣品在 30 個測試後仍然無法決定。他可以接受 3 天為規範或者選擇繼續測試，直到得到一個確定的結果。

第七章 屬性差異測試

7 Attribute Difference Tests

7.1 簡介：配對比較設計(Introduction: Paired Comparison Designs)

屬性差異測試是以衡量單一屬性(例如：甜度)來比較一樣品與其他(另一或多種)樣品的差異(第一節)。需要注意的是：在單一屬性上無差異的樣品，並不意味著沒有整體差異存在。涉及兩樣品的屬性差異測試有關的測試設計和統計處理是簡單的。

主要的困難是必須確認測試情況是單向或雙向(見實例 7.1 和 7.2)。

對兩個以上的樣品，有些設計可透過變異數進行分析，而有些則需要專門的統計處理，其複雜的程度與樣品數成正比地迅速增加，但可藉由改進測試設計而讓設計較經濟。各種多重配對測試將在下一節討論，多重樣品測試及其設計則在第四節中討論。

在第三節和第四節，我們要求受試者比較每個樣品。此類型的配對比較法在有意義的規模上，為衡量我們感興趣的屬性強度提供了一個好的方式，且其優點是可以獲得每對樣品的屬性相對強度。然而可能的配對數量與樣品的數目成正比地增加：

樣品數 t	3	4	5	6	7	8	9
可能的配對數 $N = T(T - 1) / 2$	3	6	10	15	21	28	36

在第三節和第四節中，我們提出這樣的問題：「哪種樣品更甜(更新鮮、更被喜歡)？」這種作法是基於排序數據(例如：較甜的樣品被給定為順序 2，另一個樣品則給定為順序 1)，它引入了一個人為的度數(degree of artificiality)，沒有任何一個差異度的衡量是從一個受試者直接獲得的，但其統計計算變得更為簡單。排序的數據(rating data)需要使用專門的統計方法。

7.2 定向差異測試：比較兩個樣品(Directional Difference Test: Comparing Two Samples)

A. 適用範圍及應用

當測試目標是為了確定兩個樣品之間的某一特定感官特性的差異(例如：哪個樣品較甜?)時，使用此方法。在這種模式下，該

方法也被稱為**配對比較法**(Paired comparison test)或**2 選項必選法**(2-Alternative Forced Choice, 2-AFC)。它是一種最簡單和最常用的測試法，主要用於初步確認是否需要使用其它更複雜的感官測試。其他兩個樣品配對的測試比較法是差異性對比法(Same/Different test，見第 6 章)和配對偏好測試(Paired Preference test，見第 12 章)。

在使用配對比較法時，一開始就必須區分是雙向應用(雙邊的<bilateral>，最常見)，或是單向應用(單邊的<unilateral>，當只有一個回覆是被關注的或是只有一個回覆是正確的)。(參見第 13 章)

統一方法也適用於配對比較法。測試所需的受試者的數目取決於：1. 測試是單向(使用表 T9)或雙向(使用表 T11)；2. 測試靈敏度參數 α ， β ，和 $p_{\max}$ 的選擇。在配對比較法中，參數 $p_{\max}$ 取代了第 6 章中總體差異方法所討論的參數 Pd ， $p_{\max}$ 對研究人員的意義是偏離(departure)強度相等(即 50:50 受試者認為)代表有意義的差異。例如：若研究人員認為 60:40 的受試者的差異可以判別其強度不相等，則 $p_{\max} = 0.60$ 。且研究者在適當的表(T9 或 T11)中可選擇 α 和 β 的值來決定受試者的數量。作為一個經驗法則：

- $p_{\max} < 55\%$ 表示與強度相等的小偏離；
- $55\% \leq p_{\max} \leq 65\%$ ，代表中偏離；和
- $p_{\max} > 65\%$ ，代表大偏離。

B. 原理

提供每個受試者兩個編碼樣品。準備相等數量的 AB 和 BA 組合，並且隨機分配給受試者。要求受試者從左至右品評樣品，並填寫於記錄表。明確告知受試者是否允許「沒有差異」的判斷。

只有「被迫選擇技法(forced choice technique)」是適合用於正式的統計分析。然而，在某些情況下，受試者可能會相當反對當他沒有察覺任何差異時仍需標示差異程度。感官分析人員必須再決定是否：(1) 均勻地分配其得分於兩組樣品上或(2) 忽略他們。步驟(1) 增加了找到差異的機率，而步驟(2) 則減少了找到差異的機率；因此，分析人員的選擇可能影響結果，他必須同時面對責任和抵抗誘惑。在實務中，大約有一半的分析師禁止“無差異”的判斷。另一半則認為，擁有快樂的品評小組是更好的，因此經常使用步驟(1)。

C. 測試人員

由於此測試方法簡單，甚至可以提供給經過最少訓練的人員進行測試；最基本的是測試人員需要完全熟悉要測試的屬性。或者，如果測試是特別重要的（例如：在市面上已發現一個有異味的產品），可以選擇已經展現出對此屬性特別敏銳且訓練有素的受試者。因為猜測的機率為 50%，相當大量的測試對象是必需的。表 T12 顯示：15 位品評員的測試中，如果要獲得 0.01 的顯著性水準，必須有 13 位同意。而 50 位品評員的測試，則可以用 35 位同意的判斷獲得相同意義的結果。

定向差異測試(Directional Difference Test)		
姓名：_____ 日期：_____		
樣品的類型：_____		
研究特點：_____		
說明： 從左到右品嚐每對樣品，然後寫下你的判斷 如果沒有感到明顯的差異，就算不確定也寫下你最佳的猜測。「無差異」的判斷是允許的，但只能作為最後的手段。		
測試組	其樣品較_____	
_____	_____	_____
_____	_____	_____
_____	_____	_____
評論：		

圖 7.1 定向差異測試的記錄表範例，針對不同方向的測試。配對比較法、允許「無差異」判斷。

D. 測試程序

對於測試控制和產品控制，見前面章節。情況允許的話，同時提供樣品，否則就連續地提供。準備相同數量的 A / B 和 B / A 組合，並隨機分配給各個品評員。圖 7.1 是一個典型的記錄表。請注意，不管測試是單向或雙向，記錄表都是相同的，但記錄表上必須標示是否允許「無差異」的判斷（或品評員必須知道這一點）。在單張記錄表上可以放置幾個連續的配對比較，但不要增加額外的問題，因為這些問題可能導致偏差。

計算有關的反應數量。在一個單向測驗中，計算正確反應的數目，或者有關的方向上的反應，並參考表 T10。在雙向測驗中，則計算同意的回應數量，並參考表 T12。

例 7.1：針對不同方向（雙向）-混合檸檬冰糖水

問題/狀況：針對檸檬水的消費者研究表示，消費者最感興趣的是哪種檸檬水味道最像「鮮榨檸檬汁」。

該公司已經開發出兩種口味的粉末混合系統。開發者希望知道哪一個系統更接近鮮榨檸檬汁。

計畫目的：開發接近鮮榨檸檬汁的產品。

測試目的：測量兩種系統，哪一種相對接近鮮榨檸檬汁的味道。

實驗設計：由於不同的人可能對鮮榨風味有不同的想法，一定規模的品評小組是必要的，但人員不需要經過過多的訓練。使用配對比較測試法，雇用四十名受試者，且 α -誤差設定為 5%，即 $\alpha = 0.05$ 。假設 H_0 是：新鮮度 A 等於新鮮度 B。另一種假設是 H_a ：新鮮度 A 不等於新鮮度 B。任何一個結果都是我們有興趣的，因此這測試是雙面。將樣品編碼為 691 和 812，並使用圖 7.1 中所示的記錄表收集數據。

篩選樣品：事先確認兩項樣品中的檸檬香料強度是否類似。

結果分析：二十六名品評員認為樣品 812 有更多的鮮榨檸檬香味。四位品評員認為兩個樣品“沒有差異”。從表 T12 得出以下結論：總共四十名的品評員裡有二十六名選擇樣品 812，數量足以呈現顯著差異。

結果闡釋：建議配方 812 該被用在未來的產品中，因為它在此測試中已經證實具有更接近鮮榨檸檬汁的風味。

例 7.2：方向性差異（單面）-啤酒苦味

問題／情況：一家啤酒釀造商接到來自市場的報告說，其啤酒“A”被認為不夠苦，此外使用更高的酒花添加量釀造的實驗樣品“B”正在進行中。

計畫目的：希望生產更苦的啤酒，但又不能太苦。

計畫目的：比較啤酒 A 和啤酒 B，以確定是否已經獲得了微小卻顯著增加的苦味。

實驗設計：選擇配對比較/方向性差異檢驗(paired-comparison/directional difference test)，因為感興趣的部分只有苦味是否增加。該計畫負責人偏向採用較高的確定性，即， $\alpha = 0.01$ 。感官分析師將啤酒編碼 452 和 603，並將樣品提供給有能力檢測苦味微小變化的 30 名品評員小組。品評記錄表上的問題是：「哪種樣品更苦？」（而非：「樣品 603 是否比 452 更苦？」，）以避免造成實驗目的偏差。

篩選樣品：樣品事先由六位品評員小組品嚐，**確保苦味以外的差異最小。**

結果分析：22 位品評員選擇樣品 B。虛無原假設(null hypothesis)是 H_0 ：苦味 A=苦味 B，但替代假設是 H_a ：苦味 B>苦味 A，使得測試為單方向的。分析師從表 T10 的結論是，苦味的差異被認為在 $\alpha = 0.008$ 。釀造實驗 B 是成功的。

註：在決定配對比較實驗是單面或雙面的重點是替代假設是單面或雙面，而非問受試者的問題有一個或兩個回復。單向檢驗情況的發生主要是其測試目的是確認一個明確的「改善」或「治療」效果（同時參見第 13 章）。單向或雙向檢驗情況的例子如下：

單向(One-sided)：

確認釀造試驗樣品是更苦

確認測試的產品是否首選（如我們有理由期待）

在培訓品評員方面：哪種樣品更具水果味（使用訓練樣品）

雙面(Two-sided)：

決定哪些釀造試驗樣品更苦

決定哪些產品是首選

大多數的其他測試情況...當替代假設是樣品是不同的，而不是“某樣品大於其他樣品”時。

7.3 配對排序(順位)測試：Fridman 分析(PAIRWISE RANKING TEST: FRIEDMAN ANALYSIS)－比較幾個樣品中所有可能的配對

A. 適用範圍及應用

使用這種方法測試的目的是比較幾種樣品的單一屬性、如甜味、新鮮度或偏好。此測試法特別適用於 3 至 6 種樣品，且由一組相對缺乏經驗的品評小組進行。在選定的屬性上將樣品按照強度大小重新排定，並且在這些差異上提供了數值指標及這些差異的顯著程度。

B. 測試原理

提供給每位品評員一對隨機順序的樣品，詢問如下問題：“哪種樣品更甜”（更新鮮、偏好等）。繼續此動作直到所有可能配對的樣品都被所有品評員測試過。使用 Fridman 型統計分析來評估結果。

C. 測試人員

依前面章節甄選、培訓並指導品評員。每次測試不可少於十位品評員；如果使用二十個或更多的可用品評員，則此測試的識別度將大為提高。確定該品評員能夠識別有關的屬性，例如：經過培訓可識別各種屬性中的已知強度差。根據不同測試目標，可能需要已證明有能力檢測出屬性微小差異的受試者。

D. 測試程序

對於測試控制和產品控制，見前面章節。如果可能的話，同時供應樣品，否則依序提供樣品。請確保呈現的順序是真正隨機的：不得導致品評員期待同一規律，因為這會影響判斷。同組樣品內也需隨機呈現。各組樣品只詢問品評員一個問題：「哪個樣品更加地…？」不允許「沒有差異」的判斷，如果仍然發生此情況，均勻地分配此判斷在所述樣品中。

例 7.3：玉米糖漿的口感

問題/狀況：糖漿製造商希望向市場推出低濃度的產品。已經準備四種無味的混合液(A、B、C、D)，以供評估 (Carr, 1985)。

計畫目的：為了評價 4 種糖漿混合溶液的適用性。

測試目標：建立四種混合液的感知口感濃度的主觀程度。

試驗設計：選擇 Fridman 分析(Friedman analysis)的配對排序(順位)測試(Pairwise Ranking test)：(1) 因為配對的呈現 (presentation)較少受疲勞的影響 (2) 因為這種測試可建立有意義的規模。以十二位具有能力的品評員評估六種可能組合：AB，AC，AD，BC，BD 和 CD。使用圖 7.2 的工作表，並分別記錄在圖 7.3。

結果分析：下表顯示被選擇的次數（總數 12）。若每“行”的樣品被選擇，則其被認為比每“列”的樣品更濃。例如：當樣品 B 與樣品 D 比較時， 12 名受試者中有 2 位認為更濃。

		列樣品（較稀）			
		A	B	C	D
行樣品 （較濃）	A	—	0	1	0
	B	12	—	6	2
	C	11	6	—	7
	D	12	10	5	—

日期 11-6-98		工作表 No. 78	
代碼	樣品	代碼	樣品
A	混和物 4238	C	CCSA 混和物 III
B	混和物 133.8B	D	測試樣品 11.3A
每位品評員收到六組配對樣品，每組樣品隨機編碼			
品評員編號		樣品呈現順序及其編碼	

	1st		2nd		3rd		4th		5th		6th	
1	A	D	B	D	B	C	C	D	A	C	A	B
	119	634	128	824	316	967	242	659	978	643	224	681
2	B	D	A	D	A	B	A	C	C	D	B	D
	293	781	637	945	661	153	837	131	442	839	659	718
3	A	C	B	C	C	D	A	B	A	D	B	D
	926	563	873	611	194	228	798	478	184	278	478	924
4	B	C	C	D	A	C	B	D	A	B	A	D
	455	857	764	452	975	815	523	824	556	982	737	539
5	C	D	A	B	B	D	A	D	B	C	A	C
	834	245	285	299	782	679	114	966	713	561	393	495
6	A	B	A	C	A	D	B	C	B	D	C	D
	662	196	516	777	843	581	375	313	327	415	881	242
7	A	D	B	D	B	C	C	D	A	C	A	B
	341	918	949	188	428	742	486	585	635	154	545	367
8	A	B	A	C	A	D	B	C	B	D	C	D
	787	479	491	563	259	396	659	797	899	727	112	157
9	C	D	A	B	B	D	A	D	B	C	A	C
	578	322	352	336	537	434	961	242	261	396	966	876
10	A	C	B	C	C	D	A	B	A	D	B	D
	814	952	378	381	148	297	848	383	679	165	448	781
	B	D	A	D	A	B	A	C	C	D	B	D

11	498	383	131	919	466	866	794	898	526	851	721	122
	B	C	C	D	A	C	B	D	A	B	B	D
12	675	536	495	778	622	159	263	751	953	779	296	956

圖 7.2 配對排序測試—Fridman 分析(Pairwise Ranking test—Friedman analysis)的工作表。例 7.3：玉米糖漿的口感。

Fridman 分析 (Fridman, 1937; Hollander 及 Wolfe, 1973) 的第一個步驟是計算每個樣品的排序總和。在本例中，濃度「較厚」的給定 1，「較薄」的樣品給定 2。相加所有行頻數(frequency)的總分，到兩倍列頻數的總分而得到各樣品的總排序，例如，樣品 B ($12 + 6 + 2$) + $2(0 + 6 + 10) = 52$ ：

樣品 A B C D
 排序總分 71 52 48 45

檢驗統計，Fridman 的 T，計算如下：

$$T = \frac{(4/pt) \sum_{i=1}^p R_i^2 - (9p[t - 1]^2)}{[4/(12)(4)][71^2 + 52^2 + 48^2 + 45^2] - [9(12)(3^2)]} = 34.17$$

多組配對比較實驗(MULTIPLE PAIRED COMPARISONS TEST)	
名稱：_____	日期：_____
樣品類型： <u>無味混和糖漿</u>	
差異： <u>濃度（口感）</u>	

說明： 1. 拿到樣品盤後，注意各樣品位置下面的每個樣品編碼。 2. 從左至右品嚐第一組樣品，並記錄哪些樣品是較濃厚的（更黏稠），在其編碼旁標示 X。 3. 繼續此動作直到所有 6 組樣品都評估完畢。依需要用清水沖洗口中餘味。			
樣品對	左邊樣品	右邊樣品	備註
6	_____	_____	_____
5	_____	_____	_____
4	_____	_____	_____
3	_____	_____	_____
2	_____	_____	_____
1	_____	_____	_____
如果你認為沒有什麼區別，請做一個最好的猜測。關於您的選擇的原因或樣品的特性意見可以記錄在備註欄中。			

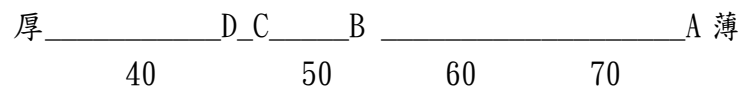
圖 7.3 為記錄表配對排序測試 Fridman 分析。例 7-2：玉米糖漿的口感。

其中， p = 基本設計的重複數目（此例中 = 12）； t = 樣品處理的數目（此例中 = 4）； R_i = 第 i 個樣品處理的排序總和； $\sum R^2$ = 所有 R 的平方總和，從 R_1 至 R_t 。

T 的臨界值已被製成表格（Skillings 和 Mike，1981）對於 $t = 3、4、5$ 及 p 的小值；至於不在表中的實驗設計，比較 T 的值與 $(t-1)$ 自由度的 χ^2 的臨界值，見表 T5。在目前的情況下，臨界值 T_s 是

顯著性水準 α	0.10	0.05	0.01
T 的臨界值	6.25	7.81	11.3

結果可以顯示在濃度厚薄比較的排序總和：



在相同的規模，HSD 值（參見第 13 章）用於比較兩個排序總和（ $\alpha = 0.05$ ），如下所示：

$$\text{HSD} = q_{\alpha, t, \infty} \sqrt{pt/4} = 3.63 \sqrt{(12)(4)(4)/4} = 12.6$$

（其中值 $q_{\alpha, t, \infty}$ 在表 7.4 中可以找到）。A 和 B 之間的差距大於 12.6，所以 A 是顯著濃度更薄，因此 A 混和液的樣品比起 B、C、D 的混和液來得理想。

7.4 簡介：多重樣品差異測試－區集設計 (MULTISAMPLE DIFFERENCE TESTS — BLOCK DESIGNS)

在 I 至 III 小節中描述的測試是根據單一選定屬性處理樣品的配對比較 (pairwise comparison)。在接下來的四個小節的測試討論大於兩個樣品組，根據一個選定的屬性（如甜度、新鮮度或優先），並使用在第 13 章介紹的區集設計 (blocking designs) 進行比較。**完全均衡設計** (Complete Block Design) — 最簡單的設計是同時排列所有的樣品（見方法 VI），但結果並不一樣精確且不適於那些更複雜的測試。其次的簡單方法是比較所有樣品，並使用排序量表 (ranking scale)。我們可以在一個完整的區集比較所有樣品（方法六，多重採樣差異檢驗 <Multisample Difference Test>，完全平衡設計 <complete balanced design>），或者我們可以透過分割成幾個較

小的區集（平衡不完全均衡設計法(Balanced Incomplete Block Design), BIB），降低小組成員對味蕾或其他感官的短期記憶負荷。
平衡不完全均衡設計（Balanced incomplete block, BIB）—在完全區集設計，每個區集（行）的大小等於正在研究樣品的處理數。一個區集在這裡指的是提供給每一位品評員的樣品組。通常品評員不能同時評估超過 4 個到 6 個的樣品。如果進行比較的樣品數量較大，例如：6 至 12 個，可以使用平衡不完全均衡設計（BIB）。除了將所有 t 樣品呈現在同一組裡，可以將他們分別呈現在不同的 b 小區集裡，各小區集含 k 個樣品，其中 $k < t$ 。各個區集裡的 k 樣品必須經過良好的安排，使得所有樣品進行評估的次數相同，也就是所有同組的樣品出現在 b 個區集裡相同的次數。Cocran 和 Cox（1957）提出了一個龐大的 BIB 設計列表，適用於大多數的實驗情況。

7.5 簡單的排序測試：Friedman 分析—隨機（完全）區集設計(FRIEDMAN ANALYSIS RANDOMIZED <COMPLETE> BLOCK DESIGN)

A. 適用範圍及應用

當測試目標是根據一個屬性(例如：甜度、新鮮度、喜好度)去比較多個樣品時，使用這種方法。排序(Ranking，過去譯為順位)是執行這種比較的簡單方式，但得到的數據僅是序數(ordinal)，並沒有方法測量各個品評員的評量差異程度。各個不同的樣品不管差異程度大小，得到的都只是一個排序單位。Pangborn(1984)針對排序測試法的優點及限制做了一個詳盡的討論。排序測試法比起其他方法耗時較少，尤其適合未來分析時要進行分類或篩選的樣品。

B. 測試原理

以平衡、隨機順序提供測試樣品給每個品評員。要求受試者根據我們所關注的屬性排序樣品。計算排序的**總數**，並使用第十三章描述的 Friedman 測試統計法評估。

C. 品評員測試

根據前述章節甄選、培訓和指導品評員。每次測試不可少於 8 位品評員。如果有 16 位以上的品評員可以大幅降低判斷的偏差。品評員可能需要視判別的樣品**屬性**再進行專業的指導和培訓。

根據測試目的，以具備辨別微小差異屬性的能力為基礎來選擇品評員。

D. 測試程序

關於測試控制和產品控制請見前面章節。如果可能的話，同時提供樣品，否則按照順序提供。品評員按照平衡隨機順序接收樣品組，其任務是重新排列其順序。各樣品組可以一次被呈現或以不同的編碼多次呈現。如果各組樣品可以兩次或更多次數以上呈現，精準度將大為改善。在偏好測試(preference tests)，指示品評員分配最佳樣品排序第一，排序二至次佳產品，依此類推等。至於強度測試(intensity tests)，指示品評員分配最低強度排序第一，次低產品排序二，等等。

建議品評員根據第一次品評的印象進行樣品排序，然後再做進一步的品評確認排列順序。指示受試者在其認為相近的樣品做一個“最佳猜測”，即使它們感覺起來是相同的。但是，如果品評員拒絕猜測，應該在「意見」欄裡註明其認為是相同的樣品。而之後進行統計分析時，我們必須分配平均序級(rank)到每個相同的樣品。例如：在一個四組樣品的測試中，如果小組成員不能區分中間兩個樣品，分配 2.5 的平均序級至兩個樣品上，即， $(2 + 3) / 2$ 。

如果同一組樣品需要依一個以上的屬性進行排序，分別對每個屬性執行以上步驟，並且使用不同的編碼，使得任一個評估不至於影響至下一次評估。使用如圖 7.4 的記錄表。提供足夠的空間給不同需求的樣品組，但請注意，每次樣品組都必須使用新的編碼，且提供給每組或每個品評員一張記錄表。

E. 分析和結果解釋

Fridman 的測試分析 (Fridman, 1937; Hollander and Wolfe, 1973)，最好使用 Kramer 的表，因為後者提供了中間序級樣品的準確評估 (Kramer 等人, 1974)。如例 7.3 計算每個樣品的排序總分 (列總和)，接著使用方程式 13.14 檢驗統計量 T 的值。如果 T 的值超過 χ^2 隨機變量在 $(T - 1)$ 自由度的臨界上限，則判定顯著差異的樣品在小組中存在。在第 13 章中我們會介紹適用於排序測試的公式 13.15 和公式 13.24，可以進一步確定哪些樣品是不同的。

例 7.3：比較四種人工甜味劑的持久性

問題/狀況：心理學實驗室希望比較 4 種人工甜味劑 A、B、C、D 的甜味持久度。

計畫/測試目的：為了確定甜味劑是否有吞嚥後的餘味在口腔中，及其中是否存在持久性的顯著差異。

實驗設計：持久的餘味感覺可能因人而異，因此建議安排多組品評員。排序實驗是合適的，因為可以簡單的進行，並不需要太多的訓練。安排 48 名學生進行 4 個樣品的測試。每個受試者接受三位數字的編碼，按照平衡、隨機的原則依序提供 4 個樣品，記錄表見圖 7.4。

*排序法是最少產生混亂的測試，前提是品評員可以按照規則分配排序給各個樣品；然而，仍然存在某些品評員，堅持配分給最高分的分數給表現最佳的產品。擺脫此困境的方法是分析師在測試後自行輸入電腦的“R 減去評分”，其中 R 是比樣品 ($R = T + I$) 的數目多一個。

排序測試(RANKING TEST)
名稱：_____ 日期：_____
樣品類型：人工甜味劑
研究特點：甜味的持久性
說明 1. 接收樣品盤，並記錄托盤上樣品下方的編碼。 2. 由左到右品嚐樣品，並注意甜味持久度，各個樣品間等待至少 30 秒，並依需要清洗口中餘味。 3. 在你覺得餘味最不持久的樣品位置填入“1”，依序填入“2”、“3”，最持久的樣品則填入“4”。 你可以先依照第一次品嚐的印象分配樣品順序，接著再通過相鄰樣品的比較確認所有順序。 4. 如果你認為兩個樣品表現相同，仍然針對他們的排序順序做一個「最佳的猜測」。

編碼	_____	_____	_____	_____
排序	☐	☐	☐	☐
備註：	_____			

圖 7.4 為簡單排序測試的記錄表。例 7.3：四種甜味劑的持久性比較。

篩選樣品：該測試需要非常仔細地準備，以確保除了預期的餘味(即那些從不同的化學組成產生的 4 種化合物之間的餘味)外，樣品之間沒有其他差異。請四位有經驗的品評員評估和調整樣品，以確保它們都具有同樣且平均的甜度，而且在溫度、粘度、外觀（顏色、濁度，和產生的泡沫等程度）不具有任何差異，或使用掩蔽等方法排除這些差異的因子。

結果分析：表 7.1 顯示了如何將編輯結果和計算排列順序。使用公式 13.14 檢驗統計量 T 的值：

$$T = ([12/(48)(4)(5)][135^2 + 103^2 + 137^2 + 105^2]) - 3(48)(5) \\ = 12.85$$

使用表 T5 發現：在自由度 3 的前提下， χ^2 的 5% 臨界值上限是 7.81。而 T 的值 12.85 大於 7.81，得出結論為樣品在 5% 的信心水準時，其甜味持久性存在顯著差異。至於如何確定哪些樣品顯著不同，可以使用多重比較等式 13.15 計算，如：

$$LSD_{rank} = 1.96\sqrt{48(4)(5)/6} \\ = 24.8$$

表 7.1

例 7.3 的結果列表：

比較四種甜味劑的持久性				
物品編號	樣品 A	樣品 B	樣品 C	樣品 D
1	3	1	4	2
2	3	2	4	1
3	3	1	2	4
4	3	1	4	2
5	1	3	2	4
——	——	——	——	——
——	——	——	——	——
——	——	——	——	——
44	4	2	3	1
45	3	1	4	2
46	3	4	1	2
47	4	1	2	3
48	4	2	3	1
排序和	135	103	137	105

任何兩個樣品的排序加總相差超過 $LSD_{rank} = 24.8$ ，即在 5% 的水平上有顯著差異。因此得出結論：樣品 B 和 D 的甜味持久度都顯著低於樣品 A 和 C 的樣品，然而樣品 B 和樣品 D 並無顯著差異，樣品 A 和樣品 C 也無顯著差異。

例 7.4：啤酒苦味差異分析(BITTERNESS IN BEER NOT AGREEING WITH ANALYSIS)

問題/狀況：一個釀酒廠的品質管理經理根據標準分析報告，得知公司的品牌 P 的啤酒花物質和競爭者相同，然而他卻得到報告指出其

產品啤酒 P 更苦。在開始調查苦味物質是否存在污染之前，他想確認苦味是否真的具有感知上的差異。

計畫/測試目的：測試啤酒 P 和競爭品牌 A、B 和 C 的苦味。

實驗設計：通過 12 個有能力偵測微小差異的品評員，將 4 個樣品進行排序檢測。虛無原假設 H_0 是：苦味 P = 苦味 A、B 或 C；而另一種假設 H_a ：苦味 P ≠ 苦味 A、B 或 C。另外事前並沒有任何資訊告知 A、B 和 C 之間存在不同物質。圖 7.4 為使用的記錄表。

結果分析：見表 7.2。需要注意的是我們可以允許有經驗的品評員分配一樣的序級。如果檢測中出現這種情況時，我們必須使用公式 13.16，來代替檢驗統計量 T' 。為了計算 T' 值，我們必須首先確定具有相同序級的樣品組數量 (i)，以及其中相同序級的樣品數量 (g_i) 和每個相同序級的樣品大小 (t_{ij})。(每個不同序級的樣品 $t_{ij} = 1$)

*沒有經驗的品評員不應當被允許分配樣品相同排序的原因有兩個：首先，分配相同排序的方式使惰性品評員不做出更多判斷，而當他們被強制規定時，這些判斷是可以存在的。第二，相同排序的樣品降低了統計測試的靈敏度。

表 7.2 為例 7.4 的結果：

啤酒的苦味差異分析

品評員編號	樣品	樣品	樣品	樣品
	A	B	C	P
1	1	2.5	2.5	4
2	2	1	4	3
3	1	3	3	3
4	2	1	3	4
5	2	3	1	4
6	2	1	4	3
7	3	1	2	4

8	1	2	3.5	3.5
9	2	3	4	1
10	2	1	3.5	3.5
11	2	3	1	4
12	2	1	4	3
排序和	22	22.5	35.5	40

只有測試中出現相同排序的區集需要被考慮，因為只有這些區集影響 T' 的計算。從表 7.2 我們可以看到，這些相同排序的區集發生在測試組 1, 3, 8, 和 10。這些區集的 g_i 值和 t_{ij} 值如下：

$$\begin{array}{llll}
 g_1 = 3, & t_{1,1} = 1 & g_3 = 2, & t_{3,1} = 1 \\
 & t_{1,2} = 2 & & t_{1,2} = 3 \\
 & t_{1,3} = 1 & & \\
 g_8 = 3, & t_{8,1} = 1 & g_{10} = 3, & t_{10,1} = 1 \\
 & t_{8,2} = 1 & & t_{10,2} = 1 \\
 & t_{8,3} = 2 & & t_{10,3} = 2
 \end{array}$$

以第 13 章的公式 13.16，這些值用來計算在 T' 的分母(denominator) 的第二名詞，如：

$$\begin{aligned}
 T' &= \frac{\sum_{j=1}^1 (X_j - G/t)^2}{[bt(t+1) - (1/(t-1)) \sum_{i=1}^b (\sum_{j=1}^1 t_{ij}^3 - t)]} \\
 &= \frac{12[(22 - 30)^2 + (22.5 - 30)^2 + (35.5 - 30)^2 + (40 - 30)^2]}{(12)(4)(5) - (1/3)(6+24+6+6)} \\
 &= 13.3
 \end{aligned}$$

$T' = 13.3$ 的值超過 χ^2 在自由度為 3 ($\chi^2_{0.05,3} = 7.81$) 時上限 5% 的臨界值，因此我們得出結論為樣品中存在顯著差異。

由於只有樣品 A、B、C 和樣品 P 是否存在差異才是我們感興趣的，因此，多重比較方法測試是適用於此排序測試(見 Hollande 及 Wolfe，

1973)。我們得出多重比較法的值是 13.1。而樣品 P 的排序總分比起樣品 A 和 B 的排序總和高出 13.1。

測試報告：品管經理得出結論為，該公司的樣品 P 比競爭對手的啤酒 A 和 B 顯著地更具苦味，所以他開始調查產品 P 上多餘的苦味是否可能存在污染。

六、多重樣品差異測試：評級測試 (Rating Approach) — 變異數分析 (ANOVA) 評量

A. 適用範圍及應用

當測試目標是為了確定 t 個樣品的特定感官屬性變化時，使用此方法。其中 t 可能為三至六個，或至多為八個，且可以視其為同一大組別同時進行比較。

註：在描述性分析（見第十章）中，當比較多個樣品時，可以應用此方法比較每一屬性。

B. 測試原理

品評員使用一個數量化的規模來判斷所選定屬性的強度等級，例如：一類別規模(category scale)（見第五章）。測試前必須指明使用的規模，並且透過變異數分析(analysis of variance)來評估結果。

C. 人員測試

參考先前章節篩選、培訓品評員。每組實驗最少使用八位品評員。如果有十六位或十六位以上的品評員，將可以大幅改進品評實驗的判斷準確性。品評員可能需要專門的培訓，使他們可重複辨識所選定的屬性（參見第 9 章）。根據測試目標的不同，可以依據屬性鑑別能力選擇受試者。

D. 測試程序

對於測試控制和產品控制，見第二章及第三章。如果情況允許的話同時提供所有樣品，否則按照順序提供。受試者依照平衡隨機的順序接受 t 個樣品，其任務是使用指定量度(scale)評定每個樣品。每組只能被呈現一次，或使用不同的編碼呈現數次。如果各組

合可以呈現兩次以上，實驗精確度將大幅改善。

假如有一個以上的屬性需要評估，理論上該樣品應按照每個測量屬性分別呈現。在實務上，因為評估多組時屬性數量仍受限於既定的樣品數量（典型為六至二十五組），為了妥善地分配樣品以評價各屬性，感官分析師接受某些屬性之間的一些相互依存關係。例如：如果在一個保鮮期研究的產品中，有**微生物霉味**（例如：酸味）或氧化味（例如，酸敗）等不同屬性，其中一個屬性的評分將會影響到另一個屬性，即使另一個屬性是不存在的。必須通過密集的培訓使品評員認識到這一點，並使他們能夠獨立地辨識及抵消各個屬性的相互助長。

E. 結果分析及解釋

使用變異數分析法分析結果，請參見第 13 章及下列範例。

例 7.6：以感官分析方法測試課程受歡迎程度—隨機完全區集設計(Randomized Complete Block Design)

問題／測試目的：食品科學系在每學期結束時例行性地要求學生評估他們的課程，採取的規模為-3 至+3，其中-3 是非常差的課程，0 是無特別感覺，+ 3 是優越的。30 名學生的評分結果記錄在表 7.3。評價的目的是確定課程是否需要改進。

結果分析：以隨機（完全）區集設計的變異數分析來整理數據。學生們被視為「區集」(Block)，評估的課程是「處理」(Treatment)。F 統計量(F-Statistic)在表 7.4 中的「評估課程」上結果是非常顯著 ($F_{3,87} = 12.91$, $P < 0.0001$)。因此，評估的結論為：沒有任何課程與平均之間有顯著的差異。若我們使用「Fisher 最小顯著差異測驗法」(LSD)的多重比較方法評估，可以找出哪個課程是顯著不同於其他課程（參見表 7.4 底部）。在 LSD 過程的結果表明：營養課程顯著較低（較差）於平均得分。其他三個課程的平均得分之間沒有顯著差異。該評估結果傳達給教授和各部門以採取進一步行動。

例 7.7：五種啤酒花特徵—裂區設計(Split plot design)

問題／情況：一個啤酒商正在籌備一個啤酒新品牌，希望其具有較高水準的啤酒花特徵。五種備選的啤酒花成本分別為\$ 1.00/磅、\$ 1.20/磅、\$ 1.40/磅、\$ 1.60/磅、和\$ 1.80 /磅。

計畫目的：選擇省錢的最佳啤酒花特性區集。

實驗目的：比較五種啤酒的啤酒花程度、確保實驗結果的可靠性。

實驗設計：一個有邏輯的方法是將五種啤酒隨機呈現給一群數量夠多的品評員，這是一種典型的多重採樣差異檢驗：20 名受試者使用圖 7.5 的記錄表記錄樣品的評價，量度規模為 0 至 9。呈現的順序是隨機的，並且使用三組不同的編碼呈現。

表 7.3 範例 7.6：多樣品差異測試的評分結果

學生編號	課程評估			
	生物	營養	感官	統計
1	2	-2	1	1
2	3	0	2	1
3	1	-3	0	0
4	2	0	1	0
5	0	1	0	0
6	-3	-3	-3	-3
7	1	3	1	1
8	-1	-1	-1	-1
9	2	-2	1	1
10	0	-3	-1	-1
11	2	0	2	2
12	-1	-2	0	1
13	3	-3	3	3
14	0	0	0	0
15	-2	2	-1	-1

16	2	-2	1	1
17	1	-1	0	0
18	0	-1	0	-1
19	3	3	3	3
20	1	-2	1	0
21	-2	-2	-2	-2
22	2	-1	1	1
23	1	0	1	1
24	3	-3	3	3
25	1	1	1	1
26	0	-1	1	-1
27	1	0	2	-1
28	2	-2	0	0
29	-2	-3	-1	-2
30	2	2	2	2

註：使用規模：-3 至+3，其中-3=非常差，0=無差別，+ 3=優秀。

表 7.4 為表 7.3 隨機完全區集設計的變異數分析的結果：食品科學課程的受歡迎程度

變異因子	自由度	平方和	平方平均數	F	P
總和	119	344.37			
學生(區集)	29	188.87			

評估課程	3	47.90	15.97	12.91	<0.0001
誤差	87	107.60	1.24		
使用 95% LSD 多重樣品比較結果得到的平均評級分數(average ratings)					
評估課程	生物學	營養學	感官學	統計學	
平均評分	0.80a	-0.83b	0.60a	0.30a	

註：平均評分後面所跟隨的字母不同時代表具有顯著差異，在 95% 信心水準下 $LSD_{95\%} = 0.57$ 。

圖 7.5 為多樣品差異測試（評級法）的紀錄表。範例 7.7：五種啤酒花特徵。

多樣品差異測試													
名稱：_____	日期：_____												
樣品類型： <u>啤酒</u>													
研究特點： <u>啤酒花特性</u>													
說明： 從左至右品評樣品，並注意研究屬性的強度。按下列比例給各樣品評分： <table> <tr> <td>0</td> <td>難以察覺</td> </tr> <tr> <td>1</td> <td></td> </tr> <tr> <td>2</td> <td>略微察覺</td> </tr> <tr> <td>3</td> <td></td> </tr> <tr> <td>4</td> <td>中度察覺</td> </tr> <tr> <td>5</td> <td></td> </tr> </table>		0	難以察覺	1		2	略微察覺	3		4	中度察覺	5	
0	難以察覺												
1													
2	略微察覺												
3													
4	中度察覺												
5													

6	強烈察覺
7	
8	極端察覺
9	
樣品編碼: _____ 評分: _____ 意見: _____	

審查樣品：要求兩名經驗豐富的品酒師品評樣品，以確定樣品足以代表要生產的同類啤酒，並且沒有除了啤酒花屬性以外的感覺差異。

結果分析：品評結果顯示於表 7.5，而相應的裂區設計的變異數分析(split-plot ANOVA)結果顯示於表 7.6。品評員與樣品之間的交互作用(The subject-by-sample interaction)則不顯著：

$$F_{交互} = 0.97, \Pr[F_{76, 190} \geq 0.97] = 0.56 > 0.05$$

樣品效應和品評員的效應均呈現高度顯著：

$$F_{樣品} = 41.88 \Pr[F_{4, 8} \geq 41.881] < 0.01$$

$$F_{品評員} = 17.79 \Pr[F_{19, 190} \geq 17.79] < 0.01$$

表 7.5

例 7.7 多樣品差異測試（評級法）－五種啤酒之啤酒花特性的結果

樣品編號.	1	2	3	4	5
1	2, 2, 1	3, 4, 5	1, 0, 2	5, 4, 3	3, 2, 4
2	0, 0, 1	1, 2, 1	0, 0, 0	2, 1, 2	2, 1, 1
3	0, 2, 1	2, 0, 2	0, 2, 0	2, 3, 2	0, 2, 2
4	3, 3, 3	4, 5, 6	2, 3, 1	5, 8, 4	5, 6, 4
5	2, 4, 3	4, 3, 1	3, 0, 3	3, 5, 6	1, 4, 3
6	2, 4, 1	3, 2, 4	3, 2, 1	4, 6, 7	3, 4, 2
7	0, 0, 1	1, 2, 1	0, 0, 0	0, 2, 1	2, 1, 1
8	6, 4, 3	4, 6, 3	3, 4, 6	4, 6, 3	3, 4, 6
9	2, 2, 2	3, 3, 5	0, 1, 1	4, 6, 5	3, 5, 3
10	1, 4, 3	2, 5, 3	2, 0, 2	5, 4, 5	5, 2, 3
11	3, 4, 2	1, 3, 4	3, 0, 3	6, 5, 3	3, 4, 1
12	1, 0, 0	1, 2, 1	0, 0, 0	1, 2, 1	1, 1, 2
13	1, 0, 0	1, 2, 1	0, 0, 0	2, 1, 2	1, 1, 2
14	3, 3, 3	6, 5, 4	1, 3, 2	4, 8, 5	4, 6, 5
15	2, 2, 2	5, 3, 3	1, 1, 0	5, 6, 4	3, 5, 3
16	1, 4, 2	4, 2, 3	1, 2, 3	7, 6, 4	2, 4, 3
17	3, 4, 1	3, 5, 2	2, 0, 2	5, 4, 5	3, 2, 5
18	1, 2, 0	2, 0, 2	0, 2, 0	2, 3, 2	2, 2, 0
19	1, 2, 2	5, 4, 3	2, 0, 1	3, 4, 5	4, 2, 3
20	3, 4, 6	3, 6, 4	6, 4, 3	3, 6, 4	6, 4, 3

說明：舉例編號 20 的品評員來說，他品評第一個樣品編號為“3”的樣品，第二次品評樣品編號“4”，接著第三次品評樣品編號“6”。

因為交互作用不顯著，可以假設品評員評定各樣品的評分是一致的。然而品評員效應的顯著差異表示各個品評員表達自己的看法時使用的比例不同。這並非不常見。然而，只要沒有相互作用，品評員之間的差異通常是次要的研究。樣品之間的差異才是我們首要關心的問題。使用 HSD 多重比較法確定哪個樣品的啤酒花特徵呈現顯著不同，如下所示：

樣品	4	2	5	1	3
平均	<u>3.9</u>	<u>3.0</u>	<u>2.9</u>	<u>2.1</u>	<u>1.4</u>

注意：若平均下方沒有共同劃線連接，則代表在 5% 的顯著性水準下具有顯著不同。 $HSD_{5\%} = q_{0.05, 5, 8} \sqrt{MSE_{\text{Error}}(A)/n} = 4.89 \sqrt{1.32/60} = 0.7$ (q 值來自表 17.4)。

表 7.6 表 7.5 裂區設計的變異數分析結果：五大啤酒之啤酒花特性

變異來源	自由度	平方和	均方	F
總計	299	975.64		
重複	2	8.89		
樣品	4	221.52	55.38	41.88 ^a
誤差(A)	8	10.58	1.32	
品評員	19	412.30	21.70	17.79 ^a
樣品 X 品評員	76	89.81	1.18	0.97
誤差 (B)	190	232.53	1.22	

註：在 1% 的顯著性水準下，誤差 (A) 是重複 X 樣品的相互作用。誤差 (B) 則通過減法得之。

樣品 4 的平均評分(average ratings)顯著地高於其他所有樣品。而樣品 2 和樣品 5 具有幾乎相同的平均評分，且啤酒花特性顯著低於樣品 4，但顯著高於樣品 1 和樣品 3。樣品 1 和 3 的啤酒花特性表現則顯著低於樣品 2、樣品 4 和樣品 5。報告結果解釋：釀酒商的感官分析報告中，含有樣品評分平均及變異數分析的結果，並得出結論為，測試的 5 個樣品中，樣品 4 的啤酒花特徵具有顯著較高的水準。而另一個成本較低廉的樣品 2 的啤酒花品種也是值得考慮的。

7.7 多樣品差異測試：BIB 排序測試(平衡不完全區集設計)(BIB Ranking Test <Balanced Incomplete Block Design>)- Friedman 分析

A. 適用範圍及應用

當實驗目的是為了確定特定的感官屬性在多個樣品中以何種方式變化，且實驗中具有大量的樣品需評估時，使用此方法。通常比較的樣品數目為 6 至 12 個，或至多 16 個時，使用此方法。

當品評員對樣品類型和(或)偏愛的較簡單統計分析較未經訓練，可選擇本方法（排序法<ranking>）。當品評員經過訓練可使用評級量度(rating scale)時，則使用 7.8 節記載的方法（評級法< rating >）。

B. 測試原理

相較於之前將所有 t 樣品呈現在同一個大區集，此測試根據 Cochran 和 Cox（1957），或 Design Express（2003 年）的設計將樣品在不同小區集中呈現。要求品評員根據所研究的屬性排序樣品。

C. 品評員測試

根據前列章節甄選、培訓品評員。確認該品評員具有足夠判別正在研究屬性的能力。例如通過判斷已知強度等級屬性等訓練。

D. 測試程序

對於測試控制和產品控制：如果可能的話同時提供樣品，否則按照順序提供。請確保樣品呈現的順序是真正隨機的：不得導致品

評員懷疑樣品呈現有一定的規律，因為這會影響判斷。測試說明文字如：「根據甜度排序樣品，最低甜度的樣品給予等級 1，次低的給予等級 2，等等。」

例 7.8：魚的種類

問題/情況：軍事口糧 XPQ-6（魚條凍）以往皆是由來自 15 個不同品種的魚所製成。然而此口糧被嚴重的投訴說有「魚腥味」。目前已追查使用某些物種的魚類會造成此魚腥味。決策者希望判別出那些物種造成魚腥味，以便能夠在糧食可用性和價格之間做權衡。

計畫目的：比較 15 種魚在定量使用下所造成的腥味問題程度。

測試目的：比較 15 種魚所製成的魚條凍的腥味程度。

測試設計：選擇平衡不完全區集設計的多樣品差異測試(Multisample Difference test)，因為它允許以三個一組的方式來比較 15 個產品。105 名品評員按照隨機原則分為 35 組，每組三個品評員。每一組的品評員按照表 7.7 得到隨機分配設計的 35 組中的一組樣品。記錄品評員根據腥味排序他所分配到的三個樣品，從最少 (= 1) 到最多 (= 3)。

表 7.7

多樣品差異測試：BIB 設計—例 7.8 - 無菌魚條凍($t= 15, k= 3, r= 7, b= 35, \iota = 1, E= 0.71$)

組別			
(1)	1	2	3
(2)	4	8	12
(3)	5	10	15
(4)	6	11	13
(5)	7	9	14
(6)	1	4	5
(7)	2	8	10
(8)	3	13	14

(9)	6	9	15
(10)	7	11	12
(11)	1	6	7
(12)	2	9	11
(13)	3	12	15
(14)	4	10	14
(15)	5	8	13
(16)	1	8	9
(17)	2	13	15
(18)	3	4	7
(19)	5	11	14
(20)	6	10	12
(21)	1	10	11
(22)	2	12	14
(23)	3	5	6
(24)	4	9	13
(25)	7	8	15
(26)	1	12	13
(27)	2	5	7
(28)	3	9	10
(29)	4	11	15
(30)	6	8	14
(31)	1	14	15

(32)	2	4	6
(33)	3	8	11
(34)	5	9	12
(35)	7	10	13

資料來源：W. G. Cochran, 及 G. M. Cox. 1957. 實驗設計, New York : Wiley. 經許可。

審查樣品：在廚師的幫助下，將各品種製成的魚條樣品盡可能擁有相似的組織、外觀和風味，並以烹調方式及次要成分的合適改變，減少品種屬性之間的差異。每個呈現的樣品要篩選外觀，並丟棄任何碎屑或顯示其他可見偏差的碎片。

結果分析：為了使結果更容易分析，表 7.8 顯示研究的排序資料。一個既定品種的魚類的排序和(rank sum)僅僅是加總所有對應於該種類的列中的數字。以 Friedman 的檢驗統計學的 T(statistic T) (見公式 13.18) 值，確定是否各品種的腥味強度之間存有差異。T = 68.53 的值超過 χ^2 的上限 5% 及 $(t-1) = 14$ 個自由度的臨界值 ($\chi^2_{14, 0.05} = 23.69$)。因此可以得出結論：在各區集中確實存在顯著差異。接著，使用公式 13.19 計算 95% 的 LSD 多重比較的值，以確定哪個品種具有顯著不同。(見表 7.9)。

解釋和報告結果：軍事領導人從表 7.9 中得出結論：在價格和可用性的考量下，樣品 5、15、13、1、6 及 9 應保留，因為這些品種產生的腥味程度最低，並彼此不具有顯著不同。另外暫時保留樣品 14、8、4，假如第一組中的品種因成本高或不可用等因素被消除的話。這樣的決定是由於樣品 14、8、4 相較於樣品 5 和 15 有顯著性更高的魚腥味，但並不顯著於第一組中剩餘的其他樣品。表 7.9 其餘的樣品 (2、11、10、12、3、7) 則由 XPQ-6 比例計算後去除。

表 7.8 例 7.8 中獲得的結果，多樣品差異測試：BIB 設計與排序數據 - 魚條凍

樣品/品種															
區集/品評員	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	1	2	3												
2				3			1			2					

3					1					3					2
4						3					2		1		
5							3		1					2	
6	3			2	1										
7		2						3		1					
8			3										2	1	
9						3			2						1
10							2				1	3			
—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
101	2													3	1
102		1		3		2									
103			1					2			3				
104					1				2			3			
105							3			2			1		
排序和	35	45	54	43	28	37	55	42	37	50	49	50	34	42	29

註：1：最不腥；2，中間；3，最腥。

表 7.9

表 7.8 魚條凍的結果總結及統計分析

樣品/品種	排序和

5	28	a				
15	29	a				
13	34	a	b			
1	35	a	b	c		
6	37	a	b	c		
9	37	a	b	c		
14	42		b	c	d	
8	42		b	c	d	
4	43		b	c	d	
2	45			c	d	e
11	49				d	e
10	50				d	e
12	50				d	e
3	54					e
7	55					e

註：平均若顯示相同字母則表示在 5% 的信心水準下 (LSD = 10.74) 無顯著不同。

7.8 多樣品差異測試：BIB 評級法（平衡不完全區集設計）－變異數分析

A. 適用範圍及應用

當實驗目標是為了確定特定的感官屬性在若干樣品中以何種方式變化時，且實驗中具有大量的樣品需評估時，使用此方法。通常待比較的樣品數目為 6 至 12 個，或至多 16 個時，使用此方法。

當品評員已經受過良好的評級訓練，且品評結果必須盡可能精確時，使用此方法。若品評成員僅經較少的訓練，或是排序法測試的結

果已經足以提供相關資訊，則使用 7.7 節的方法。

注意：在描述性分析（見第十章）中，當欲比較的樣品數量非常大時，本方法可以應用於每個屬性。

B. 測試原理

相較於之前將所有 t 樣品呈現在同一個大的區集，此測試根據 Cochran 及 Cox (1957)，或 Design Express (2003 年) 的設計將樣品在不同小區集中呈現。要求品評員使用數值化的等級將各樣品的屬性強度評級（見第五章）。使用的等級必須明確說明，並且品評後結果使用變異數分析研究。

C. 品評員測試

根據前列章節甄選、培訓品評員。確認品評員可以清楚判別研究屬性的強度，例如通過已知屬性強度的樣品等訓練。每次品評測試安排不少於 8 位品評員，16 位或更多品評員可以大為改善測試精準度。品評員可能會要求進一步的說明，使他們能夠更精準地辨認正在研究的屬性（見第 9 章）。根據測試目的，可以依照各品評員針對研究屬性的辨認能力高低選擇受試者。

D. 測試程序

對於測試控制和產品控制：如果可能的話同時提供樣品，否則按照順序提供。請確保樣品呈現的順序是真正隨機的：不得導致品評員懷疑樣品呈現有一定的規律，因為這會影響判斷。

注意：如果有多於一個的屬性需要評級，不可避免地各個屬性的評分會出現部分相互依存關係（見 4.4 節）。

E. 分析及結果解釋

使用變異數分析結果，請參見第 13 章和下面的實例 7.9。

例 7.9：冰淇淋的參考樣品

問題/情況：冰淇淋廠的品質管理經理有一個持續進行的計畫，即定期篩選每批生產的產品，以便能夠選擇將各批產品加入到品質控制

測試時的參考樣品的樣品庫中。品質控制測試的參考樣品需要定期更新，是因為舊的樣品將隨時間變化而變得不再合適。該過程也用消除在樣品庫中可能已經惡化的任何參考樣品。

表 7.10

實例 7.9：冰淇淋的參考樣品的 BIB 設計($t=6$, $k=4$, $r=10$, $b=15$, $\ell=6$, $E=0.90$)

區集				
(1)	1	2	3	4
(2)	1	4	5	6
(3)	2	3	5	6
(4)	1	2	3	5
(5)	1	2	4	6
(6)	3	4	5	6
(7)	1	2	3	6
(8)	1	3	4	5
(9)	2	4	5	6
(10)	1	2	4	5
(11)	1	3	5	6
(12)	2	3	4	6
(13)	1	2	5	6
(14)	1	3	4	6
(15)	2	3	4	5

資料來源：W. G. Cochran 和 G. M. Cox (1957) 實驗設計：Wiley，紐約。

表 7.11 例 7.8 的結果表：冰淇淋的參考樣品

區集/品評員	樣品					
	1	2	3	4	5	6
1	6	1	1	2		
2	6			1	3	3
3		4	2		5	2
4	7	2	3		2	
5	3	5		1		1
6			1	1	3	2
7	7	4	4			3
8	2		1	1	1	
9		2		2	2	3
10	4	2		2	5	
11	5		3		1	1
12		3	2	1		2
13	4	2			1	1
14	5		2	2		1
15		2	4	5	3	
調整平均	<u>5.0</u>	<u>2.5</u>	<u>2.2</u>	<u>2.0</u>	<u>2.6</u>	<u>1.9</u>

註 1：經評級法設計的 BIB。註 2：回應：10 點等級量表 0 = 無異味，9 = 極端異味。註 3：若經調整的平均下面沒有橫線的話，代表該樣品在 5% 的信心水準下 ($LSD_{5\%} = 1.1$) 顯著不同。

表 7.12 例 7.9：冰淇淋的參考樣品—以 BIB 設計的變異數分析表

變異來源	自由度	平方和	平方均	F	P
總和	59	150.98			
評審（區集）	14	39.73			
樣品（處理，根據區集調整）	5	59.89	11.98	9.33	<0.0001
誤差	40	51.36	1.28		

計畫目的：維持供應品管用的冰淇淋參考樣品庫存充足。

測試目的：每一天測試六個批次的庫存，剷除有異味且不適合作為參考樣品的庫存。

測試設計：15 位訓練有素的品評員使用 10 點等級量表，從 0（無異味）到 9（極端異味）評估六個批次的樣品整體是否有異味。品評員不能在一次測試中評估超過四個以上的樣品，因此，感官分析師選擇了 Cochran 和 Cox (1957) 的 BIB 設計（見表 7.10）。15 位品評員中的每一位被隨機分配含有 4 個樣品的區集。每個區集內的樣品的呈現順序是隨機的。

結果分析：整體異味的屬性評分結果顯示在表 7.11。這些資料由能夠執行平衡不完全區集變異數分析（參見第 13 章）的電腦程式分析。表 7.12 為 BIB ANOVA 的結果。“處理”（即冰淇淋樣品）的 F-統計相較於 F-分佈（上限 5% 的臨界值 $t - 1 = 5$ ，自由度 $tpr - t - pb + 1 = 40$ ）為顯著（F-統計 = 9.33 > F-分佈 ($F_{0.05; 5, 40}$) = 2.45）。並且使用 LSD 多重比較方法評估平均評分，以確定哪些樣品具有顯著不同的整體異味（見表 7.11 下方的註 3）。

報告結果和解釋：樣品 1 的平均異味評分比起其餘樣品的平均評分顯著更高。而其它樣品的平均評分之間並不存在顯著差異。感官分析師報告此結果給品管經理，並且建議從參考樣品庫中移除樣品 1。

第八章 感覺閾值和敏感度測試

8 Sensory Threshold and Sensitivity

感覺閾值(threshold)是指從剛好能引起感覺到剛好不能引起感覺刺激強度的一個範圍。**敏感性**(sensitivity)是感覺器官對刺激的感受、識別和分辨能力。敏感性亦稱**感受性**。

8.1 感覺閾值簡介

一般而言，**感覺閾值**(threshold)是指刺激強度的一個範圍。但在理論上感覺閾值的定義並不明確。若想要好好地定義它，需要數百種有控制的比較實驗，且其結果無法好好地複製重現。Swets 甚至懷疑感覺閾值是否存在。其首要反思是，感覺閾值研究是徒勞地投入時間和金錢。然而，下面段落中描述的方式仍是目前定義閾值最好的方法。

流動空氣的閾值是以自動流程嗅聞技術(olfactometry)測定，用來測定空氣的污染程度，並設定污染物的法定上限。化合物的閾值則多用於水源、食品、飲料、化妝品、塗料、溶劑等，是為了測定那些已知的污染物的可接受的程度。這是其最重要的用途，且測試可由數百位品評員進行以代表相對敏感度的群體分配。閾值也可以被用來作為選擇或測試品評員的一種方法，但這不應該是選擇的主要基礎（見第 9 章），除非測試目標是要求測試低刺激的樣品。添加物的閾值也可適用在食品，飲料等配方的研究工具。

必須記住若在給定化合物上設定低檢測閾值，其對應於所研究的味道則具有高敏感度。氣味單位 (O. U.) 或風味單位 (F. U.) 則使用閾值作為風味濃度的量測方式。例如：如果硫化氫(H_2S)從瓶中洩逸到房間中，當氣味濃度達到閾值 1 O. U. ;且為兩倍的硫化氫(H_2S)濃度時，其水準強度為 2 O. U. ，以此類推。必需謹慎地使用閾值，若閾值水準在 3 – 6 O. U. 時，並不適用上述閾值的水準（第 2 章）。預估感官強度的步驟在第 5 章討論。

用於測定嗅覺閾值的方法會對結果產生深遠的影響。當氣味流量低於一般的嗅覺探測量(2 升/秒)時，將引起過高的閾值。使用較大嗅瓶可能導致 10 至 20 倍低的閾值，因為品評員能夠提高氣味的探測量。而培訓可以降低閾值高達 1000 倍。

經驗證實，經過實踐和訓練，一組固定的品評小組 (> 25 人) 可以得到結果為 $\pm 20\%$ 的再現性水準，而一組人數夠多(>25)的品評小組和另一組人之間的結果為 $\pm 50\%$ 。最重要的因素是，除了如第 4 章所描述的與實際測試物的反覆訓練外，受試者可能為了討好實驗

者而盡量降低閾值，而這必須由一絲不苟的樣品製備細節抵消，以便不留下任何其他線索。

8.2 閾值定義

閾值是感官能力的極限。把剛剛能引起感覺最小的刺激量稱為**絕對閾值**的下限，又稱**刺激閾**或**察覺閾**。絕對閾值（檢測閾值，Absolute Threshold）是能夠產生感官感受的最低刺激值，是人可感受有別於正常環境但尚無法辨識的刺激濃度：最暗的光、最輕的聲音、最輕的重量、最淡的風味。低於下限的刺激稱為**閾下刺激**。剛剛導致感覺消失的最大刺激量，稱為絕對閾值的上限，高於上限的刺激稱為**閾上刺激**。

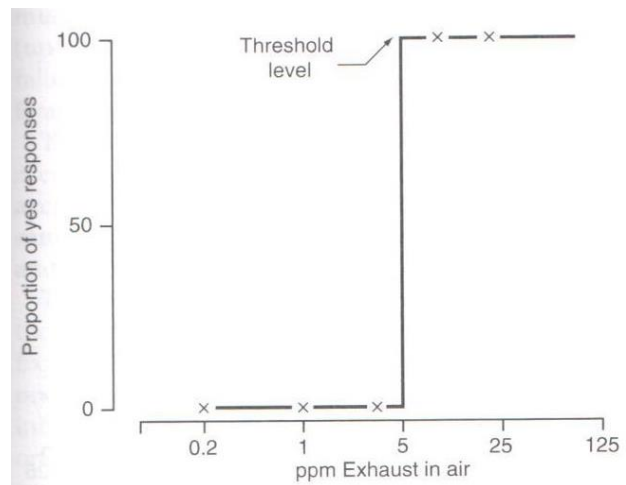
辨識閾值(Recognition Threshold)是指人可以辨識出特定感官特性的最低刺激濃度。辨識閾值通常高於絕對閾值。如果品評員品嚐含有漸增濃度的糖水時，其感覺會從“純淨水味”過渡到“非常輕微的味道”，而隨著蔗糖濃度增加，進一步轉變成“輕微的味道”。第二次感官的轉變即稱為辨識閾值。

差異閾值(Difference Threshold)則是剛剛能引起差異感覺刺激的最小變化量，是在產生可鑑別差異時需要的刺激變化程度。通常是透過比較一個標準刺激和可變刺激來確定。差異閾值又稱最小可覺差(just noticeable difference, JND)，不是一個恆定值，會隨一些因素而變化。最小可覺差(JND)這個術語指的是當差異閾值是透過少量改變的可變刺激直到受試者感覺到差異的方式來定義。此方法在第5章中有直接的討論。

終端閾值(Terminal Threshold)則是在超過可感知差異之後增加的刺激範圍；高於這個水準的刺激通常會引起疼痛。

最小可覺差(JND)通常隨著濃度比例的增加而提高，並且它們也被用作是感覺強度的比例。品評員可以區分絕對閾值和終端閾值之間大約 29 JNDs。然而，若實際應用在當作感知強度的比例單位時，閾值的變化因人而異，各品評小組的閾值也各有不同。

閾值的常見概念顯示在圖 8.1（例如：空氣中的柴油廢氣）。高於 5ppm 的排氣可以被檢出；低於 5ppm 則不能被檢測到。然而，若一位觀察者重複使用其**嗅聞**探測計(olfactometer)觀察，將產生如圖 8.2 中所示的結果。觀察者的嗅覺靈敏度將因其神經系統的記憶性及靈敏性變化而變化。在一定距離聽鐘錶的滴答聲，有時可以聽到，可能下一次滴答聲就聽不到了，而再下一次又再度聽到，等等。閾值不是一個固定的值，而是一個連續刺激的值。按照慣例，觀察者本人能夠集中時可感測到閾值刺激超過 50%的機率，而在誤差頻繁發生時不能夠檢測其閾值刺激有 X%的顯著性。作為規則，人們發現在 50% 機率可以探測的機率所描繪的實驗結果，如例 8.2 中描述的方法，變換為點線圖後則為典型的高斯劑量-反應曲線。

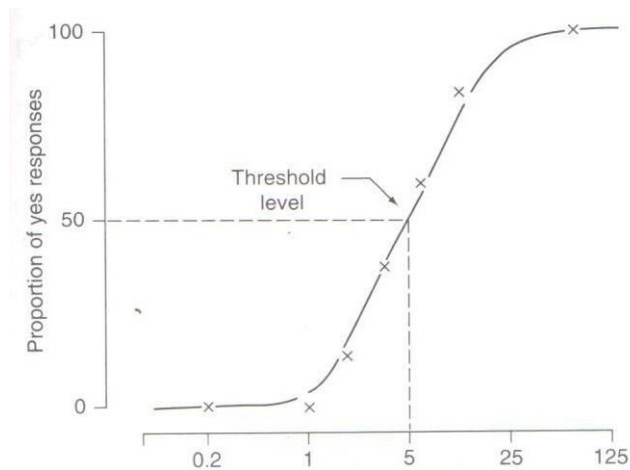


肯定回應的比例

閾值水準

空氣中的廢氣 PPM

圖 8.1 絕對閾值的常規概念（空氣中的廢氣）。



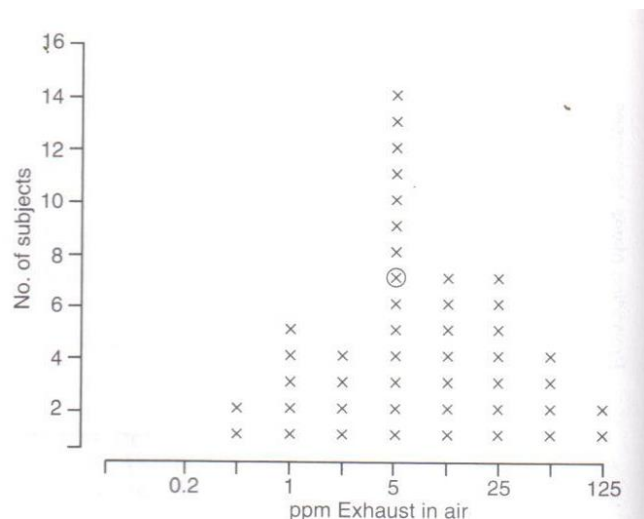
肯定回應的比例

閾值水準

空氣中的廢氣 PPM

圖 8.2 個人閾值的典型數據（空氣中的廢氣）。

收集個人閾值的數值到小組閾值的過程中，我們注意到，頻率分佈圖大多為鐘形圖。而曲線的右手尾也有比左側長的趨勢（見圖 8.3），因為大多數的小組皆可能有一定比例的成員是非常不敏感的。集中趨勢的方法可以使幾何平均值更具意義，因為它使最高的閾值有較少的權重。排序機率圖（rank probability graph，圖 8.5）是測試一組單獨閾值是否為常態分佈的一個有用的工具，如果圖中所有的點可以連成一條直線即判定為常態分佈。在這種情況下，該圖不僅可以用於判定 50% 機率定義的閾值，也可用於比如說 5% 或 90% 相應可檢測出的人數。



品評員編號

空氣中的廢氣 PPM

圖 8.3 含有 45 位受試者的品評小組的典型直方圖； f = 圖 8.2 的品評員。

8.3 閾值測定的應用

閾值可通過各種傳統心理物理學的設計來測量，例如：限制法、平均誤差法或頻率法。近年來，心理物理學傾向使用信號檢測理論（signal detection theory, SDT）來選擇不同的方式。SDT 是一種系統，其方法建立在認為閾值概念並非一個實值的思維上，而是在「兩個刺激中間的心理感官差異大小」，即為 d' 。SDT 的優點是，品評員的決定過程變得更加明確，可以使用統計建立模型。然而，SDT 的步驟比起傳統閾值設計較為費時，且已被證明若使用強迫選擇法比較兩種樣品， d' 與傳統閾值的關係為 1:1。由於這些原因，美國測試材料協會（ASTM）和國際標準化組織（ISO 2002）已決定堅持使用限制法和三選項必選法（3-AFC）：三個樣品中兩個是控制組而另一個含有待測物質。基於最小測試工作（例如：50~150 3-AFC 呈現）ASTM 快速法（E679，見例 8.1）針對測定接近閾值

的實際值。它使得每位小組成員閾值的最佳評估值變得非常相近（例如， $\pm 200\%$ ）。相反地，小組規模必須更大，使得團體閾值 and 分佈變得更為可靠，因為團體之間的變化（多達 100 倍）比起個人閾值實驗之間的變化（高達 5 倍）還要大得多。其最佳結果可能只有微小的偏差，但如果品評員的閾值落在最大值及最小值的區間之外，且實驗並沒有被重新檢視，有時也可能產生極大的偏差（見實例 8.1）。

按照圖 8.2，ASTM 居中法 (ASTM's intermediate method, E1432) 針對測定個人閾值，接著在第二步驟中，根據圖 8.3 來測定團體閾值。關於此步驟，當樣品以快速法測試呈現時給每位品評成員五次以上。而得到的回報是：團體閾值和個人閾值的分布皆無偏差。

ISO 標準 13301 (ISO 2002) 即是反應上述兩種方法的組合。曲線擬合 (Curve-fitting) 步驟使用非線性最小平方回歸法 (nonlinear least squares regression) (參見實例 8.2)。(ISO 的方法允許回歸且最大可能性法 <logistic regression and a maximum likelihood procedure>，因此我們可以使用電腦的電子表格計算)。假若研究人員需要比起這些方法所預期得到的結果還要更精確的答案，則進入了研究計畫領域的範圍，因此可能需要找尋更適當的實驗設計。例如：鮑爾斯多對測試 (Powers' Multiple Pairs test) 或信號檢測論。基於 β -二項式分佈 (Beta-Binomial distribution) 的概念，即個人閾值的數據往往比起團體閾值的數據分布來得更窄。

例 8.1 添加至啤酒中的火灼風味物質的閾值

問題/狀況：一個釀酒商注意到啤酒暴露於 UV 光下會產生火灼風味 (3-甲基-2-丁烯-1-硫醇 <3-methyl-2-butene-1-thiol>，不應存在的化合物)，因此希望測試各種類型的包裝材料以提供啤酒足夠的保護。

計畫目的：選擇能夠提供啤酒足夠的保護下最小成本的包裝，使用標準照射期間形成的火灼化合物的量，並與閾值量相比較。

測試目的：為了測定純 3-甲基-2-丁烯-1-硫醇在該廠啤酒中的閾值。

實驗設計：E679 快速檢測 (E679 rapid test) 適合掩蓋掉品評員之間的變異；由 25 位專業的品評員組成的小組，每位分別接收 6 個 3-AFC 測試，包含三種不同濃度的樣品。限制偏差：(1) 選用五位品評員的初步實驗得到的濃度範圍，和 (2) 最低閾值正確及在最高閾值錯誤的品評員，需重新測試，。

樣本準備：在初步實驗中，確定所選啤酒無火灼風味，且在所選定的實驗濃度下選定 3-甲基-2-丁烯-1-硫醇以提供純 sunstruck 風味。

執行測試：以 6 種濃度測試每位小組成員。測試誰能夠在最低濃度時正確一次或在此濃度兩次以上，包括在一或兩次濃度的組合。同樣地，測試誰在最高濃度時仍失敗一次至兩次以上。記錄及分析的結果顯示於表 8.1。每位品評員的最佳估計閾值 (best estimate

threshold ,BET) 是其錯誤的最高濃度和下一個更高濃度的幾何平均值。該團體的 BET 是個人 BET 的幾何平均值。找另一天重複此測試至少一次，並使用相同的品評員。請注意，閾值下降往往作為鑑定品評員已經習慣此物質味道的機制。如果閾值降低 20% 以上，重複測試直到值穩定。

測試報告：包括完整的表 8.1，並給小組成員的人口統計數據。

表 8.1 添加到啤酒中的火灼風味的感覺閾值

品評員	呈現的濃度 (PPB)							最佳估計閾值	
	0.27	0.80	2.41	7.28	21.7	65.2	195	Over	ppb Log (10)
01		0	0	+	+	+	+		4.19 0.622
02	0	+	+	+	+	+	+		0.46 -0.337
03		0	+	+	+	+	+		1.39 0.143
04	0	+	+	+	+	+	+		0.46 -0.337
05		0	+	0	+	+	+		12.6 1.100
06		0	+	+	+	+	+		1.39 0.143
07		+	0	+	+	+	+		4.19 0.622
08	0	+	+	+	+	+	+		0.46 -0.337
09	0	+	+	+	+	+	+		0.46 -0.337
10		0	+	0	0	+	0	+	338 2.529
11		0	+	+	+	+	+		1.39 0.143
12		0	+	+	+	+	+		1.39 0.143
13		+	0	+	+	+	+		4.19 0.622

14		0	0	+	+	+	+	4.19	0.622
15	0	+	+	+	+	+	+	0.46	-0.337
16		0	+	0	+	+	+	12.6	1.100
17		0	+	+	+	+	+	1.39	0.143
18		+	+	0	0	+	+	37.7	1.576
19	0	+	+	+	+	+	+	0.46	-0.337
20		+	0	+	+	+	+	4.19	0.622
21		0	+	+	+	+	+	1.39	0.143
22	0	+	+	+	+	+	+	0.46	-0.337
23		+	0	0	+	+	+	12.6	1.100
24	0	+	+	+	+	+	+	0.46	-0.337
25		0	0	+	+	+	+	4.19	0.622
總和 →								9.299	

團體 BET，幾何平均值(ppb) 2.35 0.3720

紀錄標準差= 0.7 19

個人閾值的直方圖

幾何平均= 2.35 ppb

24
22
19 21 25
15 17 20
09 12 14

08	11	13	23			
04	06	07	16			
02	03	01	05	18	10	
0.46	1.39	4.19	12.6	37.7	113	338

方法：ASTM E679 上升界限法—濃度系列 (Ascending concentration series method of limits)。

裝備：無色啤酒杯，250 毫升；每杯啤酒 “A” 50 毫升。

樣品：將 3-甲基-2-丁烯-1-硫醇 (3-Methyl-2-butene- 1 -thiol ,Aldrich)

純化：通過兩列製備型氣相色層分析

規模步驟：每個步驟 6 個濃度因子：3.0

受試者人數：25

高和低的結果證實了嗎？是

3-甲基-2-丁烯-1-硫醇，在啤酒的 “A” 的濃度 ppb。

例 8.2：空氣中異戊酸(Isovaleric Acid)的閾值

問題/狀況：渲染植物產生含有異戊酸的活化風味成份。鄰居們抱怨和通過條例要求減少到閾值以下。

計畫目的：各種工藝方案和更高的煙囪之間進行選擇。

測試目的：為了測定在空氣中異戊酸的閾值。

實驗設計：一個相當徹底的方法，如 ASTM 居中法 (E1432) 或 ISO 標準 13301 的第二例，因為文件精簡極為適用。使用動態嗅聞計和二十位小組成員 (CEN, 1997)。提前選定雙重隔開的五種(含)以上濃度，每位小組成員對每一種濃度進行 3-AFC 測試 6 次 (見下文)。該裝置包含三個嗅探端口，小組成員知道有兩個生產無氣味的空氣，必須選擇他認為含有添加異戊酸者。添加的濃度從最低濃度在每一後續的測試中按一個常數因子(constant factor)逐漸增加。從各濃度的正確結果的百分比，計算每位小組成員的閾值，並由此計算該組的閾值。

篩選樣本：確定該氣源無可偵測氣味，且異戊酸有感官純度且無外來氣味。以化學分析檢查嗅覺的可靠性。

進行測試：在選定的濃度依次測試每位小組成員。使事先這種選擇，通過每一組廣泛間隔濃度的單一測試（或幾個測試）（例如：2.5 %、10、40、160 和 640 ppb）。在測試中，如果小組成員應在某最低濃度得分 100% 正確，以此為最高濃度重新安排濃度系列。同樣地，如果一位小組成員分數的最高濃度小於 80% 的正確，此繼續執行該系列直到較高濃度不再發生此狀況。

分析結果：如圖 8.4，其中橫坐標是濃度，x（或 log 濃度），縱軸的數據是比例區分值（或機率以上正確的百分比）， P_d 。鈹是從正確比例（ P_c ）獲得，如下：

測試	公式	機會等級
三角試驗或 3-AFC	$P_d = 1.5 \times P_c - 0.5$	0.333
配對比較或 2-AFC	$P_d = 2.0 \times P_c - 1.0$	0.500
五選二(Two-of five)	$P_d = 1.111 \times P_c - 0.111$	0.100

以 ASTM E1432 或 ISO 13301 允許的六曲線配適法(curve-fitting method)之一計算個別閾值（例如：使用電腦程序組合的邏輯回歸，如圖 8.4）。以排序/概率圖劃出個別閾值，如圖 8.5，並以 50% 點獲得該組閾值。如果一條直線可通過點被繪製，可得出這樣的結論：小組成員表示正態分佈和其它感興趣的點可以從圖中讀出，例如：群體人口的 10% 類似於小組成員可以檢測者。（曲線配適法：通常實驗資料是一些離散資料，我們想從這堆資料中找到其規則，於是我們繪出它這些離散點的圖，因為實驗會有誤差，這些點通常不容易看出其規則，於是應用各種數值方法找出可以符合這些離散資料的最適合的曲線）。

測試報告 - 包括表 8.2 和圖 8.5 中的資料，並做成小組成員的人口統計數據。

**用途：適用 Logistic 模型； $P = (1/3 + \text{EXP}[K]) / (1 + \text{EXP}(K))$,

**其中 $K = B (T - \text{LOG} [X])$,

** P 是正確辨識的比例，

** B 是斜率，

** X 是異戊酸的在空氣中的實際濃度 (ppb) ,

**和 T 是以 $\log(\text{ppb})$ 表示的閾值。

```
PROC NLIN  Method=DUD  Data=Input;  by panelist;
PARMS B=-4  T=2
K  =  B*(T - LOG10(K));
K  =  EXP(K);
K  =  (1/3 + E) ;
E  =  (1 + E) ;
MODEL  P  =  N/D
TITLE2  'Logistic Regression Modles' ;
RUN;
```

13 位小組成員的輸出：

使用 SAS PROC NLIN 的閾值數據邏輯回歸
邏輯回歸模型。

非線性最小平方迭代階段(Non-linear Least Squares Iterative Phase)

因變量：P 方法：DUD

重述	B	T	剩餘 SS
-3	-4	2. 000000000	0. 025885700365
-2	-4. 4	2. 000000000	0. 02054415598
-1	-4	2. 200000000	0. 084958944779

0	-4.4	2.000000000	0.02054415598
1	-5.852958	1.961443385	0.010812277188
2	-6.259745	1.967823308	0.010766524899
3	-6.189164	1.951938036	0.010504941622
4	-6.283542	1.954261395	0.010481402394
5	-6.280162	1.954257276	0.010481361251
6	-6.281544	1.954068199	0.010481219887
7	-6.277816	1.953905805	0.010481193047
8	-6.280506	1.953919346	0.010481176612
9	-6.281737	1.953896400	0.010481179219
10	-6.281715	1.953899496	0.010481176193

符合收斂標準。

非線性最小平方匯總統計相關變量 p

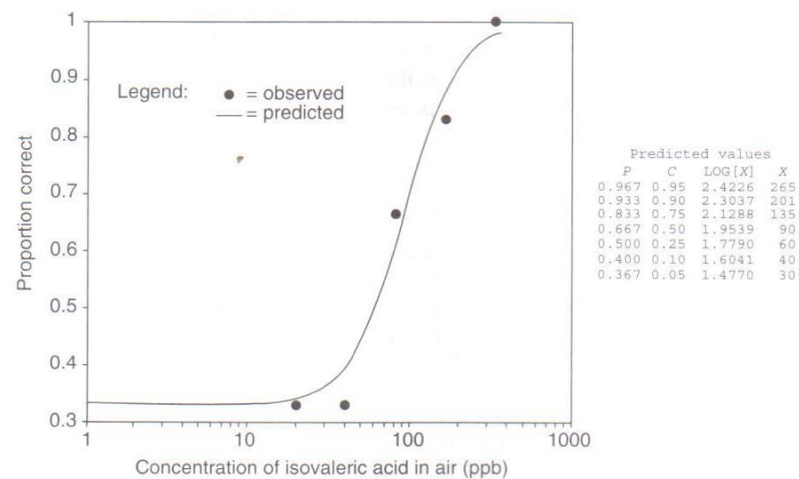
來源	DF	總和	均方
回歸	2	2.3500748238	1.1750374119
剩餘	3	0.0104811762	0.0034937254
未更正總數	5	2.3605560000	
(更正總數)	4	0.3558448000	

		漸近 (Asymptotic)	漸近 95% 置信區間

參數	估計	標準誤差	下限	上限
B	-6.281714751	1.6824126163	-11.6359929.3	-0.9274366000
T	1.953899496	0.0473965533	1.803059965	2.1047390264

預測值

P	C	LOG(X]	X
0.967	0.95	2.4226	265
0.933	0.90	2.3037	201
0.833	0.75	2.1288	135
0.667	0.50	1.9539	90
0.500	0.25	1.7790	60
0.400	0.10	1.6041	40
0.367	0.05	1.4770	30



圖例：

●=觀察值

— =預測值

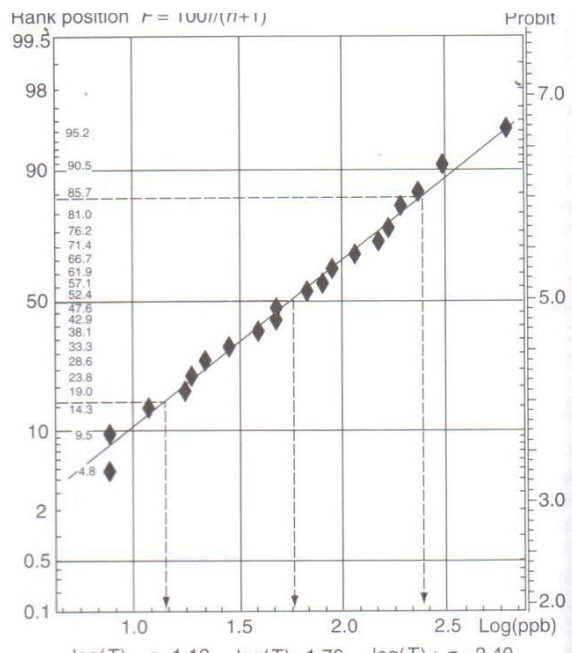
正確的比例

異戊酸在空氣的濃度 (ppb)

圖 8.4 使用 “SAS® PROC NLIN” 和邏輯法，將劑量-反應曲線配適到表 8.2 中的數據。估計值 $T = 1.954$ 是以 $\log$ (ppb) 表達的 13 位組成員閾值濃度。 P ，比例正確； X ，濃度； P_c ，比例正確，0.0-1.0； P_d 比例區分值=機率的%以上。

表 8.2 以 ASTM Intermediate Method E1 432 進行嗅覺閾值對空氣中異戊酸的測定

濃度呈現 (ppb)	小組成員					
	2	11	13	15	18	19
結果舉例，顯示 6 小組成員：正確的實驗次數（六次）						
640				6	6	5
320			6	5	6	4
160	6	6	5	4	2	3
80	5	4	4	2	2	2
40	4	4	2	3	3	1
20	6	0	2	2	1	2
10	5	3				
5.3	3					
2.5	2					
轉換為正確的比例= C/N 其中 C 是上述的數目，N = 6						
640				1.000	1.000	0.833
320			1.000	0.833	1.000	0.667
160	1.000	1.000	0.833	0.667	0.333	0.500
80	0.833	0.667	0.667	0.333	0.333	0.333
40	0.667	0.667	0.333	0.500	0.500	0.167
20	1.000	0.000	0.333	0.333	0.167	0.333
10	0.833	0.500				



排序位置 $F = 100i/(n+1)$

Probit

Log(ppb)

$\log(T) - \sigma = 1.12$

$\log(T) = 1.76$

$\log(T) + \sigma = 2.40$

圖 8.5 表 8.2 的 20 名品評員的排序概率圖。結果直線可通過點繪製；因此，品評員通常以 $\log(T) = 1.76$ ($T = 58$ ppb) 分佈； $\log(T) - \sigma = 1.12$ (13 ppb)； $\log(T) + \sigma = 2.40$ (255 ppb)。

組閾值：由排序概率圖獲得該組閾值 T，如圖 8.5。結果是 $\log(T) = 1.76$; $T = 58$ ppb。可替代地，以個體閾值的幾何平均數計算 T：
 $\log(7) = (0.84 + 0.84 + 1.04 + \dots + 2.40 + 2.52 + 2.82)/20 = 36.06/20 = 1.753$; $T = 56.6$ ppb。

8.4 敏感度測試(sensitivity test)

試驗分析型包括敏感度、差異性和描述性等測試。**敏感度測試(sensitivity test)**是分析型測試的一種，是指感覺器官對刺激的感覺、識別和分析能力。敏感性會因為個體生理狀況不同而有不同能力，人對氣體最為敏感，然後是味道，在味道中則以苦味敏感，對甜味最不敏感（敏感度：苦>酸>鹹>甜）。感覺的敏感性因人而異，受先天和後天因素的影響。人的某些感覺可以通過訓練或強化獲得特別的發展，即增大敏感性。在感官分析中，品評員的選擇，實際上主要就是對候選品評員感覺敏感性的檢測。表 8.3 為五種基本味道感官功能測試的建議濃度。

表 8.3 及 8.4 為基本味道感官功能測試的建議濃度

表 8.3 五種基本味道感官功能測試的儲備液建議濃度

基本味道	參考物質	濃度(g/L)
酸	DL 酒石酸(結晶) M=15.01	2
	檸檬酸(一水化合物結晶) M=210.1	1
苦	鹽酸奎寧(二水化合物) M=196.9	0.020
	咖啡因(一水化合物結晶) M=212.12	0.200
鹹	無水氯化鈉 M=58.46	6
甜	蔗糖 M=34.23	32
鮮	味精 M=187.13	2

味覺的敏感度測試可使用幾何平均數稀釋五種已知濃度的酸、甜、苦、鹹和鮮味來進行測試。依照濃度漸增的順序給品評員測試，如表 8.5。

表 8.4 以幾何平均數稀釋五種基本味道測試味覺敏感度的建議濃度

稀 釋 液	成分		試驗溶液濃度(g/L)						鮮
	儲 備 液 (mL)	水 (mL)	酸		苦		鹹	甜	
			酒 石 酸	檸 檬 酸	鹽 酸 奎 寧	咖 啡 因	氯 化 鈉	蔗 糖	
G6	500	稀 釋 至 1000	1	0.5	0.010	0.100	3	16	2
G5	250		0.5	0.25	0.005	0.050	1.5	8	1
G4	125		0.25	0.125	0.0025	0.025	0.75	4	0.5
G3	62		0.12	0.062	0.0012	0.012	0.37	2	0.25
G2	31		0.06	0.030	0.0006	0.006	0.18	1	0.13

第九章 品評員的選擇及訓練

9 Selection and Training of Panel Members

所謂品評組是指以感官檢查為目的的選定的具有特定資格人的集團。品評組成員稱為品評員。把參加品評組所要求的特定資格稱為品評員組成資格。

如前所述，感官分析根據不同的目的，可以分為兩大類，即分析型感官分析與偏愛型感官分析。因此，從事分析工作的品評組也相應地被分為分析型品評組和偏愛型品評組兩類。

分析型品評組的任務是檢出樣品與標準品之間，或樣品與樣品之間的差異，以及差異的程度和客觀品評樣品特性等工作。分析型品評員有無感官分析的經驗，或接受訓練的程度，會對分析結果產生很大影響。經驗是指非計畫的實際從事感官分析的知識累積；訓練是為了提高品評員的能力而進行系統性的有計畫的培訓。因此，分析型品評員按其能力，一般可分為初級品評員，訓練品評員，專家品評員三級。

嗜好型品評組僅僅是敘述其個人的喜好，這種判斷屬於感情的領域，是人的主觀品評。這類品評不需專門訓練。

9.1 簡介

本章部分是基於 ASTM 的專業技術出版物 758:選擇品評小組成員和培訓的準則(1981),及選擇、評估、培訓人員的 ISO 指南(1993)。品評小組的發展需要仔細地規劃，包括思考小組成員的本身需求、公司內部組織的支持和管理，品評候選人的可配合度和興趣，篩選訓練樣品和參考的需求，以及品評小室的可用性和條件。在食品、香水、化妝品等行業，品評小組是公司在研發和品質控制上最重要的工具。品評小組發展的成功或失敗，決定於用來選擇和培訓小組成員的嚴格標準和程序。

小組的選擇及培訓的條件取決於任何給定的感官問題或狀況的計畫目標。過去，唯一的標準常常是一個或多個基本味覺的低閾值 (ISO 1979)。當今的感官分析師則利用多種條件，針對所要研究的特性，選擇對應的培訓方案和小組成員。味覺或視覺測試只是其中一種，更重要的是描述能力，如描述「回甘」或「韻味」等特殊感官特性的印象。

本章介紹了設立品評小組的特定步驟，以及如何監督、激勵、訓練小組成員的具體程序。本章不適用於消費者測試(參見第 12 章)，它使用一般消費者來代表此產品的目標客群。雖然本文本中多使用「開發」、「生產」、「銷售」、「產品」等字眼，或是存在「高層管理人員」、「中階管理人員」和報酬結構的商業語言，所述及的系統可以簡易地透過修改以適應其他類型組織的需求，如大學、醫院、民用或軍事服務組織等。

9.2 品評小組的發展

在進行挑選和培訓品評小組之前，感官分析師必須在組織中建立一些要求，並且得到組織承諾：願意花費必要的時間和金錢來開發感官工具。對高層管理人員或計畫小組(R&D和QA / QC)來說，這些承諾來自於他們必需了解要基於可靠的感官分析數據才能做出決策，如：總體差異和屬性差異(差異品評小組)或產品標準的描述、產品隨時間的改變、配料或加工的微調、消費者問卷的建立和解讀。感官分析師另外必須定義開發和維護品評小組所需的資源。

9.2.1 人員

所需的資源列表為：(1)擁有一個足夠大的品評候選員資料庫，使我們得以針對研究計畫決定可選擇的品評員(見附錄 9.2 A 品評員候選人可能的來源)，(2)由一位品評行政職員實施篩選、培訓和維護程序，包括品評小組的領導人和技術人員，和(3)由合格的人員進行訓練。理想狀況下，品評小組的成員應來自組織內部，因為他們可能就來自樣品製備的單位(例如：研發中心或工廠)。一位好的領導人能夠有效地掌握產品開發人員或其他品評小組與客戶端之間的溝通。領導人與其品評小組一同工作，以確保品評小組對其所研究的屬性和量度(scales)有清楚的認識，並且能夠將品評小組的數據轉化為可使用的數據。一個成功的領導人也需(1)具有感官屬性的知識；(2)具有良好的團體工作能力；(3)能夠傾聽小組的意見和提供協助；(4)有創新的警惕感；(5)耐心。小組的領導人可以來自品評小組本身。如果是這樣的情況下，該領導人應另外接受領導人訓練，以妥善管理品評小組和與研究小組間的溝通，確保提供給產品開發者和其他科學家的數據是可靠、有效和有用的。如果品評小組很大，可能需要一位技術人員，負責小組內所有樣品採購、製備和呈現，以及完成計畫草案和數據輸出的所有必要的文件。

有些公司會選擇在不同的地方進行品評測試，例如在另一家公司中。為了減少實驗室的職員，許多公司也會選擇從當地社區招募

品評小組成員，而不是雇用一堆化學家和實驗室的支援人員待命。外來的人員可能可以每週工作更多小時，且費用可能會更便宜，或可以進行時間更長的測試。相對地，外部人員的主要缺點是，公司通常需要花更多的時間和精力訓練品評小組的工作技術，並且它們不具有內部員工本有的安全意識。

品評小組候選人和管理人員必須提前理解測試所需的時間（人時< personnel hours >）和特定品評小組的選擇和訓練。專家、技術人員和組長所需要的小時數應該在開發過程之前提交評估並且確認被接受。而品評小組的培訓人員往往是指定小組中最具有經驗，或是由受過特定培訓的品評員擔當，以訓練即將被培訓的品評員。

9.2.2 設備

在發展品評小組之前，必須先定義篩選、培訓品評小組的工作區域。培訓室和品評小組測試設備（品評小室或圓桌、會議室等）必須有適當的環境控制（見第3章），具有足夠的大小，以確保容納小組成員處理所有預計的測試產品，並且靠近產品準備區和小組人員辦公室。

9.2.3 數據的收集和處理

這是另一個需要被定義的資源：收集和處理品評小組產生的數據所需要的人員、硬體和軟體。例如使用電腦軟體與公司的伺服器產生的數據應該在開始累積在感官分析師的辦公桌上時就即時處理。產生這些數據和使用的特定方式（即數據、頻率、數據水準、類型，線性，估計幅度等）、屬性數量、重複再現次數以及統計分析都需要符合數據收集和處理的需求。

9.2.4 預估成本

在高層管理人員和專案小組認知到需要有一個品評小組時，他們要估計開發所需的費用、成本和收益，也可能從商業和投資的角度進行評估。這個階段是不可或缺的，以確保該品評小組的整體開發過程是基於管理人員的完整的理解及支持。在管理人員和專案小組就位後，感官分析人員可以開始計畫所有人員（包括小組成員和職員）需要被滿足的需求、設施和處理數據的支持設施。管理人員可通過公告、信件或研討會，與員工交流其對感官測試的參與和支持。當需求已透過這些方式吸引大眾的注意後，此感官評估的計劃和雇用小組成員的重要性將受到大眾的重視，因此管理人員是此系統運行良好的關鍵。如果參與感官測試的任務沒有受到中上層管理

人員的重視，品評小組隨後將會發現招聘任務異常艱鉅，且極可能因參與測試的新進人員不足而使得往後的發展計畫受阻。

經與組織溝通而得到管理人員的支持之後，並已確認了品評小組的設施和人員，感官分析人員可以開始透過告示、問卷調查或個人接觸來找尋潛在的小組成員。必須明確地傳達給候選人明白：品評測試需付出的時間和條件。一般要求包括：對測試程序的興趣、可付出的時間（約 80%）、準時、良好的健康（無過敏或影響參與的健康問題）、表達能力及在測試程序中沒有反感的產品。其他具體的標準列於第 9.3 和第 9.4 節中。

9.3 各種測試的人員選擇和培訓

品評員應依試驗的目的，方法，經驗與訓練層次等而區別為下列三大類五種基本型：

1. 消費者型(Consumer)

是產品使用者的代表者，此等人士可來自社會各階層，以喜歡或不喜歡，喜好程度及接受性作為評判產品的標準。

2. 試驗室(Laborator)

(1)無經驗型(Inexperienced)

是試驗階段之嗜好性試驗品評員。不需經過篩選、訓練，尤有興趣參加品評工作者所組成，輪流執行品評工作。一般試驗最少需 16 至 20 位品評者，通常為 30 位以上較洽當。

(2)有經驗型(Experienced)

是可從事差異試驗之品評員，由公司或團體中對品評工作有興趣之全體成員篩選試驗合格，具分辨產品差異能力者所組成。差異試驗(Difference Test)品評員人數一般為 10 人，但各個試驗室依其特殊需要而定人數。但切忌使用不合格的人員充數。一個剛建立的品評小組應維持所有合格人員至少 20 人，以應不時之需，且每次試驗至少由 5 位品評員執行。

(1)訓練型(Trained)

是描述分析品質特性之品評員。由上項之篩選合格人員，經進一步的篩選試驗，選出具有描述產品感官品質特性及差異情

形者經訓練純熟後所組成。

3. 專家型(Expert)

如品茗專家、品酒師傅。從事品質管制工作或評估產品特定屬性與記憶中預訂標準的差異。此種專家不易培養，需積數年的工作經驗始具評估能力。一般試驗室品評組織的建立不包括專家及消費者型。

9.3.1 人員選擇

假設早期招聘程序雇用了一組無明顯缺陷的候選人，例如常調整工作時間表或健康問題，感官分析師現在必須制訂一套針對不同測試篩選候選人的測試，同時儘快淘汰不適合無差別測試的候選人。這種篩選測試應該使用正在研究中的產品，並且使用將來感官測試時欲使用的品評方法。因此，它們必須依照以下的描述分類，而不是直接進行測試。篩選測試的目的是確定候選人之間的能力差異：

(1) 鑑別（和描述，如果是使用屬性差異測試法時）產品中的特性差異，和(2) 區分（和使用屬性差異測試的一度量進行描述）產品的特性強度或強度的差異。

在每章節的結尾會提供建議的評估方法。雖然具有較高成功率的候選人就整體而言是令人滿意的，但感官分析師應該進一步考慮如何選擇目前績效普通但有潛力的候選品評員。

9.3.1.1 匹配測試(Matching Tests)

匹配測試是用來確定候選人的區分能力（並進一步描述，如果測試中有要求的話），區分一些遠高於閾值強度的刺激之間的差異。初始先使用一組四到六個已編碼的樣品使候選品評員熟悉，但不進一步說明產品。接著隨機呈現已編碼的八至十個樣品，其中的幾個樣品與第一次呈現的樣品是相同的。請測試人員從第二次樣品中區分出那些是第一次呈現過的樣品，並以第一次樣品相對應的編碼來標示其編碼。

表 9.1 包含適合進行匹配測試的樣品清單。這些很可能是水中常見的物質、常見的芳香劑、具不同油系的乳液、具不同顏色的產品，類似組合物但不同重量的織物。另外應當小心避免交叉影響的效果，例如：樣品不應過於刺激。表 9.2 顯示高於閾值水準且未稀釋的芳香氣味清單。

表 9.1 匹配測試的建議樣品

味覺及化學感官因子		
味覺	刺激(克/升) ^a	濃度
甜	蔗糖	20
酸	酒石酸	0.5
苦	咖啡因	1.0
鹹	氯化鈉	2.0
澀	明礬	10
香味，香氣，氣味 ^b		
香味描述	刺激	
薄荷、薄荷味	薄荷油	
八角、茴香腦、甘草	茴香油	
杏仁、櫻桃	杏仁苯甲醛，苦杏仁油	
橘子、橘皮	橘子油	
花	芳樟醇	
薑	薑油	
茉莉花	茉莉花-74-D- 10%	
綠色	cis-3 - Hexenol	
香草	香草精	
肉桂	肉桂醛，肉桂油	
丁香，牙醫診所	丁香酚，丁香油	
冬青	水楊酸甲酯，冬青油	

^a 在無味無臭的室溫水中進行

^b 將香味吸附紙沾附有氣味的物品，加罩覆蓋 30 分鐘曬乾後，放在有緊閉蓋子中的廣口瓶中

表 9.2 香氣匹配測試的建議材料

第一次測試	第二次測試		描述字眼 ^a
079			
318			
992			
467			
134			
723			
花香	薄荷	香草	冬青
綠色	肉桂	生薑	丁香
茉莉	橙	櫻桃，杏仁	八角/甘草

注：說明：在嗅探第一組香氣時，每個樣品皆須有間隔時間讓品評員休息。接著嗅探第二組香料，並在第二組樣品中找到對應於第一組每個樣品的編碼。在第二次測試的欄位紀錄下對應於第一組的編碼。並從下方描述中選擇最適應的香氣描述。

^a 描述字眼的清單，可以在記錄表的底部進行說明。如果候選品評員將參加屬性差異測試，則選擇和使用描述字眼的能力應該再進行確認。

9.3.1.2 檢測/辨別測試(Detection/Discrimination Tests)

這項選擇測試是用來確定候選品評員評定同類產品，但成分或加工有差異的檢測能力。

呈現一系列的三角測試給候選品評員，範圍從容易到適度困難。也可使用二三點測試法。表 9.3 列出了一些常見的香味標準及其可使用量。修正樣品(“Doctored” samples)如模仿常用香味物質和異味的啤酒也可以使用。安排有經驗的品酒師進行初步測試，以確定測試系列的樣品為最佳狀況，同時控制刺激的量，以得到適當的檢測。使用標準的三角測試或兩三點測試評分表。如果需要，可以使用連續的三角測試（第 6 章）來決定是否接受此品評候選人。然而正如前面提到的，不要太依賴味覺的敏銳度。

表 9.3

用於檢測測試的建議材料

物質	濃度(公克/公升) ^a	
咖啡因	0.2 ^b	0.4 ^c
酒石酸	0.4 ^b	0.8 ^c
蔗糖	7.0 ^b	14.0 ^c
γ 癸內酯(γ -Decalactone)	0.002 ^b	0.004 ^c

^a 加入到無味無臭的水中物質的量。^b 3 倍閾值水準。^c 6 倍閾值水準。

9.3.1.3 強度的排序/評級測試(Ranking/Rating Tests for Intensity)

這些測試是用來確定候選人在一給定的屬性強度來鑑別分級水準。如果評級法是測試小組成員最終將使用的方法，請候選品評員將樣品進行評級，否則使用排序法（第七章）。以隨機順序呈現一系列樣品，其中存在不同層次的屬性參數，範圍足夠涵蓋正在研究的產品屬性。根據既定的屬性（甜度、油性、硬度、表面光滑度等）的水準要求候選入排序樣品（或使用規定的量度評分）。參見表 9.4 推薦的材料。典型的評分顯示於表 9.5 和 9.6。選擇測試的過程可以使用多個屬性的排序/評級測試，特別是如果最終品評小組將需要進行多種屬性的測試，例如：顏色、視覺、表面油性、硬度和表面光滑度。

表 9.4 排序/評級測試的建議材料

味道	感官刺激	濃度			
酸	檸檬酸/水，公克/公升	0.25	0.5	1.0	1.5

甜	蔗糖/水，公克/公升	10	20	50	100
苦	咖啡因/水，公克/公升	0.3	0.6	1.3	2.6
鹹	氯化物/水，公克/公升	1.0	2.0	5.0	10
風味					
酒精	3-甲基/水，毫克/公升	10	30	80	180
質感					
硬度	奶油奶酪 ^a 、一個美國奶酪、 ^a 花生、胡蘿蔔片 ^a				
脆度	玉米鬆餅、 ^a 全麥餅乾、芬蘭脆麵包、美國 Lifesaver 硬糖果				

^a 1/4 英寸厚。

表 9.5 強度的排序測試記錄表

	編碼
最不鹹	_____

最鹹	_____

按照鹹度排序樣品。

表 9.6 強度的評級測試記錄表

編號	
463	無_____強
318	無_____強

941	無_____強
502	無_____強

使用線性量度的度量評級每個鹹味的強度。

9.3.1.4 篩選測試結果的解讀

匹配測試：拒絕在匹配測試中正確得分低於 75% 的候選品評員。拒絕在屬性描述測試中得分低於 60% 的候選品評員。

檢測/區別測試(Detection/discrimination tests)：當使用三角測試法時，拒絕在「中等難度」(3 倍閾值) 正確率低於 40% 的候選品評員，在「容易」的測試中 (6 倍閾值) 正確率低於 60% 的候選品評員。當使用兩三點測試時，拒絕在「中等難度」測試中正確率低於 60% 的候選品評員，在「容易」的測試中 (6 倍閾值) 正確率低於 75% 的候選品評員。此外皆如第六章描述的連續測試步驟。

排序/評級測試(Ranking/rating tests)：接受正確排序樣品的候選品評員，或是僅相鄰一對排序不正確的。在評級的狀況下，使用相同的評級標準並期望候選人在刺激範圍更大時使用更大量度的評級。

9.3.1.5 不同測試方法所需品評員人數

進行感官分析試驗時，選擇不同測試方法所需品評員人數是不同的。實際工作中所需品評員的種類和人數如下：

二點測試法	初級品評員	40 名以上
	測試品評員	20 名以上
	專家品評員	6 名以上
三點測試法	初級品評員	25 名以上
	測試品評員	15 名以上
	專家品評員	6 名以上
二三點測試法	初級品評員	20 名以上

五中取二測試法	測試品評員	10 名以上
“A” “非 A 測試法	初級品評員	30 名以上
	測試品評員	20 名以上
選擇法	初級品評員	80 名以上
	測試品評員	40 名以上
排序法	初級品評員	10 名以上
	測試品評員	5 名以上
	專家品評員	2 名以上
	消費者品評員	100 名以上
配偶法	初級品評員	60 名以上
	測試品評員	30 名以上
	消費者品評員	120 名以上
評估法		
a. 為確定指標強度	初級品評員	20 名以上
	測試品評員	5 名以上
	專家品評員	1 名以上
b. 為確定嗜好的程度	初級品評員	50 名以上

(兩種樣品)

初級品評員

100 名以上

(三種以上樣品)

9.3.2 培訓

為確保建立品評小組成員的專業態度，在所有感官分析的進行皆須在專業的感官設施中進行培訓。

指導品評員如何為即將要進行的感官分析做預先準備，例如：不要使用有香味的化妝品，並在測試前 30 分鐘避免接觸食物或香氣；如何準備皮膚和手織物的觸感測試；以及哪些影響測試模式的過敏反應發生時需要通知品評小組領導人；或任何一天有品評員感冒、頭痛、睡眠不足時將如何向品評小組請假。

從一開始到測試進行時都要時時教導品評員處理樣品的正確程序。強調堅持規定的測試程序，閱讀所有的指引和嚴格遵守這些規則的重要性。

驗證消除或減少感官疲乏的方法，例如：淺嗅香味後離開幾十秒再測試下個樣品。強調不顧個人喜好和集中在檢測差異的重要性。

從正在研究產品的現行樣品中，挑選容易被感知的屬性差異。首先集中於幫助小組成員了解計畫的範圍，並獲得自信。重複使用略小但仍容易被感知的差異進行樣品測試，讓品評小組通過重複學習，直到完全有信心達成。

在屬性差異檢測的部分，仔細介紹各個屬性給小組成員，以及用來描述的詞彙(描述詞)，指示強度的度量法等。呈現一系列代表各個屬性強度差異的樣品。

持續以定期舉辦測試來培訓在職的品評員。偶爾使用模擬「異味」的訓練樣品，或主要的產品差異，以保持品評小組的專注力。

注意某些小組成員可能會因為混淆樣品或是失去對測試的興趣而有了態度或行為的轉變，或因其它問題而分心。感官測試的歷史中充滿了許多不可思議的結果，這些結果可能僅來自小組成員在測試過程中因迷失而不能參與，和“測試儀器”偵測失敗。

9.4 描述性測試的小組成員的選擇及培訓

9.4.1 選擇描述性測試小組成員

招募小組成員的過程是創造一個成功的描述性品評小組（附件 9.2A）的關鍵因素。品評小組使用於描述各個屬性強度的詞彙，其小組成員必須能夠同時使用術語和量化表達方式以充分地描述產品，儘管小組成員在此測試中被視為是一個鑑別器，此小組成員是否有思考和溝通的能力仍然是重要的。

品評小組的發展中的第一步即是儘可能招募有興趣的潛在小組成員。告知他們研究的相關細節，研究能夠給他們帶來哪些益處（金錢、知識、專業技能等）。透過張貼報紙廣告，在廣播電台做宣傳，或在他們最可能感興趣的美食、美景等報章雜誌中放置廣告。圖 9.1 是這樣一則廣告的例子。

9.4.2 描述型測試的小組成員選擇

為了描述性測試選擇小組成員時，該小組領導人或訓練人員應確定每個候選品評員的三個主要方面的能力：

Do you like food? Would you like
to know food more intimately?
Ever wanted to be a professional
“food taster?” Here’s your chance:



A major food company
in the Minneapolis
area is looking for
professional food tasters
to help in the design of
new products!



Details:

- 6-10 hours/week between 9am & 4pm M-F
 - Paid training: 4 8-hr days per month for 1st 5 months
- Minimum 2 year commitment
- Have a flexible schedule
- Enjoy critiquing food and flavors

Contact Lorraine or Jay at Kelly Technical
Services at (612)797-0771 to start your
professional food taster career now!



FIGURE 9.1

Example of a descriptive panel recruiting advertisement (Stoer, Rodriguez, and Civile 2002).

你喜歡食物嗎？你想知道食物更多的秘密？曾經想成為一名專業品評員的你，這裡是你的機會！

一間大型食品公司正在明尼阿波利斯地區尋找專業品評師，以幫助其設計新產品！

詳細聯繫方式：

- 每周 6-10 小時/上午 9 點至下午 4 點，頭五個月每月 48 小時的帶薪培訓。
- 至少 2 年的工作合約
- 靈活的工作時間
- 享受評判食物和口味

聯繫(612)797-0771 技術服務中心的 Lorraine 或 Jay，現在就開始你的食品試吃職業生涯！

圖 9.1 為描述品評測試的小組招聘廣告實例。

1. 對於每項調查中的感官特性（如香味、氣味、味覺、口感、手感或觸感等），具有發現其特性並感知強度差異的能力。
2. 具有使用（a）口頭描述感官特性（b）使用比例法判定不同強度的能力。
3. 抽象推理的能力，因為描述性分析必須依靠使用參考品，而特性必須快速地反應及聯想引用到其他產品。

除了針對以上描述的能力篩選小組成員外，品評小組領導人另外必須預審以下個人條件：

1. 有興趣全程參與嚴格的訓練、實踐以及未來進行的品評工作
2. 能夠參與 80% 以上各個階段的小組工作。當品評工作與家庭生活、工作負荷、旅行，甚至品評候選人的上司發生衝突時，最終皆可能會導致小組成員在訓練期間或之後脫隊，使得原本已經人少的 10 至 15 人品評小組因而失去一名品評專家。
3. 身體健康正常，如：
 - a. 針對食品、飲料、藥品或其他產品的香味及質感測試，品評候選人應不具備與這些感官評估測試相關的疾病，如：糖尿病、低血糖、高血壓、假牙、慢性感冒或鼻竇炎、或食物過敏。
 - b. 針對食品、飲料、個人護理產品、藥品，或家用產品香味分析，品評候選人不能有慢性感冒或鼻竇炎的情況。

C. 針對個人護理護膚產品、紡織品，或家用產品的觸覺測試，品評候選人不能有中樞神經系統疾病，或正在使用會降低神經敏感度的藥品。

檢測及描述差異的能力，應用抽象概念的能力，加上積極的態度和喜愛描述性分析任務的程度都可以通過一系列的測試得到進一步確定：

- 一組篩選問卷
- 一組視力測試
- 一組排序/評級測試
- 個人訪談

在篩選描述測試的品評員上投資時間和人力是明智的，與其進行詳盡的人員篩選過程，不要培訓不合適的候選人。

可以在 ASTM 特別技術出版物 758 (1981) 中找到針對三種描述法的（味覺、定量描述分析和質地）篩選標準列表。下面列出的標準是那些用於感官列譜描述分析(the spectrum Method of descriptive analysis)的人員篩選標準。這些可以應用在內部員工的審查，或是從當地社區招聘的情況。需要進行外部篩選的情況往往是由於固定時間內所需的量較大（每人每週 20~50 小時）。需另外使用篩選問卷來選擇可以用言語表達抽象概念的候選人。這減少了選擇擁有感官敏銳度，但無法獲得來自公司內部招募品評小組時的「技術」指導的外部小組成員的風險。

9.4.2.1 篩選問卷

要建立 15 人的品評小組，正常需使用問卷篩選 40 至 50 個候選人品評員，問卷範例如附錄 9.1 所示。附錄 9.1A 適用於觸感類的品評小組（觸感或織物手感）；附錄 9.1B 為口味型品評小組；附錄 9.1C 為口感類品評小組；附錄 9.1D 為香味型品評小組。附錄 9.1E 則是度量量度表，與任何前述的附錄 9.1 問卷共同使用。一般 40 至 50 位的候選人到最後約 20 至 30 人會通過此篩選，並繼續敏銳度測試。

9.4.2.2 敏銳度測試(Acuity Tests)

要通過這個階段，候選人應：

- 沒有因醫療或藥物而造成認知的限制。

- 可參加培訓課程。
- 正確且清楚地回答附錄 9.1A 至附錄 9.1D 篩選問卷 80%的問題。
- 在附錄 9.1E 的問卷中，分配度量評級(scalar ratings)與所有圖中的正確值距離區間 10%至 20%。

候選人應表現出的能力：

- 檢測並描述質感(qualitative sense)上的特徵。
- 檢測並描述數感(quantitative sense)的差異。

因此，儘管檢測測試（例如：在產品配方或加工方面使用變量(variations)的三角測試或兩三點測試法）可能會產生一組可以檢測小型產品變數的品評員，然而單單檢測仍不足以成為描述型測試的品評成員。為了獲得此資格，品評候選人必須能夠區分並充分描述測試產品的關鍵感官屬性，也必須證明其具備正確使用評級度量(rating scale)來形容強度差異的能力。

描述：該小組的教練提供了一系列代表產品中關鍵變數的樣品，並且選擇使用三角測試或二三點測試。樣品間的差異可能在於製作時間或烘烤溫度、成分的比例（50%，100%，150%）或包裝的差異等，皆可以用來檢測品評員的敏銳度。首先嘗試難易度低的樣品對，並依序增加難度。選擇三角測試中正確回答 50 至 60%機率的品評員，或兩三點測試 70%至 80%的正確率，這取決於每個測試的難易程度。

描述：提供一系列展示不同屬性特徵的產品（香味/香精油，幾何質感的物質，並要求候選人描述其感官印象。使用表 9.1 香味列表但不可提供描述列表供選擇。候選人必須使用其自己的話描述每種香味。這些可能包括化學術語（例如：肉桂醛）、常見的香味（例如：肉桂）或日常用語（例如：泡泡糖口味、彈牙）。候選人應能夠使用化學術語、專業術語或日常用語描述 80%的刺激，並至少應嘗試用非專門術語（例如：甜、棕色香料、熱香料）描述其他部分。

9.4.2.3 描述性分析的排序/評級篩選測試(Ranking/Rating Screening Tests for Descriptive Analysis)

通過了篩選測試和敏銳度測試後，品評候選人可以開始進行品評小組實際使用的產品和感官屬性的篩選測試。候選人使用必須選定關鍵屬性來排序或評級產品，使用未來品評小組將使用的技術。這些測試可輔以一系列加強某些屬性強度的樣品，如味道或風味（見表 9.4），或口感（附錄 11.2，質感 D 節，使用量度 5 是合適的，硬度標準包含起司奶酪= 1.0~硬糖= 14.5；或者量度為 10，包含從燕麥棒 2.0 至玉米片 14.0）。問卷如表 9.7 是合適的。對於某些觸感和織物手感的性質，使用附錄 11.2E 或 11.2F，或參照樣品可以從商業產品和實驗原型當中選擇，呈現一系列選定屬性，依序增加其強度。選擇可以將 80%樣品按照正確的量度評級所有屬性的候選

人。只允許將鄰近樣品的順序顛倒，並檢查候選人使用最多數的量度來檢查至少 50% 的所測試屬性。

9.4.2.4 個人面試

特別是針對描述型的品評小組，個人面試是需要確定候選人是否適合該團體的動能和分析方法。一般情況下是針對已通過篩選問卷和敏銳度測試的候選人，由品評小組教練或小組領導人單獨面試。面試的目的是確認該候選人對於品評工作的興趣，及其時間配合度、培訓對於該候選人各個工作階段的工作負載，其領導人或個人旅行規劃等的衝擊，或者是溝通技巧及總體個性。對於品評工作和描述型品評興趣不大的候選人應該被排除。敵對性高或膽小性格的人也必須排除，因為他們可能會造成其他小組成員的困擾。

表 9.7 用來篩選候選品評員的質地品評測試，包含兩種排序測試的紀錄表

質感描述型品評的篩選

1. 將產品放置牙齒中間，咬一次後評估硬度。將樣品依照硬度排序。

最不硬

最硬

2. 將產品放置牙齒中間，咬一次後評估脆度。將樣品依照脆度排序。

最不脆

最脆

9.4.2.5 模擬品評小組(Mock Panel)

有些公司藉由邀請候選品評人到一組「模擬品評小組」，要求他們評估兩個(含)以上產品來進一步評估並篩選候選品評員。呈現一系列產品給候選人，並要求其寫下看法(由小組領導人描述感官參數，例如：「這些餅乾的味道及口感」)。然後該小組領導人在一定時間內引導每個小組成員表達自己的看法及討論結果。觀察候選人是否能夠清楚地傳遞概念，並參與不同看法的討論，可有助於決定此候選人是否適合加入品評小組。

9.4.3 描述型測試的培訓

在任何系列的訓練中，最重要的是提供一個結構框架，以允許學生在驗證事實上有學習的根據，增長技術和信心。大多數的描述型品評小組的培訓計劃要求培訓 40 至 120 小時。所需的時間量取決於產品的複雜度(葡萄酒、啤酒、咖啡等比洗劑、霜劑或早餐穀物的評價需要更多的時間)，涵蓋的屬性數目(用於品質控制或儲存研究的簡短版描述技術，第 11 章，需要更少且更簡單的屬性)，以及有效性和可靠性的要求(一個更有經驗的品評小組可提供更大且更加詳細的再現性)。

9.4.3.1 開發術語

品評小組的負責人或教練與專案團隊必須在訓練的初始階段一起確定品評小組內使用的關鍵產品變數。專案團隊應該準備一個原型或收集市售產品做為參考的框架，儘可能表示許多不同的產品類別及可能遇到的屬性差異。品評小組首次引入影響每個產品的感知屬性的化學概念(嗅覺、味覺、化學感官因子)和物理原理(流變，幾何等)。有了這些概念和術語為基礎，專案小組隨後制定引用品評方法及術語的步驟及產品分類程序。

這個方法的運用實例有口感、葡萄酒、啤酒、雞肉、花生、牛肉、鯰魚等方面。一般來說，訓練的第一階段小組成員可能需要 15 - 20 小時理解屬於其研究類別(外觀、香味、口感等)等一些描述的屬性。第一階段的目的是為他們提供堅固的基礎，讓他們開始明顯感受在不同的產品類型上的不同特點。

9.4.3.2 描述量度法(Descriptive Scaling)的介紹

在訓練開始的 10-20 小時內可介紹量度法。藉由一組包含三到五種不同程度的屬性的產品或樣品，品評小組領導人可以同時加強感官特性及描述法，並通過在多個屬性表明不同強度(或量)的量度方法。附錄 11.2 提供幾個作為描述感官的類別的不同屬性強度的例子：味覺(香味、口味、感覺因素)，固體和半固體質感(Munoz, 1986)(硬度、黏著性、彈性等)，觸感(ASTM 1997; Civille 和 Dus, 1991)(濕度，滑度，油度等)，和織物手感(Civille 和 Dus, 1990)(滑順感、砂感、絨毛感等)。

在實踐作業中繼續使用強度參考量度是為了同時強化屬性和量度的使用能力，從而使品評小組在描述過程認識術語和數字（特性和強度）的定義及學習記錄任何分類中的產品。

9.4.3.3 初步實踐

在給定的產品類別中建立精確詞彙的發展通常是三個步驟的過程。在第一步驟中，提供所有原型品或產品特性的樣品給品評小組，作為參考的框架。從這參照框架中，每個品評員協助貢獻創造出一串長長的原始描述詞清單供品評小組未來使用。在第二階段中，因為初始的清單中包含許多重複的字詞，我們將重排並還原成一個全面的描述清單（它們能夠完全地描述所有產品類別），和謹慎的（最少化重疊字詞）工作列表。第三階段，也就是最後一個階段，包括選擇產品、原型品或是外部引用產品，用來代表所有選擇的詞彙的實例。

當品評小組具有掌握術語的把握和擁有了量度的大致理解後，品評小組教練或領導人可以提出一系列的樣品進行評估，一次一個，或是一次多個以代表在性質上（屬性）和數量上（強度）的廣泛差異。在這個持續 15 至 40 小時的早期開發階段，專案小組建立了基本的技能和信心。在不同的樣品中，小組成員得以看到詞彙和量度有效地描述和鑑別產品的差異，並幫助所有小組成員獲得作為個人和一小組的信心。

9.4.3.4 產品的微小差異

在專案(產品)團隊的幫助下，專案領導人收集各個產品代表中的微小差異，包括在生產變數和(或)產品的替補修改等變化。鼓勵品評小組改善評估術語與增修引用程序的定義，以滿足檢測和描述這些微小產品差異的需求。必須注意減少相同樣品之間的差異，因為處於培訓中的品評員往往將有差異的結果解釋成自己功力不足的原因。樣品的一致性有助於提升品評小組的信心。這個階段約占品評小組 10 到 15 個小時的時間。

9.4.3.5 最終練習

品評小組應繼續進行測試，並持續在培訓的最後階段描述數種產品（15 - 40 小時）。最終樣品與早期樣品相當不同，因為最終的產品測試應該接近該品評小組未來真實將測試的情況。在培訓計劃的所有五個階段，小組成員應在每次測試之後舉行會議討論結果，解決其中發生的問題及爭議，並另外要求審查性質或數量的參考資料。這種互動是發展品評術語、評估程序，並調整感官儀器的過程中必不可少的步驟。

9.5 品評小組的績效和激勵

任何好的測量工具都需要定期檢查，以確定其效能和一致性。在感官評估小組的概念中，個人以及作為一個整體的品評小組，也需要監測。品評小組是由擁有除了參與感官評估外還有其他責任和工作的成員所組成，因此我們有必要設法讓小組成員在長時間的產品測試過程中保持興趣和動機。

9.5.1 績效

對於差異和描述型的品評小組來說，感官分析師需要有衡量各小組成員的績效及其有效性和再現性的方法。有效性是指回答的正確性。在某些差異測試中，如在三角測試和二三點測試或定向屬性測試(directional attribute tests)中，分析員知道區分正確的答案（奇數樣品、參考編碼、較甜的樣品），且可以在一段時間內仍正確回答。回答正確的百分比可以每月定期計算，或在每兩個月的基礎上計算。加權計分也可以加入計算的基礎，對該小組成員每次參加測試的難度進行加權（Aust, 1984）。品評小組作為一個整體，其有效性可以藉由與其他品評小組的結果比較得知，或與其他感官測試機器的數據比較，或與已知變異的刺激比較，例如增加熱處理，加入化學品等。

可靠性(Reliability)，或者稱為再現性，可藉由複製測試，即使用重複樣品測試，或使用盲測等，來評估小組成員或者是整個品評小組，。

而描述性數據藉由統計的變異數分析(analysis of variance)來分析，小組成員的績效可以藉由分析每個部份的屬性來進行評估。在 QDA 中，品評小組人員可以使用不同的比例的來表達其對於同一樣品的感受。重要的是他們在各個樣品之間的相對差異，而不是絕對值。另外用於描述型品評的方法，如頻譜(Spectrum)，品評小組成員透過使用校正過的參考度量來評估相同的樣品。這種類型的描述性品評小組等同於需要定期校準檢查的分析儀器。這些方法，除了剛提到的 ASTM 準則，對於監測個人及「已校準的」品評小組的組合績效是合適的。品評小組需要監控的兩個方面是其準確度（偏差）和精確度（變異）。

偏差(Bias)—為了評估小組成員的能力是否符合「目標」，品評小組領導人可藉由其匹配(match)一對照品或參考品的屬性強度來確認小組成員的能力。目標或對照評級(control rating)的差異稱為偏差，其統計測量的方法定義為：

小組成員偏差， $d = \chi - \mu$ (9.1)

其中 d 為偏差， χ 是所觀察的小組成員的值， μ 是對照或目標屬性的值。

變異(Variability)－透過對一盲測或參考品的數次評估，該品評員的平均得分的變異採用品評成員的標準差的計算方法如下：

品評員標準差， $S = \sqrt{\Sigma} \quad (9.2)$ (公式看課本)

一個好的品評員應同時擁有低偏差和低變異。偏差的公式可以藉由去除符號修改；這將產生絕對的偏差，計算公式為

小組成員的偏差， $|d| = |x - \mu| \quad (9.3)$

低絕對偏差值是理想的，大的正偏差和負偏差不會因此相互抵消。品評員的統計結果應隨著時間繪製紀錄，以確定誰需要重新培訓或校準。

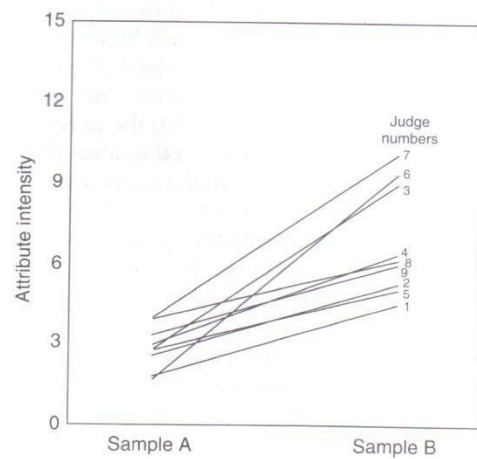
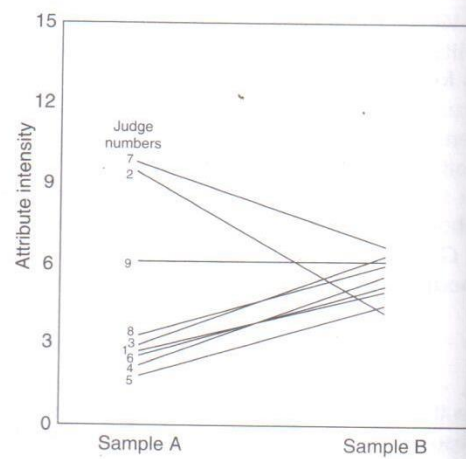
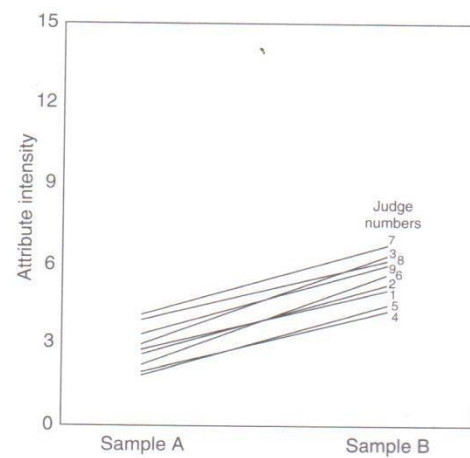
當變異數的裂區分析被用於描述性數據分析，依樣品判斷的相互作用(judge-by-sample interaction)是結果的一部分。當這種相互作用是顯著的，就要看點線圖來確定來源。圖 9.2 呈現三個隨樣品鑑定的相互作用。在每個圖中，每一行代表了該品評員對於兩個樣品的平均評級分數(average ratings)。在圖 (A) 中，判斷該樣品的相互作用是不顯著。所有品評員在相同的方向和相同的相對強度上評級樣品。因此，線在相同的方向上，並且斜度相似。第二幅圖 (B) 表示樣品具有極端相互作用的例子：品評員在一些樣品上的評分截然不同。因此，線呈現不同的方向，並具有不同的斜率。第三幅圖 (C) 顯現了某幾個品評員判定斜率和其他品評員不同。在這種情況下，雖然依樣品判斷的相互作用具有統計顯著，但問題較不極端。它只是證明品評員利用量度上有些微的不同，而不是完全相反的評分。而圖 B. 的情況，此相互作用則表明品評員需要更多的培訓，更經常地練習參考度量，或檢查術語清單。

9.5.2 品評員的維持、回饋、獎勵和激勵

小組成員的其中一個動機主要來源就是感覺工作有意義。當一個感官評估測試結束後，各個小組成員應該收到一封信，告知計畫和測試目標、測試結果，並告知通過此感官測試結果的貢獻產品如何被改變或決定。每次測試後的即時反饋也能夠讓各個小組成員的知道其績效，有些計畫負責人擔心過低的成功機率可能導致品評員洩氣或恐懼（三角測試的正確回應往往低於 50 %），這已被證明毫

無根據。品評員的確了解樣品的複雜性、測試的困難皆會影響正確機率。每個人都希望了解測試結果，並且根據過去的績效學習。強烈建議在描述型測試之後舉行小組會議。透過固定的品評小組在測試後互動，以不斷地精鍊術語及定義程序，我們可以及時地收集小組成員的反饋。

給品評員的反饋可以追溯到其測試相同產品的三至五個評估結果，讓他們可以觀察個人的績效。表 9.8 顯示了在給定樣品上每個小組成員（數字）在每個屬性（字母）上評分平均值和標準差，以及該品評小組的平均值和標準差。該品評員因此能夠藉由比較小組及其個人的差異來認同個人的績效。此外，品評員的標準差提供了她在該屬性上可靠性（變異）的資訊。應該定期提供品評員在三到五次的測試中，兩到三個樣品的資訊，例如：每 3 至 4 個月。品評員的相互作用曲線圖，如圖 9.2 中所示的那些，也可提供給小組成員，以顯示所有的品評員之間是否在某樣品上達到一致的評分，或是其相對其他小組成員的評分績效。



屬性強度
樣品 A

樣品 B
品評員數目

屬性強度
樣品 A
樣品 B
品評員數目

屬性強度
樣品 A
樣品 B
品評員數目

圖 9.2 品評員和樣品的相互作用圖（見正文）。

除了從反饋產生的心理報酬外，品評員在獎勵制度中也將會更積極地展現，並更踴躍地參加。能夠提供的成就證書有：

- 品評測試的高出席率
- 品評測試的優良績效

表 9.8 品評小組績效總結

屬性	品評員						品評組平均
	1	2	3	4	5...	14	

	/標準差						
A	7.5/0.2 ^a	7.0/2	6.8/2	6.9/1	7.9/2.5	6.2/1.9	6.9/0.5 ^b
B	4.2/1.4	4.8/2	5.5/1.6	5.0/0	4.2/1.2	4.6/1.6	4.8/0.4
C	1.4/1	3/1.3	1.5/1.2	1.0/0.9	1.1/0.8	3/1.3	1.8/0.8
D	9.0/0.5	8.0/0.7	9.0/1.0	6.4/1.2	12/1.1	10/1.3	9.4/1.6
E	4.0/0.7	4.2/0.8	3.5/1	1.9/1.2	4.4/0.9	3.8/2	3.9/1.1

14 名品評員在 3 週內界於許多樣品之間評估某一相同的樣品。在 A 屬性上，品評小組的評分均值為 6.9，在 14 位小組成員中的標準差為 0.5 或 7.2%，顯示在該屬性的評定上，小組成員之間顯示足夠的一致性。品評員 5 在評定屬性 A 和屬性 E 上的平均遠高於品評小組，並在屬性 A 上呈現出略高的標準差。

^a 品評員均值/標準差

^b 品評小組總均值/總標準差。

- 改進的績效
- 培訓計劃的完成
- 特殊計畫的完成

刺激品評小組且與其保持良好的溝通是有價值的。短期獎勵，如零食、公司產品的兌換卷和較大獎品的抽獎券可以每天提供給品評員。而長期獎勵方面，郊遊、提供小組成員午晚餐等可以讓品評員感覺公司經常贊助品評測試。如果可能的話，也可以透過公司或計劃管理人員描述品評結果如何被使用。在公司內部報紙或當地社區媒體上報導或表揚當前小組成員，也能夠刺激潛在的候選人參加。品評員的職責就是發現樣品呈現的感官差異。這種發現的能力也可以藉由鼓勵品評員加強平時的感知能力而提高。一些以提高感官認識為目的的活動也可以激勵品評小組。這些活動讓品評員在享受樂趣的同時，進一步刺激大腦（附錄 9.2B）學習新的數據。品評小組也有瓦解的可能，如果品評小組的領導人沒有清楚地設定可接受的和不可接受的行為界限。儘早建立適用於品評小組的行為準則是個好主意。由品評員審核並書面簽署這些準則做為未來作業的操守守則（附錄 9.2C）。感官測試全程有管理人員的基本支持，以及品評員的積極參與是建立一個積極且具有高素質的品評團隊的關鍵因素。

附錄 9.1 篩選問卷

在篩選問卷中的每一個問題設計是為了使得品評小組領導人或教練從一大群候選人中，篩選出那些能夠清楚地口頭描述感受，並能夠數量化表達感受的程度。針對不同類型的品評小組（觸感、香味、口感、質感或香味）使用不同類別的篩選問卷以及附錄 9.1 E 的量度練習。

A. 觸感的（或織物手感）品評小組篩選問卷

來歷：

名稱：

地址： _____

電話（家庭和企業）： _____

從什麼團體或組織聽到這個計劃？ _____

時間：

1. 是否有任何工作日（周一至周五），您將定期無法參加測試？ _____

2. 在 6 月 1 日和 9 月 30 日之間，您打算放多少個星期的假？ _____

健康：

1. 是否具有以下的疾病？

中樞神經系統紊亂 _____

異常寒冷或溫暖的手 _____

皮疹 _____

手/手指老繭 _____

敏感皮膚 _____

刺痛手指 _____

2. 是否服用任何會影響感官的藥物，尤其是觸覺？ _____

一般：

1. 您的觸覺：(二選一)

低於平均水準_____

平均 _____

高於平均水準_____

2. 是否有任何直系親屬在紙廠、織物廠或紡織公司上班？ _____

或是市場調查、廣告公司？ _____

觸感測試

1. 什麼特點使這個毛巾讓你感覺會吸水？ _____

2. 什麼是厚，油性或油膩的薄膜？ _____

3. 當你的皮膚被擦上一層油膜，你的手指怎麼移動？

滑_____或拖動_____ (二選一)

4. 毛巾中的什麼屬性使得你感覺它是柔軟的？

5. 浴巾什麼樣的具體外觀特徵影響你對它的感受？ _____

6. 舉例幾樣粘膩的東西。 _____

7. 當你的皮膚感覺滋潤，有什麼其他字眼可以代替形容呢？ _____

8. 舉例幾樣粗糙的東西。 _____什麼使得它們粗糙？ _____

9. 簡單地說，你將如何界定「實感」？ _____

10. 在織物或紙製品中，什麼使你感覺生硬？ _____

11. 你會用什麼詞彙來形容乳液是薄或厚？ _____

12. 你覺得當你觸碰織物表面，你會感覺到什麼特點？ _____手背？ _____

B 香味型品評小組的篩選問卷

來歷：

名稱： _____

地址： _____

電話（家庭和企業）： _____

從什麼團體或組織聽到這個計劃？ _____

時間：

1. 是否有任何工作日（周一至周五），您將定期無法參加測試？ _____

2. 在 6 月 1 日和 9 月 30 日之間，您打算放多少個星期的假？ _____

健康：

1. 是否具有以下的疾病？

假牙 _____

糖尿病 _____

口腔或牙齦疾病 _____

低血糖 _____

食物過敏 _____

高血壓 _____

2. 你是否正在服用會影響感官的藥物，特別是味覺和嗅覺任何藥物？ _____

飲食習慣

1. 請問您目前是否有飲食限制？如果有，說明 _____

2. 最近一個月多常外食？ _____

3. 最近一個月多常吃速食？ _____

4. 最近一個月多常吃冷凍食品？ _____

5. 什麼是你喜歡的食物？ _____

6. 什麼是你不喜歡的食物？ _____

7. 哪些食物你不能吃？ _____

8. 那些食物你不喜歡吃？ _____

9. 您辨別氣味和口味的能力

	聞	味覺
高於平均水平	_____	_____
平均	_____	_____
低於平均水平	_____	_____

10. 你的直系親屬中是否有人在食品公司工作？ _____

11. 你的直系親屬中是否有人在廣告公司或行銷研究機構工作？ _____

味覺測驗

1. 如果食譜要求加入百里香但你並沒有，你會使用什麼替代呢？ _____

2. 什麼是味覺像酸奶的食物？ _____

3. 為什麼人們經常建議加入咖啡至肉汁使其口感豐富呢？ _____

4. 你會如何形容味覺和香氣之間的差異？ _____

5. 你會如何形容味覺和質感的差異？ _____

6. 形容磨碎的意大利起司（帕瑪森或羅馬諾）最好的一個或兩個詞？ _____

7. 描述美乃滋中不可或缺的一些味覺。 _____

8. 描述可樂中引人注目的一些味覺。 _____

9. 描述香腸中明顯的一些味覺。 _____

10. 描述在 Ritz 餅乾中明顯的一些味覺。 _____

口感型品評小組的篩選問卷

來歷：

名稱： _____

地址： _____

電話（家庭和企業）： _____

從什麼團體或組織聽到這個計劃？ _____

時間：

1. 是否有任何工作日（周一至周五），您將定期無法參加測試？ _____

2. 在 6 月 1 日和 9 月 30 日之間，您打算放多少個星期的假？ _____

健康：

1. 是否具有以下的疾病？

假牙 _____

糖尿病 _____

口腔或牙齦疾病 _____

低血糖 _____

食物過敏 _____

高血壓 _____

2. 你是否正在服用影響感官的藥物，特別是味覺和嗅覺？ _____

飲食習慣

1. 請問您目前是否有飲食限制？如果有，說明 _____

2. 最近一個月多常外食？ _____
3. 最近一個月多常吃速食？ _____
4. 最近一個月多常吃冷凍食品？ _____
5. 什麼是你喜歡的食物？ _____
6. 什麼是你不喜歡的食物？ _____
7. 哪些食物你不能吃？ _____
8. 那些食物你不喜歡吃？ _____
9. 你對食物結構特徵的敏感性 _____
比平均更好 _____
平均 _____
比平均更糟 _____
10. 你的直系親屬中是否有人在食品公司工作？ _____
11. 你的直系親屬中是否有人在廣告公司或行銷研究機構工作？ _____

口感測驗

1. 你怎樣描述香味和質感的差異？ _____
2. 描述在一般食品中的質感屬性。 _____
3. 描述在一般食品中可發現的顆粒。 _____
4. 描述當我們咀嚼食物食顯而易見的屬性。 _____
5. 描述酥和脆之間的差異。 _____
6. 什麼是薯片的質地口感？ _____
7. 什麼是花生醬的質地口感？ _____
8. 什麼是燕麥片的質地口感？ _____

9. 什麼是麵包的質地口感？ _____

10. 什麼類型的產品質地口感很重要？ _____

D 氣味型品評小組的篩選問卷

來歷

名稱： _____

地址： _____

電話（家庭和企業）： _____

從什麼團體或組織聽到這個計劃？ _____

時間：

1. 是否有任何工作日（周一至周五），您將定期無法參加測試？ _____

2. 在6月1日和9月30日之間，您打算放多少個星期的假？ _____

健康：

1. 是否具有以下的疾病？

鼻疾 _____

低血糖 _____

過敏 _____

經常性感冒或鼻竇炎 _____

2. 你是否正在服用會影響感官的任何藥物，尤其是氣味？ _____

日常的生活習慣

1. a. 您是否經常噴香水或鬚後水/古龍水？ _____

- b. 如果是，什麼品牌？ _____
2. a. 你喜歡有香味或無香味的肥皂、洗滌劑、織物柔軟劑？ _____
b. 為什麼？ _____
3. 哪些產品有你喜歡的香味呢？類型或品牌 _____
4. 哪些產品有你不喜歡的香味呢？類型或品牌 _____
5. a. 列舉一些讓你感覺不舒服氣味。 _____
b. 它們為何讓您感到不舒服？ _____
6. 什麼樣的氣味、香味或臭氣最吸引你？ _____
7. 你辨別氣味的能力是，高於平均水準 _____
平均 _____ 比平均糟 _____
8. 你的直系親屬中是否有人在肥皂、食品、個人用品公司或廣告公司工作？ _____ 如果有，是哪間公司？ _____
9. 品評員在訓練中不能夠使用過重的香水/古龍水，也能在品評測試之前一小時抽菸。
如果你選擇一個品評員，你是否可以做到以上要求？ _____

氣味測驗

1. 如果香水類型是「花香」，還可以使用其他什麼詞來形容呢？ _____
2. 哪些產品有草藥的氣味？ _____
3. 哪些產品有香甜的氣味？ _____
4. 哪些類型的氣味與乾淨清爽有關？ _____
5. 你會如何形容果味和檸檬味的差異？ _____
6. 簡單地說，你會用什麼詞語來形容女性和男性之間的香味差異？ _____
7. 哪些話可以完整地描述曬衣架的味覺？ _____
8. 描述一家麵包店中明顯的氣味。 _____

9. 描述在液體洗碗精中明顯的氣味。 _____

10. 描述肥皂中明顯的氣味。 _____

11. 描述在地下室的明顯的氣味。 _____

12. 描述一家麥當勞餐廳中明顯的氣味。 _____

E 量度練習

(將包含在每個篩選問卷中)

EXAMPLES



None |-----/-----| All



None |-----/-----| All



None |-----/-----| All

1.



None |-----| All

2.



None |-----| All

3.



None |-----| All

4.



None |-----| All

5.



None |-----| All

6.



None |-----| All

7.



None |-----| All

例子

無_____全部

無_____全部

無_____全部

1. 無_____全部

2. 無_____全部

3. 無_____全部

4. 無_____全部

5. 無_____全部

6. 無_____全部

7. 無_____全部

8. 無_____全部

9. 無_____全部

10. 無_____全部

篩選問卷：量度練習答案：

1	7/8
2	1/8
3	1/6
4	1/4
5	7/8
6	1/8
7	3/4

8	1/8
9	1/2
10	1/2

附錄 9.2 領導品評小組的建議

A 招募品評員提示

招聘品評員需要創意和毅力。創造力發揮於設計和廣告刊登時。以下是可以打廣告的地方清單：

- 當地報紙的家庭、食物、或周末版面
- 郵寄優惠券
- 雜貨店、健身俱樂部、社區等佈告欄
- 線上招募，如 Craig 的名單 (<http://www.craigslist.com>)
- 從現有的品評小組或員工內找人介紹
- 當地的廣播電台。
- 社區有線電視台
- 學校
- 洗衣店

B 為提高感官意識和動機的品評小組活動

• 共享感覺記憶。請品評員分享一個感覺記憶，藉著在畫架紙上或白板上大字寫“我記得.....”，並鼓勵品評員使用自己的話講一個故事。準備好分享你的感官記憶。

- 設立一個「感官展示」的日子。邀請品評員互相分享自己的感官體驗。每月一次，並要求品評員註冊。
- 建立一個前十大清單(David Letterman)。例如：「我們很高興不用評估的前十大物品。」
- 做一個百事可樂、可口可樂的三角測試或其他常見物品（泡芙與舒潔等），然後再討論。

- 建立一個跟感官相關的漫畫/動畫牆（或公告板），每個月定期出刊漫畫/動畫。要求品評員一同參與：紐約客雜誌是個很好的資源。
- 去戶外散步。配對品評員，並要求他們出去戶外進行 10 分鐘的步行路程，請他們記錄從離開房間那一刻至返回的路程中發現的氣味。
- 畫出聲音的內容。播放一些音效，請品評員繪製聲音的樣子（鼓勵品評員不要畫出是什麼產生聲音），如果聲音有一個外型，畫出它的形狀、形式或色彩。然後再問聲音會聞起來像什麼，嘗起來和/或感覺起來像什麼。
- 蒙眼測試。將品評員蒙上眼睛，並呈現一些感官刺激，要求他們不僅猜它們是什麼，也描述其感受到的感官特性。
- 與品評員共同建立感官體驗的願望清單。他們希望能夠體驗到什麼樣的感官刺激。共同發想 100 條，然後選擇 12 條（每月一個）並執行他們。記住你將有很多機會在發想清單的過程中觀察到不同的判斷過程。當然有些項目會是遙不可及的（例如吞毛皮），但是這是樂趣的一部分。
- 請每個品評員與整個小組分享一個他們認為很酷的感官體驗。不要限制在人類的感官知覺上，也可以分享其他生物的事實。進行一個最酷（或最不起眼的）事實比賽。
- 設立「害怕的食物」品嚐日。要求每個品評員帶他們最不願意吃的食物（如豬腳）來品嚐和描述。然後問一些問題：哪些食品嘗起來味覺最好？哪個食物讓你吃驚？
- 香薰經驗。選擇一種芳香（如薰衣草）並找尋使用此香味的各種產品。進行薰衣草香味產品之間的比較。它們以何種方式相似？又以何種方式不同？
- 舉行每週（或每月）的「感官讀書會」，請品評員帶相關的書並大聲朗讀關於感官描述的段落。這不應超過 5 分鐘。我們都知道普魯斯特說明瑪德蓮蛋糕的段落，鼓勵品評員建立一個感官閱讀的推薦清單。
- 光學幻覺日：進行不同的光學錯覺遊戲。
- 請到網站 <http://puzzlemaker.school.discovery.com/>，並請品評小組猜猜感官相關的字謎。看看誰可以最快完成。
- 收集電影中跟感官有關的場面，並在每次品評測試結束後播放。要求品評員帶來其他電影片段。如果你不知道從哪裡開始，可以使用情定巴黎(French Kiss)中凱文·克萊恩教授教導梅格·瑞恩關於葡萄酒味覺的細微差別的片段；他甚至翻出一個參考詞彙組。
- 設立一個感性日。要求每個品評員將他們喜歡的東西帶來，並讓所有品評員傳閱並描述。你可以設立專注在不同類型的主題，比如說軟的物品、黏的東西等。
- 與品評員約在當地市集，並設立一個感官尋寶目標。配對品評員，使他們進行團隊工作。

- 與品評員一同創建感官的填詞遊戲（書店有相關書籍）。如果你不能找到一個適合自己的書籍，創建你自己的目錄。

如果你不知道什麼是填詞遊戲：找一位 8 歲孩童，他們保證知道。

- 請品評員製作一個複雜的感官體驗的拼貼，如奶味、清新、清爽、滋潤、柔軟等，要求每個品評員解釋他或她的拼貼畫，然後掛在牆上，以創建一個大的拼貼畫。請品評員分享他們所看到的感官特性？對這些感覺的見解？
- 邀請專家與品評小組進行座談會。專家包括調香師、廚師、侍酒師，花卉設計師，時裝設計師，調音師等。
- 玩感官字謎遊戲。
- 舉辦「_____（產品）的設計最差」的比賽。讓品評員進行團隊工作，以創建產品最壞的感官體驗。進行約 10 分鐘。
- 收集掉漆的碎片（許多不同的顏色），並要求品評員幫牠們取名。設定主題如：只有食品的名稱、度假地名等。
- 另外，只是為了好玩，可以與品評員一起玩「名稱遊戲」，準備品評員的名稱標籤，放置在他們的背上（不要把該品評員的名字放在他們的背後）。遊戲的目的是透過詢問只有是或否的問題，找出他背後的名字。

C 品評小組指南

公司提供品評員一個安全，舒適的工作環境，並提供品評員靈活地選擇他們希望參與研究的時間。

到達時間

- 要求品評員必須在測試開始前的 10 分鐘到達。
- 大衣和毛衣需要掛在大廳的壁櫥裡。
- 準備測試的地點。
- 要求品評員記錄其進出時間。由於薪資是基於時間表支付，品評員必須記錄他們的參與以便計算薪資。

產品練習和研討會

- 這些測試的目的是審查相關屬性，並提供研究期間出現的相關問題的反饋。

- 練習測試依需求安排。
- 參與這項測試的研討會之前，產品練習的測試必須全勤。如果沒有全勤的人將無法參與後續的研究。

研究計劃

- 客戶與品評小組的領導人會議排定測試需求。
- 小組的領導人通知品評小組任何即將到來的研究要求。通常情況下，研究需要 10 位評委的品評小組以提供樣品數據。

- 潛在的品評員必須能夠參加所有預定的測試，以參與這項研究。
- 為了滿足客戶的期限，一般無法安排個人補測試。不過品評小組領導人有自由裁量權。
- 如果整個測試由於天氣或品評小組生病而取消，則另找一天進行補試。

假期時間

- 品評小組的領導人必需知道品評員的休假計劃。這是必要的，以滿足客戶的計畫時間，並避免成員之間休假衝突。
- 品評小組在感恩節的一周，聖誕節和元旦之間的一周將不會排定測試。

健康食品的品評小組範例

薪水

- 每月一次支付，支付期上在本月 20 日結束的附近。如果可能的話，將在排定的測試結束後直接給予品評員支票。有些支票可能要郵寄。
- 薪級表如下：
 - 培訓期間\$X/小時（注：培訓包括一系列密集的測試，以及持續三個月兩週一次的測試）。
 - 第一年訓練後：每小時\$ 1.2X。

- 第二年以後：每小時\$ 1.35X。
- 品評員的小時計算是以最接近的半小時為基礎，並納入 1/4 時。
- 保證薪酬兩個小時，依據品評員到達每個排定測試的時間計算，除非另外有專題研究。
- 品評員必須參加所有預定的測試以獲得薪水支付。個人緊急情況將根據情況的性質進行審查。
- 守時是重要的，以免延誤或打亂小組會議。拖延準時到達的品評員是不公平的。
- 如果品評員遲到超過 5 分鐘，付款會從下一刻鐘開始。
- 品評員應避免在小組測試前有個人約會。
- 如果品評員需要提早離開，應盡快通知品評領導人。付款將僅基於品評員在品評小組的時間，而最低支付兩小時薪資的條款將不適用。

獎金

- 品評員每年一次有獎金審查的機會。
- 獎金的計算標準包括績效、遵守小組協議、考勤和數據的有效性。
- 包括準時出席，並參與預定的研究（80%）。
- 品評小組的協議包括準備區域，盡量少說話，有組織的工作空間，注意細節等等。
- 數據的有效性將由客戶端的研究和驗證研究來確定。
- 數據可靠性是非常重要的。

注：品評員審查健康食品的績效都將被評量和紀錄。不良記錄的品評員將不會收到獎金，並進行審查且停薪。

品評結果的完整性

- 繳交品評結果給小組領導人之前必須完整地檢查。
- 品評記錄表包括名稱、品評員 ID、樣品編碼，所有的屬性成績等數據應該由同組品評員校對。
- 數據不能在事後被重建。遺失的數據無法再使用，因為可能會導致品評小組的研究與公司的費用衝突。

- 如果研究由於不完整的選票需要再次進行，那些寫了不完整選票的品評員將不被支付重複研究的費用。

驗證

- 驗證研究在於記錄品評員的平均值和標準差，以及整個品評小組的平均值和標準差。
- 這包括品評員重複品評樣品的可靠性。
- 這些研究涵蓋的領域從觸感到氣味評估，包括化妝水、洗手液、香皂等
- 品評小組完整的記錄結果依照客戶需求提供。

學習規劃

- 利用現有的品評員組成的小組人員庫進行研究。
- 每個研究需要至少十位品評員。多於十位的品評員更好。
- 所有品評員都將參加練習賽。
- 品評員必須參加所有測試才有資格獲得研究結果。

討論

- 品評小組房間內的討論僅限於品評領導人領導的。
- 品評員必須專注於評等樣品。說話時會從手頭的任務分心，提高犯錯誤的機會，忘記記錄數據等。
- 在評估期間，品評員可以攜帶閱讀材料或進行其他安靜的個人工作。
- 某一些閱讀則可能不被允許，必須等到自由的時間，或是等待下一次評價的時間。

品評小室

- 品評小室中不允許任何點心。
- 禁止吸煙。

- 每次測試結束後必須清理品評小組區域。
- 所有的雜誌和報紙使用後都必須歸位

取消品評

- 惡劣的天氣出現時，盡快做出決定，且若取消的話，品評小組的領導人必須聯繫有關的品評員。
- 冬季天氣通常遵循當地學校的政策取消。如果道路有學校開通，則排定的品評計劃須進行，除非品評小組領導人有不同的決定。該小組負責人可以使用通話鏈的聯絡方式通知品評員變更。
- 通話鏈用來通知品評員在日程安排有變化時。
- 品評小組領導人是通話鏈的起始者。
- 接電話的那個人將聯絡通話鏈上的下一個人。
- 如果手機無法聯絡到人，留下語音訊息，並繼續聯絡下一位在名單上的人，直到被接通。
- 上述過程繼續，直到最後一個人打電話告知品評領導人已完成通話鏈。
- 列表更新是必要的。品評員應告知領導人假如有任何電話號碼和地址的變更。

緊急呼叫

- 如果你不能參加一個計劃小組會議，請致電並留言。
- 非工作時間並沒有答錄機。
- 請限制辦公室中的私人通話。這款手機應該只用於緊急聯繫人。

外部準備

- 對於質感的品評小組：
 - 建議品評員複習之前的學習和練習課程的相關協議。
- 測試時間小心保養皮膚。天氣可能會損傷皮膚表面而影響評估。

- 測試評估的前一天不要在皮膚上使用乳液或面霜。包括揮之不去的香料如洗髮精、髮膠，香水等。
- 用洗滌劑洗碗時請使用橡膠手套。
- 從事園藝時使用手套保護，以免長出老繭和水泡。
- 外出時使用防曬品，防止曬傷。
- 請注意品評測試之後你的皮膚質感和表面的變化。
- 品評員若有任何過敏反應，應該報告給組長。

第十章 描述性分析技巧

10 Descriptive Analysis Techniques

描述性測試是最複雜的感官測試方法，雖與差異性測試同屬分析性的工具，但描述性測試提供的是比較完整的感官描述。

10.1 定義

描述性測試是一種描述產品中可感受感官特性的步驟，它納入所有的感官-視覺、聽覺、嗅覺、味覺、觸覺等的描述。所有描述性分析法都是藉由五至一百位受過訓練的品評員對樣品的品質及數量偵測（區別< discrimination >）和描述。五到十人的小型品評小組通常用於典型的商店物品，而大型的品評小組則往往用於大量製造的商品，如啤酒、汽水等，其中若有微小的變異都會產生重大的影響。

品評員必須偵測和描述所感受到的樣品的感官屬性。這些產品的品質組合定義了這項產品，包含它所能於其他產品產生區別的外觀、風味、質地或聲音特性等。除此之外，品評小組必須學習區別並評級產品的數量或強度，並且定義此樣品中有多少數量的特性及其強度。兩項產品可能有相同的質性特徵，但卻有不同的強度，因此造成相當不同且容易檢測出的感官印象。下面兩項產品有一樣的質性特性，但在此特性上卻有潛在不同的強度。用於評定此強度的等級是 15 公分的線性量度(line scale)，其中 0 代表沒有任何量被檢測出，而 15 代表檢測出非常多的量。

下面兩項樣品(385 和 408)是市售薯片：

特性	385	408
炸薯片	7.5	4.8
生馬鈴薯	1.1	3.7
植物油	3.6	1.1
鹹度	6.2	13.5

甜度	2.2	1.0

雖然這兩項薯片的樣品有一樣的描述性屬性，各個屬性上的強度卻不同。樣品 385 因為植物油、甜度及生馬鈴薯等特性造成其特徵，而樣品 408 則是鹹度特性占主要的特徵，而其在蔬菜油和甜度上的表現都相當的低。

10.1.1 方法組成

- 1.質的描述：如屬性、特質/特性、描述語、描述詞等。一般將五官感受系統分為七大類。
- 2.量的描述：每一感官特性，感受的強度不同。通常以一個“值”來表達，相對量由參考樣品來規範。
- 3.出現時間的描述：如餘後味、餘後感、殘留時間的不同。要求品評員依時間順序進行評估。如魚丸先評估硬度，再評估彈性，最後評估咀嚼感。
- 4.整體印象的描述：對產品或目標的整體印象的描述。通常分為氣味/風味的綜合強度、整體平衡感、整體差異性等三類。

10.1.2 方法原理

- 1.先確定描述測試的目標：包括測試樣品的範圍，如不同廠牌、不同原料來源、不同製程、不同生產線等。
- 2.篩選品評員：需 10-15 位，儲備 20 位適用品評員。
- 3.由品評員品評樣品，發展出所有可能的重要或足以描述目標的感官特性或描述用語。
- 4.經過團體討論和參考樣品的協助，整合描述用語。
- 5.依列表中的感官特性，以具體的描述統一定義，並統一品評作業的動作和評估時間。
- 6.使用標準樣品或參考樣品，訓練品評員熟悉用所要求的標示法來表達各感官特性的強度。
- 7.依已建立的感官特性列表，執行描述性測試，每一試驗做三重複。
- 8.進行數據分析，建立剖析圖。

10.2 應用範圍

使用描述性測試獲得食物及飲料的香氣、風味、口感，以及個人用品的觸感或織品或紙類的手感，或所有產品的聲音及外表等詳細的描述。這些感官輪廓被用在研究發展及製造中：

1. 定義新產品開發中的目標產品的感官特性。
2. 定義品質保證/品質管制(QA/QC)及研發應用的一對照品或一標準特性及規格。
3. 在進行消費者測試之前先建立產品屬性，以幫助從消費者問卷內各屬性中選擇屬性，並幫助解釋消費者測試的結果。
4. 追蹤產品的感官變化以確立保存期限及包裝條件。
5. 定位感知到的產品屬性及其相關的儀器、化學、成分或物理特性。
6. 量測特定屬性在短期時間內的強度變化(時間-強度分析)。

10.3 描述性分析的要素

10.3.1 特性：質的方面

產品的可感知感官要素(perceived sensory parameters)意謂以各種名詞來定義產品，如屬性、特性、描述詞、術語。這些質性因素（如同第五章分類法討論過的要素）包含了該樣品的感官輪廓、感官印象等條件。其中重要的是，除非是由受過訓練的品評員測試，這些條件的解讀可能會因人而異。感官屬性的選擇及其相關的定義都應該與此產品相對應且能夠被檢測出的化學及物理特性有關。遵循瞭解產品真實的物理特性以及化學性質的路徑，可以使得描述性資料更容易解讀，並對決策過程更有幫助。統計方法如變異數分析以及多變量分析（multivariate analysis）可用來選擇區別性更高的條件。

下列提供了一系列不同的描述輪廓的要素（範例顯示在括號中）。注意此清單是理解第十一章的附錄 11.1 至附錄 11.3 中更完整的清單的關鍵。某些特性及範例重複地出現是有原因的。

1. 外觀特性

- a. 顏色（色調、飽和度、一致性、明暗度）
- b. 表面質地（明亮、平滑度／粗糙度）
- c. 尺寸及形狀（容積及幾何形狀）

- d. 塊或分子間的互動（黏稠度、聚集情況、鬆散狀況）
- 2. 氣味類特徵
 - a. 嗅覺（香草味、水果味、花香、臭味）
 - b. 鼻腔感受因子（清涼、刺激）
- 3. 風味類特徵
 - a. 嗅覺（香草味、水果味、花香、巧克力味、臭味、腐爛味）
 - b. 口味（鹹味、甜味、苦味）
 - c. 口腔感受因子（熱、冷、燙、澀、金屬感）
- 4. 口感質地特性
 - a. 機械性參數：產品對應力(stress)的反應（硬度、黏度、型變／脆度）
 - b. 幾何參數，即產品的大小、形狀和顆粒（砂狀，粒狀，片狀，線狀）
 - c. 油/水參數，即脂肪、油或水的存在、吸附、釋放（油、油膩、多汁、潮、濕）
- 5. 皮膚觸感特性
 - a. 機械參數：產品對應力的反應（厚度、緩和展開、滑溜度、密度）
 - b. 幾何參數：即產品顆粒或皮膚上使用後的大小，形狀（砂狀、泡沫、片狀）
 - c. 油/水參數，即脂肪、油或水的存在、吸附、釋放（油膩感、油感、乾、濕）
 - d. 外觀參數;產品使用過程中的視覺變化（光澤、美白、消瘦）
- 6. 織物和非織造的紋理、手感
 - a. 機械特性：產品對應力的反應（剛度、壓縮力或撐力、韌性）
 - b. 幾何特性：即顆粒的尺寸，形狀和方向（砂狀、崎嶇、粒狀、稜紋的、毛絨狀）
 - c. 濕度特性：水分的存在及吸收度（乾、濕、油性、吸水性）。

另外必須重複的是，描述性分析測試的有效性和可信度倚賴於：

- 對於風味、質地或外觀的技術和生理原理的徹底理解；
- 所有品評員皆徹底受過訓練，且充分了解相同的術語，以及使用相同的方式引用它們；
- 術語的應用（參見第 11 章，附錄 11.2），以確保品評員以一致的程序使用描述性用語。

10.3.2 強度：量的方面

上述質的描述在產品並不只是有或無而已，不同的產品可能含有相同的感官特性描述語，但感受的強度不同，這就是所謂量的描述。這種程度或強度的不同，我們通常以一個“值”來表達，當然此處所謂“值”的概念是相對的，不是絕對值，而這個相對概念或相對量(箱對值)是由參考樣品來規範，而參考樣品也是由品評小組的共識而來

描述性分析的強度，或數量方面，是表示對每個特性（術語，品質成分）的存在程度。這是藉由在固定量度上分配某些值來表達。而術語及強度測量的可信度及有效性是高度依賴於：

- 選擇的量度技術足以涵蓋全部的參數強度，並且有足夠的分散點可表現樣品間強度的所有細微差異；
- 品評員的全面訓練，以相同的方法使用量度花同樣的時間去評估所有樣品（見第 9 章品評員培訓）；
- 針對不同屬性的強度使用的參考量度（見附錄 11.2），以確保在不同強度的感官性能上，不同品評小組使用一致的量度，並具有重現性。。

描述性分析中常用三種類型的量度：

1. 類別量度（**Category Scaling**）是一組有限的單詞和數字，由保持各個類別之間相等的間隔建構而成（以作為最佳解釋）。可以在第 5 章中找到全面的描述。類別量度從 0 到 9 也許是最常用的描述性分析，但使用更長的量度往往有其判斷。此經驗法則是評估品評員需要使用多少量度的量度，並設置兩倍長度量度。有時使用 100 分的量度，例如在視覺和聽覺的研究中。
2. 線性量度（**Line Scaling**）是讓品評員在線長 6 吋或 15 公分的直線上做標記。在第 5 章中有詳細的描述。線性量度幾乎與類別量度一樣流行。其優點是強度可以更準確地分級，因為沒有步驟或「數字偏好」。使用線性量度的主要缺點是，品評員更難一致，因為沒有了數字，他們不容易記住線段上的距離。
3. 幅度預測量度（**Magnitude Estimation Scaling**）是基於自由分配第一個數字的原理，在此之後的所有後續數字按比例分配（見第五章）。幅度預測量度大多用在學術研究上，重點在觀察一個屬性上感官強度的大幅度變化。

第 11 章的附錄 11.2 包含一系列各種氣味和口味的參考樣品，作為建立機械的、幾何的、濕度和口感屬性的量度。附錄 11.2 所有的數值是基於一個 15 公分長的線性量度；然而，同樣的標準可以分布在任何長度的線或數值量度。這些量度採用標準的水溶液，例如蔗糖、氯化鈉、檸檬酸和咖啡因，已經足夠一致性地測試超市常見的物品，例如：Hellmann's® 蛋黃醬和 Welch's® 葡萄汁。

10.3.3 出現的順序：時間的觀念

除了樣品的屬性（質）和每個屬性的強度（量），品評小組可能因產品順序導致檢測出參數間的差異-時間的因子，也就是在消費或接觸產品的過程中，不是每個感官特性都在同一時間出現。物理特性呈現的順序，如口感、皮膚觸感和質地等一般是由產品的操作方式而預先確定的。通過控制操作方式（咀嚼、用手擠壓），品評員在一段時間內誘導有限屬性（硬度，緻密性，變形）的數量上表現。例如同樣是咀嚼過程產生的感官特性，但硬度的感受一定在咀嚼度之前發生，因為硬度通常是咬 1~2 下時感受，而咀嚼度通常必須在咀嚼 5~6 下以上時才能評估。但是，樣品的化學組成（香氣和風味）及其一些物理性質（溫度、體積、濃度）可能會在檢測的過程中改變某些屬性的順序。例如飲料類產品，特徵出現的順序常常是代表此產品的資訊，如個別的香氣和風味特點，及其各自的強度。樣品經過處理後，特定屬性出現的順序，像是餘味或用後感等，仍然視為此產品所能被感知的屬性。這類的特性都是刺激(食品)離開口腔後所產生的特性，且殘留的時間長短不同。一個產品的完整感官圖像是產品使用後感知得到的所有特徵及其強度。

餘味或用後感的屬性描述並不一定是負面的註解。例如：薄荷糖的涼爽餘味是必要且預期的特徵。另一方面，可樂的金屬餘味可能暗示著包裝污染或特定甜味劑有問題。

在一段指定的時間內反覆測量及追蹤一個或多個（通常不超過三個）感覺屬性的強度，此技術稱為**時間強度分析**(Time-intensity analysis)。本章稍後會更詳細介紹這種技術。

10.3.4 整體印象:整合觀念

人與儀器的不同，有一個重點就是整合的能力-對產品或目標的整體印象的描述。除了藉由檢測及描述產品的質性、量性及時間因素來定義產品的感官特性外，品評員能夠綜合評估一些產品性能，且通常管理人員對此較有興趣。這種嘗試整合方式的包括以下四種：**氣味或風味的綜合強度**(Total intensity of aroma or flavor): 指的是產品中所有影響整體氣味或風味(感知到的揮發物)的綜合強度(Overall intensity) 或綜合衝擊 (Overall impact)，包括氣味、口味，以及貢獻風味的感覺因素。這種評估方法可以評估整體香味或風味的衝擊，

因為有時提供給消費者的產品難以被品評員描述其在整體風味中的氣味或味道貢獻的細微差別。質地部分則是比較離散的特性，並不存在「整體質地」這種屬性。此處指的是綜合，而非總和，並不是把所有的氣味強度加起來的意思，而是經過人類大腦整合過的整體強度。

整體平衡感 (Balance/Blend)：或稱為振幅(Amplitude)。經常有人要求經良好訓練的品評小組評估各種風味或香味特性在產品上的結合程度。這種評估各種屬性各自的強度，以及是否平衡且協調地混和，涉及了複雜的理解過程，是一半學術一半直覺的。整體平衡感（或在風味剖析法[Flavor profile]中稱作振幅[amplitud]）的測試甚至對經過完整培訓的品評員都相當困難，不應該要求初學的品評員嘗試。此外，使用整體平衡感的數據必須特別小心。常常有些產品並不具有平衡的混合：麻辣感或焦味可能是此產品充滿個性的關鍵優勢。因此在這些產品上，消費者可能不會欣賞存在均衡的組成，儘管其整體平衡感是由經訓練的品評小組所確定的。因此，要測量和使用這樣的數據之前，必須先理解此共混物的平衡之於消費者的相對重要性。

整體差異性(Overall difference)：在某些產品的情況中，關鍵決策依賴於樣品和一些控制組或標準產品的相對差異。雖然每個產品之間的屬性差異的統計分析和描述技術有許多種，計劃領導人經常關心的只是樣品和標準樣品有什麼樣的不同。整體差異性的確立（見第 6 章的第 6.8 節，參考比較法）允許計劃領導人得以根據樣品及其控制組的差異來源及大小來決定相關的處置。

喜好評分法(Hedonic ratings)：它是在描述性品評小組完成描述測試後，嘗試詢問品評員對該產品的總體接受度。在大多數情況下，這種嘗試必須要抵制，因為作為品評員，在其培訓過程中，已經刪除其與大部份消費者共同的觀點，因此他不再具有代表公眾的任何部分。培訓往往會改變品評員的個人喜好。當他們更加了解產品的各種屬性，品評員將會從不同的方式衡量屬性，便不再是個普通消費者。

10.4 常用的描述性測試方法

在過去的 50 年中已經制定了許多描述性分析方法，有的已經獲得廣泛使用並成為流行的標準方法。這些方法如下所述的事實是其受歡迎的反映，但並不建議使用。相反地，感官分析師為了特定的產品和計畫，如果希望開發一套描述性系統，應該研究描述性方法的文獻，並在建立描述性分析系統之前審查所有能夠包含四種方法的案例研究，另外在 IFT 感官評估指南中（IFT 1981 年），包含 109

種不同領域的參考案例。最近在感官科學文獻和演示文稿中並沒有新的描述性分析方法。但針對現有方法的修改則是常見的，以促進品評小組針對一個產品或產品類別的感官特性建立客製化的描述性分析法和小組記錄。

10.4.1 風味剖析法（Flavor Profile Method）

風味剖析法是由 Arthur Little 博士在 1940 年代後期所開發，利用描述用語對產品的風味進行描述。定量部分以 0、X、1、2、3、4 或 0、X、0.5、1、1.5、2 或+號的數量來表示強度。它包括由四到六個經過訓練的品評員所組成的品評小組鑑定產品的氣味和風味特徵及其強度，且按照氣味、風味、餘味的順序評估分析。振幅評級（見前頁）通常佔此方法的一部分。

按照品評員的生理測試決定其鑑別風味、鑑別風味強度，鑑別嗅覺及描述。安排個人面試以確認品評員的興趣、時間配合度及潛在的工作情況。在培訓過程中提供品評員各種不同產品種類的參考樣品，以及廣泛的產品類型成分和加工變數。品評員，藉由品評小組領導者提供和維持許多參考樣品的幫助下，開發和定義整個品評小組共用的描述性術語。品評小組也可以開發共同的參照框架，如第五章所討論的，使用七點風味輪廓強度的量度（Seven-pint Flavor Profile intensity scale）。

品評員就座於圓桌或六角桌邊，各自評估樣品的香氣和風味，並按照外觀和餘味的順序記錄其屬性，又稱為「特性紀錄」(character notes)，及其強度。將結果報告給品評小組的組長，測試之後組長將帶領品評小組進行一般性討論，並且在每個樣品的描述上達到共識。該數據通常以表格形式報告，不過圖示也是可以的。

風味剖析法適用於當品評小組必須評估許多不同的非主線生產的產品時。此法主要的優勢，但也是其重大的缺點，是它僅使用七至十五位品評員。缺乏一致性及重現性在某種程度上可以藉由培訓和共識法(consensus method)克服。然而，後者一直被批評過於片面性。品評小組中可能被某些年長的或具支配個性的品評員掌控了整個小組的意見，使得品評小組的意見輸出並不太公正。另外對風味剖析法的批評包括在評定特定氣味或風味的差異時屏除了篩選法，然而這可能在具體的產品應用中是很重要的，另外七點量度的限制也可能忽略了產品中微小程度但極為重要的變異。

10.4.2 質地剖析法（Texture profile method）

基於上風味剖析法的原理，質地剖析法是美國通用食品(General Food)公司透過產品評估和質地技術小組所制定的食品質地參數而建立。該方法之後經由 Civille 和 Szczesniak 和 Civille 和 Liska 擴展，包括描述特定產品，如半固體食品、飲料等的特定描述術語。每個

產品類型都有其特定的術語，但它是第一本質地資料出版物。

在針對特定的產品應用上（固體食物、飲料等）接受過訓練的品評小組中，依照其鑑別質地的能力來選擇適合的品評員。與大多數其他描述性分析技術相同，安排品評員面試，以確定其興趣、時間配合度和態度。選擇培訓的專業品評成員將接觸到正在研究的類別中廣泛的產品，以提供寬廣的參考框架。此外，品評員需接觸所涉及研究的產品結構的相關質地原則。這樣的學習經驗提供了品評員基礎的概念以避免曲解產品的判讀。相反地，品評員能夠避免在多餘的術語上冗長的討論，並且選擇評估產品最合適的技術和描述性用語。品評員也必須定義進行評估的所有程序和所有術語，藉此可減少在大多數描述測試所遇到的變異。在品評員的培訓中使用的參考量度，可以在實際操作時當作類似的刻度值，進一步降低品評小組的變異。

每位專業品評員使用前面所討論的任一量度技術各自評估樣品。原始質地剖析法使用的刻度是風味剖析量度的擴大 13 點版。然而在過去的幾年裡，質地剖析法的品評員卻持續被教導使用類別、線性、幅度預測量度等方法進行評估（參見第 11 章，附錄 11.2，使用 15 點或 15 公分的線性量度評估食品質地參考樣品）。根據由品評小組呈現的數據，以及使用哪種量度的類型，品評小組的最終判讀可以與風味剖析法相同，透過會議達成共識，或者透過對數據進行統計分析。最終報告中，數據可以以表格或圖形呈現。

10.4.3 定量描述性分析(The Quantitative Descriptive Analysis) (QDA®)法

許多感官分析師認為在風味剖析法或其相關方法中的數據缺少統計處理，為了解決這種不足，Tragon 公司開發了用於描述性分析的 QDA®法（Stone 等人. 1974;Stone 及 Sidel 1992）。該方法大量地依賴統計分析，以確保專業品評員分析特定產品時採用適當的術語及程序。

根據區分特定類型產品的差異能力，從大型的品評員庫中的候選人間挑出品評員，並安排相關培訓。QDA 品評小組的培訓要求使用真實產品和相關成分，與其他描述方法相同，是為了刺激術語的產生。該小組的領導者的角色是作為一個促進者，而不是作為一個教練，並須注意不要影響品評小組。小組的目標是發展一致的術語，但品評員可以在提供的 15 公分（6 英寸）線性量度的方法上自由使用自己的方式評分。

QDA 品評員在分開的品評小室單一地評估產品，以減少分心及品評員之間的互動。品評員將數據輸入電腦中，或是各自的數位化評分卡中，再統一收集並由讀卡機自動讀取計算。品評員不得討論數據、術語或是直接讀評分卡，且他們必須仰賴小組領導人的自由裁決，才能進一步得知自己的表現與其他的品評員的相對差異。

QDA 測試的結果將會進行統計分析，並通常以蜘蛛網狀的數據圖形報告，表示從中心點延伸至每個屬性的表現。QDA 方法是與加州大學戴維斯分校大學部的食品科學部合作開發的。它主要呈現將人類智能被視為測量儀器使用的理想跨越了一大步，在第一章中我們特別討論過，使用圖形量度圖（視覺模擬），可以減少使用數字而產生偏見的量度。數據的統計處理、品評員獨立地評分，加上提出數據圖形的方式使得感官科學家向其客戶在視覺描述方式上發生很大的改變。以下是從此變化或進一步發展可能受益的領域：

- 1.品評小組，由於缺乏正規的指導，可能發展出錯誤的數據。例如：天然香草和純香草醛之間的差異應該很容易被經過訓練的品評小組檢測並描述出來，然而非經導引的品評小組可能會選擇使用香草一詞來形容香草醛的味道。缺乏指導的方向卻可能允許資深品評員或個性強的品評員主宰品評小組全部的術語發展。
- 2.自由地使用量度的方式可能會導致結果不一致，有可能是品評可能在某一天測量產品但在另一天沒有測量，也有可能是一個可見產品跟另一沒有內在量度參考品比較而產生的環境影響。
- 3.定期即時反饋的機制給品評員減少了學習和擴充了區別或描述更多差異的機會。
- 4.另外，蜘蛛網量度的表現方式可能會誤導一些經過技術培訓的使用者，他們期望曲線下面積有一定的意義。在現實中，此圖表所示的感官尺寸是相互無關的，或是不能以這種方式來表示其有關。

10.4.4 頻譜™描述性分析法（Spectrum™ Descriptive Analysis Method）

此方法由 Civile 設計，在第十一章中描述了相關細節。其主要特點是品評員使用先前培訓中的絕對強度量度來評分其感受到的樣品屬性強度。目的是使得結果剖析圖獲得普遍理解，而且之後數據可以在任何實驗室使用。基於此目的，該方法提供了標準屬性名稱（詞彙）的陣列，每個屬性都有其定義強度的量度，通常是從 0 到 15 的標準，可以使用 15 公分線長的量度測量，或簡單地使用數字記錄。

10.4.5 時間-強度的描述性分析（Time-Intensity Descriptive Analysis）

對於某些產品來說，感知的強度隨時間變化，不管是長時間或短時間。屬性的時間-強度曲線因此變成此產品的關鍵層面。長時間－強度的研究週期性地測量使用乳液後的數天內皮膚乾燥的減少程度。唇膏的顏色強度可以在幾個小時內定期評估。短時間－強度的研究追踪了口香糖味道在幾分鐘內的質地屬性。最短時間的研究可以在 1-3 分鐘內完成，反應可以連續地記錄下來。例子包括甜味劑甜度、啤酒苦味、茶的回甘等效果。餐飲大多屬於短時間－強度的研究。反應可使用紙筆紀錄、滾動圖表記錄器或電腦系統。品評員不應該

看到不斷變化的回應曲線，因為這可能導致其先入為主的印象而產生偏差。

時間—強度的線型研究已由 Lee 和 Pangborn（1986 年）全面建立了方法論，特別在甜味劑方面（Booth 1989）。要考慮的重要變數是：

- 評估草案(Protocols for evaluation)-產品的呈現形式、數量、在嘴裡的時間、人工操作或吞嚥的需要都該明確定義。
- 連結評估產品（控制維持）和回應記錄（輸入數據）的草案需要提前制定並演練以減少呈現模式的偏差，必須訂定相關協議。
- 品評員可能需要經過數次培訓，以良好學習和建立所有控制時間強度的研究中所必須的草案。是紀錄時間—強度研究的參數的一個例子。

表 10.1 三種甜味劑所獲得的反應實例。

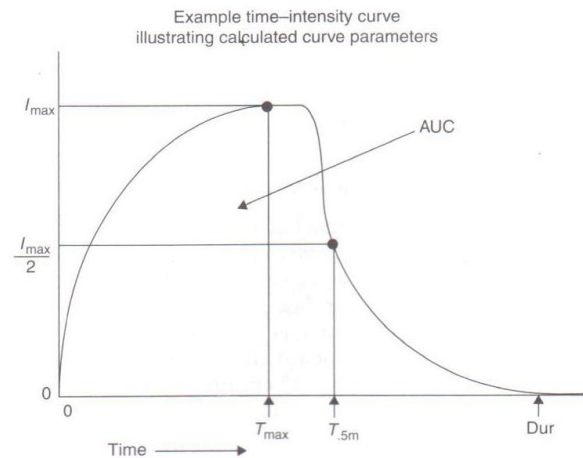


圖 10.1 圖解時間—強度曲線的計算參數。I_{max} 為觀測到的最大強度; T_{max} 為最大強度發生的時間; AUC 為曲線下的面積; DUR，強度持續時間：直到強度回落到零的時間; T_{5m}（T_{max} 後）是強度已經下降到最大強度的一半的時間

10.4.6 自由選擇剖析法（Free choice profiling）

自由選擇剖析是農業食品委員會的 Williams 和 Arnold（英國）作為解決消費者對既定屬性使用不同術語的問題而開發。自由選擇剖析使品評員依照自己需要去創造和使用許多術語，去描述一組樣品的感官屬性。在相同類別產品的樣品上，品評員開發其自己的記分表。使用廣義的 Procrustes 分析數據，一種多變數量度技術被不同的品評員調整及使用，然後微調數據去結合測量相同特性時出現的術語。這些組合提供了單一產品的剖析輪廓。

將自由選擇剖析和其他描述技術結合的研究正在進行中。比起其他描述技術來說，這項新技術的主要優點是品評員不需要經過任何訓練，大約一個小時的指導即可上場操作，節省許多時間。

表 10.1 三種甜味劑的強度數據

參數	7.5%蔗糖 (條件樣品)	0.05% 阿斯巴甜	0.4% 乙酰磺胺-K	7.5%蔗糖
曲線下面積 cm2	121.2	153.7	98.6	154.2
最大強度, Imax	7.2	7.6	7.8	7.6
最大強度時間	7.4	8.2	4.8	6.2
持續時間	28.3	33.3	24.7	33.4

第二個優點是品評員並沒有經過訓練，因此仍可視為消費者的代表。然而，關於如何結合品評員解釋感官所得結果的能力，必須加以解決。為了提供產品研究可靠的指導，實驗者/感官分析師必須決定每個組合詞的實際意義。因此，每個生成的參數或術語來自實驗者或感官分析，而不是品評員。結果可能因為分析者的角度，而非品評員的判斷而有所不同。

10.4.7 稀釋描述分析法

此方法是學者 Tilgner 於 1962 年提出，其動機為由於進行風味描述分析時，食品系統中常有風味重疊、遮蓋或交互作用的現象，因此設計將樣品稀釋至風味化合物閾值的濃度時來觀察，看每一個描述用語可以獨立呈現的狀況或其強度，但是此法顯然是以儀器分析的概念設計，而將感官評估的特點忽略了，人之所以與儀器不同就是有“整合”與交互作用的觀念，因此許多專家學者認為研究食品的風味，應該在其正常食用的狀態與濃度下才有意義，而對此方法提出批評，也許在對產品進行基礎研究時，此方法仍有應用之處，因為有些產品中具有風味特別強的化合物，而蓋過其他的風味物質，若想了解這些強度低的化合物在產品中的角度或貢獻，或者想了解這些化合物於不同濃度的影響，此方法仍有其可取之處，但在應用研究上，目前較少人採用此一方法。

10.4.8 標準差離描述分析法

此方法是由學者 Rose Mary Pangborn 於 1967 年提出，其發展描述與訓練品評員部份皆與其他方法相同，但其標示法的精神則融入參考比較法的概念，此法乃針對每一個發展出來的描述語提供一個標準品，將標準品的感官強度定為 0，品評員評估樣品並與標準樣品的強度比較後，與標準樣品的感官強度相同者定為 0，較標準樣品強的以+1、+2、+3…表達，而較弱者則以-1、-2、-3…表達，數據經統計後，以條狀圖方式表現樣品間的整體差異。許多蔬果類的研究皆採用此方法進行描述測試。

10.5 描述性分析品評小組數據的應用

描述性分析數據的產品數據是理解產品資訊的多元來源，不管是在政府和學術研究機構或企業行銷上。描述性分析結果為尋求在特定產品上的感覺特性的專業人員提供指南。這些結果可用於：

1. 產品的感官特性的紀錄文件。這是描述性分析數據的主要用途。在每一次產品的詳細屬性和屬性強度方面的描述性測試，提供了此產品在香氣、風味、外觀、質地和(或)聲音的文字記錄、輪廓圖或剖析圖。每個描述是每個產品唯一，並且被認為是該樣品的藍圖。在表 10.2 中，將三種橙汁的描述並排顯示，以展示數據細節和實際產品的關係。表 10.3 提供了商業橙汁的完整描述，某些屬性的評分為零。同樣的產品輪廓也以圖形的方式顯示在圖 10.2。

表 10.2 對比鮮榨柳橙汁、冷凍濃縮柳橙汁，和罐頭柳橙汁的風味描述輪廓

	鮮榨柳橙汁	Minute Maid® 冷凍濃縮柳橙汁	罐頭柳橙汁
氣味			
綜合柳橙味	9.5	7.0	4.0
生味	6.0	1.0	0.0
熟味	0.0	5.0	4.0
蒸餾橙油	0.0	0.0	2.0
手榨橙油	3.5	2.0	0.0
果香／花香	4.0	0.0	0.0
其他柑橘類	2.0	1.5	2.0
類型	柑橘	葡萄柚	葡萄柚
強度類型	1.0		
烯			
其他類型	0.0	1.5	
熱帶鳳梨、香蕉	2.0		
甜味(焦糖/麥芽 酚)	0.0	0.0	0.0
青草味	1.0	0.0	0.0
維他命	0.0	0.0	0.0
紙板／氧化	0.0	1.5	. 0.0
水解油	0.0	0.0	6.0

發酵	0.0	0.0	0.0
煙燻味／苯酚	0.0	0.0	0.0
紙張／明膠	0.0	0.0	0.0
基本口味：			
甜度	8.0	7.5	7.0
鹹度	0.0	0.0	0.0

以下屬性不存在此範例中，故強度為零：果香／花香；焦味；水解油；蒸餾的橙油；紙/明膠；煙燻味/苯酚；發酵；維他命；綠色；甜味；金屬味；化學（穩定劑）；鹹度。

表 10.3 Minute Maid®冷凍濃縮橙汁完整氣味及化學感覺因子輪廓

屬性	強度
氣味	
綜合柳橙味	7.0
生味	1.0
熟味	5.0
蒸餾橙油	0.0
手榨橙油	2.0
果香/花香	0.0
其他柑橘	1.5
類型：	葡萄柚

其他水果	1.5
類型：	熱帶
甜味（焦糖/麥芽酚）	0.0
青草味	0.0
維他命	0.0
紙板／氧化	1.5
水解油	0.0
發酵	0.0
煙燻味／苯酚	0.0
紙張／明膠	
基本口味：	0.0
甜度	7.5
鹹度	0.0
酸味	4.0
苦	0.5
化學感覺因子	
澀	2.0
焦味	1.0
化學（穩定劑）	0.0
刺癢	0.0
餘味	
金屬味	0.0

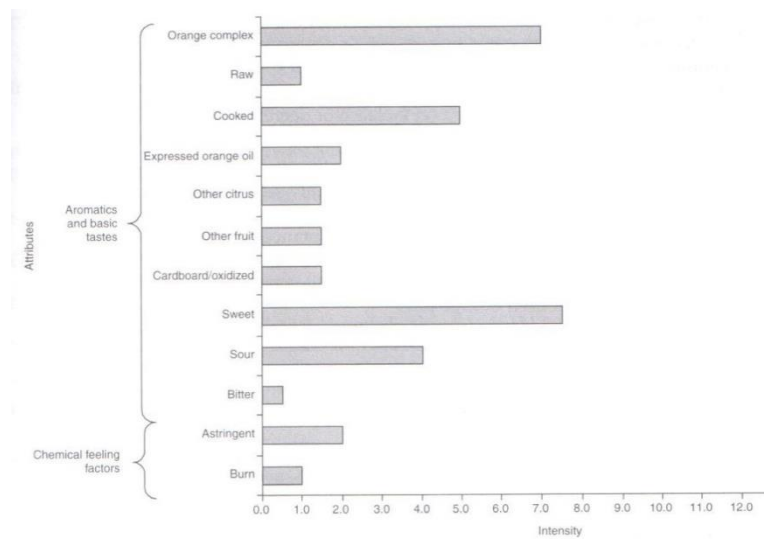


FIGURE 10.2
Minute Maid® frozen concentrate complete descriptive profile.

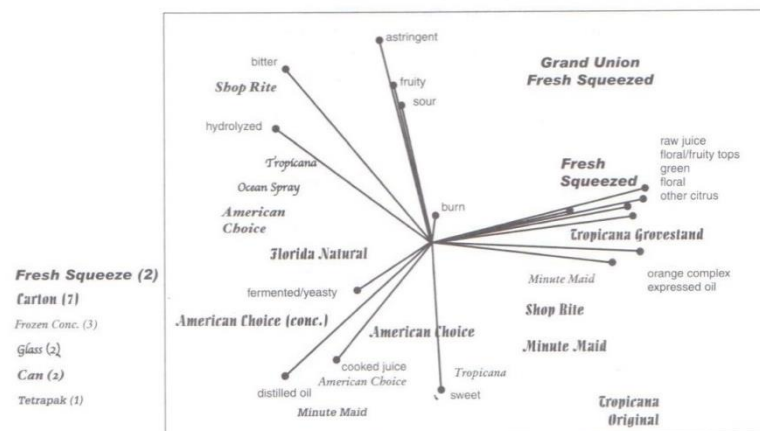


FIGURE 10.3
PCA map of orange juice products.

圖 10.2 Minute Maid®冷凍濃縮橙汁完整的描述輪廓

Aromatics and basic tastes 氣味及基本風味

Chemical feeling factors 化學感覺因子

Orange complex 綜合柳橙味

Raw 生味

Cooked 熟味

Expressed orange oil 手榨橙油

Other citrus 其他柑橘

Other fruit 其他水果

Cardboard/oxidized 紙板／氧化

Sweet 甜味

Sour 酸味

Bitter 苦味

Astringent 澀味

Burn 焦味

Intensity 強度

圖 10.3 橙汁產品的 PCA 地圖

鮮榨（2）

紙包裝（7）

冰凍濃縮（3）

玻璃瓶裝（2）

利樂包（1）

品牌名：Shop Rite, Tropicana, Ocean Spray, American choice, Florida Natural, American choice(濃縮), Minute maid, Tropicana grovestand, Fresh squeeze, Grand Union Fresh Squeeze

2. 產品屬性的比較。這些感官特性的紀錄提供了商業決策，如建立消費者在特定屬性上的接受度所設立的品質管制的規範；基於高接受度產品的比較預測市場（表 10.2）；將產品在增高特定屬性或降低負面屬性的數值做出市場產品利基，進行廣告宣傳等。
3. 找出現有市場競爭者中的基準產品，描述基準點與消費者接受度相關，以定義了哪些產品屬性是決定消費者接受度、表現、利潤或虧損是類別評估計畫(Category Appraisal projects)的關鍵要素。圖 10.3 是橙汁分析地圖主要的要素，在比較兩種鮮榨橙汁來源在不同的包裝及儲存條件下的表現。

10.6 應用

10.6.1 茶葉品質官能品評作業

1. 沖泡方法：秤取 3.0 公克茶葉放入評茶杯，沖入沸騰之開水約 150 毫升，加蓋靜置 5~6 分鐘，將茶湯倒入評茶碗中供湯質品評，茶渣仍留於杯中供香氣品評。

2. 評茶項目：分為茶外觀（形狀、色澤）、湯質（水色、香氣、滋味）及葉底等，各項目品評標準會因茶類不同而異。

(1) 外觀品評：審視茶葉之外形、條索、色澤、芽尖白毫，及副茶與雜夾等。

(2) 水色品評：審視茶湯之顏色，湯液是否明亮、混濁、晦暗或帶雜色等。

(3) 香氣品評：判別香之種類、高低、強弱、清濁、純雜以及是否帶油臭、焦味、煙味、菁臭味、霉味或其他異臭。

(4) 滋味：品評茶湯的濃稠、淡薄、甘醇、苦澀及活性、刺激性、收斂性等。

(5) 葉底：審視茶葉開湯後茶渣的色澤、葉面展開度、葉片芽尖是否完整無破碎等，並可做為判別茶菁原料品種、老嫩、均一性及發酵程度是否適當之參考。

3.評茶設備：

- (1)評茶室（評茶時採自然光，必要時可裝設日光燈補充光線，並應有空調設備）。
- (2)評茶台（黑色台面，表面防水處理）。
- (3)秤量器（感度至少達 0.01 公克）。
- (4)計時器或砂漏（計時最小單位為秒；至少可計時 10 分鐘）。
- (5)白色磁製評茶杯（容量 140~150 毫升）、評茶碗（容量 200 毫升）至少各 10 組〔參照 ISO3103-1980(E)所訂規格〕。
- (6)其他評茶用具如評茶匙、網匙、煮水器、茶渣桶等。

第十一章 喜好性測試：消費者測試及內部品評小組接受度

測試

11 Affective Tests: Consumer Tests and In-House Panel Acceptance Tests

喜好性測試的受測者應使用非訓練型品評員，但並不是受測者不須經過篩選和學習，實際上不論是使用組織內員工、社區人士或街頭消費者進行產品的喜好性測試，皆應符合兩點原則：一為受測者應能代表產品的消費者/使用者群，若為新問世產品，也應篩選潛在消費者，否則非產品消費者的數據會干擾結果，也浪費成本；另一原則為同一受測者應避免在短時間內接觸類似產品三次以上，以避免學習效應。

11.1 目的及應用

喜好性測試的主要目的是評估目前或潛在使用者針對某個產品、某個產品創意或某個產品特性的個人反應（偏好或接受度）。

喜好性測試主要由商品製造商(如食品製造商、餐廳)所使用，但有時服務提供商(如市調公司)也需要使用，如醫院、銀行或者是軍方等建立許多第一次測試的地方。消費者測試每年變得愈來愈常用。為了提高銷售量或售價而設計產品及服務，它們被證實是有用的工具。生意好的公司往往具有優良的消費者測試知識，藉此他們也更加了解其消費者。

本章在設計消費者測試及內部喜好性測試上提供大致的指導。許多著作有更詳細的討論。區分這些著作的方法是是否接受使用內部品評小組的接受度測試。本章基於測試產品類別，適當接納使用內部品評小組的意見。紅酒上不依賴消費者測試，但是某些食品則需要。對於一般公司的產品來說，意圖改變或是程序和材料無法避免的改變所產生的測試數量超過消費者測試的可能時，將大部分的測試工作轉向內部品評小組就變得較為合適，並且經過校準後盡可能地接近消費者測試。

消費者測試的應用變得前所未有地廣泛。隨著市場持續的變化、利基產品以及逐漸增加的差異型消費者，要預測消費者的喜好變

得更為困難。這些市場的改變增加了進行消費者測試的重要性，消費者研究是昂貴的，每年增加 5~10% 的成本。這促成了許多可能的選項，且新的替代方案正突飛猛進地開發中。大部分人在今天都以某種形式參加了消費者測試。一項典型的測試需要 100 到 500 個目標使用者，分散在三至四個城市中。目標使用者代表的是此產品欲銷售的客群。質性或量性的測試應用正在擴張，包括內部品評小組、家庭使用測試（Homeuse test）、焦點族群（Focus group）以及線上研究。研究者有責任保證測試是妥善且有成本效益的。

適當的消費者研究，包括在產品產量漸增前市調的篩選測試，評估未上市新產品的替代原料或是降低成本的可行性，或是取得產品開發的遠見。不適當的測試包括來自簡陋的測試系統（協議）或問卷調查、不正確的消費者篩選，或是結果的濫用：基於錯誤的產品檢測，在錯誤的時間進行測試，或者使用測試作為市場調查研究的替代案。

在碳酸飲料的一個典型的消費者研究的例子是，在兩個城市中招募過去兩周內曾在便利商店買過碳酸飲料的 18-34 歲的男性。透過面試、電話或在商場中篩選潛在的受訪者。提供這些被選中並願意參加的候選人各種飲料與記分卡，詢問其健康、喜好和參加的理由，並調查以往的購買習慣和各種基本資料，如：年齡、收入、職業、種族、背景等資訊。計算各個子群中整體偏好的分數結果。

當發現真正的遠見時，消費者測試才會變得更有價值。所謂的遠見是指以前所不知道的見解，可以從量性及質性的科學研究中找到其變化多端的型態。例如：遠見可能是孩子打開花生醬罐子將花生醬抹在一片吐司上的方式，或是從一杯熱咖啡上看見蒸汽而有了新鮮的感官訊號。從來沒有聽說過的新資訊，或是不同的數據被合併組合的方式都是構成遠見的來源。

近年來在消費者感官研究領域的一個令人振奮的變化，包含了在產品開發過程的**模糊前置作業**（Fuzzy Front End）。模糊前置作業提估新產品問世的機會，並在開發過程的早期揭露消費者的需求和保持產品的動態。模糊前置作業工作在之後幾頁會進一步地詳細討論。產品的全球化通常需要為不同的目標客群制定不同的研究設計。質性及量性方面的全球化研究近年來顯著性地增加。近年來，ASTM E18 的工作組正在制定跨越不同國家和文化的消費者研究的指導方針。

在偏好性或接受測試中，最有效的是根據產品及嚴格篩選的品評人員所精心制定的測試協議。品評人員和測試協議的選擇是依據計畫的目標。沒有根據關鍵計畫目標而設計的消費者感官測試往往增加 \$ 10,000 到 \$ 100,000 美金的成本，甚至更多。內部品評測試通常是昂貴的；例如涉及 20-40 人，20 分鐘的測試時間，計算工資和開銷的總成本可能達到 \$ 400- \$ 2000 美金。

從計畫的角度來看，進行消費者測試的原因通常分為以下幾類：

- 產品維護
- 產品改進/最適化

- 新產品開發
- 市場潛力評估
- 產品類別審核
- 廣告宣傳的支持

11.1.1 產品維護

在一家典型的食品公司中，處理產品的維護、市場佔額和銷售量等大部分的工作，是透過研究開發部門和行銷部門的共同合作。研發計畫可能涉及降低成本、替代原料、加工過程和配方的改變、包裝的變化等。每個計畫案都在不影響產品總特性和產品接受度的前提下進行。在這些案例中的感官測試通常都是區別測試，以評估產品的差異性或相似性。然而，當沒有能夠匹配目標的結果時，有必要選擇出一個或多個「幾乎符合」的樣品出來，以進一步確定樣品原型對於消費者來說，至少等同於當前產品（接受度或偏好度），或者競爭者的產品。

產品維護是在品質管制／品質保證，和儲存期限／儲存計畫中的一個關鍵因素。最初，建立消費者對於標準或對照組產品的「喜好狀況」是必要的。完成此作業後，在品質控制或儲存測試的幫助下，內部測試有助於決定隨時間所變化的測量幅度或類型、條件、製造地點、原物料來源等。通過內部測試檢測出不論大或小的感官差異，接著可以進行消費者測試，以評估相較於標準產品或對照組產品，有多少的差異可以減少（或增加）所測試產品的接受度或偏好度。

11.1.2 產品改進/最適化

由於貨架空間的激烈競爭，企業也不斷地尋求改善及最適化其產品，以求提供比競爭對手更好的產品給消費者。產品改進計畫一般希望「修改」或升級一到兩個消費者反應需要改進的關鍵產品屬性。產品最適化計畫則通常藉由控制幾個成份或生產過程的變數，進而提高所需的屬性和整體產品的消費者接受度。這兩種類型的計畫需要描述型的品評小組，以（1）驗證最初的消費者需求，（2）成功的原型樣品的特性記錄及歸檔。改進產品屬性的計畫例子如：

- 增加關鍵香味和(或)風味屬性，如：檸檬、花生、咖啡、巧克力等

- 增加重要的質地屬性，如：酥脆、濕潤等，或降低負面屬性，如濕軟或發白。
- 改進已感知的屬性（例如：鬆散乾燥的質地、腐敗味、人工而非自然水果味）
- 提高感知的特性表現，如更持久的香味、光澤更明亮、更滋潤皮膚等

在改進產品的過程中，製成的原型樣品，由一個描述型或屬性型品評小組進行測試，以確定預期的屬性差異是顯著的，接著再進行消費者測試，以確立感知的產品改善程度，及消費者對產品總體接受度或偏好分數的影響。

至於產品最適化，原料或製程變數經微調，受影響的關鍵感官屬性經描述分析驗證，接著進行消費者測試，以確定消費者是否感受到屬性的變化，以及這種變化是否提高了整體的評級(overall ratings)。

改變屬性的研究加上消費者的評分，使得公司能夠鑑別和理解那些影響市場接受度的屬性、成分或製造過程的變數。

11.1.3 新產品開發

在典型的新產品開發週期中，幾個關鍵時刻需要進行喜好性測試，例如：焦點小組(focus group)評估概念或原型；將測試產品呈現給消費者測試，使他們可以見到或觸摸的可行性研究；在產品開發階段中進行集中定點測試法 (Central location test, CLT)，以確立產品特性確實提供了預期的競爭優勢；在測試市場期間與競爭者的控制比較；在付諸實行的過程中更新比較，以確認預期的產品特性存活到量產時期；最後，在成長階段進行的集中定點測試法和家庭使用測試 (Home use test, HUT)，以確定試圖趕上競爭者的成功程度。

根據每個過程中的測試結果，以及研發人員重新制定且在各階段產量漸增(scale up)的能力，新產品的開發週期可能是幾個月甚至是幾年的時間。此方法需要使用數種不同的喜好性測試，例如：第一個概念的反應、被選擇的概念樣品與原型樣品的比較、不同的原型樣品、原型樣品與競爭者的比較。在任何給定時間的開發過程中，測試目標可能類似於產品維護的計畫，例如：中間工廠(pilot plant)產量漸增、或產品最適化的計畫，如上所述。需要快速開發原型品的時機是在當上市時間短，而且迫切需要做產品的決定時。利用與目標消費客群持續頻繁接觸的方法，即時收集反饋。提供反饋給產品的開發者，然後迅速改變，並再次提供給目標客群以收集進一步的反饋，包括小型的 CLT 研究或質性研究都利用此方法進行測試。

11.1.4 市場潛力評估(Assessment of Market Potential)

市場潛力的評估通常是行銷部門的工作。當他們在設計問卷問題（如描述產品差異的關鍵屬性）時，會諮詢感官評估部門測試方法和之前感官評估的數據文件等。關於購買意向、購買價格、目前購買的習慣、消費者的飲食習慣等問題，以及包裝和廣告的效果、便利性的作用都是品牌產品的接受度的關鍵。感官分析師的主要功能是引導研究開發。無論感官分析師是否應該指引市場導向的消費者測試問題，行銷部門提供這樣的數據的能力和感官分析師承擔評估市場條件責任的能力，都是公司組織中的一項功能。

11.1.5 產品類別審核(Category Review)

為了要了解其品牌在競爭組合中的定位或鑑別其在一項產品類別中的區域(商機可能存在)，因而公司想要了解產品類別時，建議使用產品類別審核(Category Review)。產品或原型品的最大排列(可定義或涵蓋類別)的描述性分析可得出一產品類別圖(category map)。使用多元分析技術，產品和屬性的相對位置可使用圖形格式顯示（見第 14 章）。這使研究人員得以瞭解（1）產品和屬性群集(cluster)如何在產品／屬性空間散佈（2）哪裡有新產品的機會（3）哪種產品的最佳定義是哪些屬性。

針對消費者，相同產品的幾種額外的測試可以允許其他面向(vectors) 投射在空間。這些其他的面相可能代表了消費者總體喜愛和/或消費者的整體用語，如濃密的、豐富的、新鮮的、柔軟的。在不同產品上，消費者族群表現出一致性的喜好度也是有可能的。在每一區段(segment)內連結的消費者可用研究中的產品類別的關鍵描述性屬性加以區分。

11.1.6 廣告宣傳的支持

在印刷刊物、廣播、電視、網站上宣稱的產品和服務廣告都需要有實際的數據支撐。宣稱感官相似性（如：嚐起來跟領導品牌一樣好）或優越性（如：比領導品牌更乾淨）都需要以消費者研究和(或)品評小組測試的數據做為根基以使這些宣稱具有說服力。

11.1.7 發掘消費者的需求

發掘新產品機會，或為符合消費者需求而改進產品，是感官評估正在發展的領域。基於概念和產品需求的更佳目標定義及較強的企業案例(建立在事實及消費者導向資訊)，了解消費者對產品闡明和未闡明的需要、想要、願望和行為，引導產品經設計而符合特定需求。用來收集這些資訊的技術通常專注於質性測試，如觀察研究、一對一面談、販售點訪談、日誌、家訪等。這些方法的應用揭露

了此產品類別是如何被「真正地」使用，並確立包括「一定要有」(must have) 和「有也很好」(nice to have)特點的關鍵感官特性。

11.2 喜好性測試中的測試對象／消費者

11.2.1 取樣及人口統計特徵(Demographics)

如何從廣大的人群中取樣是喜好性測試的最重要事項。不管感官測試時是何時執行，一組測試對象被挑選出代表一些更大的群體，感官分析師藉此希望能夠得出一些結論。在一些區別測試中（異同性測試和描述性測試），感官分析師挑選出偵測能力平均或高於平均的候選人作為**測試對象樣本**（sample subject）。假設這些個體無法偵測出產品的差異，則代表更大的族群也將無法偵測出差異。然而在喜好性測試中，從廣大的人群中挑選候選人是沒有效率的。符合目標客群的產品或服務的消費者、選擇的市場或是小心切分出的族群，這些條件要求感官分析師首先必須定義此產品的目標客群；例如：對於一個較甜的早餐麥片，目標客群有可能是4至12歲的兒童；對於壽司或者優格，目標市場可能是南加州；而對於高價的珠寶品、衣物、或是汽車，區隔的人口族群可能是年輕人，25至35歲，事業正值上坡，已婚或未婚。

面對在鑑別和使用消費者(以非常精確的人口統計學模型代表目標客群)樣本的需求方面取得平衡的任務，消費者研究人員必須使用**篩選器(screener)**。妥善的篩選器需要思考，並輸入來自消費者研究員及使用者的產品研發和市場研究等數據。從篩選過程中得出的資訊能夠幫助確認候選人是否與廣大客群存在異同，以及這些相似性和差異性隨後是否影響產品偏好或購買決策。有效的篩選從清晰地認識計畫目標開始。質性和量性測試的篩選可能有各自不同的詳細標準，然而仍有一系列通常會被提問的問題，如年齡、性別、打工／職業、種族、收入、一般習慣、敏感性、時間可配合度、參加意願等。在一般性產品如麥片、飲料、啤酒、餅乾等，消費者測試的研究指導可能只需要選擇此產品類別或此品牌的使用者或潛在使用者。隨後的消費者研究指導或市場行銷研究測試要求更嚴格的人口統計特徵是合理的。選擇測試對象樣本需要考慮的人口統計特徵有：

使用族群：基於不同人口族群在消費或使用產品速率的差異，品牌經理通常將之歸類為輕度使用者、中等使用者或重度使用者。這些條件在很大程度上取決於產品類型及其正常消耗速度（見表11.1）。對於不斷變化的產品，如電子用品，前端使用者（lead user）將會在新產品概念上提供最有用的資訊。前端使用者比起普通人更早發現需求，並且試圖填補這一需求。另一個目標客群包括不滿的使

用者，可能由於在市場上找不到一個更好的替代產品而使用本產品。市場研究人員可能從此群組中找出一些創新的概念。對於特色產品或較不吸引人群的新產品，消費者測試的成本從根本上增加，因為必須聯繫很多人才能發掘適當的測試對象樣本。

年齡：4-12 歲是測試玩具、糖果和穀物最好的年齡；12-19 歲的青少年買衣服、雜誌、零食、飲料和娛樂。20-35 歲的年輕人在消費者測試中最受關注：（1）由於人口數；（2）由於缺乏家庭成本而有較高的消費能力；（3）由於終生的習慣和品牌忠誠度都在這個年齡段形成。35 歲以上，買房子和養家的消費者；65 歲以上的人口則由於照顧健康的關係，選擇產品的眼光變得挑剔。

表 11.1 不同消費產品的典型使用頻率

使用者分類	咖啡	花生醬、空氣 清新劑	通心粉產品和 起司	地毯清潔劑
輕度	最多一杯／天	1-4 罐／月	兩個月一包	1 次／年
中等	2-5 杯／天	1-6 罐／周	1-4 包／月	2-4 次／年
重度	5 杯／天	1 罐以上／天	2 包以上／周	1 次以上／月

如果產品具有廣泛年齡的吸引力，例如碳酸飲料，應該按照各年齡層的客群比例選擇受試者。

性別：雖然女性較常購買消費品、服裝，男性較常購買汽車，酒類和娛樂，兩性之間的生活習慣和購買差異正不斷縮小中。研究人員若要將產品類別按照性別區分，如：方便食品、零食、個人護理產品和酒等，應該使用非常近期的數據。國外學者建議在食品類測試時，選擇受測者女性 60%及男性 40%，因為通常女性為購買者（及使用者），但這並不一體適用所有產品。

收入：對大多數產品有意義的人口收入的區分是：

- 低於\$ 20,000
- \$ 20,000- \$ 40,000
- \$ 40,000- \$ 70,000
- 超過\$ 80,000

根據產品有時也會有不同的收入區分方式，例如超過 50 英尺遊艇的產品，使用\$ 200,000、\$ 300,000 的區分。

地理位置：區域差異也造成許多產品的偏好差異，例如：在美國，產品通常需要測試一個以上的地點，並避免在有類似偏好和類似人口組成的區域測試（或使用比例測試）產品，如紐約、南加州。此外，注意城市、郊區和農村也會影響測試結果。

國籍、地域、民族、宗教、教育、職業，以及其他因素，如婚姻狀況，家中兒童的數量及年齡、寵物、住所的大小等，可能對某些產品或服務的採樣非常重要。產品研究人員、品牌經理或感官分析師必須仔細考慮所有的人口統計數據後，界定目標客群，為既定的測試選擇測試代表。通過行銷研究人員篩選後，對受訪者一步步地進行調查。

11.2.2 測試人員來源

消費者測試需要從使用該產品的人口中抽樣。有三種來源的人可以被選為參加此測試：員工、被招募進測試人員庫的當地居民，以及一般民眾。

11.2.2.1 員工

為了從消費群體中正確抽樣，原則上採用公司辦公室、技術部或工廠的員工和當地居民的選項通常是被排除的。然而，由於消費者測試成本高以及周轉時間長的緣故，有時候公司在部分的喜好性測試中會選擇使用員工和當地居民做為測試代表。

若測試目標是產品維護，直覺上使用員工和當地居民並不會帶來很大的風險。在目標是維持感官整體協調的計畫中，請員工和當地居民作為產品固定使用者來評定產品的整體特性是不錯的。在這樣的情況下，請員工和當地居民可經由已知的標準品或對照品，判斷對待測樣品的接受性或喜好性的相對差異。

僅限於產品維護以及正確使用的情況，員工可以算是喜好性測試的一個寶貴的資源。由於他們熟悉產品和測試流程，員工可以在相同時間內處理更多的樣品，更準確地區別差異，提供更快的回覆，更便宜的服務。員工接受性測試可以在實驗室中或集中地點進行測試，或是可以讓員工帶產品回家測試。然而，若在新產品開發、產品最適化或產品改善的階段，員工或當地居民不應該被用來代表消費者。以下是使用員工進行喜好性測試可能會產生偏見的一些例子：

1. 員工往往會找理由喜歡他們及同事製造的產品，或者假若他們感情不好，則會找理由盡可能拒絕此新產品。因此，掩飾測試產品是必要的。如果這是不可能的，必須使用消費者作為品評員。

2. 員工可能無法使用與消費者相同的方式綜合評量產品特性的好壞。例如：員工可能知道，最近的製造過程中改動了顏色，這可能使他們更注意產品的外觀，並對其他產品特性給予太少的權重。因此，在這種情況下，必須掩飾產品的顏色，如果無法掩飾，必須進行外部測試。

3. 假如公司針對不同市場生產不同的產品，外部測試將被配發到目標客群中，但這不能使用員工來完成。如果需要以員工進行測試，建議告訴他們該產品是注定在 X 市場，但有時這無法在不違反盲測的原則來完成。如果是這樣，則必須使用外部測試。

11.2.2.2 當地居民

另一種消費者測試的招募方法是公司發展自己的當地居民數據庫。這種方法雖然比較符合成本效益，仍需要公司內部支持開發及維護。要小心不要過度使用在數據庫中的消費者。機密性可能難以保持，特別是當此測試是安排在製造現場。決定哪個計畫階段可以使用消費者數據庫是很重要的，以確保所收集的資訊是符合計畫目標的。

總之，測試組織者必須有想像力地計劃測試，且必須注意所有偏差的來源。此外，回應的有效性必須通過與使用同一樣品在更廣泛的消費人群上進行的消費者測試的頻繁比較來確認，才比較安心。這樣一來，組織者和小組成員慢慢建立市場需求的知識；這一點，隨後可以更容易地衡量陷阱並避開它們。

11.2.2.3 普通民眾

普通民眾的消費者測試，傳統上是指收集符合預先定義的目標使用者的群體所得的回應。這些受訪者最常從數據庫中挑選，並直接聯繫參與測試，或者他們是集中招募，如購物商場。這種方法的優點是產品的使用者或潛在使用者的測試能力；缺點是增加招募的成本。建立適當的篩選標準以消除過於專業的評估是很重要的。

11.3 測試地點的選擇

測試地點或測試網站對結果的影響很多，不僅是因其地理位置，還因為進行測試的地方會影響樣品許多方面的產品取樣和可感知感官特性。同樣的一組樣品和測試人員，有可能因為測試地點的不同，而得到不同的測試結果。這些產生結果差異的地方發生在：

- 產品使用／測試的時間長度

- 受控制製備 v. s. 正常製備的產品
- 在集中定點測試單一產品 v. s. 與其他產品一同測試，或是在充斥個人用品的家中測試
- 在家測試時其他家庭人員的影響
- 問卷的長度及複雜性

通常喜好性測試的型態可分為實驗室測試(laboratory test)、集中定點測試(central location test, CLT)及家庭使用測試(home use test, HUT)三種，後兩者為所謂的消費者測試(consumer test)。在選擇何種型態時，首先最主要的考量因素仍是目的，因為不同的測試型態提供的數據不同，若以研發新產品的步驟而言，通常在篩選及測試原型產品時，可以使用較低成本的實驗室測試法，準備量產上市時則應進行集中測試或家庭使用測試。

11.3.1 實驗室測試

在實驗室測試的優點是：

- 產品的製作和呈現可以被小心地控制。
- 可以在短時間內聯繫員工參加。
- 在一個原型樣品的製造過程中，顏色和其他視覺方面等無法充分控制的差異可以被掩蓋，使得受試者可以集中測試和評定味道或質地的差異。

實驗室測試的缺點則是：

- 在公司或特定工廠的位置測試暗示了測試產品的來源，和在此地的過往經驗都可能影響測試人員產生預期和偏見。經驗和產品等知識往往會導致差異敏感性增加。這些感知到的差異反應可能不會反映在目標客群上。
- 缺乏正常消費體驗，例如：啜飲測試 (Sip test)，而非完整的消費體驗，可能影響正面或負面屬性的檢測或評價。
- 標準化的準備程序和產品處理協議不可能完全模擬在家的消費行為和經驗。

11.3.2 集中定點測試 (Central location test)

集中定點測試通常在潛在購買者聚集，或可以集合的地方進行。主辦者在商場、教堂，或測試機構設置一個攤位，或租用一個房間。學童使用的產品可能會在學校操場上進行測試；分析化學家使用的產品可以在專業會議室進行測試。在開放地點攔截受試者並進行篩選，接著帶選定的候選人到密閉空間內。也可以通過手機預先篩選受試者，並邀請到一個測試場地。通常情況下，每個位置收集 50-300 個回應。在受試者看不到的地方準備產品，並使用三位數編碼的相同托盤（紙杯、玻璃杯）呈現樣品。受試者分心的可能性很高，所以測試說明和問題應簡明扼要；在附錄 11.3 提供評分表的例子。另一種變型的測試過程是，將產品從原包裝公開取出並分配，然後提供受試者此產品的廣告及市場定位的例子。

集中位置檢測的優點是：

- 受試者在主辦方控制的條件下測試產品；任何誤解可以馬上解釋並獲得更真實的反應。
- 產品由最終使用者自行測試；這確保了結果的有效性。
- 此條件有利於從大型人口群體中抽樣調查中得到較高回應率。
- 可以在一次消費者測試中測試多種產品，因此可以在測試單個消費者的成本下得到大量訊息。

集中定點測試的主要缺點是：

- 比起在家中，或在聚會、餐館等的正常使用，集中位置的樣品準備方式、消費數量和消費時間長度等方面的測試條件都較不自然。
- 相對於在家中測試，集中位置可以詢問受試者的問題較為受限。因此又限制了對於不同的年齡組別、社會經濟群體等的喜好資訊。

11.3.3 家庭使用測試 (Home use test)

在大多數情況下，家庭使用測試（或家庭布置測試[Home placement test]）代表著消費者研究的極致。產品是在正常使用條件下測試。參加者是作為代表的目標客群被選擇。考慮家庭成員的互相影響，因而能夠獲得整個家庭的意見。除了產品本身，家用測試在適用的情況下，提供了關於產品包裝的確認點以及準備產品的說明。典型的品評測試大小為，在每 3 或 4 個城市收集 75-300 個回應。通常情況下，進行兩個產品的比較。在第一次測試完成且相應記錄表也遞交後的 4-7 天，進行第二次產品的測試並評分。兩個產品不

應該同時提供，因為有可能遵循錯誤的指導，或記錄到錯誤的評分表，而使得測試在錯誤的基礎上進行。附錄 11.3 提供了一個評分表的例子。

家庭使用測試的優點是 (Moskowitz 1983; Resurreccion 1998)：

- 產品的準備和使用是在自然條件下進行。
- 比起在商場攔截測試，受試者對於產品的資訊或偏好較為穩定（重複），而不是單次的第一印象。
- 從受試者重複使用產品的累積效果，可以提供潛在重複銷售的相關資訊。
- 統計的抽樣方案可以被充分利用。
- 因為有更多的時間完成記錄表，可以收集消費者對於產品的各種特性，包括感官屬性、包裝、價格的態度等多種資訊。

家庭使用測試的缺點是：

- 一次測試非常耗時，約須一到四週完成。
- 比起集中定點測試，每次家庭使用測試的規模小很多；聯繫許多住宅是冗長和昂貴的。
- 沒有回應的可能性較大；除非經常提醒受試者，否則他們經常容易忘記自己的任務；當測試接近尾聲時，他們可能給予雜亂無章的回應。
- 最多可以比較三個樣品；任何較大數量的測試會擾亂選擇家庭使用測試的初衷：自然的使用情況。因此，多樣樣品的測試，如最適化和類別審核 (category review)，不適合使用家庭使用測試。
- 樣品被錯誤準備的容忍度。從準備樣品的差異性，到測試產品的時間，以及搭配使用的其他食品等差異，都可能使較小抽樣的結果產生極大的變異性。

11.4 喜好性測試方法—模糊前置作業 (Affective Test Methods -Fuzz front end)

在喜好性測試方法中，可分為偏好性測試 (preference test) 及接受性測試 (acceptance test) 兩大類。

11.4.1 定義、目的、結果

揭露消費者需求經常發生在初期的時候，也就是在模糊前置作業中。通常研究是在計畫的早期階段進行，在計劃正在進行時，正在評估最初的市場和技術上的可行性，並且突破性創意正在探索。模糊前置作業的研究在金錢投入詳細的技術評估、成本概念測試，大量人力投入之前進行。這並不意味著這些投入到掌握消費者的早期工具和技術不能在產品開發過程的所有階段使用。使用的方法非常特別，因為它們深入地收集消費者的資訊，他們如何及為何使用此產品，他們真正喜歡或不喜歡的特性。為了捕捉這種級別的資訊，必須使用超越標準以及常用的量性和質性的方法。

11.4.2 應用

在模糊前製作業的研究中允許：

- 探索消費者購買產品的原因，是特定產品特色或可區別的感官特性
- 研究產品的功能性和人體工程學
- 探索消費者是如何修改產品或改變使用習慣，以適應其需求
- 發現文化中的態度、行為和動機
- 通過觀察研究消費者在其自身的環境中的行為

11.4.3 工具與技術

有許多方法及技巧可用於發掘消費者對於新產品的想法和意見。最常用的技術在本質上是質性的；然而，量性方法也是有效的。消費者可能在自己的家中、大街上、商店裡，或在購買行為發生時和花錢處被研究。去現場觀察消費者的研究通常稱為人種學（Ethnography）或沉浸法（Immersion）。沉浸到消費者的環境中，通過觀察和對話收集的資訊。比起傳統的技巧，從一對一單獨的訪談中獲取消費者資訊，模糊前置作業收集資訊的方法經常是以想像力作為基礎，並且包括但不限於心智圖（Mind map）、字彙網（Word-web）和拼貼法（Collage）。量性技術則超越了集中定點測試（CLT）和家庭使用測試（HUT），包括線上研究和內部/外部研究。線上研究提供了早期的探索態度和行為研究的概念。內部/外部研究探討產品的基本概念與外部動機。模糊前置作業使用網路研究的詳細討論請參

見 11.7 節。

如前所述，通過在特定背景中觀察消費者，更可能取得更深入的了解。這個在真實世界實施的方法有助於發現實際行為。在進行這種類型的研究時，有兩種不同的途徑可獲取資訊。一種方法是研究者作為一個觀察者，並沒有與消費者進行談話。另一種方法則是研究者作為一個觀察面試官，積極地與消費者互動，深入探測其有興趣的領域。觀察研究的兩個例子：

例 11.1：前往不同家庭中以觀察在提供午餐給孩子的過程中有什麼樣主要的照護行為，研究人員觀察他們使用什麼成分（麵包、肉、佐料、和起司），這些產品在什麼樣的儲存條件（廚房和冰箱），和如何搭配三明治（步驟），並且如何提供。多個家庭訪問會發現一些喜好行為的差異以及需求。

範例 11.2：觀察和採訪 18-21 歲、30-35 歲、60-65 歲等不同年齡層的女性，在選擇和購買口紅、光澤液，或粉底霜時，其年齡、生活方式和皮膚類型等的不同呈現什麼樣不同的喜好模式。進一步探討的資訊如使用哪種產品、將會詢問什麼問題、將會什麼顏色等。

另一種方法是，提供給受訪者日記本，使其在自己的環境將使用產品的過程中完成。消費者記錄使用過程，以及在使用產品的各個階段產生的情緒，從決策到購買和處置。隨時記錄產生的想法和步驟，使消費者不需要靠記憶力。因此，日記提供更完整的真實畫面。添加照片或影音使日記更趨完整。

社群敘述（Community narratives），也就是故事分享（Storytelling），是一種質性的研究方法，鼓勵創意的消費小組分享產品經驗和表達自己的感受。消費者使用他們自己的話描述經驗或情況。該組中的參與者通常是一個產品類別的使用者，而不是一個特定品牌的使用者。成熟和創新的質性數據收集方法使用在與目標消費群體具有長期合作關係的情況。雖然傳統的焦點小組只允許短時間探索代表目標客群的測試小組的感情和行動，故事分享則允許舉行多次測試，重複安排有創意、善於表達的消費者組成一個社群。這個過程引導更多誠實的溝通，增強創造力，並提高了目標產品類別和/或概念的洞察。

社群敘述的技術探索了超越表面的消費觀念、行為和情感，比起傳統的感官方法，讓研究人員得以了解更深層次的情感。鼓勵消費者表達自己的經歷和感受，讓動機和未闡明的需要被發現，並常常產生建立在其他成員的想法上的新見解。藉由文字和比喻練習的使用，故事分享的過程中因為允許群體成員自發地使用自己的描述方式，而使其更專注於研究人員的初步問題。了解消費者對產品的感官屬性的反應是一個重要的結果。

11.4.4 模糊前置作業研究的設計

模糊前置作業的關鍵步驟可以分解成框架，稱為 I-SIGHT。這是一個動態的框架，它可以在整個產品開發過程中使用。該框架的具體步驟是：

1. 創新：通過團隊發展激發創造力和創新力
2. 整合現有知識：總結所有已知和未知的事項，釐清事實和觀點
3. 確立目標：定義機會並設定研究目標
4. 經過精心設計的研究產生數據
 - a. 選擇或創建符合計畫目標的正確方法
 - b. 採取研究的路徑，準備階段
 - c. 收集資訊
5. 收穫想法：經過組織、排序和連結相關資訊的數據發現藏在其中的真相
6. 採取行動：確立移動到專案下一步的行動計劃

專案在此階段可以花費更多的時間設計研究，因為此階段比起標準的質性或量化研究更為客製化，或者相關概念正在研究中。在模糊前置作業的研究設計中，其他注意事項將在下面的章節中討論。

11.4.4.1 招募和篩選

最常使用的是符合特定招聘要求的消費者。很多時候選擇這些消費者，是由於其特定的行為、產品的使用習慣或需求。舉一個例子，製造商正在考慮開發一個新的外帶餐點，放在一個餐盒中而在超市中販賣，使所有家庭成員都能享用。產品的潛在使用者是成年人、青少年和兒童。每個人都有特定的偏好和習慣。理想的測試人員可以透過選擇傳統和非傳統的使用者、非使用者、不滿的使用者，和前端使用者。

11.4.4.2 研究地點選擇

選擇在何處進行研究，取決於產品類別和所需的資訊。前往此類產品供應的地方是必要的。如果目標是如何生產感覺像家庭自製

的冰淇淋，可以選擇前往冰淇淋店、雪糕攤、餐館和家庭。使用不同的位置提供了更廣泛、更複雜的角度來研究。飛機餐的研究需要去機場研究當前的產品。按摩油的研究則是去按摩店。

11.4.5 數據分析與挖掘

模糊前置作業的數據分析涉及轉化訊息為遠見、可能性和機會。模糊前置作業的數據主要是質性的，因此解釋方式是開放且不同的。相對於量性數據，其結果是文字，圖片，或帶有非量性圖表和統計分析的故事。重點是要記住，模糊前置作業的研究是探索性的，因此允許研究人員有數種不同的解釋。數據的轉化由擱置個人偏好開始，解讀所收集的數據並詢問問題，如：

- 共同點：比如說，在收集個人拼貼畫時，是否有一個潛在的主題？是否有一種或兩種顏色持續出現在整個收集中？是否有圖像、形狀或物體持續出現？不同的受試者的拼貼畫之間是否有共同的物體，他們是否使用類似的隱喻或字詞？
- 缺少的資訊：有什麼東西沒人提起？什麼被避免了？什麼事情讓人不願意討論？受試者如何替補這些事物？
- 感官屬性：什麼屬性最明顯或最常被提起？什麼屬性從未被討論過？什麼樣的產品屬性讓人失望？
- 有趣的聯繫：數據如何被連結？什麼樣的感官屬性被連結到關鍵的情緒？什麼樣的連結有意思？例如：在商場裡每個獨特的商店都有黑木和金色調；這種組合是否暗示獨特性？你可以藉由把數據放在一起而創造什麼不尋常的關係？

使用拼圖技術（Mapping technique）往往是有用的（如序列拼圖[Sequence mapping]，見 11.8.4）將收集的數據轉為可管理且有主題的組別。拼圖方法主要是組織技術。這些技術變得非常有用，並在完成組別分類後得到進一步的見解。從較小的數據組合內得出見解較為容易，更不用說變化不同的組織數據的方式是回答上述問題的機會。

11.5 喜好性方法：質性(Affective Methods: Qualitative)

1 1.5.1 應用

質性喜好性測試是透過某些方式（例如：訪談和焦點小組）使代表消費者的受試人員表達對產品特定屬性的主觀感受。質性的方法在下列情況下使用：

- 發現和了解消費者從未表達的需求；例如：「人為什麼要購買 4 輪車來開在柏油路上？」研究人員，包括人類學家和人種學家進行開放式的採訪。更多資訊請參見第 11.4 節。
- 為了評估消費者對一個產品概念和/或樣品的最初反應。當產品研究人員需要確定一個概念是否具有普遍接受度，或者相反，有一些明顯的問題時，質性測試可以讓消費者自由討論產品和/或一些早期原型品的概念。討論的結果或總結可以讓研究人員更充分地了解消費者對概念或原型的最初反應。專案方向可以在此時進行調整，以回應所獲得的資訊。
- 為了學習消費者描述一個概念、樣品、商用產品、或產品類別的感官特性時使用的術語。在消費者問卷調查和廣告的設計中，使用消費者端的用語，而不是來自市場行銷或產品開發的術語是極為關鍵的。質性測試，允許消費者使用他們自己的話公開討論產品的屬性。
- 為了釐清和擴大量性研究中的消費者反應。量性研究在確立消費者如何喜歡產品或對感官特性有何反應上是理想的。然而，它並不能充分捕捉細微差別或背後評級的原因。更深入的資訊，可以透過要求消費者在量性部分完成後留下來進行一對一訪談，或讓他們晚點參加焦點小組討論來獲得。
- 為了了解關於使用某產品的消費行為。當產品研究人員想確定消費者如何使用某些產品（包裝方面）或消費者在使用過程中如何動作（牙線、女性護理產品），質性測試可以探索原因，以及消費行為的做法。

在以下討論的質性方法中，需要訓練有素的訪談者／主持人。由於訪談者/主持人和消費者之間需要較高級別的互動，訪談者／主持人必須學會掌握小組動態、探詢技術、表現中立態度、總結和報告等技巧。

11.5.2 質性篩選開發 (Qualitative Screener Development)

開發篩選標準的最佳資訊來源是使用者端。除了在第 11.2.1 節所列的廣泛領域或類別外，一系列問題探索使用習慣、購買條件、對產品或成分過敏或敏感，以及概念接受性都需要詢問。由於質性研究的受訪者較少，制定具體的態度或使用標準是很重要的，以確保受訪者代表一個多元且符合關鍵標準的群體。質性篩選最關鍵的要素是消費者在小組討論做出貢獻的意願。訪談人員會詢問參與者的是否願意公開說出其意見及看法。顯然地，若此候選人不願意敞開心扉分享其感受，將不會是質性討論的正確選擇。最後一個選擇標準是應聘者有效表達思想和感受的能力。在訪談的最後詢問一個開放式的問題是典型測試此能力的方法。這個問題可能是：「如果你

能夠與歷史上某一個人見面，那會是誰？為什麼？」至於為了保障討論富有成效的，以及確保沒有人代替測試，在每個參與者抵達工廠時問一些其他問題。此項重複篩選的過程只需要幾分鐘就可以完成。

11.5.3 質性喜好性測試的類型

11.5.3.1 焦點小組 (Focus group)

在特定標準的基礎（產品使用習慣、人口統計等）上，選擇一組 8-12 位消費者與焦點小組的主持人見面 1-2 小時。主持人須掌握團體動態，藉由介紹各個小組成員，引導互動討論至會議的焦點，並盡可能討論具體的事項且發現盡可能多的資訊。

通常情況下，一個專案焦點需要舉行兩到三次這樣的會議，確認小組成員對產品和/或原型品的概念有任何整體趨勢的回應。除了整體趨勢外也需要記錄下一些獨特的反應。這些反應記錄，加上 DVD 或磁帶（音頻或視頻）的摘要將被提供給使用者端研究員。純粹主義者會說， $3 \times 12 = 36$ 的判斷太少，不足以代表任何消費潮流；然而在實踐中，任一個趨勢出現都是有道理的，應該基於此進行修改。而修改的事項可以在隨後的小組會議中或者量性研究中測試。

11.5.3.2 密集焦點小組 (Focus panel)

由焦點小組變化而來，使用相同的一群消費者訪談兩三次。目標是讓同組組員有一些初步的接觸，對主題開始進行一些討論，接著讓他們回家使用產品，然後再回來小組討論經驗。進行快速原型品開發時，此方法非常有效。它可以讓消費者參與產品的開發，並持續地提供反饋和方向。

11.5.3.3 迷你組，二人或三人組 (Mini group, Diads, Triads)

迷你組，二人或三人組是 8-12 人消費小組的替代方案。迷你組通常由 4-6 個受訪者組成，三人組含 3 位受訪者，二人組則含 2 位受訪者。這種方法通常使用在有必要深入探討一個特定主題時，但正在討論的主題過於敏感，或者是很難找到符合篩選標準的受訪者。通常遵循焦點小組的框架進行。

11.5.3.4 一對一面談

質性喜好性測試的一對一面談使用在當研究人員需要深入探討每個消費者，或其中的主題對於焦點小組太過敏感的情況下。這通常被稱為深度訪談 (in-depth interviews) 或 IDIs。在任何地方連續深入訪談 12 至 50 位消費者，每次訪談使用類似的框架，但深入探索每位消費者的回答。

這種方法的另一種獨特的變形是只採訪一個人，在集中定點測試現場或在消費者家中準備產品。將該過程記錄或錄影，然後之後與消費者對此記錄進行更多討論。探索消費者有關他們如何使用洗滌劑或準備打包的晚餐，其中產生了關於消費者行為的資訊，可能與公司原先預期的不同。

一對一的消費者訪談或觀察可以使研究人員發現未闡明或潛在的消費者需求，而這又可能引導出符合這樣需求的創新產品或服務。

11.6 喜好性測試：量性(Affective Methods: Quantitative)

11.6.1 應用

量性喜好性測試是一大群消費者（50 至數百人）就偏好性問題、喜好，感官屬性等的回應。量性的喜好性測試方法在下列情況下適用：

- 為了確定產品對代表其目標客群的受試者是否有總體的偏好或喜好。是否使用接受度和/或偏好度問題的決定將在每個測試方法中進一步討論。
- 為了確定產品在感官特徵（香氣、味道、外觀和質地）在廣義上的偏好性或喜好性。研究產品廣義的特色可以提供關於影響總體偏好或者喜好的因素的遠見。
- 要測量消費者對產品特定感官特性的反應。使用強度量度(Intensity scale)、喜好量度(Hedonic scale)，或是適合量度(Just right scale)產生的數據可以與先前所述的喜好評級(Hedonic rating)和描述性分析的數據連結。

11.6.2 量性喜好性測試的設計

11.6.2.1 量性篩選開發

與篩選質性討論小組相同，為了確定量性研究所需要的代表群體，了解研究目標是至關重要的。基於使用者提供的資料及產品類別的一般知識，可以建立篩選標準的條件。巧克力調味乳，應該針對的人口段是男孩（50%）和女孩（50%），5-10 歲（50%）和 11-16

歲（50％）。量性研究可能需要數天才能完成，原因有很多，比如說必須測量的樣本總數以及單次測試所能進行的數目。因此，在面試中，候選人必須同意參加，並願意來測試地點一次以上的測試以完成研究。為了避免候選人半途退出研究，告知候選人他們必須完成所有必需的測試才能得到報酬。

11.6.2.2 問卷設計

在設計喜好性測試問卷中，建議遵循以下原則：

1. 按照受試者預估測試的時間比例維持問卷的長度。受試者可能須按照合約花數小時測試接受數款產品的問卷調查。而另一個極端例子則是，某些專案只需受試者回答幾個問題便提供足夠的資訊。設計問卷調查應該詢問最少的問題實現專案的目標；進而構建測試，以便受試者預計空出適當的時間。
2. 保持清楚且類似的問題。在同一節問卷中使用相同類型的量度，無論偏好、喜好、適合度，或強度。強度和喜好性的問題可能出現在相同的問卷調查中（見附錄 11.3 實例），但應該明確區分。這些問題及其回答在調查問卷中的每一個部分都應該按照大致相同的步調。為了確保答覆一致性和準確性，每個屬性的強度問題應該都設計為相同方向，例如，[太少·.....·太多]，使受試者不必停頓理解每一個問題。
3. 每個產品在測試中的主要區別應該使用直接的問題。屬性的問題應該與那些在產品中可檢測及區分的屬性相關聯。這可以通過預先進行的描述性測試來確定。受試者將不會對他們不能感知的屬性或他們無法檢測差異的問題給出明確的答案。
4. 僅使用之後可以進行操作的問題。不要問沒有提供相應操作數據的問題。假設一個問卷的問題是詢問包裝的吸引力，收到的答案是包裝有點沒吸引力，研究人員知道要怎麼因應這個結果更改包裝嗎？
5. 在問卷中始終提供空位以回答開放式的問題。例如：在詢問受試者是否對產品有偏好度或接受度的問題之後，馬上接著詢問原因。
6. 將偏愛度或接受度的所有問題放置在可引起最深思熟慮回答的位置。在許多情況下，整體接受度(overall acceptance)是最重要的，分析師理所當然地傾向放在調查表的開頭。然而，在實際消費前，消費者被問及產品的外觀和/或香味等具體問題的情況下，有必要等到這些屬性進行評估和評分後，再詢問全盤接受或偏好的問題。附錄 11.3 提供接受度問卷的兩個例子。

11.6.2.3 協議(草案)設計(Protocol Design)

在實驗室環境中（見第3章，第2節），感官測試非常難以控制。在實驗室之外，在集中位置或家庭使用環境，控制測試設計、產品處理，消費者選擇的需要就更難了。開發和設計外部喜好性測試建議遵循以下原則：

測試設施：在集中定點測試，設施以及測試管理員必須遵守關於尺寸、靈活性、地點及環境控制等嚴格的協議。測試位置應該設置在具有多數目標客群的位置，且受試者應該能輕鬆地到達。

基於該研究的設計，應該考慮各個設施能提供足夠空間的能力，照顧每個消費者/受試者的隱私，適當的環境控制（照明、噪音控制、氣味控制等），產品處理和準備的空間，以及足夠數量的管理人員和調查員。

測試管理員：管理員應該是通過感官分析師設計特定類型的測試培訓經歷的合適者。除了熟悉測試設計，必須提供測試管理員一套關於具體研究的詳細說明：怎麼處理問卷、受試者和樣品。

受試對象：每個測試點都必須依照當地人口統計的條件仔細選擇符合標準的受試者（見11.2）。一旦選定受試者，必須告知測試的位置、測試類型、測試產品數量、持續時間和報酬的類型等。假如測試內容不是完全符合其期待，消費者並不會有妥當的回應。

篩選樣品：任何喜好性測試之前，提供的樣品都必須經過篩選以確定：

- 真實的待測樣品來源（批次、試產廠、生產代碼）
- 樣品在運輸前的儲存條件
- 用於儲存和運輸的包裝要求
- 運輸方式（空運、貨車、冷藏等）
- 使用描述性分析的產品感官屬性於問卷設計中和在研究的最後解讀

樣品處理：正如測試協議對於測試現場的嚴謹，儲存條件、拿取、準備和樣品的呈現的詳細及具體的說明是進行適當的測試所必要的，附錄11.4提供喜好性測試正在發展中的協議，以及已完成協議的例子。

11.6.3 量性喜好性測試的類型

喜好性測試可依照測試的主要任務大致分成兩種類別：

任務測試及類型問題

任務	測試類型	問題
選擇	偏好測試(preference test)	你偏好哪個樣品？ 你比較喜歡哪個樣品？
評級	接受度測試(acceptance test)	你多喜歡這個樣品？ 這樣品的接受程度如何？

除了這些問題，還可以使用各種形式（見如下）來問，測試設計經常詢問表示贊成或接受的原因來接續問題。

11.6.3.1 偏好性測試(Preference Tests)

關於喜好性測試該選擇偏好性測試或接受測試，應該根據專案(project)目標來決定。如果該專案是專門設計來與另一個產品競爭，如產品改進，或者是與同業競爭，則選擇偏好性測試。偏好性測試迫使選擇其中之一。但並不指明是否有任何產品都喜歡或都不喜歡。因此，研究人員必須具有當前產品或競品(競爭產品)的初階知識，以及它們被喜好的狀況。

偏好性測試可以分為如下幾類：

測試類型	樣品數	偏好
配對偏好 (Paired preference)	2	兩個樣品中選擇一個 (A-B)
排序偏好 (Rank preference)	3 或更多	按照偏好順序排序 (A-B-C-D)
多組配對偏好 (Multiple)	3 或更多	一系列兩兩配對的樣品

paired preference) —

全部配對(all pairs)

多組配對偏好—選擇配 3 或更多

對(selected pairs)

(A-B, A-C, A-D, B-C,

B-D, C-D)

將一或兩個樣品與其他

的兩兩配對 (A-C, A-D,

A-E, B-C, B-D, B-E)

見第 7 章，多組配對的原則、程序及分析的討論。

1. 配對偏好：又稱為配對比較，此法的目的是用來找尋受測者對 2 項樣品之間整體或某項感官特性的偏好傾向，其實施方式為給予受測者 2 個樣品，請受測者選擇較偏好者。此方法適用於產品匹配(matching test)與消費者測試(consumer test)。與其他偏好性測試方法比較，配對比較法的優點有(1)執行程序簡易，設計不易混淆(2)易與其他市場調查設計搭配(3)可增加無偏好的選擇；而缺點則包括(1)可提供的樣品間資訊較少(2)若樣品數多時，執行效率低(3)若針對特定感官特性時，較易受其他因素的交互作用影響。其統計方法與前述差異性測試相同，主要為查閱配對試驗檢定表。關於可有無偏好的選擇，與差異性配對比較法不同，因為“都喜歡”與“都不喜歡”在此是很重要的訊息，但是須注意的是若允許“平手”的選擇時，方法的辨別力會因而降低，通常受測者在 50 人以下時不建議有此選擇。

芝麻醬	
說明	
1. 首先品嚐左側的產品，接著再品嚐第二個產品	
2. 現在，你已經嚐過這兩種產品，你更喜歡哪一個？ 請選擇一個：	
☐	☐

463	189
2. 請寫下你喜歡的原因 _____	
名稱_____日期_____	

圖 11.1 範例 11.1 配對偏好性測試：芝麻醬改善的評分表。

市場希望確認該原型品確實是比起現有產品更得到消費者的偏好。

測試目標：要確定是否原型品優於目前的產品。

測試設計：這是一個單向測試，新開發的原型品是為了因應消費者要求而增加更濃的芝麻味。預先篩選芝麻醬的使用者，選擇 100 位受試者，並邀請到集中定點測試，他們同時收到兩個樣品，一半呈現的順序為 A-B，另一半則為 B-A。所有樣品以三位數隨機編碼。鼓勵受試者做出選擇（見被迫選擇的討論，第 7 章）。記錄表如圖 11.1。虛無假設 H_0 是：對於較濃芝麻味的原型品偏好比率 $\leq 50\%$ 。替代假設是 H_a ：對於原型品的偏好比率 $> 50\%$ 。

篩選樣品：使用的樣品是那些已通過前面描述的屬性差異檢驗，其中的芝麻風味較高是得到證實的。

測試執行：使用在第 7 章第 2.4 節中描述的方法，62 位受試者選擇原型品。由表 17.8 顯示該原型品比起現有產品存在顯著性的偏好。

結果解讀—該新產品可以取代現有產品，並標示「具有更多的芝麻味」在市場上銷售。

2. 排序偏好：又稱為順位偏好，此法則是用來找尋受測者對 2 項以上樣品間的整體或某項感官特性的偏好傾向，其實施方式為給予受測者多個樣品，請受測者按偏好程度將樣品順序排列，如偏好程度最高者排最右邊或最左邊，依次排列。

範例如下：

排序偏好法測試

說明：以下有 3 個樣品，請由左到右品嚐(每一樣品間隔請漱口)，並按您對樣品(芝麻醬)的喜好程度由左至右排列之，愈喜好者排愈左。

芝麻醬 _____

此方法適用於研發、產品匹配、定位(mapping)、找尋最適化產品或條件(optimization)與消費者測試等。與其他偏好性測試方法比較，排序偏好法的優點有(1)執行程序較簡易(2)可使用參考樣品比較；而缺點則包括(1)可提供的樣品間資訊僅有順序，而無程度的量(2)不同次的結果，不能互相比較(3)必須一定要排序，不能有相同位置者。

3. 喜好評分法(Hedonic test)：此方法則是用於找尋受測試對 1 或多像樣品間整體或某項感官特性的偏好傾向及程度，其實施方式為給予受測者 1 或多個樣品，請受測者按偏好程度依 9 分制標示法標示，如 1 分代表最不喜歡，9 分代表最喜歡，或反之亦可，品評單範例如下：

喜好性評分測試

說明：以下有 4 個芝麻醬樣品，請依序品嚐樣品後，依下列數字代表寫下您對該樣品的喜好程度，在品嚐各樣品間隔中，請漱口以維持準確度。

9-極喜歡 8-很喜歡 7-喜歡 6-略喜歡
5-無意見 4-略不喜歡 3-不喜歡 2-很不喜歡 1-極不喜歡

樣品代碼 _____

喜好程度 _____

此方法適用於研發、產品匹配、定位、找尋最適化產品或條件、找尋市場潛力與消費者測試等。與其他偏好性測試方法比較，喜

好評分法的優點有(1)最通行常用的測試法，對受測者是容易理解的方法(2)可提供有用而穩定的結果(3)亦可使用 5 分制或 7 分制(4)不同次測試的結果可一起分析(條件須相同)；而缺點則為(1)可提供的樣品資訊中，類別間隔(喜好程度差距)並不一定相等(2)類別描述語的影響需考慮文化與社經等因素。

16.3.2 接受度測試(*Acceptance Tests*)

當產品研究員需要確定一個產品是如何深受消費者喜愛，也就是它的「喜好性狀態」，接受度測試是正確的選擇。接受性測試包括適合度法(just right scale)與接受度法(acceptability scale)。將產品與很受歡迎公司的產品或競爭者的產品相比，使用喜好性量度，如在圖 11.2 中所示，被用於指出可接受或不可接受，喜歡或不喜歡。較低的兩個量度，「P&K 孩童」量度和「史努比」量度，通常使用於學齡兒童。

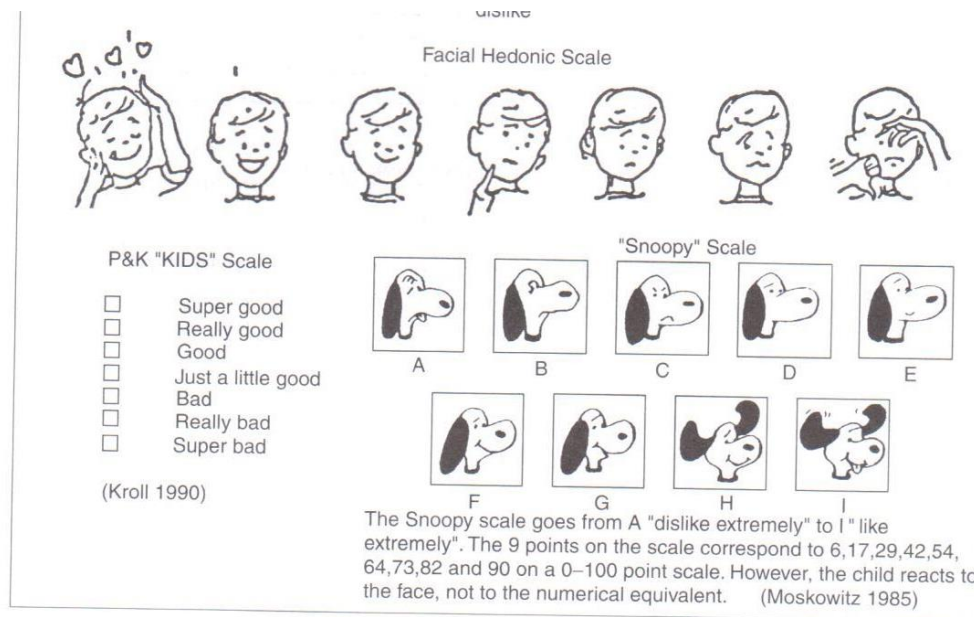


FIGURE 12.2

Scales used in acceptance tests. The last two scales are used with children.

喜好量度語言

- ☐ 極度喜歡
- ☐ 非常喜歡
- ☐ 中等喜歡
- ☐ 輕微喜歡
- ☐ 說不上喜歡不喜歡
- ☐ 輕微不喜歡
- ☐ 中等不喜歡

購買意圖量度

- ☐ 絕對會購買
- ☐ 可能會購買
- ☐ 可能會也可能不會購買
- ☐ 可能不會購買
- ☐ 絕對不會購買

☐ 非常不喜歡

☐ 極度不喜歡

喜好量度分類



極度不喜歡

普通

極度喜歡

臉部喜好量度



P&K 孩童喜好量度

史努比量度

P&K 孩童量度

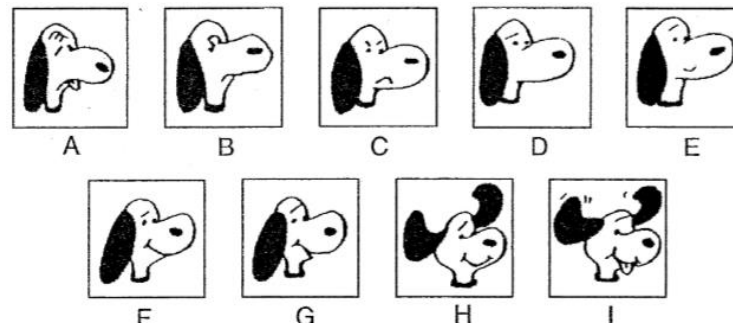
☐ 超級好

☐ 很好

☐ 不錯

☐ 還可以

- ☐ 不好
 - ☐ 很不好
 - ☐ 超級不好
- (Kroll 1990)



史努比量度是從 A「非常不喜歡」到 I「非常喜歡」。9 個量度分別對應到 0-100 量度的 6, 17, 29, 42, 54, 64, 73, 82, 90，然而孩童多使用臉部表情，而非數字。
(Moskowitz, 1985)

圖 11.2 在接受度測試中使用的量度。最後兩個量度使用於兒童。

從相對的接受度得分，可以推斷偏好度；具有較高分數的樣品有較高的偏好度。欲獲得最好（最具差別性、最可行的）的結果必須使用平衡的量度，即正負類別的數量相等並具有相等大小的步驟。圖 11.3 顯示的量度並不被廣泛使用，因為它們是不平衡的、不均勻的間隔。例如在圖 11.3 的六點優良量度(six-point excellent scale)，正面類別（「良好」到「優秀」）明顯過多，且從「糟」到「尚可」的距離顯然大於「極度優良」及「優秀」。接受度測試其實非常相似於屬性差異試驗（見第 7 章），除了這裡的屬性是指接受或喜歡。有許多不同類型的量度，如類別量度（如在圖 11.2 和圖 11.3 所示）、線性量度，或幅度預測量度可用於測量對於產品的喜好程度。

八點喜好量度
覺得它很棒
很喜歡它

九點軍需官量度
(Nine-point
quartmaster) (unbal<

六點喜好量度（不平衡）
我覺得它很棒
我很喜歡它

有點喜歡它	不平衡>.)	不知為何有點喜歡它
一點點喜歡它	極度喜歡	普通
一點點不喜歡它	強烈喜歡	我並不特別喜歡它
有點不喜歡它	非常喜歡	我完全不喜歡它
非常不喜歡它	頗喜歡	
覺得它很糟糕	中等喜歡	
	輕微喜歡	
	輕微不喜歡	
	中等不喜歡	
	強烈不喜歡	
七點優秀量度	五點（不平衡）	六點優良量度
優良	優良	優良
非常好	好	極度優良
好	尚可	非常好
尚可	差	好
差	糟糕	尚可
非常差		糟
糟糕透了		

圖 11.3 喜好量度不平衡或間距不明確的例子。

範例 11.4：競爭的高纖早餐麥片—兩個原型品的接受度測試

問題/情況—一家主要穀物製造商已決定進入高纖穀物市場，並且已經準備了兩個原型品。另一家主要的穀物生產已經在高纖早餐麥片市場擁有領導品牌，且其市占率持續增長中。研究人員需要獲得他的兩個原型品與競爭品牌相比的接受度評級。

專案目標—要確定兩個原型品之一，與市面上銷售的領導品牌相比，具有足夠的接受度。

測試目標—測量兩個原型品和市場領導品牌在高纖穀物使用者中的接受度。

篩選樣品—在審查樣品時，研究人員、品牌行銷人員、感官分析師分別品嚐將提交給家庭使用測試的樣品，以及競爭的產品。

測試設計—每個原型品與競爭者產品做一系列的單獨配對，且每個產品進行一周測試。原型品和競品各自在 150 家受試家庭中的一半先評估，並要求受試者使用 11.2 圖中所示的九點語言式喜好量度 (nine-point verbally anchored hedonic scale) 評估產品。

測試進行—產品（原型品或競品）各自在各個通過預篩選受試者的家被放置一個星期。在問卷完成後，取走第一個產品，並提供第二個產品給受試者，在第二周使用。在第二周的末尾取走第二張問卷，以及第二個樣品。

結果分析—每個原型品與競爭對手進行獨立成對樣品 t 檢定測試（見第 13 章）。樣品的平均接受度得分如下：

	原型品	競品	差異
原型品 1	6.6	7.0	-0.4
原型品 2	7.0	6.9	+0.1

例 1 屬性診斷：屬性偏好問題(Attribute-by-Preference Questions)的例子

1. 整體上你偏好哪個樣品？ 467-- 813--
2. 你喜歡哪一個的顏色？ 467-- 813--
3. 你喜歡哪個可樂的刺激？ 467 - 813--
4. 你喜歡哪一個的柑橘味道？ 467-- 813--
5. 你喜歡哪一個的辛辣味道？ 467-- 813--
6. 你喜歡哪一個的甜味？ 467-- 813--

例 2 屬性診斷問卷—使用喜好性量度評測單個樣品的每個屬性

- ☐ 極度喜歡
- ☐ 非常喜歡

- ☐ 中等喜歡
- ☐ 稍微喜歡
- ☐ 既不喜歡也沒有不喜歡
- ☐ 稍微不喜歡
- ☐ 中等不喜歡
- ☐ 非常不喜歡
- ☐ 極度不喜歡

使用上面的量度評測下列的問題：(量度可以在每個問題後重複)

你覺得整體上這種飲料如何？-----

你覺得它的顏色如何？-----

你覺得可樂影響如何？-----

你覺得柑橘味道如何？-----

你覺得香料味道如何？-----

你覺得甜味味道如何？-----

你覺得稠度如何？-----

例3 使用「適合度法」的屬性評測（燉物）

請說明您對以下特點意見：

肉汁顏色	☐	☐	☐	☐	☐
	太輕		恰到好處		太暗
蔬菜量	☐	☐	☐	☐	☐
	太少		恰到好處		太多

牛肉香精數量	☐	☐	☐	☐	☐
	過低		恰到好處		過高
鹽度	☐	☐	☐	☐	☐
	過低		恰到好處		過高
肉汁的厚度	☐	☐	☐	☐	☐
	過薄		恰到好處		過濃

例 4 屬性診斷：含蓄的「適合度法」

1. 顏色	☐	☐	☐	☐	☐	☐	☐
	太亮						太暗
2. 可樂味	☐	☐	☐	☐	☐	☐	☐
	太薄						太濃
3. 柑橘味	☐	☐	☐	☐	☐	☐	☐
	太薄						太濃
4. 甜度	☐	☐	☐	☐	☐	☐	☐
	完全不甜						太甜
5. 濃度	☐	☐	☐	☐	☐	☐	☐
	太薄						太濃
6. 二氧化碳量	☐	☐	☐	☐	☐	☐	☐
	不足						過多

圖 11.4

在屬性診斷測試中使用的量度例子。

例5 屬性診斷：簡單的強度量度

請評估麵食樣品的下列各屬性的強度：

外表

- [illegible]

風味

4. 熟麵味	☐	☐	☐	☐	☐	☐	☐	☐	☐
	無								濃
5. 鹹味	☐	☐	☐	☐	☐	☐	☐	☐	☐
	無								重
6. 蛋香味	☐	☐	☐	☐	☐	☐	☐	☐	☐
	無								濃
7. 新鮮風味	☐	☐	☐	☐	☐	☐	☐	☐	☐
	無								濃

質地

8. 初始黏度 ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐
- 不粘 很粘
9. 硬度 ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐
- 很軟 很硬

10. 彈性 □ □ □ □ □ □ □ □
很軟 很有彈性

11. 澱粉性 □ □ □ □ □ □ □ □
沒有 很潮

原型品 1 和競品之間的平均差異顯著不同於零，即原型品 1 比起競爭對手的平均接受度有顯著性地低。原型品 2 和競品之間沒有顯著差異。

結果解析—專案經理的結論是，原型品 2 的表現和競品一樣好，建議該集團將它作為公司進入高纖穀物領域的產品。

11.6.4 評估單項屬性 (屬性診斷<Attribute Diagnostics>)

作為消費者測試的一部分，研究人員經常努力透過詢問感官屬性（外觀、香氣/香味、聲音、香味、質地和手感）等問題，以確定任何偏好或拒絕的理由。這些問題可分為以下幾組：

1. 對屬性的情感反應：

偏好：你喜歡哪一個樣品的香味？

喜好：你喜歡這款產品的質地？

[極度不喜歡·····極度喜歡]

2. 對屬性的強度：

強度：這個餅乾是如何脆？[不脆.....非常脆]

3. 強度の適度：

適合度：評價這個穀物的甜度：[不夠甜.....太甜]

圖 11.4 顯示了屬性問題的例子；其他例子在 11.6.1. 討論。在第一個例子中，問卷調查的是兩個樣品對每個屬性的偏好，他們更喜歡哪個樣品。在第二個例子中是單一產品的「屬性診斷」調查問卷，要求受訪者使用「極端喜歡」到「極端不喜歡」的量度評測每個屬性。這樣的問卷被認為在確定每個屬性的重要性上並不那麼有效，因為受試者常常在各個屬性給予相似的評價，其結果是一系列具有一般反應「光環」的屬性。此外，如果一個屬性得到負面的評級，研究者沒有辦法確定產品被討厭的部分。如果產品質地不被喜歡，是「太硬」或「太軟」？「太粗」或「太薄」？在第三個和第四個例子中的「適合度("just right")」量度中，允許研究者在相對屬性的強度上評估受試者的心理標準。「適合度」無法透過計算平均反應進行分析，因為每個消費者心中對每個屬性的相對強度以及相對適當性可能是不平衡或不均勻的間隔量度。下面的步驟建議：

1. 計算消費者回應每個類別屬性的百分比

適合度量度的屬性結果例子

回應百分 比	5	15	40	25	15
	太少	有點少	恰到好處	有點多	太多

2. 使用 X² 檢驗（見第 13 章），與成功品牌獲得的回應分佈比較。

另一種類似的方法是在每個屬性上使用強度量度（沒有中間點）（第五個例子）。為了評估這些屬性的適當性，強度值必須與整體接受度或該屬性接受度相關。通用食品將消費者個人理想的強度等級連結到消費者質地剖面法進行研究，表明了接受度和各種產品接近消費者理想的程度之間的高度相關性。

11.6.5 其他資訊

人們對於一個品牌或產品類別的態度和印象可能會隨時間而改變。市場研究人員展開頻繁且定期的監測消費者認知和行為的跟踪

研究，被稱為態度及習慣（Attitude and usage, A&U）、態度意識及習慣（Attitude Awareness Usage, AAU），以及習慣和態度（Usage and Attitude, U&A）。A&U 研究週期性地重複，提供了研究人員捕捉行銷活動正在如何影響消費者對於一個品牌或產品的態度。

一項 A&U 研究經週期性地重複，提供了研究人員獲得行銷活動正在如何影響消費者對於一個品牌或產品的察覺(awareness)。相對於競爭者，察覺(awareness)被追蹤：品牌形象如何變化？目標市場的使用習慣模式如何變化？以及哪些變數(包括人口統計數據)定義目標市場？這些研究的目的是為了獲得特定的使用資訊，比如去速食店及點套餐的頻率，或者如購買巧克力曲奇餅等特定購買行為的頻率。收集的資訊可以作為行銷策劃，新產品開發和競爭情報的基礎。雖然測量態度及習慣的研究設置非常嚴格，仍可以併入這些問題至產品研究中，以確保該樣品被典型的目標客群以相同的方式追蹤。

11.7 網際網路研究(Internet Research)

11.7.1 介紹/定義/用途

網際網路是連結目標客群的另一種方式。網際網路研究採用分佈式調查（surveys distributed）或透過在全球網路上發佈訊息來收集消費者對某一特定產品或產品概念的資料調查。使用網際網路進行研究的方式在過去十年已普及。儘管一些研究人員對於收集到的數據的有效性感到懷疑，利用網際網路進行的網絡調查仍持續增長。據 MacEvoy（2000 年），在網際網路普及率超過 20% 人口的國家，線上調查的結果往往與傳統的紙筆方式的結果非常相似。此外，美國已報導傳統方式的低回應率，以及其數據分析幾近零成長率，而使用網際網路的調查則持續增加中。

在進行消費者調查時，網際網路是一個有價值的工具。使用網際網路進行消費者研究有一些潛在的好處。關於利用網際網路調查如何超越傳統的調查方法主要有四種假設：（1）耗時更少；（2）同樣有效，甚至可能比傳統的調查方法更有效；（3）更具成本效益來進行；（4）更容易執行。這些假設可能或可能不是真實的，並在很大的程度上取決於調查的具體情況。

使用網絡調查收集數據並自動輸入至數據庫的好處是多方面的。不僅是數據自動化輸入大幅降低了手工輸入錯誤的數據機率，數據的周轉率也比傳統的方法快得多。

網際網路研究的另一個關鍵優勢是能夠與更廣的群眾連結，以及更多樣化的人口。這些人可能是難以用傳統方法連結的。由於沒

有地理障礙，也不需要消費者浪費交通時間，更多的人都能夠參與線上調查，離測試設施再遠的住宅都能參加。另一個好處是能夠針對成千上萬的消費者取樣，這對於傳統的方法是完全無法想像的能力。研究人員必須牢記，若目標消費者的整體人口特徵是不上網或使用電子郵件時，將無法參加調查。某些特定種族和社會經濟群體是沒有接觸網路的，因此比起白種人，中產階級的消費者來說更不適用於網際網路研究。

利用網際網路進行研究並不是沒有缺點。相較於郵件調查，網際網路調查有較低的回應率。不請自來的商業電子郵件在當今的網際網路淹沒個人電子郵件帳戶的情形不斷增加（也稱為垃圾郵件），消費者更可能直接刪除類似於垃圾郵件和所有看不懂的網路調查邀請的電子郵件。此外，由於防火牆保護的進步，要發送給潛在參加者的電子郵件變得很困難。有許多方法來提高回應率，包括簡單的電子郵件和調查格式，個性化的電子郵件求職信，跟進提醒和意圖發送調查問卷的預先通知等。

11.7.2 應用

網際網路研究在正確使用的前提下，是一個強大的工具，它可以用來進行早期概念的研究，以及決定早期產品的發展方向、聯合分析、量性和質性的研究。

早期的概念研究測試可能涉及一個想法或概念，可能包括圖像、音頻或視頻的描述。消費者被問及有關他們的思想、感情和概念等印象的量性及開放的問題。

聯合分析識別那些對消費者影響最大的產品元素及概念。這是藉由呈現給受訪者許多關於產品元素的組合資訊而完成的。

傳統的調查和問卷調查可以在網際網路研究（如產品特性強度評級或喜好度評級）使用，可以產生量性的數據。質性數據可以利用網際網路成立的焦點小組或面試進行。消費者被邀請參加「聊天室」，討論某個特定的產品或產品概念。主持人必須具備更多及高度熟練的對話技巧，如透過打字和控制談話，以調節網路上的焦點小組的討論方向、促進討論及打斷壟斷式的發言。

11.7.3 網際網路研究的考慮

有許多細節是需要考慮的。例如：什麼類型的測試是感官小組有興趣進行的？是進行消費者意見調查？或者是從預先篩選的消費者人力庫中進行檢索已獲得資訊？感官團隊可以透過網際網路研究在公司內部進行，或團隊可外包此研究給外部市調公司。在沒有外部公司的幫助下進行網際網路研究時，可以使用一些研究工具或軟體引導，如 Zoomerang。這些方案透過電子郵件或網站指導調查的

設計和配置，以及分析數據。關於網際網路研究存在許多各式各樣的公司，它們各有自己的方式來進行招募、傳輸數據和數據分析。在選擇公司時，要考慮團隊和專案的需求。

與消費測試參與者的溝通分為四個接觸點：預先通知、聯繫、回應和後續跟進。雖然並非每個測試都需要運用此四個步驟，然而使用全部的四個步驟將提供更高的回應速率和更高的使用者滿意度。如前面提到的，使用電子郵件預先通知調查的消費者，將不太可能被誤認為垃圾的調查電子郵件（聯繫），並且消費者將更有可能參與。同樣地，預先通知可以產生消費者的興趣，而再次增加參加的可能性。後續跟進可以使用各種模式進行，如手機、電子郵件、直郵信封等，並能提醒消費者調查已發送到他們的電子郵件。這些通訊模式也可用於感謝消費者完成上述調查。

透過網絡進行的研究設計有若干個步驟時需要思考。在設計和實施一個線上調查過程的基本步驟概述如下。研究的成功與否取決於所有步驟。

1. 決定測試調查目標，包括

- 切分感興趣的人群（例如：35歲以上女性或貓的主人）
- 決定收集數據的類型（量性或質性）
- 確定期望結果的精確度

2. 確定進行採樣的目標，例如，要使用的採樣方法

3. 創建測試儀器，包括

- 選擇回應模式（郵件，網路或其他）
- 起草問題
- 預先測試和修改調查工具

4. 整個調查過程聯繫參加者：

- 調查前的預先通知
- 調查後的提醒與答謝
- 針對無回應的參加者的後續跟進

5. 數據收集、數據整合和分析

11.8 使用其他感官方法來發掘見解

11.8.1 關於喜好性和描述數據

產品開發專業人員在一個產品週期中同時接觸研發和行銷角度，因而認識到消費者對整體的接受度反應和購買意向是決定產品或概念繼續或停止的底線。

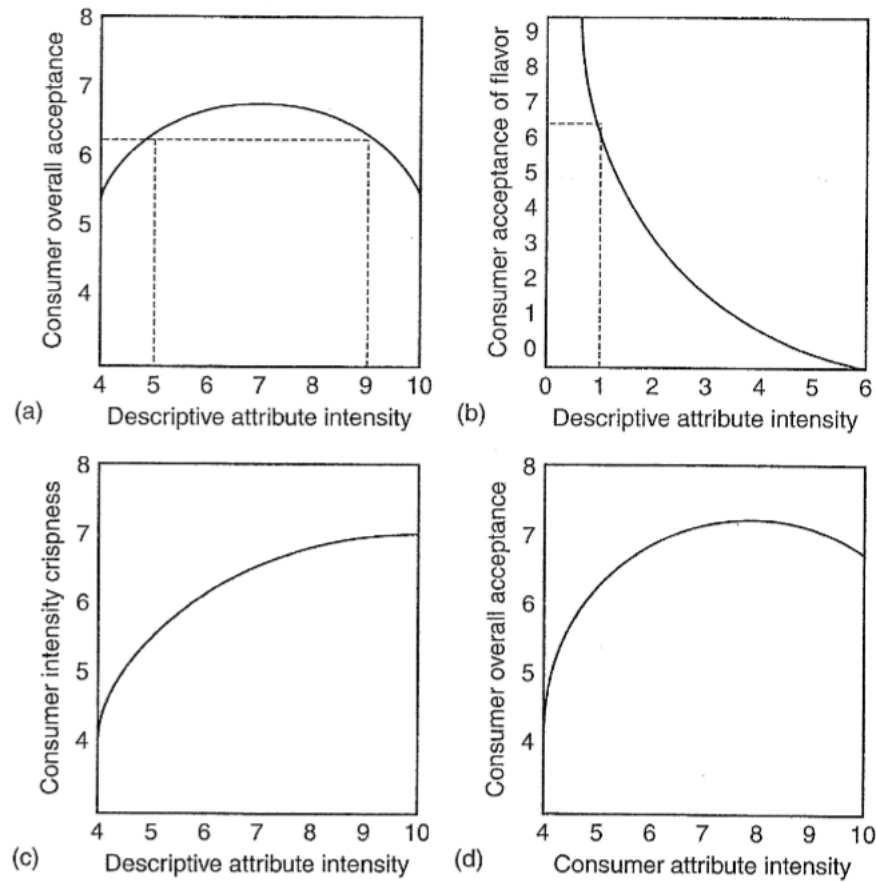
儘管承認需要喜好性數據，產品開發團隊一般不太能確定消費者回應的意思。當消費者表示產品「太乾」或「沒有足夠的巧克力味道」，他是真的因為感知到濕潤/乾燥，或感知到巧克力味而做此回應？或者說，是他在回應那些在他的腦海中與產品的優劣有關的印象呢？太多的研究人員只看到消費者的表現回應（作為研究者使用感官詞彙而言），而這些研究人員落得修改那些可能並不違背消費者意見的屬性。

解碼消費者診斷和消費者的接受程度的關鍵，是使用比較客觀的感官工具來衡量一個產品的感官特性。訓練有素的描述或專家小組提供了產品感官特性的指紋圖（Thumbprint）或頻譜圖（Spectrum）。由這些感官檔案所構成的真實屬性特性與產品之間差異的列表可用於設計有關的問卷，以及在測試完成之後來解釋所得到的使用者數據。透過對比品評小組以及消費者的測試數據，並在可能的情況下與材料、製造過程的器材或化學分析等比對，研究者可以發現產品屬性和最終的底線：消費者接受度，之間的關係。

當屬性的強度數據的範圍可跨用在數種樣品上（15-30），（見附錄 11.1 及附錄 11.2 手部和身體洗劑的例子），可以使用統計方法來研究數據的關係，如第 14 章所描述。圖 11.5 顯示了四個例子。圖 A 顯示了消費者的整體接受度如何因描述性品評屬性的強度而變化（例如：顏色強度）；這使研究者得以理解不同特性的強度的作用並確定可接受範圍的限度。在圖 B 中，橫軸顯示不期望的屬性，例如：臭味的強度；縱軸是風味的消費者可接受度；陡的斜度表示該產品的某方面對消費者喜好度有很強的影響。從曲線 C 的類型關係，研究者可以學習在比較技術性的術語上，消費者如何描述及表達。需注意，描述型品評小組在脆度的評級上與消費者的評級呈現高相關性，但後者較無急劇上升。最後，曲線 D 將兩種消費者評級做關聯，顯示在一個屬性強度的範圍內是消費者認為可以接受的。這種關係，無異於「適合度」（"just right"）的評測。

圖 11.5 的數據關係是單變數的。消費者數據通常為多個變數之間的相互作用（產品、品評員、一個或多個屬性）。這種類型的數據的需要多元的統計方法，如主成分分析（Principal component analysis, PCA）或部分最小二乘法（Partial least square, PLS）

(見第 14 章)。



A) 縱軸：消費者整體接受度 橫軸：描述性屬性的強度

B) 縱軸：消費者對於風味的接受度 橫軸：描述性屬性的強度

C) 縱軸：消費者評量脆度的強度 橫軸：描述性屬性的強度

D) 縱軸：消費者整體接受度 橫軸：消費者的屬性強度

圖 11.5

從消費者研究中提取數據關係的例子。(A)(左上)消費者整體接受度與描述性屬性強度(顏色強度);(B)(右上)消費者對於風味的接受度與描述性屬性強度(臭味);(C)(左下)消費評量脆度的強度與描述性屬性強度(脆度);(D)(右下)消費整體接受度與消費屬性強度(甜度)。

11.8.2 使用喜好性數據定義保存期限或品質限制

在第 11 章中,「改良」的或短期的描述過程的主要用途是定義品質保證/品質管理,或保存期限。在典型的情況下,第一步是寄送新鮮產品至典型的使用者處以進行接受度測試。這張最初的問卷可能包含要求消費者回答一些關於重要屬性的其他問題。

該產品也接受品評小組的接受度測試及關鍵屬性測試,而這種評測在商品保存期內定期重複,每次與控制儲存組的產品比較,也就是在可能的情況下,在抑制惡化的儲存條件下(如氮氣深凍< deep-freeze>儲存)的相同產品,或者如果這是不可能的,就與目前生產的新鮮產品比較。

當品評小組發現與對照組和/或在一些主要屬性(多個)上具有顯著的總體差異,將樣品再次發送到使用者組,以確定此統計學顯著差異是否在消費者測試上有所不同。重複此測試,因為產品將隨著儲存的時間而有所變化。品評小組差異的大小可以與降低消費者的接受或偏好度建立關聯之後,可以使用內部品評小組以監測在保存期限中正常生產的產品,並保證結果能夠準確預測消費者的反應。

範例 11.5: 芝麻餅乾的保質期

問題/情形—一個芝麻餅乾新產品將在包裝上印上生產當天日期,該生產公司希望定義其保存期限

計畫目標—確定該產品的貨架儲存期間限制,此時間過後消費者感到產品是「過期」、「不新鮮的」、「放過久的」。

測試目標:(一)使用經過專業訓練的品評小組以鑑別不同貨架儲存期的產品差異(二)將產品呈現給消費者進行接受度測試,在(1)起始時間;(2)品評小組確立差異之後(3)重複一定的間隔時間直到消費者確立差異。

測試設計：同一批次的芝麻餅乾樣品分別放置 2、4、6、8、12 週，並且儲存在以下四組不同的條件：1「對照組」：在接近冰點的密閉容器中；2「正常條件」= 70°華氏溫度和 50%相對溼度的條件；3「濕潤條件」= 85°華氏溫度和 70%相對濕度；以及 4「乾熱條件」= 100°華氏溫度和 30%相對溼度。

測試對象：從 R&D 實驗室中依照辨別過期芝麻餅乾的能力挑選二十五位成員。

，即能夠辨識餅乾表面氣味與氧化的芝麻油混合的味道。挑選兩百五十位零食餅乾的消費者，並依照人口統計特徵選擇各階層目標客群代表。

感官方法—該研究小組使用圖 11.6 的調查問卷，並且經過訓練使用七點線性尺度(seven line scales)評定有關芝麻餅乾貨架壽命的樣品外觀、關鍵風味屬性和質地。研究小組成員並收到標記為「對照組」的樣品，並指導其使用該表格的最後一行記錄進行的參考比較法（見第 6 章第 8 節）。告知品評小組成員這些樣品是保存期限研究的一部分，測試樣品偶爾會包括比對照組還新鮮的生產樣品（這些信息減少了品評小組成員在保存期限研究中預測產品劣化的傾向）。

每次消費者收到兩個連續的編碼樣品（採隨機順序呈現的樣品和對照組），使用圖 11.7 的評分表，完成後立即交回給主審。

分析結果：最初的接受度測試獲得了 250 位消費者針對 2 個新鮮樣品平均 7.2 的評分，並附帶說明屬性的評級，餅乾被認為是新鮮且酥脆的。

芝麻餅乾的評價

1. 說明：評價芝麻餅乾的外表、風味及質地，在下列各列尺度標示出您的評分

外表

表面顏色 淡 |-----| 深

風味

烘焙麥味 無 |-----| 強

芝麻 無 |-----| 強

紙板感 無 |-----| 強

色素 無 |-----| 強

質地

硬度 軟 |-----| 硬

脆度 濕軟 |-----| 脆

2. 與對照組樣品做比較，並在下列尺度標示差異程度

無差異 |-----| 差異很大

備註：_____

名字：_____ 日期：_____

圖 11.6 為例 11.3：芝麻餅乾的保存期限，研究小組使用評價屬性等級和差異的評分表。

與對照組量度的差異(difference-from-control scale)，相同的兩個樣品在研究小組中得到評分為 3.2 (滿分 15)。2 週和 4 週的樣品顯示沒有顯著差異。在 6 週的時間點，「濕潤條件」的樣品獲得與控制樣品 5.9 分的差異，這與 3.2 分有著顯著不同。此外，「濕潤條件」樣品紙板味上得到 4.2 分 (對比於新鮮組的 0 分) 和鬆脆性 5.1 分 (對比於新鮮組的 8.3 分)，兩者透過 ANOVA 分析皆有顯著差異。

接著將 6 週的「濕潤條件」樣進行消費者測試，得到 6.7 的接受度，對比對照組 7.1 ($P < 0.05$)。在「新鮮烘烤」的風味屬性也表現出顯著下跌。

該產品研究者決定其他兩個測試樣品 (「正常環境」和「乾熱條件」) 在品評小組的差異測試超過 5.0 時，盡快進行消費者測試。隨後的測試表明，消費者僅對由研究小組評為 5.5 或以上版本的差異敏感。所有關於芝麻保存期限研究，使用 5.5 做為參考比較法中的關鍵點，超過此標準的樣品不僅在統計上有顯著的差異，也對消費者來說有潛在的差異意義。

芝麻餅乾

1. 說明：整體評估後，在您覺得最能描述產品的下列方格標示

☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐ ☐

極度喜歡 非常喜歡 中等喜歡 稍微喜歡 普通 稍微不喜歡 中等不喜歡 非常不喜歡

極度不喜歡

外表

表面顏色 ☐ ☐ ☐ ☐ ☐ ☐

淡 深

風味

鹹味 ☐ ☐ ☐ ☐ ☐ ☐

無鹹味 非常鹹

芝麻味 ☐ ☐ ☐ ☐ ☐ ☐

無芝麻味 非常濃的風味

新鮮烘焙味 ☐ ☐ ☐ ☐ ☐ ☐

不新鮮 非常新鮮

質地

脆度 ☐ ☐ ☐ ☐ ☐ ☐

濕軟 脆

後味 ☐ ☐ ☐ ☐ ☐ ☐

不舒服 舒服

備註： _____

名字：_____ 日期：_____

圖 11.7 消費者評分表，例 11.3：芝麻餅乾的保存期。

11.8.3 快速原型品開發

有一個持續努力以鑑別和實踐的測試方法能夠提供快速反饋到產品開發，並允許縮短開發週期。現有的許多方法是很容易實現的；然而，我們仍建議結果需要被大規模市場研究驗證後再進行產品發佈。鑒於開發時間表的緊湊及快速，我們使用定義較為鬆散並能夠快速收集消費者資訊的快速原型開發的量性和質性的技術，以求在產品開發過程及時得到反饋和見解。有效的快速原型開發需要包括可操作的資訊需求和快速的反饋；有日程規劃，並且能夠處理多個樣品，使用低到適中的成本和更小的規模。所用的量性和質性的技術可以彼此獨立或共同執行。快速原型的三種情形和測試計劃如下：

情形一：招募受訪對象參加牙刷的焦點小組。二到四組的受訪者在固定時間點返回參加三到四次連續多輪的測試。每一輪的重點著重在牙刷的新特性，如刷毛的數目、長度和剛度，或尺寸和形狀。該小組專注於評測各種屬性的強度範圍內的無數刺激的感官特性表現。產品開發、市場行銷，和感官分析的專業人員從後面的房間裡觀看監測產品改進的感官線索。並將想法帶回實驗室以創造新的原型品。

情形二：涉及量性檢測，招募少數的消費者（N = 50-75），從商場招募或從預先篩選的人員庫選擇參加甜味和鹹味餅乾的小規模口味測試。受訪者測試一系列 5-6 種味道的餅乾，代表不同層次的甜味印象和鹹味的特性。在品評測試結束後，進行一對一的單獨面試，讓受訪者使用其語言表達對餅乾的想法。利用電子數據採集技術，將收集的評論迅速審查並轉換成資訊。產品開發因而能夠依照消費者顯示的反應修改原型品。這個過程被重複三至四次，或更多次，重複此短期的循環直到在產品中發現可測量的改善。

情形三：使用社群敘述或講故事的方式。希望創建一種新的或改進青少年飲料，能夠提供能量及失去的營養價值，並且能夠在運動和遊戲中飲用。招募有在進行體育運動的 10-15 位青少年加入為期兩個月的測試方案。每星期有一天，這些青少年需要來到工廠兩個小時。此兩個小時被分為三部分，用三分之一花在會議或聚集談論其需求和慾望，第二段時間進行體育鍛煉，最後的第三部分品嚐產品和討論。這種社群開發的方法，使得青少年建立互動的模式，並在真實環境中獲得產品反饋。

利用快速原型開發的方法意味著和消費者保持持續的溝通基礎；這使他們能夠持續提供反饋。雖然這是一個比較實務上的做法，它仍然是一個能夠提供直接的反饋並澄清反應的快速方法。

11.8.4 序列定位 (Sequence Mapping)

序列定位是透過納入整個產品的消費經驗，包括決策、購買行為、產品的使用、情感和感官特性，並建立成一個完整故事的工具。透過一系列與符合目標客群的消費者代表的質性方式的應用，如接受者、拒絕者或前端使用者，收集詳細信息和開發產品體驗的地圖。一個序列定位的開發技術通常包括日誌、購買點的訪談、觀察研究，一對一的訪談，焦點小組和社群敘述。序列定位的最終結果是思想、情感、行為的合併，以及從早期決策到產品處置整個產品生命週期所被感知的感官特性。

範例 11.6：案例討論：一個序列定位的建立

一家領先的製造商希望為年齡 24-50 歲的女性創造其生活方式產品。這個族群的人希望在消費甜點時同時能夠符合健康生活的一部分。包括在雜貨店、便利商店、藥店和大眾商品店的觀察研究、購買點的一對一的訪談，針對滿足兩個區段的婦女進行：同時想要放縱和想要健康替代方案的婦女。所得的地圖結果表明了動機、情緒和感覺是對消費者重要的屬性。

地圖 11.1 及地圖 11.2 顯示案例 11.1 的第一層次的結果輸出，也就是研究的亮點，而地圖 11.2 則顯示事件、動機、情感和消費者感官屬性等細節。

此種序列定位結果顯示，可以創造同時滿足這兩個利基的產品：放縱和健康的生活。為了取得成功，產品必須具有特定的素質，包括：

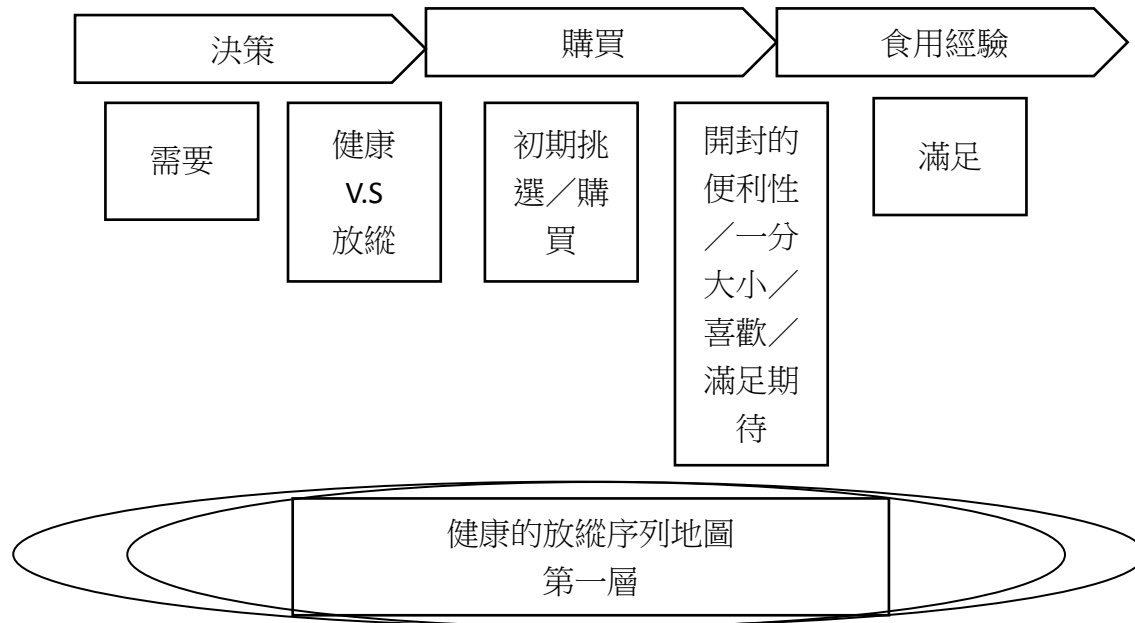
- 一種乾淨的、放縱又有足夠風味的味道
- 一種奶油和/或酥脆的放縱質地
- 一種乾淨的香氣，由濃蛋白特性或其他營養物質，如大豆、酪蛋白、維生素、礦物質或等組成
- 高品質的牛奶巧克力
- 不具有與蛋白質，大豆，或維生素相關的不良餘味
- 較小尺寸，以一片 2-至 3 盎司為最佳
- 採用新鮮和營養均衡的食材
- 增加纖維更為加分
- 巧克力之外的風味選項有：

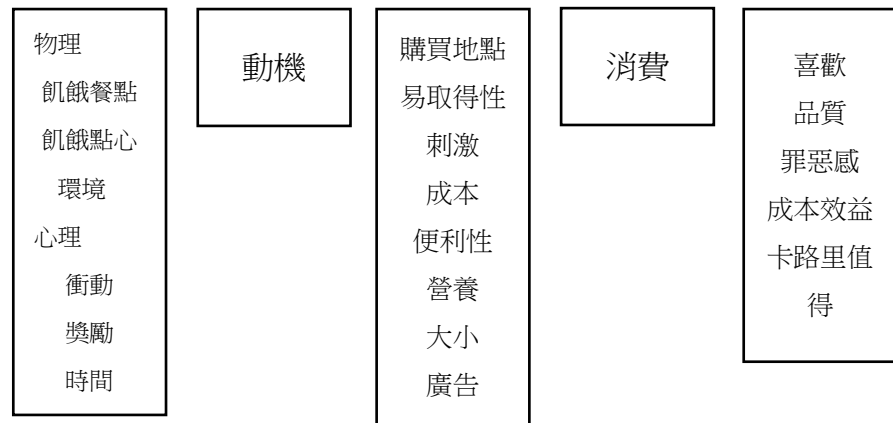
烤堅果的特色，如鹽和花生、烤黃豆或花生醬；巧克力與任何東西的組合，包括焦糖、莓果、花生醬，或優酪乳等都被認為是既放縱且健康的組合

• 其他建議形式有：

為了方便放在布丁管的優酪乳

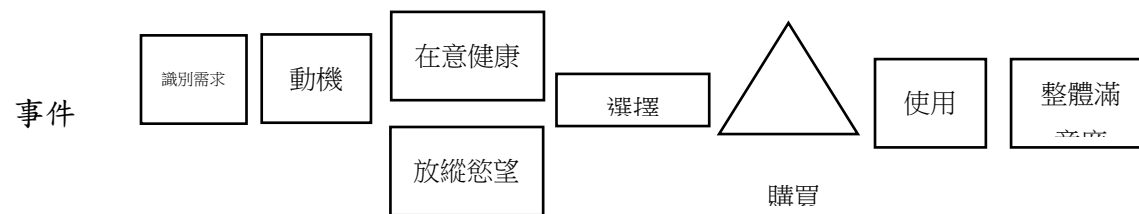
多層次的產品，如具有香脆中心的甘納許巧克力



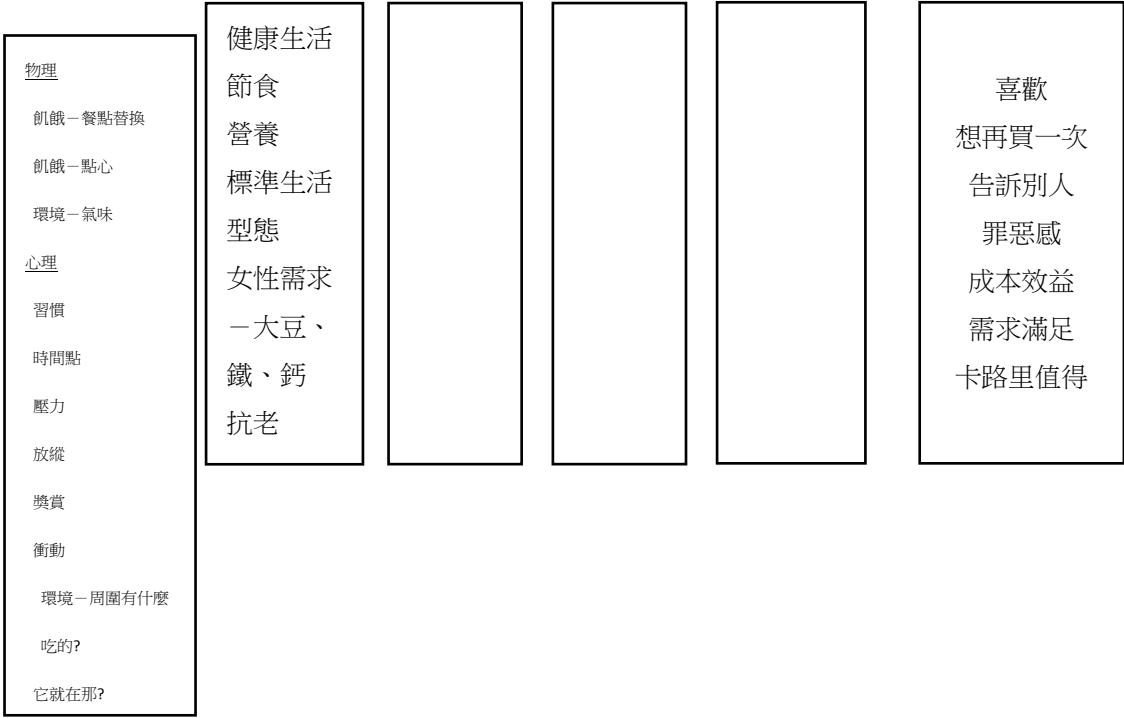


地圖 11.1 第一層序列定位

模糊前置作業與序列定位的日誌－健康的放縱



動機



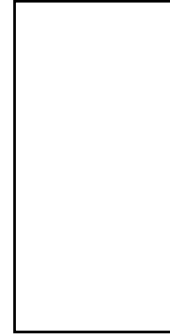
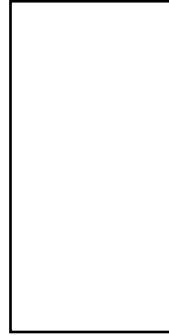
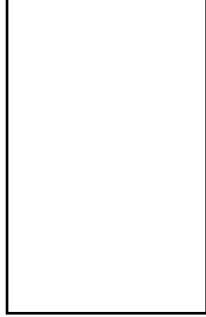
對所有事件的
情緒

壓力、飢餓、滿足、不滿足、生氣、失望、需求未滿足、空虛、不值得、容易動搖、理想的、完美的、需要的

感官反
應

包裝

產品



整體欣賞

地圖 11.2 詳細的序列定位

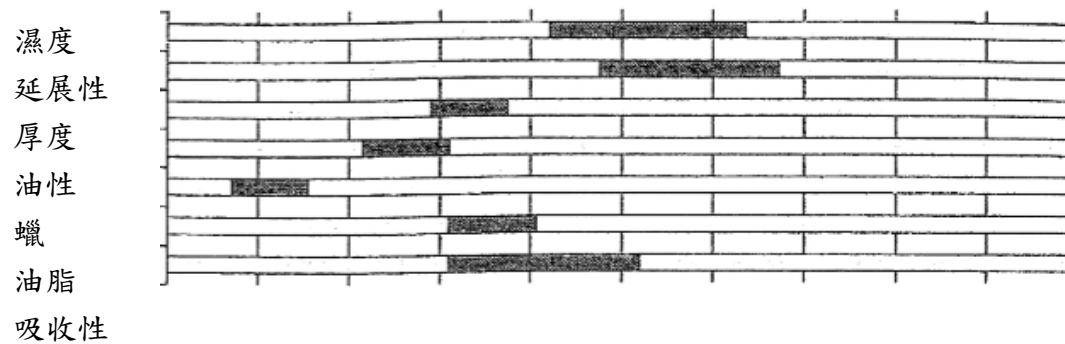


圖 11.8 描述性分析的特點數據範圍圖表。

範例 11.7：案例研究：描述性分析數據與消費者質性信息

一家防曬霜的製造商正在開發一種新的防曬乳液產品。模糊前製作業的研究成果發現，防曬乳液有一個推出新產品成功的機會，那就是提供有滋潤性又同時能夠防曬的乳液。製造商正在尋找產品開發的初步資訊和方向。

第一步是記錄的手部乳液在市面上的產品的陣列（15-30）中，從其中的多樣化子集（4-8）選擇為與消費者討論的市場。

一個訓練有素的頻譜描述性分析品評小組記錄零售市場的乳液產品陣列（16-20）中的膚感特性，因此精準地定義產品在手上及皮膚上塗抹的感覺（見第 11 章）。圖 11.8 顯示了 16 個樣品的初步產品特徵數據範圍。

使用多元統計技術分析描述品評小組結果，例如主成分分析。數據地圖允許研究人員尋找涵蓋商業潤手乳液的所有空間。從地圖上選擇五種多樣化的乳液子集合與消費者討論。

與選定的消費者進行質性訪談。消費者描述了最佳的防曬乳液應給予肌膚「柔軟、有彈性、舒緩和水分」的感覺。消費者使用這些術語來描述被描述性品評小組評定有較高的皮膚柔軟度，低的皮膚紋理能見度，並且比起油性或油膩有更多的矽脂感的產品。透過思考原來的感覺強度矩陣的已知範圍，感官科學家開發了指引發展方向的強度說明。顯示在圖 11.9。產品開發者現在能夠進一步為新創的原型品進行測試，無論是消費者接受度或偏好度，或感知的效率。

11.8.5 關鍵驅動、偏好地圖，和區隔分析 (Key Drivers, Preference Mapping, Segment Analysis)

應用於外部偏好地圖的關鍵驅動分析，是用來確定最重要的消費者感官屬性和建立目標產品的預測感官曲線。

光澤
緊繃
黏
濕潤
滑順
阻塞
柔軟
乾燥/粗糙
殘留數量
油度 %
蠟度 %
脂度 %
矽膠感 %
皮膚紋理的能見度

圖 11.9 產品使用後感覺特性的建議方向。

目標產品是在偏好地圖上被預測為最容易被任何人口區段或是研究人員感興趣的人群偏好的位置。目標產品可能在地圖上落於目前沒有實際產品存在的位置。在通過逆向工程的任何情況下，仍可以得到此虛擬目標產品的感官檔案。透過瞭解目前的產品與目標產

品是如何的不同，以及透過了解每一個感官屬性對接受度的重要性，研究人員可以優先思考其產品改進的機會並最大化收益。

識別偏好區隔是組成關鍵驅動分析的一個部分。偏好區隔是指一群喜好性模式相似的受訪者，但與其他群體有所不同。識別每個偏好區隔的目標產品使研究人員得以擁有更實際的機會滿足最大比例的消費者，以及在放棄一些產品概念上擁有更實際的目標。關鍵驅動分析以及其他用於偏好區隔的方法在第 14 章，第 4 節中結合案例分析，有更詳細的討論。

表 11.3 滿意度得分及重要性

	重要性>滿意度	目標 (%)	滿意度>重要性 (%)
給自己一個好處		110	
作為一個放縱的好處		113	
對吃不感到內疚		112	
過癮		111	
有我喜歡的味道			
是我最喜歡的食物			120
為了填補情感需求			126
有些健康的好處			132

三種牛奶巧克力和三種黑巧克力，在第一階段測試評估相同的屬性，並使用 9 點喜好性尺度。第一階段透過基本消費需求的洞察力了解驅動產品成功的關鍵因素。在第一層的比較分析，詢問消費者「什麼比起巧克力更好？」以提供理想巧克力定位的設計框架。結果表明，一個較有說服力的產品定位是巧克力比起大多數食品提供更高的滿意度，但不高於浪漫的夜晚（見表 11,2）：...

使用重要性和滿意度得分計算一個期望性差距，以確定如何透過揭示有意義的消費者動機（如表 11.3 所示），開拓市場中獨特的定位。這一分析匹配產品表現（滿意度）及預期（重要性），以幫助建立吸引消費者的產品語言。在這項研究中，只有一個屬性 — 「有我喜歡的味道」符合目標，因為這個屬性的重要性等於消費者對目前產品的滿意度。不太重要但對於目前的產品感到滿意的屬性是「最

喜愛的食物」、「為了填補情感需求」和「有些健康的好處」。

後產品分析 (Post-product analysis) 是為了達到總體喜好性得分較高的產品而發明的。牛奶巧克力的喜好性比黑巧克力高。產品前後的經驗提供了另一種滿意度及重要性的差距分析，以探討消費者預期（見表 11.4）。在前測中，「乳霜的光滑質地」被認定為最關鍵的特性。

表 11.4 實際產品的滿意度及重要性－滿意度得分的比較

屬性	前測		後測					
	重要性	滿意度	Dove 牛奶巧克力	Dove 黑巧克力	Godiva 牛奶巧克力	Godiva 黑巧克力	Valrhona 黑巧克力	Lindt 牛奶巧克力
有光滑的乳霜質地	3.85	3.98	3.85	3.60	3.34	3.08	3.06	3.32
濃的巧克力風味	3.71	3.98	3.62	3.67	3.17	3.37	3.45	2.71
在嘴中融化的方式	3.61	4.01	3.78	3.58	3.40	3.10	2.99	3.38
揮之不去的味道	3.03	3.62	3.53	3.36	3.10	3.11	3.20	2.93
精緻的	2.81	3.82	3.19	3.42	2.78	2.89	3.09	2.36

味道

黑巧克 2.39 3.70 2.47 4.02 2.98 3.84 3.97 1.79

力的味

道

在產品後測中，獲勝的牛奶巧克力具有與滿意度相當的重要性，因此符合目標。產品後測的滿意度得分低於前測成績的並沒有滿足消費者需求：「巧克力濃度」、「在口中融化的方式」和「揮之不去的味道。」沒有任何產品被認為具有複雜的味道，然而，當牛奶巧克力和黑巧克力的消費者被區隔時，各種黑巧克力的產品被認為有較為複雜的味道。

從網際網路研究得到的見解是：

1. 對巧克力放縱的想像比滿意度更大。產品的前測分析顯示巧克力類別產品中的情感和智力驅動的重要性，而後測分析則表明，巧克力的食用經驗可能不符合消費者的期望及滿足其慾望。
2. 包括融化速率的質地，在定義巧克力的感官特性上是比味道更為重要的。質地提供了產品複雜性及放縱感。
3. 品牌名稱是選擇巧克力的驅動因素；然而品牌知名度在這項研究中並不清楚。

附錄 11.1 消費者研究的篩選—焦點小組、集中地點測試（CLT）和家庭使用測試（HUT）

篩選

手和身體乳液

一般：質性（焦點小組）或量性（CLT 或 HUT）

名稱_____日期_____

電話_____時間_____
(白天) (晚上)

街道_____

城市_____國家_____郵遞區號_____

面試官_____地點_____

預約：

日期_____

時間_____

向受試者簡介：

你好，我是一間全國性調查研究公司的_____。我們正在進行一項調查；請問你是否有空回答一些問題？

如果沒有：好的，謝謝您的時間。

如果有的話：太好了。如果您在調查結束時符合資格的話，將會被要求參加一項研究。我們屆時會與您預約前來的時間。

所以，讓我們開始吧

廣泛的問題

1. 記錄受訪者的性別：

男 ☐ 基於配額終止或繼續

女 ☐ 基於配額終止或繼續

2. 在過去的 3 個月，你是否有參與調查、小組討論或消費者測試？

是 ☐ 終止

否 ☐ 繼續

3. 你，或者你的直系親屬的是否在以下任何類型的企業工作？

是 否

廣告代理或電視品牌行銷研究公司 ☐ ☐

一家公關公司 ☐ ☐

科研或相關領域 ☐ ☐

一個零售、批發或生產個人護理產品公司 ☐ ☐

一種化妝品折扣店 ☐ ☐

如果回答是、不知道或拒絕回答任何問題的話則終止測試

4. 為了分類，請告訴我下列哪個類別最符合您的年齡。(混合招募)

25 歲以下 ☐ 視配額終止或繼續

25-34 歲 ☐

35-45 歲 ☐

46-55 歲 ☐

56 歲以上 ☐

5. 下列哪個收入階層最符合您的家庭總收入是多少？（混合招募）

\$ 30,000 以下 ☐ 視配額終止或繼續

\$ 30,000-55,000 ☐

\$ 55,000-80,000 ☐

\$ 80,000-100,000 ☐

超過\$ 100,000 ☐

特定問題

6. 下列哪個項目過去 6 個月中你曾經購買過並且定期使用？（標記所有符合的）

手部及身體乳液 ☐ 若沒有選中任何產品則終止

洗衣粉 ☐

脆餅 ☐

面紙 ☐

碳酸飲料 ☐

7. 你有皮膚過敏或敏感性？

肥皂 ☐

洗衣精 ☐

手部及身體的香味乳液 ☐ 如果選中則終止

洗髮精 ☐

無香味的手部和身體乳液 ☐ 如果選中則終止

8. 你是否經常在一天中擦手部及身體乳液？

0-3 次 ☐

4-6 次 ☐ 基於配額終止或繼續

7-10 次 ☐

10 次以上 ☐

9. 哪個牌子的手部和身體乳液你最常使用？

品牌 A ☐

品牌 B ☐ 必須勾選以繼續

品牌 C ☐

10. 我將陳述一系列語句，告訴我在以下說法中你是非常同意、同意、既不同意也不反對、不同意或強烈不同意。

焦點小組	非常同意	同意	既不同意	不同意	強烈不同
			也不反對		意

a. 皮膚護

理對我是

很重要的

b. 我吃提

供皮膚維

生素和營

養素的食

物

c. 我使用
提供多種
維生素和
礦物質滋
養皮膚的
手部和身
體乳液

d. 在冬天
使用手部
及身體乳
液比夏天
更為重要

e. 手部及
身體的香
味

我想要你對一些陳述的反應 [閱讀下列清單]

對於表達我的意見和信仰我感到舒服 ☐是 ☐否

我喜歡在小組討論中，每個人表達自己的意見 ☐是 ☐否

如果被要求描述東西，我通常可以詳細地這麼做 ☐是 ☐否

[受訪者必須對以上陳述皆回答「是」才符合條件]

我的下一個問題與比到目前為止的問題還要困難回答，但請給我你最好的答案。如果你可以與任何人吃飯，那會是誰，你為什麼會選擇他們，你會講什麼？

為什麼？-----

會講什麼？-----

注意：這個問題是用來篩選受試者的表達能力。注意聽受試者如何回答這些問題，而不是他/她的回答內容。我們需要清楚地表達自己的受訪者，並且能夠輕鬆地使用詞彙表達抽象的概念及其想法。

D 邀請

焦點小組

公司邀請像你一樣的男性和女性參加市場調查。焦點小組討論將持續大約 75 分鐘，作為補償將依您的投入時間支付對應的報酬。如果你在測試開始時缺席，或者是無法前來參加整個測試的話，您將不會收到款項。你願意參加？

是 () 否 () 時間：

集中地點測試 (CLT)

我想邀請你參加一個將在我們的辦公室_____舉行的有趣研究。你是否願意來我們的辦公室 2 天嘗試幾種手部和身體乳液？每一天你將會給 4 種不同的乳液意見。每次測試將持續約 45 分鐘。作為補償將依您的投入時間支付對應的報酬。你是否願意參加？

是 () 否 () 時間：

家庭使用測試 (HUT)

我想邀請你參加一個在家測試的手部及身體乳液的市場研究。在接下來的一箇月中，你將被要求使用兩種不同的手部及身體乳液。你將被要求在家測試我們的工廠的第一款乳液產品 14 天。在此期間，你將被要求回答一些問題。我們將提供給你一本日記，以記錄你的反應和評論。在 14 天結束時，你必須將產品和日記交給我們，並試用第二個樣品 14 天。在此期間，你將再次被要求回答問題。我們將提供給你一本日記，以記錄你的反應和評論。為了補償你的時間和參與，您將收到相對應的報酬。

你是否願意參加？

是 ()

否 ()

附錄 11.2 小組或一對一面談的討論指南

(簡單) 討論指南

手部及身體的營養乳液

小組一 25-34 歲的女性；每天使用手部及身體乳液

小組二 35-44 歲的女性；每天使用手部及身體乳液

小組三 45-54 歲的女性；只有需要時使用手部及身體乳液

小組四 55-64 歲的女性；每天使用手部及身體乳液

+用途/介紹/暖身/基本規則 (15 分鐘)

+感謝大家參與；我們非常有興趣聽到每個人的發言；有沒有錯誤的答案；

大家的意見我們都有興趣

+討論規則：一次一個人；等待發言；使用視頻及磁帶記錄；禁止手機；設

施位置；討論的長度；在房間應考慮別人；保密協議

+說明小組討論的目的，以便更好地了解使用產品及說明產品預期

- +在討論前的介紹：姓名、年齡、職業、皮膚、皮膚在意的事物，以及接受什麼樣的皮膚保養，以及使用多久
- +介紹和審查概念（10 分鐘）：一種富含維生素和礦物質的手部身體乳液，以滋潤你的皮膚。
- +對概念的回應（20 分鐘）
 - +討論對概念的反應
 - + 在你面前的紙上寫下三個詞來形容基於這種理念的理想產品特點
- +探索：
 - +此概念對你具有什麼意義？
 - +預期表現
 - +「滋潤」的含義
 - +讓你感覺如何
 - +感受到的整體益處；從維生素和礦物質
 - +何時會使用此產品
 - +此產品還有什麼樣額外的益處可以提供？
- +分類和選擇產品的標準（15 分鐘）
 - +觀察桌子上的手部及身體乳液（8-10 產品）的集合
 - +你將如何分類和組合這些產品？選擇其中的領導產品。（觀察選擇、排序的過程，以及放置的決定等）
- +假如有任何產品優於其他產品，或是特殊且與眾不同，定義其產品特性。桌上是否有能夠最恰當地匹配概念的產品？

可能探索：

- +一般習慣
 - +通常購買的產品種類
 - +品牌
 - +品質
 - +必要性 V. S. 放縱
 - +價值
 - +價格
 - +附加的期望值
-

附錄 11.3 消費者研究的問卷

消費者研究問卷

A. 糖果棒問卷

姓名_____

產品 # _____

糖果棒

- 在開始測試之前請漱口。
- 透過觀察及品嚐，評估在你面前的產品。

- 考慮所有產品特性（外觀、風味和質地），在下列空格中勾選總體意見

□ □ □ □ □ □ □ □ □ □

非常不喜歡

不喜歡也不討厭

非常喜歡

- 註釋：請具體說明你喜歡或不喜歡這個產品。（使用單詞而不是句子。）

喜歡

不喜歡

1. 糖果條喜好性問題

請根據需要重新品嚐產品，並在下列方格勾選你喜歡或不喜歡的程度。

整體外觀

非常不喜歡

不喜歡也不討厭

非常喜歡

總體味道

A number line from 0 to 10. The number 1 is highlighted with a grey box. The numbers 2 through 10 are each inside a square box.

非常不喜歡

不喜歡也不討厭

非常喜歡

整體質感

☐☐☒☐☐☐☐☐☐☐☐☐

非常不喜歡

不喜歡也不討厭

非常喜歡

2. 糖果條特定評測

請根據需要重新品嚐產品，並在下列方格勾選以表示在各個產品特性上的強度及喜好度及喜好性問題。

外表

喜好性

強度／水準

顏色

☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐

非常不喜歡

不喜歡也不討厭

非常喜歡

淡

深

顏色一致性

☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐

非常不喜歡

不喜歡也不討厭

非常喜歡

不一致

一致

破掉的氣泡數量

☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐

非常不喜歡

不喜歡也不討厭

非常喜歡

無

有

風味

巧克力

☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐

非常不喜歡

不喜歡也不討厭

非常喜歡

無

濃

花生

☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐

非常不喜歡

不喜歡也不討厭

非常喜歡

無

濃

烘烤味

非常不喜歡

不喜歡也不討厭

非常喜歡

無

濃

甜度

非常不喜歡

不喜歡也不討厭

非常喜歡

無

濃

糖果條特定評測

質地

整條很紮實

非常不喜歡

不喜歡也不討厭

非常喜歡

軟

硬

堅果很脆

非常不喜歡

不喜歡也不討厭

非常喜歡

不脆

脆

融化率

非常不喜歡

不喜歡也不討厭

非常喜歡

慢慢融化

快速融化

油脂包覆口腔感

非常不喜歡

不喜歡也不討厭

非常喜歡

無

濃

B. 餐巾紙問卷

姓名_____

產品#_____

紙餐巾

- 請確保在開始前你的手是乾淨的。
- 評估在你面前的產品。
- 觀察這餐巾紙，打開並感覺它，並回答下列問題。

總體意見

請說明對此產品的整體來說，你多麼喜歡或多麼不喜歡（考慮到所有的外觀、質地、觸感等特性）圈選以下數字來表達你對此產品的整體意見。

0 1 2 3 4 5 6 7 8 9 10

極度不喜歡 既不喜歡也不討厭 極度喜歡

- 註：請說明您特別喜歡或不喜歡這個產品的什麼。（使用單字而不是句子，並盡可能具體。）

喜歡的

不喜歡的

1. 紙餐巾喜好性問題

請根據需要重新測試產品，並註明你喜歡或不喜歡的程度，圈選以下的數字以表示你的回應。

整體外觀

	0	1	2	3	4	5	6	7	8	9	10
極度不喜歡	不喜歡也不討厭			極度喜歡							

整體質感

	0	1	2	3	4	5	6	7	8	9	10
極度不喜歡	不喜歡也不討厭			極度喜歡							

2. 紙餐巾特定測試

請根據需要重新品嚐產品，並圈選對應數字以表示在各個產品特性上的強度及喜好度。

喜好度

強度／水準

表面光滑度	0	1	2	3	4	5	6	7	8	9	10	0	1	2	3	4	5	6	7	8	9	10
	極度不喜歡			不喜歡也不討厭							極度喜歡 無光澤											

光亮 顏色／亮度	0	1	2	3	4	5	6	7	8	9	10	0	1	2	3	4	5	6	7	8	9	10
	極度不喜歡			不喜歡也不討厭							極度喜歡 暗											

亮 表面花紋	0	1	2	3	4	5	6	7	8	9	10	0	1	2	3	4	5	6	7	8	9	10
	極度不喜歡			不喜歡也不討厭							極度喜歡 無花紋											

許多花紋

僵硬感	0	1	2	3	4	5	6	7	8	9	10	0	1	2	3	4	5	6	7	8	9	10
	極度不喜歡						不喜歡也不討厭						極度喜歡						無僵硬感			
非常僵硬																						
表面光滑度	0	1	2	3	4	5	6	7	8	9	10	0	1	2	3	4	5	6	7	8	9	10
	極度不喜歡						不喜歡也不討厭						極度喜歡						粗糙／不光滑			
光滑																						
本體	0	1	2	3	4	5	6	7	8	9	10	0	1	2	3	4	5	6	7	8	9	10
	極度不喜歡						不喜歡也不討厭						極度喜歡						輕薄的			
豐滿的																						
柔軟度	0	1	2	3	4	5	6	7	8	9	10	0	1	2	3	4	5	6	7	8	9	10
	極度不喜歡						不喜歡也不討厭						極度喜歡						不軟			
非常軟																						

告知監考人員你已經完成問卷，謝謝！

附錄 11.4 消費者研究的協議設計

A. 協議設計工作表的格式

1. 產品篩選

(1). 測試目標

(2). 樣本選擇

a. 變數

b. 產品/品牌

(3). 原因

2. 樣品資訊

樣品條件

(1). 樣品來源

年紀

地點

編碼

包裝條件

(2). 樣品控制_____

(3). 其他_____

3. 樣品準備

全部數量_____

其他材料_____

溫度（儲存或準備）_____

準備／復原時間_____

持有時間_____

包裝_____

其他_____

特別指導_____

4. 樣品呈現

數量_____

包裝／器具_____

編碼_____

準備大小_____

溫度_____

呈現方式_____

編碼_____

5. 品評人員

年紀範圍_____

性別_____

產品用法_____

產品消費頻率_____

可用性

B. 協議篩選 : 糖果棒

1. 產品篩選

(1) 測試產品

為測定糖果棒(具不同比例的巧克力和花生，以及某些烘烤差異的花生)的相對接受度和屬性診斷

(2) 樣品篩選

a. 變數 糖果棒上標準 1050 的覆蓋量、花生的重量、花生的烘烤色澤程度

b. 產品/品牌 篩選 18 至 22 原型品(實驗設計)和 2 個競品，有可提供的描述性數據以鑑別

彼此有稍微至無差異的產品，選擇 12 至 15 糖果棒待測

(3) 理由

14 個被選定樣品，宣稱在花生/巧克力平衡、烘烤風味強度和果粒鬆脆度

2. 樣品資訊

樣品條件

1. 樣品來源 試用原型樣品(3 oz)、相同時期小心儲存批次的競品

儲存期 3 個月

地點 訂製生產、競品來自中西部配銷商。

編碼 自有產品 L432-439、競品 A4192，7425S

包裝條件 自有產品 L432-439、競品 A4192，7425S 所有樣品以白箔帶

(732 設備

訂製)密封

2. 樣品維持

在測試前所有以白箔帶密封的樣品，放入 24 號箱子維持 3 週，

在運送到

測試地點前儲存在 50%RH 房間。

3. 其他

3. 樣品準備

全部數量

其他材料

溫度（儲存／準備）

準備／回復時間

持有時間

包裝

其他

特殊情況

原文書 308

4. 樣品呈現

數量

包裝／用具

編碼

提供大小

溫度

呈現方式

其他

5. 品評人員

年紀範圍

性別

產品用法

產品消費頻率

可用性

第十二章 技術選擇及報告結果指南

12 Guidelines for Choice of Technique and Reporting Results

12.1 技術選擇簡介

隨後的 5 張表格是一種資訊儲存方式。他們不是這本書中描述的各個方法研究的替代品。然而，由於這些方法已經普遍，透過實際動手的大多測試表可以用來檢查是否有可能是解決給定的問題的一個更好的方式。多數分析師優先考慮幾個其信賴且偏好的測試法，有時也可能為了使用其偏好的測試手法而曲解一些測試目的——這是個危險的習慣。

為了避免這種做法，並且另外找尋解決方法，筆者建議採取以下實際步驟。

12.1.1 定義計畫目的

閱讀第一章的文字，然後參考表 12.1 將計畫類型進行分類。回顧 13 項內容。寫下這個計畫的目的，然後查找哪張測試表能夠引用。

12.1.2 定義測試目的

四張表可用於這一目的：

- 表 12.2：差異測試—樣品之間是否感官差異存在？
- 表 12.3：屬性差異測試—樣品之間的屬性 x 有什麼不同？
- 表 12.4：喜好測試—哪個樣品是較偏好的？如何接受一個樣本 X ？
- 表 12.5：描述測試—紀錄完整的產品屬性。

寫下測試目的，並列出所需的測試。然後與項目負責人及其他參與該項目者共同討論並細化測試的設計。

12.1.3 重新擬定計畫目的和測試目的一測試修訂設計

在感官檢驗中的一個給定的問題，通常需要有相當的想法才能選擇相對應的實際測試（IFT 1981）。這是因為該問題可能需要釐清一些初始概念。在可接受的設計出現前，重新定義許多次問題和測試目的並非不尋常的。

表 12.1 感官分析會遇到的問題類型

問題類型	適用測試
1. 新產品開發:產品開發團隊需要試驗產品的感官特性,以及與市場上現有產品相比的消費者接受度等資訊,	本書的所有測試。
2. 產品匹配:此處最重視的就是證明現有的產品和發展中的產品之間不存在差異	在相似模式的差異測試，第六章。
3. 產品改進:第 1 步:定義到底是什麼感官特性需要改進;步驟 2:確定測試產品確實不同;第 3 步:確認試驗組喜好度優於對照組	所有差異測試。
4. 流程的變化:第 1 步:確認不存在差異;第 2 步:如果差異確實存在,確認消費者如何看待差異	在相似模式的差異測試，第六章;喜好性測，表 12.4；見說明。
5. 降低成本和/或新的供應源的選擇：	在相似模式的差異測試，第六章；喜好

- 第 1 步：確認不存在差異；第 2 步：如果差異確實存在，確定消費者如何看待差異
6. 品質控制：在生產、分配和銷售過程中抽樣產品進行測試，以確保它們和標準品依樣好；描述性測試（經良好訓練的品評小組），可以同時監看多個屬性
7. 儲存穩定性：在標準老化測試後，測試當前產品及試驗品；第 1 步：確定何時差異變得明顯；第 2 步：；描述性測試（經良好訓練的品評小組），可以同時監看多個屬性；步驟 3：喜好性測試可以確定存儲產品的相對接收度
8. 產品等級或評級：用於由製造者和使用者之間的協議存在分級方法時，確認是否接受。往往由政府監管
9. 揭開消費前的產品概念和產品開發的需要
10. 了解消費者的語言，產品使用和最初的產品原型品回應
11. 消費者接受度和/或觀點：由試驗室篩選後，可能希望提交產品到集中位置測試，或家庭測試，以確定消費者的反
- 性測，表 12.4；見說明。
- 差異測試，表 12.2；描述性測試，表 12.5。
- 差異測試，表 12.2；描述性測試，表 12.5；喜好性測試，表 12.4。
- 評級法，第五章。
- 模糊前置作業技巧，第十二章。
- 焦點小組，第十二章。
- 喜好性測試，表 12.4；第十二章。

應;接受度測試將指示目前的產品是否可銷售，或者需要改進

12. 消費者的偏好：全面的消費者偏好測試是測試市場前的最後一步;員工偏好研究不能代替消費者進行調查,但可以減少它們的數量和成本,只要產品的期望關鍵屬性從以前的消費者測試得知。

13. 品評員的選拔和培訓：品評小組的基本活動;可以由（1）面試；（2）敏感度測試；（3）差異測試;和（4）描述性測試

14. 感官相關性的物理及化學測試：需要相關性研究的時刻為：（1）減少品評小組的樣品負擔,藉由透過試驗分析取代部分試驗;（2）建立每個感官屬性的物理和化學背景知識。
（繼續）

喜好性測試，表 12.4；第十二章。

第九章。

描述性測試，表 12.5；屬性差異測試；表 12.3。

表 12.1 續

問題類型

適用測試

12. 增加物質的閾值：（1）在解決問題

第八章

時確認異味（S）來源；（2）制定感官屬性的化學原因和消費者偏好的背景知識。

注意：在 3-5 中，如果新產品是不同的，描述性測試（表 12.5）可能在特徵化差異上是有用的。如果差異被發現是在一個單獨的屬性中，屬性的差異測試（表 12.3）是可以在進一步的研究中使用的工具。

表 12.2 總體差異測試的應用範圍：樣本之間是否存在感官差異？

在此表中的測試是合適應用如下如：

1. 要確定任何產品的差異是否因材料、加工、包裝或貯存的變化而導致
2. 在沒有特定屬性（多個）可以被識別為已經受到影響下，要確定整體差異是否存在，
3. 為了測定兩個樣品是否足夠類似可以互換使用
4. 要選擇和培訓專家組成員並監測其測試樣本應用程序的測試區域區分能力。

測試

應用範圍

1. 三角測試

兩個樣品沒有明顯不同；最常用的差異測試的其中一種；統計學上有效，但有些受感官疲勞和記憶效應干擾；一般需要 20-40 位品評員，最少需要 5-8 位品評員；品評員需要簡短的培訓。

2. 兩三點測試

兩個樣品沒有明顯不同；測試具有低統計效率，但是比起三角測試較不受到感官疲勞影響；當產品是出名的可以作為

參考樣品而相當實用；通常需要 30 或更多位品評人員，最少可以只有 12-15 位；品評員需要簡短的培訓。

3. 五取二測試

兩個樣品無明顯可見的差異；統計學上有高效率，但強烈地被感官疲勞所影響，因此須利用受限的視覺、聽覺、和觸覺；通常需要 8-12 品評者，最少可用 5 位；需要簡短的培訓。

4. 差異性對比法（也稱作簡易異同測試）

兩個樣品沒有明顯不同；測試具有低統計效率，但適合使用在樣品具有揮之不去的後味、需要應用到皮膚上的測試，或是非常複雜的刺激，容易使品評人員混淆的樣品；通常需要 30 位或更多的品評人員，可以用至少 12-15 位；需要短暫的培訓。

5. A 非 A 測試

如同第四點，但用在其中一個樣品具有重要意義如標準品或參考產品，對受試者來說是熟悉的，或者在計畫中是必不可少的樣品，作為當前樣本對抗所有其他測試的樣品。

6. 差異程度測試（也稱作參考比較法）

兩個樣本可能有輕微視覺上的差異，例如由肉、蔬菜、沙拉和烘焙品等正常異質性所造成的產品；測試情況下的差異

大小影響到關於測試目的的決定，例如，在品質控制和儲存研究中；一般需要 30-50 種樣品對呈現；所需的培訓量適中。

(繼續)

表 12.2 (續)

測試

7. 連續性測試

應用領域

使用任何上述 1-3 測試以使用最少的測試確定在預定的顯著性水準下這兩個樣本是否是顯著的 (1) 相同或 (2) 不同。

8. 相似性測試

用於測試 1-3 或 7，當測試的目的是要證明這兩個產品之間沒有可感知的差異存在；在如下情況使用：(1) 一個取代舊的新成分，或者 (2) 加工變化，為了替換一個舊的或低效率的設備。

表 12.3 屬性差異性測試的應用領域：請問 x 屬性之間是如何不同？

測試

1. 成對比較試驗 (2-AFC 測試)

應用領域

屬性差異中最常用的測試；用於顯示兩個樣品中哪個有更多的測試中屬性(定

向差異測試)，或兩個樣本中偏好哪個（配對偏好測試）；測試存在單面或雙面的應用；通常需要 30 位或更多的品評員，可以使用至少 15 位。

2. 成對排名測試

用於排序 3-6 個樣品，根據一個屬性強度；配對排名的執行是簡單的，且統計分析並不複雜，但結果並不與評級獲得的結果來得實用；通常需要 20 位或更多的品評員，可以用至少 10 位。

3. 簡單排名測試

根據屬性強度排名 3-6 個樣品，最多不超過 8 個；排名的執行是簡單的，但結果並不如評級所獲得的結果來得實用；兩個樣品之間或大或小的差異顯示在排名上都是相同的差異（即，一個等級單位）；排名對於預先分類樣品或篩選樣品上非常有用，以供後續進行更詳細的測試；一般 16 位或以上品評員，可以只有 8 位。

4. 多樣品評級

用來測試 3-6 個樣品，最多不超過 8 個，按屬性數值強度評級樣品；要求在一個大的集合內進行比較樣品、通常 16 位或更多的品評員，可以使用至少 8 位；可用於比較幾個樣品之間的屬性描

述性分析，但請注意（第七章），將有一些光暈效應。

5. 平衡不完全區塊

如 4，但在有太多的樣品時使用（例如，7-15），且在一個集合中一起呈現。

6. 多樣品評級，平衡不完全區塊

如 5 號，但在有太多的樣品時使用（例如，7-15），且在一個集合中一起呈現。

此表中的測試被用來確定是否，或在何種程度上，兩個或更多的樣品在一個定義的屬性上不同。這可以是一個單一的屬性，諸如甜味，或多個相關的屬性，例如新鮮度，或整體評價，如偏好測試。除了偏好測試以外，品評人員必須小心地訓練以識別所選擇的屬性，並僅只有在品評小組成員皆理解和遵守這樣的指令，所得的結果才是有效的。在所選擇的屬性上缺乏差異，並不意味著沒有整體的差異存在。樣品不必是明顯相同，因為只有在所選擇的屬性進行評價。

表 12.4 在消費者研究及員工接受度測試中的喜好測試的應用範圍

測試	質性測試中的典型問題	應用範圍
A. 焦點小組/品評小組	什麼是與這些概念或產品相關的詞彙、行為或意見？	焦點小組小團體/二單元組/黑社會
1. 焦點小組/品評小組	針對這個個人產品，你有什麼樣的意見和行為？	小團體/二人/三人。
2. 小團體/二人/三人	當在商店或在家遇到此產品時，消費者如何表現，？	購買點的採訪、日誌觀測、影音、拼貼，序列地圖、日記。
B. 模糊前端研究	消費者如何更深層次的回應此產品或概念？	
1. 觀測值		
2. 人種論		

量性測試

A. 偏好測試	你喜歡哪一個樣本？哪個你更喜歡？	
1. 配對偏好	根據你的偏好排序產品，1 =最喜歡，2 =	比較 2 種產品
2. 排名偏好	次喜歡的，等等。	比較 3-6 種產品
3. 多個配對偏好	如 1	比較 3-6 種產品
4. 多個配對偏好，選擇配對	如 1	比較 5-8 種產品
B. 接受度測試		
1. 簡單的驗收測試	你接受/不接受此樣品？	在員工的接受度測試前的第一次篩選。
2. 喜好評級	第 12 章，圖 12.2 和圖 12.3	一個或多個產品，研究接受度是如何分佈在調查對象所代表的人口群體上。
C. 屬性診斷		
1. 按屬性偏好測試	你更喜歡哪種香水樣品？	比較 2-6 個產品，以確定哪些屬性「驅動」偏好。
2. 個人的屬性喜好評級	使用所提供的喜好尺度評價以下屬性	研究一個或多個產品，以確定哪些屬性，以及在什麼水準，「驅動」偏好。
3. 單個屬性的強度評級	使用所提供的強度評價以下屬性等級，與您的理想等級比較。	研究一個或多個產品，在受試者群體的偏好不同的情況。

喜好性測試可分為質性和量性的喜好性方法。質性方法包括焦點小組和一對一（深入）的訪談等透過團體和個人的採訪收集消費者資訊的技術。訓練的主持人、協調者和/或產品開發人員將消費者的詞彙、態度和行為指導解讀成影音或文件形式輸出。量性喜好測試可分為偏好測試，其中任務是按照偏好順序排序產品；接受度測試，其中任務是評定產品的可接受性；以及「屬性診斷」，其中的任務是依照影響產品的偏好度或接受度的主要屬性來排名或評分；在統計學分析上，偏好測試和接受度測試可被視為一種屬性差異測試（表 12.3）的特殊情況，其中所關注的屬性是偏好或接受度。理論上，在表 12.3 列出的所有測試也可以當作偏好測試和/或接受度測試。而實務中，在喜好性測試中的測試人員是缺乏經驗的，因此複雜的設計，如平衡不完全區集，是不可行的。此表中的測試也同樣適合於試驗室測試的呈現、員工的接受度測試、集中位置消費者測試、或家庭使用的消費者測試，除非另有說明。

表 12.5 描述測試的應用領域
測試

1. 風味檔案
(Arthur D. Little)
2. 質地檔案
(通用食品公司)
3. QDA®方法
(Tragon Corp.)

應用領域

此情況下許多不同的樣品必須由少數訓練有素的品評師測試。

此情況下許多不同的樣品必須由少數訓練有素的品評師來判斷質地。

如在大企業中的情況，其中有大量的同類產品必須由訓練有素的品評小組日復一日的判斷以保證品質一致；在產品開發階段的情況下則不需要要求再現性。

對於測試風味的感知強度在入口後會隨時間變化的樣品特別有用，例如：啤酒苦味，人造甜味劑的甜度。

在消費者測試中，當我們希望受試者不是使用共同的尺度時。

適用於大多數應用的客製化系統，包括 1-3 的測試；適合要求隨時間和地點重複性的測試。

為了監測在保質期研究中的產品的幾個關鍵屬性；檢查可能的製造缺陷和產品投訴；進行例行的質量保證。

描述性測試是非常多樣的，經常因為每個應用而修改設計或修改程序，因此很難在像這樣的表中進行分類。發明者的分類也許是最有幫助的。

感官測試是昂貴的，且它們經常提供出無法被清楚了解的結果。如果發生這種情況，該設計可能是有缺陷的。試點測試 (Pilot test) 通常是精化設計的一個好方法。比如說，沒有事先確認感知差異是否存在就進行數百名消費者偏好測試可能是沒有意義的；而感知差異測試可透過 10 或 20 位品評人員進行。舉另一個例子來說，相反的偏好可能是存在的，而導致正常偏好測試無效，解決此情況的方案可能是在試點研究中，使不同類型的客戶接受單一樣品測試。

12.2 報告結果簡介

對於感官結果的用戶來說，最重要的考慮因素是他或她可以在其中放置多少信心。兩個主要因素決定此點 (Larmond 1981)：

1. 可靠性(信度): 如果使用相同的品評人員測試, 是否能夠獲得類似的結果? 如果使用不同的品評人員測試呢?
2. 有效性(效度): 結論是如何有效的? 此測試是否測量了它原先要測量的東西?

由於從人類受試者所得的結果有許多變異性和偏差的機會, 感官測試的報告必須包含比物理或化學的測量報告更多細節。要決定納入多少資訊可能是困難的; 下面的建議主要是由 Prell (1976) 和食品技術研究所的感官評估部門 (1981) 提出。本章的後續將在實例中顯示建議的應用指導。

12.2.1 摘要

此測試提供什麼樣的資訊? 一個重要的環節是, 用戶不要強迫自己透過數頁文字陳述所有結果的本質。結論對於感官分析師來說是顯而易見的, 他或她應該在開放且短暫的總結中簡潔地陳述試驗結果。摘要不應該超過 110 字, 並應該回答下列四個什麼:

- 目的是什麼?
- 完成了什麼?
- 結果是什麼?
- 有什麼可以總結?

12.2.2 目的

在這本書中多次重申, 明確且書面記錄下的計畫目的和測試目的是任何感官試驗的成功基礎。

報告 (如果將呈現給計畫負責人) 應當說明並解釋測試目的; 如果該報告涵蓋了全部計畫, 應該說明和解釋各個計劃目的以及所包含的部分計畫, 以及每個測試的目的。在某些情況下, 例如, 如果報告將要出版, 解釋應該採取包括引用的期刊、相關前期工作以及參考等形式。之後, 應當對問題進行簡要的定義。陳述解決問題的辦法是非常重要的; 第十二章, 表 12.1, 遵循 IFT (1981), 應可協助這方面。如果研究是基於一個假設, 該假設應該在簡介中, 且對於讀者來說是明顯的, 且報告的後續章節應提供假設檢驗。

12.2.3 試驗

試驗部分中應提供足夠的細節, 以使得此部分工作得以重現。使用的方法應該引用足夠的參考文獻。有時候查閱試驗部分的副標題能夠幫助讀者找到特定資訊。此部分應說明收集感官數據的重要步驟, 通常會包括以下內容:

試驗設計。假設目的已經在前面清楚的陳述過了，現在文本應該解釋就目的而言試驗的「佈局」。如果有主要和次要目的，報告應該表明這是如何體現在設計上。如果使用的是進階的設計（完全隨機區集，平衡不完全區集，拉丁方塊等），可以藉由引用 Cochran 和 Cox（1957：469）的相應部分來描述。接著，說明測量物（例如，感官、物理、化學）中的樣品變量和樣品水準（在適當情況下），重複的次數，和設計的限制（例如，在測試階段可用的採樣批次、性質和數量）。描述為了減少試驗誤差所做出的努力。

感官方法。當描述所採用的方法時，使用本書中的術語（請參閱第 12 章，表 12.5、表 12.2），這和國際標準組織（1985 年，2004 年）和 IFT（1981）是相同的。

品評小組：每個試驗條件下的品評人員數量都應該陳述，因為它影響所獲得的統計學意義結果。如果使用過少的小組成員，需要很大的差異才具有統計學顯著性，而如果使用太多人（例如，三角測試種使用 1000 位品評人員），則可能會導致統計學上有意義但實際差異卻過小。應避免在試驗過程中改變品評小組，但如果發生這種情況，必須充分地描述。先前的培訓和當前測試的準備方法，小組成員的程度，包括使用任何參考標準等都需要有完整的說明，這些都是判斷結果的有效性的重要資訊。如果部分試驗有喜好性測試，應說明品評小組的組成條件（年齡，性別等）。

測試條件：測試地點的物理條件，以及製備和樣品呈現的方式，都是影響結果的可靠性和有效性的重要變量。報告應包括以下信息：

1. 測試區：測試位置（展位、商店、家、巴士）應當說明，以及任何存在的干擾（氣味、噪音、熱、冷、照明）應連同盡量減少其影響力的努力一併描述。

2. 樣品準備：應該描述設備和樣品製備的方法（時間、溫度、所使用的任何載具）。如果可以，指示及描述原料和配方。

3. 樣品呈現：描述應該使得讀者得以判斷可能被包含在結果的偏差程度，並且可以包括以下任何能夠影響的因素：

- 品評小組是否單獨工作，或是以團體進行
- 使用照明，如果是與平常不同的
- 樣本量、容器、器具、溫度
- 呈現順序（隨機、均衡）
- 樣品容器編碼，例如：三位數的隨機數目
- 任何特殊說明，如漱口，約被測樣品或變量的身份信息；樣本之間的時間間隔；樣品被吞嚥或咳出
- 任何其他可能影響結果的變量，例如，一天中的時刻，高或低的濕度，樣品的保存時間等。

統計技術：應該定義從實際測試得出的回應所轉換的報告數據的方式，例如，從分數轉換成排名。使用的統計類型以及所使用的假設（例如，常態）的程度等都应该放入討論，另外虛無假設和替代假設也應該納入，如果沒有太瑣碎。

12.2.4 結果與討論

結果應該使用表和圖的形式簡單明瞭的呈現，並且提供足夠的數據以證明結論。然而，同樣的資訊不能使用兩種形式呈現。表格數據一般都比較簡潔，除了趨勢和互動可能會更容易從數字中看出來。

結果的部分應該總結相關的收集數據和統計分析。所有的結果都應該顯示，包括那些違反假設的結果。顯著性測試（F、C²、t、r 等）的結果應列出機率水準、自由度、所獲得的檢驗統計量的值，以及效果的方向。

在討論部分中，結果的理論和實踐意義都應當指出，並和以前的知識連結。討論應該從簡要說明結果是否支持或不支持任何原假設開始。數據的解釋應該有邏輯的組織，並應該遵守試驗設計。應該解釋、比較和對比（有指示限制）結果，以及報告應該以明確的結論結束。

見表 12.1，其中說明了描述性研究的術語和尺度的發展。

表 12.6 報告範例：五種啤酒的特性

概要	
目的是什麼？	在他們很可能會提供的啤酒花基礎上選擇五種啤酒花進行測試，試點釀造使用啤酒花樣品 1、2、3、4 和 5，成本分別為\$ 1.00、\$ 1.20、\$ 1.40、\$ 1.60、\$ 1.80/磅；20 個訓練有素的啤酒廠品評人員使用由 0 到 9 的尺度判斷每個啤酒三次，樣品 4 所獲得評分為 3.9，比起樣品 2 和樣品 5 顯著較高，評分分別為 3.0 和 2.9。樣品 1 和樣品 3 的評分為 2.1 和 1.4，具有顯著最低。可以得出結論為樣品 4 和樣品 2，比起其他樣品來說，每美元提供更多啤酒花特性。
完成了什麼？	
結果是什麼？	
可以得出什麼結論？	
計畫目的、測試目的、試驗前協議	啤酒廠從多個供應商處獲得大量有代表性的樣品。該計畫的目的是基於他們提供啤酒花特色的能力進行選擇。測試目的是（1）在有意義的尺度上比較五個啤酒中啤酒花特色的程度和（2）測量結果的可靠性

試驗	
完成目的 1 和 2 的設計	設計—使用 5 個測試樣品釀造產出 14 BU 的標準苦味水準的啤酒。由啤酒廠品評小組 20 位成員評價測試的啤酒；測試共在不同天執行了三次

描述使用的感官測試

感官評估—受試者使用 0-9 的尺度評測啤酒花特色數量；參考標準如下；合成啤酒花在 1.0 毫克/升= 3.0 尺度單位，並且在 3 毫克/升= 6.0 尺度單位

描述品評小組：數量、培訓等

品評小組—依照過去評測啤酒花的績效選出 20 位小組成員；全部 20 名品評人員各測試所有樣品三次。

描述測試的條件：篩選品評小組的樣品資訊，品評小組的樣品呈現區域

樣品準備及呈現—測試啤酒儲存在 12°C，並在裝瓶後 7-10 天進行評測。樣品由兩位資深品評員篩選，使得各自代表的啤酒類型在顏色、泡沫，或風味上沒有除了啤酒花以外的差異。品評成員被告知樣品是由不同的啤酒花釀造，除此之外，沒有其他的啤酒特性被揭露。成員各自單獨在其展位進行工作，測試後並沒有進行討論。70 毫升的樣品在 8 盎司的玻璃容器，12°C 下呈現。這五個樣品是以平衡、隨機的順序同時呈現。樣品被品評人員吞下。

統計技巧

統計評測—結果透過變異數的裂區分析評價。

結果與討論

明確呈現結果

五個啤酒的平均結果顯示於表 1，且相應的統計分析顯示在表 2，其中樣品 4 的啤酒花特色比起其餘樣品得到顯著較高的評價（3.9）

給予足夠的數據來證明結論

表格 1

五種啤酒樣品的平均啤酒花評分

樣品	4	2	5	1	3
平均	<u>3.9^a</u>	<u>3.0</u>	<u>2.9</u>	<u>2.1</u>	<u>1.4</u>
使用的啤酒花，磅/桶	0.36	0.38	0.34	0.32	0.35

a 下方不是由一個共同的線連接的樣品在 5% 的顯著性水準下顯著不同。

（繼續）

表 12.6（續）

給定的機率水準、自由度、檢驗統計量所得的值

表 2 裂區 ANOVA 結果
廣場廣場°F

變異來源	自由度	平方和	均方	F
總計	299	975.64		
重複	2	8.89		
樣品	4	221.52	55.38	41.88 ^a
誤差 (A)	8	10.58	1.32	
品評人員	19	412.30	21.70	17.79 ^a
樣品 x 品評人員	76	89.81	1.18	0.97
誤差 (B)	190	232.53	1.22	

注：誤差 (A) 是由重複乘以樣品。誤差 (B) 通過減法計算。

a 在 1% 的水準顯著。

解釋數據，在試驗的設計以下

樣品 2 和 5 中，評分 3.0 和 2.9 幾乎相同，且啤酒花特性顯著低於樣品 4 的，但比樣品 1 和 3 顯著性高。統計評估顯示了受試者和樣品的相互作用沒有顯著意義 ($F = 0.97$)；因此可以假定，該小組成員在他們的評分上是一致的；品評人員效應 ($F = 17.79$) 顯著，表示小組成員使用不同部位的尺度來表達自己的看法；這並非少見；此外，當沒有相互作用時，受試者之間的差異是次要關注的。過程中的首要關注的問題是樣品之間的差異，使用 HSD 多重比較評價； $HSD_{5\%} = 0.7$ ，顯示於表 1 中的下方畫線的差異。使用不同啤酒花以釀造 14 BU 苦味的啤酒間的差異比起感知的啤酒花強度差異要小得多

以明確的結論結束

結論

檢測的 5 個樣品中，樣品 4 (\$ 1.60 / 磅) 產生的啤酒花特性有顯著較高的水準。樣品 2 (\$ 1.20) 是值得考慮的較便宜的啤酒

注：本報告涵蓋例 7.6，第 7 章中所描述的測試。

第十三章 基本統計方法

13 Basic Statistical Methods

13.1 簡介

應用統計的目的是從**母體**(population, 或譯為群體)的樣本(sample)所包含的資訊中得到一些結論。結論的類型分為兩大類：估計(Estimates)和推論(Inferences)。此外，從母體中取出的樣本大小和方式影響了由此產生的估計和推論的精準度。這些問題都在感官研究的實驗設計中有所解決。由於在感官評估中使用的方法與一些基本統計的方法相關，本章介紹了估計、推論和實驗設計的概念和技術。各節主題都包含了理論中最簡單的概念。

本章節中提出的幾種定義是基於讓讀者更容易理解其原則而進行相關討論。母體(Population)指的是所要探討元素的整體集合。在感官分析中所討論的母體在不同研究中是不同的。在一些情況下，母體可能是指人群(例如：特定食品的消費群)，而在其他案例，可能是指產品(例如，某批次玉米糖漿)。母體中的一個元素或單位可能是特殊的：單一消費者或特定批次糖漿。母體中的單一元素的測量有可能是不連續的，亦即，只採取特定的值(例如對於品牌A的偏好)，或連續的，即，取連續值中的任何值(例如甜味的強度)。測量值由機率分配所控制，通常是以數學方程式的形式表示一個特定的值發生的機率。與分佈相關的某些固定數稱為參數(parameter)。參數的值提供了有關母體的資訊。在連續分佈中，比如位於測量中心的平均值(μ , mean)。標準差(σ , Standard deviation)則是所有測量值與平均值的離散和分散程度。在不連續的分佈中，母體的比例(Proportion)含有某種我們要探討的特性。例如：一個二項式分佈的母體比例(p)可能引出結論為兩種產品的偏好分佈。

只有在非常罕見的情況下，才有可能進行母體普查，並直接計算母體參數的精確值。典型上我們將母體的元素子集合，稱為一個樣本(sample)，收集樣本，對樣本的每一元素進行我們感興趣的測量。這些測量的數學函數，稱為統計(statistics)，用來求未知母體參數的近似值。因此一個統計量的值稱為一個**估計值**(estimate)。

通常情況下，一個研究者想要確定一個母體所擁有的特殊特性(例如，越來越多的人比起B產品更喜歡A產品)。將母體視為整體，但其實可知的資訊只是一個樣本，而從其中得出的結論是有風險的。正規程序稱為**假設檢定**(tests of hypotheses)，設置了得出不正確結論的機率極限。然後，根據實驗的實際結果，研究人員的風險由這些已知的限制所框限。假設檢定是一種統計上的推斷，讓感官研究人員更加確信能夠做出正確的決策。

得出完美的統計結論所需的資訊數量取決於幾個因素(例如：研究者願意承擔的風險水準、所需的資訊精度、正在研究的母體的固有變異等等)。這些需要

解決的問題以及被稱為**實驗設計**（Experimental design）的行動計劃，必須在進行研究之前制定。實驗設計，基於技術和常識的原則，確保實驗資源集中在研究中的關鍵問題上，收集正確的資訊，以及沒有過多採樣客群和產品的狀況發生。

本章的其餘部分致力於新想法的進一步發展。13.2 節提出了將數據彙總成表格和圖形形式的一些基本技巧。13.3 節融合估計機率的一些基本概念，提出檢驗統計假設的一些方法。13.4 節介紹塞斯通模型（Thurstonian Model）在感官評估中的應用。塞斯通模型提供了測量樣品之間差異的另一種方法，並提供相關評估是如何執行的獨到見解。部分 13.5 涵蓋感官研究中所有最常見的實驗設計，包括用於改進品評小組，用於檢測產品之間差異的敏感度技術。在附錄中是從一些常見計算機率的分佈的基本技巧（見 13.6）。

13.2 彙總感官數據(Summarizing Sensory Data)

品評小組的評測數據應該在正式的統計分析（即，假設檢定）進行之前彙總成圖表。檢視圖表資料可能會發現在統計和機率的計算中數據會流失的特性。事實上，在圖表中顯示的特性可能指出此次使用標準統計分析程序對所收集的數據是不適當的。

不管收集了多少合理的觀察，任何數據分析的第一步是建立回應的頻率分佈（見圖 13.1）。然後計算一組基本的彙總統計(summary statistics)，包括為了估計分佈中心值的算術或樣本平均值， $\bar{x}$ ，以及為了估計數據相對於中間值的分散狀況，即樣本標準差(standard deviation) s 。樣本平均值的計算公式為

$$\bar{x} = (\sum_{i=1}^n X_i / n) = (X_1 + X_2 + \dots + X_n) / n, \quad (13.1)$$

儲存在 45 度及 90 度的沙拉醬的喜好性評分四周的頻率分佈

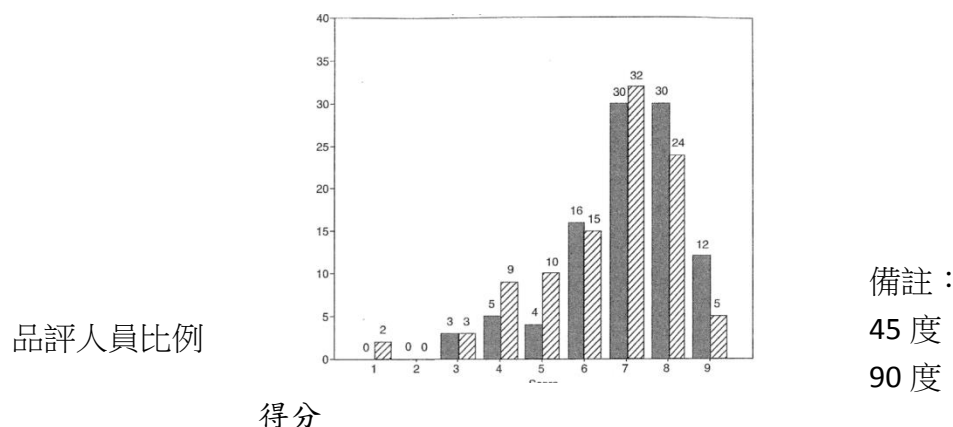


圖 13.1

兩種沙拉醬樣品的總體喜好分數直方圖（以頻率）。

其中， Σ 代表加總函數。下標 ($i=1$) 和上標 (n) 表示要加總的範圍。方程 13.1 表示樣品中所有 n 個元素的加總。樣本標準差的計算公式為

$$S = \sqrt{\frac{\sum_{i=1}^n X_i^2 - \frac{(\sum_{i=1}^n x_i)^2}{n}}{n-1}} \quad (13.2)$$

這些基本的統計數據有時可能會產生誤導。他們應該在數據是多峰 (multimodal) 等情況下謹慎使用 (在不同位置收集, 即回應量度不同的幾組數據的組合), 或是數據出現極端值 (即, 異常值) 時。多峰數據可能表示幾個具有不同平均值的子集的存在。在這種情況下, 所有數據的樣本平均值可能是毫無意義的; 樣本標準差也是相同無意義的 (因為採取的計算方式是圍繞平均值的散佈)。應進一步檢查多峰數據以確定是否可以打破母集合並形成新的數個子集 (例如: 按性別、年齡、地域、工廠、批次等)。然後分開計算各個子集的統計結論。如果整個集合是不可能拆散的, 則研究人員必須確定哪個片段的數據統計仍然有意義。例如: 中間值以下 50% 的數據, 或者是中間值以上 50% 的數據。這可能是識別一組多峰數據中心的一個有意義的方式。

同樣地, 數據可以從回應中的第一個和第三個四分位數之間分布的差異進行測量 (即: 分別低於 25% 及 75% 的值)。這種差異被稱為四分位範圍 (Interquartile range)。

樣本平均值, $\bar{x}$, 對於存在極端值的數據是非常敏感的。中位數則對極端值較不敏感, 因此它可以代替樣品平均作為代表數據中心的值。另一種估計中心趨勢的替代方案稱作**截尾平均** (Trimmed mean)。截尾平均以同樣的方式計算樣本均值, 但特定比例 (例如, 5%) 之外的最高和最低的數據值均已被淘汰。有數種電腦統計分析軟體可以計算各種常規的集中趨勢和分散方式。

很多統計分析程式是假設數據為標準分佈 (normally distributed)。如果計算 $\bar{x}$ 所使用的原始數據是標準分佈, 那麼 $\bar{x}$ 也是。事實上, 即使在原始數據不是像隨機變數一樣擁有標準分佈, $\bar{x}$ 仍然大致正常, 只要該樣本大小大於 25 左右。分佈平均 $\bar{x}$ 與原始數據的分佈平均, 即 μ , 相同。且如果 σ 是原始數據的標準差, 則 $\sigma/\sqrt{n}$ 是 $\bar{x}$ 的標準差。 $\sigma/\sqrt{n}$ 稱為**平均值的標準差** (Standard error of the mean)。

注意當樣本 n 的大小增加, 平均的標準差則減小。因此, 當樣本的大小變大時, $\bar{x}$ 是越來越有可能接近 μ 的真實值。平均值的標準誤差 (SE) 是由 $SE = S/\sqrt{n}$ 計算得出, 其中 s 是方程式 13.2 所計算出的樣本的標準差。

13.2.1 評級數據的總結分析

在分別四個不同城市所收集的 30 位個體的總體喜好回應呈現在表 13.1 中。回應的頻率分佈使用表格的方式呈現，如表 13.2，並使用如圖 13.2 簡單的點圖（dot-plots）呈現。在各個城市中並沒有強烈的多峰跡象。這些數據的彙總統計呈現於表 13.3 中。圖 13.3 的盒鬚圖（Box-and-whisker plots，見 Danzart 1986）提供了從各個城市的評級分佈的額外資訊，可能還能夠觀察到一些較不在意的極端值。（盒鬚圖(box-and-whisker plot，簡稱 boxplot)又稱長鬚圖，是一很有效的表示資料的方法。A 稱為資料的最小值，E 稱為最大值，B 與 D 則分別為資料的下四分位數及上四分位數，因此圖中盒子包含資料的中間 50%部分。又 C 為資料的中位數。上圖包含一個盒子，及二凸出來的鬚，這是此圖命名的由來。)

13.2.2 估計帶有特定特性的母體比例

用於估計一個二項式分佈的母體比例 p 的統計值是 $\hat{p}$ 其中，

$$\hat{p} = \frac{\text{成功次數}}{\text{測試次數}}$$

(13.3)

假設 150 位消費者參與 A 和 B 兩種樣品的偏好測試，接著，假設 86 位參與者表示他們更偏好樣品 A，而樣品 A 在試驗前便先定義為成功，所以從公式 13.3， $\hat{p} = 86/150 = 0.573$ 。即，0.573 或 57.3% 的消費者偏好樣品 A，假如此測試可以在多個城市中進行，則對於樣品 A 估計的偏好可以使用圖形方式如圖 13.4 的條型圖（bar chart）呈現。

13.2.3 μ 和 p 的信賴區間 (Confidence Intervals)

先前計算的單值統計，稱為**點估計** (Point estimates)，關於它們自身的精準度等並沒有提供更進一步的資訊。信賴區間彌補了這種缺少的資訊。

表 13.1 多個城市的單一產品消費者測試 (Multicity Monadic Consumer test) 數據

屬性：總體喜好度 ^a				
回應	亞特蘭大	波士頓	芝加哥	丹佛
1	12.6	10.4	7.9	10.3
2	9.8	10.4	7.8	11.7
3	8.6	8.9	6.3	11.5
4	9.8	8.0	11.1	9.9
5	15.0	10.4	5.5	11.7
6	12.7	11.0	6.5	10.3
7	12.8	7.4	8.8	11.6

8	9.5	10.5	5.2	12.1
9	12.4	9.2	7.8	11.6
10	9.6	9.2	7.6	12.3
11	9.2	9.8	6.3	12.4
12	7.1	9.1	7.1	10.5
13	9.9	9.7	8.0	12.4
14	12.4	10.3	5.7	14.4
15	8.7	9.1	5.5	11.1
16	11.9	10.3	5.2	9.9
17	9.9	11.7	7.2	11.9
18	11.3	9.8	8.0	8.8
19	10.4	10.2	9.1	12.3
20	11.8	9.5	8.4	8.6
21	11.5	12.4	4.0	11.9
22	8.9	9.5	6.9	9.3
23	11.4	12.9	6.6	10.0
24	6.9	11.1	7.4	10.2
25	8.8	13.3	7.3	10.8
26	11.6	12.9	7.5	12.7
27	11.3	11.4	9.1	11.1
28	9.7	9.0	6.9	11.9
29	10.0	10.1	8.4	10.2
30	11.2	11.2	6.1	10.1

a 使用 15 公分未組織的量度測量

信賴區間是值的一個範圍，其中參數的實值位於一個已知的機率中。信賴區間使得研究人員得以確定點估計是否足夠精確滿足調查的需求。

存在三種類型的信賴區間：上單尾信賴區間 (one-tailed upper confidence interval)、下單尾信賴區間 (one-tailed lower confidence interval) 及雙尾信賴區間 (two-tailed confidence interval)。用於計算 μ ，和 p 的這些信賴區間方程式呈現在表 13.4 中。在一般情況下，雙尾信賴區間是最有用的，但如果研究的主旨是只在單一方向，例如判斷平均值「過大」或「過小」，則使用適當的單尾信賴區間即可。

表 13.4 中的 $t_{\alpha, n-1}$ 及 $t_{\alpha/2, n-1}$ 的數值是 t 值統計 (t -statistics)。 α 的數值是測量的信賴程度。例如：如果 $\alpha = 0.05$ ，則信賴區間為 $100(1-\alpha)\% = 95\%$ 信賴區間。表 13.4 中的 $(n-1)$ 數值是與 t 分佈相關的參數，稱為**自由度** (Degree of freedom)。 t 的值取決於 α 的值和自由度 $(n-1)$ 的多寡。表 15.3 介紹 t 的臨界值。

表 13.4 數值 z 是標準常態分佈的變數臨界值。(標準常態分佈具有均值 $\mu = 0$ ，及標準差 $\sigma = 1$)

表 13.2 在表 13.1 中在多個城市中的消費者測試數據的頻率分佈

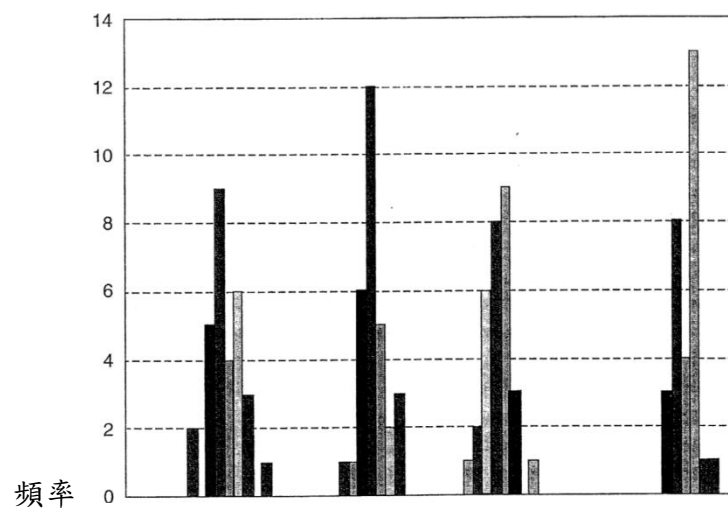
屬性：總體喜好度				
在各城市中的頻率				
類別中點	亞特蘭大	波士頓	芝加哥	丹佛
1	0	0	0	0
2	0	0	0	0
3	0	0	0	0
4	0	0	1	0

5	0	0	2	0
6	0	0	6	0
7	2	1	8	0
8	0	1	9	0
9	5	6	3	3
10	9	12	0	8
11	4	5	1	4
12	6	2	0	13
13	3	3	0	1
14	0	0	0	1
15	1	0	0	0

a 假設在亞特蘭大，總共 9 位參與者回應界於 9.5 及 10.4 之間的喜好
評
級

Z 在一些常用的 α 水準下的臨界值呈現在表 15.3 的最後一行。(即， ∞ 自由度所對應的那行)。

思考表 13.1 中的母體喜好度。亞特蘭大的樣本平均強度為 $\bar{x} = 10.56$ ，而樣本標準差為 $S = 1.79$ 。



城市： 亞特蘭大 波士頓 芝加哥 丹佛

圖 13.2 在表 13.1 中在多個城市中的消費者測試數據的總體喜好度直方圖

表 13.3 在表 13.1 中的多城市消費者測試數據的統計彙總

屬性：總體喜好度							
	城市	數量	平均	中位數	截尾平均	標準差	標準誤差
總體喜 好度	亞特蘭大	30	10.557	10.200	10.573	1.793	0.327
	波士頓	30	10.290	10.250	10.273	1.401	0.256
	芝加哥	30	7.173	7.250	7.146	1.448	0.264
	丹佛	30	11.117	11.300	11.115	1.276	0.233
	城市	最小值	最大值	第一分位	第三分		

		點 Q1		位點 Q3	
總體喜好 度	亞特蘭大	6.900	15.000	9.425	11.825
	波士頓	7.400	13.300	9.200	11.125
	芝加哥	4.000	11.100	6.250	8.000
	丹佛	8.600	14.400	10.175	11.950

為了建立整體平均值的下單尾 95%信賴區間，我們使用表 13.4 及表 15.3 以獲得

$$\bar{x} - \frac{t_{\alpha, n-1} s}{\sqrt{n}}$$

其中 $\alpha = 0.05$, $n=30$ ，所以 $t_{\alpha, n-1}$ 即為 $t_{0.05, 29} = 1.699$ ，得出

$$10.56 - 1.699(1.79)/\sqrt{30} = 10.56 - 0.56 = 10.00$$

此界限意即研究人員有 95%的信心水準認為在亞特蘭大的總體喜好度不低於 10.00

城市

亞特蘭大

波士頓

芝加哥

丹佛

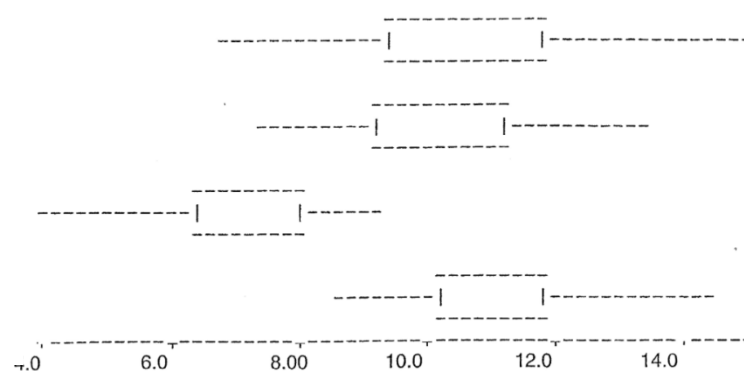


圖 13.3 表 13.1 中的多城市總體喜好度消費者測試數據的箱根圖

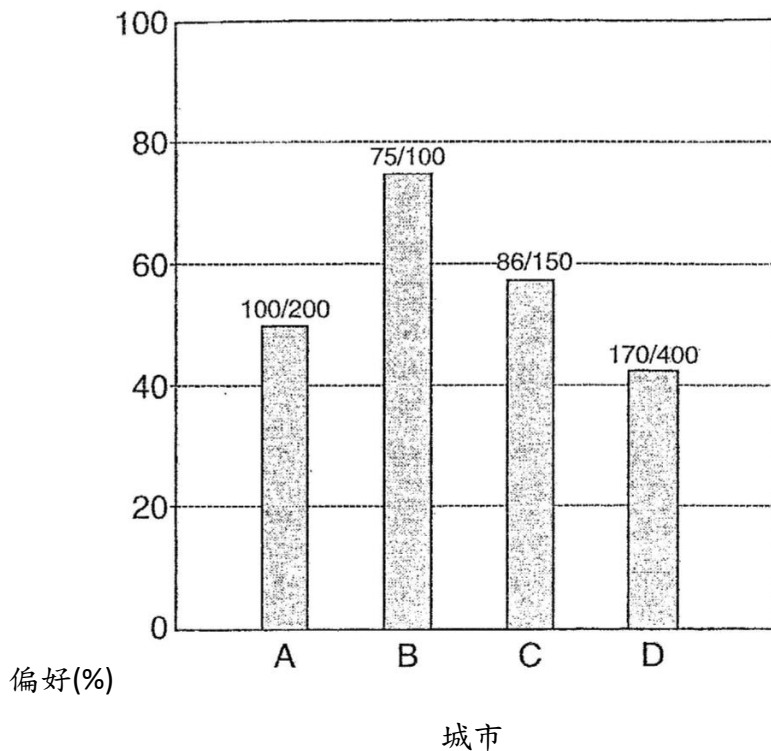


圖 13.4 在四個城市中進行的兩種樣品的偏好結果的長條圖，圖中顯示各城市有相對差異。每個城市皆包含了真實偏好結果以及全部回應的基礎。
平均數的 95% 雙尾信賴區間計算如下：

$$\bar{x} \pm t_{\frac{\alpha}{2}, n-1} s / \sqrt{n},$$

其中 $\alpha=0.05$, $n=30$ ，所以 $t_{\frac{\alpha}{2}, n-1}$ 即為 $t_{0.025, 29} = 2.045$ ，得出

$$10.56 \pm 2.045(1.79) / \sqrt{30} = 10.56 \pm 0.67 \text{ or } (9.89, 11.23)$$

也就是說，研究人員有 95% 的信心認為在亞特蘭大的總體喜好度平均介於 9.89 及 11.23 之間。在圖 13.5 中在四個城市進行的總體喜好測試的樣本平均及其相對應的 95% 信賴區間呈現在表 13.1 中。分析師可以開始建構關於此四個城市中的喜好平均差異的相關案子。

思考先前所舉的消費者偏好的討論，150 位受試者中有 86 位偏好 A 產品 ($\hat{p}=0.573$)，為了建立在按母體比例 p 的實值的 95% 雙尾信賴區間，我們使用表 13.4 以及表 15.3 以得出：

$$\hat{p} \pm z_{\alpha/2} \sqrt{\hat{p}(1-p)/n},$$

其中 $n=150$ ， $\alpha=0.05$ ，所以 $z_{\alpha/2}=z_{\frac{\alpha}{2}, \infty} = 1.96$ ，得出：

$$0.573 \pm 1.96 \sqrt{(0.573)(0.427)/150} \text{ or } (0.494, 0.652)$$

表 13.4 信賴區間的計算格式

信賴區間類型	參數	
	μ	P
上單尾	$\bar{x} + t_{\alpha, n-1} s / \sqrt{n}$	$\hat{p} + Z_{\alpha} \sqrt{\hat{p}(1-\hat{p}) / n}$
下單尾	$\bar{x} - t_{\alpha, n-1} s / \sqrt{n}$	$\hat{p} - Z_{\alpha} \sqrt{\hat{p}(1-\hat{p}) / n}$
雙尾	$\bar{x} \pm t_{\alpha, n-1} s / \sqrt{n}$	$\hat{p} \pm Z_{\alpha} \sqrt{\hat{p}(1-\hat{p}) / n}$

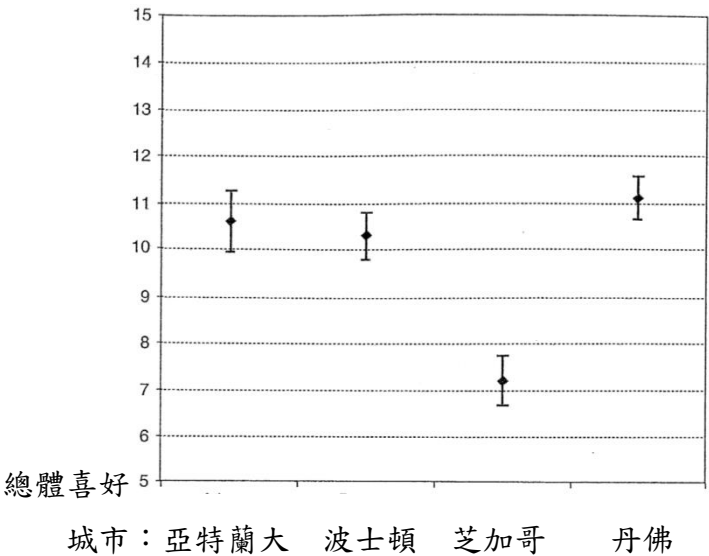
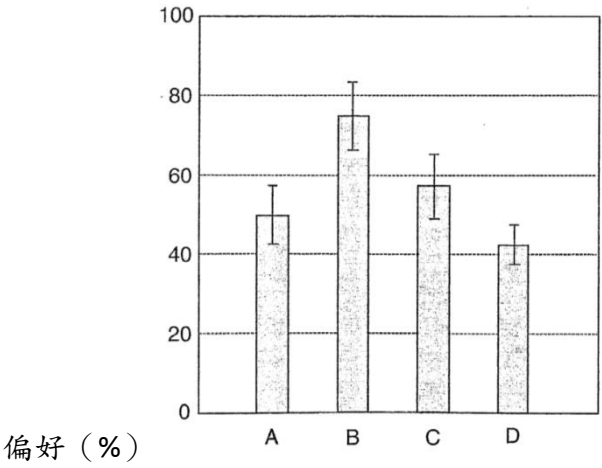


圖 13.5

表 13.1 多城市消費者測試數據的平均總體喜好分數的 95% 信賴區間。注意亞特蘭大、波士頓和丹佛之間有很大程度的重疊，而芝加哥的平均值則低得多。

研究人員有 95% 的信心得出結論，更喜歡樣品 A 的人口真實比例位於 49.4 % 和 65.2 % 之間。信賴區間和比例也可以以圖形方式描繪，如圖 13.6，其中 95 % 的雙尾信賴區間已被彙總到圖 13.4 的數據中。



城市

圖 13.6 包括在四個城市進行定期抽樣調查偏好結果的 95% 信賴區間的條形圖。與 50% 重疊的信賴區間表示在那個城市沒有顯著的偏好存在 ($\alpha = 0.05$)。兩個城市的信賴區間並沒有重疊，大致表示這兩個城市各自對產品的喜好程度有所不同。

13.2.4 其他的區間估計

信賴區間表示一個在已知機率下包含母體參數的實值範圍。研究人員可能並不總是想得出這樣的結論，因此另外存在其他類型的統計區間估計。

例如，**預測區間** (Prediction interval) 是在已知的機率下，一個具有包含 k 個未來觀察值的平均值的範圍。研究者可以選擇 $k = 1$ ，以在已知機率下，計算出下一個回應觀測值的區間 (例如，在 95% 信心下一批薯片的感知鹹味將位於 7.2 和 10.4 之間)。雙面預測區間計算為：

$$\bar{x} \pm t_{\alpha/2, n-1} s \sqrt{(1/K) + (1/n)}$$

另一個統計區間，稱為容忍區間 (tolerance interval)，是在已知機率下含有母體特定比例的值的範圍。一個單面容忍區間的例子是，研究者 95% 肯定此批次的 90% 的硬度評分小於 6.3。雙面容忍區間也可以計算的出來 (見 Dixon 和 Massey, 1969)。

13.2.5 數據轉換

偶爾研究人員可能希望轉換收集數據所用到測量的量度，變成另一種對於呈現上更有意義的量度。這種變換是很容易進行的，稱為**線性變換** (linear transformation)。假如原始變數 x ，變換成一個新的變數， y ，使用 $y = a(x) + b$ ，則 y 是 x 的線性轉換。線性轉換受限於恆定的原始變數乘以一個變數 a ，和/或添加一個常數 b 。假若希望將原始變數轉換成乘冪函數、對數、弦數、反函數等都是非線性變換。如果 x 具有均值 μ 和標準差 σ ，那麼 y 的均值和標準差分別是 $a\mu + b$ 及 $a\sigma$ 。這些計算轉化變數 y 的平均值和標準差的方程式僅適用於線性變換。樣本平均值， $\bar{y}$ 和樣本標準差， s_y ，通過 μ 替換 $\bar{x}$ 及 σ 替換 s_x 得出。

這種數據轉換技術的例子時常發生在整體差異測試，如三角測試法、二三點測試法和五取二測試法，其中原來測量正確反應的比例為 P_c 。以三角測試法為例， P_c 可轉換成可以區分樣本的母體比例， P_d ，藉由使用 $P_d = 1.5(P_c) - 0.5$ 的方程式。 P_d 的公式是藉由反轉在三角形測試得到正確答案的機率方程式所獲得， $P_c = 1(P_d) + 1/3(1 - P_d)$ ；也就是說，一個正確答案的機率是選擇一個有區別性的機率， P_d (總是會給出正確答案的人)，再加上無區別性的機率 ($1 - P_d$)，且其具有猜中的機率 (其中有 $1/3$ 的機率)。注意當在一個三角形測試樣品之間無感受

差異時，猜中正確答案的預期比例為 $P_c = 1/3$ ，這轉變為區分機率的預期比例 $P_d = 0$ （即，每個人都用猜的）。

在有 n 位受訪者的三角測試中，如果 x 位正確地選擇奇樣本，那麼 P_c 的估計值是 $\hat{P}_c = x/n$ ，且 P_c 的估計標準差為 $S_c = \sqrt{\hat{P}_c(1 - \hat{P}_c)/n}$ 。估計的區分比例則為的 $\hat{P}_d = 1.5(x/n) - 0.5$ ，其估計標準差 $S_d = 1.5S_c$ 。這些轉換應用出現在第六章的數個地方。

這些數據轉換在第六章介紹的統一差異測試方法中特別有用。信賴區間可以建構在品評小組的母體中的區分比例， P_d ，藉由使用

下單尾極限： $\hat{P}_d - Z_\alpha S_d$ ，以及

上單尾極限： $\hat{P}_d + Z_\beta S_d$

其中 $\hat{P}_d$ 是估計的區分比例， S_d 是區分比例的樣本標準差，而 Z_α 和 Z_β 是標準常態分佈的臨界值。 $\hat{P}_d$ 和 S_d 的數值由下列公式中的 $\hat{P}_c$ 和 S_c 求得：

方法	$\hat{P}_d$	S_d
三角測試	$1.5\hat{p}_c - 0.5$	$1.5S_c$
兩三點測試法及配對比較法	$2\hat{p}_c - 1$	$2S_c$
五取二測試法	$(10/9)\hat{p}_c - (1/9)$	$(10/9)S_c$

假如單尾信賴區間的下限是 0 或更少，則沒有明顯的差異的虛無假設不能被拒絕（在 $1-\alpha$ 信賴水準）。如果單尾信賴區間的下限大於零，則該樣品是明顯不同。如果信賴區間上限小於區分比例，也就是研究人員希望能夠檢測的比例， P_{max} ，則產品有足夠相似性（在 $1-\beta$ 信賴水準）。如果信賴區間上限大於 P_{max} ，則樣品不夠相似。（參見第六章使用這些信賴區間的例子。）

13.3 統計的假設檢定

調查的目的通常是為了確認假設是否合理，一個參數的未知值是否等於一些指定值，或可能兩個未知參數彼此是否相等。對於樣本中不完整且變化的資訊的情況下，通常在統計上使用假設檢定。統計假設檢定的方法，由以下五個步驟總結：

1. 使用數學術語中指出調查的目的，稱為虛無假設（ H_0 ），（例如， $H_0: \mu = 8$ ）。
2. 根據研究員的主要關注項目，另一個替代的數學術語叫做替代假設（ H_a ）（例如， $H_a: \mu > 8$ ， $H_a: \mu < 8$ ，或 $H_a: \mu \neq 8$ ）。
3. 從母體中收集樣本的元素，並且測量所採取的樣本中的每個元素。
4. 統計值被用來估算所關注的參數。
5. 根據測量的假定機率分佈，以及虛無假設， H_0 ，統計透過步驟 4 中計算出機

率值。如果此機率是比預定的機率（ α ）值小時，必需拒絕虛無假設，並有利於替代假設。

13.3.1 統計假設

在大多數的感官研究中，統計假設指明一個機率分佈中一些參數的值，例如平均值 μ 或人口比例 p 。由研究目標來確定虛無假設，其作為在實驗運行前就存在的基本假設條件。虛無假設所指定的特定值被用於計算假設檢定中的統計量（並且得出 p 值）。替代假設則是基於研究者的主要關注部分而建立的。例如，如果一個公司正在與備用供應商討論更便宜的成分，以替換當前產品的原料，感官分析師加入該研究只是為了確定在高水準的信賴度下，該產品使用更便宜的成分所製造的產品的顧客喜好度不低於該公司目前的產品。此研究的虛無假設和替代假設則為：

$$H_0 : P_{\text{當前}} = P_{\text{較便宜}}$$

與

$$H_a : P_{\text{當前}} > P_{\text{較便宜}}$$

其中 P_i 是母體中的喜歡 i 產品的比例。進行測試之前，虛無假設和替代假設都必須定義好。假如替代假設是在審查數據後才進行定義，統計測試的結果被往往偏向於拒絕虛無假設。

13.3.2 單面和雙面假設

有兩種類型的替代假設：單面替代（One-sided alternatives）和雙面替代（two-sides alternatives）。下面是使用單面和雙面替代的一些例子：

單面	雙面
確認釀造測試批是更苦	決定哪些釀造測試更苦
確認測試的產品是否優於控制組	決定哪個測試產品是首選
通常 H_a 的形式為：A 大於（小於）B，其中 A 和 B 同時指定	通常 H_a 的形式為：A 和 B 是不同的

研究人員經常很難決定另使用單面或雙面的假設。對於一個人的準則可能誤導另一個人。關於決定使用單面或雙面替代假設並沒有統計上的標準。替代假設的形式是由研究者的主要關注事項所決定的。如果研究者僅感興趣在兩個樣品是否不同時，則使用雙面替代假設。另一方面，如果研究人員希望測試兩個樣本之間的特定差異，即一個（指定的）樣品比起另一個（指定的）樣品，是更被偏好或更甜等，那麼使用單面替代假設。大多數替代假設是雙面的，除非研究者在研

究進行之前便指出其感興趣特定類型的差異。

對於單面與雙面的選擇上可能產生混亂，因為在幾種常見的感官測試情況，單面測試的統計數據常常用於雙面的替代假設測試中。

例如：在三角測試中，虛無假設只在大量正確的選擇情況下被拒絕（即，單尾測試標準）。然而，此替代假設是雙面的（即 H_a ：樣品可察覺出不同）。類似的情況會出現在 X^2 和F-測試中。

在實務中，研究者應該使用他們自己的話表達他們感興趣的研究面向（即虛無和替代假設）。如果研究人員的關注點明確的說明，則選擇替代假設是單面或雙面的便很容易決定。假若無法做到，那麼進一步的探測是必要的。感官分析師應該以研究人員陳述的關注（單面或雙面）面報告研究的結果，無論該統計方法是單尾或雙尾。

13.3.3 I 型和 II 型誤差

在測試統計假設時必須得出一些結論。有些結論可能是正確或不正確的。有兩種方式可能得出不正確的結論。首先，研究人員可以得出結論，虛無假設是錯誤的，而事實上這是對的（例如：當不存在差異但卻被認為有差異時）。這種誤差被稱為 **I 型誤差**。另外，研究者可以斷定虛無假設為真，或者更正確的說法是，虛無假設不被拒絕，事實上它卻是錯的（例如：沒有檢測到存在差異）。這種誤差被稱為 **II 型誤差**（見圖 13.7）。圖 13.8 介紹 I 型和 II 型誤差的實際暗示意義。

導致 I 型和 II 型誤差的機率在調查執行就已被指明。這些機率被用來確定研究所需的樣本大小。[參見，例如，Snedecor 和 Cochran (1980:102)。I 型誤差的機率等於 α ，而 II 型誤差的機率等於 β 。儘管 α 和 β 是機率（即，數字），它目前可與 I 型誤差和 II 型誤差互換使用。這個術語通用的作法在實務中常常造成一些混亂。

		決策	
		拒絕 H_0	不拒絕 H_0
真實	H_0 正確	I 型誤差 I 型誤差的機率= α	正確決策
	H_0 錯誤	正確決策	II 型誤差 II 型誤差的機率= β

圖 13.7 α 和 β 大小的 I 型誤差和 II 型誤差

真實	拒絕 H_0	不拒絕 H_0
H_0 正確	I 型誤差 ● 不應該選擇但卻選擇替代品	正確決策

	● 使用相同方法促銷新產品 ● 因為失去顧客信任而經銷失敗	
H_0 錯誤	正確決策	II 型誤差 ● 應該選擇替代品但卻沒選擇 ● 錯失候選產品 ● 浪費時間、金錢及付出的努力 ● 錯過黃金機會

(a) 在測試差異時

真實	拒絕 H_0	不拒絕 H_0
H_0 正確	I 型誤差 ● 應該選擇替代品但卻沒選擇 ● 錯失候選產品 ● 浪費時間、金錢及付出的努力 ● 錯過黃金機會	正確決策
H_0 錯誤	正確決策	II 型誤差 ● 不應該選擇但卻選擇替代品 ● 使用相同方法促銷新產品 ● 因為失去顧客信任而經銷失敗

(b) 在測試相似時

圖 13.8 I 型和 II 型誤差的現實意義。

II 型誤差的互補值，即 $1 - \beta$ ，被稱為**統計檢驗的功效**（power of the statistical test）。功效就是檢測該測試從虛無假設到一個給定的大小的偏離的（即：正確地拒絕錯誤的虛無假設）機率。在差異測試中，例如，虛無假設為 $H_0: P_d = 0\%$ 。測量從虛無假設偏離的值作為 $p_d > 0\%$ 。假設研究人員使用 40 位品評員進行兩三點測試，並且設定顯著性水準 $\alpha = 0.05$ 。如果母體的品評員的區分真實比例為 $P_d = 25\%$ ，則檢驗功效為 $1 - \beta = 0.44$ ，也就是說，此實驗設計在 $\alpha = 0.05$ 的顯著性水準下，品評人員能夠測試出樣品差異時，有 44% 的機率拒絕

虛無假設。檢驗功效被偏離虛無假設的大小，(即 P_d)、I 型誤差的大小(α 風險)、以及品評人員數， n ，所影響。

13.3.4 範例 13.4：平均值、標準差和比例的測試

本節介紹在進行整體平均下常態分佈的平均值、標準差(或成功的機率函數)的例行檢驗假設程序。

例 13.1：測試一個分佈的均值等於指定的值

假設 13.2.1 節的消費者測試例子中，感官分析師想要測試芝加哥的整體平均喜好平均是否等於 6 或大於 6。虛無假設及替代假設的數學形式如下：

$$H_o: \mu = 6$$

與

$$H_a: \mu > 6$$

替代假設是單面的。用於測試該假說的統計過程是單尾單樣本的 t 檢驗。檢驗統計的形式是：

$$t = (\bar{x} - \mu_{H_o}) / \left(\frac{s}{\sqrt{n}} \right) \quad (13.4)$$

$\bar{x}$ 和 s 的值由表 13.3 計算出。代入 13.4 的公式中得出：

$$t = \frac{(7.17-6)}{1.45\sqrt{30}} = 4.42 \quad (13.5)$$

比較 t 值和 t -分佈的上 α -臨界值，在 $(N-1)$ 的自由度下(表示為 $t_{\alpha, n-1}$)。 $t_{\alpha, n-1}$ 的值代表的意義是在 t -分佈中(具有 $(N-1)$ 自由度)所觀察大於 t 值的機率是 α 。如果公式 13.5 得到的值大於 $t_{\alpha, n-1}$ ，則虛無假設在 α -顯著性水準上拒絕。假設感官分析師決定使用 5% 控制 I 型誤差(即， $\alpha = 0.05$)。則從表 15.3 中，對應於 29 個自由度的那一排， $t_{0.05, 29} = 1.699$ 的值；因此，感官分析拒絕虛無假設，即在 5% 的顯著性水準上 $\mu = 6$ ，而接受替代假設的 $\mu > 6$ 。如果當時替代假設為 $H_a: \mu \neq 6$ (即，雙面替代假設)，則虛無假設也將會被拒絕，因 t 的絕對值(方程 13.5)大於 $t_{\alpha/2, n-1}$ ，也就是說， $|t| > t_{0.025, n-1} = 2.045$ (從表 15.3)。

範例 13.2：比較兩個平均—配對樣本案例

感官分析通常由單一品小組評估兩個樣品。當品評小組的每個成員評測兩種樣品時，使用成對 t 檢驗法(paired t -test)是適當的統計方法。在一般情況下，虛無假設可以指定任何感興趣的差異(即 $H_o: \delta = \mu_1 - \mu_2 = \delta_0$ ；設置 $\delta = 0$ 等同於測試 $H_o: \mu_1 = \mu_2$)。替代假設可以是雙面(即 $H_a: \delta \neq \delta_0$)，或單面(即

$H_a: \delta > \delta_0$ 或 $H_a: \delta < \delta_0$) 。)

表 13.5 例 13.2 成對 t 檢驗的數據和統計彙總

品評員	樣品 1	樣品 2	差異
1	7.3	5.7	1.6
2	8.4	5.2	3.2
3	8.7	5.9	2.8
4	7.6	5.3	2.3
5	8.0	6.1	1.9
6	7.1	4.3	2.8
7	8.0	5.7	2.3
8	7.5	3.8	3.7
9	6.9	4.5	2.4
10	7.4	5.0	2.4
			$\bar{\delta} = 2.54$
			$S_{\delta} = 0.61$

在這兩種情況下，成對的 t-統計的形式是

$$t = \frac{\bar{\delta} - \delta_0}{S_{\delta}/\sqrt{n}} \quad (13.6)$$

其中 $\bar{\delta}$ 是兩個樣品的平均差異， S_{δ} 是差異的樣本標準差。表 13.5 彙總所有品評員在單一屬性上的評分。分析師希望測試樣品 1 的平均得分是否比樣品 2 的平均得分大於兩個單位以上。在這種情況下，虛無假設是 $H_0: \delta \leq 2$ ，而替代假設 $H_a: \delta > 2$ 。檢驗統計的計算公式為：

$$t = \frac{2.54 - 2.00}{0.61/\sqrt{10}} = 2.79$$

其中， $n = 10$ ，因為有 10 位品評員，每位各貢獻樣本集一個差異的數據。如果 t 值超過了 t 分佈在 (N-1) 的自由度的上 α -臨界值（即， $t_{\alpha, n-1}$ ），拒絕虛無假設。分析師決定設置 $\alpha = 0.05$ ，並在表 15.3 找到 $t_{0.05, 9} = 1.833$ 的價值。t = 2.79 比 1.833 大，因此分析師拒絕虛無假設，並在 5% 的顯著性水準下出結論為，樣本 1 的平均得分是比樣品 2 的平均得分大兩個單位以上。

範例 13.3：比較兩個平均 - 獨立（雙樣本）案例

假設感官分析師已經在不同的時間點訓練了兩組不同的描述型品評小組，現在希望合併兩組人員。分析師希望在高度信賴水準下，兩組品評人員的評測成績是足夠相似而能夠把它們當作一個品評小組。

感官分析師進行幾個屬性品評，以確保這兩個小組是相似的。對於每個屬性，分析師呈現給所有品評人員同樣產品的樣品，並記錄他們的分數和它們所屬的組別。表 13.6 呈現此研究中的其中一組數據。該測試的虛無假設是 $H_0: \mu_1 = \mu_2$

表 13.3 例 13.35 中的雙樣本 t-檢驗數據和統計彙總

第 1 組		第 2 組	
品評員	分數	品評員	分數
1	6.2	1	6.7
2	7.5	2	7.6
3	5.9	3	6.3
4	6.8	4	7.2
5	6.5	5	6.7
6	6.0	6	6.5
7	7.0	7	7.0
		8	6.9
		9	6.1
$n_1 = 7$		$n_2 = 9$	
$\bar{x}_1 = 6.557$		$\bar{x}_2 = 6.778$	
$s_1 = 0.580$		$s_2 = 0.460$	

(或相等的， $H_0: \mu_1 - \mu_2 = 0$)。替代假設是 $H_a: \mu_1 \neq \mu_2$ (即雙面替代假設)。用於測試假設的檢驗統計是雙樣本 t-檢驗。檢驗統計的形式為：

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - \delta_0}{\sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}} \quad (13.7)$$

其中， δ_0 是在虛無假設 (在本例中 $\delta_0 = 0$) 中指明的差異。代入表 13.6 的值到公式 13.7 中得出：

$$t = \frac{(6.557 - 6.778) - 0}{\sqrt{\frac{(7 - 1)(0.580)^2 + (9 - 1)(0.460)^2}{7 + 9 - 2} \sqrt{\frac{1}{7} + \frac{1}{9}}}} = \frac{-0.221}{\sqrt{0.265} \sqrt{0.254}} = -0.85$$

比較 $t = -0.85$ 的值與 t-分佈在 $\alpha / 2$ 顯著性水準 (因為是雙面替代假設)，以及在自由度 $(n_1 + n_2 - 2)$ 的臨界值。本例中 (使用 $\alpha = 0.05$) 從表 15.3 找到 $t_{0.025, 14} = 2.145$ 。如果絕對值 (即忽略正負號) 的 t 大於 2.145，拒絕虛無假設。因為 $t = -0.85$ 的絕對值 (即， $|t| = 0.85$) 小於 $t_{0.025, 14} = 2.145$ ，感官分析師不

拒絕虛無假設，並得出結論為，兩個小組在這個屬性上就平均而言有類似的評級。

範例 13.4：比較兩個常態母體的標準差

在範例 13.3 中的感官分析師同時應該關注兩個小組的到的分數變異性是否是相同的。為了測試這兩個組的變異性是否相等，我們必須比較它們的標準差。此測試的虛無假設是 $H_0: \sigma_1 = \sigma_2$ 。替代假設是 $H_a: \sigma_1 \neq \sigma_2$ （即，雙面替代假設）。

用於檢驗這個假設的檢驗統計量是

$$F = \frac{S_{Larger}^2}{S_{Smaller}^2} \quad (13.8)$$

其中， S_{Larger}^2 是兩個樣本較大的標準差的平方，而 $S_{Smaller}^2$ 是兩個較小的樣本標準差的平方。而在本例中，從表 13.6 可看到，第一組具有較大的樣本標準差，所以 $S_{Larger}^2 = S_1^2$ ，而 $S_{Smaller}^2 = S_2^2$ 。因此方程式 13.8 的 F 值為：

$$F = (0.58)^2 / (0.46)^2 = 1.59。$$

比較 F 值和 F-分佈在 (n_1-1) 和 (n_2-1) 的自由度的上 $\alpha/2$ 的臨界值。（該自由度的分子是 (n_1-1) ，是因為這個例子中 $S_{Larger}^2 = S_1^2$ 。如果 $S_{Larger}^2 = S_2^2$ ，則自由度將是 (n_2-1) 和 (n_1-1) 。使用 $\alpha = 0.05$ 作為顯著性水準， $F_{0.025,6,8}$ 的值可在表 15.6 找到，即為 4.65。假若 $F > F_{\alpha/2, (n_1-1), (n_2-1)}$ ，則拒絕虛無假設。

因為 $F = 1.59 < F_{0.025,6,8} = 4.65$ ，虛無假設沒有在 5% 的顯著性水準下拒絕。感官分析師的結論是，沒有足夠的理由相信這兩個組在這個屬性上的得分變異性不同。

這是另外一個使用單尾統計檢驗測試雙面替代假設的例子。雙面替代拒絕虛無假設的標準是在 13.8 公式中的 F 值大於 $F_{\alpha/2, df_1, df_2}$ ，其中 df_1, df_2 分別是自由度的分子和分母。公式 13.8 仍用於單面替代假設（即 $H_a: s_1 > s_2$ ），但標準變成了「如果 $f > F_{\alpha/2, df_1, df_2}$ ，拒絕虛無假設。」

範例 13.5：測試母體比例是否等於特定的值

假設兩個樣品（A 和 B）在一個偏好測試中進行了比較。測試的目標是為了確定任一個樣品的是否受到 50% 以上的整體偏好。感官分析師收集 $n = 200$ 的隨機樣本，並且按照平衡、隨機的順序呈現給每位受試者兩項樣品，並詢問他們更喜歡哪個樣品。對於那些拒絕陳述偏好的受訪者，也就是「無偏好」的反應，我們將其得分平分至兩個樣本之間。研究發現，125 位受訪者表示他們偏好樣品

A，偏好樣品 A 的母體比例估計則根據方程式 13.3 為 $\hat{p}_A = \frac{125}{200} = 62.5\%$ 。

感官分析師任意地挑選「偏愛樣品 A」為「成功」的定義，並檢驗假設 $H_0: P_A = 50\%$ ，以及替代假設 $H_a: P_A \neq 50\%$ 。分析師選擇在 $\alpha = 0.01$ 的顯著性

水準下檢測此假設，並使用適當的 z-檢驗：

$$Z = \frac{\hat{p} - p_0}{\sqrt{(p_0)(1-p_0)/n}} \text{ 對於 } \hat{p} \text{ 以及 } p_0 \text{ 比例}$$

或者是

$$Z = \frac{\hat{p} - p_0}{\sqrt{(p_0)(100-p_0)/n}} \text{ 對於 } \hat{p} \text{ 以及 } p_0 \text{ 百分比} \quad (13.9)$$

其中， $\hat{p}$ 和 P_0 分別是 p 的觀測值和假設值。代入觀察值和假設值到公式 13.9 中得出：

$$Z = (62.5 - 50.0) / \sqrt{(50)(100 - 50) / 200} = 3.54。$$

比較 z 值與標準常態分佈的臨界值。對於雙面替代假設，比較 z 的絕對值以及

$z_{\alpha/2} = t_{\alpha/2, \infty}$ （至於單面替代假設，比較 z 值和 $z_{\alpha} = t_{\alpha, \infty}$ ），對照表 15.3，

$z_{0.005} = t_{0.005, \infty} = 2.576$ 。由於 $Z = 3.54$ 大於 2.576，拒絕虛無假設，分析師在 1% 的顯著性水準的結論是，50% 以上的母體偏好樣品 A。

範例 13.6：比較兩個母體比例

將範例 13.5 擴展到區域性的偏好。假設一個公司希望在兩個區域推出新產品（A），並想知道該產品在這兩個地區是否與其首要競爭對手的產品（B）具有同樣的偏好性。感官分析師在每個區域中進行 200 人的偏好測試並得到如表 13.7 的結果。

在本實例中的虛無假設是 $H_0: P_1 = P_2$ 與替代假設 $H_a: P_1 \neq P_2$ ，其中的 P_i 定義為在 i 區域中偏好產品 A 的母體比例。使用 χ^2 -檢驗測試此假設：

$$\chi^2 = \sum_{i=1}^r \sum_{j=1}^c (O_{ij} - E_{ij})^2 / E_{ij}, \quad (13.10)$$

其中 r 和 c 是如表 13.7 的數據表中的行和列的編號。 O_{ij} 是數據表中 i 行及 j 列的觀察值。 E_{ij} 是數據表中 i 行及 j 列「預期」的輸入價值。 E_{ij} 的計算方法為：

$$E_{ij} = (i \text{ 行的小計})(j \text{ 列的小計}) / (\text{總計})。$$

表 13.7 中的值代入公式 13.10：

$$\chi^2 = \frac{\left(125 - \frac{(200)(227)}{400}\right)^2}{\frac{(200)(227)}{400}} + \frac{\left(75 - \frac{(200)(173)}{400}\right)^2}{(200)(173)/400} + \frac{\left(102 - \frac{(200)(227)}{400}\right)^2}{(200)(227)/400} + \frac{\left(98 - \frac{(200)(173)}{400}\right)^2}{(200)(173)/400} =$$

$$\frac{(125-113.5)^2}{113.5} + \frac{(75-86.5)^2}{86.5} + \frac{(102-113.5)^2}{113.5} + \frac{(98-86.5)^2}{86.5} = 5.39$$

表 13.7 例 13.6 在兩個地區的口味偏好測試結果

偏好

區域	產品 A	產品 B	小計
1	125	200	75
2	102	98	200
總計	227	173	400

比較等式 13.10 的 x^2 的值和自由度 $(r-1)(c-1)$ 的 x^2 -分佈的上 α -臨界值。如果分析師選擇 $\alpha = 0.10$ (即, 10% 的顯著性水準), 則臨界值 $x_{0.10,1}^2 = 2.71$ (見表 15.5)。由於 $x^2 = 5.39 > x_{0.10,1}^2 = 2.71$, 分析師結論為在 10% 的顯著性水準該產品在這兩個地區的並沒有和產品 B 有一樣的偏好。可能必須考慮區域性配方。

13.3.5 計算差異測試中的樣本大小

差異測試中的樣本大小是測試的靈敏度參數的函數, α 、 β 以及 P_d , 或在方向差測試的情況下, p_{max} 。表 15.7、表 15.9、表 15.11、表 15.13 可以用於查找常選擇使用參數的樣本大小。另外, 研究人員可以使用電子表格來進行必要的計算。「敏感度分析測試("Test Sensitivity Analyzer")」已經併入到 Microsoft Excel 中, 可以讓研究人員觀察不同的 α 、 β 以及 P_d (或 P_{max}) 是如何影響樣本大小, 以及需要多少的正確回應才能判斷存在差異或者樣品是相似的 (見圖 13.9)。敏感度分析測試透過讓研究者在相同的樣本大小, n , 中, 選擇不同的值, 如: 正確反應數目, x , 和區別的最大允許比例, p_d (雖然 p_d 在方向差異測試中不具有意義, 鍵入的 p_d 的值實際是以 P_{max} 來計算) 來間接執行。敏感度分析測試接著計算 α 、 β 的值。藉著調整 n 、 x 和 P_d 的值, 研究者可以找到一組符合測試的敏感度要求同時可利用的資源之間的最佳折衷參數數值。

差異測試所基於的二項式分佈是, 是一種離散的機率分佈。樣本大小, n , 正確的反應, x , 必須是整數才是有效的。 n 和 x 的微小變化可能對 α 、 β 的機率產生巨大的影響 (特別是對於小的 n 值。一般來說是不可能選擇 n 、 x 、 p_d (或 p_{max}), 使得 α 、 β 正好等於其目標值。取而代之的是, 研究者必須適當選擇 n 、 x 、 p_d (或 p_{max}) 使得產生的 α 、 β 接近其目標的值。

如圖 13.9 所示, 研究人員希望進行相似性的兩三點測試, 並且符合目標的敏感度: $\alpha = 0.25$, $\beta = 0.10$ 且 $p_d = 25\%$ 。嚴格地說, 應該選擇使 α 、 β 不超過其目標值的 n 和 x 。然而, 研究人員只能找到 60 位人員進行評估。設定 $N = 60$, 研究者發現正確反應數目 $x = 33$ 得到的 α 、 β 值相當接近他們的目標, 雖然 $\alpha = 0.26$ 的值略大於所需的值。透過調整 n 和 x , 研究者發現, 將需要 $N = 67$ 評估, 並且正確反應數目 $x = 37$ 得到的 α 、 β 值是符合且低於他們的目標。研究者最後決定 60 位評估的測試是足夠的, 因為這是可進行評估的最大數目, 並且所得到的 α 風險只比目標值大 1%。

敏感度分析是規劃差異測試的有用工具。研究人員可以很快地選擇各種不同的 n 、 x 、 p_d (或 p_{max}) 值, 並觀察不同的情況中產生的影響和得到的 α 、 β 風險的變化, 最後選擇最好的折衷解決方案的值。敏感度分析測試可在 Excel 中通

過表 13.8 所顯示的編程在表格中執行。輸出的說明性文本可以選擇字體和大小，以達到所需的視覺效果。

在相似性測試(testing for similarity)或在統一方法(unified approach)中，正確反應數目， x ，不應選擇小於單獨的隨機預期機率（例如：在三角測試中， $n/3$ ；在二三點測試中， $n/2$ ）。

輸入				輸出			
受訪者數目	正確回應數目	猜測正確的機率	區分比例	P_{max} 或者是正確回答數機率 P_d	I 型誤差	II 型誤差	檢驗功效
n	x	p_0	p_d	p_{max}	α -風險	B-風險	$1-\beta$
60	33	0.50	0.25	0.625	0.2595	0.0923	0.9077

解釋：

33 個或更多的正確回應表示在 $\alpha=0.26$ 的顯著性下有差異存在

32 個或更少的正確回應表示有 91% 的把握 25% 的品評人員可以測試出差異，也就是在 $\beta=0.09$ 的顯著下有相對的相似性， $p_d = 25\%$ 。

指引：

- 請只在第四排的欄 A 到 D 中輸入數字！
 - 在格 A4 中輸入受訪者的數目 (n)。
 - 在格 B4 輸入正確反應的數量 (x)。
 - 在格 C4 中輸入正確猜測的機率 (p_0)。
 - (例如，在三角測試中， $c4=1/3$ ；在二三點測試中， $c4=1/2$)。
 - 在格 D4 中輸入你希望能夠檢測的區分比例。
- 評估第四排的欄 F 到 H 的結果，決定測試是否具有足夠的敏感度。
 - 在差異測試中，選擇 α —風險較小的值。（調整 n 和 x 以獲得所需的敏感度（即， α —風險））。
 - 在相似性測試中，選擇， β —風險較小的值。（調整 n ， p_d 以實現希望的敏感度（在 p_d 和 β —風險兩者之一取得平衡。）。（不要選擇 x 值低於預測猜測的機率（例如，在三角測試中， $n/3$ ；在二三點測試中， $n/2$ ）。（需要的話可以增加 n ）
 - 當同時測試相似性和差異性時，選擇 α 和 B-風險都可以接受的小值。（調整 n 、 x 和 p_d 以達到所需的敏感度（即， α —風險、 β —風險、 p_d 之間的平衡）。
 - 在雙面方向性差異測試中，加倍 α —風險的計算值以將測試的雙面性質納入考量。

圖 13.9 敏感度分析測試，說明 n ， x ，和 p_d 在 $\alpha=0.25$ 、 $\beta=0.1$ 的兩三點測試中的值。注意 α —風險略大於指定的目標值。

表 13.8 Excel 的敏感度分析測試編程資訊

格	輸入
E4	=D4+C4 (1- D4) (如果需要， 可以隱藏)
F4	=1-BINOMDIST (B4-1, A4, C4, TRUE)
G4	=BINOMDIST (B4-1, A4, E4, TRUE)
H4	=1-G4
A7	=B4
F7	= F4
A9	=B4-1
F9	=H4 (使用%格式)
B10	=D4 (使用%格式)
D11	=B10 (使用%格式)
F11	=G4

這些值等同於區分的比例值為負數 ($PD < 0$)。邏輯上是不可能被用作測試的決定條件。在使用敏感度測試器進行雙面方向性差異測試時，研究人員必須記住 α 一風險要加倍計算，以考慮測試的雙面性。

13.4 賽斯通量度 (Thurstonian Scaling)

雖然百分比區分模型由於便於解釋而出現，它卻不是人類行為的理論模型。它用於規劃差異測試是有用的，但在一般情況下，它過度簡化了感官品評的行為。如果事實上一個人必然是能夠區分或不能區分的兩者之一，則在三角測試中應該也重複出現此行為，例如：一組評估者應該有 100% 的正確反應（能夠區分），而另一組則應該具有 1/3 正確回答（不能區分）。但這在實務中並不是如此。一個更精準預測人類行為的模型是必要的，以便更加了解差異測試觀察到的結果。塞斯通量度 (Thurstonian 1927) 是一個這樣的模型。下面所概述的賽斯通量度是 ASTM (2003) 中由 Ennis (2001) 整理的最新的釋義。

13.4.1 感官差異的基本方法

塞斯通模型是基於兩個假設。第一個假設是，感知有遵循常態機率規則的機率元素。第二個假設是，品評員可以忠實地執行他們被要求執行的感官任務，以及相關的決策規則。

塞斯通模型認識到，當一個品評員重複執行同一個產品時產生的評價和看法會有所變化。這樣的變化可能來自於在評估產品時瞬時產生的生理或心理的變

化，或這些變化的組合。不管來源為何，感知在每次評估中皆會變化。塞斯通模型假設感知的變化機率遵循非特定量度的感官規模的常態分佈。為了不失去普遍性，最簡單的塞斯通模型假定感知的分佈落入有關樣本的平均幅度與一個標準差（見圖 13.10）。

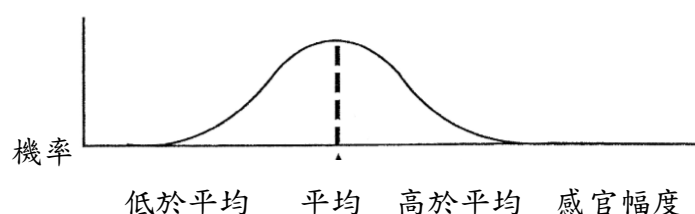


圖 13.10

感官知覺的常態分佈。一個品評員的感知變化根據常態機率分佈而變化

正是因為感知的變異過程，導致品評人員混淆產品有著「平均」不同的變化。當兩個產品具有非常不同的平均感覺幅度時，它們將不會被混淆，因其感知分佈並沒有重疊。例如：在圖 13.11a，所有產品 B 的感知幅度高於產品 A 的感知幅度。然而，當產品之間的差異較小時，兩個分佈的重疊區域更多，這兩種產品便有可能產生混淆。在圖 13.11b，雖然產品 B 的平均比產品 A 更高，在一次單一的評測中產品 B 可能被認為更接近於產品 A 分佈的平均值，比起其自身的分佈平均值。發生這種可能性是與兩個分佈的平均值之間的距離成比例關係。兩個平均值的距離，稱之為 δ ，也就是在賽斯通模式中的是感官差異的基本方法。 δ 是分開兩個分佈平均值的標準差的數量。如果 δ 小，兩款產品都差不多。如果 δ 大，這兩種產品是顯然的不同的。 δ 的值可使用任一種必選法（forced-choice）來估計，或是類別量度法（category-scaling）。用於估計 δ 的統計值是 d' 。

13.4.2 感官差異測試的決策規則

品評人員在必選法或者類別量度法中所用來架構其測試反應的決策規則在每個測試中皆不同。例如，在一個二三點測試中，決策規則是選擇那些被認為是最接近參考樣品的樣本編碼，而在二選項必選法（2-AFC）中，決策規則是挑選「較強的樣品」。表 13.9 介紹幾個必選法測試的決策規則。其中的重要假設是品評人員在評估產品時，總是正確地使用所屬的決策規則。換句話說，品評員總會給出他們認為是正確的答案，即使感知的變異可能導致他們的答案不符合產品實際的差異。

我們使用三角測試的例子說明這一點。在三角測試中，提供品評人員三個編碼樣品。假設兩個樣品中的其中之一是產品 A，而另一個是產品 B。評估人員的任務是確定出「奇異」樣本一即，識別三個樣品中哪個是與其他兩個最不同。評估員將始終選擇他或她在三角測試中所感覺到最「奇異」的樣品。

圖 13.12 顯示三角測試中的兩種可能結果。在第一次測試中，品評人員選擇 B 樣品，因為他感知的差異遠離另外兩個 A 樣品的感知幅度。

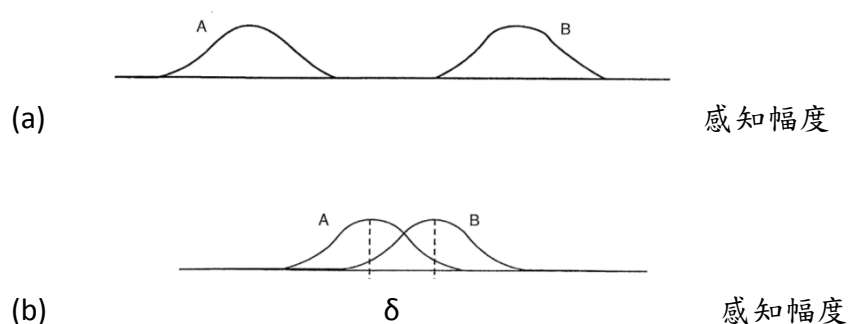
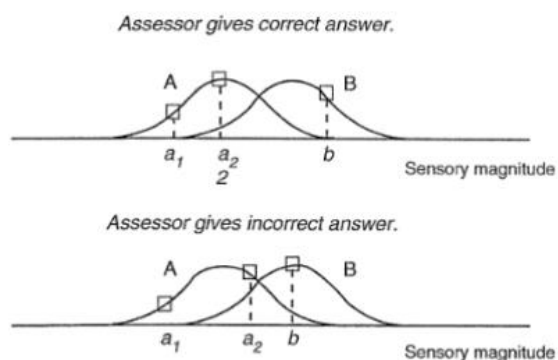


圖 13.11 (a) 兩種產品塞斯通表現有很大的不同。 (二) 兩種產品的賽斯通表現是混淆的，因為他們的感官分佈重疊。重疊度和兩個平均值之間的距離成比例關係，即 δ 。

表 13.9 二選項必選法 (2-AFC)、三選項必選法 (3-AFC)、三角測試法、兩三點測試法、A 非 A 檢驗法，以及同樣差異性對比法 (Same-difference method) 的決策規則

方法		決策規則
二選項必選法 (2-AFC)		$P_c = P(b > a)$
三選項必選法 (3-AFC)		$P_c = P(b > a_1 \text{ 及 } b > a_2)$
三角測試法		$P_c = P(a_1 - a_2 < a_1 - b \text{ 及 } a_1 - a_2 < a_2 - b)$
兩三點測試		$P_c = P(a_R - a < a_R - b)$
A 非 A 檢驗法		$P_a = P(a > c)$ $P_{na} = P(na > c)$
同樣差異性對比法		$P(S/U) = P(b - a < \tau)$ $P(S/M) = P(x_1 - x_2 < \tau), x =$ $a \text{ 或 } b$

在這次測試中，品評員給出一個正確的答案，因為評估的看法是與產品之間的實際差異相符的。在第二次測試中，評估選擇 A_1 樣品，因為感知上，它遠離 A_2 和 B 樣品，以及 A_2 和 B 樣品與其它樣品的距離。在這次測試中，評估人員給出了一個不正確的答案，因為評估的看法是與產品之間的實際差異不符合。在這兩次測試中，品評人員皆正確遵守了決策規則。然而，由於感知機率的性質，



品評人員給出正確答案

感官幅度

品評人員給出不正確答案

感官幅度

圖 13.12 在三角測試中的正確和不正確的答案。在上面的三角測試中，品評人員給出了正確的答案，即正確回答 b 是奇異樣品，因為感知上， A_1 和 A_2 是更遠離 B，而且比他們彼此的距離還遠。然而在底部的測試中，品評人員給出錯誤的答案，即選擇 A_1 為奇異樣品。因為他感知樣品 A_1 距離樣品 A_2 和樣品 B 的差異更遠，也遠過 A_2 和 B 彼此的差異。在這兩種情況中，品評員皆正確地使用決策規則。在一次測試中的答案是正確的，在另外一次的答案卻不正確。

Ennis 表示由於決策規則的差異，差異測試在評定一個給定大小的感官差異， δ ，的判別能力不同。在一般情況下，更複雜的決策規則需要更多的評估，以在測試中提供相同的統計功效。例如，2 選項必選法和 3 選項必選法測試比起兩三點測試和三角測試更強大。Frijters (1979) 採用這種方法來解決「無鑑別的差異測試矛盾」。

13.4.3 估計 δ 值

13.4.3.1 必選法(Forced-Choice Methods)

在必選法中正確答案的比例隨著產品之間的距離比例增加。正確答案的比例可以與測試的決策規則組合以估計 δ 值。d' 表，為廣泛使用於估計 δ 的統計資訊(參見 ASTM(2003)、Elliott(1964)，和 Ennis(1993))。此外，ASTM(2003)，和 Bi, Ennis 和 O' Mahony (1997) 發表了 d' 的變異表。瞭解估計的變異使得研究人員能夠建立信賴區間，以及關於 δ 的實際值的統計檢驗。也可以使用軟體計算 d' 和其相關聯的統計數據，例如：從感知機構。

獲得 d' 值的過程在所有必選法 (2-AFC, 3-AFC, 二人，三人，和三角形) 中都是同樣的：設定一個固定、虛無假設的正確回應的機率。對每個測試來說，d' 的值是與所觀察到的正確反應成比例關係的。但這個情況在 A 非 A 檢驗及差異性

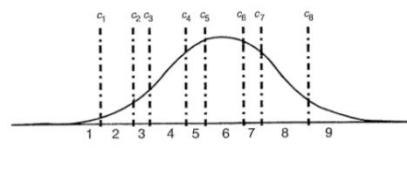
對比法中並不成立。這兩種測試方法涉及一種安慰劑效應。例如，在 A 非 A 檢驗中，品評人員會容忍 A-分布平均值周圍的偏差，並且認為感知的樣品是從產品 A 所製成。當感知距離 A-分布的平均太遠，品評人員將會清楚地將樣品劃分為「非 A」，該品評人員做出此決定時所使用的邊界被稱為 c 。同樣地，在差異性對比法中，兩個樣品仍然被歸類為「相同」，但之間存在的最大差異值為 τ （見表 13.9）。這些新的參數稱為認知標準，影響了在這些方法中獲得 d' 值的過程。此外，認知標準可以揭露品評人員在容忍他們自然感知變化中的變化程度。

13.4.3.2 使用量度的方法

塞斯通模型在類別量度的 A 非 A 檢驗使用了 c 標準。多個 c 值被定義為數個類別量度的界限。例如，在一個 9 點量度中，從 c_1 到 c_8 總共有 8 個 c 值。若接收的評級為 5，則樣品的感知強度必須大於 c_4 的界線，但比 c_5 界線更小，如圖 13.13 所示。 c 值在感官維度中往往不是等距。間距不相等是由於品評人員的行為差異。例如，在 9 分制中中性的評級 5 比起相鄰的評級更窄，那是因為品評人員傾向給予正面或負面的回應，儘管是只感受到很微弱的差異。

此外，藉由估計樣品之間的差異 d' ，識別類別之間的界限揭露了品評人員如何使用量度等有趣的資訊。

圖 13.13 多個 c 標準形成類別量度的界線。為了接收評級 5，樣品的感知強度必須大於 c_4 但小於 c_5 。



在全部類別量度中計算 d' 的方式太複雜，不好在這裡解釋。然而，ASTM (2003) 提供了分類成類別量度數據以及 A 非 A 檢驗法數據的兩種快速查表法。（收集的數據是不論多少物理量度）。由於數種方法拆分成兩大類，許多細節已流失，但 d' 及其變異數的值仍然準確。

13.5 感官研究小組的統計設計

在本節中介紹在感官評測中常用的實驗設計。我們結構化討論以避免圍繞這主題太多的混淆。在第 13.5.1 節，實驗的獨立重複和單個樣品的多次觀察是區分開來的。沒有認識到母體關注的多個觀測的混淆複製，可能會導致感官數據的統計分析不正確，並且造成測量誤差。當這種情況發生時，樣品經常被宣稱為具有顯著不同，但它們並沒有。在第 13.5.2 節，呈現了感官小組研究中最常用的實驗設計。這些方法包括隨機區集（完成）設計，平衡不完全區集設計，拉丁方塊設計（Latin-square design）和裂區設計（Split-plot design）。

1 3.5.1 抽樣：重複與多次觀察

統計分析的實驗設計基本目的是要準確和精確的估計實驗誤差。所有的假設檢定和信心度聲明都基於此原則。實驗誤差是研究中的母體無法解釋的自然變異。實驗誤差以數量表示即為變異，或是母體的標準差。從母體中提取一個單位的測量無法提供平均，因此無法估計實驗誤差。實際上，相同單位的多次觀測也無法提供平均估計實驗誤差。而一個單位的多次觀測之間的差異是由測量誤差造成的。同一個母體中的數個單位需要被採樣以建立估計實驗誤差的有效估計。針對不同單位取得的測量被稱為重複。它是包含有關母體變異性資訊（即，在實驗誤差）的單位跟單位之間的差異。

感官研究的共同目標是基於某些屬性的感知強度的差異來區分產品。如果每個產品只有一個樣品供評估（批次、罐、製備等），也沒有辦法估計母體產品的實驗誤差。通常情況下測量誤差，也就是品評員之間的變異，在感官評估數據的統計分析中以實驗誤差替代。這是一個非常危險的錯誤，因為忽略了實驗誤差，並使用測量誤差替代它，可能會導致分析師錯誤地得出產品之間存在顯著差異的結論，但事實上並沒有如此的差異存在。評估單一批產品忽略批次間的差異，可能在本質上貢獻了產品的變異度。正如對一個單獨個體的身高重複測量無法告訴我們人和人之間的差異，一個單一產品（不管品評小組大小的）的重複評估也無法告訴我們產品批次與批次之間的變化。

測量誤差是真實存在的（專業感官人士都非常瞭解）。然而，測量誤差不能隨便的以實驗誤差替代，因為誤導統計分析結果的風險非常大。如果在味覺測試中，同一罐蛋黃醬被分成 20 份呈現給品評人員，或同一甜味劑溶液被分成 20 杯評估，則測試的結果是等同於重複測量同一個人的高度。研究得到的變異性的估計是測量誤差的估計。它並不是產品變異性（有效的實驗誤差），因為並未呈現該產品的獨立重複（例如，不同批次）的量度。能從這樣的研究中得出的唯一合理的結論是品評人員是否能夠評估特定樣品之間的差異。這和產品是否有差異是不一樣的結論，因為任何觀察到的差異都是相同的產品，並沒有辦法來評估未來不同批次的產品差異。

為了避免混淆多個觀測與獨立重複，感官分析師必須清楚地瞭解正在研究哪個母體。假如一項研究的目的是比較幾個品牌的產品，那麼必須進行評估各個品牌的產品。如果已知一種成分是非常均勻的，那麼最起碼，每個樣品的單獨製備加上此樣品應個別呈現給每位品評員。（對於非常均勻的產品，變異的主要來源很可能是在準備過程中的差異）。

不可否認地，從單罐產品中準備每份樣品比起從不同罐產品中取出樣品還來得方便。但是，感官分析師必須比較測試的方便性和得到錯誤結論的代價，例如一個原型品製劑因為基於單一批次的產品評價，而得出錯誤的結論為顯著優於當前製劑因而替換，造成的市場失敗的損失。

13.5.2 區集一實驗設計(Blocking an Experimental Design)

實驗設計的區集結構是如何將處理施加到實驗材料的描述。要理解區集結構，兩個概念必須知道：「區集」和「實驗單位」。一個區集簡單來說就是一組同質的實驗材料。理論上來說，一個區集內的任何單位將對於一個給定的處理有著相同的反應。反應的水準隨著區集的不同而變化，但是在同一區集中應用的任何兩種處理之間的差異將在所有區集中維持常數。一個區集內的實驗材料被分成更小的組別稱為實驗單位。一個實驗單位是處理單獨應用在整個實驗材料的一部分。

在實驗運行之前便先考慮到區集間的差異，則研究的敏感度增加。如果處理的應用適當的區集效果可以從處理效果和從實驗誤差分離出來，從而我們能夠「清楚」地觀察處理效果，同時減少研究中原因不明的變異性。

在感官測試中用更通俗的術語來說，實驗材料是由品評員進行評估的一組大的評估。這些評估通常根據品評員被安排成區集。舉例說明，由於不同的閾值，品評員可以使用評量量度的不同部位來表達他們的看法。任意兩個樣品之間的感知差別的大小對於不同品評員來說是假設相同的。每個品評員（即區集）的單一評測為實驗單位。處理在這一點上可以被認為是產品，必須在每個評估中獨立使用。為了讓回應返回感知的一些基本水準（所有評價為常數），必須透過以上所述技術，加上隨機呈現的順序，連續單元的呈現，以及間斷並足夠的測試時間而達成。

13.5.2.1 完全隨機設計 (Completely Randomized Designs)

最簡單的區集結構是完全隨機設計 (CRD)。在 CRD 中，所有的實驗材料是均勻的；即 CRD 包含一大組實驗單位。使用 CRD 的時機為，舉例來說，當單一產品可以在數個位置被不同組品評人員（例如，單一產品多城市的消費者測試）評估。在此情況下，因為位置而造成差異的顯著性在每個位置內發生的變異被清楚地確定。

表 13.11 為符合 CRD 總體喜好性數據。圖 13.1 給出的箱根圖和圖 13.5 的信賴區間表明，一些城市跟城市之間的差異可能存在。平均喜好回應可以用來總結城市和城市之間的差異。變異數分析 (ANOVA) 被用於確定在平均喜好性上所觀察到的差異是否在統計上是顯著的。這些數據列於表 13.10 的 ANOVA 表中。

在表 13.10 中的城市 F-比例是高度顯著 ($F_{0.01,3,116} = 3.95$)，這表示至少一些城市之間觀察到的差異是真實的。為了確定哪些平均值是顯著不同，必須應用另一種稱為**多重比較過程** (Multiple comparisons procedure) 的統計方法。在此實例中，使用被稱為 (Fisher's Least Significant Difference, LSD) 的多重比較技術。在一般情況下，使用 LSD 比較兩個平均值，： $\bar{x}_i$ 和 $\bar{x}_j$ ，計算如下：

$$LSD_{\alpha} = t_{\frac{\alpha}{2}, df_E} \sqrt{MS_E \left(\frac{1}{n_i} + \frac{1}{n_j} \right)}, \quad (13.11)$$

其中 $t_{\frac{\alpha}{2}, df_E}$ 是自由度為 df_E 值的 t-分佈的上 α -臨界值 (即，ANOVA 中的誤差自由

度)， MS_E 是 ANOVA 分析的誤差均方。

表 13.10 於表 13.1 的多城市消費者測試數據－完全隨機設計的 ANOVA 分析表

變異度來源	自由度	平方和	均方	F
總計	119	541.56		
城市	3	283.34	94.45	42.43
誤差	116	258.22	2.23	

而 n_i 和 n_j 分別是用於計算 $\bar{x}_i$ 和 $\bar{x}_j$ 的觀測值數目。如果所有的 $\bar{x}$ 樣本大小都相同的話，那麼等式 13.11 簡化為

$$LSD_{\alpha} = t_{\alpha/2, df_E} \sqrt{2MS_E/n} \quad (13.12)$$

其中 n 是共同的樣本大小。在此例中， $n = 30$ ，所以 $LSD_{0.05}$
 $= 1.96 \sqrt{2(2.23)/30} = 0.76$ 。任何兩個樣品的平均差異大於 0.76，則代表在 5% 的顯著水準下是不同的。如表 13.11 所示，芝加哥比起其他三個城市有顯著性較低的平均值，而波士頓則是比丹佛的平均值來的顯著性低。
 完全隨機的設計很少用於涉及感官小組的多重樣品研究，因為使每個小組成員只評估一個樣品是低效率的，但應當知道的是，不同的品評人員可能使用不同部位的量度來表達自己的看法。進行這樣的研究需要更精細的品評小組設計。最常用的四種品評小組的設計在本節的剩餘部分討論。

13.5.3 隨機（完全）區集設計

如果採樣的數量足夠小，使得感覺疲勞不是問題，那麼隨機的（完全）區集設計是合適的。品評員是「區集」；樣品是「處理」。每個小組成員評估（通過評級或排名）所有樣品（因此稱為「完全的區集」）。

當感官分析師有足夠信心認為品評小組的品評員排名的樣品皆是一致，但仍須承認品評員可能使用不同部分的量度來表達他們的看法，因此隨機區集設計是有效的。從隨機區集設計應用於數據分析時，須考慮到品評員之間這種類型的差異而產生的實驗誤差，以做出更準確的估計和假設，進而進行敏感度更高的測試。

獨立地複製測試產品，以隨機順序呈現樣品給品評人員（使用每個品評員是各自單獨地隨機化）。品評員的評價數據顯示在如表 13.12 的雙向表（two-way table）中。

13.5.3.1 隨機區集評級分析

從隨機區集設計所獲得的評級數據進行 ANOVA 分析。適合隨機區集設計的

ANOVA 的形式列於表 13.13。虛無假設是，所有樣品的平均評分相同($H_0: \mu_i = \mu_j$ ，對於所有 i 到 j 的樣品)，替代假設則是至少兩個樣品的平均評分不同 ($H_a: \mu_i \neq \mu_j$ ，i 到 j 的其中幾對樣品)。

表 13.11 為表 13.1 從多城市中收集的單一產品消費者測試數據的平均整體喜愛評分

城市	亞特蘭大	波士頓	芝加哥	丹佛
平均評級	10.56BC	10.29B	7.17A	11.12C

註：平均後面的字母不同代表在 5% 的水準下顯著不同。

表 13.12 隨機（完全）區集設計數據表

樣品							
區集（品評員）	1	2	.	.	.	t	行小計
1	X_{11}	X_{12}	$\dots$	.	.	X_{1t}	$X_{1.} = \sum_{j=1}^t x_{1j}$
2	X_{21}	X_{22}	.	.	.	X_{2t}	$X_{2.} = \sum_{j=1}^t x_{2j}$
...	...	...	...	...	...	...	
B	X_{b1}	X_{b2}	.	.	.	X_{bt}	$X_{b.} = \sum_{j=1}^t x_{bj}$
欄小計	$X_{.1} = \sum_{i=1}^t x_{i1}$	$X_{.2} = \sum_{i=1}^t x_{i2}$	.	.	.	$X_{.t} = \sum_{i=1}^t x_{it}$	

如果表 13.13 計算出的 F—統計值超過了 F 在 (t-1) 及 (b-1)(t-1) 自由度的臨界值（見表 15.6），則拒絕虛無假設，並選擇替代假設。

如果在表 13.13 中的 F—統計是顯著的，則使用多重比較程序，以確定哪些樣品具有顯著不同的平均評分。Fisher LSD 在隨機（完全）區集的設計的應用為：

$$LSD = t_{\frac{\alpha}{2}, df_E} \sqrt{2MS_E/b} \quad (13.13)$$

其中 b 是研究中的區集（通常為品評人員）數量， $t_{\frac{\alpha}{2}, df_E}$ 和 MS_E 則如之前所定義。

13.5.3.2 隨機區集分析的排名數據

如果從隨機區集設計的數據是排名的形式，則使用弗里德曼式統計 (Friedman-type statistic) 進行非參數分析。數據整理在表 13.12 中，但和

評分數據不同的是，在表中的每一行包含了由每個品評員分配給該樣品的排名。在表 13.12 底部的欄小計為樣本的等級總和。

表 13.13 使用評級的隨機區集設計—ANOVA 表

變異度來源	自由度	平方和	均方	F
總計	bt-1	SS_T		
區集（品評員）	b-1	SS_j		
樣品	$Df_s = t - 1$	SS_s	$MS_s = SS_s/df_s$	MS_s / MS_E
誤差	df_E $= (b - 1)(t - 1)$	SS_E	$MS_E = \frac{SS_E}{df_E}$	

弗里德曼式統計的排名數據，使用 F—統計的評分分析，是

$$T = \left(\left[\frac{12}{bt(t+1)} \right] \sum_{j=1}^t x_{.j}^2 \right) - 3b(t+1), \quad (13.14)$$

其中 b 是品評人員的數目，t 是樣品的數量，並且 $x_{.j}$ 是樣品 j 的排名總和（即，在表 13.12 該列樣本 j 的合計）。在 $x_{.j}$ 中的「點」表示該過已經加總過，即 $x_{.j} = \sum_{i=1}^b x_{ij}$ 。

測試過程中，如果方程式 13.14 中的 T 值超過 $\chi_{\alpha, t-1}^2$ ，拒絕在 α 顯著水準下無樣品差異的虛無假設。要不然就接受。其中 $\chi_{\alpha, t-1}^2$ 是自由度 t-1 的 χ^2 —分佈的上 α 百分位數（見表 15.5）。此方法假定比較多的品評人員參加研究。12 名或更多的品評員的測試結果相當準確。

假如 χ^2 —統計量是顯著的，那麼接著執行多重比較方法以確定哪個樣品是顯著不同的。使用 Fisher 的 LSD 的進行隨機（完整）區集設計的排名加總的非參數模擬

$$LSD_{rank} = Z_{\alpha/2} \sqrt{bt(t-1)/6} = t_{\alpha/2, \infty} \sqrt{bt(t+1)/6} \quad (13.15)$$

假如兩個樣品的總和高於公式 13.15 的 LSD_{rank} 值，則這兩個樣品在 α 水準下是顯著不同的。

如果品評人員被允許分配不同的樣品相同的排名，那麼必須使用稍微複雜一點的檢驗統計量 T' （見 Hollander, Wolfe, 1973）。平均分配排名在一個無法區分的樣品上。例如，在一個四組樣品的測試中，如果中間的兩個樣品（通常為排名 2 和 3）不能被區分的話，那麼分配兩個樣品平均 2.5 的排名。替代公式 13.14 中的 T 值。

$$T' = \frac{12 \sum_{j=1}^t (x \cdot j - G/t)^2}{bt(t+1) - [\frac{1}{t-1} \sum_{i=1}^b (\sum_{j=1}^t t_{ij}^3) - t]}, \quad (13.16)$$

其中， $G = bt(t+1)/2$ ， g_i 是區集 i 中並列小組的數目； $t_{i,j}$ 是區集 i 中並列小組的第 j 個樣本。（無並列的樣品則各自算單獨的一組， $t_{i,j} = 1$ ）。

13.5.4 平衡不完全區集設計

平衡不完全區集設計（BIB）允許感官分析師從其品評成員中獲得一致且可靠的數據，即使當研究的樣品總數比起感官疲勞發生時的樣品數量還多時。在 BIB 設計中，小組成員只評估樣品總數的一部分（也就是說，每個小組成員在總數 t 個樣本中的 k 個， $k < t$ ）。每個品評小組成員評估 k 個樣本的特定組別，使得在 BIB 設計中的每個樣品被評估的次數相等（ γ ），以及所有樣品對一起進行評估的次數相同（ λ ）。 γ 和 λ 實際上在 BIB 設計中所有的樣品皆是一致的，以確保每個樣品平均值是以相等的精確度估計，以及兩個樣品之間的所有成對比較具有同樣的敏感度。完成單次重複的 BIB 設計所需的區集數為 b 。表 13.14 是一個典型的 BIB 格式。BIB 設計的列表，如 Cochran 和 Cox（1957）所提出的，在選擇研究具體的設計時非常有幫助。

為了獲得足夠大的數目的全部重複，整個 BIB 設計（ b 個區集）可能不得不重複幾次。基本設計的重複數目使用 p 表示。因此區集的總數為 pb ，得到每個樣品總共重複的次數 pr ，以及在全部 BIB 設計中每對樣品發生的總次數 $p\lambda$ 。從 9 點類別量度和非結構化的線性量度的經驗中顯示，複製的總次數（ pr ）至少應該為 18 次，以獲得足夠精確的樣本均值估計。這只是個普遍的規則，在確定研究需要多少品評人員時提供了一個起始點。樣品是否有足夠的差異使得在統計上具有顯著性，受多種因素的影響：產品、品評人員敏銳度、訓練水準等，只有經驗、測試和誤差可以回答任何給定的研究需要多少重複次數。

在感官研究中具有兩種常用的方法使用 BIB 設計。首先，如果區集的數量是相對小的（例如，四個或五個），可能可以使少數的品評人員（所有 p ）回來重複測試，直到每個小組成員完成了設計中的所有重複。（區集的呈現順序應該分別隨機地提供給每個小組成員，就像每個區集中的樣品應該隨機呈現）第二，對於 b 值大情形，通常的做法是調用大量的品評人員（所有 PB ），並且讓每個品評員在單個區集中評測樣品。特定品評人員收到的樣品區集應該隨機分配，而樣品在各個區集內的呈現順序應該在所有情況下都是隨機的。

13.5.4.1 評級的 BIB 分析

BIB 評級形式的數據使用 ANOVA 來分析（見表 13.15）。如在隨機（完全）區集設計的情況下，總變異度拆分在各個區塊的效果、樣本和誤差裡。然而，在 BIB 分析中用於計算平方和的公式

表 13.14 平衡不完全區集設計數據表（ $t = 7$ ， $k = 3$ ， $b = 7$ ， $r = 3$ ， $\lambda = 1$ ， $p = 1$ ）

樣品 區集	1	2	3	4	5	6	7	區集總 和
1	X	X		X				B_1
2		X	X		X			B_2
3			X	X		X		B_3
4				X	X		X	B_4
5	X				X	X		B_5
6		X				X	X	B_6
7	X		X				X	B_7
處理的 總和	R_1	R_2	R_3	R_4	R_5	R_6	R_7	G

註：X，個人觀察； B_i ，i 行觀察的總和； R_j ，j 欄觀察的總和；G，所有觀察的總和。

表 13.15 平衡不完全區集設計的 ANOVA 表

變異來源	自由度	平方和	均方	F
在所有 b 區集中每位品評員 p 的評估				
小計	$tpr-1$	SS_r		
品評員	$p-1$	SS_p		
區集（品評員中）	$p(b-1)$	$SS_{B(p)}$		
樣品（區集附屬）	$df_s = t - 1$	SS_s	$MS_s = SS_s/df_s$	MS_s/MS_E
誤差	df_E $= tpr - t - pb$ $+ 1$	SS_E	$MS_E = SS_E/df_E$	
單一區集中每位品評員 pb 的評估				
小計	$tpr-1$	SS_T		
區集	$Pb-1$	SS_B		
樣品（區集附屬）	$df_s = t - 1$	SS_s	$MS_s = SS_s/df_s$	MS_s/MS_E
誤差	df_E $= tpr - 1 - pb$ $+ 1$	SS_E	$MS_E = \frac{SS_E}{df_E}$	

比起任何隨機（完整）區集分析更加複雜。感官分析員應該確認其使用來進行分析的統計軟體是能夠處理 BIB 設計的。不然則必須特別開發執行 BIB 分析程序。用於分析 BIB 數據的 ANOVA 形式取決於設計如何執行。如果每個小組成員在基本設計中評估每個區集，那麼「品評人員效應」可以從總變異性分割出來（見表 13.15a）。如果每個品評人員只有評價一個樣品區集，那麼品評人員效應會與區集效果混合再一起（見表 13.15b）。在兩種情況下都有評估到品評人員效應的

影響，因此無論使用哪種方法，都不誇張地提供了實驗誤差的估計。

如果表 13.15 的 F-統計量超過相應自由度的 F-分佈的臨界值，則拒絕樣品之間的評分平均是相等的虛無假設。在 BIB 設計的 Fisher LSD 形式為：

$$LSD = t_{\alpha/2, df_E} \sqrt{2MS_E/pr} \sqrt{[k(t-1)]/[(k-1)t]} \quad (13.17)$$

其中 t 是樣品的總數，k 是在單次測試期間每個品評人員評測的樣本數，r 是在基本設計中（即，在 B 區集中的一次重複）每個樣品進行評價的次數，p 則是基本設計重複的數目。MS_E和 $t_{\alpha/2, df_E}$ 定義如前所述。

13.5.4.2 排名數據的 BIB 分析

在 BIB 設計所產生的排名數據弗里德曼式的統計應用排名。檢驗統計量的形式是其中 t、K、r、λ、p 如前面所定義，而 R_j 是第 j 個樣本的排名和（即，表 13.14 最後一行的樣本 j 的值）。方程式 13.18 中 T 的臨界值表格可使用 t = 3-6，k = 2-5 和 p = 1-7 的組合選擇。然而，在大多數感官研究中，區集的總數超過了表中的值。對於這些情形來說，如果在等式 13.18 的 t 值超過 (t-1) 自由度下的 X²-統計量的臨界值（見表 15.5），則測試程序是必須拒絕樣品為相等的假設。

假如 x²-統計值是顯著的，執行多重比較方法以進一步確認是哪些樣品顯著不同。使用對於 BIB 設計中排名數據的 Fisher 的 LSD 進行非參數模擬：

$$LSD_{rank} = Z_{\alpha/2} \sqrt{p(k+1)(rk-r+\lambda)/6} = t_{\alpha/2, \infty} \sqrt{p(k+1)(rk-r+\lambda)/6} \quad (13.19)$$

13.5.5 拉丁方塊設計

在隨機區集和平衡不完全區集設計中，變異性的單一來源（即品評員）在感官小組研究開展之前便是已知且獲得補償的。當品評小組進行評估前便已知道兩種變異性的兩個來源，則應該使用拉丁方塊設計。例如，我們普遍知道，品評人員可能因為在測試皆有變化（即，測試效應[session effect]），並且根據指示而改變他們評估樣品的方式（即，背景效應[context effect]）拉丁方塊設計可用於補償這兩個來源的變異性，並對於樣品之間的差異具有更高敏感度的比較。

樣品的數量必須足夠小，使得所有的樣品都可以在每次測試中被評估（典型上，t ≤ 5）。此外，每個品評人員必須能夠重複返回進行一系列的測試，使得研究中所有樣品都得以被評估（即，t 在拉丁方塊設計中代表樣品和測試的數目）。對於每個品評人員，每個樣品在每次測試中都提交一次。在整個 T-測試中，每個樣品在每個位置呈現一次。如圖 13.14 所示，每個樣品在各個呈現順序/測試組合相等的配置可以顯示在一個方陣中，因此稱為「拉丁方塊」。

呈現給每位品評員的順序採取獨立隨機原則。這可以透過先各自隨機分配給

各個品評員的樣品編碼，如圖 13.14，S1、S2、S3、S4 和 S5，接著再次為每位品評員每次測試中隨機選擇特定的順序（即，圖 13.14 的每一行）。

使用 ANOVA 來分析拉丁方塊設計數據（見表 13.16）。母體變異被劃分各個品評員、品評測試、測試順序、樣品及誤差的效果。如果表 13.16 的 F—統計值超過相對應自由度的 F 臨界值，那麼拒絕樣品之間的平均評分相等的虛無假設。對於拉丁方塊設計的 Fisher 的 LSD 形式為：

$$LSD = t_{\alpha/2, df_E} \sqrt{2MS_E/pt}$$

其中，p 是品評人員的數量。

	1	2	3	4	5
1	S ₁	S ₂	S ₃	S ₄	S ₅
2	S ₅	S ₁	S ₂	S ₃	S ₄
3	S ₄	S ₅	S ₁	S ₂	S ₃
4	S ₃	S ₄	S ₅	S ₁	S ₂
5	S ₂	S ₃	S ₄	S ₅	S ₁

圖 13.14 拉丁方塊在一次測試中五種樣品提供給一位品評員的呈現順序。研究中樣品代碼 S1、S2、...、S5 被隨機分配到 5 個樣品上。呈現順序和測試順序各自獨立隨機分配。

1 3.5.6 裂區設計

在隨機平衡不完全區集的設計，品評人員被視為是一種區塊因子；也就是說，假設品評人員是在研究進行前就已知存在可識別的變異性來源，因此應該在品評小組的設計中補償。在變異數分析中，環境因素（例如，品評員）和處理因子（例如，產品）的影響是被視為額外添加的。在實務上，這種假設意味著儘管品評人員可以使用感官評測量度的不同部位來表達自己的看法，樣品之間的差異大小和方向被認為對於所有品評人員都是以相同的方式報告。當然，研究實際收集到的數據與假設路徑由於實驗誤差的關係略微發散。說明這個假設的另一種方式是，在隨機區集或 BIB 設計中，區集和處理（例如，品評員和樣本）之間沒有任何「互動」。對於一批訓練有素、動機性強、「經過校準」的品評人員，品評員和樣品之間的相互作用是合理的。不過若是在訓練中，感官分析師可能懷疑這個假設的有效性。裂區設計即是用來確定品評員及樣品間的相互作用是否存在。

表 13.16 拉丁方塊設計的評級 ANOVA 表

變異來源	自由度	平方和	均方	F

小計	$pt^2 - 1$	SS_r		
品評員	$p-1$	SS_j		
測試（單個品評員內）	$p(t-1)$	SS_p		
順序（單個品評員內）	$p(t-1)$	SS_o		
樣品	$df_s = t - 1$	SS_s	$MS_E = SS_s/df_s$	MS_s/MS_E
誤差	df_E $= (pt - p - 1)(t - 1)$	SS_E	$MS_E = SS_E/df_E$	

在裂區設計，品評員被視為是樣品的第二種實驗處理。在至少兩組 ($p \leq 2$) 的 b 位品評人員中呈現 t 個樣品（呈現給每位品評員的順序皆單獨地隨機化）。 P 個品評小組是實驗設計中的「區塊」或「重複」。呈現給每個品評小組的樣品來自隨機選擇的批次或單獨準備。這是隨機化的裂區研究的第一層。接著品評人員收到特定集合的樣品（基於每個品評小組的到達時間而呈現的隨機順序）。由於隨機化的自然連續性，當第一個處理因子（樣本）在重複過程內隨機化，接著第二個處理因子（品評員）在第一個處理因子（即樣本）後再次隨機化，裂區設計是適當的。

13.5.6.1 評級的裂區分析

使用 ANOVA 的特殊形式來分析從裂區設計得到的數據（見表 13.17）。樣品效應被稱為「全區效應」（Whole-plot effect）。品評員之間，以及品評員和樣品之間的相互作用被稱為副區效應（subplot effect）。測試的全區效應和副區效應有各自的誤差數與（因為先前描述的隨機化方式的順序性）。

全區效應的誤差（表 13.17 的誤差 (A)）和品評人員及樣品之間的相互作用是以相同的方式計算，如果其存在的話。表 13.17 F_1 的統計是用來測試樣品效應的顯著性。如果 F_1 的值比起自由度 $(p-1)(t-1)$ 的 F -分佈的上 α -臨界值大的話，則可以得出結論：樣品的平均值之間有顯著差異。

表 13.17 中的 F_2 和 F_3 統計值用於測試副區效應、品評人員之間，以及品評人員和樣品的相互作用的顯著性。 F_2 和 F_3 的分母是副區誤差項 $MS_{E(B)}$ 。如果 F_3 超過自由度 $(b-1)(t-1)$ 以及 $t(p-1)(b-1)$ 的 F 分佈的上 α -臨界值，則品評人員和樣品的相互作用的存在是顯著的。相互作用的意義表示品評人員以他們各自不同的方式表達對樣品之間差異的理解。品評人員與樣品的相互作用來自於訓練不足、對於屬性定義的混淆，或不熟悉評級技術。

表 13.17 使用評級的裂區設計—ANOVA 表

變異來源	自由度	平方和	均方	F
小計	$pbt - 1$	SS_r		

品評員	p-1	SS_p		
樣品	t-1	SS_s	MS_E $= SS_s/(t-1)$	F_1 $= MS_s/MS_{E(A)}$
誤差 (A)	$df_{E(A)}$ $= (p-1)(t-1)$	$SS_{E(A)}$	$MS_{E(A)}$ $= SS_{E(A)}/df_{E(A)}$	
品評員之間	b-1	SS_j	MS_j $= SS_j/(b-1)$	F_2 $= MS_s/MS_{E(B)}$
品評員和樣品之間	df_{js} $= (b-1)(t-1)$	SS_{js}	MS_{js} $= SS_{js}/df_{js}$	F_3 $= MS_{js}/MS_{E(B)}$
誤差 (B)	$df_{E(B)}$ $= t(p-1)(b-1)$	$SS_{E(B)}$	$MS_{E(B)}$ $= SS_{E(B)}/df_{E(B)}$	

當品評員與樣品的相互作用是顯著的存在時，審查樣品的母體效果是沒有意義的（表 13.17 中的 F_1 測試），因為相互作用的存在表示樣品的差異方式取決於品評人員。應該檢視個別品評員的平均評級表，以及品評員和樣品的相互作用標繪圖，以確定哪些品評員會使得相互作用是顯著的（參見第 9 章）。

假如 F_3 不是顯著的，但 F_2 是的話，那麼整體品評員效果依然存在。一個顯著的品評員效果表示有品評員使用不同部位的評級量度來表達他們的看法。這並不比起品評員和樣品之間的相互作用的顯著來得嚴重。但是，根據品評員之間差異的大小，可以指示品評小組需要使用參考量度以重新校準。

如果 F_3 並不顯著，但 F_1 是，那麼母體樣品效果是存在。為了確定哪些樣品是顯著不同，使用針對裂區設計的 Fisher 的 LSD：

$$LSD = t_{\alpha/2, df_{E(A)}} \sqrt{2MS_{E(A)}/pb} \quad (13.21)$$

其中 p 是獨立重複品評小組的次數，而 $df_{E(A)}$ 和 $MS_{E(A)}$ 分別是自由度的數量以及誤差 (A) 的均方。

13.5.7 同時的多重比較方法

到目前為止，我們只使用了 Fisher 的 LSD 多重比較方法，以確定哪些樣品在設計的感官小組研究中顯著不同。事實上，還有，兩種多重比較方法。第一種包括 Fisher 的 LSD，控制配對類的誤差率；即每一次的平均比較或排名總和所應用的 I 型誤差 (α 大小)。控制配對誤差率的方法被稱為單一性 (One-at-a-time) 多重比較方法。第二種是控制實驗類的誤差率；也就是說，同時在所有平均配對和排名總和上的 I 型誤差。控制實驗類的誤差率方式被稱作是同時多重比較方法。

Tukey 的誠實顯著差異 (Honestly Significant Difference, HSD) 是一種同時多重比較方法。Tukey 的 HSD 可以不管樣品間的差異以及整體測試的結果而使用。杜克的 HSD 在相同樣品大小的評及數據的一般形式是：

$$HSD = q_{\alpha, t, df_E} \sqrt{MS_E/n} \quad (13.22)$$

其中， q_{α, t, df_E} 是用於比較 t 樣品的均值，自由度 df_E 的學生化範圍分佈 (Studentized range distribution) 的上 α - 臨界值 (見表 15.4)。如同 LSD， df_E 和 MS_E 分別是來自於 ANOVA 的自由度和誤差均方 (裂區設計 ANOVA 的誤差(A))； n 是所有被比較平均的樣品共同大小。對於隨機 (完成) 區集設計， $n = b$ ；對於裂區設計， $n = p$ 。Tukey 的 HSD 在 BIB 設計的形式為：

$$HSD = q_{\alpha, t, df_E} \sqrt{MS_E/pr} \sqrt{k(t-1)/(k-1)t} \quad (13.23)$$

Tukey 的 HSD 在排名總和的非參數模擬為：

$$HSD_{\text{排名}} = q_{\alpha, t, \infty} \sqrt{bt(t+1)/12} \quad (13.24)$$

對於隨機 (完全) 區集設計，以及 BIB 設計

$$HSD_{\text{排名}} = q_{\alpha, t, \infty} \sqrt{p(k+1)(rk-r+\lambda)/12} \quad (13.25)$$

13.6 附錄機率

本節的目的是介紹一些常用的計算機率的技巧，基於常見的機率分佈。這些技術是在第 13.2 節和第 13.3 節，以及在第 14 章中討論的更進階的主題，的統計估計和推論的基礎。

13.6.1 常態分佈

常態分佈是機率統計中最常用的分佈。常態分佈函數的形式是：

$$f(x) = \left[\frac{1}{\sqrt{2\pi}\sigma} \right] \exp\left(-\left[x - \frac{\mu}{\sigma}\right]^2\right),$$

其中， $\exp$ 是以 e 為底的指數函數。常態分佈的參數是均值 μ ($-\infty < \mu < \infty$) 和標準差 σ ($\sigma > 0$)。標準分佈是對稱於 μ ，即 $F(x - \mu) = F(\mu - x)$ 。均值 μ 是測量分佈的中心位置。標準差 σ 是測量常態分佈基於 μ 的分散程度。 σ 小的值，分佈圖狹窄而尖； σ 值較大的分佈圖則較寬扁 (見圖 13.15)。如同所

有的連續機率圖，不管參數的值是多少，分佈曲線下的總面積皆等於1，。

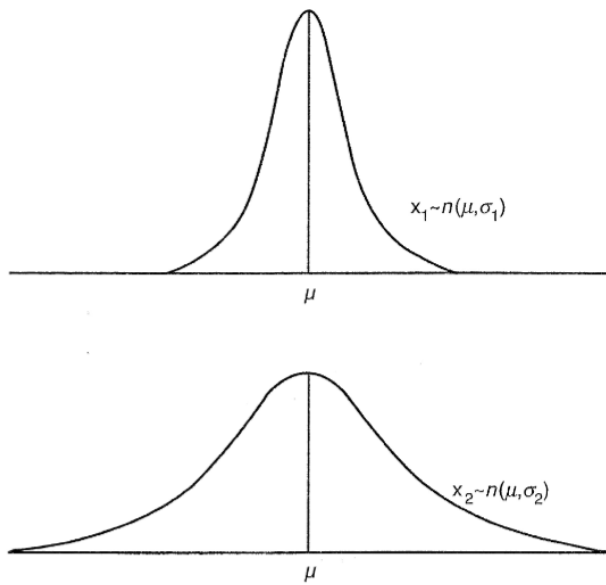


圖 13.15

兩個均值相等的常態分佈，但是 $\sigma_1 < \sigma_2$

令 x 是具有均值 μ 和標準偏差 σ 的常態分佈中的隨機變數(通常縮寫為 $x \sim n(\mu, \sigma)$)。定義變數 z 為：

$$z = \frac{(x - \mu)}{\sigma} \quad (13.26)$$

隨機變數 z 也具有常態分佈。 z 的平均值是零，其標準偏差是1[即 $Z \sim n(0, 1)$]。 z 被稱為是標準常態分佈，或經常 z 被稱為標準常態偏離。給定常態隨機變數 x 的 μ 值和 σ 值，以及標準常態分佈的機率表(見表 15.2)，它能夠計算各種感興趣的機率。

範例 13.7：在一個區間內計算常態機率

計算均值 $\mu = 50$ ，標準差 $\sigma = 5$ 常態隨機變數 x ，處於 50 和 60 之間的值(也就是說， $\Pr[50 < X < 60]$)。解決該問題的第一步是使用(等式 13.26)「標準化」 X ：

$$\begin{aligned} \Pr[50 < X < 60] &= \Pr[(50 - 50) / 5 < (X - 50) / 5 < (60 - 50) / 5] \\ &= \Pr[0 < z < 2]。 \end{aligned}$$

表 15.2 給出了標準常態偏差基於零的值(即，平均值)到某種特定數目的

機率。因此，對照表 15.2 中對應於 2.0 的行，以及對應 0.00 的列，分析師發現尋求的機率等於 0.4772（見圖 13.16）。接著，思考找尋 $PR [45 < X < 50]$ 的問題在哪裡，如同先前的操作步驟， $X \sim n(50, 5)$ 。

標準化：

$$\begin{aligned} Pr [45 < X < 50] &= Pr [(45 - 50) / 5 < (X - 50) / 5 \\ &< (50 - 50) / 5] \\ &= PR [-1 < Z < 0]。 \end{aligned}$$

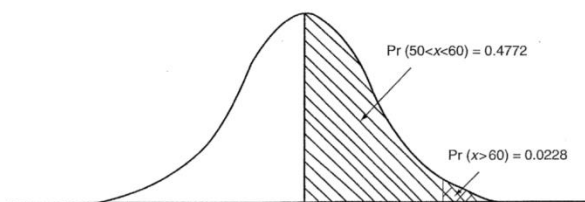


圖 13.16 在一常態分佈尾部的區間中計算機率的描繪圖。

由於標準常態分佈是在其平均值，零，上左右對稱的，對於任何常數 c ，它遵循 $PR [-c < Z < 0] = PR[0 < Z < c]$ 。因此，通過表 15.2：

$$PR [-1 < Z < 0] = PR[0 < Z < 1] = 0.3413 \quad (13.27)$$

因此， $PR [45 < X < 50] = 0.3413$ 。

最後，考慮在相同的隨機變數 $x \sim n(50, 5)$ 的機率 $PR [45 < X < 60]$ 。解決這個問題的方式如下：

$$\begin{aligned} Pr [45 < X < 60] &= Pr [-1 < Z < 2] \\ &\text{（透過[13.26]的標準化步驟）} \\ &= Pr (-1 < Z < 0] + Pr[0 < Z < 2] \\ &= Pr[0 < Z < 1] + Pr[0 < Z < 2] \quad \text{（按[13.27]）} \\ &= 0.3413 + 0.4772 \\ &= 0.8185。 \end{aligned}$$

範例 13.8：計算常態尾部機率

尾部機率與分佈極值的機率曲線下面積相關聯（見圖 13.17）。也就是說，尾部機率對於一些常數 c ，可表示為 $Pr [x > c]$ 或 $Pr [x < c]$ 。尾部機率廣泛應用於檢驗統計假設。

考慮尋找 $Pr[X > 60]$ 的問題，其中 $x \sim n(50, 5)$ 。注意任何機率曲線下的總面積（即，機率）都是一，因此從常態分佈的對稱性得出 $Pr[X < \mu] = Pr[X > \mu] = 0.50$ 。

因此，

$$\begin{aligned}\Pr[X > 60] &= \Pr[(X - 50) / 5 > (60 - 50) / 5] \\&= \Pr[z > 2] && \text{(按[13.26])} \\&= 0.50 - \Pr[0 < Z < 2] \\&= 0.50 - 0.4772 && \text{(來自範例 13.7)} \\&= 0.0228\end{aligned}$$

(參見圖 13.16 的交叉陰影線區域以理解第三步驟。)

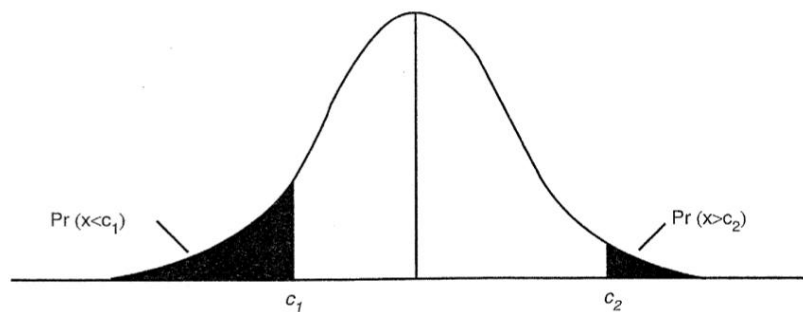


圖 13.17 常態分佈的尾部機率。

13.6.2 二項式分佈

二項式分佈函數為

$$\Pr[x = k] = b[k] = \binom{n}{k} p^k (1 - p)^{n-k} \quad (13.28)$$

對於 $k = 0, 1, 2, \dots, n$; $n > 0$ 且為整數；且 $0 \leq p \leq 1$ 。

二項式分佈的參數是： n =測試次數； P =任何測試中「成功」的機率。每次測試中成功的定義是任意選擇的。例如，在一個雙樣本偏好測試中（A 與 B 的比較），偏好 A 可以定義為成功，或是偏好 B 也可以定義為成功。不管如何，在 $i=1, 2, \dots, n$ 中，如果第 i 個試驗的結果是成功的， $K_i = 1$ ；否則， $K_i = 0$ 。在公式 13.28， $k = \sum_{i=1}^n k_i$ 是 n 次測試成功的總數。精確的二項式機率可以用電子表格功能計算，如 Excel 的 BINOMDIST。近似二項式的機率可以使用二項式的正常逼近法計算。

範例 13.9：計算精確的二項式機率

假設 $n = 16$ 的品評員參加五取二的差異測試。從五個中正確選擇兩個奇異樣品的成功機率遵循成功機率 $p = 0.10$ （當樣品中沒有可感知的差異）的二項式分布。要找出剛好兩個品評人員（ $K = 2$ ）做出正確選擇的機率（即，在 16 次測

試中正好 2 次成功)，請在 Excel 電子表格中的其中一格欄位輸入 = BINOMDIST (2, 16, 0.10, FALSE)。顯示的回應將是 0.2745，這就是所求的機率。
要查找兩位到六位（含）品評員做出正確選擇的機率，須知道：

$$\Pr[2 \leq x \leq 6] = \Pr[x \leq 6] - \Pr[x \leq 1] \quad (13.29)$$

在 Excel 中計算，透過在電子表格中輸入以下：

1. 在 A1 格中輸入：= BINOMDIST (6, 16, 0.10, TRUE)
2. 在 A2 格中輸入：= BINOMDIST (1, 16, 0.10, TRUE)
3. 在 A3 格中輸入：=A1-A2

所求的機率值，0.4848，顯示在 A3 格中。

使用電子表格計算尾部機率有兩種方法，這取決於您是否要計算分佈下尾部或上尾部。計算分佈下尾部的機率（例如，少於三位品評員作出正確的選擇的機率），使用以下技巧：

$$\Pr[x < 3] = \Pr[X \leq 2] \quad (13.30)$$

因此，在 Excel 電子表格中的其中一格輸入：= BINOMDIST (2, 16, 0.10, TRUE)，所求的機率，0.7892，就會顯示出來。

另一方面，計算至少三位（即，三位或更多位）品評員作出正確的選擇，

$$\Pr[x \geq 3] = 1 - \Pr[x < 3] = 1 - \Pr[x < 2] \quad (13.31)$$

因此，在 Excel 電子表格中輸入：= 1- BINOMDIST (2, 16, 0.10, TRUE)，將顯示出所得機率：0.21081。

範例 13.10：常態逼近二項式

當手邊沒有電子式表格計算二項式機率函數時，也可以使用常態分佈來逼近二項式機率。使用第 13.5.1 節的方法，我們需要知道 μ 和 σ 的值。對於成功的數目：

$$\begin{aligned} \mu &= np \\ \sigma &= \sqrt{np(1-P)} \end{aligned} \quad (13.32)$$

設定 $n = 36$ 和 $p = 1/3$ ，並考慮計算至少 16 個成功的機率問題。使用公式 13.32 計算：

$$\begin{aligned} \mu &= (36)(1/3) = 12 \\ \sigma &= \sqrt{(36)(1/3)(1 - 1/3)} = \sqrt{8} = 2.828 \end{aligned}$$

因此，使用例 13.8 的方法：

$$\begin{aligned}\Pr[X \geq 16] &= \Pr[(X - 12)/2.828 \geq (16 - 12)/2.828] \\ &= \Pr[Z \geq 1.41] \\ &= 0.5 - \Pr[0 < z < 1.41] \\ &= 0.50 - 0.4207 \\ &= 0.0793.\end{aligned}$$

我們也可以使用常態分佈逼近二項式來計算成功比例的相關機率。在這個例子中：

$$\begin{aligned}\mu &= p \\ \sigma &= \sqrt{p(1-p)/n} \quad (13.33)\end{aligned}$$

在大多數的感官評估中，測試的數目應該要足夠大，使得常態分佈的逼近給出充分精確的近似值。一個常見的普遍規則是如果 $np > 5$ 且 $n(1-p) > 5$ ，常態逼近二項式可以是足夠精準的。也就是說，若希望常態逼近二項式是合理的，樣本大小 n 應該要足夠大，使得我們能夠觀察到至少五個成功和至少五個失敗的結果。

第十四章 進階統計方法

14 Advanced Statistical Methods

14.1 簡介

在第十三章所提供的基本統計技術都是在感官測試中經常需要使用來分析結果的方法。然而，當研究的目標超出簡單的估計或區別的範圍時，可能需要應用更複雜的統計方法。本章介紹了一些比較常見的進階技術。計算方法的複雜度使用手動計算是不切實際的。因此我們假設讀者能夠使用電腦等資源以進行必要的計算。

感官研究很少只包括單一的反應變數。更多的時候，在每個樣品上測量許多變數經常是研究的目標之一，以確定「多元」測量是如何彼此地相關以及如何地不同。14.2 節介紹研究多變數數據關係的方法。首先討論的是相關性分析

(Correlation analysis)、主成分分析法(Principal component analysis, PCA)、群聚分析(Cluster analysis)。這些技術被用於研究多元的數據組，其中所有的變數都是平等的狀況。接著呈現的是迴歸分析(Regression analysis)、主成分迴歸(Principal component regression)、部分最小平方法(Partial least-squares)以及區隔分析(Discriminant analysis)。這些方法適用於當

數據組中的變數可以被分類為獨立或非獨立時，以及分析目標是使用非獨立變數以預測獨立變數的值。14.3 節則介紹多種偏好定位法 (preference mapping)，其中 14.2 節介紹的所有方法應用在連結測試產品的感官特性以及消費者接受度。在 14.4 節中，系統性地研究個人以及多個實驗變數結合作用的實驗計畫。其中討論了包括因子實驗 (Factorial experiments)、部分因子 (Fractional factorial) (或稱「篩選研究」[screening studies]) 和反應曲面法 (Response surface methodology) (或「產品優化研究」)。

14.2 數據關係

感官評估中經常需要確定不同的變數之間存在的關係。描述屬性以何種方式和程度一起增加或減少、消費者對各種產品的相似喜好模式，以及基於年齡預測產品的感知屬性的值的能力，是這種類型問題的三種例子。

繪製變數之間關係的數種統計方法可以分為兩組。第一組方法是處理的數據組中的所有變數都是獨立的，意思是它們都被視為是同等重要的，且其中沒有任何變數是被其他變數所驅策或影響 (例如：一組描述性的風味屬性)。第二組方法是數據組中同時含有獨立及非獨立變數。這些數據組的其中一個或多個變數是相對於其他變數來得特殊且重要 (例如：總體喜好 vs. 描述性屬性評級)。由於這兩組方法都在同一時間內處理多個變數，因此它們是隸屬於「多元變數 (Multivariate)」統計方法的範疇。

14.2.1 所有獨立變數

當所有的變數都被看作是同樣重要的，統計分析的目的就變得是確定變數之間的關係的性質和程度，以確定是否存在相關變數的組別，或者確定觀測的不同組別是否存在。

14.2.1.1 相關性分析

在多元技術中，相關性分析是最簡單的。它用於測量兩個變數之間的線性關係強度。例如：屬性 X 和 Y 之間的關係強度，總結成相關係數 r：

$$r = \frac{\sum (y_i - \bar{y})(x_i - \bar{x})}{\sqrt{\sum (y_i - \bar{y})^2 \sum (x_i - \bar{x})^2}}$$

r 的價值介於 -1 及 +1 之間。 -1 表示完全地相反線性關係 (即：一個變數減少，則另一個變數增加)，而值為 +1 指示完全直接的線性關係 (即：這兩個變數同時增加或同時減少)。若 r 值接近零，則意味著兩個變數之間存在小的線性關係。很強的相關性並不意味著因果關係，也就是說，不能自動假設其中一個變數「驅使」另一個變數，而是兩個變數在一定程度上共同變化。

相關係數是綜合性指標。分析師應該永遠先檢查配對變數的散佈圖

(Scatterplots) 以決定 r 值是否合適代表其相對關係。強烈的線性趨勢並且觀察點之間有著相對均勻的間隔，如圖 14.1a，適合使用相關係數總結，同樣適合使用的還有跨越兩個變數範圍的觀察點呈現無趨勢圖形的擴散，如圖 14.1b。有些關係可能具有高 r 值，但顯然可以更好地使用非線性術語做總結，如圖 14.1c。在其他情況下，具有非常低的 r 值是顯而易見的，如在圖 14.1d 和圖 14.1e。相反地，相關係數有時在不同的觀測值（沒有內部相關性）仍可能誤導性地大，如在圖 14.1f'。散佈圖矩陣（見圖 14.2）允許所有的數據成對地顯示在一個密集的格式中。

相關分析可以用來鑑定一組反應是否以類似的方式變化，或是以相反的方式。此外，相關性分析可以用來確定來自不同來源（例如：消費者評級和一個經過訓練的品評小組針對描述性屬性的評級，以及儀器測量等等）所產生的數據之間的關係的強度。第 12 章包含這些關係類型的一些例子。

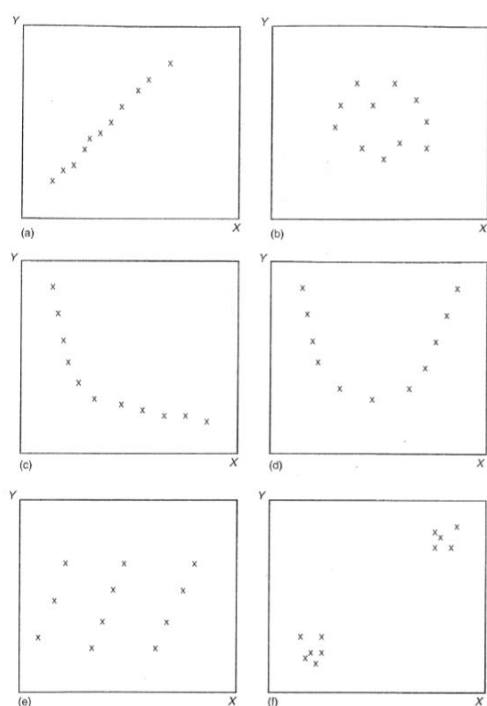


圖 14.1 x 和 y 兩個變數的散佈圖，顯示樣品相關係數 r ，有時是一個很適合用於總結的量度[即，圖 (a) 和 (b)]，有時卻不是[即，圖 (c) 通過 (f)]。

相關性的模式可以跨越許多不同的產品類別、區域，或者市場區隔。應該注意的是使用於進行相關分析的數據應該來自單一總體，而不是異質性的混合。

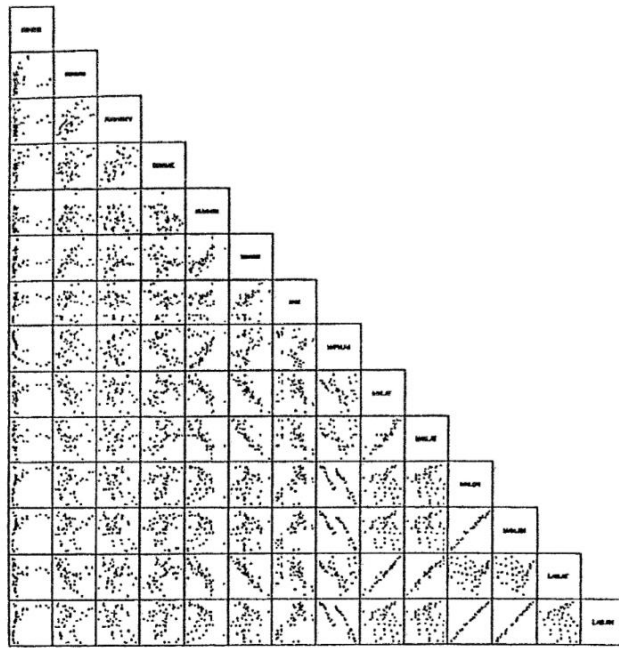


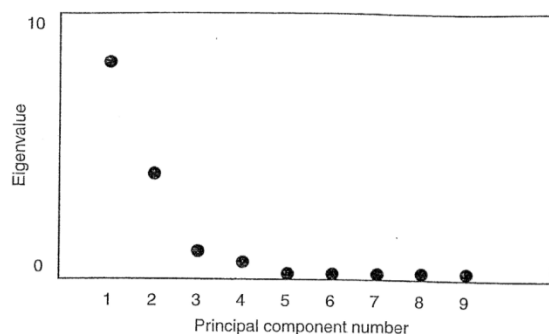
圖 14.2 用於識別多組回應中相關或不相關的變數組的散佈圖矩陣。

14.2.1.2 主成分分析 (Principal Components Analysis ,PCA)

初始的相關性分析可以確定幾組高度相關的變數（以及不是高度相關的變數）。這表示各組中包含相關資訊的變數，以及少數未觀測（或「潛在」）的變數可能提供的總體變異度。PCA 是用於識別所有甚至最小的潛在變數，稱為「主成分」，的統計技術，以解釋觀察最大的變異性。通常可 25-30 個變數與 2-3 個主成分的數據組可以解釋高達 75-90% 的總變異。

從一組多變數數據中提取主成分的電腦程序是被廣泛使用的，而以下討論並未包括此程序的理論和計算細節。

PCA 分析一組具有多個變數的觀測值的相關性結構，並識別此數據中最大變異性的主軸。此主軸被稱為第一主要元素 (First principal component)。第二主要元素 (Second principal component) 是組成同主體剩餘最大量變異性的軸線，軸線彼此必須是垂直（成直角）的（即：正交或不相關的）。選擇每個附加的主成分並且與所有其他的主成分正交，使得每個連續的主成分能夠針對其餘不明原因的變異盡可能的解釋。主成分的數目不能超過觀察到的變數數目，在實務中常常要少得多。提取主分量的過程結束於當預先確定的總變異量已經能夠解釋時（該數量始終包含在 PCA 的電腦輸出中），或者在當提取另外一個主成分仍只平凡地解釋在已知變數時。這種情況描繪在圖 14.3 的平坦陡坡圖 (Scree plot) 中。



X: 主要元素數量

Y: 特徵值

圖 14.3 使用主要元素說明變異性的陡坡圖，透過確定多少主要組分應該被納入研究。只有當陡坡圖變平之前提取的主成分需要被保留。

通過各個主成分， y_i 定義軸的方向，主成分是觀察到的變數， x_j 的線性組合，如下：

$$y_i = a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{ip}x_p \quad (14.1)$$

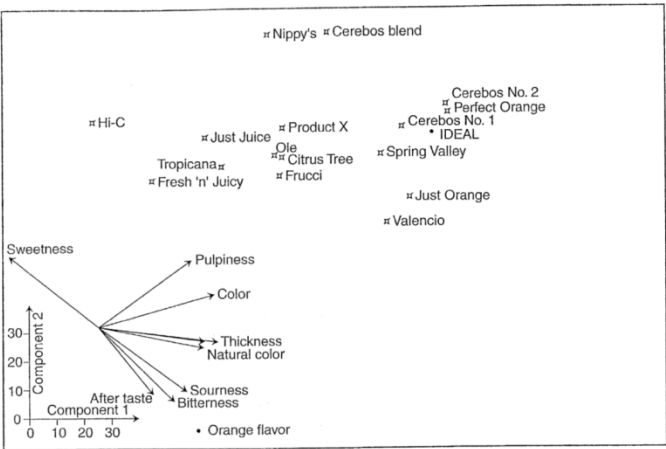
係數， a_{ij} ，被稱為權重 (Weight) 或負荷因素 (loading factors)。它們用於測量各個主成分上的原始變數的重要性。如同相關係數， a_{ij} 的值界於 -1 及 +1 之間。若值接近 -1 或 +1，代表對應的變數對主成分的值有很大的影響；值若接近 0 則表示對應變數對主成分上的影響不大。通常情況下，高度相關的觀測變數組將被隔離到與特定主成分相關的非重疊組別中。

負荷因素的檢視揭露了觀測變數如何組合在一起，並可能有意義的提出各個主成分總結的變異類型的解釋。為了幫助進一步詮釋，有時會旋轉主成分軸以增加其對原始變數軸的排列。旋轉後的第一主要元素將不再是最大變異性的方向，其次是第二元素，而獲得的優勢則是藉由這樣的旋轉效果而得到少數可闡明的潛在變數。

除了描繪原始變數之間的相關性，PCA 也可以用來顯示樣品的相對「位置」。一組產品的主要成分得分圖可以揭露此樣品的分組及極化，此現象在大量的原始變數的檢查時並非如此的顯而易見。Cooper, Earle 和 Triggs 使用 PCA 在二維圖形中描繪 16 種橙汁產品最初的十項屬性彼此的關係 (請參閱圖 14.4)。在其分析中，前兩個主要成了解釋了 79% 的原始變異性。Piggott 和 Sharman 提出了另外的例子。Powers 則呈現了在描述性分析使用 PCA 的多種參考。

PCA 提供了在低維度圖形上總結大量的收集數據的方法。我們很容易對於是否有必要繼續評估所有的原始變數，而不是只要研究少數幾個「代表」變數即可的過程感到疑問。研究原始變數的數量不應該基於 PCA 的結果而降低。如等式 14.1，每個原始變數包含在每個主成分的計算中。僅保留一小群感官的代表變數

並忽略原始變數效應的多元性質，將無法在未來進行主要成分穩定性的驗證，並可能導致今後評估的誤導結果。



圖譯：

Nippy's Cerebos blend：Nippy' s 綜和飲品

Component 1 第一主要元素

Component 2 第二主要元素

Sweetness 甜度

Pulpiness 果肉

Color 顏色

Thickness 濃度

Natural color 自然顏色

Sourness 酸

Bitterness 苦

剩餘皆橙汁品牌名

圖 14.4 橙色飲料產品的 PCA 結果，顯示產品彼此的關係以及其對應的原始描述性屬性。產品圖和屬性圖的區隔幫助視覺呈現，同時維持相對方向及幅度。例如，Hi-C 甜度較高但苦味較低，而 Cerebos 2 較不甜，且果肉和顏色的分數較高。

14.2.1.3 群集分析

本著同樣的精神，PCA 根據它們的相關行為識別屬性組別，多元統計法中的群集分析根據他們的評級相識性識別觀察值的組別。評級可以是在單一的樣品上收集的多元屬性，或是在不同樣品上收集的單個屬性。目前有大量的群集分析算式普遍使用中；因此，群集分析的計算細節無法在這裡公平的介紹和討論。Godwin, Barmann 和 Power 提出了群集分析的一個有趣的應用，其中感官屬性和儀器測量是根據它們之間的關係而分組，並且收集喜好性回應。

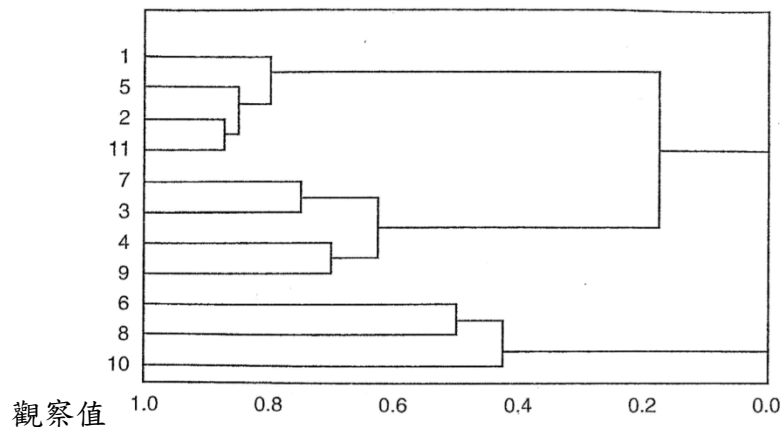


圖 14.5 群集分析的樹狀圖，呈現觀察值組合以及分離的群集之間的程度。

雖然統計上不是完全正確（屬性並非從一些現存群體隨機取樣觀察），他們的方法仍然是一個研究數據關係的有趣技巧，不應該被忽視。

有兩種類型的群集分析的算法：階層和非階層方法。它們之間的實際區別是觀察值由階層法分配給一個群集之後，它不能被移動到另一個群集。而非階層法的觀察值則能夠從一個群集移動到另一個。

階層方法往兩個方向中的其中一個方向進行。在比較常見的方法，每個觀察值最初被分別認為是大小為一的群集，接著分析合併其他觀察值（或者中間群集的觀察值），直到只剩一個群集存在。另外，分析也可能將所有的觀察值認定為一個大的群集，然後開始進行拆分觀察組，直到只剩一個觀察值存在。連續合併或拆分以圖形描繪成樹狀圖（或樹形圖）（參見圖 14.5）。樹狀圖中的觀察值階層結構，測量了群集標準的變化程度，並用於決定多少群集真正存在。

群集分析算法的一般區別是測量其中兩個群集之間的距離（或連結）的方式。常用的算式包括平均連結（average linkage）、重心連結（centroid linkage）、中位數連結（median linkage）、最遠相鄰（或全部）連結（furthest neighbor linkage）、最近相鄰（或單一）連結（nearest neighbor linkage），以及 Ward 的最小變異連結（Ward's minimum variance linkage）（見 SAS 1989）。Ward 的方法使用 ANOVA 平方和當作測量「距離」。每種方法都有其優缺點。目前並沒有普遍偏好使用的方法。

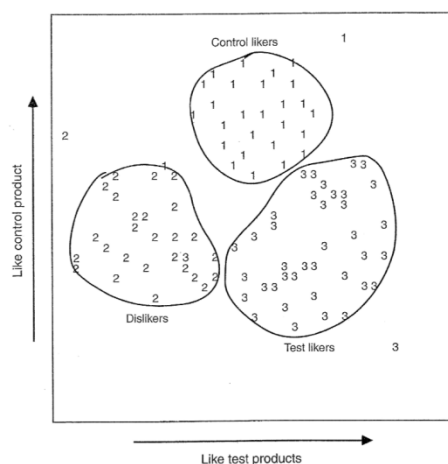
非階層方法包括 k-均值法（K-means method）和模糊目標函數（Fuzzy objective function 或 FCV）方法（Bezdek 1981）。在兩者中都使用了反覆的數學技巧。兩種方法中，使用者皆必須指示存在的群集數目。在 K-均值法中，每個觀測值都基於從群集中心到此觀察值的（歐幾里得）距離分配群集。中心隨著越來越多的觀測被添加到群集中而移動；因此，觀察值的分配必須重複，直到不再發生進一步的變化。FCV 則使用「成員度」（degree of membership）替換「群

集成員」(cluster membership) 的概念。這個方法分配每個預定的群集數目的每一個觀察值「成員權重」(Membership weight)，界於 0 和 1 之間。

此方法不重新分配觀測值到不同的群集中，而是持續調整成員的權重，直到收斂條件得到滿足（例如：群集中心的位置最小偏移）。

FCV 適合區分強烈連結到特定群集（即會員權重接近+1.0）的觀察值上，以及連結多個群集（即，連結到兩個或多個群集以上的會員權重幾乎相等）的觀察值。此外，Jacobsen 和 Gunderson 進行一些關於確定區別原始變數的重要性討論，其中有使用 FCV 群集的例子，例如基於氣相色譜法(gas chromatography)數據的挪威啤酒。

群集分析在感官接受度測試中特別重要的一個應用是識別具有跨產品性的不同接受度模式的受訪者群體。當某些風味強度增加而使得部分受訪者接受度提高，另一群可能會導致反感。合併這樣差異的組別可能誤導產品的接受度，因為，以統計術語來說，即沒有識別出群集而導致計算多峰數據的平均值。這樣的一組觀察值的中心可能無法代表任何真正的受訪小組，因此，是一個不適當的總體喜好結論（見圖 14.6）。



Like control product 喜好控制產品

Like test product 喜好測試產品

Dislikers 不喜歡者

Control likers 控制組喜歡者

Test likers 測試組喜歡者

圖 14.6 三組受訪者群集的圖，透過優格製品的整體喜愛的模式區分。該圖顯示了三組的總平均將會得到此產品的普遍喜歡的不適當結論。

使用群集分析鑑別喜好模式可能揭露不同的受訪者群體，跨越常用人口統計的界線。因此，群集分析在「區隔」(Segmentation) 上比起這個經典的方法來得有優勢。

群集分析的另一個重要的應用常出現在外部偏好地圖研究(見第 14.3.2 節)。

研究中包含的產品數量可能超過可供消費者測試的資源。如果描述檔案可用於所有潛在的產品，產品的感知地圖可以使用 PCA 建構。產品的因子得分 (Factor scores) 可以提交到群集分析，以進一步確定類似產品的集合。舉例來說，假如 18 個產品中只有 10 個可供消費者測試，10 個群集的方案將被選中。每個群集只選擇一個產品代表其群集參加消費者測試。以這種方式使用群集分析是為了確保保留的 10 個參加消費者測試的產品在最大程度上呈現了所有 18 種產品的感官差異的範圍。

在受訪者的群集被確定後，可以開始進行相關性分析、主成分分析，和/或迴歸分析（在下一節討論），以確定群集之間的相似性和差異，以及感知屬性是如何與喜好性相關聯。多個產品、品種，利基市場 (Niche market)，或是線性延伸皆可能被指認出。最後，可以進行群集內的人口統計總結，以確定群集內是否有成員符合營銷目的的市場目標。

14.2.2 依賴變數和獨立變數 (Dependent and independent variables)

在這組方法中，某些變數的值被看作是依賴於該組中的其他（獨立的）變數的值。對於這樣的數據，統計方法使用獨立變數來預測一個連續的依賴變數的值，或者它們使用獨立變數將觀測值分組到特定類別的離散變數。即使當依賴變數和獨立變數皆存在，研究者應首先應用上一節中所描述的方法來發現所有變數之間的基本關係（使用相關性分析和主成分分析 [PCA]）以確定是否所有的觀測值可以作為單組分析，或者存在顯示明顯不同模式的群集（通過群集分析）。這些初步分析有助於確保獲取的數據包含有意義的資訊和完整的摘要。

14.2.2.1 迴歸分析

基於一個或多個變數值預測另一個變數的值已變得非常常見。例如，使用描述性數據或者配方和處理過程的價值預測消費者的接受度，以及使用實驗結果預測描述性數據。各種反應的感覺強度基於刺激的強度預測（或濃度），使用心理模式（例如，Stevens 的定律）或動力學模型（例如，Michaelis-Menten/Beidler 方程式）這些例子皆使用迴歸分析將一個連續依賴變數的值關連到一個或多個獨立變數的值。

迴歸可以簡單地用於預測反應的值，或者不那麼簡單的，確定一個變數是如何變化且影響另一個變數的值。迴歸分析本身沒有產生因果關係。如果研究者準備了系統動態的相關假設而來，則迴歸分析可以用於測試假設關係的有效性。然而，在一般情況下，

由迴歸分析得到的只是一個極度精準的預測模型。極度精準的模型並不意味著獨立變數驅動著依賴變數。研究人員必須提供數據背後的含義，而含義是無法透過研究數據的數值分析步驟得到的。

將數據製圖對於得到一個成功的迴歸分析至關重要。我們在相關性分析中討論過同樣的原因，研究人員很容易盲目地運用電腦程式而得到誤導的結果，如果沒有先檢查依賴變數與獨立變數的區域（見圖 14.1）。以下小節介紹能夠確定迴歸模

型品質的一些有用的圖。

14.2.2.1.1 簡單線性迴歸

在簡單的線性迴歸中，一個單一的依賴變數的值， y ，是用一個單一的獨立變數的值， x ，使用以下形式的線性模型預測：

$$y = \beta_0 + \beta_1 x + \varepsilon \quad (14.2)$$

其中， β_0 和 β_1 是迴歸方程式的係數，將在分析中估計，而 ε 是 y 的觀測值和其預測值之間不明原因的偏差參數，稱為餘值 (Residual)。

度量的原始單位不必在簡單線性迴歸中保留。如果檢查 y 與 x 的曲線揭示了一種非線性的關係，通常可能藉由替換 x 或 y ，或兩者，而得到線性關係。變換 y 和 x 的值，接著代入方程式 14.2 而得到 β_0 和 β_1 （在轉換的尺度上）的估計。例如，在圖 14.1c 的數據可能通過取 y 的對數進行線性化。係數 β_0 和 β_1 估算方式為：

$$\hat{\beta}_1 = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2} \quad (14.3)$$

和

$$\hat{\beta}_0 = \bar{y} - \beta_1 \bar{x} \quad (14.4)$$

基於估計的迴歸係數，依賴變數的預測（或期待）值， $\hat{y}$ 為

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i \quad (14.5)$$

公式 14.3 和公式 14.4 中的估計值應該在最小化 y 的觀察值和預測值之間的平方差之和的意義上是「最佳的」，即，它們減少了餘值的平方和：

$$SS_{Res} = \sum(y_i - \hat{y}_i)^2 = \sum \varepsilon_i^2 \quad (14.6)$$

這也就是為什麼我們又稱迴歸方程式像是使用「最小平方法」的數據分析法。

用於評估迴歸方程式品質的基本準則是，確定轉換的線性函數是否大量減少了依賴變數的變異性。

y 圍繞線（亦， $\hat{y}$ ）的變異性與 y 圍繞在其平均值 $\bar{y}$ 的變異性相比（即 y 原來的「預期」值）（參見圖 14.7）。這個概念通過統計學形式化的方式即為假設迴歸分析的殘值為常態分佈且相互獨立的，有相同的平均值，0，以及

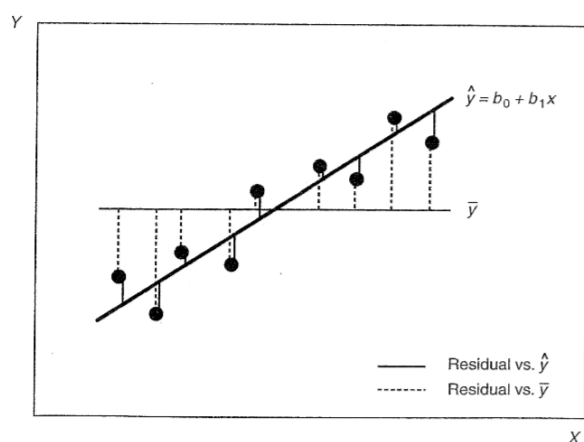
相同的變異數， σ^2 ，即， $\varepsilon \sim n(0, \sigma^2)$ 。接著可以使用 ANOVA 以確定不明原因的變異中是否在使用了最小平方估計法，即由 x 值預測 y 後，顯著地減少。ANOVA 表中針對簡單線性迴歸使用的 F-比率，如表 14.1，實際測試 $H_0: \beta_1 = 0$ 與 $H_1: \beta_1 \neq 0$ ，發現如前所述的變異性以相等的程度降低（如果 $\beta_1 = 0$ ，則為水平線，且 $\hat{y} = \bar{y}$ ，因此並沒有發生變異性減少）。

其他用來來評估迴歸方程式的品質的標準有，確定係數（Coefficient of determination）：

$$R^2 = 1 - \frac{SS_{Res}}{SS_{Tot}} \quad (14.7)$$

總結了透過使用 x 預測 y 解釋的總變異性比例。在公式 14.7 中的 SS_{Res} 和 SS_{Tot} 從表 14.1 分別為餘值以及 ANOVA 的總平方和。感官評估中， $R^2 > 0.75$ 的值通常被認為是可接受的。然而，實際上仍然取決於預期使用的迴歸方程式。使用其他標準可能可以使得資訊更加豐富。使用於 β_1 的信賴區間可使用下列算式建構：

$$\hat{\beta}_1 \pm t_{\alpha/2, n-2} \sqrt{MS_E / SS_E} \quad (14.8)$$



Residual 餘值

圖 14.7 擬合迴歸線（fitted regression line）所得到的餘值與從 $\bar{y}$ 得到的餘值的比較，用於確定擬合線是否顯著地減少了回應中無法解釋的變異度。在 $\bar{y}$ 和 $\hat{y}$ 之間餘值大小減小（即：與「預期」值的距離）顯示了迴歸線良好地匯總此組數據。

表 14.1 簡單線性迴歸的 ANOVA

變異來源	自由度	平方和	均方	F
總計	$n-1$	SS_T		
迴歸	1	SS_{Reg}	$MS_{Reg} = SS_{Reg}$	MS_{Reg} / MS_E

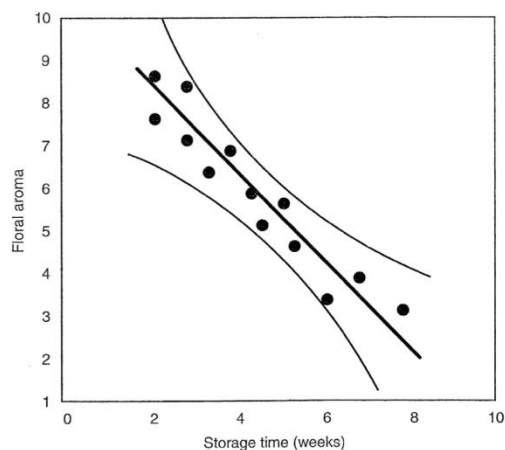
誤差	$df_E = n - 2$	SS_{Res}	MS_{Res} $= SSE/df_E$

其中， $SS_x = \sum(x_i - \bar{x})^2$ 且 $t_{\frac{\alpha}{2}, n-2}$ 是具有 $n-2$ 個自由度的學生 t 分佈 (student's t-distribution) 的上 α -臨界值。ANOVA 的 F-比率只告訴分析師 β_1 是否不等於零。公式 14.8 的信賴區間則告訴分析師 β_1 的估計是否在應用中足夠精確。 β_1 的信賴區間概念可以延伸到 y 的預測值的信賴帶 (Confidence band)：

$$\hat{y}_0 \pm t_{\frac{\alpha}{2}, n-2} \sqrt{MS_E \left[\left(\frac{1}{n} \right) + \frac{(x_0 - \bar{x})^2}{SS_x} \right]} \quad (14.9)$$

該信賴帶可以沿著預測值繪製，使得我們可以視覺評估擬合品質 (見圖 14.8)。如果信賴帶太寬，那麼無論是 F-比率測試或 R^2 的值，擬合的簡單線性迴歸方程式都不夠好。

存在好幾種可能性來解釋一個不佳的擬合迴歸方程。大致上可以從迴歸剩餘圖進行研究。剩餘, ε ，有 95% 的信賴帶的擬合迴歸線。信賴帶的寬度使得我們可以視覺評估擬合線的品質。窄的信賴帶顯示該擬合線適當的代表數據，而寬的信賴帶則表示仍有大量原因不明的變異性存在。



橫軸：存放時間（周）

縱軸：花香氣味

圖 14.8 95% 的信賴帶的擬合迴歸線

應該和預測值， $\hat{y}$ ，以及獨立變數， x ，一同繪製成圖。剩餘應隨機分散在預測值和獨立變數的範圍之間 (見圖 14.9a)。任何明顯的趨勢皆代表簡單線性迴歸是不夠的，且 x 和 y 之間存在一個更複雜的關係。可能需要高階項 (例如， x^2) (見圖 14.9b)，或者可能需要再執行一些數據轉換 (參見圖 14.9c)。不管在

垂直或水平方向上，存在與其他觀測值距離遙遠的單點，其可能是對擬合模型產生巨大不合理的影響的離群值（見圖 14.9d）。應當審查這樣的觀察值，並適當地予以消除。

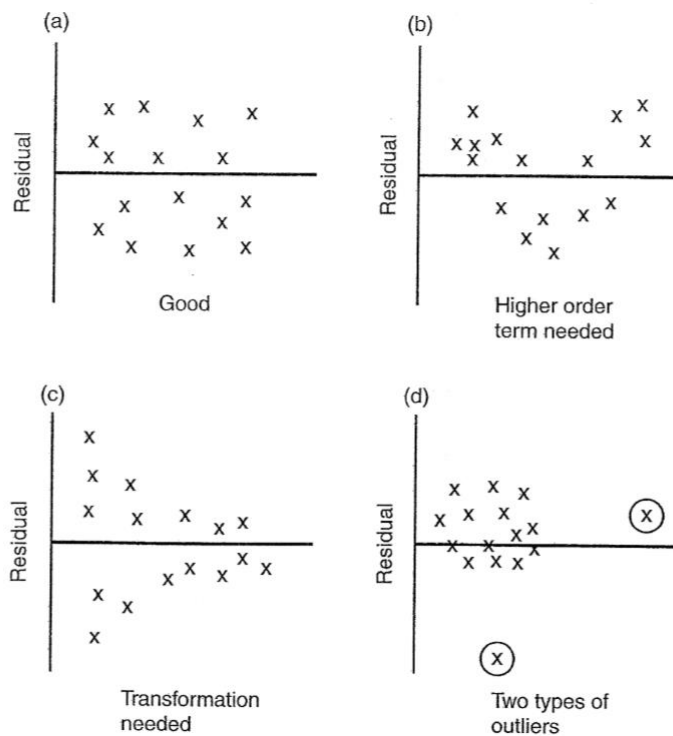
14.2.2.1.2 多元線性迴歸

有時候為了得到可接受的預測值， y ，需要多於一個的獨立變數。這可能發生在單一變數 x 的多項式上，因為 x 和 y 之間的關係不是直線，如在

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3$$

在其他情況下，回應可能由一個以上的獨立變數所影響，如在：

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$$



A: 良好

B: 需要高階項

C: 需要數據轉換

D: 兩種離群值

圖 14.9 餘值與簡易迴歸曲線圖顯示了 (a) 符合期望且隨機的排列，以及一些非期望的模式，即，圖 (b)、(c)、(d)，以及它們相對應的解釋。

或這兩種情況的組合可能存在。無論如何，多元迴歸分析是簡易的線性迴歸的直接延伸，它允許多個獨立變數被包含在 y 的迴歸方程式中。組成此分析的整體包括評估模型中納入的每個條件的價值。使用多元迴歸相關的缺陷是，所產生的迴歸方程式會影響原有的依賴變數之間存在的任何關係。多重共線性

(Multicollinearity) 是用於描述兩個或更多個獨立的變數彼此高度相關的情

況的術語。此相關性將影響估計係數， b_i' ，的值且可能導致對於模型中每個項的重要性不正確的結論。重要的是，在進行多重線性迴歸分析之前先進行研究依賴變數的相關性結構。相關性分析將更有意義如果伴有獨立（和依賴）變數的散點圖矩陣，如圖 14.2。多重迴歸的一個避免多重共線性的設計，稱為反應曲面法（Response surface methodology, RSM），將在 14.3 節討論。

不是所有的獨立變數都需要被包括在多元線性迴歸模型中。某些獨立變數可能在預測回應上較差，或者由於多元共線性，兩個或多個獨立變數可以解釋相同的變異性。選擇納入模型中的變數有數種方法可供選擇（參見 Draper 和 Smith, 1981）。大多數電腦軟體組包括一個以上的方法。

一個「蠻幹」的方法是「所有可能的迴歸」的方法。正如其名稱所暗示的，所有獨立變數的可能子集都將被考慮，從單變數模型開始，接著考慮所有可能的雙變數模型等。電腦輸出通常僅呈現各種尺寸的組別少數幾個最佳模式。有數種標準可以確定在所有可能的迴歸中，哪些模型是最好的。多重 R^2 （公式 14.7）是一種常見的方法，其中我們偏好較大的值。另一個標準是 ANOVA 剩餘均方的大小， MS_{Res} 。 MS_{Res} 是估計的餘值，所以偏好較小的值。

與剩餘均方標準相關的是調整的 R^2 ，其中

$$R_{adj}^2 = 1 - \frac{(n-1)MS_{Res}}{SS_{Tot}} \quad (14.10)$$

闡釋 R_{adj}^2 的方式和 R^2 相同。在多重迴歸中，剩餘均方最初將因為新項增加到模型中而降低； R^2 則將增加。某些研究會發現新項的進一步增加將導致剩餘均方值升高（不好的跡象），但 R^2 將繼續增加（好的跡象）。當剩餘均方開始增加，調整 R^2 將開始降低時，兩個統計值才達到質性上的一致。

在所有可能的迴歸中常用的最後一個標準是 Mallows 的 C_p 統計（1973）：

$$C_p = \frac{SS_{Res(p)}}{MS_{Full}} + 2p - n \quad (14.11)$$

其中， $SS_{Res(p)}$ 為含 p 項模型的剩餘平方和，且 MS_{Full} 是從包含所有獨立變數的模型的剩餘均方。不像 R^2 和 MS_{Res} 中， C_p 考慮完全模型是如何好，並使用此作為比較的基礎，以測量僅含有獨立變數的子集的模型的品質。 C_p 的一個缺點是，它只能在當數據組的觀察值的數量大於獨立變數的數量時才能計算。舉例來說，如果描述性屬性的數目大於樣品的數量越多，則 Mallows 的 C_p 標準不能使用。

另一個在多元迴歸中使用的變數選擇法是逐步包含法（Forward inclusion）、逐步淘汰法（Backward elimination）和逐步選擇法（Stepwise selection procedure）。這些方法是使用類似的標準來決定獨立變數是否該包含在模型中。

逐步包含法首先增加降低不明原因變異性最多的獨立變數到模型中，由剩餘的總平方 (SS_{Res}) 衡量。基於 SS_{Res} 的額外減少的結果而增加此新項。電腦軟體組使用一個 “f-to-enter” 統計確定加入特定項是否將導致無法解釋的變異在統計顯著上的減少。具有最大的 F-to-enter 值的獨立變數是添加到模型中的下一項。在 F-to-enter 值可以被設置為相對應分析師預期的顯著性水準。例如，在含有 30-50 觀測數據組中，4.0 的值大致對應於顯著性水準 $\alpha = 0.05$ 。逐步包含法持續增加新項直到沒有 f-to-enter 的值足夠大（與在程序中設定的值相比）。

逐步淘汰法從納入模型中所有的獨立變數的開始，基於它們解釋依賴變數的變異性的貧乏度消除此項。電腦軟體組使用 “F-to-remove” 統計值來衡量一個特定值是如何不重要的。擁有最小的 F-to-remove 值的項是下一個該從模型中排除的項。同樣地，分析人員可以設定 F-to-remove 的值，而該過程將繼續刪除條款，直到沒有過小的 F-to-remove 值（4.0 是 F-to-remove 良好的初始值）。

逐步選擇法是在構建多元線性迴歸模型上推薦的變數選擇過程。逐步選擇法結合了逐步包含法和逐步淘汰法，允許程序中每一個步驟能夠選擇從模型中消除項或增加新項至模型中（從兩項已加入模型後開始）。最初使用逐步包含法的 F-to-enter 標準選擇加入模型的項。由於多元共線性，每一個項的重要性根據模型中納入的其他項而變化。正如一項加入模型中可能導致另一項會變得多余。逐步選擇使用逐步淘汰中 F-to-remove 標準刪除這樣的項。在過程中每個步驟每個獨立變數的 F-to-enter 值和 F-to-remove 皆重新計算。當沒有統計值滿足由分析員先前所設置的值時，結束分析。

所有用於評估簡單的線性迴歸模型品質的診斷皆可用來確定多重線性迴歸模型的擬合度。這些包括 R^2 （現在還包括了 R^2_{adj} ）、 MS_{Res} ，對個別係數的信賴區間，以及最重要的，剩餘曲線圖。漏掉非線性關係或異常值的可能性在多元線性迴歸中更高，因為在三維以上視覺化的困難度。剩餘與獨立變數的曲線圖是解釋問題，以及/或確定該模型是否該進一步改進的最簡單的方法。

14.2.2.2 主成分迴歸(Principal Component Regression)

多元線性迴歸的一個弱點是它預測相關變數（即變數 x ）的方法，涉及多元共線性問題。如前一節所指出的，如果兩個高度相關的預測變數包括在迴歸模型中，大小、甚至斜率係數都可能會被誤導。此問題透過使用任一種變數選擇法，例如逐步選擇法，得以在一定程度上克服。

但是，從逐步選擇法產生的迴歸模型並不能整體說明所有涉及到預測的變數， y 的 x 值。例如，使用描述性品評小組所做出的屬性強度結果的數據來預測消費者在特定的屬性上的接受度，如甜味可能是和接受度高度相關的，但它可能不會出現在迴歸模型中，因為它與另一個屬性，如甜後味是高度相關的，而此屬性可能已經在模型中。逐步選擇過程將不會把甜味和甜後味一起納入在迴歸模型中，因為它們是高度彼此相關的（因此被解釋為在接受度上有相同變異性）。這使得研究者得到不正確的印象，認為只有甜後味而非甜味對於接受度非常重要。更正

確的解釋是，在迴歸模型中出現的項，例如，甜後味，實際上代表所有與它高度相關的描述屬性。然而，迴歸模型沒有透露哪些屬性是相關的。因此，應用在相關預測變數的逐步迴歸的過程本身並無法發現所有影響接受度的屬性。

主成分迴歸(PCR)是克服這個弱點的方法。PCR 是 PCA 和迴歸的簡單組合。同樣在使用描述性數據來預測接受度的例子中，PCA 是應用在研究中的樣品平均屬性。每個樣品獲得一組因子分數。因子分數被當作預測變數，即， x 值，使用，在迴歸分析中預測消費者接受度， y 。從 PCA 得到的因子表示樣品中感官變異性的基本維度，並且通常可以使用因子負荷 (factor loading) 解釋。在單一因子上具有大的正或負的負荷量 (即，接近 +1 或 -1) 的屬性是定義此因子的屬性，因此，如果因子被發現是影響接受度的一個顯著驅動因子，研究者即知道所有與該因子相關的屬性，作為一個整體都影響了接受度。此外，從 PCA 得到的因子不彼此相關。因此避免了多元共線性的問題。Popper, Heyman, Rossi (1997) 提出了一個很好的例子，應用在消費者接受度上的 PCA，12 個蜂蜜芥末沙拉醬基於其描述性檔案而預測的消費者接受度。

PCA 確定的因子是根據多少原始數據可解釋變異性的排序。第一個因素最大影響變異性，其次為第二因子，以此類推，有些情況下，在原始 x 變數最大影響變異性的因素可能不是 y 的最高度相關的因子。實際上，某些因子可能完全與 y 沒有相關，因此不需要被包括在 PCR 模型中。逐步迴歸可以用於產生只包含有統計上顯著能預測 y 的因子的 PCR 模型。以描述性分析為例，逐步的 PCR 允許研究者得以確定哪些感官變異性 (及其所有相關屬性) 的基本維是影響接受度，哪些沒有影響。

PCR 法可進一步延伸，透過識別該因子可能與反應變數， y ，在非線性和交互的方式相關。非線性和交互關係可以在 PCR 中通過包括和因子得分的交互相乘和平方，如同反應曲面分析 (見 14.3.3) 中的迴歸變數分別計算以完成。這允許研究者識別因子分數相關聯，例如，最喜歡的產品，的程度。

14.2.2.3 部分最小平方迴歸(Partial Least Squares Regression)

如上所述，PCA 可能產生與迴歸中的獨立變數， y ，不相關的因素。部分最小平方 (PLS) 迴歸 (見 Martens 1986) 是與 PCR 相關的多元技術，並且克服了這個弱點。其中，PCA 僅集中於說明由相關預測變數 (x -變數) 表現出的變異性，PLS 則導出因子 (即， x 變數的線性組合) (1) 解釋 x -變數的大部分變異性，(2) 同時盡可能與依賴變數 y 相關聯。雖然 PLS 的因子可能不能如同在 PCA 中的因子盡可能的解釋 x 值的變異性，PLS 確保識別每個因子最大預測 y 的能力。

PLS 擁有與 PCR 不同的兩個重要的優勢。第一，它產生的圖形輸出清楚地顯示的預測變數， x 值和欲預測變數 y 之間的關係。其次，PLS 容易擴展到同時預測多個依賴變數。當欲預測單個依賴變數時，分析被稱為 PLS1。當欲預測多個依賴變數時，分析被稱為 PLS2。PLS2 圖形關係可以用來說明，例如，消費者詞彙如何涉及描述性屬性的評分 (見圖 14.10)。

PLS 迴歸中所執行的計算已經超出了我們討論範圍。有數種電腦程式可以執

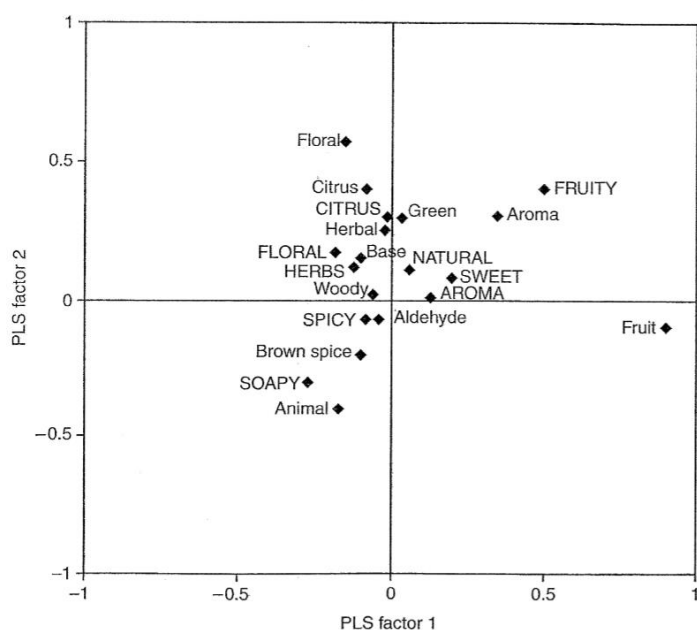


圖 14.10

使用 PLS2 的圖形關係說明消費者詞彙和描述性屬性評分之間的關係。大寫字母的項目是消費者回應，小寫字母的項則是描述性的屬性。需注意的是消費者和描述性的詞彙並不總是相似的。越來越多的出版物與程式在感官評估中應用 PLS 的範例，使得這種技術被越來越多感興趣的研究人員使用。

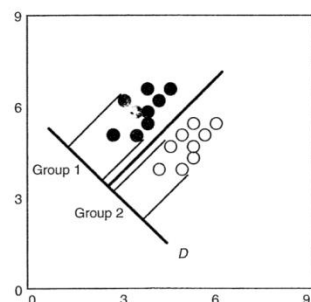
14.2.2.4 區隔分析(Discriminant Analysis)

區隔分析是用於將項分類到預先存在的類別（由離散依賴變數所定義）的多元技術。由連續獨立變數所建構的數學函數最能區隔從該類別中產生的項。例如，描述性屬性數據可以用於，從品質控制的角度，分類「接受」或「不接受」完成品，或者描述和實驗法可能被用來確定原料成分的來源（例如，國家或製造商）。

區隔分析是類似於我們已經討論過的幾個多變數技術。一方面，區隔分析與迴歸是相似的，使用一組連續的獨立變數預測依賴變數的「值」。在迴歸分析中，該值是指使用「迴歸方程式」預測的連續依賴變數的範圍。而在區隔分析中，此值是指使用「區隔函數」預測離散的依賴變數的類別。另一方面，區隔分析在使用獨立變數的相關性質開發新軸（即，原始變數的加權線性組合）上，則被認為類似主成分分析法。在主成分分析法中，軸是依照解釋變異性的多寡而被選擇。而區隔分析中，軸則是依照最大化依賴變數的離散類別與中心之間的差異而被選擇。區隔分析的簡單圖形可參見圖 14.11，其中顯示可接受和不可接受的產品在兩個描述屬性：庫存時間和鬆脆度，所組成的圖上。區隔函數定義新軸線，D，在其上兩個組別的平均之間的差異是最大的。

如果依賴變數只包含兩個類別，則僅需要一個區隔函數。如果依賴變數包含兩個以上的類別，則可能需要多於一個的區隔函數來準確分類觀察值。可能的判別函

數數目比起類別的數量少一個。



橫軸：庫存時間

縱軸：鬆脆度

圖 14.11 區隔分析的圖形，描繪兩個樣品在兩種原始屬性（庫存時間和鬆脆度）以及「最大區隔軸」，D。

無論如何，最佳分開類別的原始變數的線性組合（s）是最大化下列比率：

$$\frac{\text{各類別平均的變異}}{\text{類別之中的變異}}$$

除了這個比例，區隔函數的品質是由正確地分類的项目比例所測定。該評定可以透過使用被用來建立區隔函數的相同觀測值所實行。然而，當足夠的數據是可用的，預留一些建構模型的觀察值是最佳的，並在該區隔函數執行分類任務後才使用。這樣，驗證的過程則更為客觀。

不是所有的原始獨立變數都需要對每個項目準確分類。如同在迴歸分析中，有四個常用的變數選擇標準：逐步包含、逐步淘汰、逐步選擇，和所有的可能函數。在每個標準中，決定變數值的標準是它有助於類別之間的識別的程度。Powers 和 Ware（1986）提出了一種逐步區隔分析的匯總，其中 6 種藍乳酪僅使用了 28 種原始屬性中的 14 個最佳地分類此產品。

Powers 和 Ware 在感官評估中針對區隔分析的應用進行了全面的討論。其中討論了兩種線性區隔分析的選擇：典型區隔分析（canonical discriminant analysis）和鄰近區隔分析（nearest-neighbor discriminant analysis）。關係到其他多元技巧的各種工業應用，以及相關電腦軟體也都有介紹到。鼓勵有興趣進行區隔分析的分析師可以研讀他們提出的材料。

14.3 偏好地圖

偏好地圖是一系列用於展示感官（有時為儀器）數據和消費者接受度的關係的多元技術的集合。偏好地圖的研究有很多名字，包括：類別評價（Category appraisals），競爭性評估（Competitive assessments），產品空間地圖（Product

space mapping) 和關鍵驅動因素分析 (Key drivers analysis) 等等。各式各樣的名稱會導致混亂。例如，當一個組織提及偏好地圖研究，可能是作為類別評價時，另一個組織可能保留類別評估研究這個詞作為追蹤市場資訊，如銷售份額、發起廣告、擴展新生產線等的方法。由於某些人認為「偏好地圖」這個術語太過於技術，討論這種方法時應注意同事是否適當的了解，而沒有與其他測試方法的術語混淆。

產生混淆的另一個原因是有數種偏好地圖存在。兩種主要的分別是內部偏好地圖和外部偏好地圖。在這一章中，PLS 是作為第三種偏好地圖呈現。判別三種方法的特徵在於，用於定位地圖上的測試產品的資訊。內部偏好地圖使用消費者接受評級定位在地圖上的產品。外部偏好地圖則使用感官描述性屬性評級定位在地圖上的產品。PLS 地圖同時使用消費者和感官數據定位在地圖上的產品。

為了搞清楚「內部」和「外部」之間的差別，記得「偏好地圖」這個術語是由進行接受度測試的行銷研究人員所創造的，這應該會有所幫助。消費者接受的評級是供「內部」進行研究；因此，使用消費者評級在地圖上定位樣品的方法是「內部偏好地圖」。對於行銷研究者來說，感官屬性的評級是「外部」的資源，因此使用描述性屬性評級在地圖上定位樣品的方法是「外部偏好地圖」。內部、外部和 PLS 是偏好地圖的三大類。落入此三類中的方法有許多細節上的變異性。因此，再次提醒，在使用偏好地圖研究前，必須再次確認你和你的同事對於將要使用的方法的概念是一樣的。

單一數據組可以用來說明三種方法的偏好地圖。這些數據包括對 10 種餐點產品評估 30 種感官屬性。100 位消費者對於相同 10 種產品進行整體接受度測試。每個消費者評估所有 10 種產品。

在第 14.3.1 節，消費者數據是用來建立一個內部偏好地圖。地圖透過相關性分析的創意運用而延伸，說明感官數據與接受度之間的關係。在第 14.3.2 節中，外部優先地圖使用 PCA 和迴歸分析的發展，如同上一節中所討論。另外包括在第 14.3.2 節的是「偏好區隔」，其中群集分析被用於區隔在十個產品上具有類似喜好模式的消費者。平均接受度評分在外部偏好地圖之前進行是很重要的，以確定消費者群體是同質性的。各組若有不同的偏好平均可能掩蓋掉影響消費者接受度的感官屬性。最後，在第 14.3.3 節中，PLS 地圖應用在合併後的感官和消費者數據。包括在第 14.3.2 節中所確定的偏好區隔。

14.3.1 內部偏好地圖

內部偏好地圖第 14 章中的 14.2 節裡所討論的主成分分析 (PCA) 的一種應用。測試產品組成了數據即的橫列，而消費者的總體喜好程度評分則形成直欄(見表 14.2)。由於 PCA 需要完整的數據，消費者測試中，每個受訪者評估

表 14.2 內部偏好地圖的輸入數據的格式

產品	回應者 1	回應者 2	回應者 3	...	回應者 100

A	6	8	9	...	8
B	3	7	7	...	1
C	8	8	9	...	7
D	2	7	9	...	3
E	3	7	7	...	3
F	2	6	6	...	2
G	9	8	7	...	7
H	2	8	9	...	5
I	7	8	9	...	4
J	5	9	9	...	7

產品構成數據集的橫列。每個受訪者的整體喜好度形成直欄。

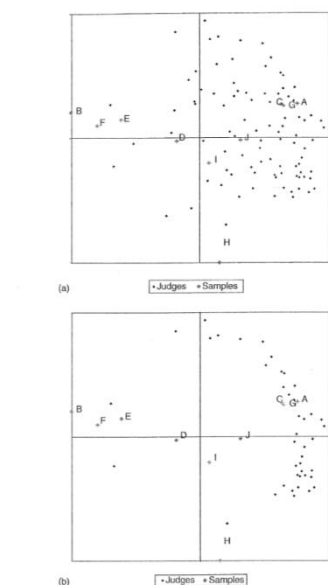
所有偏好的產品，以避免過度估算需要的遺失值，假如使用不完整的供應設計。將產品的因子得分和二維的 PCA 中的受訪者的因子負荷共同繪製在同一張圖上。

建構地圖是簡單的四個步驟：

1. 將產品的因子得分和受訪者的因子負荷繪製在同一張圖上（見圖 14.12a）。產品的因子得分需要被重新調整在 -1 到 +1 之間，因此它們落入與受訪者因子負荷相同的範圍內。對於每個維度來說，透過將原始因子分數除以最大的因子得分（忽略符號）來完成此計算。
 2. 刪除不能很好配合到地圖上的受訪者的喜好回應（見圖 14.12b）。這些受訪可以以兩種方式之一來標識。如果使用因子分析方法，如 SAS 中 (2005) PROC FACTOR，來擬合 PCA 模型中，輸出則包括每個回應者共同度 (Communality) 的統計信息。就像迴歸分析中的 R^2 值，共同度是受訪者喜好數據的變異性百分比，由 PCA 模型所解釋。刪除共同度小於預定值（通常是 0.50 和 0.75 之間的某處）的受訪者回應。或者，刪除最接近地圖上原點（即，(0,0) 點）的回應。從原點到回應的距離是共同度的平方根。藉由計算每個因子負荷的平方和得到從原點到每個受訪者的距離的平方。刪除回應的平方距離從原點小於預定值（再次地，典型上為 0.50 和 0.75 之間的某處）。
 3. 重新調整剩下的受訪者回應，使得其同樣地遠遠散落在距離圖的原點上（見圖 14.12c）。每個代表受訪者回應的點被從距離原點一個單位的距離重新調整。為了進行此步驟，將每個受訪者回應的因子負荷除此回應與原點的距離（即，距離 = $\text{SQRT}(F_1^2 + F_2^2)$ ，其中 F_1 即 F_2 是受訪者的因子負荷）。此步驟不是必需的，但在實踐中是常用的。
- 內部偏好地圖是獨立區隔的。受訪者傾向於在地圖上形成多個群集。一個在地圖上落在接近彼此的受訪者形成一個區隔。最接近他們的測試產品，在一般情況下，是他們最喜歡的產品。落在地圖上相對側的產品則是他們喜歡最少的產品。

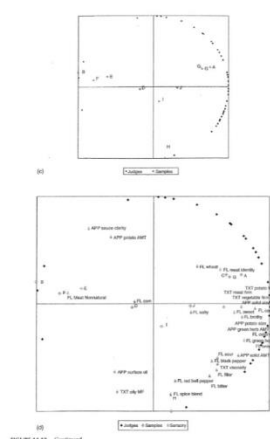
例如，在圖 14.12c 中，在右下角象限的一個區隔最喜歡產品 A、C、G 和 H，且最不喜歡產品 B、E 和 F。另外在右上角象限中有一個較為分散的區隔，他們也最喜歡產品 A、C 和 G，較能接受產品 B、E 和 F，也最不喜歡產品 H。

4. 將產品的屬性強度與其因子分數相關聯，以納入感官描述資訊，並在地圖上標繪所得到的相關係數（見圖 14.12d）。這種相關性分析的創意運用最早是由 Mc Ewan (1998) 提出。每個產品擁有全套屬性強度和兩個因子得分，一個是因子 1（相對於地圖的橫軸）和一個因子 2（圖的垂直軸）。對於每一個屬性來說，使用 x 座標作為屬性得分與因子 1 的相關性，使用 y 座標作為屬性得分與因子 2 的相關性。



Judges:受訪者 samples:樣品

圖 14.12(a)產品因子得分和所有受訪者的因子負荷皆重新調整過後的圖。(b)產品的因子得分重新調整過後與受訪者的共同度大於 0.50，即，該模型中的受訪者至少解釋了 50%自己的喜好評級變異。(c)重新調整產品的因子得分以及受訪者的因子負荷，使得其距離原點相等的圖。此步驟有時可以使我們更容易識別受訪者群集。(d)內部偏好地圖的最後包含了感官屬性。感官屬性的坐標是此樣品在兩個維度上的因子得分的相關度。



Judges 受訪者 samples 樣品 sensory 感官屬性

表 14.12 續

屬性彼此的方向揭示了其內部的相關結構。地圖上落在接近彼此的屬性是彼此正相關。落在地圖上相對側的屬性則彼此為負相關。落在近直角的屬性彼此不相關。更重要的是，樣品的屬性和對於受訪者群集的位置揭示了其正面影響或負面影響接受度。接近受訪者群集的屬性是該群集的正面影響。在這些屬性上我們偏好較高的強度。落在地圖上相對側的屬性則對於接受度是負面影響（這些屬性我們則偏好低強度）。屬性與群集呈現直角，則此屬性不主要影響該群集的接受度。例如，在圖 14.12d，在右下角象限的受訪者傾向於高強度的綠色香草、洋蔥、芹菜、胡蘿蔔口味。他們不喜歡太過清澈的醬、以馬鈴薯為主的外觀、以及非天然肉的風味。

要知道在所有形式的偏好地圖中，當我們說消費者更喜歡「低」或「高」強度的屬性，應該解釋為研究中呈現產品的「低」或「高」的強度，這一點是很重要的。偏好地圖中提及的「低」或「高」強度並不意味著用於評估產品的評定尺寸的極端值。產品屬性的強度再同一類別的範圍可能在 15 級尺寸中只佔了 2-5 的單位。對於產品接受度來說這是很大的差異，例如，一個關鍵屬性強度 2.5 的產品是與關鍵屬性強度 4.5 相反的。

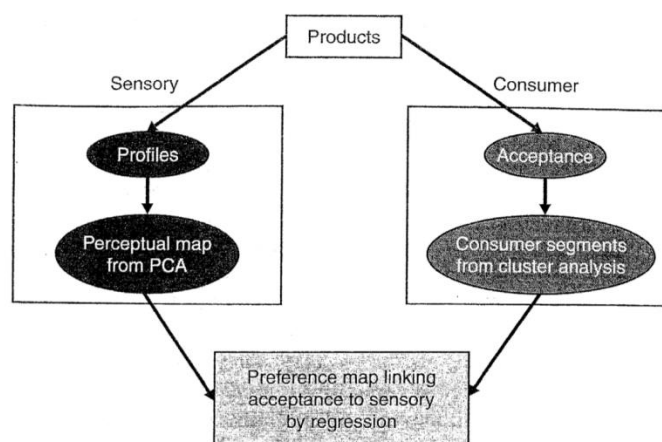
14.3.2 外部偏好地圖

外部偏好地圖是比起內部偏好地圖的構造要複雜得多。在此分析中使用到三種統計方法。PCA（第 14.2.1.2 討論）僅用於建立基於產品感官特性的產品空間的感知地圖。群集分析（在第 14.2.1.3 節所討論過）用於識別消費者的偏好區隔。偏好區隔是在測試完所有研究中產品的一組顯示出類似喜好模式，且不同於另外群集的受訪者。最後，使用迴歸分析（在第 14.2.2.1 節討論）以連結感官資訊以及消費者接受度，透過使用模型中的樣品因子得分，由感官地圖預測在總體受訪者的接受度，並在群集分析中確定是否有任何偏好區隔（見圖 14.13）。

14.3.2.1 建構產品空間的感官地圖

產品空間的感知地圖由產品－屬性數據的 PCA 獲得。產品構成數據集的各橫行（觀測值），且感官特性構成直欄（變數）。為了保持是為了創建地圖的想法，從分析中得到的主成分被稱為感知空間的關鍵感官維度（Key sensory dimensions of the perceptual space）。

在進行 PCA 之前，必須先決定是否將所有的產品和所有的屬性納入在分析中。假如研究中只有一個產品具有一個或多個屬性的超強閾值強度，則 PCA 很有可能得出一個完全得以與所有其他產品區分的因子。一個或多個這種「單樣品維度」在研究之中具有掩蔽其他研究中的產品差異的負面作用。例如，在研究中納入十個起司披薩和一個辣香腸口味披薩可能會掩蓋一些



Products 產品

Sensory 感官

Profiles 檔案

Perceptual map from PCA 從 PCA 獲得的感知地圖

Consumer 消費者

Acceptance 接受度

Consumer segments from cluster analysis 從群集分析獲得的消費者區隔

Preference map linking acceptance to sensory by regression 偏好地圖使用迴歸連結感官及接受度

圖 14.13 外部偏好地圖分析的示意圖。產品同時呈現至感官描述性評測和消費者測試。感官結果彙總在一個感知地圖上。消費者接受度的評級提交做群集分析，以確定偏好區隔。使用迴歸分析的連結感官和消費者資訊。

起司披薩餅之間細微但有意義的差異。若刪除香腸口味的披薩可能獲得關於驅動喜好度的屬性更深入的資訊，並且將所關注的類別重新定義為純起司披薩餅的屬性。如果只有一個屬性被納入模型中，一種選擇便是是刪除此屬性，特別是

當先前的研究已經表明此屬性與接受度沒有顯著作用。另外一種選擇是進行偏好地圖研究兩次。第一次可以包括所有的產品，並針對香腸相關屬性的效果進行研究。第二次則只包括純起司披薩產品，更具體地專注於驅動接受度的屬性。在消除可能的屬性上須要有同樣審慎的考慮。如果屬性的強度在所有的測試產品上相同，或者如果強度在一個平凡小值的範圍內變化（例如，在 15 個單位尺度上，此屬性強度範圍 ≤ 0.5 單位），則此屬性應當從分析中消除。同樣地，如果屬性的強度在所有的產品上皆低於閾值，此屬性也應被消除。如果 PCA 使用相對於反應的共變異數矩陣（covariance matrix）的相關性矩陣進行，這點是特別重要的。若含有小的數值範圍的屬性則可能發生大但是假的相關性。一個基於相關性的 PCA 在計算因子負荷和因子得分時，無法區分小範圍強度的屬性和那些有意義的大範圍。小範圍強度的屬性可能在定義產品的感知空間時扮演著關鍵的作用。這一點似乎支持 PCA 利用共變異數矩陣，因為共變異矩陣在低權重屬性上表現出很小的變異。然而，人們普遍認識到，在某些屬性（例如，異味）的微小差異可能對接受度有極大的影響。基於共變異矩陣的 PCA 能調整這些低權重屬性以反映他們對接受度真正的重要性。在消除小範圍的強度屬性後，使用相關矩陣保留這些屬性的重要性。正因為如此，我們推薦在外部偏好地圖的感知地圖的步驟使用相關性為基礎的 PCA。

任何刪除產品或屬性的動作必須謹慎進行。如果可能的話，納入所有的產品和特性進行分析，然後持續重複分析並且淘汰越來越多的產品和屬性。多重分析往往讓我們對單一產品和屬性有更清晰的認識和揭示更多有趣的見解，關於什麼在整體類別中驅使喜歡或接受度等。

候選產品的初步篩選可用於確定產品在類別的感官空間中的定位。研究者可決定是否消除極端的產品，或者消除具有獨特屬性的產品，以獲得新的、更狹義的類別，以及一個更均勻的全面性。可替代的、附加的產品或專門配製的原型品可以被添加到該研究中，以填補在感官空間中的任何「空缺」。這種篩選也提供研究者一些數據分析和開發資訊和見解的背景資料。

對於餐點準備的數據，土豆味、紅椒味和甜味屬性都因變異過小而淘汰。淘汰玉米味，因為只有樣品 D 有超過閾值的強度，初步相關分析表示，玉米香精對接受度無顯著影響。其餘 26 個屬性提交至 PCA。

必須在感知地圖的發展前作出的第一個決定是多少維度（即因子）將包括在地圖上。有三種標準來做出這個決定。第一種方法是再也沒有特徵值（Eigenvalue）的幅度大大下降。特徵值測量每個維度解釋的變異值。若加入另一個維度到地圖基本上不增加解釋變異度的量，也就是停止的時候了。這個最終的決定標準透過使用陡坡圖評估。當圖中的特徵值開始變平了，也就是時候停止添加維度（見圖 14.14）。第二種標準是在解釋了至少 75% 數據的總變異數時停止一次。第三個標準是停止加入新維度當單個特徵值低於 1.0 時。所有標準 PCA 輸出包括以下兩條信息（見表 14.3）。

圖 14.14 表明，特徵值開始在第三維變平。表 14.3 表明，只有兩個維度時

便已解釋了 75% 的變異性，以及單點特徵值低於 1.0 時是在第六個維度加入時。決定進行二維、三維和四維的方案，並檢查哪種結果的因子負荷使得感官及產品看法最有意義。

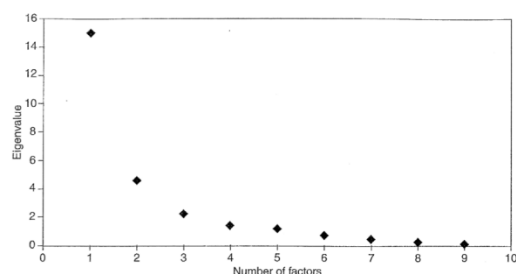
TAB LE 14.3 特徵值和從 PCA 獲得的解釋變異的匯總

因子	特徵值	解釋變異 (%)	累計 (%)
1	14.98	58	58
2	4.59	18	75
3	2.22	9	84
4	1.41	5	89
5	1.23	5	94
6	0.76	3	97
7	0.44	99	2
8	0.24	1	100
9	0.12	0	100

變異性的百分之七十五是由前兩個因子所解釋。直到因子 5 時特徵值皆大於一。

圖 14.14

從 PCA 的特徵值的陡坡圖可以幫助決定多少因素應該納入感知地圖。



Eigenvalue 特徵值

Number of factors 因子數

在感官分析師和產品開發者審查結果後，選擇三維的方案。三個維度的解釋可以透過檢查每個維度的屬性因子負荷。因子負荷類似於相關係數。它們的值範圍從-1 到+ 1。接近兩個極端值指示屬性和維度之間有很強的關聯。小的值（例如 ± 0.6 ）顯示弱的相關性。集中在較大的因子負荷有助於解釋每個維度。檢查表 14.4 的因子負荷，發現第一感官維度，在一般情況下，是關於產品的“有益健康”的特性。在第一維度的產品範圍有在非天然肉味和香精等得分高的產品，也有含有許多大塊和堅實的肉、馬鈴薯、蔬菜的產品，以及在綠香草味，胡蘿蔔、芹菜味道上具有高的強度。

第二感官維度是醬的外觀和質地，以及辣味的組合。產品範圍從輕薄和清澈

的醬料到粘稠的醬料產品。粘稠的醬料產品往往在辣味/黑胡椒味上表現較高。為了簡便起見，第二維稱作是「醬：清澈／黏稠」。第三個感覺維度則清楚的表現感知產品的「油性」差異。

這點出了關於感知地圖很重要的一點。在地圖上出現的維度是被包括在分析中的感官屬性之間存在的相關性。這些相關有時是在產品類別中所固有的。例如，它可能在某產品類別中是自然的，較薄調味的醬汁具有更高的清澈度，而較濃的醬汁則較不透明。然而，一些相關性可能是僅出現在分析中的巧合，或是該組特定的產品組合。例如，濃厚、半透明的醬汁不一定是辛辣和黑胡椒味重的。當解釋感官維度時，區分研究中的相關性是固有或是巧合是非常重要的。

表 14.4 感官品質的因素負荷，根據因子順序排序

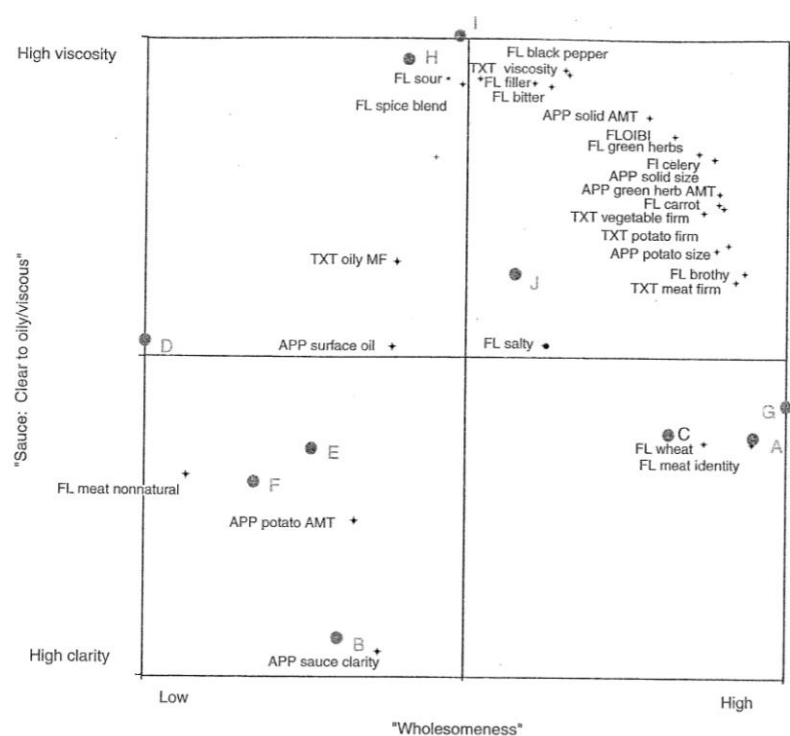
屬性	有益健康	醬：清澈／黏稠	油性
FL 肉塊感	0.90		
FL 高湯	0.87		
TXT 堅實的肉	0.84		
APP 馬鈴薯大小	0.82		
APP 香草 AMT	0.80		
FL 胡蘿蔔	0.79		
APP 固體大小	0.79		
TXT 堅實的馬鈴薯	0.78		
FL 芹菜	0.77	0.62	
FL 小麥	0.75		
TXT 堅實的蔬菜	0.74		
FL 綠色香草	0.72	0.64	
FL 非天然肉	-0.86		
TXT 粘度		0.89	
FL 黑胡椒		0.89	
FL 酸味		0.87	
FL 填料		0.86	
FL 混合香料		0.85	
FL 苦		0.85	
APP 固體 AMT		0.75	
FL 洋蔥	0.65	0.70	
APP 醬清澈度		-0.92	
APP 表面油感			0.92
TXT 油性 MF			0.88
FL 鹹			

APP 馬鈴薯 AMT			

只有顯示負荷大於 ± 0.6 。因子負荷類似於相關係數。值接近 ± 1.0 是重要的。一個因素大負荷的屬性有助於解釋感官變異性。

命名感官維度有好有壞。名稱有助於解釋且較容易熟記結果。參看感性地圖 DIM1, DIM2 等經常關閉計畫團隊技術較少的成員，因此資訊不完全地被使用。然而，幫維度取名等於畫下一個界線。如果一個重要的屬性沒有在維度的名稱中提到的，資訊的使用者可能會忘記它扮演的角色。在命名維度時要盡量使用具有最廣泛吸引力描繪感知地圖的名稱，且技術和非技術人員都能清楚了解的詞。但要強調的是，名稱不是所有涉及到的屬性的全面總結。像表 14.4 中，其中的屬性的因素負荷量 ($>\pm 0.6$) 已被替換成自己的符號，是為了說明所有在感官維度扮演顯著角色的屬性的一個好方法。

屬性和產品之間的關係，現在可以在感知地圖上顯示。圖 14.15 顯示屬性和測試產品的前兩種感官維度的關係。由此可以看出，例如，產品 B 有高醬料清晰度但低粘稠及辣味的醬料，而產品 H 和 I 則是醬粘度和辛辣味較高和醬清晰度較低。產品 D, E, 在非天然肉味和香精上得分高，以及肉的大小和堅實感、堅實的馬鈴薯和蔬菜片上得分低。產品 A, C 和 G 在肉塊感以及小麥風味上高，和小麥香料，以及肉塊大小和堅實感，馬鈴薯，和蔬菜片的堅實感，和蔬菜風味。



橫軸：「有益健康」

縱軸：「醬料：清澈／黏稠」

圖 14.15 在準備餐點研究的前兩個維度的感知圖。產品 A, C 和 G 是肉塊感和

小麥口味高。產品 B 是醬清晰度高和低粘稠質地、酸、苦、辣和黑胡椒的味道。產品 H 和 I 有一套相反於產品 B 的特點

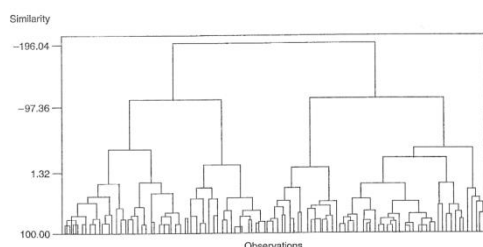
14.3.2.2 識別偏好區隔

現在，產品空間的感知地圖已經被開發出來，感官資訊可以被連結到消費者的接受度。在一個典型的偏好地圖的研究中，消費者的接受評級的平均跨越所有受訪者以及基於人口統計的各小組內的態度和使用模式等（如按年齡或性別，運動愛好者或久坐者，重或輕類別用戶等）。了解不同的人口、態度和使用習慣的喜好模式是在定位產品，在市場中找尋利基等機會上非常重要。然而，尋找這些類型的區隔，潛在的假設是，每個在區隔中的人都對產品的類別有著相同的偏好。這可能不是真的，並且，因此，個體的真實偏好可能通過那些具有不同的產品偏好而抵銷了他們的喜好評分平均。因此，在執行連結感關和消費者資訊的分析之前，重要的是要確保產品的平均喜好評分來自具有相似偏好的消費者群組。群集分析是用於完成此任務的工具。

如前一節中所討論的，有數種方法可用於執行群集分析，且每種主要方法中有許多變型。在準備餐點的數據分析中，採用 Ward 的階級群集法。分析的數據是消費者的整體喜好評分。受訪者構成數據集的橫行（觀察值）；產品構成的直欄（變數）。透過從個人的評分減去每個受訪者的平均喜好評分來中心化喜好評分。中心化移除了原始數據使用的尺度效果。例如，兩個受訪者可能在測試產品上有相同的喜好，但一位受訪者傾向於使用刻度的中間部分（4、5、6）評價產品，而另一位受訪者則傾向使用較高端的尺度（7、8、9）評價產品。喜好評分的中心化使得群集分析根據評價的模式分組，而不是根據評分的絕對水準分組受訪者。

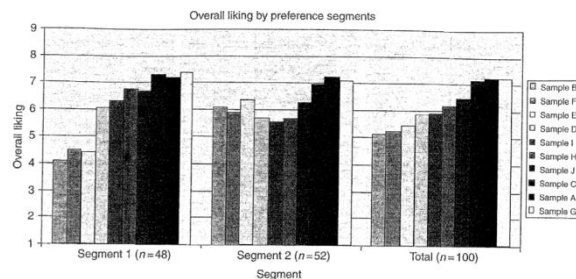
14.16 的樹形圖表示，兩個或四個偏好區隔存在。多少偏好區隔包括在偏好地圖上的最後決定需要針對區隔的大小以及區隔內部均勻性做平衡。平均評分不到 25 到 30 位受訪者的區隔應避免使用，因為從這樣的小團體得出的平均值缺乏精度。在本例子中，四個區隔中有兩個區隔少於 25 位受訪者。出於這個原因，我們選擇二個偏好區隔。

區隔之間的差異在他們的平均總體喜好評分的曲線圖（見圖 14.17）中顯示。區隔 1 具有較廣泛的平均評分。



Similarity 相似性 Observation 觀察值

圖 14.16 中心化過的消費者整體喜好評分的樹狀圖，建議兩個或四個區隔。最後選擇兩個區隔，因為四個區隔的方案中有兩個區隔少於 25 位受訪者。



Overall liking by preference segments 偏好區隔的整體喜好

Overall liking 整體喜好

Segment 區隔

Sample 樣品

圖 14.17 根據偏好區隔排序的產品平均喜好評分，顯示了這兩個區隔都最喜歡產品 A，C 和 G。區隔 1 不喜歡產品 B，E 和 F，區隔 2 並沒有不喜歡任何產品，但較小程度的喜歡產品 D，H，I。

在區隔 1 的受訪者最喜歡產品 A，C 和 G，且最不喜歡品 B，E 和 F。在區隔 2 中的受訪者的評級表現出較窄的範圍，區隔 2 也很喜歡產品 A，C 和 G，但不同於區隔 1 的受訪者，他們顯示出最少喜好的产品是 D，H，I。

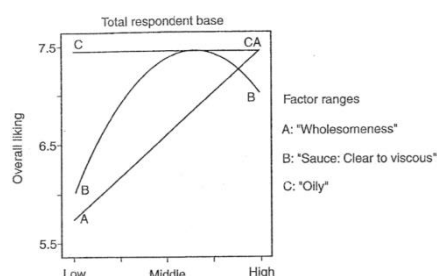
14.3.2.3 從感知地圖到偏好地圖

在外部偏好地圖開發的最後一步是基於整體平均喜好評級和研究中所有感興趣的消費群體（人口統計、態度、使用習慣和偏好區隔）進行擬合迴歸方程式。依賴變數（即，預測變數）是從 PCA 得到的測試結果的因子分數。因子分數的線性和二次項形式都包含在迴歸分析中。將二次項包含在迴歸模型中創造了識別比起極端值更受歡迎的維度中間點的機會。正因為如此，包括二次項的迴歸模型，被稱為理想點模型（ideal point model）。僅包括線性項的迴歸模型被稱為**向量模型**（vector model），因為它們只能指出喜好增加的方向。

如逐步迴歸或逐步淘汰的變數選擇法被用來鑑定與總體喜好度具有顯著關係的感官維度。當產品在測試中的數量足夠大，逐步淘汰優於逐步迴歸，因為它提供了所有的預測因子留在最終模型（Anderson, Whitcomb 2005）中一個平等的機會。

基於全部受訪者的迴歸分析結果顯示在圖 14.18 中。圖上的每個線條表示此感關維度對整體喜好有顯著影響。陡峭的線條表示對喜好度有大的影響；平坦的線條代表較小的影響。若線條的斜率向上表示，在較高的感官維度比起較低的更受人喜歡。相反的，若斜率下降則表明，較低的水準的感官維度更受人喜歡。彎

曲線表示中間點最受歡迎。要生成圖表上的線，保持所有感官維度，但將其中之一變為常數，並將剩餘維度的整體喜好結果從高到低的變化繪製在圖上。在所有的感官顯著維度上重複此過程。圖 14.18 表明，「有益健康」維度對於整體喜好有較強的積極影響，因此更高水準是首選。「醬料：清澈／黏稠」維度也對整體喜好有顯著的影響，因此這個維度的中高層水準是較佳的。「油性」維度的線條是平的，這表明該維度對喜好沒有顯著影響。



Total respondent base 整體受訪者基礎

Factor ranges 因子範圍

A 「有益健康」

B 「醬料：清澈／黏稠」

C 「油性」

圖 14.18 基於整體受訪者的關鍵驅動模型的擾動圖 (Perturbation chart)。(A)「有益健康」和 (B)「醬料：清澈／黏稠」對整體喜好度同樣重要。(C)「油性」對於喜好度不具有顯著影響。「有益健康」的高水準和「醬料：清澈／黏稠」的中高端水準是較佳的。

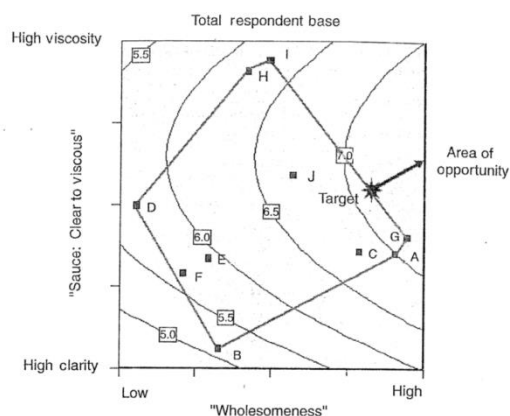
預測喜好評分在偏好地圖上的任何一點可以使用等高線圖（第 14.4 節中討論）來說明（見圖 14.19）。包括在圖中還包括由測試產品形成的凸體。凸體代表產品空間的限制。落在產品空間外的預測點是外推值，應謹慎看待。雖然迴歸模型可以表示在某個方向移動更遠應該會對喜好產生積極的影響，然而產品空間之外的點並沒有數據支持，因此我們沒有辦法知道此趨勢會繼續到多遠。因此產品空間外的點預測值是相當不可靠的。

被預測為最受歡迎的點如圖 14.19 中被標識為「目標」的產品。當目標位於產品空間的邊緣，如同它在本範例中那樣，由該迴歸模型所指示的喜好增加的方向可被解讀為機會區域。雖然沒有給出喜歡的預測值，但分析確實表明這一區域可能值得一些額外的探索。

14.3.2.4 目標產品檔案的逆向工程

在開發總體喜好數據的迴歸模型的同時，建立使用樣品的因子分數來預測原始感官屬性的迴歸模型。只有線性項被包括在模型中，且沒有使用任何變量選擇法，所以每個感官屬性是被所有感官維度所預測。對應於目標產品的因子分數安置在模型中的每個感官屬性上，以獲得對應於目標產品的預測感官檔案（見表

14.5)。



Total respondent base 整體受訪者基礎

Area of opportunity 機會區域

High viscosity 高度濃稠

High clarity 高度清澈

“Wholesomeness” 「有益健康」

Target 目標

圖 14.19 在準備餐點的研究中的偏好地圖的首先兩個維度的等高線圖，在線上的任何一點的被預測為具有總體喜好評級相對的值。最大化的總體喜好是在「有益健康」]上的高水準，以及「醬料：清澈\濃稠」的中高水準。為了保持在產品空間內，目標產品位於曲線凸角上。趨勢表示兩個維度較高的水準使產品可能更受歡迎。

14.3.2.5 個別受訪者的外部偏好地圖

另一種方法是在每個受訪者的整體喜好數據上做個別迴歸模型擬合。正如內部偏好地圖，差的擬合模型的受訪者可以在進一步的分析中刪除。對於其餘受訪者，定義為表示「滿意」的標準動作。例如，在偏好地圖上預測的喜好評價是 5.0 以上的任何一點可以說是「滿意」的受訪者。另外，在地圖上的任何一點其最大的預測喜好平極為 0.5 個單位的受訪者也可以被認為是滿意的。在偏好地圖上的滿意的受訪者百分比繪製可以繪製成如圖 14.20 的等高線圖。這種方法的一個優點是，像內部偏好地圖，它是自我區隔的。多目標的產品可以被識別為地圖上對應於高滿意度區域的單獨的點。這種方法與內部偏好地圖不同的優點是，每一個受訪者數據點都可以使用一個理想的擬合模型，以便在地圖上的內部點可以被確認為最大整體喜好程度的點。（內部偏好地圖，從某種意義上說，是一個向量模型，因為它只能指認喜好增加的方向。）該方法的缺點是個別受訪者的數據模型往往不佳。無論是從分析中刪除許多受訪者，或是使用非常寬鬆的定義都需要被用來保持在模型中有足夠比例的受訪者。

14.3.3 部分最小平方地圖 (PLS mapping)

PLS 地圖是第 14.2.2.3 描述部分最小平方迴歸的直接應用。作為一個偏好地圖工具，在 PLS 模型中的依賴變量 (y) 是消費者和獨立變量 (x) 的總體喜愛的評級是感官屬性評級。因為 PLS 可以在同一個模型中處理多個依賴變量，總體受訪者的總體喜愛程度的評分，以及感興趣的消費者區隔，都可以在單個分析中擬合。這有助於確認在區隔中驅動喜好模式的屬性相似性及差異性。

表 14.5 透過逆向工程測定目標產品的檔案

感官屬性	目標檔案
FL 肉塊感	4.9
FL 高湯	4.9
TXT 堅實的肉	6.2
APP 馬鈴薯大小	10.8
APP 香草 AMT	6.1
APP 固體大小	8.5
TXT 堅實的馬鈴薯	4.3
FL 胡蘿蔔	3.0
FL 芹菜	2.9
FL 小麥	3.7
TXT 堅實的蔬菜	2.7
FL 綠色香草	3.3
FL 非天然肉	2.3
TXT 粘度	3.1
FL 黑胡椒	1.7
FL 綜合辣度	0.2
FL 苦	2.8
FL 酸味	1.7
FL 填料	4.4
APP 固體 AMT	7.1
FL 洋蔥	2.4
APP 醬清澈度	4.1
APP 表面油感	8.8
TXT 油性 MF	5.6
FL 鹹	7.4
APP 馬鈴薯 AMT	7.4
FL 玉米	0.4

FL 馬鈴薯	0.0
FL 紅椒	0.1
FL 甜度	0.7

圖 14. 21. 是準備餐點例子在 PLS 的地圖上呈現所有受訪者的總體喜愛程度評分數據，以及圖 14. 17 介紹的兩種偏好區隔，分別是 PLS 分析的依賴變量。在外部偏好地圖中使用的感官特性同樣是 PLS 模型中的獨立變量。產品 A、C 和 G 出現在與消費者偏好區隔 1 和區隔 2 在相同象限上，這表明這些產品在兩個區隔中都是一樣受歡迎的。產品 B、E 和 F 落在地圖上與區隔 1 的相對側，說明這些產品沒有廣受那個區隔的消費者的喜歡。相反的，產品 H 和 I 落在從區隔 2 地圖的相反側，這表明這些產品對於這些消費者來說喜歡得最少。

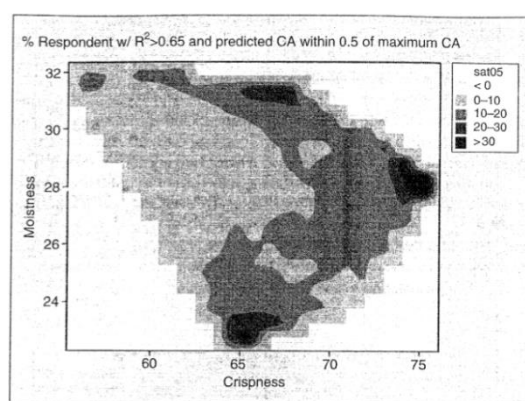


圖 14. 20 受訪者的等高線圖（每個個別模型具有 $R^2 > 0.65$ ）。注意高滿意的區域出現在多個地方，在中脆度/低濕度、高脆度/中濕度，和中脆度/高濕度。

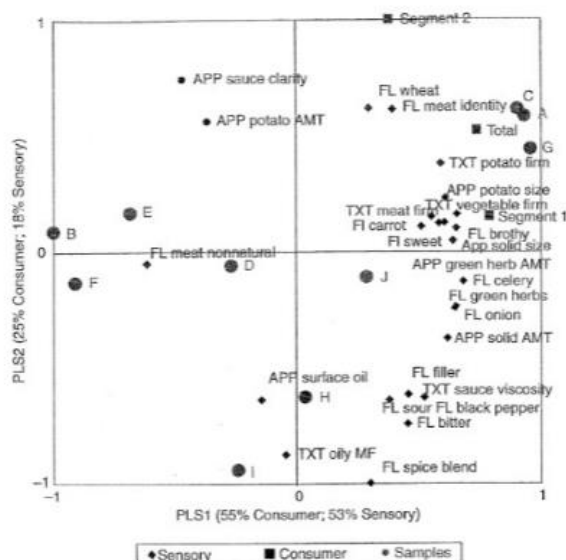
在圖 14. 21 中的區隔 1 和區隔 2 之間的點距離表示，在兩個區隔中不同的感官屬性驅動喜好度。落在接近區隔點的感官屬性的點對於該區隔具有正面的喜好度驅動。落於地圖上區隔的相對側的屬性則為負的喜好驅動。圖 14. 22a 和 b 的測試揭示了兩個區隔之間的主要驅動的差異。區隔 1 喜歡的產品，有很多大的和堅實的肉、馬鈴薯和蔬菜片，以及高強度的肉塊感、香草、胡蘿蔔、洋蔥、芹菜的味道和低強度的非天然肉味和香精。而另一方面，區隔 2 更喜歡許多堅實的馬鈴薯、清澈低粘度的醬料，高強度的肉塊感、低油性和低辣度（尤其是黑胡椒）。這些發現與內部和外部偏好地圖的結果非常吻合。

藉由這種方式，研究人員可以嘗試不同的統計方法來挖掘努力確認結果，並尋找關於產品與消費者的任何額外資訊的數據。

採用不同的統計方法來挖掘數據，使得研究人員能夠交叉驗證研究的主要成果，並揭開產品與消費者的額外資訊。

14.4 實驗設計的處理結構

在第 13 章中所討論的實驗設計中，處理（或產品）被視為一組性質不同的對象，相互之間沒有特別的關聯。這樣的設計可以說具有一個單向處理結構。單向實驗通常朝向研究方案的終點，因為其目標是決定應選擇哪個產品進一步發展。



PLS2(25%消費者；15%感官)

感官

消費者

樣品

圖 14.21 準備餐點一例中的 PLS 地圖，顯示偏好區隔之間的差異以及測試產品的分佈。兩個區隔同時喜歡 A、C 和 G 產品，區隔 1 不喜歡產品 B、E 和 F，區隔 2 則不喜歡產品 H 和 I。

然而在許多實驗的情況，研究的重點不在於特定的樣品，而是一些因子或已施加到樣品上的因子的影響。例如，研究者可以針對不同的麵粉和糖對一個特定蛋糕配方的味道和質感的效果感興趣，或者他可能對於烹調時間和溫度在肉品準備上的風味和外觀的影響有興趣。在諸如此類的情況下，也有針對所有因素和影響的強度進行的全面且準確的比較，並且同時最小化研究所需的實驗材料總量的具體計劃。

兩個「多方面」的處理結構在本節中討論。他們的階乘處理結構（通常稱為階乘實驗[factorial experiments]）和反應曲面處理結構（通常稱為反應曲面法[response surface methodology]，或 RSM）。

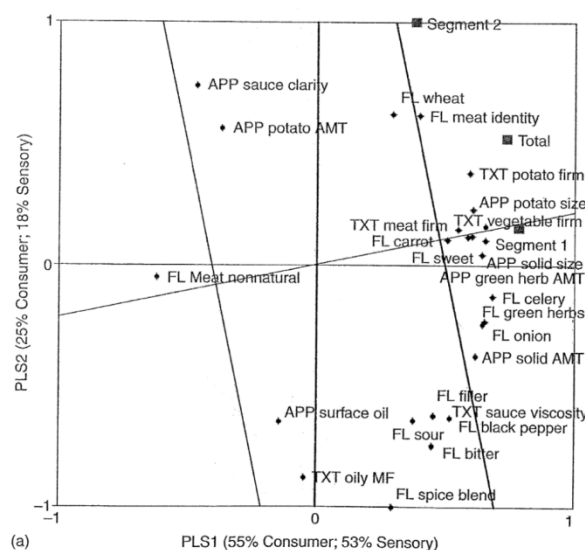
14.4.1 階乘處理結構

研究者通常對於兩個或更多因素對一組回應的影響有興趣。階乘處理結構是進行這樣子的研究最有效的方式。在階乘實驗中的每個因素的特定水準都已定義。階乘實驗的單次複製實驗是由因子的所有可能的級別所組合而成。例如，一個釀酒商可能在比較兩種沸騰時間影響他的啤酒的啤酒花香氣的效果有興趣。此外，如果此釀酒商當前正使用兩種啤酒花，他可能不確定兩種啤酒花對於同樣的沸騰時間的變化反應是相同的。結合第一因子(沸騰時間)的兩個水準和第二因子(啤酒花品種)的兩個水準，交互產生了四個不同處理的組合，即個析因實驗的單次複製(見表 14.6)。在階乘實驗中的實驗變量可以是量性的(例如，煮沸時間)或質性的(例如：啤酒花品種)。量性和質性因子的任意組合可以在相同的階乘實驗中運用。

一個因子的「效應」是從因子的水準變化而產生的反應變化(或差異)。各別因子的效應稱為主效應。例如，如果在表 14.6 中的內容項所代表的是四個啤酒樣品中平均啤酒花香氣的評價，由沸騰時間造成的主效應是：

$$\frac{(T_{1A}-T_{2A})+(T_{1B}-T_{2B})}{2}$$

(14.12)



PLS1 (55%消費者；53%感官)

PLS2 (25%消費者；18%感官)

圖 14.22

PLS 地圖表示區隔的正面驅向和負面驅向。圖 (a) 說明區隔 1 偏好許多大和堅實的肉感、馬鈴薯和蔬菜片，以及高強度的肉塊感、青菜、胡蘿蔔、洋蔥和芹菜的味道，以及低強度的非天然肉味、香精。圖 (b) 說明區隔 2 喜歡很多大且堅實的馬鈴薯、清澈且低粘度的醬料，高強度的肉塊感、低油性及低辣度(尤其是黑胡椒)。

缺圖

圖 14.22 (續)

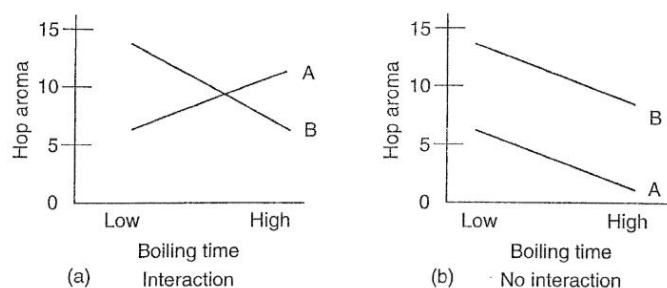
同樣，由啤酒花品種造成的主效應是

$$\frac{(T_{1A}-T_{1B})+(T_{2A}-T_{2B})}{2} \quad (14.13)$$

在一些研究中，一個因子的效應取決於第二因子的水準。發生這種情況時，是因為這兩個因素之間的相互作用所導致。假設使用啤酒花品種 A 的啤酒隨著沸騰時間增加而提高了啤酒花香氣的等級，但對於相同沸騰時間的變化時，啤酒花品種 B (見表 14.6) 的啤酒卻因此而降低啤酒花香氣。沸騰時間和啤酒花品種之間的相互作用，因為沸騰時間的效果取決於使用的啤酒花品種。

表 14.6 階乘處理結構的兩個因素分別為含有二個級別

		啤酒花品種	
		A	B
沸騰時間	低 (1)	$T_{1A} = 6$	$T_{1B} = 13$
	高 (2)	$T_{2A} = 12$	$T_{2B} = 7$



Hop aroma 啤酒花香味

(a) boiling time interaction 煮沸時間相互作用

(b) boiling time no interaction 煮沸時間無相互作用

圖 14.23 平均啤酒花香氣回應的曲線顯示研究的因子之間(a)有相互作用和(b)沒有相互作用的

圖可以被用來說明相互作用。圖 14.2 a 顯示出沸騰時間和啤酒花品種之間的相互作用。圖上的點是在表 14.6 呈現的四個實驗條件下平均啤酒花香氣的評分。這兩個因素之間的相互作用是由兩線之間缺乏平行關係所指出。如果兩個因素之間沒有相互作用，線條將是幾乎平行的（僅由於實驗誤差而偏離），如圖

14.23b。相互作用存在時研究人員必須在解釋主效應上非常謹慎。表 14.6 的數據說明了「沸騰時間 x 啤酒花品種」的互動。方程式 14.12 是估計由沸騰時間所導致的主效應，帶入後得到 $(6 - 12) + (13 - 7) / 2 = 0$ ，表示沸騰時間沒有任何影響。然而，圖 14.23a 清楚地表明，對於啤酒花品種對於沸騰時間有相當的影響。因為單獨的品種效果是相反的，它們在計算中互相抵消由於沸騰時間的主效應。在相互作用的情況下，一個因子的效應只能透過固定第二因子的水準的影響才有意義。

研究人員有時會使用階乘處理結構的替代方案，稱為一次一個處理結構 (one-at-a-time treatment structures)，錯誤地認為他們節省了研究程序。假設在啤酒釀造的例子中，該啤酒已經只準備了三種樣品：低沸騰時間/品種 A，標示 T_{1A} ，低沸騰時間/品種 B，標示 T_{1B} ，和高沸騰時間/品種 B，標示 T_{2B} （省略高沸騰時間/品種 A， T_{2A} ）。因為只有準備三種樣品，一次一個方法似乎比起全階乘方法更經濟。然而，如果考慮到主效應的估計值的精度，這是不正確的。只有一個不同點，由於沸騰時間導致的主效應在一次一個的研究中只剩一個差異（即， $T_{1B} - T_{2B}$ ）。同樣的情況發生在品種上（即， $T_{1A} - T_{1B}$ ）。方程 14.12 和公式 14.13 表明，對於階乘處理結構，兩個差異值才可用於估計一個效果。整個一次一個實驗不得不被複製兩次而得到 6 個實驗點，以得到如階乘實驗中同樣精確地從四個點獲得的那些主效應的估計。

階乘處理結構的另一個優點是，比起一次一個實驗，具有檢測相互作用的能力。如果高溫/品種 A 觀察值 $T_{2A} = 12$ 從表 14.6 的數據忽略（如同在一次一個的研究中），人們會觀察到啤酒在高沸騰時間釀製的啤酒花香氣小於低沸騰時間的啤酒，而且啤酒花品種 A 的啤酒花香氣少於啤酒花品種 B。最明顯的結論是，高沸騰時間的啤酒使用啤酒花品種 A 的組合有最少的啤酒花香氣。表 14.6 的完整的數據，且圖 14.23a 的相互作用圖顯示這將是一個不正確的結論。

表 14.7 階層實驗的 ANOVA 表

變異來源	自由度	平方和	均方	F
總共	$rab-1$	SS_T		
A	$a-1$	SS_A	$MS_A = SS_A/(a-1)$	$F_A = MS_A/MS_E$
B	$b-1$	SS_B	$MS_B = SS_B/(b-1)$	$F_B = MS_B/MS_E$
AB	$df_{AB} = (a-1)(b-1)$	SS_{AB}	$MS_{AB} = SS_{AB}/df_{AB}$	$F_{AB} = MS_{AB}/MS_E$
誤差	$df_E = ab(r-1)$	SS_E	$MS_E = SS_E/df_E$	

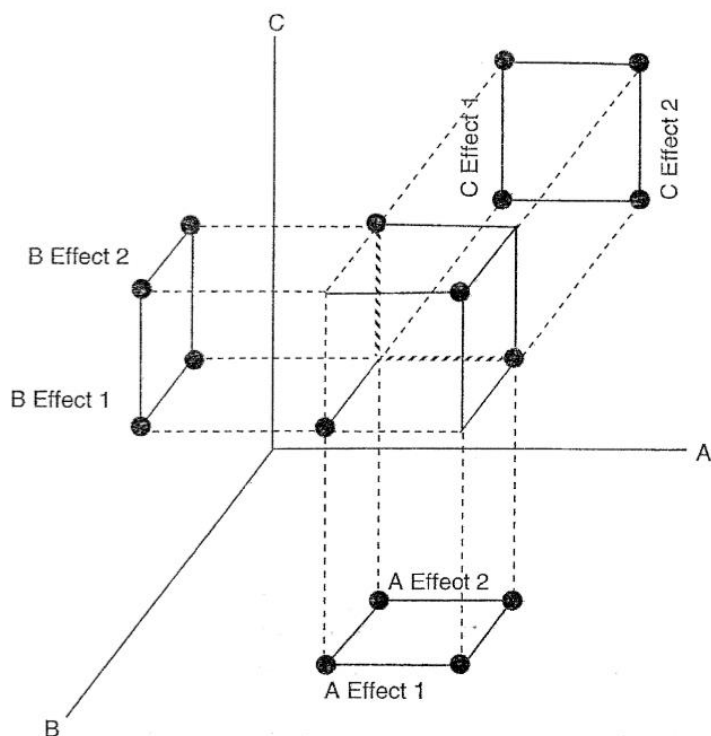
注：因子 A 具有“a”的水準，因子 B 具有“b”的水準，並且整個實驗被複製“r”次。樣品按照完全隨機區集結構製備。

在感官評估中的應用階乘處理結構所建議的步驟如下。準備至少兩次獨立重複的全面階乘實驗。使用如第 13 章所描述的適當的區集結構提交樣品給品評小組進行評價。從品評小組的品評數據中取均值，並利用它們作為 ANOVA 的原始數

據。ANOVA 的輸出包括實驗因子的主效應和相互作用（見表 14.7）的測試。此過程可避免混淆測量誤差與實驗誤差，前者是從品評小組數據的分析中得到的，後者是只可能從獨立地複製處理組合之間的差異所獲得。

14.4.2 部分階乘和篩選研究

在研究計畫的初期，許多變量被提出為可能在重要反應上具有有意義的效果。要執行一個有效的研究計畫，實驗者需要一種能讓他們篩選出與那些對反應很少或沒有影響的重要變量的方法。這必須確定並以最小的工作量執行，從而在計畫完成前有足夠的資源對最終原型品做必要的微調和「精化」的工作。有一類實驗計畫稱為部分階乘，允許研究者篩選許多變量圖同時使用最少數量的實驗樣品。當階乘實驗中的實驗變量的數量成長時，每個主效應的估計因越來越多的「隱藏複製」而增加。例如，如同前一節所指出，在一個 2×2 （或 2^2 ）階乘中，各主效應是由兩個差異（即，兩個隱藏重複）所估計。在 2^6 階乘中（即六個因子，各有兩個水準），用來估算各個主效應的隱藏重複的數量已發展到 32。這可能是過度的。 2^6 階乘的單次複製包括 64 個實驗樣品。如果研究關注主要集中在識別具有顯著主效應的個別實驗變量，則隱藏重複的數目可以安全地減少到 16 或甚至 8 個而沒有過份地犧牲了敏感度。實驗樣品的數數將可同時減少到 32 或 16，從而產生一個可管理的實驗。圖 14.24 表明樣品在一個 2^3 階乘的數量可被切成一半，從八個至四個，而仍同時提供了用於估算每個主效應的兩個差異。



Effect 效應

圖 14.24 2^3 個實驗的 1/2 重複部分階乘的圖形透過投影顯示即使總實驗已經從

八個減少至四個樣品，仍保留估計每個主效應的兩個差異。

14.4.2.1 建構部分階乘

大多數篩選研究通過選擇每個實驗變量的兩個層次，高和低，來進行。高和低層次的各種處理的組合構成了實驗設計。也就是說，處理的組合定義了應被用來產生每個實驗樣品的實驗變量的等級。為了識別每個處理的因素水準組合而創建了一種方便的符號。A 變量的高水準由其小寫 a 表示，B 變量的高水準則以 b 表示，以此類推。而一個變量的低水準則由缺少其小寫字母表示。例如，在 2^3 階層中，處理組合高-A、高-B、高-C 將被表示為 abc；處理組合低-A、高-B、高-C 將被表示為 bc；處理組合低-A、低-B、高-C 將表示為 c。用來表示在其低水平的所有因素的綜合標誌是 (I)。

八個處理組合成 2^3 階乘實驗的單一複製在表 14.8 的第一直欄中呈現。其餘直欄包含將被用於估計每個階乘效應的係數符號（係數是-1 或+1；因此只需要符號即可）。由係數符號所組成的處理組合用來估算三向相互作用 ABC。以這種方式形成的兩個小組是 2^3 階乘的每個 1/2-重複。此兩種四個處理組合都可以選擇用在篩選研究中。Cochran 和 Cox (1957) 介紹了在 2^n 和 3^n 實驗的部分階乘計畫，其中因子的數量，n，至多為八。

表 14.8 在 2^3 階乘測試中的因子效應，如同部分因子實驗 1/2 重複的排列 (ABC+ 和 ABC-)

處理 組合	階乘效應						
	A	B	C	AB	AC	BC	ABC
a	+	-	-	-	-	+	+
b	-	+	-	-	+	-	-
c	-	-	+	+	-	-	-
abc	+	+	+	+	+	+	+
(i)	-	-	-	+	+	+	-
ab	+	+	-	+	-	-	-
ac	+	-	+	-	+	-	-
bc	-	+	+	-	-	+	-

透過選擇具有在三向相互作用中的係數相同符號的處理組合，任何估計這種效應的符度的能力都被犧牲了。ABC 被稱為定義對照 (defining contrast)，因為它是被用來區分階乘成兩個 1/2-重複的標準。

注意在表 14.8 中，每個組內有兩個+的和兩個-以估計任何效應。這些是即使在只運行全階乘的一半時，仍然存在的隱藏複製。假設選擇四個處理組構成第一組並運行（即，ABC+組）。然後，變量 A 的主效應會（除了 2 的除數）透過估算

$$A = abc + a - b - c$$

然而，在雙向互動 BC 的估計也是

$$BC = abc + a - b - c$$

A 的主效應被化名為是 BC 的雙向相互作用，表示為 $A = BC$ 。同樣地， $B = AC$ 和 $C = AB$ 。在實際應用中，如果兩個階乘效果已有化名了，那麼就不可能將其相關反應與個別影響分開。A 的表面效果可能是來自 A 或來自 BC，或者甚至可能是由兩者的組合所導致。

將主效應與相互作用化名化的代價是分裂階乘實驗。通常情況下，主效應的幅度是假定比相互作用的量值更大，在這種情況下，部分階乘可以安全地用於篩選重要的實驗變量。但是，如果大型交互效應存在，那麼研究人員可能會被誤導而得出一個變量對於一個反應具有重要影響的結論，而事實上它並沒有。這份告誡並不是打算嚇唬研究人員使其不要使用部分階乘，而只是讓他們意識到這個問題，因為它可能可以用來解釋研究計畫的進展中出現的不協調的結果。

14.4.2.2 Plackett-Burman 實驗

部分階乘並不是唯一可用於篩選有影響的變量的方案。Plackett-Burman (1946) 的實驗在樣品數量的需求上是更為經濟的。在 Plackett-Burman 實驗中，樣本的數目始終是 4 的倍數。可使用 Plackett-Burman 實驗篩選的實驗變量數量，最多只能低於樣品一個（即，4、5、6 或 7 變量是可以用 8 個樣品篩選的數量；8、9、10，或 11 的變量可以用 12 個樣品進行篩選；等等）Box 和 Draper (1987) 呈現的 Plackett-Burman 實驗架橫跨了從 4 個至 27 個實驗因子的範圍（即研究 8-28 個實驗樣品）。

14.4.2.3 篩選研究分析

無論部分階乘和 Plackett-Burman 實驗都可以藉由 ANOVA 分析。然而，由於涉及少數的樣本，有時不可能計算 F-比率來測試效果的顯著性。這是因為，在一些篩選實驗中，沒有可用的自由度來估計實驗誤差。無論如何，即使能夠進行 ANOVA 計算，測試也不是很敏感，所以缺少真正效應的可能性（即，II 型誤差）相對較高。

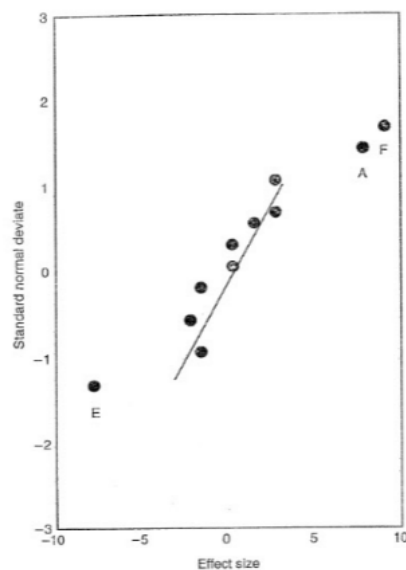
分析篩選實驗的圖形技術使得研究人員得以更投入在決定哪些變量影響了反應。此技術驅動的邏輯是，如果沒有一個變量對反應有影響，那麼他們的估計效應的值實際上是一個平均為零的分配（假設為常態分配）中的隨機觀測值。如果這些估計效應是針對其相應的正常隨機偏離值作圖，他們應該會形成一條直線（在缺焚任何真正的效應下）。然而，如果一些變量影響反應，那麼所估計的效應就不只是隨機觀察。真正的效應將掉落在圖中的直線上，可能是在較高和右處（正效應），或較低且左處（負效應）。研究者可以檢查估計效應的「常態機率圖」，如在圖 14.25 所介紹的，以決定哪些變量實際影響反應。

建構常態機率圖有四個步驟：

1. 使用 ANOVA 估計實驗變量的影響。
2. 依照 $i = 1$ 到 n 的順序遞增排名估計的反應。
3. 將有順序的估計與新變量 $z = \Phi^{-1}[P]$ 配對，其中 $p = \frac{(3i-1)}{(3n+1)}$ ，且 Φ^{-1} 是標準常態分佈函數的反函數。許多統計電腦軟體組包含用於使用 p 值計算 z 值（有時稱為“PROBIT”）的功能。
4. 繪製 z 與估計效應的關係，藉由使用標準繪圖程序擬合（由眼睛）數據為直線，並查找落於高且右，或者低且左的點。這些點被識別為「顯著」變量。

14.4.3 反應曲面法

處理結構如 RSM 基本上是一個設計的迴歸分析（見 Montgomery 1976; Giovanni 1983）。不像階乘處理結構，其目標是確定因子是否（以及如何）影響反應，RSM 實驗的目標是基於實驗因子的控制值（稱為獨立變量）預測反應變量的值（稱為依賴變量），在所有的 RSM 實驗中，因子必須量性的。



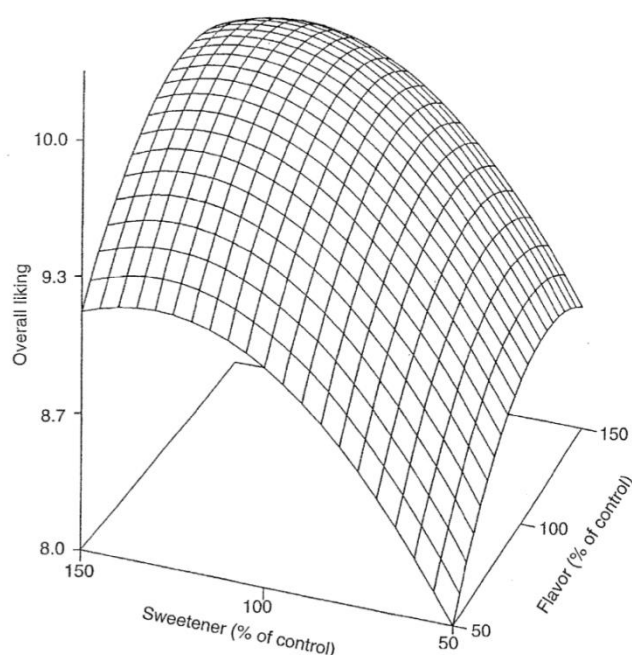
Effect size 效應大小 standard normal deviate 標準常態偏離

圖 14.25 常態機率圖顯示了「不顯著」階乘效果落在線上，而「顯著」的效應在線的較高且右處，或是較低且左處。

RSM 處理結構提供一種經濟的方式，在依賴變量的值的範圍內，預測一個或多個反應的值。一組樣品（即，試驗點）根據選定的 RSM 處理結構中指定的條件製備。樣品透過一個感官小組分析，將得到的平均反應提交給逐步迴歸分析。該分析產生了由（多個）反應連結到獨立變量的值的預測方程式。預測方程式可以

使用圖形描繪成如圖 14.26 的反應曲面圖。另外，所預測的關係可以顯示成如圖 14.27 的「等高線圖」。等高線圖很容易理解。他們允許研究者確定在實驗區域的任何點處的反應的預測值，而不需要在該處準備樣品。

有數類處理結構可作為 RSM 實驗。最廣泛的一類，在此處所討論的，是非常相似於階乘實驗的。該計劃中的一部分是由高和低層次的獨立變量的所有可能的組合所組成。(在兩個因子的 RSM 實驗中，這部分由四個點所組成：[低，低]，[低，高]，[高，低]和[高，高]) RSM 實驗的階乘部分是由中心點增強（即，採取所有的因子的平均值，[低+高]/2)。典型地，該中心點被複製多次（不少於三）以提供實驗誤差的獨立估計（見圖 14.28)。在 RSM 實驗的常規做法是分配所有低水準的因子-1 的編碼；高水準的因子分配+1 的編碼；中心點則分配編碼值 0。



Overall liking 整體喜好度

Sweetener(% of control) 甜度（控制百分比）

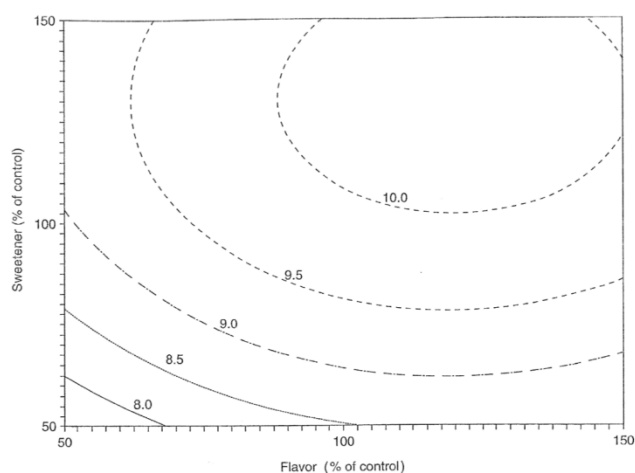
Flavor(% of control) 風味（控制百分比）

圖 14.26 反應曲面圖顯示了產品整體喜好度和甜度含量及味道之間的預測關係。

圖 14.28 中所繪的 RSM 實驗的處理結構被稱為一階的 RSM 實驗(first-order RSM experiment)。可以通過處理結構來擬合的完整迴歸方程式具有以下形式：

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k \quad (14.14)$$

其中 β_i 是要估計的迴歸方程式的係數，且 x_i 是在實驗中第 k 個因子的編碼水準。一階 RSM 實驗用於識別總體趨勢，並且用來確定獨立變量的正確範圍是否被選定。一階模型用於研究計劃的早期，以確定獨立變量的水準剛往哪個方向移動，以影響依賴變量至期望的改變（例如，增加期望反應或減少不想要的反應）。如果依賴變量和獨立變量之間存在複雜關係，一階模型可能無法充分預測反應。這種情況下需要使用二階 RSM 處理結構。



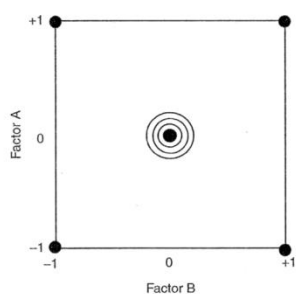
Sweetener(% of control) 甜度（控制百分比）

Flavor(% of control) 風味（控制百分比）

圖 14. 27 產品總體喜愛程度和甜味劑水準以及風味之間的預測關係的等高線圖。等高線圖提供了產品的敏感度的量性評估，當成分發生水準的變化時。

可以使用二階 RSM 處理結構擬合的完整的迴歸模型具有如下形式：

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \beta_{11} x_1^2 + \beta_{22} x_2^2 + \cdots + \beta_{kk} x_k^2 + \beta_{12} x_1 x_2 + \beta_{13} x_1 x_3 + \cdots + \beta_{k-1,k} x_{k-1} x_k \quad (14.15)$$



Factor A 因子 A

Factor B 因子 B

圖 14.28 兩因子的一階 RSM 實驗。此圖說明了在方程式 14.14 中的一階迴歸模型的估計，在兩個獨立的變量的 RSM 實驗對於中心及兩個因子的安排。

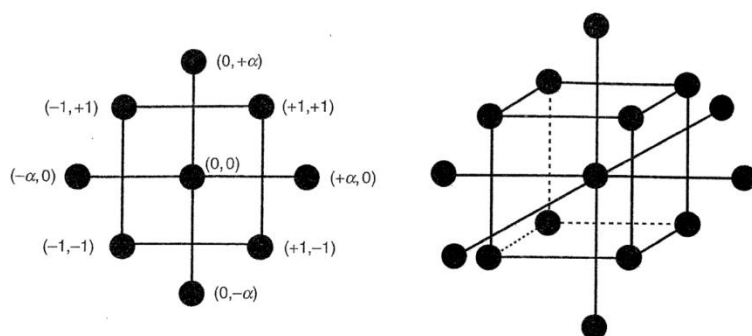


圖 14.29 中央複合 RSM 實驗。圖中顯示出具有兩個和三個獨立變量，允許如等式 14.15 二階迴歸模型的估計，的 RSM 實驗的軸向、階乘、中心點的安排。

模型中加入的平方項和交叉乘積項，允許預測反應曲面「彎曲」和「柔軟」，因而可以就複雜的關係改進預測。

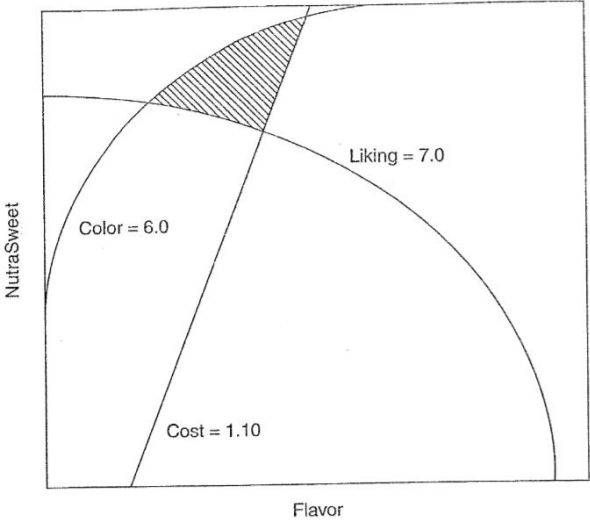
二階 RSM 實驗其中一個流行的分類是中央複合，可旋轉的處理結構。中央複合實驗藉由在一階 RSM 處理結構中添加一組軸向或「起始」點（見圖 14.29）而建立。在 k -因子的 RSM 實驗中存在 $2k$ 的軸向點。一般使用 $-1, 0, +1$ 為因子水準編碼，軸向點為 $(\pm\alpha, 0, \dots, 0), (0, \pm\alpha, 0, \dots, 0), \dots, (0, 0, \dots, 0, \pm\alpha)$ ，

其中 α 是從軸向點到實驗區的中心（即，中心點）的距離。 α 的值是 $(F)^{1/4}$ ，其中 F 是在一階實驗中的非中心階乘點的數目。例如，在一個雙因子實驗中， $F = 4$ ，而 $\alpha = (4)^{1/4} = 1.414$ 。

二階 RSM 模型有優於一階模型的幾個優點。如前面所提到的，二階模型能夠更好地擬合依賴變量和獨立變量之間的複雜關係。此外，二階模型可以用於定位在依賴變量的水準方面的反應的預測最大值或最小值。

執行 RSM 實驗建議的步驟如下（見 Carr 1989）：首先，實驗樣品應根據 RSM 計劃製備。其次，在 RSM 處理結構中的樣品上執行常規的 BIB 分析，忽略樣品之間的關聯。（建議使用 BIB 區集結構，因為通常在 RSM 實驗中有過多的樣品需要進行評價。然而，如果能夠同時評估所有的樣品，則可以使用隨機[完整]區集設計）從 BIB 分析中得到的唯一關注的輸出是一組調整樣品均值。整體測試統計量的顯著（或不顯著）不是所關注的。接著，提交樣品平均至一個逐步迴歸分析以建立和實驗因子的水準有關反應的值的預測方程式。接著使用預測方式產生等高線圖，它提供關於反應的因子效應的描繪圖形。如果只有一個反應，其中該反應具

有可接受的值（或達到的最小值或最大值）的區域在等高線圖上顯而易見。當數個反應同時被考慮過時，即為個別等高線圖中的覆蓋區域。所有反應的可接受值的區域希望可以如同圖 14.30 中被識別出。



Nutrasweet 代糖

Flavor 風味

Color 顏色

Cost 成本

Liking 喜好度

圖 14.30 幾個反應變量的關鍵限制所覆蓋的等高線圖，顯示預測此配方水準為同時滿足所有限制區域。

第十五章 統計表

15 Statistical Tables

表 15.1 數字 1-9 的隨機順序：以三欄為一組排列。

說明

(1) 為了產生三位數的隨機數(亂數)的序列，在任意位置進入該表中，例如，閉上眼睛指向任一點。不檢查號碼的向上或向下移動，並記錄數組數字。丟棄不適合的任何數字（超出範圍、之前用過等等）。以這種方式得到的數列是隨機順序。

(2) 為了產生二位數的隨機數序列，如步驟(1)所示，但首先得先決定，該選取表中每個號碼的前兩位或後兩位數字，例如，拋硬幣。與三位數字的處理採相

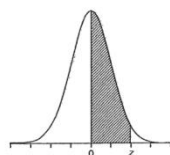
同的方式，即，如果兩位數字出現一次以上，只保留第一次出現的，丟棄第二次出現的數字。

(3) 隨機數表對於問題像是：「將 15 至 50 以隨機順序放置」是不實用的。若要處理這種問題，將每個數字寫在卡片上，並將所有卡片放在袋子裡蒙眼抽取，或使用電腦中的亂數產生器，如從 SAS.®PROC PLAN

862	245	458	396	522	498	298	665	635	665	113	917	365	332	896	314	688	468	663	712	585	351	847
223	398	183	765	138	369	163	743	593	252	581	355	542	691	537	222	746	636	478	368	949	797	295
756	954	266	174	496	133	759	488	854	187	228	824	881	549	759	169	122	919	946	293	874	289	452
544	537	522	459	984	585	946	127	711	549	445	793	734	855	121	885	595	152	237	574	611	145	784
681	829	614	547	869	742	822	554	448	813	976	688	959	714	912	646	873	397	159	155	136	463	363
199	113	941	933	375	651	414	891	129	938	862	572	698	128	363	478	214	841	314	437	792	874	926
918	481	797	621	743	827	377	916	966	429	657	246	423	277	685	533	937	223	582	946	323	626	519
335	662	875	282	617	274	635	379	287	791	334	139	117	963	448	957	451	585	821	829	267	512	638
477	776	339	818	251	916	581	232	372	374	799	461	276	486	274	791	369	774	795	681	458	938	171
653	489	538	216	446	849	914	337	993	459	325	614	771	244	429	874	557	119	122	417	882	714	769
749	824	721	967	287	556	628	843	725	731	553	253	183	653	988	431	788	426	875	838	457	927	475
522	967	259	532	618	624	396	562	134	563	932	441	834	787	231	958	232	537	439	956	531	345	352
475	172	986	859	925	932	282	924	842	642	797	565	399	896	596	282	441	784	258	684	625	662	291
894	333	612	728	869	487	741	259	476	127	286	736	257	168	847	316	969	692	786	549	949	559	526
116	218	464	191	132	218	573	786	258	296	471	372	618	935	353	747	123	863	644	161	793	196	847
381	641	393	375	354	193	165	615	587	384	187	965	572	112	695	615	941	361	375	376	871		633
968	755	847	643	773	765	439	478	611	978	868	898	546	319	775	169	896	275	513	222	114	233	184
742	421	226	286	522	618	471	218	397	745	461	477	478	535	957	674	132	228	442	225	444	171	151
859	878	392	311	659	772	935	447	834	117	658	161	754	654	176	883	855	195	637	751	586	948	513
964	593	137	574	288	994	582	961	746	336	983	782	611	988	833	265	969	584	564	683	197	214	326
177	636	674	897	167	157	856	524	662	598	145	926	362	777	415	931	313	317	195	137	959	536	985
228	755	915	955	946	233	647	653	425	674	719	543	549	826	669	429	576	773	756	392	632	725	879
591	214	851	669	394	349	299	192	179	264	332	294	896	299	782	397	791	659	921	569	811	683	762
636	167	789	438	413	565	118	889	253	452	577	859	125	141	241	746	444	841	313	446	225	362	248
415	982	543	743	835	826	364	776	988	923	224	615	283	462	328	512	228	466	278	874	373	499	437
383	349	468	122	771	481	723	335	511	889	896	338	937	313	594	158	687	932	889	918	768	857	694

資料來源：W. G. Cochran 和 G. M. Cox. 1957。試驗設計，John Wiley 父子，紐約。

表 15.2 標準常態分佈

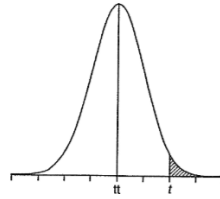


說明：見第十三章的範例

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.000	0.00	0.00	0.012	0.01	0.01	0.023	0.027	0.03	0.035
0.1	0.03	0.04	0.04	0.05	0.055	0.059	0.063	0.067	0.07	0.07
0.2	0.07	0.08	0.087	0.091	0.094	0.098	0.10	0.10	0.110	0.11
0.3	0.117	0.12	0.12	0.12	0.13	0.13	0.14	0.14	0.14	0.1
0.4	0.15	0.15	0.16	0.166	0.17	0.17	0.17	0.18	0.184	0.18
0.5	0.19	0.19	0.19	0.20	0.205	0.208	0.21	0.21	0.21	0.22
0.6	0.22	0.22	0.232	0.235	0.23	0.242	0.24	0.248	0.25	0.25
0.7	0.25	0.261	0.26	0.267	0.270	0.273	0.27	0.279	0.282	0.28
0.8	0.28	0.29	0.293	0.296	0.29	0.302	0.30	0.307	0.31	0.31
0.9	0.31	0.31	0.32	0.323	0.326	0.328	0.33	0.334	0.336	0.338
1.0	0.34	0.343	0.346	0.348	0.35	0.35	0.355	0.35	0.35	0.36
1.1	0.36	0.36	0.368	0.3708	0.372	0.374	0.37	0.379	0.381	0.383
1.2	0.38	0.38	0.38	0.390	0.39	0.39	0.39	0.39	0.399	0.40
1.3	0.403	0.404	0.406	0.408	0.40	0.41	0.41	0.41	0.416	0.41
1.4	0.41	0.420	0.422	0.423	0.425	0.426	0.427	0.429	-0.43	0.43

1.5	2	0.433 5	0.434 7	0.435 0	0.437 2	0.438 4	0.439 06	0.44 18	0.44 9	0.442 41	0.44
1.6	2	0.445 63	0.44 4	0.447 84	0.44 5	0.449 5	0.450 15	0.45 5	0.452 5	0.453 45	0.45
1.7	4	0.455 64	0.45 3	0.457 82	0.45 1	0.459 9	0.459 08	0.46 16	0.46 5	0.462 33	0.46
1.8	41	0.46 9	0.464 6	0.465 4	0.466 71	0.46 8	0.467 6	0.468 93	0.46 9	0.469 6	0.470
1.9	13	0.47 9	0.471 26	0.47 2	0.473 8	0.473 4	0.474 0	0.475 6	0.475 1	0.476 67	0.47
2.0	2	0.477 78	0.47 3	0.478 8	0.478 3	0.479 8	0.479 03	0.48 8	0.480 12	0.48 17	0.48
2.1	21	0.48 26	0.48 0	0.483 4	0.483 38	0.48 2	0.484 6	0.484 50	0.48 4	0.485 57	0.48
2.2	61	0.48 4	0.486 8	0.486 1	0.487 5	0.487 8	0.487 81	0.48 4	0.488 7	0.488 90	0.48
2.3	93	0.48 96	0.48 8	0.489 1	0.490 4	0.490 06	0.49 9	0.490 1	0.491 13	0.49 16	0.49
2.4	18	0.49 0	0.492 2	0.492 5	0.492 27	0.49 9	0.492 931	0.4 2	0.493 34	0.49 36	0.49
2.5	38	0.49 0	0.494 1	0.494 3	0.494 5	0.494 6	0.494 48	0.49 9	0.494 1	0.495 52	0.49
2.6	53	0.49 5	0.495 6	0.495 7	0.495 59	0.49 0	0.496 61	0.49 2	0.496 3	0.496 64	0.49
2.7	5	0.496 6	0.496 7	0.496 8	0.496 9	0.496 0	0.497 71	0.49 2	0.497 3	0.497 74	0.49
2.8	4	0.497 75	0.49 6	0.497 7	0.497 7	0.497 8	0.497 79	0.49 9	0.497 0	0.498 81	0.49
2.9	1	0.498 2	0.498 2	0.498 83	0.49 4	0.498 4	0.498 85	0.49 85	0.49 6	0.498 6	0.498
3.0	7	0.498 87	0.49 7	0.498 8	0.498 8	0.498 9	0.498 89	0.49 9	0.498 0	0.499 90	0.49

表 15.3 學生 t-分佈（表格內容為 $t_{\alpha,v}$ ）上 α - 機率點



說明：(1) 查看誤差自由度(v)相對應的行

(2) 查看預定的 α 值出現的欄，接著挑出相對應的 t 值

V	α						
	0.25	0.10	0.05	0.025	0.01	0.005	0.0005
1	1.000	3.078	6.314	12.706	31.821	63.657	636.619
2	0.816	1.886	2.920	4.303	6.965	9.925	31.598
3	0.765	1.638	2.353	3.182	4.541	5.841	12.941
4	0.741	1.533	2.132	2.776	3.747	4.604	8.610
5	0.727	1.476	2.015	2.571	3.365	4.032	6.859
6	0.718	1.440	1.943	2.447	3.143	3.707	5.959
7	0.711	1.415	1.895	2.365	2.998	3.499	5.405
8	0.706	1.397	1.860	2.306	2.896	3.355	5.041
9	0.703	1.383	1.833	2.262	2.821	3.250	4.781
10	0.700	1.372	1.812	2.228	2.764	3.169	4.587
11	0.697	1.363	1.796	2.201	2.718	3.106	4.437
12	0.695	1.356	1.782	2.179	2.681	3.055	4.318
13	0.694	1.350	1.771	2.160	2.650	3.012	4.221
14	0.692	1.345	1.761	2.145	2.624	2.977	4.140
15	0.691	1.341	1.753	2.131	2.602	2.947	4.073
16	0.690	1.337	1.746	2.120	2.583	2.921	4.015
17	0.689	1.333	1.740	2.110	2.567	2.898	3.965
18	0.688	1.330	1.734	2.101	2.552	2.878	3.922
19	0.688	1.328	1.729	2.093	2.539	2.861	3.883
20	0.687	1.325	1.725	2.086	2.528	2.845	3.850
21	0.686	1.323	1.721	2.080	2.518	2.831	3.819
22	0.686	1.321	1.717	2.074	2.508	2.819	3.792
23	0.685	1.319	1.714	2.069	2.500	2.807	3.767
24	0.685	1.318	1.711	2.064	2.492	2.797	3.745
25	0.684	1.316	1.708	2.060	2.485	2.787	3.725
26	0.684	1.315	1.706	2.056	2.479	2.779	3.707
27	0.684	1.314	1.703	2.052	2.473	2.771	3.690
28	0.683	1.313	1.701	2.048	2.467	2.763	3.674
29	0.683	1.311	1.699	2.045	2.462	2.756	3.659

30	0.683	1.310	1.697	2.042	2.457	2.750	3.646
∞	0.674	1.282	1.645	1.960	2.326	2.576	3.291

表 15.4 學生等級的百分點：對於 Tukey 的 HSD 多重比較法的上－ α 臨界值

說明：(1)查看預定的 α 值相對應的表

(2)查看 ANOVA 誤差自由度相對應的列，

(3)查看處理數目相對應的欄位，接著挑出相對應的 q 值

df	df																			
	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	2
表格內容為 $q_{0.01}$ 當 $P(q < q_{0.01}) = 0.99$																				
3	90	1					2		2		2	2	2	2		2		2		
4	35.0	164.3	185.6	202.2	215.8		27.2	237.0	45.6	253.2	260.0		66.2	71.8	77.0	281.8	286.3	90.4	294.3	290.0
5	14	1					2		3		3	3	3	3		3		3		
6	9.02	22.29	24.72	26.63	28.29		9.53	30.68	1.69	32.59	33.40		4.13	4.81	5.43	36.00	36.53	7.03	37.50	37.95
7	8.2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
8	0.62	2.17	3.33	4.24	5.00	5.64	6.20	6.69	7.13	7.53	7.89	8.22	8.52	8.81	9.07	9.32	9.55	9.77		
9	6.	8	9		1	1	1		1	1	1	1	1	1	1	1	1	1	1	1
10	.12	.17	9.96		0.58	1.10	1.55	11.93	2.27	2.57	2.84	3.09	3.32	3.53	3.73	3.91	4.08	4.24	4.40	
11																				
12	5.	6.98	7.80	8.42		8	9	9	9	1	1	1	10.89	11.08		1	11.40	11.55	11.68	1
13	70				.91	.32	.67	.97	0.24	0.48	0.70		1.24					1.81		11.93

[illegible]

6	13	4. 4.79	.19	5.49	.72	5 .92	5 .08	6 6.22	6.35	6.46	6.56	6.66	6.74	.82	6 6.90	.97	6 .03	7 7.09	.15	7
7	0	4.1 4.74	.14	5 .43	5 .66	5 .85	5 .01	6 .15	6 .27	6 .38	6 6.48	.57	6 6.66	.73	6 .81	6 .87	6 .94	6 7.00	.05	7.
8	7	4.0 4.70	5.09	.38	5 .60	5 .79	5 .94	5 6.08	6.20	.31	6 .41	6 6.50	6.58	.65	6 .73	6 6.79	.85	6 .91	6 97	6.
9	4.05	4.67	.05	5 .33	5 .55	5 .73	5 .89	5 6.02	.14	6 .25	6 .34	6 .43	6 .51	6 .58	6 6.65	6.72	6.78	.84	6 6.89	
20	2	4.0 4.64	.02	5 5.29	.51	5 5.69	.84	5 .97	5 6.09	.19	6 6.28	.37	6 6.45	.52	6 6.59	6.65	.71	6 .77	6 6.82	
24	3.96	4.55	4.91	5.17	5.37	5.54	5.69	5.81	5.92	6.02	6.11	6.19	6.26	6.33	6.39	6.45	6.51	6.56	6.61	
30	3.89	4.45	4.80	5.05	5.24	5.40	5.54	5.65	5.76	5.8 5	5.9 3	6.01	6.08	6.14	6.20	6.26	6.31	6.36	6.41	
40	3.82	4.37	4.70	4.9 3	5.11	5.26	5.3 9	5.50	5.60	5.69	5.76	5.83	5.90	5.96	6.02	6.07	6.12	6.16	6.21	
60	3.76	4.28	4.59	4.82	4.99	5.13	5.25	5.36	5.45	5.53	5.60	5.67	5.73	5.78	5.84	5.89	5.93	5.97	6.01	
120	3.70	4.20	4.50	4.71	4.87	5.0 1	5.1 2	5.21	5.30	5.37	5.44	5.50	5.56	5.61	5.66	5.71	5.75	5.79	5.83	
∞	3.64	4.12	4.40	4.60	4.76	4.88	4.99	5.08	5.16	5.23	5.29	5.35	5.40	5.45	4.49	5.54	5.57	5.61	5.65	

表格內容為 $q_{0.05}$ 當 $P(q < q_{0.05}) = 0.95$

1	1	2	3	3	4	4	4	4	4	5	5	5	5	5	5	5	5	5	5
	7.97	6.98	2.82	7.08	0.41	3.12	5.40	7.36	9.07	0.59	1.96	3.20	4.33	5.36	6.32	7.22	8.04	8.83	9.56
2	6	8	9	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	.08	.33	.80	0.88	1.74	2.44	3.03	3.54	3.99	4.39	4.75	5.08	5.38	5.65	5.91	6.14	6.37	6.57	6.77
3	4	5	6	7	8	8	8	9	9	9	9	1	1	1	1	1	1	1	1
	.50	.91	.82	.50	.04	.48	.85	.18	.46	.72	.95	0.15	0.35	0.53	0.69	0.84	0.98	1.11	1.24
4	3	5	5	6	6	7	7	7	7	8	8	8	8	8	8	8	9	9	9
	.93	.04	.76	.29	.71	.05	.35	.60	.83	.03	.21	.37	.82	.66	.79	.91	.03	.13	.23
5	3	4	5	5	6	6	6	5	6	7.	7	7	7	7	7	7	7	8	8.
	.64	.60	.22	.67	.03	.33	.58	.80	.99	17	.32	.47	.60	.72	.83	.93	.03	12	.21
6	3	4	4	6	5	5	6	6	4	6	6	6	7	7.	7	7	7	7	7
	.46	.34	.90	.30	.63	.90	.12	.32	.49	.65	.79	.92	.03	14	.24	.34	.43	.51	.59
7	3	4	4	5	5	5	5	6	6	6	6	6	6	6	6	6	7	7	7
	.34	.16	.68	.06	.36	.61	.82	.00	.16	.30	.43	.55	.66	.76	.85	.94	.02	.10	.17
8	3	4	4	4	5	5	5	5	5	6	6	6	6	6	6	6	6	6	6
	.26	.04	.53	.89	.17	.40	.60	.77	.92	.05	.18	.29	.39	.48	.57	.65	.73	.80	.87
9	3	3	4	4	5	5	5	5	5	5	5	6	6.	6	6	6	6	6	6
	.20	.95	.41	.76	.02	.24	.43	.59	.74	.87	.98	.09	19	.28	.36	.44	.51	.58	.64
10	3.	3	4	4	4	5	5	5	5	5	5	5	6	6	6	6	6	6	6
	15	.88	.33	.65	.91	.12	.30	.46	.60	.72	.83	.93	.03	.11	.19	.27	.34	.40	.47
1	3	3	4	4	4	5	5	5	5	5	5	5	5	5	6	6	6	6	

1	.11	.82	.26	.57	.82	.03	.20	.35	.49	.61	.71	.81	.90	.98	.13	.20	.27	.33
														6.06				
12	3	3	4	4	4	4	5	5	5	5	5	5	5	5	5	6	6	6
	.08	.77	.20	.51	.75	.95	.12	.27	.39	.51	.61	.71	.80	.88	.95	.02	.09	.15
1		3	4	4	4	4	5	5	5	5	5	5	5	5	5	5	5	6
3	3.06	.73	.15	.45	.69	.88	.05	.19	.32	.43	.53	.63	.71	.79	.86	.93	.99	.05
1	3	3	4	4	4	4	4	5	5	5	5	5	5	5	5	5	5	6
4	.03	.70	.11	.41	.64	.83	.99	.13	.25	.36	.46	.55	.64	.71	.79	.85	.91	.97
15	3	3	4	4	4	4	4	5	5	5	5	5	5	5	5	5	5	5
	.01	.67	.08	.37	.59	.78	.94	.08	.20	.31	.40	.49	.57	.65	.72	.78	.85	.90
1	3	3	4	4	4	4	4	5	5	5	5	5	5	5	5	5	5	5
6	.00	.65	.05	.33	.56	.74	.90	.03	.15	.26	.35	.44	.52	.59	.66	.73	.79	.84
1	2	3	4	4	4	4	4	4	5	5	5	5	5	5	5	5	5	5
7	.98	.63	.02	.30	.52	.70	.86	.99	.11	.21	.31	.39	.47	.54	.61	.67	.73	.79
1	2	3	4	4	4	4	4	4	5	5	5	5	5	5	5	5	5	5
8	.97	.61	.00	.28	.49	.67	.82	.96	.07	.17	.27	.35	.43	.50	.57	.63	.69	.74
19	2	3	3	4	4	4	4	4	5	5	5	5	5	5	5	5	5	5
	.96	.59	.98	.25	.47	.65	.79	.92	.04	.14	.23	.31	.39	.46	.53	.59	.65	.70
2	2	3	3	4	4	4	4	4	5	5	5	5	5	5	5	5	5	5
0	.95	.58	.96	.23	.45	.62	.77	.90	.01	.11	.20	.28	.36	.43	.49	.55	.61	.66

表格内容为 $q_{0.10}$ 当 $P(q < q_{0.10}) = 0.90$

24	2.92	3.	3	4	4.37	4	4	4.4	4.92	5.	5.	5	5.	5.	5.	5.	5.	5.

		53	.90	.17		.54	.68	81		01	10	.18	25	32	38	44	49	55	59	
30	2.89	3.49		3.	4.4.30	4.46	4.60	4.72	4.82	4.92		5.	5.	5	5.	5.	5	5	5.	5.
			85	10							00	08	.15	21	27	.33	.38	43	47	
40	2.86		3.	3.	4	4.4.39		4.	4	4 4.82		4	4.	5.4.11		5	5	5.	5	5
		44	79	.04	23		52	.63	.73		.90	98	04		.16	.22	27	.31	.36	
60		2	3.	3.	3.	4.	4	4.	4.	4.	4	4.	4.	4.	5.5.06	5.11		5 5.20		5
	.83	40	74	98	16	.31	44	55	65	.73	81	88	94	00			.15		.24	
1	2.80		3.	3.	3.	4.	4.	4	4.4.56	4.64		4.	4.	4.	4.	4.	5.	5.5.09		5
20		36	68	92	10	24	.36	47			71	78	84	90	95	00	04		.13	
∞	2.77		3.	3.	3.	4	4.	4	4	4.	4.	4.	4.4.74		4.	4.4.89		4	4.	5.
		31	63	86	.03	17	.29	.39	47	55	62	68		80	85		.93	97	01	

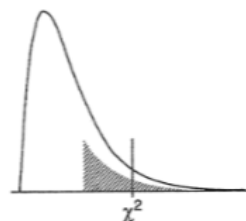
表格內容為 $q_{0.10}$ 當 $P(q < q_{0.10}) = 0.90$

1	8	1	1 18.49		2	2	22. 23.62	24.48	25.24		2 26.54		2 27.62	28.10	28.54	28.96		2	2
	.93	3.44	6.36		0.15	1.51	64			5.92		7.10						9.35	9.71
2		4	5.6.77	7.54		8	8.	9.	9.9.72		1 10.26		1	1 10.89		1	1	11.54	11.68
	.13	73			.14	63	05	41		0.01		0.49	0.70		1.07	1.24	1.39		
3	34.47	5.20		5.	6	6.6.81	7.06	7.29	7.49	7.67		7.7.98		8.	8.	8.	8.8.58		8.
	.33			74	.16	51					83		12	25	27	48		68	
4		3.98	4.59	4.03	5.39		5.	5.	6.6.33	6.49		6.6.78	6.91	7.02		7.	7.	7.	7.
	.01					68	93	14			65				13	23	33	41	50
5	23.72	4.26	4.66	4.98		5.	5.	5.5.82	5.97	6.10	6.22	6.34	6.44	6.54		6.6.71	6.79	6.86	
	.85					24	46	65							63				

	6	2.75		3.4.07		4	4.	4.	5.	5.5.50		5	5.	5.5.98		6.	6.	5 6.32	6.40		6.
		56		.44	73	97	17	34		.64	76	87		07	16	.25				47	
	7	2.68		3.	34.28		4.	4	4.	5	5 5.41		5	5	5.	5.	5 5.99	6.06		6	6.
		45		.93		55	.78	97	.14	.28		.53	.64	74	83	.91			.13	19	
	8	2.63		3.	3.	4.	4.	4.4.83	4.99		5	5.	5 5.46	5.56		5.5.72	5.80		5.	5.	6.
		37	83	17	43	65			.13	25	.36				64			87	93	00	
	9	2.59	3.32	3.76	4.08	4.34	4.54		4.	4.	5.	5	5	5 5.42		5 5.58		5.	5.5.79		5.
								72	87	01	.13	.23	.33		.51		66	72		85	
	10	2.56		3.3.70	4.02	4.26		4.4.64	4.78		4.	5	5	5	5	5.	5.	5.	5.	5.	5
		27					47			91	.03	.13	.23	.32	40	47	54	61	67	.73	
	11	2.54		3.3.66	3.96	4.20	4.40		4.	4.4.84	4.95		5.	5	5.	5	5	5.	5	5.	5.
		23						57	71			05	.15	23	.31	.38	45	.51	57	63	
		12.52	3.20	3.62		3.4.16	4.35	4.51		4.4.78	4.89	4.99	5.08		5.	5.	5	5.	5.5.49		5.
2					92				65					16	24	.31	37	44		55	
		12.50	3.18	3.59	3.88		4.4.20	4.46	4.60		4 4.83	4.93		5.5.10		5.	5	5	5.	5.	5.
3						12				.72			02		28	.25	.31	37	43	48	
	1	2.49	3.16	3.56		3	4.	4.4.42	4.56	4.68	4.79	4.88		4.5.05		5.	5	5 5.32		5.	5.
4					.85	08	27						97		12	.19	.26		37	43	
		12.48		3.	3 3.83		4.	4.4.39		4 4.64	4.75	4.84	4.93		5.	5.	5	5.	5.5.32		5.
5			14	.54		05	23		.52					01	08	.15	21	27		38	
	1	2.47		3.	3.	3	4.	4.	4.4.49		4.	4.	4.4.89		4.	5.	5	5	5.	5.	5.

6	12	52	.80	03	21	36			61	71	81		97	04	.11	.17	23	28	33	
	1 2.46	3.11	3.50		3 4.00		4 4.33	4.46	4.58	4.68	4.77	4.86	4.93		5.	5.	5	5.	5.	5.
7				.78		.18								01	07	.13	19	24	30	
	1 2.45		3 3.49		3.	3	4	4	4.	4 4.65	4.75	4.83	4.90	4.98	5.04		5.	5.	5.	5.
8		.10		.77	.98	.16	.31	.44	.55							10	16	21	26	
19	2.45	3.09	3.47	3.75	3.97	4.14	4.29	4.42	4.53	4.63	4.72	4.80	4.88	4.95	5.01	5.07	5.13	5.18	5.23	
20	2.44	3.08	3.46	3.74	3.95	4.12	4.27	4.40	4.51	4.61	4.70	4.78	4.85	4.92	4.99	5.05	5.10	5.16	5.20	
24	2.42	3.05	3.42	3.69	3.90	4.07	4.21	4.34	4.44	4.54	4.63	4.71	4.78	4.85	4.91	4.97	5.02	5.07	5.12	
30	2.40	3.02	3.39	3.65	3.85	4.02	4.16	4.28	4.38	4.47	4.56	4.64	4.71	4.77	4.83	4.89	4.94	4.99	5.03	
40	2.38	2.99	3.35	3.60	3.80	3.96	4.10	4.21	4.32	4.41	4.49	4.56	4.63	4.69	4.75	4.81	4.86	4.90	4.95	
60	2.36	2.96	3.31	3.56	3.75	3.91	4.04	4.16	4.25	4.34	4.42	4.49	4.56	4.62	4.67	4.73	4.78	4.82	4.86	
120	2.34	2.93	3.28	3.52	3.71	3.86	3.99	4.10	4.19	4.28	4.35	4.42	4.48	4.54	4.60	4.65	4.69	4.74	4.78	
∞	2.33	2.90	3.24	3.48	3.66	3.81	3.93	4.04	4.13	4.21	4.28	4.35	4.41	4.47	4.52	4.57	4.61	4.65	4.69	

表 15.5 χ^2 分配的上 α 點機率 (表格內容為 $\chi^2_{\alpha:v}$)



說明：(1) 查看自由度(v) 的 χ^2 分配相對應的行

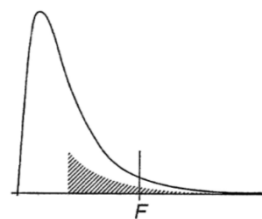
(2) 查看預定的 α 值出現的欄，接著挑出相對應的 χ^2 值

α														
V	0.995	0.99	0.975	0.950	0.900	0.75	0.500	0.25	0.100	0.05	0.025	0.010	0.005	
	0					0		0		0				
1	0.0000393	0.00	0.000982	0.00	0.01	0.10	0.45	1.32	2.71	3.84	5.02	6.63	7.88	
		0.157		0.393	0.58	2	5							
2	0.0100	0.02	0.0506	0.103	0.211	0.575	1.39	2.77	4.61	5.99	7.38	9.21	10.6	
		0.1												
3	0.0717	0.115	0.21	0.35	0.58	1.21	2.37	4.11	6.25	7.81	9.35	11.3	12.8	
			0.6	0.2	0.4									
4	0.20	0.29	0.484	0.711	1.06	1.92	3.36	5.39	7.78	9.49	11.1	13.	14.	
	0.7	0.7									13	14.9		
5	0.41	0.554	0.83	1.1	1.61	2.67	4.35	6.63	9.24	11.1	12.	15.	16.	
	0.2		0.1	0.5							8	1	7	
6	0.67	0.87	1.24	1.6	2.20	3.45	5.35	7.84	10.6	12.6	14.	16.	18.	
	0.6	0.2		0.4							4	8	5	
7	0.98	1.2	1.65	2.17	2.83	4.25	6.35	9.04	12.0	14.	16.	18.	20.3	
	0.9	0.4								1	0	5		
8	1.34	1.65	2.1	2.73	3.49	5.07	7.34	10.2	13.	15.	17.	20.1	22.0	
			0.8					4	5	5				

9	1.7	2.09	2.70	3.33	4.17	5.90	8.34	11.4	14.	16.	19.0	21.7	23.6	
3								7	9					
10	2.1	2.56	3.25		3.94	4.87	6.74	9.34	12.	16.	18.	20.5	23.2	25.2
6							5	0	3					
11	2.60		3.05	3.82	4.57	5.58	7.58	10.3	13.	17.	19.	21.9	24.7	26.8
							7	3	7					
12	3.07	3.57	4.40		5.23	6.30	8.44	11.3	14.	18.	21.0	23.3	26.2	28.
							8	5					3	
13	3.57	4.11	5.01	5.89	7.04	9.30	12.	16.	19.	22.4	24.7	27.7	29.8	
						3	0	8						
14	4.07	4.66		5.63	6.57	7.79	10.	13.	17.	21.1	23.7	26.1	29.1	31.
					2	3	1						3	
15	4.60		5.23	6.26	7.26	8.55	11.0	4.3	18.	22.3	25.0	27.5	30.6	32.8
							2							
16	5.14	5.81	6.91	7.96		9.31	11.9	15.	19.	23.5	26.3	28.8	32.0	34.
						3	4						3	
17	5.7	6.4	7.5	8.6	10.	12.	16.	20.	24.	27.	30.	33.	35.	
0	1	6	7	1	8	3	5	8	6	2	4	7		
18	6.26	7.0	8.23	9.39	10.9	13.	17.3	21.6	26.0	28.9	31.5	34.8	37.2	
	1				7									
19	6.84	7.63	8.9	10.	11.7	14.	18.3	22.7	27.2	30.1	32.9	36.2	38.6	
		1	1		6									
20	7.4	8.26	9.59	10.	12.	15	9.3	23.8	28.4	31.	34.2	37.6	40.0	

[illegible]

表 15.6 F 分配 (表格內容為 $F_{\alpha;v_1,v_2}$)



- 說明：(1) 查看預定 α 值相對應的表
 (2) 查看分母自由度 (v_2) 所對應的列
 (3) 查看分子自由度 (v_1) 所對應的欄，挑出相應的 F 值

v_1																			
v_2	12	3	4	5	6	7	8	9	10	15	20	25	30	40	50	60	80	100	∞
$\alpha = 0.10$																			
1	349.50	5.01	5.57	5.72	5.83	5.91	5.98	6.04	6.09	6.15	6.20	6.25	6.29	6.34	6.38	6.42	6.46	6.50	6.53
2	9.86	3.59	3.83	3.95	4.04	4.11	4.17	4.22	4.26	4.31	4.35	4.39	4.42	4.46	4.49	4.52	4.55	4.58	4.61
3	8.45	3.34	3.57	3.68	3.76	3.83	3.88	3.93	3.96	4.00	4.04	4.07	4.10	4.13	4.16	4.19	4.22	4.25	4.28
4	7.71	3.18	3.39	3.50	3.57	3.63	3.68	3.72	3.75	3.79	3.82	3.85	3.88	3.91	3.94	3.96	3.99	4.02	4.04
5	7.17	3.07	3.26	3.36	3.43	3.49	3.53	3.57	3.60	3.63	3.66	3.69	3.71	3.74	3.77	3.79	3.82	3.84	3.86
6	6.76	2.98	3.16	3.25	3.32	3.37	3.41	3.44	3.47	3.50	3.53	3.55	3.58	3.60	3.62	3.64	3.66	3.68	3.70
7	6.41	2.91	3.08	3.17	3.23	3.28	3.32	3.35	3.38	3.41	3.43	3.46	3.48	3.50	3.52	3.54	3.56	3.58	3.60
8	6.11	2.85	3.01	3.10	3.16	3.21	3.24	3.27	3.30	3.33	3.35	3.37	3.39	3.41	3.43	3.45	3.47	3.49	3.51
9	5.86	2.80	2.95	3.04	3.10	3.14	3.17	3.20	3.23	3.25	3.27	3.29	3.31	3.33	3.35	3.37	3.39	3.41	3.43
10	5.64	2.75	2.90	2.98	3.04	3.08	3.11	3.14	3.17	3.19	3.21	3.23	3.25	3.27	3.29	3.31	3.33	3.35	3.37
15	5.05	2.60	2.74	2.82	2.87	2.91	2.94	2.96	2.98	3.00	3.02	3.04	3.05	3.07	3.08	3.10	3.11	3.13	3.14
20	4.68	2.50	2.63	2.70	2.75	2.78	2.81	2.83	2.85	2.87	2.88	2.90	2.91	2.93	2.94	2.95	2.96	2.98	2.99
25	4.41	2.43	2.55	2.62	2.66	2.69	2.72	2.74	2.75	2.77	2.78	2.79	2.80	2.81	2.82	2.83	2.84	2.85	2.86
30	4.22	2.37	2.48	2.55	2.59	2.62	2.64	2.66	2.67	2.69	2.70	2.71	2.72	2.73	2.74	2.75	2.76	2.77	2.78
40	3.92	2.28	2.38	2.45	2.49	2.52	2.54	2.56	2.57	2.58	2.59	2.60	2.61	2.62	2.63	2.64	2.65	2.66	2.67
50	3.72	2.21	2.31	2.37	2.41	2.44	2.46	2.48	2.49	2.50	2.51	2.52	2.53	2.54	2.55	2.56	2.57	2.58	2.59
60	3.58	2.16	2.25	2.32	2.35	2.38	2.40	2.42	2.43	2.44	2.45	2.46	2.47	2.48	2.49	2.50	2.51	2.52	2.53
80	3.40	2.10	2.18	2.24	2.27	2.30	2.32	2.34	2.35	2.36	2.37	2.38	2.39	2.40	2.41	2.42	2.43	2.44	2.45
100	3.28	2.06	2.14	2.20	2.23	2.25	2.27	2.28	2.29	2.30	2.31	2.32	2.33	2.34	2.35	2.36	2.37	2.38	2.39
∞	3.12	2.00	2.07	2.13	2.16	2.18	2.20	2.21	2.22	2.23	2.24	2.25	2.26	2.27	2.28	2.29	2.30	2.31	2.32

		.54	.32	.19	.11	.05	.01	.98	.95	.94	.92	.90	.87	.84	.83	.82	.80	.79	.78	.76	
	5	4	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	
		.06	.78	.62	.52	.45	.40	.37	.34	.32	.30	.27	.24	.21	.19	.17	.16	.14	.12	.10	
	6	3.	3	3	3	3	3	3	2	2	2.	2	2	2	2	2	22.78		2	2.	2
		.78	.46	.29	.18	.11	.05	.01	.98	.96	.94	.90	.87	.84	.82	.80		.76	.74	.72	
	7	3.	3	3	2.	2	22.78		2	2	2	2	2.	2	2.58		2.54		2	2.	2.
		.59	.26	.07	.96	.88	.83		.75	.72	.70	.67	.63	.59		2.56		.51	.49	.47	
	8	3	3	2	2	2	2.	2.	2	2	2	2.50	2.46		2.2.40		2	2	2.	2.	2
		.46	.11	.92	.81	.73	.67	.62	.59	.56	.54			.42		.38	.36	.34	.32	.29	
	9	3	3	2	2	2	22.55		2	2.	2.	2.	2	2	2	2	2	2	2	2	2
		.36	.01	.81	.69	.61		.51	.47	.44	.42	.38	.34	.30	.28	.25	.23	.21	.18	.16	
	1	3	2	2	2	2	2	2	2	2	2	2	2	2.	2	2		2	2	2	
0		.29	.92	.73	.61	.52	.46	.41	.38	.35	.32	.28	.24	.20	.18	.16	.13	.11	.08	.06	
	1	3	2	2	2	2.2.45		2	2	2	2.	2	2.	2	2	2	2.	2	2	2	1
1		.23	.86	.66	.54		.39	.34	.30	.27	.25	.21	.17	.12	.10	.08	.05	.03	.00	.97	
	1	3	2	2	2	2.	2	2	2.	2	2	2	2	2.	2.	2	1	1	1	1	
2		.18	.81	.61	.48	.39	.33	.28	.24	.21	.19	.15	.10	.06	.04	.01	.99	.96	.93	.90	
			3	2	2.2.43		2	2	2	2.	2	2	2.2.05		2	1	1	1	1	1	
3		.14	.76	.56		.35	.28	.23	.20	.16	.14	.10		.01	.98	.96	.93	.90	.88	.85	
	1	3	2	2	22.39		2.2.24		2	2	22.10		2	2	1	1	1	1	1.	1.	
4		.10	.73	.52		.31		.19	.15	.12		.05	.01	.96	.94	.91	.89	.86	.83	.80	
		3	2.	2.	2.	2	2	2	2	2	2	2.	2.	1	1	1	1	1	1	1	

5	.07	70	49	36	.27	.21	.16	.12	.09	06	02	.97	.92	.90	.87	.85	.82	.79	.76
	13.05	2.67	2.46		2	2.	2.	2.2.09	2.06	2.03		1	1.	1	1	1.	1.	1.	11.72
6				.33	24	18	13				.99	94	.89	.87	84	81	78	.75	
	1	3.2.64	2.44		22.22		22.10	2.06		2	2.	1.	1.	1.	1.	1.	1	1.1.72	1.69
7	03			.31		.15			.03	00	96	91	86	84	81	.78	75		
	1	3.2.62	2.42	2.29	2.20		2.2.08	2.04	2.00		1.	1.1.89		1.	1	1.	1	1.1.69	1.
8	01					13				98	93		84	.81	78	.75	72		66
	12.99		2.2.40	2.27		2.2.11	2.06	2.02		1.	1.	1.	1.	1.	1	1.	1.	1	1.
9		61			18				98	96	91	86	81	.79	76	73	.70	.67	63
20		2.2.59	2.38	2.25	2.16	2.09	2.04	2.00		1	1.	1	1.	1	1.	1.	1.	1	1.
	97								.96	94	.89	84	.79	77	74	71	.68	64	61
	2	2.96	2.57	2.36	2.23		2.2.08	2.02		1.	1.	1.	1	1.	1	1	1.	11.66	1.
1					14				98	95	92	.87	83	.78	.75	72	.69		62
	2	2.95	2.56		2	2.22		2.	2.	2.	1.	1.	1.	1.	1.	11.70		1.	11.60
2			.35		13	06	01	97	93	90	86	81	76	.73		67	.64		.57
	2	2.94	2.55	2.34		2.	2.2.05	1.99		1.	1.1.89		11.80		1.	1.1.69		1.	1.1.59
3				21	11				95	92		.84		74	72		66	62	.55
24	2.93		2. 2.33		2.	2. 2.04		1.	1	1	1.	1.	1.	1.1.70		1.	1.	1.	1.
		54		19	10			98	.94	.91	88	83	78	73		67	64	61	57
	2	2.92	2.53	2.32		2.2.09	2.02		1.	1	1.	1.	1.	1	1	1.	1.	1.1.56	1.
5				18				97	.93	89	87	82	77	.72	.69	66	63	59	52
26		2.2.52		2.	2.2.08		2.	1.	1	1.	1.	1	1.	1.	1.	1.	1	1	1.

	91		31	17		01	96	.92	88	86	.81	76	71	68	65	.61	.58	54	.50	
	2 2.90	2.51		2	2.2.07	2.00		1.	1	1.	1.1.80		1.1.70		1.	1.	1	1.	1.	1.
7			.30	17			95	.91	87	85		75		67	64	.60	.57	53	49	
	2 2.89	2.50	2.29		2.2.06	2.00		1.	1	1.	1.1.79		1.1.69		1.	1.	1	1	1.	1.
8				16			94	.90	87	84		74		66	63	.59	.56	52	48	
29	2.89	2.50		2	2.2.06		1.	1.1.89		1	1.	1	1	1.	1.	1	1	1	1.	1
			.28	15		99	93		.86	83	.78	.73	68	65	.62	.58	.55	51	.47	
30	2.88	2.49		2.	22.05		1.	1.	1.	1.	1.	1	1	1.	1.	1.	1	11.50		1.
			28	.14		98	93	88	85	82	.77	.72	67	64	61	.57	.54		46	
40	2.84	2.44	2.23	2.09	2.00		1	1.	1.	1.1.76		1	1	1.	1	1.	1	11.42		1.
						.93	87	83	79		.71	.66	61	.57	54	.51	.47		38	
60	2.79	2.39		2.2.04	1.95		1.	1.	1	1	1.	1 1.60		1	1	1.	1	1	1.	1
			18			87	82	.77	.74	71	.66		.54	.51	48	.44	.40	35	.29	
	12.75	2.35		2	1	1	1.	1.	1	:	1.	1	1	1.	1.	1.	1	1.1.26	1.19	
20			.13	.99	.90	82	77	.72	:68	65	.60	.55	48	45	41	.37	32			
00		2.	2. 2.08		1.	1.	1.	1.	1	1.1.60		1.1.49		1.	1	1.	1.	1.	1.1.00	
	71	30		94	85	77	72	.67	63		55		42	.38	34	30	24	17		

a=0.05

	1	1	2	224.6		2	234.0		2	238.9	240.5	241.9	243.9	245.9	248.0	2	2	251.1	252.2	253.3	254.3
	61.4	99.5	15.7		30.2		36.8									49.1	50.1				
2	1	1	19.16		1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	8.51	9.00		9.25	9.30	9.33	9.35	9.37	9.38	9.40	9.41	9.43	9.45	9.45	9.46	9.47	9.48	9.49	9.50		
3	19.55		9.	9.9.01		8	8.89		8	8.8.79	8.74		8.8.66		8.8.62	8.59		8.	8.	8.	

	0.13	28	12		.94		.85	81		70		64		57	55	53	
	4 7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.5.86	5.	5.5.75	5.5.69	5.66	5.	
										91		80	77	72		63	
	56.61	5.79	5.	5	5 4.95	4.88	4.82	4.77	4.74	4.68	4.62	4.56	4.4.50	4	4 4.40	4.36	
			41	.19	.05								53	.46	.43		
6	5.99	5.14	4 4.53	4.39	4.28	4.21		44.10	4.06	4.00	3.94	3.87	3	3.	3	3 3.70	3.67
			.76				.15						.84	81	.77	.74	
	75.59	4.74	4 4.12	3.97	3.87	3.79		3.	3.3.64	3.	33.44	3.3.38	3.34	3.30	3.27	3.23	
			.35				73	68		57	.51	41					
	85.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3	3	3	3	3 3.08	3	3.2.97	2.93
									.35	.28	.22	.15	.12	.04	01		
9	5.12	4.26	3.86	3.3.48	3.	3	3.	3 3.14	3.	3	2 2.90	2.86	2.83		2	2 2.71	
			63		37	.29	23	.18		07	.01	.94		.79	.75		
1	4.96	4 3.71	3.	3.3.22	3 3.07	3:02	2.98		2.	2.	2.	2.2.70	2.66	2.62	2.58	2.54	
0		.10	48	33		.14				91	85	77	74				
	1 4.84	3.3.59	3.36	3.20	3.09		3.2.95	2.90	2.85		2 2.72	2.65	2.	2.2.53	2.49	2. 2.40	
1		98				01				.79		61	57		45		
	14.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.	2 2.69	2.62	2.54	2.	2.	2	2	2
2								80	.75			51	47	.43	.38	.34	30
	1	4.3.81	3.	33.03	2.92	2.83	2.77	2.	2.2.60		2.2.46	2.42	2.38	2.34	2.30	2.25	2.
3	67		41	.18				71	67		53						21
	14.60	3 3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.53	2.46	2	2 2.31	2.27	2.22	2.	2
4		.74										.39	.35			18	.13

	14.54	3	3.	3	2 2.79	2.	2 2.59	2.54	2.48	2.40	2.33	2.29	2.25	2.	2.2.11	2.		
5		.68	29	.06	.90	71	.64							20	16	07		
	14.49	3.63	3.24		3.2.85	2.74	2.66	2.59	2.54	2.49	2.42	2 2.28	2.24	2	2	2.	2.	2.
6				01							.35		.19	.15	11	06	01	

表 15.6 續

v_1																				
v_2	1	2	3	4	5	6	7	8	9	1 12	15	20	2 30	4	6 120	∞				
									0				4		0	0				
7	4.	3.	3	2	2.2.70		2	2	2.2.45		2	2	2.	2	22.10	2.06	2.	1.		
	45	59	.20	.96	81	.61	.55	49		.38	.31	23	.19	.15		01	96			
8	4	3.	3	2	2.2.66		2.	2	2.	2.	2.	2.	2	22.11		2.	2.	1.	1.	
	41	55	.16	.93	77		58	.51	46	.41	34	27	.19	.15		0b	02	97	92	
9	4	3	32.90		2.	2.	2	2.	2.	2	2	2.	22.11	2.07		2	1.	1	1	
	.38	.52	.13		74	63	.54	48	42	.38	.31	23	.16		.03	98	.93	.88		
0	2	4	3.	3	2.	2	2	2	2.	2	2	2.	22.08	2.04		1	1	1.	1.	
	.35	49	.10	87	.71	.60	.51	45	.39	.35	.28	20	.12		.99	.95	90	84		
1	4.	3.	32.84		2	2.	2.	2.	2	2.2.25		2	2.	2	2.	1.	1.	1	1	
	32	47	.07		.68	57	49	42	.37	32		.18	10	.05	01	96	92	.87	.81	
2	2	4	3.	3.	2.2.66	2.55	2.46	2.40		2	2.	2.	22.07		2.	1.	1	1.	1.	1.
	.30	44	05	82					.34	30	23	.15		03	98	.94	89	84	78	
2	2	4.	3.	32.80		2.	2.2.44		2	2	2.	2	22.05		2	1.	1	1.	1.	1.

3	28	42	.03		64	53		.37	.32	27	.20	.13		.01	96	.91	86	.81	76	
	2	4.26		3.	3	2	2	2.	2.	2	2	2	2.	2.	1.	1.	1	1.	1	1
4		40	.01	.78	.62	.51	42	36	.30	.25	.18	11	03	98	94	.89	84	.79	.73	
	2	4.	3	2	2.	2	2	2.40		2.	2	2.24		2	2.	2.	1.	1.	1	1.
5	24	.39	.99	76	.60	.49		34	.28		.16	09	01	96	92	.87	82	77	.71	
	2	4.	3	2.	2.	2.247		2	2.	2.	2.	2	2.	1.	1	1.	1	1	1.	1
6	23	.37	98	74	59		.39	32	27	22	.15	07	99	.95	90	.85	.80	75	.69	
	2	4	3	2	2.257	2.46		2	2	2.2.20		2	2.	1.	1	1	1	1.	1	1
7	.21	.35	.96	.73			.37	.31	25		.13	06	97	.93	.88	.84	79	.73	.67	
	2	4.	3.	2	2	2.56	2.45	2.36		2.2.24		2	2	2.04		1.	1	1.	1	1
8	20	34	.95	.71			.29		.19	.12		96	.91	87	.82	.77	.71	.65		
	2	4	3	2	2.2.55	2.43		2	2.	2.	2	2	2.	1	1	1	1	1	1.	1
9	.18	.33	.93	70			.35	28	22	.18	.10	03	.94	.90	.85	.81	.75	70	.64	
	3	4	3	2.	2.	2.2.42		2	2	2.	2	2	2	1.	1.	1.	1.	1.	1	1
0	.17	.32	92	69	53		.33	.27	21	.16	.09	.01	93	89	84	79	74	.68	.62	
	4	4	3	2	2	2.45		2.2.25		2	2	2.08	2.00		1	1.	1.	1.	1.	1
0	.08	.23	.84	.61		.34		.18	.12			.92	84	79	74	69	64	.58	.51	
	6	4.		2.	2.	2.2.25		2	2.2.04	1.99		1.	1	1.	1.	1.	1	1	1.	1.
0	00	.15	76	53	37		.17	.10			92	.84	75	70	65	.59	.53	47	39	
	1	3	3.	2	2.	2.	2.2.09		2	1.	1.	1	1	1.	1	1	1	1	1	1
20	.92	07	.68	45	29	.17		.02	96	91	.83	.75	66	.61	.55	.50	.43	.35	.25	
	0	3.	3.	2.	2.	2	2	2.	1.	1.	1.	1	1	1	1.	1.	1	1.	1.	1
0	84	00	60	37	.21	.10	01	94	88	83	.75	.67	.57	52	46	.39	32	22	.00	

$\alpha = 0.01$																			
	4	4999.5	5	5	5	5	5	5	6	6	6	66209	6	6	6	6	6	6	6
	052		403	625	764	859	928	982	022	056	106	157		235	261	287	313	339	366
2	9	9	9	9	9	9	99.36		999.39		9	9	999.45	99.46	99.47		9	9	99
	8.50	9.00	9.17	9.25	9.30	9.33		9.37		9.40	9.42	9.43				9.47	9.48	9.49	.50
3	330.82		2		228.24		2	227.49		227.23	27.05		226.69		2	2	2	2	26
	4.12		9.46	8.71		7.91	7.67		7.35			6.87		6.60	6.50	6.41	6.32	6.22	.13
4	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
	1.20	8.00	6.69	5.98	5.52	5.21	4.98	4.80	4.66	4.55	4.37	4.20	4.02	3.93	3.84	3.75	3.65	3.56	3.46
5	1	1	1	1	1	1	1	1	1	1	9.	9	9	9.	9	9	9	9.	9
	6.26	3.27	2.06	1.39	0.97	0.67	0.46	0.29	0.16	0.05	89	.72	.55	47	.38	.29	.20	11	.02
6	1	1	9	9	8	8.	8.	8.	7	7.	7.	7.7.40		7	7	7	7.	6.	6.
	3.75	0.92	.78	.15	.75	47	26	10	.98	87	72	56		.31	.23	.14	06	97	88
7	19.55		8	77.46		7	6.	6	6	6	6.	6	6.	6	5	5	5	5.	5.
	2.25		.45	.85		.19	99	.84	.72	.62	47	.31	16	.07	.99	.91	.82	74	65
8	1	8	7	7	6	6	6	6.	5	5	5	5	5.	55.20		5	4.	4.4.86	
	1.26	.65	.59	.01	.63	.37	.18	03	.91	.81	.67	.52	36	.28		.12	03	95	
9	1	8	6.	6.6.06	5.80		5.	5	5	5.	5.	4	4.	4	4.	4	4.	4	4.
	0.56	.02	99	42			61	.47	.35	26	11	.96	81	.73	65	.57	48	.40	31
1	1	7.	6	5	5.	5	5.	5.	4.	4.	4.	44.41		44.25		44.08		4.	3.
0	0.04	56	.55	.99	64	.39	20	06	94	85	71	.56		.33		.17		00	91
	1	9.	7.	6	5.	5.	54.89		4.	4.	4.	4	4.	4.	4	3.	3	3	3
1	65	21	.22	67	32	.07		74	63	54	.40	25	10	.02	94	.86	.78	.69	.60

	1	9	6.	5	5	5.4.82	4.64		4.4.39		4	4	4	3	3	3	3.	3	3
2	.33	93	.95	.41	06			50		.30	.16	.01	.86	.78	.70	.62	54	.45	.36
	1	9	6.	5.	5	4	4.4.44		4	4	4	3	3	3	3.	3	3	3	3
3	.07	70	74	.21	.86	.62		.30	.19	.10	.96	.82	.66	59	.51	.43	.34	.25	.17
	1	8	6	5.	5.	4	4.	4.	4	4.	3.	3.	3.	3.	3.	3.	3	3.	3.
4	.86	.51	56	04	.69	46	28	.14	03	94	80	66	51	43	35	.27	18	09	.00
	1	8.	6.	5	4	4.	4.	4	4.	3.	3.	3.	3.	3.	3.	3	3	3.2.96	2.
5	68	36	.42	.89	56	32	.14	00	89	80	67	52	37	29	.21	.13	05		87
	1	8	6.	5.	4.	4.	4.	4.	3.	3.	3.	3.	3	3.	3	3.	3	2.2.84	2.75
6	.53	23	29	77	44	20	03	89	78	69	55	.41	26	.18	10	.02	93		
	1	8.6.11		5	4	4.	4	3.	3.	3.	3.	3	3	3	3	3.	2.	2.	2.2.65
7	40		.18	.67	34	.10	93	79	68	59	.46	.31	.16	.08	00	92	83	75	
	1	8.	6.5.09		4.	4.	4.	3.	3	3.	3.	3.	3.	3.	3.2.92		2.	2	2.2.57
8	29	01		58	25	01	84	.71	60	51	37	23	08	00		84	.75	66	
	1	8	5.	5.	4.	4	3.	3	3.	3.	3.	3.	3	3.	2.	2.2.76		2.	2.2.49
9	.18	93	01	50	.17	94	.77	63	52	43	30	.15	00	92	84		67	58	
20		8	5.4.94		4.	4	3	3.	3.	3.	3.	3.	3.	2.2.86		2.	2.	2.2.52	2.42
	.10	85		43	.10	.87	70	56	46	37	23	09	94		78	69	61		
	2	8.	5.	4	4	4	3	3	3	3.	3.	3	3.	2.2.80		2.	2	2.2.46	2.36
1	02	78	.87	.37	.04	.81	.64	.51	40	31	.17	03	88		72	.64	55		
	2	7.	54.82		43.99		36.59		3.	3	3.	3	2.	2.	2.	2.	2.2.50	2.40	2
2	95	.72		.31		.76		45	.35	26	.12	98	83	75	67	58			.31
	2	7.88		5.4.76	4.26		3.	3.	3.	3	3.	3.	2.	2.	2.	2.	2.	2.	2.2.26

3		66			94	71	54	.41	30	21	07	93	78	70	62	54	45	.35		
	2	7	5.	4	4.	3.	3.	3.	3.	3	3	3.2.89		2.	2.	2.2.49	2.40		2	2.
4	.82	61	.72	22	90	67	50	36	.26	.17	03		74	66	58			.31	21	
	2	7.	5	4	4	3.	3	3.	3.	3.	3	2.99		2.	2.	2.	2	2.2.36	2.	2
5	77	.57	.68	.18	85	.63	46	32	22	.13		85	70	62	.54	45		27	.17	
	2	7.	5.4.64		4	3	3	3.	3.	3	3.2.96		2.	2	2.	2.	2.	2.	2.	2
6	72	53		.14	.82	.59	42	29	.18	09		81	.66	58	50	42	33	23	.13	
	2	7.	5.4.60	4.11		3.	3.	3	3	3.2.06		2.	2	2.	2.	2.	2.2.29	2.2.0	2.10	
7	68	49			78	56	.39	.26	.15		93	.78	63	55	47	.38				
	2	7.64		5.	4.	4.	3.	3.	3	3.	3	3.2.90		2.2.60		2.	2.	2	2.	2
8		45	57	07	75	53	.36	23	.12	03		75		52	44	.35	26	17	.06	
29	7.60		5.4.54		4	3	3.	3.	3	3.	3.	2.	2.	2.	2.	2.	2.	2.	2.2.03	
		42		.04	.73	50	33	.20	09	00	87	73	57	49	41	33	23	14		
	3	7.5.39		4.	4.	3.	3.	3.	3	3.	2.	2.	2	2.	2.	2.	2.	2.2.11		2.
0	56		51	02	70	47	30	.17	07	98	84	.70	55	47	39	30	21		01	
40		7	5	4	3.	3.	3.	3.2.99	2.89	2.80		2.	2.	2.	2.	2.	2	2.	1.1.80	
	.31	.18	.31	83	51	29	.12				66	52	37	29	20	.11	02	92		
	6	7.08	4.98		4	3.	3.	3	2	2.	2.	2.2.50		2.2.20		2	2.	1.	1	1.1.60
0			.13	65	34	.12	.95	82	72	63		35		.12	03	94	.84	73		
	1	6.4.79		3	3.	3	2.2.79	2.66		2.	2.2.34		2	2.	1	1.	1	1.	1.	1.
20	85		.95	48	.17	96			56	47		.19	03	.95	86	.76	66	53	38	
∞		6.4.61		3	3	3.	2.	2.	2.	2.	2	2.	1	1	1.	1	1	1	1.1.00	
	63		.78	.32	02	80	64	51	41	32	.18	04	.88	.79	70	.59	.47	32		

表 15.7 三角測試的最小評估數（表格內容為 n_{α,β,p_d} ）

表格內容為三角測試中要求的，由 α 、 β 、 p_d 值所決定的，相等尺寸（n）的敏感度。進入 p_d 值相對應的表格段落，以及選定的 α 值所對應的列。讀取所需要的樣本大小，n，以及它所對應的 β 值。

α	β							
	0.50	0.40	0.30	0.20	0.10	0.05	0.01	0.001
Pd =50%								
0.40	3	3	3	6	8	9	15	26
0.30	3	3	3	7	8	11	19	30
0.20	4	6	7	7	12	16	25	36
0.10	7	8	8	12	15	20	30	43
0.05	7	9	11	16	20	23	35	48
0.01	13	15	19	25	30	35	47	62
0.001	22	26	30	36	43	48	62	81
pd=40%								
0.40	3	3	6	6	9	15	26	41
0.30	3	3	7	8	11	19	30	47
0.20	6	7	7	12	17	25	36	55
0.10	8	10	15	17	25	30	46	67
0.05	11	15	16	23	30	40	57	79
0.01	21	26	30	35	47	56	76	102
0.001	36	39	48	55	68	76	102	130
pd=30%								
0.40	3	6	6	9	15	26	44	73
0.30	3	8	8	16	22	30	53	84
0.20	7	12	17	20	28	39	64	97
0.10	15	15	20	30	43	54	81	119
0.05	16	23	30	40	53	66	98	136
0.01	33	40	52	62	82	97	131	181
0.001	61	69	81	93	120	138	181	233
pd=20%								
0.40	6	9	12	18	35	50	94	153
0.30	8	11	19	30	47	67	116	183
0.20	12	20	28	39	64	86	140	212
0.10	25	33	46	62	89	119	178	260
0.05	40	48	66	87	117	147	213	305

0.01	72	92	110	136	176	211	292	397
0.001	130	148	176	207	257	302	396	513
pd=10%								
0.40	9	18	38	70	132	197	360	598
0.30	19	36	64	102	180	256	430	690
0.20	39	64	103	149	238	325	439	819
0.10	89	125	175	240	348	457	683	1011
0.05	144	191	249	325	447	572	828	1178
0.01	284	350	425	525	680	824	1132	1539
0.001	494	579	681	803	996	1165	1530	1992

表 15.8 三角測試的正確回答數臨界值（表格內容為 $x_{\alpha,n}$ ）

表格內容為對於在預定的 α －顯著性（即，直欄）以及相對應的回應者數量（即，橫列）所必要的最少正確回答數。假如正確回答數大於或等於表格內所對應的值，拒絕「無差別」的假設。

α							α							
(	(	(	(	0	1	2	(	(	(	(	0	1	2	
.4	.3	.2	.1	.05	.01	.001	.4	.3	.2	.1	.05	.01	.001	
											1	1	2	
							1	2	3	4	5	6	8	0
											1	1	2	
							2	2	3	4	5	6	8	0
2			3	3	-	—					1	1	2	
							3	3	3	4	5	7	8	1
3			4	4	-	—					1	1	2	
							4	3	4	5	6	7	9	1
3			4	4	5	—					1	1	2	
							5	3	4	5	6	7	9	2
3			5	5	6	—					1	2	2	
							6	4	4	5	7	8	0	2
4			5	5	6	7					2	2	2	
							2	6	7	8	9	0	2	5
4			5	6	7	8					2	2	2	
							8	8	9	0	1	2	5	7
4			6	6	7	8					2	2	3	

						4	0	1	2	3	5	7	0		
0	€		€	7	€	9						2	€	3	
1	€		7	7	€	1						2	€	3	
2	€		7	€	€	1						€	€	3	
3	€		€	€	€	1						€	€	4	
4	€		€	€	1	1						€	€	4	
5	€		€	€	1	1						€	4	4	
6	7		€	€	1	1						4	4	4	
7	7		€	1	1	1						4	4	5	
8	7		1	1	1	1						4	4	5	
9	€		1	1	1	1						4	€	5	
0	€		1	1	1	1						€	€	5	
1	€		1	1	1	1						€	€	6	
2	€		1	1	1	1						€	€	6	
3	0	1	2	2	4	6	38	8	0	2	4	6	0	4	
4	0	0	1	2	3	5	44	0	2	4	6	8	2	7	
5	0	1	1	2	3	5	50	2	4	6	8	1	5	9	
6	0	1	2	3	4	5	56	4	6	8	1	3	7	2	
7	1	1	2	3	4	6	62	6	8	0	3	5	9	4	
	1			1	1	1	1						€	7	7

8	1	2	2	4	5	6	8	68	8	0	2	5	7	1	6
		1			1	1	1	1						6	7
9	1	2	3	4	5	7	9	74	1	2	4	7	9	4	9
		1			1	1	1	1						7	8
0	2	2	3	4	5	7	9	80	3	4	6	9	1	6	1

註：對於不在表中的 n 值，計算 $z = \left(k - 1\left(\frac{1}{3}\right)n\right) \sqrt{\left(\frac{2}{9}\right)n}$ ，其中 k 是正確回應的數目。比較 z 值及標準常態分佈的 α -臨界值，即，表 17.3 的最後一橫列 ($z_\infty = t_{\alpha,\infty}$)

表 15.9 在兩三點測試法或者是單面方向性差異測試 (表格內容項為 n_{α, β, p_d})

表格內容是在二三點測試法或單面定向差異所需的樣本大小 (n)，透過選擇 α, β, p_d 的值定義的敏感性。在兩三點測試中，進入相對應的 p_d 值的表格段，或在方向性差異測試中進入 p_{max} 所對應的表格段，且選取 α 值所對應的橫列。且選取 β 值所對應的直欄，讀取所需的樣本大小， n 。

[illegible]

0.40	4	6	8	14	29	41	76	120
0.30	7	9	13	24	39	53	88	144
0.20	10	17	21	32	49	68	110	166
0.10	21	28	37	53	72	96	145	208
0.05	30	42	53	69	93	119	173	243
0.01	64	78	89	112	143	174	235	319
0.001	107	126	144	172	210	246	318	412
$p_d = 20\%$								
$p_{max} = 60\%$								
0.40	6	10	23	35	59	94	171	282
0.30	11	22	30	49	84	119	205	327
0.20	21	32	49	77	112	158	253	384
0.10	46	66	85	115	168	214	322	471
0.05	71	93	119	158	213	268	392	554
0.01	141	167	207	252	325	391	535	726
0.001	241	281	327	386	479	556	731	944
$p_d = 10\%$								
$p_{max} = 55\%$								
0.40	10	35	61	124	237	362	672	1124
0.30	30	72	117	199	333	479	810	1302
0.20	81	129	193	294	451	618	1006	1555
0.10	170	239	337	461	658	861	1310	1905
0.05	281	369	475	620	866	1092	1583	2237
0.01	550	665	820	1007	1301	1582	2170	2927
0.001	961	1125	1309	1551	1908	2248	2937	3812

表 15.10 兩三點測試以及單面方向性差異測試的正確回應數臨界值（表格內容項為 $x_{\alpha,n}$ ）

表格內容項是在 α 自由度（即，直欄）的統計顯著的最低正確回應數目，以及所相對應的受試者人數， n （即，橫列）。假如正確回應數大於或等於表格值，拒絕「無差別」的假設。

n	α							n	α						
	0	0	0	0	0	0	0		0	0	0	0	0	0	0
	.40	.30	.20	.10	.05	.01	.001		.40	.30	.20	.10	.05	.01	.001

								31	17	18	19	20	21	23	25
2	2	2	—	—	—	—	—	32	18	18	19	21	22	24	26
3	3	3	3	—	—	—	—	33	18	19	20	21	22	24	26
4	3	4	4	4	—	—	—	34	19	20	20	22	23	25	27
5	4	4	4	5	5	—	—	35	19	20	21	22	23	25	27
6	4	5	5	6	6	—	—	36	20	21	22	23	24	26	28
7	5	5	6	6	7	7	—	40	22	23	24	25	26	28	31
8	5	6	6	7	7	8	—	44	24	25	26	27	28	31	33
9	6	6	7	7	8	9	—	48	26	27	28	29	31	33	36
10	6	7	7	8	9	10	10	52	28	29	30	32	33	35	38
11	7	7	8	9	9	10	11	56	30	31	32	34	35	38	40
12	7	8	8	9	10	11	12	60	32	33	34	36	37	40	43
13	8	8	9	10	10	12	13	64	34	35	36	38	40	42	45
14	8	9	10	10	11	12	13	68	36	37	38	40	42	45	48
15	9	10	10	11	12	13	14	72	38	39	41	42	44	47	50
16	10	10	11	12	12	14	15	76	40	41	43	45	46	49	52
17	10	11	11	12	13	14	16	80	42	43	45	47	48	51	55
18	11	11	12	13	13	15	16	84	44	45	47	49	51	54	57
19	11	12	12	13	14	15	17	88	46	47	49	51	53	56	59
20	12	12	13	14	15	16	18	92	48	50	51	53	55	58	62
21	12	13	13	14	15	17	18	96	50	52	53	55	57	60	64
22	13	13	14	15	16	17	19	100	52	54	55	57	59	63	66
23	13	14	15	16	16	18	20	104	54	56	57	60	61	65	69
24	14	14	15	16	17	19	20	108	56	58	59	62	64	67	71
25	14	15	16	17	18	19	21	112	58	60	61	64	66	69	73
26	15	15	16	17	18	20	22	116	60	62	64	66	68	71	76
27	15	16	17	18	19	20	22	122	63	65	67	69	71	75	79
28	16	16	17	18	19	21	23	128	66	68	70	72	74	78	82
29	16	17	18	19	20	22	24	134	69	71	73	75	78	81	86
30	17	17	18	20	20	22	24	140	72	74	76	79	81	85	89

注：對於表中沒有出現的 n 值，計算 $z = \frac{(k-0.5n)}{\sqrt{0.25n}}$ ，其中 K 是正確回應數目。比較 z 值與標準常態分佈在 α 自由度的值，即，表 17.3 的最後一列 ($z_\alpha = t_{\alpha,\infty}$)。

表 15.11 雙面方向性差異測試的最小評估數目（表格內容項為 $n_{\alpha,\beta,p_{max}}$ ）

表格內容是在雙面方向性差異測試所需的樣本大小 (n)，透過選擇 α, β, p_{max} 的值定義的敏感度。進入 p_{max} 所對應的表格段，且選取 α 值所對應的橫列。且選取 β 值所對應的直欄，讀取所需的樣本大小， n 。

α	β							
	0.50	0.40	0.30	0.20	0.10	0.05	0.01	0.001
$p_{max} = 75\%$								
0.40	5	5	10	12	19	26	39	58
0.30	6	8	11	16	22	29	42	64
0.20	9	9	14	19	26	33	48	70
0.10	13	16	18	23	33	42	58	82
0.05	17	20	25	30	42	49	67	92
0.01	26	34	39	44	57	66	87	117
0.001	42	50	58	66	78	90	117	149
$p_{max} = 70\%$								
0.40	5	10	12	19	30	39	60	94
0.30	8	13	18	22	33	44	68	102
0.20	14	19	21	28	39	53	79	113
0.10	18	23	30	37	53	67	93	132
0.05	25	35	40	49	65	79	110	149
0.01	44	49	59	73	92	108	144	191
0.001	68	78	90	102	126	147	188	240
$p_{max} = 65\%$								
0.40	10	17	21	32	49	68	110	166
0.30	13	20	29	42	59	81	125	188
0.20	21	28	37	53	72	96	145	208
0.10	30	42	53	69	93	119	173	243
0.05	44	56	67	90	114	145	199	176
0.01	73	92	108	131	164	195	261	345
0.001	121	140	161	188	229	267	342	440
$p_{max} = 60\%$								
0.40	21	32	49	77	112	158	253	384
0.30	31	44	66	89	133	179	283	425
0.20	46	66	85	115	168	214	322	471
0.10	71	93	119	158	213	268	392	554
0.05	101	125	158	199	263	327	455	635
0.01	171	204	241	291	373	446	596	796
0.001	276	318	364	425	520	604	781	1010
$p_{max} = 55\%$								
0.40	81	129	193	294	451	618	1006	1555
0.30	110	173	254	359	550	721	1130	1702

0.20	170	239	337	461	658	861	1310	1905
0.10	281	369	475	620	866	1092	1583	2237
0.05	390	497	620	786	1055	1302	1833	2544
0.01	670	802	963	1167	1493	1782	2408	3203
0.001	1090	1260	1461	1707	2094	2440	3152	4063

表 15.12 雙面方向性差異測試的正確回應數臨界值（表格內容項為 $\chi_{\alpha,n}$ ）
 表格內容項是在 α 自由度（即，直欄）的統計顯著的最低正確回應數目，以及
 所相對應的受試者人數， n （即，橫列）。假如正確回應數大於或等於表格值，拒
 絕「無差別」的假設。

	α							n	α						
	.4	.3	.2	.1	.05	.01	.001		.4	.3	.2	.1	.05	.01	.001
								31	19	19	20	21	22	24	25
2	—	—	—	—	—	—	—	32	19	20	21	22	23	24	26
3	3	3	—	—	—	—	—	33	20	20	21	22	23	25	27
4	4	4	4	—	—	—	—	34	20	21	22	23	24	25	27
5	4	5	5	5	—	—	—	35	21	22	22	23	24	26	28
6	5	5	6	6	6	—	—	36	22	22	23	24	25	27	29
7	6	6	6	7	7	—	—	40	24	24	25	26	27	29	31
8	6	6	7	7	8	8	—	44	26	26	27	28	29	31	34
9	7	7	7	8	8	9	—	48	28	29	29	31	32	34	36
10	7	8	8	9	9	10	—	52	30	31	32	33	34	36	39
11	8	8	9	9	10	11	11	56	32	33	34	35	36	39	41
12	8	9	9	10	10	11	12	60	34	35	36	37	39	41	44
13	9	9	10	10	11	12	13	64	36	37	38	40	41	43	46
14	10	10	10	11	12	13	14	68	38	39	40	42	43	46	48
15	10	11	11	12	12	13	14	72	41	41	42	44	45	48	51
16	11	11	12	12	13	14	15	76	43	44	45	46	48	50	53
17	11	12	12	13	13	15	16	80	45	46	47	48	50	52	56
18	12	12	13	13	14	15	17	84	47	48	49	51	52	55	58
19	12	13	13	14	15	16	17	88	49	50	51	53	54	57	60
20	13	13	14	15	15	17	18	92	51	52	53	55	56	59	63
21	13	14	14	15	16	17	19	96	53	54	55	57	59	62	65
22	14	14	15	16	17	18	19	100	55	56	57	59	61	64	67

0.40	5	5	6	8	9	11	20	30
0.30	5-	5	6	8	9	15	24	35
0.20	5	5	6	8	13	15	28	39
0.10	5	5	9	11	17	22	32	47
0.05	7	8	12	14	20	26	39	54
0.01	13	14	18	23	30	36	49	69
0.001	21	22	27	32	42	49	66	87
$p_d=20\%$								
0.40	6	7	8	10	13	21	38	59
0.30	6	7	8	10	18	26	43	69
0.20	6	7	8	15	22	30	53	79
0.10	10	11	17	23	31	40	62	94
0.05	13	19	24	27	40	53	76	108
0.01	24	30	36	43	57	70	99	136
0.001	38	48	55	67	81	99	129	172
$p_d =10\%$								
0.40	9	11	13	22	40	60	108	184
0.30	9	16	19	34	54	80	128	212
0.20	14	22	31	47	73	99	161	245
0.10	25	38	54	70	103	130	206	297
0.05	41	55	70	94	127	167	244	249
0.01	77	98	121	145	192	233	330	449
0.001	135	158	187	224	278	332	438	572

表 15.14 五取二測試的正確回應數臨界值（表格內容項為 $x_{\alpha,n}$ ）

表格內容項是在 α 自由度（即，直欄）的統計顯著的最低正確回應數目，以及所相對應的受試者人數， n （即，橫列）。假如正確回應數大於或等於表格值，拒絕「無差別」的假設。

α								α							
								</							

6	2	2	2	3	3	4	5	36	5	5	6	7	8	9	11
7	2	2	2	3	3	4	5	37	5	6	6	7	8	9	11
8	2	2	2	3	3	4	5	38	5	6	6	7	8	10	11
9	2	2	3	3	4	4	5	39	5	6	6	7	8	10	12
10	2	2	3	3	4	5	6	40	5	6	7	7	8	10	12
11	2	3	3	3	4	5	6	41	5	6	7	8	8	10	12
12	2	3	3	4	4	5	6	42	6	6	7	8	9	10	12
13	2	3	3	4	4	5	6	43	6	6	7	8	9	10	12
14	3	3	3	4	4	5	7	44	6	6	7	8	9	11	12
15	3	3	3	4	5	6	7	45	6	6	7	8	9	11	13
16	3	3	4	4	5	6	7	46	6	7	7	8	9	11	13
17	3	3	4	4	5	6	7	47	6	7	7	8	9	11	13
18	3	3	4	4	5	6	8	48	6	7	8	9	9	11	13
19	3	3	4	5	5	6	8	49	6	7	8	9	10	11	13
20	3	4	4	5	5	7	8	50	6	7	8	9	10	11	14
21	3	4	4	5	6	7	8	51	7	7	8	9	10	12	14
22	3	4	4	5	6	7	8	52	7	7	8	9	10	12	14
23	4	4	4	5	6	7	9	53	7	7	8	9	10	12	14
24	4	4	5	5	6	7	9	54	7	7	8	9	10	12	14
25	4	4	5	5	6	7	9	55	7	8	8	9	10	12	14
26	4	4	5	6	6	8	9	56	7	8	8	10	10	12	14
27	4	4	5	6	6	8	9	57	7	8	9	10	11	12	15
28	4	5	5	6	7	8	10	58	7	8	9	10	11	13	15
29	4	5	5	6	7	8	10	59	7	8	9	10	11	13	15
30	4	5	5	6	7	8	10	60	7	8	9	10	11	13	15

注：對於表中沒有出現的 n 值，計算 $z = \frac{(k-0.1n)}{\sqrt{0.09n}}$ ，其中 K 是正確回應數目。比較 z 值與標準常態分佈在 $\alpha/2$ 自由度的值，即，表 17.3 的最後一列 ($z_{\alpha/2} = t_{\alpha/2, \infty}$)

參考文獻

Anderson , M.J. and Whitcomb , P.J. 2005. RSM Simplified: Optimizing Processes Using Response Surface Methods for Design of Experiments, Connecticut: Productivity, Inc.

Bezdek , J. 1981. Pattern Recognition with Fuzzy Objective Function Algorithms, New York: Plenum Press.

Box , G.E.P. and Drape, N.R. 1987. Empirical Model Building and Response Surfaces, New York: Wiley.

Carr , B.T. 1989. "An integrated system for consumer-guided product optimization," in Product Testing with Consumers for Research Guidance, L.S. Cattell , R.B. 1966. "The screen test for the number of factors/ Multivariate Behavioral Research, 1:245-276.

Wu, ed., Philadelphia, PA: ASTM International, pp. 41-53.

Cochran , W.G. and Cox, G.M. 1957. Experimental Designs, New York: Wiley.

IFT. 1981. "Guidelines for the preparation and review of papers reporting sensory evaluation data," Food Technology, Sensory Evaluation Division, Institute of Food Technologists, 35:11, 50.

Cooper, H.R., Earle, M.D. and Triggs, C.M. 1989. "Ratio of ideals-A new twist on an old idea," in Product Testing with Consumers for Research Guidance, L.S. Wu, ed., Philadelphia, PA: ASTM International, pp. 54--63.

Draper , N. and Smith, H. 1981. Applied Regression Analysis, 2nd Ed., New York: Wiley.

Giovanni , M. 1983. "Response surface methodology and product optimization," Food Technology, 37:11, 41-43.

Godwin, D.R., Bargmann, R.E. and Powers , J.J. 1978. "Use of cluster analysis to evaluate sensory objective relations of processed green beans," Journal of Food Science, 43: 1229-1230.

ISO. 1985. "International Standard ISO 6658:1985," Sensory Analysis-Methodology-General Guidance, International Organization for Standardization, Available from American National Standards Institute, 11 West 42nd St., New York, NY 10036, or from ISO, 1 rue Varembe, CH 1211 Geneve 20, Switzerland.

ISO/CD 5492. 2004. "Sensory Analysis- Vocabulary".

Jacobsen, T. and Gunderson, R.W. 1986. "Applied cluster analysis/" in Statistical Procedures in Food Research, J.R. Piggott, ed., Essex, UK: Elsevier Science, pp. 361-408.

Larmond , E. 1981. "Better reports of sensory evaluation," Technical Quarterly of the Master Brewers Association of America, 18: 7-10.

Melilgaard, M.C., Civille, G.V. and Carr, B.T. 2007. Sensory Evaluation Techniques. 4th ed. CRC Press. New York.

Prel, P.A. 1976. "Preparation of reports and manuscripts which include sensory evaluation data," Food Technology, 30:11, 40.

, T. and Gunderson, R.W. 1986. "Applied cluster analysis/" in Statistical Procedures in

Food Research, J.R. Piggott, ed., Essex, UK: Elsevier Science, pp. 361-408.

李振登. 2013. 食品科學精要. 第三版. 偉明圖書出版社. 台北. 台灣.

區少梅. 2007. 食品感官品評學及實習. 二版. 華格那出版社. 台中. 台灣.

姚念周. 2001. 感官品評-基礎與應用. 樞紐顧問公司. 新竹市. 台灣.

劉伯康和莊朝琪. 2014. 食品感官品評-理論與實務. 新文京開發出版社. 新北. 台灣.