

台灣析數資訊系列叢書

統計學：SPSS分析導論

生活中的統計應用14篇

梁德馨 王智立 著

台灣析數資訊系列叢書

統計學：SPSS分析導論

生活中的統計應用14篇

梁德馨 王智立 著

視野無界・心智無限
Open Your Eyes, Stretch Your Mind.



華泰文化
HWA TAI PUBLISHING
since 1974

統計學：SPSS 分析導論 生活中的統計應用 14 篇

作者：梁德馨 王智立 台灣析數資訊顧問群

發行人：吳昭慧

責任編輯：謝茹堯

封面設計：謝茹堯 王素真

發行所：華泰文化事業股份有限公司

地址：台北市 11494 內湖區新湖二路 201 號

電話：(02)2162-1217

傳真：(02)8791-0757

網址：www.hwatai.com.tw

E-mail：business@hwatai.com.tw

登記證：行政院新聞局局版北市業字第 282 號

出版：西元 2014 年 1 月 初版

I S B N：978-957-609-934-2

基本定價：壹拾壹元壹角貳分

國家圖書館出版品預行編目資料

統計學：SPSS 分析導論 生活中的統計應用 14 篇／梁德馨

王智立 台灣析數資訊顧問群 著 — 初版. —

臺北市：華泰，2014.01

面：公分

ISBN 978-957-609-934-2（平裝）

1.統計套裝軟體 2.統計分析

512.4

102025094

著作權所有 翻印必究

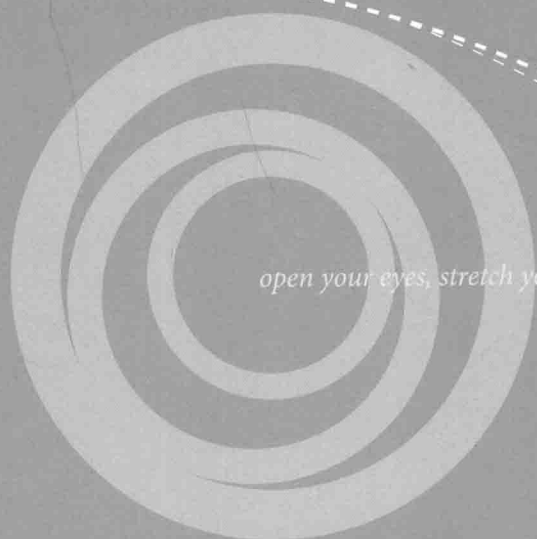
(本書若有缺頁、破損或裝訂錯誤，請寄回本公司更換。)



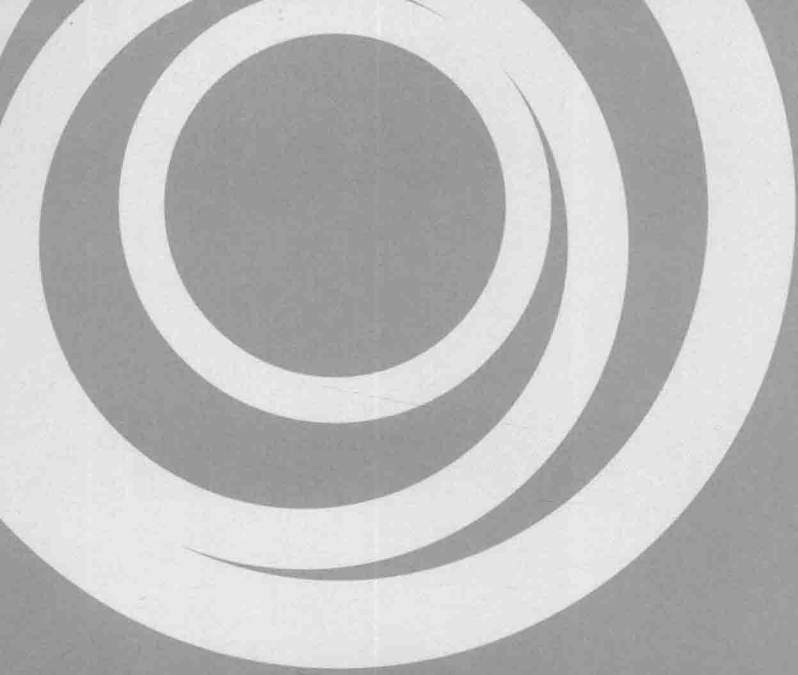
open

your eyes, stretch your mind

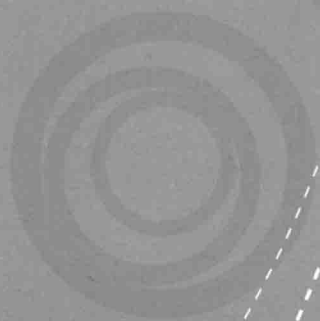




open your eyes, stretch your mind



Hwa Tai Publishing



推薦序

統計應用範圍廣泛，雖然市面上統計學的書已經是琳瑯滿目、五花八門，不勝枚舉，好友德馨也要出一本，我還是很高興。因為我知道她是學統計出身的管理學博士，一切講求實務應用，著重觀念，沒有不必要的公式證明，只求簡單易懂、好學好用。這是本書第一個特點，我稱之為「簡化」。

德馨在輔仁大學統資系任教多年，身兼統計諮詢顧問中心主持人，指導學生從事無數專案研究，了解學生學習困難，也知道產業界需要統計之處，故能以深入淺出的方式，將統計理論以業務實例加以闡釋。不僅學生容易明白，也體會到學統計的用途。藉由諸多的實例演練，配合軟體程式操作，達到活學活用的目的。這是本書第二個特點，我稱之為「活化」。

原子彈之父奧本海默 (Robert Oppenheimer) 和二十世紀最偉大詩人艾略特 (Thomas Stearns Eliot) 見面說：「我們科學家愛把複雜的東西簡單化。」艾略特回答說：「我們這類人剛好相反，喜歡把簡單的東西複雜化」。統計就是歸納，是要「化繁為簡」，找出複雜數據背後的真相。

有了這本著重觀念與應用的書，我相信你會喜歡上統計！

饒志堅

交通部統計處 副處長

2013.11.06

作者序

對統計學的第一次接觸是在大一時，坐在輔仁大學管理學院伯達樓一樓的大教室，聽著已故張照營老師的初統課，就深深地迷上了統計。當年，學到一個新的方法及觀念時，什麼實務經驗都沒有的我，腦子就不斷地轉著諸如：這個觀念在實務上應如何應用、資料應如何蒐集及建置……等問題。當年 18 歲時，將統計應用於實務上解決問題的理念，就已在心中萌芽了。而讓我走入統計應用殿堂的，則要感謝我在淡江管理科學研究所的指導老師張紘炬教授，在他的指導及提攜下，我開始了統計應用之旅。所以我的應用統計人生，啟萌於張教授，也奠基於張教授。

30 多年前的統計教學是十分理論及紮實的，而統計系也是出名的「當舖」。在那個人電腦都還在初步發展的時代，如果不會撰寫程式，就無法揮舞統計這把應用之器，也因此侷限了統計在實務上的發展。在那個理論掛帥的世代，我就堅信未來是統計應用的世界，因此，在 1994 年，為了協助進行數據分析的人能應用統計，我與友人合作寫了二本如何以 SAS 軟體進行分析的書。

隨著資訊科技快速發展，統計軟體的日漸親和使用，統計在政府及商業上也愈來愈仰賴數據來進行決策。尤其在 Big Data 的時代，統計在數據挖掘上扮演著舉足輕重的角色。然而，統計在不少教師及學子的心中，仍然是「很數學的」及「很理論的」一門高深學問。每當我告訴別人我是統計系的教授時，大部分人的第一個反應是「你數學一定很強，統計很有用但很難」。在我 20 多年來，近 200 個政府及商業的應用統計實務計畫經驗中，85% 都只用到了初等統計學的知識，只要能正確且活用這些知識，就可以創造出很好的決策數據。因此，為了能讓更多人輕鬆地以「做中學」的方式來學會及應用統計，我編寫這一本將統計簡化及活化的書。

時間飛逝，一眨眼離我的上一本書，已過了 20 年。在這 20 年間，我對統計之熱愛更深，對統計應用之推廣使命感更強。這一本書是採用在社會及管理學門很普遍的 SPSS 軟體，來帶領大家走上應用統計之路。其實，採用什麼軟體來做統計分析只是一種選擇，推廣統計之應用，讓更多人能揮舞統計，並活化統計之魂，才是我最核心的理念。

～成為應用統計專家，是我生涯發展的目標；
推廣統計應用，是我一生的職志。～

梁德馨

輔仁大學統計資訊學系 教授

2013.11

推薦序	i
-----	---

作者序	ii
-----	----

Chapter 1 緒 論

1.1 統計學的定義	2
1.2 變數的衡量尺度	5
1.3 如何運用統計軟體協助學習	6
習題	31

Chapter 2 統計圖表

2.1 類別型資料圖表	34
2.2 連續型資料圖表	36
2.3 雙變數統計圖表	39
2.4 範例	42
習題	74

Chapter 3 敘述統計

3.1 位置量數	76
3.2 離散量數	77
3.3 形態量數	79
3.4 Z 分數與盒形圖	79
3.5 柴比雪夫定理與經驗法則	80
3.6 範例	81
習題	95

Chapter 4 機率概論

4.1 樣本空間與事件	98
4.2 機率概念	101

4.3 聯合機率	103
4.4 邊際機率	107
4.5 條件機率與獨立事件	108
4.6 貝氏定理	109
習題	112

Chapter 5 離散型機率分配

5.1 隨機變數	114
5.2 離散型隨機變數	115
5.3 伯努力分配	116
5.4 二項分配	119
5.5 卜瓦松分配	122
5.6 超幾何分配	125
習題	128

Chapter 6 連續型機率分配

6.1 連續型隨機變數	130
6.2 均勻分配	133
6.3 常態分配	137
6.4 指數分配	141
6.5 其他連續型分配	143
習題	146

Chapter 7 抽樣分配

7.1 樣本平均數的抽樣分配： $\bar{X}$	148
7.2 樣本比例的抽樣分配： $\hat{p}$	149
7.3 樣本變異數的抽樣分配： S^2	150
7.4 中央極限定理	151
7.5 範例	152
習題	159

Chapter 8 估計

8.1 點估計	162
8.2 區間估計	164
8.3 母體平均數的信賴區間	167
8.4 母體比例的信賴區間	169
8.5 母體變異數的信賴區間	171
8.6 有效估計之樣本數	171
8.7 範例	172
習題	176

Chapter 9 假設檢定

9.1 統計假設之基本概念	178
9.2 型 I 誤差與型 II 誤差	180
9.3 母體平均數的檢定	182
9.4 母體比例的檢定	189
9.5 母體變異數的檢定	191
9.6 範例	193
習題	197

Chapter 10 兩母體假設檢定

10.1 兩獨立樣本母體平均值之比較	200
10.2 成對樣本母體平均值之比較	207
10.3 兩個獨立母體比例之比較	210
10.4 兩個獨立母體變異數之比較	213
10.5 範例	215
習題	222

Chapter 11 變異數分析

11.1 變異數分析簡介	224
11.2 變異數分析檢定原理	225

11.3 多重比較程序	228
11.4 範例	230
習題	235

Chapter 12 相關與迴歸

12.1 相關分析	238
12.2 簡單線性迴歸	242
12.3 線性複迴歸	252
12.4 範例	258
習題	269

Chapter 13 卡方分析

13.1 卡方檢定——獨立性檢定	272
13.2 卡方檢定——適合度檢定	277
13.3 範例	278
習題	284

Chapter 14 無母數統計

14.1 前言	286
14.2 符號檢定	286
14.3 隨機性檢定、連檢定	289
14.4 適合度檢定	291
14.5 等級總和檢定法	291
14.6 符號等級檢定法	294
14.7 獨立 k 組樣本之檢定	297
14.8 範例——SPSS 執行相關分析	298
習題	312

附 表 313

習題解答 323

緒論

1

CHAPTER

生活中的統計

想像你是……

1. 某電信公司的行銷研究員，最近要向主管進行一個客戶區隔的分析報告。
2. 某大學的學聯會主席，想要在編撰畢業紀念冊時，有一些不同的創意及新作法，所以你必須了解畢業生對於原有的畢業紀念冊及現有創意的一些想法及接受度。
3. 中華民國消費者文教基金會的職員，你想要做一個民眾對微笑單車的接受度及使用情況的專題。
4. 飲料公司的行銷部門主管，想要了解新推出飲料的廣告片在電視上播放的民眾接觸率、喜好度及對銷售提升的效果。
5. 某連鎖量販業的業者，想要了解各分店、各區域或各品類的銷售數字並進行比較。
6. A 銀行的風險部門想要針對借款倒帳顧客之基本背景特性，探討與一般顧客有何不同。
7. C 人壽保險公司的精算部門想要了解目前台灣地區的出生率，以及不同疾病患者或不同年齡層之死亡率。
8. 某公司的顧客關係管理部門主管，想明白現有會員的主要購買商品及付費行為。
9. 某地方政府的首長，想要了解該地居民對其施政的滿意度。
10. 交通部想要了解車流量與車禍次數之間有沒有什麼樣的關聯性。
11. 某藥廠研發部的研究員，想製作新藥的安全性評估報告，證明其與學名藥並無太大差異。
12. 析數農場的負責人，想要了解新的肥料對農作物產量的增進影響。

1.1 統計學的定義

1.1.1 統計學 (Statistics)

一般而言，對於資料的蒐集、整理、呈現、分析及解釋，並透過適當的機率模型，由抽取之樣本資料去推估母體未知之特性，希望能在不確定情況下，作成決策之科學方法稱之為統計學。統計學分為敘述統計學與推論統計學兩大類型。

1.1.2 敘述統計學 (Descriptive statistics)

探討如何簡化與表示現成之統計資料，稱為敘述統計學（又稱為描述統計學），亦即敘述統計學乃是僅就統計資料之蒐集、整理、呈現、分析與解釋部分，因此，其解釋範圍僅止於統計資料本身作討論。

1.1.3 推論統計學 (Inferential statistics)

探討如何利用抽取之研究對象的量化答案，去推估所有想要研究對象的答案稱為推論統計學。亦即如何以所抽出的樣本資料，經整理後去估計、檢定或預測母體中未知特性之科學方法，稱為推論統計學。推論統計學主要是探討樣本與母體間的關係，利用已知的樣本資訊推論到未知的母體特性，是從局部推論到全體的統計方法。

1.1.4 母體與樣本

人們在研究某一自然或社會現象及問題時，必須針對發生此一現象或問題的對象進行調查研究，調查研究的全體對象即是所謂的母體 (Population)。但是由於某些理由（如下所述），母體資料無法或不易取得，因此必須藉著抽樣方法 (Sampling)，由母體中抽取某些樣本 (Sample)，再針對這些樣本進行研究分析，以了解母體的特質。母體與樣本的意義可分別說明如下：

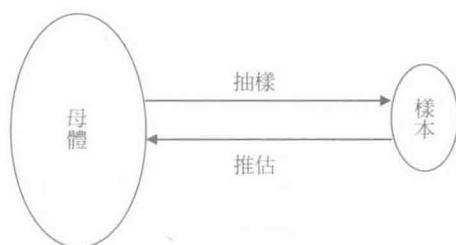
1. 母體

所謂母體是由 N 個具有某些共同特質 (Characteristic) 的元素 (Element) 或個體所組成的群體，是研究人員所要研究觀察的對象的全體集合。例如：想要針對台灣地區中低收入戶進行調查，其研究的母體為台灣地區家庭年所得在 100 萬元新台幣以下的所有家庭；想要了解台灣製造業廠商目前的能源使用狀況，則其母體為台灣地區

所有製造廠商；想要了解罕見疾病家人的精神壓力，則所有母體為所有罹患罕見疾病者之家庭。

2. 樣本

樣本是由母體中抽取部分個體而組成的集合，是母體的一部分。例如：從某大學企管學系二年級 100 多位學生的母體中，隨機抽取 15 位學生作智力測驗，這 15 位學生即是樣本；又如從 14 萬多家製造業廠商中，隨機抽出 1,000 家廠商，這 1,000 家廠商即組成一組樣本。



3. 母體參數 (Population parameter)

用以表達或呈現出母體之特性的一些數值稱為母體參數。例如：母體為台灣地區家庭年所得在 100 萬元新台幣以下的所有家庭，而這些家庭的平均年所得收入即為母體參數，這些家庭的雙薪家庭的比例也為一個母體參數，因為平均年所得（母體平均數）及雙薪家庭的比例（母體比例）皆是以數字呈現或表達出母體的一些特值。一般最常被使用的母體參數有母體平均數、母體變異數、母體標準差及母體比例等。而當母體數太大時，母體參數通常是未知的。

4. 樣本統計量 (Statistics)

用以表達或呈現出樣本之特性的一些數值稱為樣本統計量。例如：母體為台灣地區家庭年所得在 100 萬元新台幣以下的所有家庭，由其中抽出 1,000 個家庭，計算 1,000 個家庭的平均年所得收入即為樣本統計量，這些家庭的雙薪家庭的比例也為一個樣本統計量，因為平均年所得（樣本平均數）及雙薪家庭的比例（樣本比例）皆是以數字呈現或表達出樣本的一些特值。一般最常被使用的樣本統計量有樣本平均數、樣本變異數、樣本標準差及樣本比例等。在母體參數未知的狀況下，研究者常會以抽取樣本計算出樣本統計量的方式來推估母體參數。例如以 1,000 個家庭的平均年所得收入（樣本平均數）去估計所有平均年收入 100 萬元新台幣以下的家庭（母體參數）；用 1,000 個家庭中雙薪家庭的比例（樣本比例），去估計所有平均年收入 100 萬元新台幣以下的家庭中雙薪家庭的比例（母體比例）。

1.1.5 實驗與隨機實驗

1. 實驗 (Experiment)

舉凡各類「資料蒐集的過程」皆為實驗之一種。舉例來說，若進行一次民意調查，以問卷方式了解台灣地區民眾對電價上漲的看法，訪問了 1,068 位受訪者，則可以說，此次調查共進行了 1,068 筆的資料蒐集，亦即進行了 1,068 次的實驗；或者說，此次調查共進行了一系列的實驗，共針對 1,068 個實驗單位進行資料蒐集。

2. 隨機實驗 (Random experiment)

實驗若在相同的狀況下一再重複，但每次實驗均不能預知其結果，但所有可能的可能結果為已知，這樣的實驗就稱為隨機實驗。舉例來說，擲一個銅板 3 次，即表示進行了 3 次實驗，擲銅板的所有可能結果是已知的，只有人頭及花二種結果，但還沒有擲出銅板時，會出現哪一面是不確定的，所以此 3 次實驗即為隨機實驗。又再舉例，針對 1,068 位擁有智慧型手機的民眾，訪問其是否使用智慧型手機上網，所有可能答案有：「曾經使用上網但現在沒有、過去及現在都沒有上網、現在有使用智慧型手機上網」，但在還未訪問某一位受訪者之前，並不知道該位受訪者會回答什麼答案，所以表示執行了 1,068 次的隨機實驗。隨機實驗常能提供給我們資料以合理地推測不確定問題的答案。

如前述而言，一個隨機實驗中，每次實驗結果都不能預知，但所有可能出現的結果所成的集合是可預知的。對這個集合，稱之為樣本空間 (Sample space)。亦即樣本空間即為隨機實驗所有可能發生結果所成的集合。前述的二個例子其隨機實驗之樣本空間分別為：「花、人頭」及「曾經使用上網但現在沒有、過去及現在都沒有上網、現在有使用智慧型手機上網」。

3. 變數 (Variable)

隨機實驗可能針對每一個實驗單位去蒐集不同的特性，例如先前提到的訪問 1,068 位民眾其智慧型手機上網行為的實驗中，針對每一個受訪民眾（即實驗單位）可以詢問多個問題，例如：是否使用智慧型手機上網？每天行動上網的時數？是否使用 Line？……等問題，每一個問題皆為一個變數。

1.1.6 資料的蒐集方法

資料蒐集的方法可分為四大類，分別為次級資料、實驗設計、調查及觀察等四種方法。

所謂次級資料即表示資料之蒐集者為其他人，且可獲得的資訊為整理過後之數值、圖及表。其可能的資料來源有來自各政府發佈的統計報告及資訊、期刊或雜誌中已發佈的數據或圖表、書籍內的各類數據及圖表、網頁上的各類資訊……等。在進行產業分析或總體經濟環境分析時，常需蒐集此類資料。

所謂實驗設計法，是透過統計上的實驗設計法 (Design of experiment, DOE) 設計出實驗組及對照組，並按照計畫去進行資料的記錄及蒐集。在進行醫學、藥學、教學方法及教材適合度之研究中，較常採用此類方法去蒐集資料。

所謂調查法是以問卷形式，透過人員、電話或網路的各種不同可能溝通管道去蒐集資料，在各類分析及研究中都可能採用調查法。調查法又因其調查的範圍不同可分為二類，一是普查，一是抽樣調查。所謂普查是只針對所欲研究的對象（母體）全部進行調查訪問。所謂抽樣調查，是指由母體中抽出部分對象來進行調查訪問。

所謂觀察法，是指透過有規劃地去觀察研究對象，並有條理及系統的記錄下所觀察到的結果及數據。常用在兒童或動物的行為研究上。一般而言，觀察法也可用於人的行為研究，例如：消費者對商品不同陳列位置之注意度。

1.2 變數的衡量尺度

變數就是指所蒐集的資料。不同的資料特質就代表著不同的變數。例如：學生的性別、系別、成績皆是不同的資料特質（變數）。在蒐集資料時，必須加以測度與衡量，所謂衡量 (Measurement) 通常是用數字來記錄或表示所測得的事物或性質的量，以利統計分析。舉例來說，性別在記錄時以 1 代表男性，2 代表女性，又如成績的記錄是以 0 分為最低分，100 分為滿分（最高分）。衡量變數的特質或分類的工具，稱為衡量尺度 (Measurement scale)。換言之，資料的衡量尺度是給予資料一個實數值，作為計算或比較的基礎。

一般的資料分類粗略可區分為低階資料與高階資料，通常將名目尺度 (Nominal scale) 與順序尺度 (Ordinal scale) 的資料歸類為低階資料，而將等距尺度 (Interval scale) 與比率尺度 (Ratio scale) 的資料歸類為高階資料。四種衡量尺度說明如下：

1.2.1 名目尺度 —— 低階資料

名目尺度係為了標示目的而指定之數字，亦即當採用數字給予不同事物或類別代號，以方便辨認時，這類資料便稱為名目變數。以名目尺度測量下所得之資料，數字