學門與資源規劃報告

統計科學

行政院國家科學委員會

自然科學發展處

中華民國八十二年十二月規劃

中華民國九十四年十月修訂

目　　錄

摘　　要

學門規劃之主要目的是集結國內有限資源,規劃出對國家及本學門發展最有利之研究發展重點方向與分工合作模式,作為國科會推動研究之參考。本規劃報告盼能提供國內從事統計研究的人員，了解各領域在國內的發展現況與未來趨勢，並進而與國內實際環境配合，形成研究群，完成具有特色的研究，並提昇研究至國際水準。

八十二年國科會完成第一次的「統計學門規劃報告」，並在八十四年進行修訂，至九十五年間無再做修正，由於考慮到統計界發展日新月異，故重新集結本學門之各方意見，完成此一「統計科學學門規劃」。

本規劃報告格式參考了八十四年修訂之「統計科學學門規劃報告」，各領域的座談及意見徵詢、國內外發展狀況、國內本土需要、及是否具整合性、是否有較活躍的研究人力，擬定了統計科學學門現階段重點研究發展方向。

本報告參考 A Report on the Future of Statistics (Bruce G. Lindsay, Jon Kettenring and David O. Siegmund (2004)) 文中的分類法,將統計學門分為理論統計及應用統計兩大類，並在應用統計下，又細分成應用機率、工程及工業統計、生物統計與生物資訊、資訊科技相關之統計、地球科學及環境相關之統計、社會及經濟統計六大領域，此外，亦將人才培育納入本報告規劃的要點之一。

為了落實重點研究之推動，乃依據現有資源及人力配合狀況，擬訂現階段較可行的推動方法，主要為成立研究群，執行整合性計畫，促進研究之分工合作精神，加強資深研究人員與新進研究人員的互動，以加速研究人才之培育與研究成果質與量的提升。

並擬推動與其他學門的整合，配合國內現階段幾個重點科技之發展，推動相關統計發法之研究，並整合應用於相關領域。

由附錄之各項調查統計，可明顯看出國內統計科學研究人口成長快速，研究活動日趨活躍，研究成果亦有相當的進步，但在研究資源方面卻十分缺乏。為使重點研究之推動能有實際的發展與突破，充分的人力配備及其設備支援是必須的，而為加強統計科學在各領域之應用，合作聯繫費用的增加亦是需要的。我們盼望統計科學學門的研究經費能有合理的成長,這對學門的發展及與其他領域的互動與整合發展有重大的意義。

壹、學門規劃之目的

學門規劃之主要目的是集結國內有限資源,規劃出對國家及本學門發展最有利之研究發展重點方向與分工合作模式，作為國科會推動研究之參考。

學門規劃掌握的基本精神為重點研究與自由研究同等重要。對重點研究方向，國科會將協助推展，期能達到研究人員的臨界值量;而對優秀的自由研究，國科會亦同等重視，俾良性互補健康發展。

學門規劃之功能係將有限資源下原本無法進行的研發工作，藉分工合作方式，發揮綜合效益以提升研發層次與績效。

本規劃並著重於配合國內統計研究人員現況,因應國際發展趨勢及國內科技發展需求,擬定統計科學學門之重點研究方向。盼能提供國內從事統計研究的人員，了解各領域在國內外的發展現況與未來趨勢，並進而與國內實際環境配合，形成研究群，完成具有特色的研究，並提昇研究至國際水準。

貳、規劃修訂之情形

學門規劃委員名單：

召集人：傅承德。

委員：吳鐵肩、陳宏、曾勝滄、黃文璋、熊昭、樊采虹、鄭清水、盧鴻興（張源俊代）。

發展重點：

- 統計方法與理論
- 應用機率
- 工程及工業統計
- 生物統計與生物資訊
- 資訊科技相關之統計
- 地球科學及環境相關之統計
- 社會及經濟統計
- 人才培育

各領域負責人及發展重點：

- 吳鐵肩 (成功大學統計所)– 統計方法與理論、社會及經濟統計。
- 陳宏 (台灣大學數學所)– 統計方法與理論。
- 傅承德 (中研院統計所)– 統計方法與理論。
- 曾勝滄 (清華大學統計所)– 工程及工業統計。
- 黃文璋 (高雄大學統計所)– 應用機率。
- 熊昭 (國衛院生統組)– 生物統計與生物資訊。
- 樊采虹 (中央大學統計所)– 地球科學及環境相關之統計。
- 鄭清水 (中研院統計所)– 統計方法與理論、人才培育。
- 張源俊、盧鴻興(中研院統計所, 交通大學統計所）– 資訊科技相關之統計。

參、國內外研究現況

一、 統計方法與理論

在實證研究或實驗科學中，統計方法幾乎是不可或缺的。而量化研究又是眾多學門的主要發展方向，所以統計方法與理論的發展極為重要，近年來為能處理極大型資料及模型複雜系統，統計科學領域產生相當大及快速的變化。由於統計學研究的定位，乃在於發展正確及有效的統計方法，以利科學之前瞻研究。因應新的數據結構及新的計算工具，統計方法與理論皆有快速之進展，故核心理論之發展益發重要。舉例來說，由於新的資料型態及其特性，不偏估計或近不偏估計這些原則就須被重新評估；又由於多重假設檢定之數目可達到千計，傳統之 Neyman-Pearson 假設檢定之架構已經不再恰當，傳統由第一型式誤差所衍生之 familywise error rate 並不適宜，而改採偽發現率(false discovery rate)，此時多重假設檢定方法及其相關理論就須被重新評估。以下將分成迴歸分理論與方法、非參數與半參數模型分析、時間序列與空間統計、模型選取與識別、逐次分析、決策與貝氏理論、多變量分析及離散分析、及實驗研究等八個研究方向來討論。

I. 迴歸分析理論與方法

迴歸分析是統計方法中最重要的分析工具之一，在經濟、社會、天文、醫學等量化研究中幾乎都需要利用迴歸分析的方法，來分析數據及探索量與量之間的關係。迴歸分析相關的理論與方法雖已經過了長期的發展，但仍持續是此領域中最蓬勃的研究領域之一。

現在僅就國際上對迴歸分析的理論與方法學研究上的幾個新方向及國內研究現況，概分為半參數或非參數迴歸模式、反應變數或/及解釋變數數據是不完全觀測之情況下的迴歸分析、迴歸模式在特殊研究設計下的應用、模式診斷與選取、相依觀測值等五個主題加以說明。

(a)將傳統的參數迴歸模式推廣至半參數或非參數迴歸模式，如將線性迴歸或廣義線性迴歸推廣至部分線性模型（partly linear model）、廣義加性模型（generalized additive model）、迴歸樹（regression tree）、Cox 模型及線性轉換模型（linear transformation model）等，其後兩類模型之反應變數常為生存時間。其目的在於使分析的結果能建立在更少的模式假設上而增加推論結果之穩健性。半參數與非參數迴歸常需要深入的理論探討，且其大樣本性質常需要根據一些新發展的機率論理論體系，諸如鞅論（martingale theory）及經驗過程論（empirical process theory）。而在實際的執行上，也需要發展有效率的演算法及新的近似理論（approximation theory），因此引發相當大量的研究課題。國內部份統計學者也在此領域有所貢獻。

(b)在反應變數或/及解釋變數數據是不完全觀測之情況下的迴歸分析。這裡的「不完全觀測」包括了測量誤差（measurement）及缺失值（missing value）。傳統的測量誤差及缺失值問題較集中在線性常態模型。近來的發展則較集中於廣義

線性迴歸、半參數及非參數迴歸。特別是對不完全觀測的變數，近來的方法論傾向不對此類變數作分布上的假設，即使原本的迴歸模式為參數迴歸，因此形成一個新的半參數問題，其吸引了包括國內學者在內的許多統計學家加以探討。

(c)考慮迴歸模式在特殊研究設計下的應用。最顯著的例子是選樣與反應變數有關的研究設計（outcome dependent sampling design），這常見於流行病學或遺傳相關之研究，如病例—對照（case-control）研究設計。在此類研究設計下，樣本不再是一隨機樣本，因此傳統的迴歸分析方法極可能導致偏誤的結論。另一個例子是二階段研究設計（two-stage design），其導致部分個體的數據被不完全觀測，因此上述（二）中的問題在此亦會發生。此方向的研究在理論及實務上均有其重要價值，國內亦有若干研究人員進行相關問題的研究。

(d)模式診斷與模式選取。迴歸模式的統計推論建立在迴歸模式的假設之上。模式診斷的方法學提供檢視此假設是否合理的工具。針對新型態的迴歸模式及研究設計，需要發展相應的模式診斷工具。國內已有數位學者從事相關的研究工作。模式選取的方法學則提供選取合適模型的量化依據。在從事大量變數的迴歸分析時，模式選取是篩選重要變數的有效工具。與此相關的是維度縮減（dimension reduction）的方法學，其著重於篩選重要的變數維度（或變數的線性組合）。

(e)傳統迴歸分析的基本假設之一，是觀測值(observations) 之間互為獨立。 然而許多實際情形下，觀測值之間為相依的；如時空群聚資料、家族性資料等。目前文獻中處理相依觀察值之迴歸分析方法學主要分為兩類：隨機效應模式（random effects models）及廣義估計方程（generalized estimating equations），國內目前在此兩方法學領域均有學者進行研究，但人數相對較少。反觀國外，隨著時空資料及家族性資料等相依資料的日益增加，國際間對相依資料迴歸分析方法學的探討卻日益熱絡。

II. 非參數及半參數模型與方法

資料探索及理論模型建構係經驗科學的核心，而模型建構提供了進行資料探索及理論模型建構的方法，而非參數及半參數方法是近三十年來，統計科學發展最為快速的一個領域。現就此領域的國內(外)研究現況說明如下：

非參數統計模型的宗旨在於不對產生數據的未知母體做任何參數模型假設。其模型建立完全歸依於數據本身，而所根據的假設為適當的平滑程度。當參數模型的假設成立時，它簡化易解釋也有優越的估計有效性，反之則全盤皆墨。非參數模型如果平滑參數選擇恰當，往往能夠適切的呈現未知的母體，因而除了模型建構之外還有檢驗參數模型的功能。

早期計算機尚未發達，非參數模型停滯在直方圖（histogram）、核密度函數估計及核迴歸估計等原型。1980 年代至今，計算機發展一日千里，非參數模型與分析方法的計算瞬間即得，非參數模型強大的可塑彈性，加上社會與科學進展中產生大量構造複雜的數據，其中伴隨著各種問題等待有效合理的解釋，國際統

計學家投注大量精力專注研發非參數模型，至今成果斐然,已成統計學門主流領域之一。

非參數模型的建造方法早先有核平滑、樣條平滑（smoothing spline）、多項式樣條（polynomial spline）等。收斂速度、平滑參數的選擇、估計方法改善（如數據疏稀、降低變異量、減少誤差）等嚴肅的課題一直極受注目。另外，定性（如單調或俱唯一眾數）平滑函數估計與檢定、斷點或轉折點的估計與檢定、各處具不同平滑程度的函數估計、相關數據下的非參數函數估計、以非參數模型為基礎的分類法（classification）、群聚法（clustering）及主結構建立法、影像處理等等都是目前極度重要且受重視的研究主題。

高維度數據下，非參數模型受制於 curse of dimensionality，降低維度（dimension reduction）的方法與理論備受關心，當前的解決方法有 additive model, projection pursuit, sliced inverse regression, single index model 等方向，其中的估計與模型選擇仍待探討。

半參數模型主意於同時俱有非參數模型的彈性以及參數模型的解釋能力，在實際應用時廣受歡迎，目前受矚目的有 Cox proportional hazard model, varying coefficient model, partial linear model 等。隨著科技進展，目前泛函資料（functional data）與長期追蹤資料分析與歸論亟待解決，其中 within cluster correlation 與泛函主成份的漸進表現最為關鍵。

III. 時間序列與空間統計

在本報告中首先探討就時間序列及空間統計分別討論。

時間序列分析的研究方向，近二十餘年，由於受到許多應用領域的影響，已產生結構性的變化。核心課題已脫離傳統 ARMA（或 ARIMA）過程，轉向更複雜更具挑戰性的模型。此一發展趨勢，大致以下列三個方向為主軸。

(一) 非平穩序列

相關的問題像模型選取、單位根檢定、結構改變…種種議題，都在非平穩假設下，重新檢討。特別在計量經濟學裡，由於共積理論的提出，非平穩序列更是引發廣泛而深入的探討。

(二) 非線性或非高斯序列

諸種非線性模型，如 TAR，雙線性，ARCH（或 GARCH），紛紛被提出，而非高斯時的預測與遞迴計算法也有相當的進展。

(三) 長相關序列

傳統時間序列相關函數大多以指數速度趨於零，這和一些水文資料，電腦網路訊息流量，以及財務市場報酬序列的波動性等，具有長時間效應的現象不吻合。長相關序列的結構性研究已有了基礎，其線性模型統計推論課題的探討，亦漸趨完備。

此外，許多偏重計算方法的研究，如在貝氏架構下的 MCMC、小波、神經網路，以及最近的機器學習等也深受重視。

自一九九0年後，研究的重心更進而指向前述三大方向之間互相的結合，例

如非線性的 I(1)序列，非線性的長相關序列，或是非平穩序列與長相關序列的分辨等，都是目前及未來十年甚受重視及有待進一步研究的課題。

國際上較出色的時間序列研究群，散於美國、日本、德國及瑞士的大學或研究機構裡。國內的研究者對上述題材都有涉及，專業領域包括統計、經濟、財經、地科、工程等，在理論、方法或應用都有不錯的表現。

空間統計隨著各種精密量測儀器的開發，使我們對環境能做更廣泛且連續性地監測，在時間與空間中產生大量的環境資料，也導致環境科學及地球科學等領域對此類資料分析的需求日益增加。例如我們不僅對空氣污染在空間中的濃度有興趣，也想知道污染隨時間改變的狀況，類似的問題也出現在分析許多大氣、水文、農業、流行病及生態等資料上。然而時空資料通常在時間及空間上呈現複雜的統計關係，非傳統忽略時空關係的統計方法能有效處理。欲適當分析時空資料，而得到較精確的推論或預測，時空統計模式因而扮演關鍵的角色。

在本報告中就國內(外)研究現況、重要研究課題、及重要研究課題的推動進行說明。

IV. 模型選取與識別

由迴歸分析理論與方法、非參數與半參數模型分析、及時空統計此三各主題，皆須考慮到模型選取問題。在本報告中首先就國內(外)研究現況行說明：

1970 年以來，模型選取(model selection)這個議題持續受到各領域研究者的關心，幾乎所有要用到數學(統計)模型的科學，都無法避免這個問題。1980 年以前，統計學者在選模上的研究大多偏重在證(說)明赤池訊息準則(Akaike's Information Criterion, AIC)與交互驗證法(Cross Validation, CV)或 Mallow's Cp 的(漸進)等價性，以及證明貝氏訊息準則(Bayes Information Criterion, BIC)在選取時間序列模型及線性迴歸模型時的一致性。

然而，由於討論一致性需假設真實模型包含在候選模型中，為了避免這一個特殊的假設，1980 年以後考慮真正模型不必然包含在候選模型中的研究漸漸受到重視。Shibata (1980, 1981)在 *Annals of Statistic*s 及 *Biometrika* 上的兩篇文章開啟了這個方向研究的先河。他主要的想法是用複雜(高難度)的模型逼近真模(極複雜需用到無窮多個參數)，並容許逼近模型的複雜度隨著資料量增加而增加以降低逼近誤差。在此架構下，他證明了：AIC 及 Cp 最終可以選到預測能力最好(稱為漸進有效性)的逼近模型，但 BIC 傾向選擇過分簡潔的模型而不具漸進有效性，這個結果同時也對無母數迴歸中帶寬(bandwidth)選取的研究產生深遠的影響。奠基在解碼理論的研究上，Rissanen 在 1986 年提出了另一類廣被討論的選模原則，稱為累積預測誤差(Accumulated Prediction Error, APE)，它的特色是完全根據模型的線上(on-line)預測表現來選模。

國內近十餘年來在選模問題上的研究大多放在上述二大類選模準則上，也解決了一些頗為重要的難題。這兩年年來，在選模後的統計論上亦有一些好成績。然而隨著微陣列資料及龐大商業數據的湧現，大部分的選模方法卻面臨一個計算上的弱點，即是當解釋變數個數很多時，在所有可能的候選模型中尋找準則值最

小的模型，幾乎是不可能做到的。因此，一個能大幅減少搜尋模型次數，且具有一定程度的理論上優勢的選模方式，已隱隱然成為相關研究的焦點。

Tibshirani (1996)提出，Least Absolute Shrinkage and Selection Operator (LASSO)，是在上述需求下的開創性嘗試之一。透過 penalized least squares，LASSO 不僅可以避免 least squares 在高維度環境下的不穩定性，而且能將不顯著的參數自動估計為 0，達到選模的目的。換句話說，估計和選模是同時被完成的。最近，Efron 等人(2004)提出的，Least Angle Regression(LARS)，則是另一個重要的嘗試。LARS 最大的特色是能夠被快速運算，當訂出適當的停止法則時，LARS 並與 LASSO 等價，它與 Cp 也有某種程度的關聯。然而，LARS 在理論上的優缺點並未被深入討論。另一個相關研究是，在維度極高(觀察值的指數倍)的線性迴歸模型下，Buhlmann(2006)也證明了 L2-Boosting 選出來的模型可以某種速度逼近真實模型。

貝氏選模則是近年來另一個被注視研究方向，特別是配合馬可夫鏈蒙地卡羅法後，尋找後驗分配函數值最大的模型已不再因計算的困難而遙不可及。

V. 逐次分析

二次大戰期間，在武器的設計與改良上，需要以有限的試射次數來精確估計新武器之有效性，此一問題以統計術語來描述，即為如何以最小的樣本數檢定虛無假設或估計未知參數，由此而孕育出逐次分析的觀念。

Wald 於 1943 年提出 Sequential Probability Ratio Test (SPRT) 並與 Wolfowitz 證明其最優性，使得逐次分析的方法與理論受到重視而成為重要的研究領域，並影響到生物醫學、工程、經濟等應用學科，以下就四個方面簡要敘述一些相關發展。

(a)逐次假設檢定

在貝氏決策理論架構下，當取樣成本 (sampling cost) 小時 (對應到大樣本)，最佳終止邊界 (optimal stopping boundary) 可利用偏微分方程的自由邊界問題 (free boundary problem) 來決定其漸近行為 (asymptotic behavior)，此一漸近性質可在 frequentist 架構下，與 generalized likelihood test 結合而得到漸近最優檢定。

(b)臨床試驗 (clinical trials)

SPRT 之最優性是建立在虛無及對立假設皆為 simple 上，當對立假設為 composite 時，Armitage 等人提出 Repeated Significance Test (RST)，原始的 SPRT 與 RST 均為 fully sequential，在臨床試驗的應用上，必須考慮許多實際因素，故 Pocock 引進了 group sequential 觀念，而 Lan 與 DeMets 提出了 type I error spending function 以適當決定終止邊界，此外 ethical issue 亦為臨床試驗上困難而重要的課題。

(c)逐次改變點偵測 (sequential change-point detection)

品質管制與隨機系統控制的主要課題之一是逐次改變點偵測，早期的 Shewhart 控制圖簡單明瞭、容易分析，但由於未有效利用所有資訊，故對小

量改變無法即時偵測出來，有鑒於此，Page 經由 likelihood ratio 提出 CUSUM 檢定，並由 Lorden 等人證明其具有最優性，而在貝氏架構下導出的 Shiryaev-Roberts 偵測亦具有相似的最優性。

(d)Multi-armed bandit

在控制工程及經濟學中經常面臨如下的問題：有些決策(decision)有明顯的短期效益但卻無法獲得較多的資訊，而有些決策雖缺短期效益卻可對未知的情況有較多的瞭解，從而有可能獲得較佳的長期效益，multi-armed bandit 問題被視為這類問題的一個代表，主要的發展包括三十年前 Gittins-Jones 在幾何折現架構下成功的利用 Gittins index 來描述最優策略，而相當完整的漸近最優策略大樣本理論也陸續於近二十年內由黎子良等人完成。

史丹佛大學黎子良教授於 2001 年在中華統計學誌 (Statistica Sinica) 發表論文"Sequential Analysis: Some Classical Problems and New Challenges" (pp. 303-408)，對逐次分析的歷史及未來展望有精闢的見解。

國內在逐次分析方面的研究，包括傳統的漸近最優檢定與估計，最佳終止時間(包括 secretary problem 及 parking problem)，隱藏式馬可夫過程 CUSUM 改變點偵測最優性，與校正隨機漫步逼近自由邊界問題等。

VI. 決策與貝氏理論

決策理論認為每一個統計問題都是一種決策的問題，任何行動都可能帶來損失。在此一架構之下，可將統計推論轉化成以下的問題，依推論之目的確立一損失函數後，再基於某一原則，如最大風險最小化、貝氏風險最小化…等，找出最佳的對策，或找出可容的（admissible）對策。

決策理論的基本架構與早期發展，受到對局論（Game Theory）的影響很大。理論的演進，也由單一決策，發展成多重決策（排序與擇優）及逐次決策。在 60 年代以決策方法為底的漸近理論也發展起來，統計決策在極限實驗裡是否最佳，變成一種判別原則。到了 70 年代，探討多變量時最大概似法是否可容的 Stein 現象成為主導題材，到了 90 年代，這個問題又在非參數或半參數模型中的參數估計上重新得到新的生命力，而有蓬勃的發展。在 80~90 年代裡，條件推論（conditional inference）和半參數法的漸近最優理論亦曾風光一時。

近幾年來有一些國外學者在研究多重比較程序（multiple comparison procedure）的決策理論，而目前在國內仍有少數學者涉及排序與擇優及其運用和條件推論的研究。

在統計的許多領域中，均能見到貝氏或經驗貝氏的處理方法。貝氏統計對未知參數放上一先驗（prior）分佈，進而建構出貝氏意義下最佳統計方法。貝氏方法在結合主觀與客觀的資訊上提供了一個作法，此方法可進而提供未來數據的預測分析。

經驗貝氏法由 H. Robbins 在 1952 年的先啟工作提出了對貝氏法的一個修正，貝氏法中完全主觀的先驗分佈由過去的資料估計取代。經驗貝氏法中

對先驗分佈的假設，有參數型、非參數型的，其中參數型的經驗貝氏估計與 James-Stein 估計、ridge regression 估計均有極密切的關係。近年來經驗貝氏法更被應用到高維度的問題（如 microarray 的分析及 multiple comparisons）上。

多年以來國內在貝氏領域的相關研究有單參數及多參數的穩健貝氏法、貝氏統計計算、貝氏逐次理論等。關於經驗貝氏法的大樣本理論亦有許多研究，包括估計或檢定統計量之收斂速度，漸近最佳性，並針對選擇與排序（selection & ranking）等多重決策統計問題提出經驗貝氏的處理方法。此外小樣本性質、經驗貝氏與階層貝氏法之比較、生長曲線模型的貝氏法亦有人在研究。另外，在測量理論及醫學統計中，經驗貝氏模型受到相當重視並成功地被採用。

近十年內由於許多有效的 MCMC（Markov Chain Monte Carlo）抽樣方法被提出，再加上計算速度之突飛猛進，使得貝氏方法受到更廣泛的重視與應用。這些以 Monte Carlo 為基礎的抽樣方法主要包括 MC important sampling, Gibbs sampling, Hit and Run sampling, Metropolis-Hastings sampling 和 hybrid methods 等。

VII. 多變量分析及離散分析

如何處理大量資料？如何探索複雜系統及其模型之建構？如何處理不確定性？這三者是統計科學最核心的三個問題。統計科學的起源與生命及生存的大自然離不開關係，如由處理誤差所發展之最小平方法，這是與十八、十九世紀之天文學發展息息相關；而為了理解生命之延續，十九世紀的 Galton 爵士對於遺傳性狀之關連性，而開始了多變量分析－遺傳數據之相關係數分析。因著量化研究進入每一個領域，由於量化，如何分析多個變量之間的關係成為重要的科學問題，如在人類學及植物學上的分類促使區別分析(discriminant analysis)的發展，而對精神及智商分數的研究促使對因子分析(factor analysis)的探討。同時要對農業、工程及經濟上之多變量資料進行描述及探索，除使用高維常態外，也必須發展其它的多維分佈。這些都對二十世紀的前七、八十年間的統計發展提供了豐碩的動機及素材。而在二十世紀的最後三十年，因著計算機的發展，快速的計算及高速電腦的普及化，再加上大量資料庫的存在，使得多變量分析的進入嶄新的時代，也成為近年及未來發展最為蓬勃之重要領域，尤其這個領域已與資訊科學領域之機器學習產生相輔相成互相輝映之局面，在近年的多篇文章中，也因此探討是否會產生出新的 paradigm？及新一代統計人才的核心知識及工具等問題。

在離散分析部份，因數據為離散之型式，所以此領域之多變量分析，最早就以列聯表(contingency table) 形式出現在統計文獻上。雖然 Bartlett 有關交互作用檢定早在 1935 年即已出版，但對多維度之列聯表分析，直到計算機之普及化後才開始廣泛應用在醫學、流行病學及社會學科上。一般最常採用的模型是對數線性模型 (1oglinear model)與羅吉斯迴歸模型。

以下就十年前的規劃評估、國(內)外研究現況、加以說明：

(一) 十年前的規劃評估

台灣的統計發展的起步及茁壯正在二十世紀的最後三十年中，十年前的規劃正好驗證了這個發展。在該規劃中，多變量分析的重點研究方向為動畫法、有效維度簡化法、吉氏抽樣與重抽法在多變量的關係、生長曲線模型的參數推論與預測分析、方向資料、組合資料與時空序列分法、有效的逼近法、非精確檢定、穩健高維迴歸、跨領域的共同研究等九個方向。在離散分析方面的重點研究方向為對數線性模型及羅吉斯迴歸模型、樣本之 ROC 曲線分析、長期資料、重複測量資料與具測量誤差資料等模型、廣義估計函數之理論與應用、貝氏與經驗貝氏法在離散分析的應用、缺失資料的處理等六個方向。

在多變量分析中的吉氏抽樣與重抽法在多變量的關係、有效的逼近法及跨領域的共同研究，而在離散分析中的對數線性模型及羅吉斯迴歸模型、樣本之 ROC 曲線分析、長期資料，重複測量資料與具測量誤差資料等模型及缺失資料的處理，皆已成為重要的獨立研究課題。而在重點研究方向，在過去的十年中，除「吉氏抽樣與重抽法在多變量的關係」及「樣本之 ROC 曲線分析」外，國內的研究者在人數及研究成果上，皆有相當程度的成長及出色的結果。這不但說明當時的規劃確實能掌握到與世界脈動一致的研究方向，且國內也有適當的研究人力之投入。但國科會並未因著當初這些規劃方向而投入更多的資源，所以這些成果，當然歸諸於國內有相當比例的研究人員具有極佳的研究品味，故可與國際的重要研究課題維持同步，同時能維持國內在這些重要領域的研究活力。

在離散分析方面，由於數據型態、來源及其使用領域之多樣化，而對 flexible modeling 之需求極高，其中相當一部分已成為貝氏理論、非參數方法與廣義線性模型中的主流研究課題；而對於 ROC (receiver operator characteristic)曲線分析，在醫學診斷、regulatory agencies 及近日之生物晶片應用日益普及，已成為生物統計之重要研究課題；至於缺失資料(missing data)之處理，由於缺失資料在統計各領域經常發生，較適宜納入迴歸模型之主題探討。而連續型多變量分析中的生長曲線模型的參數推論與預測分析，更適宜於貝氏分析或長期資料中處理。

因這六個主題皆已成為重要的獨立研究課題，且可適宜的併入其他領域，故在本次的規劃中，不宜再列入於此主題中討論。而在跨領域的共同研究部份，也因發展之快速及其多樣化，也不宜於此討論，而置於本報告之跨領域部份闡述，方不易流於偏頗。再加上眾多數據之呈現，常是混合型態，故在本次之規劃，不再將離散分析單獨陳述。

(二) 國(內)外研究現況

在最近的十年，由於量化研究已更深入於各個研究領域，如在後基因體時代的生命科學、基因流行病學、神經造影（PET, fMRI）、大型資料庫、computational models、地球物理及環境科學、通訊、網路等，新的挑戰為不論在數據的變量（p）及其個數（n）都有極大幅度的成長，但更有趣的是變量數遠高於其個數（$p>>n$），這樣的問題是過往的統計分析所不太處理的。且數據之來源舉如天文、地理、物理、化學、醫學、農業、工業、商業等，這對從事統計核心理論研究者，也需要

較以往更多的與各科學領域之研究工作者進行溝通，方易於掌握問題之重心。

在多變量分析領域，國內目前主要的研究方向為 :

(a)資訊視覺化(information visualization):

為了從多變量數據中有效取得資訊，如何藉由計算機的幫助及認知，在螢幕上展示各種多維資料的處理，這改變了過去古典靜態展示敘據的傳統，也有助於探索多維資料。

(b)有效維度及數據簡化 (effective dimension and data reduction):

如何將高維度的數據投影到較低的維度而不損失過多訊息，一直是多變量分析中的重要領域。古典的主成分分析(principal component analysis)，因子分析(factor analysis)，多向度量尺法(multidimensional scaling)均可視為維度簡化的早期工作，這些想法也被引進於 functional data analysis 中；而八零年代的投影追蹤法(projection pursuit)及九零年代的反切迴歸法(sliced inverse regression)之延伸，持續引導有效維度之發展。此外在相對應分析(correspondence analysis)，分類樹形 (classification tree)，及 ACE(alternating conditional expectation)等方法，在變量數遠高於其個數的情況之下，其是否仍為有效的資料探索是當代最重要的問題。

(c)統計及機器學習(statistical and machine learning):

此一領域涵蓋了分類、迴歸、群聚等重要的三個方法，常用的手法包含決策樹、神經網路、Support vector machine、hidden Markov models、關聯法則(association rule)、graphical model、貝氏學習等。國內有一羣研究者形成研究群，已舉行過兩次研討會，也有出色的研究成果。但在學習理論的理論研究方面雖有不錯的結果，但人力過於單薄。

(d)方向性資料(directional data) 與組合資料(compositional data):

這些課題均與環境統計有關而日益成 為多變量分析之重要研究領域。

(e)非精確檢定 (non-exact testing):

由於應用上的需要 (如大量特高維度數據的出現)和理論上的準備,在理論上已經証明了非精確檢定通常會優於精確檢定。更多的理論工作待加強。

(f)穩健高維迴歸 (robust high dimensional regression):

在穩健迴歸分析中尋找新的統計方法，使得統計量既是穩健又是在通常許可的限制下，對於維度及樣本大小之比有較寬的容許度。

台灣目前在以上各方面都有研究人員之參與，同時與其他學門保持良好的互動及合作研究，在數個研究方向有相當大的成功機會。

VIII.實驗設計

在本報告中，我們將實驗設計的相關討論分為兩部份，分別置於 “統計方法與理論”與“工業統計”中。在統計方法與理論的部份，我們的討論重點著重在實驗設計的理論研究上。關於實驗設計在各種工業問題上的應用及未來的發展，請見工業統計。

實驗設計的理論發展，常隨著不同型態實驗的出現，而有新的研究課題及方

向。例如自早期的農業實驗，到後來的工業實驗、品質改進(田口方法)，乃至目前受重視的生物技術實驗及電腦模擬實驗等,皆對實驗設計理論的發展有深遠的影響。國內有關實驗設計的理論研究方向包括：區集設計、行列設計、迴歸設計及混合配方實驗等的最適設計問題、穩健參數設計、因子設計、各種設計的組合建構與計算機建構等。

最適設計(optimal design) 與因子設計為國內實驗設計理論方面頗為活躍的兩個重要領域。最適設計探討的問題為針對給定的統計模型及比較準則(如 *D*-、*A*-、*E*-criterion 等)，尋求將該準則最優化的設計。這方面的研究有近似理論(approximate theory)與精準理論(exact theory)兩個主要的方向。在近似理論(或稱連續理論)中為了數學運算上的便利,將設計視為實驗區域上的一個機率分布；所得的連續解與觀察數無關，但觀察數給定後可藉其獲得近似解。精準理論則對給定的觀察數尋求最佳解，因此處理的是一個離散的問題，代數與組合方法在此扮演重要的角色，應用到的工具包括群論、圖論、有限投影幾何學 (finite projective geometry)、 編碼學 (coding theory) 等。很多最適設計具有對稱性與直交性(orthogonality)，這類設計如拉丁方陣、BIB 設計、*t*-設計、以及直交表等均為組合設計或其應用，有關這些設計之建構，在過去的文獻中屢見不鮮，其中直交表為因子實驗中常用的設計。所以如何在組合設計中繼續發掘最適設計或高效率設計，亦是值得研究的方向。由於在工業實驗中重要的應用，因子設計的研究在近年來相當受到重視。

IX. 非參數與半參數模型分析

非參數統計模型的宗旨在於不對產生數據的未知母體做任何參數模型假設。其模型建立完全歸依於數據本身,而所根據的假設為適當的平滑程度。當參數模型的假設成立時，它簡化易解釋也有優越的估計有效性，反之則全盤皆墨。非參數模型如果平滑參數選擇恰當,往往能夠適切的呈現未知的母體，因而除了模型建構之外還有檢驗參數模型的功能。

早期計算機尚未發達，非參數模型停滯在直方圖（histogram）、核密度函數估計及核迴歸估計等原型。1980 年代至今，計算機發展一日千里，非參數模型與分析方法的計算瞬間即得，非參數模型強大的可塑彈性，加上社會與科學進展中產生大量構造複雜的數據，其中伴隨著各種問題等待有效合理的解釋，國際統計學家投注大量精力專注研發非參數模型,至今成果斐然，已成統計學門主流領域之一。

非參數模型的建造方法早先有核平滑、樣條平滑（smoothing spline）、多項式樣條（polynomial spline）等。收斂速度、平滑參數的選擇、估計方法改善（如數據疏稀、降低變異量、減少誤差）等嚴肅的課題一直極受注目。另外，定性（如單調或俱唯一眾數）平滑函數估計與檢定，斷點或轉折點的估計與檢定，各處俱不同平滑程度的函數估計、相關數據下的非參數函數估計、以非參數模型為基礎的分類法（classification）、群聚法（clustering）及主結構建立法、影像處理等等都是目前極度重要且受重視的研究主題。高維度數據下,非參數模型受制於

curse of dimensionality，降低維度（dimension reduction）的方法與理論備受關心，當前的解決方法有 additivemodel, projection pursuit, sliced inverse regression, single index model 等方向，其中的估計與模型選擇仍待探討。半參數模型主意於同時俱有非參數模型的彈性以及參數模型的解釋能力，在實際應用時廣受歡迎，目前受矚目的有 Cox proportional hazard model,varying coefficient model, partial linear model 等，隨著科技進展,目前泛函資料（functional data）與長期追蹤資料分析與歸論亟待解決，其中 within cluster correlation 與泛函主成份的漸進表現最為關鍵。

X. 時空統計

由於對周遭環境之關心、財務計量學之發展及生醫資料之蒐集等，多產生時空資料，而這些資料常在時間及空間上呈現複雜的交互影響關係，如何提出合宜的分析方法及模型，是近年來值得專注的問題。近幾年國內針對建構時空統計模式的研究包括以頻譜法構造不可分離時空互變異函數、多尺度時空模式、臭氧及其傳輸之時空模式、二元資料之馬可夫時空模式、及以小波 (wavelet) 建構之時空模式等。另外與時空統計相關的研究主題有「地震活動與前兆之統計分析」及「颱風降雨量與風速之統計預測」等問題。

隨著各種精密量測儀器的開發，使我們對環境能做更廣泛且連續性地監測，在時間與空間中產生大量的環境資料,也導致環境科學及地球科學等領域對此類資料分析的需求日益增加。例如我們不僅對空氣污染在空間中的濃度有興趣，也想知道污染隨時間改變的狀況,類似的問題也出現在分析許多大氣、水文、農業、流行病及生態等資料上。然而時空資料通常在時間及空間上呈現複雜的統計關係，非傳統忽略時空關係的統計方法能有效處理。欲適當分析時空資料，而得到較精確的推論或預測，時空統計模式因而扮演關鍵的角色。

過去分析時空資料主要有幾種方法：(1) 對個別地點採取分別或共同的時間序列分析；(2) 對個別時間點採取分別或共同的空間模式分析；(3) 以多變量時序列模式分析n 個地點所獲得的時間序列資料。前兩種方法的缺點在於忽略空間或時間關係，或需要以某種兩階段的方式分別研究時間和空間的關係，且僅適用於觀測點較規則的時空資料型態。第三種方法則針對空間中n 個固定點(或區域)有興趣的問題。

近十年來時空統計模式有著相當大的進展,許多與此相關的國際性會議也在世界各地進行。主要的時空統計模式可大致歸類如下：

1. **時空互變異函數模式**：

這類模式假設觀測資料由一個連續空間及離散時間的時空隨機過程採樣而得，資料間的時空關係主要運用時空互變異函數建構。早期的方法是簡單將時間互變異函數與空間互變異函數相乘得到一個具分離結構(separable) 的時空互變異函數。近幾年來針對具分離結構的時空互變異函數忽略時間與空間交互關係的缺陷，有相當多的研究在於開發具不可分離(non-separable) 結構時空互變異函數的構造方法，包括頻譜法、完全單調函數(completely monotone function) 法、隨機微分方程法、混合(mixture) 法、及捲積(convolution) 法等。另外有關於時空

互變異函數之平滑程度與其對應之頻譜密度函數關係也有一些理論結果。

2. **動態空間模式**：

這類模式將資料表示為時間序列的狀態空間(state space) 形式，可以分析在時間上具自迴歸結構且時空互變異呈現可分離關係的資料，也可以透過空間經驗正交函數(empirical orthogonal function) 分析具非平穩(non-stationary) 空間結構及不可分離時空互變異關係的資料。這類模式有計算上的優點，因為其概似函數及最佳時空預測皆可採用Kalman filter演算法計算。

3. **階層貝氏(Hierarchical Bayesian) 模式**：

此模式基本上將資料的統計關係分為三個階層建構，第一層以量測方程式表示資料與狀態變數的關係，第二層則對每一個地點的狀態變數以迴歸或時間序列模式表示其統計關係，第三層則對第二層模式的參數建構空間統計關係。這類模式具備相當彈性，可以分析連續(如溫度、雨量) 及離散(如某時段某地區流行病發生個數) 型態的資料。其後驗分佈通常以馬可夫蒙地卡羅(Markov chain Monte Carlo) 法計算。

4. **其他**：

分析時空事件(如地震、流行病案例) 之時空點過程(point process)模式。近幾年國內針對建構時空統計模式的研究包括以頻譜法構造不可分離時空互變異函數、多尺度時空模式、臭氧及其傳輸之時空模式、二元資料之馬可夫時空模式、及以小波(wavelet) 建構之時空模式等。另外與時空統計相關的研究主題有「地震活動與前兆之統計分析」及「颱風降雨量與風速之統計預測」等。目前時空統計的發展相較於純時間序列或純空間統計方法，尚未成熟且仍有很大的發展空間，在應用上時空統計分析的軟體更是極為缺乏。

XI. 實驗研究

國內方面較偏重於理論發展，研究方向包括：區集設計、行列設計、迴歸設計及混合配方實驗等的最適設計問題、穩健設計、設計組合與計算建構等。目前較受重視的計算機模擬實驗等所引進的新的實驗設計理論的發展，帶來新的研究方向。

XII. 模型選取與識別

1970 年以來，模型選取(model selection)這個議題持續受到各領域研究者的關心，幾乎所有要用到數學(統計)模型的科學，都無法避免這個問題。1980 年以前，統計學者在選模上的研究大多偏重在證(說)明赤池訊息準則(Akaike's Information Criterion, AIC)與交互驗證法(Cross Validation, CV)或Mallow's Cp的(漸進)等價性，以及證明貝氏訊息準則(Bayes Information Criterion, BIC)在選取時間序列模型及線性迴歸模型時的一致性。然而，由於討論一致性需假設真實模型包含在候選模型中，為了避免這一個特殊的假設，1980 年以後考慮真正模型不必然包含在候選模型中的研究漸漸受到重視。Shibata(1980, 1981)在Annals of Statistics 及Biometrika 上的兩篇文章開啟了這個方向研究的先河。他主要的想

法是用複雜(高難度)的模型逼近真模(極複雜需用到無窮多個參數)，並容許逼近模型的複雜度隨著資料量增加而增加以降低逼近誤差。在此架構下，他證明了：AIC 及Cp 最終可以選到預測能力最好(稱為漸進有效性)的逼近模型，但BIC 傾向選擇過分簡潔的模型而不具漸進有效性，這個結果同時也對無母數迴歸中帶寬(bandwidth)選取的研究產生深遠的影響。奠基在解碼理論的研究上，Rissanen 在1986 年提出了另一類廣被討論的選模原則，稱為累積預測誤差(Accumulated Prediction Error, APE)，它的特色是完全根據模型的線上(on-line)預測表現來選模。

國內近十餘年來在選模問題上的研究大多放在上述二大類選模準則上，也解決了一些頗為重要的難題。這兩年年來，在選模後的統計論上亦有一些好成績。

然而隨著微陣列資料及龐大商業數據的湧現，大部分的選模方法卻面臨一個計算上的弱點，即是當解釋變數個數很多時，在所有可能的候選模型中尋找準則值最小的模型，幾乎是不可能做到的。因此，一個能大幅減少搜尋模型次數，且具有一定程度的理論上優勢的選模方式，已隱隱然成為相關研究的焦點。

Tibshirani(1996)提出，Least Absolute Shrinkage and Selection Operator (LASSO)，是在上述需求下的開創性嘗試之一。透過penalized least squares，LASSO 不僅可以避免least squares 在高維度環境下的不穩定性，而且能將不顯著的參數自動估計為0，達到選模的目的。換句話說，估計和選模是同時被完成的。

最近，Efron 等人(2004)提出的，Least Angle Regression(LARS)，則是另一個重要的嘗試。LARS 最大的特色是能夠被快速運算，當訂出適當的停止法則時，LARS並與LASSO 等價，它與Cp 也有某種程度的關聯。然而，LARS 在理論上的優缺點並未被深入討論。另一個相關研究是，在維度極高(觀察值的指數倍)的線性迴歸模型下，Buhlmann(2006)也證明了L2-Boosting 選出來的模型可以某種速度逼近真實模型。

貝氏選模則是近年來另一個被注視研究方向，特別是配合馬可夫鏈蒙地卡羅法後，尋找後驗分配函數值最大的模型已不再因計算的困難而遙不可及。

二、 應用機率

應用機率此一領域，近十年可說進展很大。以美國數理統計學會(Institute of Mathematical Statistics)自 1991 年發行的 Annals of Applied Probability 為例，其篇幅逐年增厚，便可得知此領域的研究很蓬勃。當然歷史較悠久些 Journal of Applied Probability、Advances in Applied Probability 等應用機率方面的雜誌，也仍繼續吸引並刊登應用機率方面最新發展的文章。

廣泛地講，很多領域均可視為機率的應用。不過較側重統計的題材，如資料分析、估計、檢定等，不論是統計的理論或應用，通常不納入應用機率的範疇。我們將應用機率方面近年來的發展，約略分為排隊網路、財務數學、複雜隨機系統，及演算法隨機分析等四大方向。分別陳述如下。

I. 排隊網路

排隊網路是排隊理論中分析上最為複雜，應用卻也最為廣泛的模型。這個領域的第一根基石是在 50 到 60 年代初期由 R. R. P. Jackson 和 J. R. Jackson 等所奠下的，他們分析出具有乘積解(product-form solution)的排隊網路，學界便以 Jackson Networks 來稱呼此類網路模型。

在接下來的三十多年中，比較重要的理論結果有：Walrand and Varaiya 擴大了具有乘積解的網路種類、Kelly 利用可逆性(reversibility)研究一些有趣的網路模型，如對稱模型(symmetric queues)、閉網路(closed networks)等。另外，由於電話網路的應用，有限等候線(finite buffer)的網路模型之佔線機率(blocking probability)亦吸引了包括電機、通訊相關研究人員的興趣。有關 1990 年以前排隊模型的發展過程，可參閱 Wolff (1989)。

(a) 國內現況

過去十年來國內學者在排隊網路領域的研究成果十分豐碩，有超過 140 篇的學術論文刊登在 SCI/EI 的國際期刊上，其中大部份研究的議題與資訊科學有關：早先五年(1996~2000)最熱門的議題是探討 ATM(Asynchronous Transfer Mode)網路的控制與評估，近五年(2001~2006)研究的焦點則是無線網路(wireless networks) 與研究頻寬(bandwidth)的最佳分配與利用；十年來只有不到十分之一(11/142)的論文屬於排隊網路的基礎研究，如探討排隊網路的穩定性(stationarity)、收斂速度(convergence rate, asymptotic behavior)、敏感度分析(sensitivity analysis)等隨機性質。相較於國外這方面的基礎研究比例達五分之一(607/3118)，我們可以觀察到一些國內的研究現況：優秀研究人才比較集中於電機、資訊相關學門；應用科學的研究比基礎研究更具吸引力。雖然以上兩點與國內的產業環境密切相關，然而基礎研究畢竟是最上游的產業，需要國科會運用有限的研究資源來加以培植、保護。

同樣的失衡情況亦出現在模擬學(simulation)方面的研究。模擬學是研究機率、統計，和複雜的隨機模型(如排隊網路)之重要方法。事實上，模擬方法亦大量的應用在許多不同學門上的研究。十年來國內學者有超過 800 篇的研究論文使用到模擬方法，但是國內目前研究模擬理論(simulation methodology)的學者卻很有限。十年來所發表相關的研究論文約 20 篇左右，有產生隨機變數或估計值的有效方法，探討測度變換法(change-of-probability measure)對於模擬效果的影響，信賴區間的建構與統計意義，結合模擬與迴歸分析或是 bootstrap 抽樣法等研究。

目前國內有一個由「國科會數學中心」補助的應用機率研討會，每年春季輪流在政大、中興及東華等校的應用數學系舉辦，已辦了將近十屆。這個研討會的規模雖然不大，但是每次研討會的主題比較集中，主要是排隊理論(queueing theory)、隨機排程(stochastic scheduling)、模擬學(simulation) 與應用機率方法論(applied probability methodology)等。研討會每次都會邀請順道來台訪問的國外知名學者參加，而國內參加的學者通常都會與所指導的研究生同行，因此會期中除了老師之間(P2P)的交流，還有老師與學生(P2S)，學生與學生(S2S)之間的交流。

此研討會的另外一個特色是在議程安排上,參與者交換意見與研究心得的機會比較充分。因此幾年下來，這個研討會催生了許多合作研究的成果，成績斐然。

(b) 國外現況

近十年來排隊網路的主要發展方向，有下列幾個：在高使用率(heavy traffic)和多種類顧客(multiclass customers)時網路的穩定性(stability)、最佳控制法(optimal control)、優先次序法(priority rule)和隨機模型法(stochastic simulation methodology)等。

多種類顧客的網路模型主要是應用在半導體晶片的製造流程,而高使用率則是許多實際生產系統的現狀。近來對於這類問題有下面兩種較新的研究方法：

流體模型(fluid Models)是對於離散型的隨機網路模型的一個連續型、常數性的相似體。最早是由 Rybko and Stolyar (1992)，開始利用流體模型來探討研究排隊網路。由於它是建構在一組聯立方程式上，只需以相關隨機變數的期望值當做參數，因此，流體模型近年來已成為研究排隊網路最有效的工具之一，尤其是在探討排隊網路的穩定性方面。粗略估計，近十年在一些主要機率統計期刊中，至少有 35 篇直接相關的研究文獻。

布朗近似法(Brownian approximation)，也稱做 semi-martingale reflected Brownian motion (SRBM)，是以連續性的布朗運動來近似網路模型的隨機行為，用來估計網路模型等候人數和工作量等的衡態機率分佈(stationary distribution)。最早的代表作是 Harrison (1988)。這個方法的理論基礎是泛函中央極限定理(functional central limit theorem) ---亦稱擴散近似法(diffusion approximations)，和強近似定理(strong approximation theorem)。此一領域目前的研究重點是找出那些網路模型可以布朗近似法估計之。近十年關於以布朗近似法研究網路隨機行為的論文至少有 46 篇。

排隊網路另外一個重要的應用是電子通訊及電腦網路,研究的重點是找尋最佳控制法(optimal control)，或是找出瓶頸(bottleneck)，平衡分配服務資源(resource allocation)，解決網路壅塞的問題，達到最大的產出率(throughput rate)。近十年相關的論文有 30 篇以上。

優先次序法則是以排程(scheduling)或是定價(pricing)的方法，讓性質或是付費不同的客戶各自得到合理的服務。分析的方法主要仍是流體模型或是布朗運動模型。近十年與排程相關的論文高達 70 篇以上。至於與定價相關的論文雖然不多，有鑑於「使用者付費」的系統日漸普及，可以預見將會是未來十年的熱門議題之一。

當排隊網路的問題複雜到上述的幾種分析法都無效時,我們常訴諸的數值方法便是隨機模擬法。由於網路的組成繁複，建構模擬模型十分困難，電腦模擬所須的時間亦十分冗長，因此，開發有效可行的模擬方法是研究的重點，近年來發展出的方法有平行法(parallel simulation)、測度變換法、分解法(splitting method)等。近十年有關模擬排隊網路的論文有 100 篇以上。

II. 財務數學

在 1970 年代初期，Black, Scholes 與 Merton 三位財務經濟學家藉由 arbitrage 及 hedging 等基本觀念建構了原創性的動態資產訂價理論(dynamic asset pricing theory)，從而促成財務數學快速且蓬勃的發展，在這三十年的發展過程中，由於 filtration 的概念能夠適當且自然的用來描述經濟市場上資訊隨時間的變化(information flow)，機率論因而扮演關鍵性的角色。例如，Harrison, Kreps 與 Pliska 在 arbitrage-free 的條件下，推導出 equivalent martingale measure，對(frictionless) complete market 模型之任意 contingent claim，均可利用 equivalent martingale measure 來表示其價格。

在動態資產訂價理論架構下，選擇權訂價(option pricing)已被廣泛的應用到實務上，主要包括歐式選擇權及美式選擇權，另有各種較新奇、複雜的選擇權，如 barrier option,可作為資產分配的避險工具。歐式選擇權為一合約，明訂某一支股票為標的，價格 K(strike price, exercise price)與期限 T，當期限到時，合約的買方有權利以一股 K 元向賣方買標的股票(call option)或以一股 K 元將標的股票賣給合約的賣方(put option)。選擇權訂價理論可決定合約的合理價格(買方需付給賣方的價錢)，美式選擇權與歐式選擇權不同之處是前者允許買方在期限之前的任何時期均可執行其權利，故美式選擇權的價格較高，可表示成歐式選擇權價格加上所謂的 early exercise premium，牽涉到 early exercise boundary，而轉化為 optimal stopping 問題，再轉化為偏微分方程的自由邊界(free boundary)問題，可利用數值分析的 finite difference 方法求數值解，或者經由機率論的 random walk approximations 來估計美式選擇權價格及 early exercise boundary。

最佳投資組合(optimal portfolio)與消費選擇(consumption choice)亦為財務數學的重要課題，與隨機控制理論有密切的關聯，主要的數學工具包括 dynamic programming 及 viscosity solutions of the Hamilton-Jacobi-Bellman equation，對某些特定形式的 utility 函數，可得到明確的(explicit)最佳投資組合與消費選擇。

(a) 國內現況

國內在財務數學方面的研究人員多來自機率學界，屬於數學與統計學門，對財務金融的實務面較少接觸，由於人數少，有待進一步整合有限的人力與資源，並與國內外財務學家互動及合作；最近幾年的研究課題包括最佳投資組合與消費選擇，有交易成本之選擇權訂價理論，模型不確定情況下的最佳投資，隱藏式馬可夫模型在選擇權訂價的應用，校正隨機漫步逼近自由邊界問題在選擇權訂價的應用。

(b) 國外現況

財務數學由早期的 Black-Scholes 模型(即股價波動呈現幾何布朗運動)及 Cox-Ross-Rubinstein 二項式模型推廣到 Levy 隨機過程及有交易成本(transaction cost)的情況，選擇權及一般 contingent claims 訂價課題極為複雜，而尚無完整理論，相應的(perfect) hedging 觀念引伸為 superhedging, quantile hedging 及 hedging under constraints，並引進風險量度(risk measures, 例如 Value-at-Risk)，牽涉到的

機率論課題包括 martingale 與 semimartingale 理論，擴散過程理論，Levy 隨機過程理論，隨機控制與最佳化理論，backward 隨機微分方程理論及 Malliavin calculus，此外電腦模擬計算相關理論也是重要的研究課題。Mathematical Finance 與 Finance and Stochastics 為以財務數學為主的重要學術期刊，而 Annals of Applied Probability, Probability Theory and Related Fields, Journal of Applied Probability, Advances in Applied Probability, SIAM Journal on Control and Optimization, Mathematics of Operations Research 等期刊亦經常刊登相關論文。

III. 複雜隨機系統

複雜隨機系統是一個廣泛的領域，範圍包括特定應用的模型建構觀點和模型配適、估計步驟、模擬和計算的應用機率分析相關議題。尤其近十年來不斷進步的電腦強大計算能力更造成了統計與機率分析的變革。複雜隨機系統的觀點是在一個已知系統預測介入的影響。它區別了限制於 intervention 和 observation 的不同。應用的語言指出了其中的馬可夫性質；並且可推廣到隨機場(random fields)上。

(a) 國內現況

目前執行的研究計畫，包含：

(i) 狀態空間和隱藏馬可夫模型。這是用來建構不完整資料的模型和一些不可觀測的簡單馬可夫動力系統的雜訊函數。其應用包含工程、生物、財務和地球物理的時間序列。它包含一些應用機率上建構複雜隨機系統的觀點，及一些在統計物理的題材，即交互擴散作用無窮系統的大量空間-時間(large space-time)行為。近來發展的配適此模型的蒙地卡羅法如預測、濾波(filtering)和平滑法(smoothing)等方法。隱藏式馬可夫模型之有效濾波及平滑法之執行，一般而言，是針對非線性模型，而非線性模型實際上是不可實行的，因此，發展最佳濾波及平滑法之效率逼近法是有其必要性的。最近發展的粒子濾波(particle filter)為一有趣的課題。

(ii) 馬可夫蒙地卡羅術 (MCMC)已經是一個用來建構模型和分析複雜隨機系統的普及統計運算技術。這一領域的研究不僅是設計給想使用標準 MCMC 方法做分析以及合理化分析結果的人，而是那些使用他們自己的程式碼利用標準 MCMC 方法於一些新的應用上面。主要強調於了解此一方法的原理和模擬成果的主要概念。其所產生應用機率上的相關問題具有挑戰性。

(b) 國外現況

這方面的研究，包含：

(i). 時間序列模型，如隨機波動模型、GARCH 模型、跳躍擾動模型，經濟、財務、水文、流行病學及其他相關領域之應用。

(ii). 隱藏式馬可夫模型，(隨機)神經網路及 SVM 模型，在生物資訊方面之應用。

(iii). 其他議題：計算機網路、混沌模型、圖論模型、馬可夫蒙地卡羅術等。

更明確地說，從統計及應用機率的觀點考慮這些問題，即模型建立、參數估計和有效執行法及預測等方向著手，從而產生應用機率上的相關問題。

VI. 演算法隨機分析

我們先說明何謂演算法分析。由於計算機在各科學領域之應用幾乎是無遠弗屆，它所伴隨而來的演算法設計與對應的分析，便成為一重要課題。要解決一個組合問題，經常有不同的演算法，這些不同的演算法之間孰優孰劣，何時某法優及何種條件下另法較佳，便需進一步的模型假設與數學分析。而由於計算機本質上只能處理離散的資料，自然地在進一步的分析中，不可避免地引出一些離散組合結構的研究。所以演算法分析所需要的工具來源除了原出發問題外，亦廣義的涵蓋了組合、機率與分析。

在美國數學學會(AMS)的數學學科分類中，與“algorithm”一辭有關的類別包含邏輯(02E10)、圖論(05C85)、數論(10A,10K,11K,11Y)、代數結構(16A)、符號計算(33F10)、離散幾何(52B55)、數值分析(65D,65F,65G,65K,65V,65Y)、計算機(68A,68C,68Q,68W)與流體力學(76M27)，可謂牽連甚廣。其中演算法分析被歸類在 68Q25 與 68W40 (2000 年後新增)。底下的統計數字可粗略地反應出演算法分析的重要性(搜尋日期為 2006.3.28 日)，在 AMS Math Reviews 所收錄的文章中共有 108317 篇其類別歸入 68(計算機)，而這其中有 12288 篇(約 1.1 成)與 68W40 或 68Q25 有關。17350 篇是 2000 年後發表，這些文章中有 3557(約 2 成)篇歸入 68W40 或 68Q25。另一方面，在美國計算機組織(ACM:Association for Computing Machinery)的分類下，演算法分析歸在 F2，其下細分為數值性與非數值性的演算法(承襲 Knuth 較傳統的分類)。而任何演算法皆免不了要選用適當的資料結構。所以另一大類 E(資料結構及其他相關)便與演算法密不可分。

在演算法分析及相關文獻中，絕大多數的分析模式皆為非隨機且結果僅止於所謂的“大 O 分析”。傳統上，隨機分析需較多的數學工具與模型假設，所以相關的分析在文獻上並不多。但近年來極大量資料在各領域的處理，使得較精確的漸近分析更形重要。同時在諸多數學工具的發展下，更細的隨機分析也得到許多重視。更重要的是演算法所引出的一些隨機問題，不但提供傳統機率學家一全新的方向，且經常極具挑戰性。同時很多新現象的多樣與普通性，使得相關研究愈加有趣。

(a) 國內現況

台灣計算機與組合界研究演算法的人數相當多，但研究其隨機行為的卻極少數。底下的統計數字可粗略地反應出大致的現狀。在 AMS 的 MathSciNet 資料庫中，類別為 68(計算機)且至少有一作者的任職單位在台灣的文章數目共有 1085 篇(搜尋日期為 2006.3.28 日)，其中與 68Q25 或 68W40 相關者有 19 篇。這其中僅 27 篇含“average”或“random”兩個關鍵字。如何吸引更多人才從事相關或合作研究，是未來亟待努力的方向。

(b) 國外現況

演算法隨機分析自 1960 年代初期由 Kunth 開端，目前公認的世界權威，非法國的 Philippe Flajolet 莫屬，他在過去 20 餘年的工作中，成功地結合了組合、分析與機率上的技巧，提出許多演算法分析及相關領域上系統性的工具（很多工

具甚至已自動化，可由其團隊開發的軟體來處理）。在他的引導下，許多人皆投入此研究領域。而自 1993 年開始舉辦的“演算法分析會議”（早期兩年一次，1997 年後每年一次），今年已辦第 11 次。每年大約有 50 至 90 人與會（視場地大小決定邀請人數）。除了兩次在美國(Princeton 及 Berkeley)外，其餘皆在歐洲，充分說明其本質上較歐化的特性。這十幾年來，最大的改變便是新問題與新人才的加入，使得研究社群更加擴展。

三、 工業統計

它是工業界用來處理產品研發、製程監控及售後維修服務等所衍生的相關品質問題之所有數據分析方法的總稱。利用這些統計方法配合工程師專業知識，將可有效地改善產品品質/可靠，進而提昇產品競爭力。因此工業統計是目前工業界十分重視的數據分析工具。

工業統計的研究領域十分廣泛，但主要包括線上品管與監控 (on-line quality control and monitoring) 、線外品管改善 (off-line quality improvement) 及可靠度分析等應用領域。其中線上品管與監控 (又稱品質管制) 包括機台設備與製程監控、統計製程管制、回饋與控制及製程能力分析。線外品管改善 (及實驗設計) 包括參數穩健設計及反應曲面方法。而可靠度分析包括設限及加速壽命模型、衰變模型、系統可靠度及可維修模型等。此外國外新近的工業統計發展，尚有機台缺失，偵測及分類、剖面資料製程監控、高可靠度統計推論、奈米元件可靠度分析、軟體可靠度、不規則實驗區間、貝氏方法等，茲分述如下：

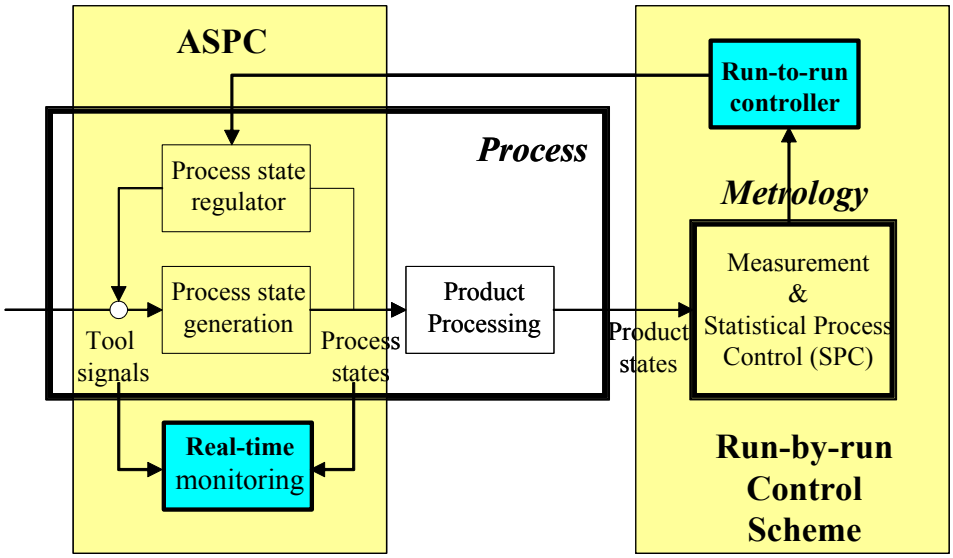
I. 線上品質管制與監控

在傳統的品質管制領域中，EPC (Engineering Process Control) 和 SPC (Statistical Process Control) 通常是分屬於兩個不同的應用與研究範疇。前者主要應用於連續製程產業 (process industry) 如化工業，此方面之研究則由化工與控制領域的學者所主導；後者主要應用於零件製造業 (parts manufacturing industry)，研究則由工業工程及應用統計領域的學者所主導。MacGregor (1988) 與 Box & Kramer (1992) 等學者開始討論兩者的異同與可能的整合。1990 年代起，由於高科技產業如半導體與液晶面板製造產業同時擁有連續製程產業與零件製造業的特性，研究學者開始試著整合兩者，Vander Wiel 與 Tucker 等學者首先提出了 ASPC (Algorithmic Statistical Process Control) 的觀念。而針對半導體產業，Ingolfsson & Sachs (1993) 等人亦提出批次控制(Run-to-Run Control, R2R Control) 的架構。2000 年起，在半導體製造的研究領域更興起了 AEC/APC (Advanced Equipment Control/Advanced Process Control) 的風潮，AEC/APC 架構進一步整合了 ASPC 與批次控制如圖一所示。

(a) 機台設備與製程監控

AEC/APC 之架構與傳統 EPC 或 SPC，在觀念及做法上有兩個很重要的改變，一是將傳統製程回饋調控的 EPC 觀念用於後製程量測 (metrology) 後，回

饋 (feedback) 至製程自身或前饋 (feed-forward) 至下一個製程配方 (recipe) 的調節，此又稱為批次控制。另一則是將傳統製程量測後的 SPC 用於機台設備的即時監控，又稱為機台缺失偵測與分類 (Fault Detection and Classification, FDC)。除此之外，傳統的 SPC 在整個架構中亦持續扮演不可或缺的角色。



圖一、Advanced Equipment/Process Control 架構

(b) 統計製程管制 (Statistical Process Control)

統計管制圖 (control charts) 約可分為下列三大類:

(i) Shewhart 管制圖

分為計量管制圖和計數管制圖。計量管制圖以 $\bar{X}$, R 和 S 管制圖為主；而計數管制圖則以 np, p, c 和 u 管制圖為主。

(ii) 記憶型 (Memory) 管制圖

主要分為 CUSUM (Cumulative Sum) 和 EWMA (Exponentially Weighted Moving Average) 管制圖。這兩種管制圖對於製程平均數或變異數發生微小偏移之監控工作，比 Shewhart 管制圖具有更佳的偵測能力及靈敏度。

(iii) 調適型 (Adaptive) 管制圖

不同於 Shewhart 和記憶型管制圖，調適型管制圖的管制參數是由前一次製程資料所決定，因此為一動態管制圖。這種管制圖被證實對製程變化之監控具有較高的靈敏度，主要有 VSI (Variable Sampling Interval), VSS (Variable Sample Size), VSSI (Variable Sample Size and Interval) 和 VP (Variable Parameters) 等管制圖。

在前述管制圖中，如何決定適當樣本數、抽樣間隔及管制界限是目前此領域極重要之研究課題。一般而言，管制圖設計主要可分為統計設計 (statistical design) 和經濟設計 (economic design)。統計設計通常是以 ARL (Average Run Length) 和 ATS (Average Time to Signal) 為準則，來決定管制圖中的相關參數值。經濟設計

是以成本最優化的觀點來考慮管制圖的參數設計。由於管制圖經濟設計通常有較高的錯誤警訊率 (false alarm rate) 或較低的偵測能力，因此有些學者建議在模式中加入一些統計限制條件，這就是所謂的經濟-統計設計 (economic-statistical design)（詳見 Montgomery, 2005）。

(c) 批次控制 (Run-to-Run Control)

批次控制的發展始於 1990 年代初期，2000 年代批次控制又被稱為先進製程控制(Advanced Process Control, APC)。過去的研究發展主要是針對製程的投入及產出 (Input-Output, I-O) 模型為單一投入及單一產出 (Single-Input and Single-Output, SISO) 或多重投入及單一產出 (Multiple-Input and Single-Output, MISO) 系統，在製程是否漂移 (drift) 或老化之限制下，先建立製程產出之預測模型並分別探討 single, double EWMA 控制器的穩定條件 (stability conditions) 及其最適折扣因子之挑選。除此之外，當 I-O 模型為一動態系統及非線性系統時，此控制器產出的穩定條件將極複雜，此時大多改採用自我微調 (self-tuning) 及遞迴最小平方法 (recursive least squares) 來監控製程。

(d) 製程能力分析

在上述 (b) 和 (c) 中所討論之研究大多是針對機台設備與製程是否穩定而進行監控或分析；另一有關製程的重要研究課題為製程能力之分析 (process capability analysis)。品質保證攸關一個產業的生存與發展，為了使製程能符合顧客的需求以保證產品之品質，製程必須具有產出符合工程規格零件的能力。美國品質協會對製程能力的定義為：「對一指定特性的固有製程變異性 (inherent process variability) 的統計量測」。 一般業界都採用製程能力指標作為製程表現的量測標準，而此標準是依據製程能否達成規格要求為基準。

常用的製程能力指標有 C_p, C_{pk}, C_{pm}, C_{pmk} 等；其中 C_p 只與規格上下限與品質特性 (quality characteristic) 之標準差有關，C_{pk} 同時考慮了品質特性的平均值，而 C_{pm} 和 C_{pmk} 則再加上平均值是否偏離目標值的考量。初期的研究大多假設常態分配、雙邊規格且目標值位於規格正中間；爾後研究逐漸推廣至種種不同情況。

目前此領域之研究包括下列主題：

(i) 製程能力指標之估計與檢定
(ii) 貝氏製程能力分析
(iii) 不同製程能力指標之比較
(iv) 製程規格為非對稱時之製程能力分析
(v) 製程資料為非常態分配時之製程能力分析
(vi) 製程資料有相關性時之製程能力分析
(vii) 製程資料為混合分配時之製程能力分析
(viii) 製程資料為多變量時之製程能力分析
(x) 製程資料有量測誤差時之製程能力分析

相關文獻可參考 Kotz & Johnson (1993), Kotz & Lovelace (1998) 和 Pearn &

Kotz (2006)。

II. 可靠度分析

可靠度 (reliability) 又可稱為時間導向品質 (time-oriented quality)，它是消費者評估是否採購某特定商品之重要考慮因素，因此生產者在推出任何新的產品(或系統) 時，經常面臨的決策問題是如何提供消費者有關產品可靠度相關資訊。一般而言，產品可區分為不可維修 (non-repairable) 及可維修 (repairable) 產品。具體而言，不可維修產品所面臨的可靠度決策問題是如何執行壽命測試實驗 (lifetime testing) 以便快速地提供消費者有關其產品或系統的平均壽命 (Mean Time To Failure, MTTF) 或 50 百分點壽命 (Median Life, ML) 等相關可靠度資訊。而可維修產品所面臨的問題是如何估計產品的現場 (field) 失效率、可用度 (availability) 及其維修和更換策略。下面我們將分別從此二種不同角度來回顧國內外可靠度研究之現況。

(a) 設限及加速壽命測試 (不可維修產品)

在進行壽命試驗時，受限於試驗時間，僅能收集到部分不完整的實驗數據，我們稱此類型資料為設限資料 (censored data)。典型的設限資料有型 I 和型 II 設限資料 (type-I and type-II censoring)。Lawless (1983) 曾對此領域之研究做非常詳盡之介紹。近年來有許多學者如 Balakrishnan (2006) 將此一階段設限資料擴展為多階段的逐步設限 (progressively censoring) 或更一般化的逐步設限 (general progressively censoring)，例如逐步區間設限 (progressively interval censoring)，type-I and type-II 的混合設限 (hybrid or mixed censoring) 等。另外，在方法論方面，也由原來的參數推論及無母數推論擴展到半參數推論 (semi-parametric method)。

隨著科技水準的提升，現今的工業產品大都具有「高可靠度」之特性，在合理的試驗時間內欲觀測到產品壽命之失效資料，將愈來愈不可能。此時較可行的做法是將產品置於較高環境應力下進行壽命實驗，並藉由產品壽命中之未知參數與應力變數的物理/化學關係式，來推估產品在正常使用條件下之壽命分配，這就是所謂加速壽命試驗 (Accelerated Life Test, ALT)。有關 ALT 的參數模型之統計推論工作，必須先建構產品的壽命-應力 (life-stress) 之統計模型。常用工業產品的壽命分配有 Weibull 及 Lognormal 分配等; 另外常見的環境應力有以溫度為加速變數的 Arrhenius Reaction 及以電壓為加速變數的 Inverse Power 模型。因此典型的產品壽命-應力統計模型有 Lognormal-Arrhenius 及 Weibull-Inverse Power 模型等。此外若同時以溫度及電壓為雙加速變數的產品壽命-應力統計模型則有 Lognormal-Generalized Eyring 模型等。

在前述的產品壽命-應力統計模型下，典型的統計決策問題是有關加速壽命實驗之最適設計與分析，亦即如何決定較高環境應力變數之適當水準及在給定之應力水準時，決定適當的樣本數及試驗時間 (life-testing time)，以便在最低試驗成本之條件下，獲得產品在正常應力時之最精確的 MTTF 或 p-th 百分點等估計值。有關加速壽命試驗的相關統計決策問題可參閱 Nelson (1990) 及 Meeker & Escobar (1998)。

(b) 衰變模型 (不可維修產品)

高可靠度產品之壽命推論除了加速壽命模型外，亦可改用衰變(degradation)模型及加速衰變(accelerated degradation) 模型來估計產品壽命相關資訊。所謂「衰變模型」是假設此產品存在與其可靠度有高度相關之品質特性，它隨時間會逐漸衰變；若其衰變量達到某一給定水準，則可判定此產品失效。典型的例子如日光燈及 LED 之壽命推估，它可藉由其亮度衰變到原始亮度的一半值時來獲得。

有關衰變模型及加速衰變模型之統計推論，必須先建構產品的衰變統計量及其第一次通過臨界值 (First Passage Time, FPT) 之壽命分配。一般常用線性或非線性模型來描述產品的衰變路徑，典型例子如：混合效應(mixed-effect) 模型以及隨機過程為主的推論模型，如 Wiener process 或 Gamma process 等。其中混合效應模型的優點是可考量產品具有 unit-to-unit variation 之現象；而隨機過程模型的優點是可適當地描述衰變資料隨時間具有相關之現象。此方面的重要參考文獻有 Lu & Meeker (1993), Lawless & Crowder (2004) 等。

典型衰變模型的統計決策問題是如何在最低試驗成本之限制條件下，決定衰變試驗之最適配置，以便獲得產品的 MTTF 或 p-th 百分點等之最精確估計值。衰變試驗的實驗配置有下列三個重要參數需考慮：樣本個數、觀測頻率及終止觀測時間。此方面的重要參考文獻有 Boulanger & Escobar (1994), Yu & Tseng (1999) 等。

(c) 系統可靠度分析 (可維修產品)

近年來，有關系統可靠度之研究，大致有下列幾點：

(i). 任何系統之可靠度取決於個別元件的可靠度及其組成結構。在實務上，當計算複雜系統之可靠度頗為費時之情況下，如何獲得該系統可靠度之較精確的上下界或其信賴區間，將是此領域之重要研究課題。參考文獻見 Barlow & Proschan (1981) 與 Meeker & Escobar (1998)。

(ii). 在有限資源限制下，如何經由某些元件可靠度之加強 (包括並聯複組 (active redundancy) 與備份複組 (standby redundancy)) 等來獲得整個系統可靠度之最優化是一個相當重要的研究課題。除此之外，元件重要性測度及排序 (component importance measure and ranking) 亦可提供我們挑選元件之參考，關於這方面的介紹可參考 Rausand & Hoyland (2004)。

(iii). 當系統元件的壽命分配彼此之間並非互為獨立時，如何建立適當的模型來描述元件的相關程度是近年來一個相當受重視的研究主題，見 Barlow & Proschan (1981), Rausand & Hoyland (2004)。

(iv). 當傳統的二元 (binary) 系統模型不足以描述現實上面對的狀況時，我們可用近年來慢慢建立的多元可靠度系統 (multi-state system) 理論來進行分析。Lisnianski & Levitin (2003) 對這方面的研究有相當詳細之介紹。

(d) 可維修模型 (可維修產品)

前述不可維修產品之壽命數據分析通常假設所處理的失效資料為獨立且同佈分配。但是對可維修產品在很多情況下，失效資料是相依而非同佈分配，此時

可靠度工程師所重視的問題是推估該產品的失效率函數及可用度等相關資訊。由於多次重複失效事件是可維修模型經常面臨的實際狀況，因此可採用隨機模型之重現數據分析 (Recurrent Data Analysis, RDA) 來處理。

重現數據的分析方法可略分為參數和無母數方法。Lindqvist (2003) 將母數方法依不同失效率函數之型態歸納為：Homogeneous Poisson Process (HPP), Nonhomogeneous Poisson Process (NHPP), Renewal Process (RP), Perfect Renewal Process (PRP), Trend Renewal Process (TRP) 等。傳統上，一般採用 PRP 模型來處理完全可維修系統的數據，而用 NHPP 來處理最少維修問題，近年來則採用更一般化的 Superimposed PRP 及 TRP 來處理較複雜的維修問題。

無母數方法則由 Nelson (1988) 提出平均累積函數 (Mean Cumulative Function, MCF) 的估計量。Nelson (1995) 更進一步提供 MCF 最小變異不偏估計量，並介紹如何計算 MCF 之信賴區間。另外 Lawless & Nadeau (1995) 亦提供有關 MCF 非負數之偏估計量。同時亦考慮到具有共變數(covariates) 的半參數迴歸模型重現數據的分析。有關無母數方法可參閱 Nelson (2003)。

III. 實驗設計與品質改善

在本報告中，我們將實驗設計的相關討論分為兩部份，分別置於“工業統計”與“統計方法與理論”中。在工業統計的部份，我們的討論重點著重在實驗設計與分析應用於解決各種工業問題上的發展。關於實驗設計的各種理論層面的討論，請見統計方法與理論。

工業界藉由做實驗來收集資訊，以解決研發、品管、製程上所產生的問題已行之多年。因為實驗設計這個學科主要關注的問題，便是應該如何收集及分析實驗數據，才能在顧及經濟效益之餘，有效地獲得正確的實驗資訊，故實驗設計在工業統計中，一直佔有重要的地位。

早期的實驗設計研究，起源於農業實驗。在農業實驗時期，實驗設計主要著重於比較各種處理方式 (treatment comparison) 之間是否有差異的問題。自二次世界大戰後，隨著工業的蓬勃發展，實驗設計也逐漸被應用到工業實驗上。但因工業與農業實驗，在目的與本質上有很大的不同，使得工業上的實驗設計，在研究焦點上，與早期的實驗設計有很大的差別。以下我們將介紹幾個工業實驗設計上的重要發展。

(a) 反應曲面法 (Response Surface Methodology)

反應曲面法為 Box 等人在二次世界大戰後，將實驗設計應用於化學工業的問題上，所開發出來的一系列方法。這些方法中的一些概念跟精神，對後來實驗設計在工業上的應用與發展，產生了重大的影響。以下列出反應曲面法中的幾個重要的主題：

(i) 序列實驗法 (Sequential Experimentation)

實驗者不應期待一個實驗便能解決所有的問題，故將所有的實驗資源集中在一個大實驗上，是較為不智的做法。比較好的策略是將資源分配到幾個小實驗上，而每一個小實驗都利用之前的實驗所收集到的資訊，來決定目前的實驗要解

決什麼樣的問題。在反應曲面法中便充份運用這個概念，利用一連串的實驗來探索反應曲面的最佳值位在何處。反應曲面法可分為因子篩選 (factor screening)、最佳值區間搜尋 (search of optimum region)、二階實驗 (second-order experiment) 三個步驟；每一個步驟裡，都包含了不同的實驗法、反應曲面建模、資料分析、以及實驗目的。

(ii) 反應曲面建模 (Response Surface Modeling)

在較小的實驗區間裡，我們可用低階多項式模型來逼近反應曲面。反應曲面法裡運用這個想法，提出一階模型 (first-order model) 與二階模型 (second-order model)。前者適用於實驗區間尚未接近最佳值，反應曲面曲率不大的情況，而後者則適用於實驗區間已經接近或包含最佳值的情形。

(iii) 二階設計 (Second-order Design)

針對二階模型，在文獻上有許多論文提出或討論各種的二階設計。其中較為著名的有中央複合設計(central composite designs), Box-Behnken designs, uniform shell designs, small composite designs 等。

(iv) 最佳值區間搜尋 (Search of Optimum Region)

當實驗區間遠離最佳值時，我們可藉由一連串的實驗來搜尋其位置。目前較有名的方法有最陡上升搜尋 (steepest ascent search) 法和直方格搜尋 (rectangular grid search) 法。

關於反應曲面法更進一步的資訊，可參見 Myers & Montgomery (1995)。

(b) 穩健參數設計 (Robust Parameter Design)

隨著近代工業的技術越來越成熟以及顧客對品質的要求越來越高，工業界關注的問題，除了如何製造出品質達到標準的產品外，如何能使大部份的成品有一致的品質，也成為工業界 (尤其是製造業) 越來越關心的一個課題。針對這個課題，Taguchi 提出穩健參數實驗的方法，藉由應用實驗設計及分析，找出能降低品質特性變異 (variation reduction) 的參數設定，以使產品的品質特性除達到目標值之外還能趨於一致。就實驗設計的角度來看，Taguchi 方法最重要的一個貢獻，便是提出了干擾因子 (noise factor) 的概念。干擾因子是指那些在實驗中可以控制，但在真正的製程中卻無法控制的因子。了解控制因子 (control factor) 對於干擾因子所造成的變異之影響是穩健參數實驗的主要目的。簡而言之，在穩健參數實驗之前，實驗的主要目的皆是探討因子對反應變數的期望值的影響，而穩健參數實驗則更進一步探討因子對反應變數的變異數的影響。

因為在一個穩健參數實驗中有兩種不同角色的因子，所以穩健參數實驗的設計與分析，與傳統因子實驗有些不同，也因此引出不少有趣的研究成果。比如在設計上，有運用交叉表 (cross array) 與單一表 (single array) 兩種做法來建構設計矩陣。而在分析上，因為必須對反應變數的期望值和變異數分別建構模型來使得產品的品質特性達到最佳化，故主要有 location and dispersion modeling 與 response modeling 兩種做法。前者較適合用於交叉表的資料分析，後者則適合用於單一表的資料分析。有關穩健參數設計的更進一步參考資料，請參見 Wu &

Hamada (2000)。

四、 生物統計

生物醫學研究中，因研究的主體多是有生命的，因此個體間的差異就使得統計科學在科學研究要探討整體現象時，扮演了重要的角色。生物醫學的研究範圍廣泛，而其中不能缺少統計科學的部分即為生物統計研究的主要領域。

由於生物醫學研究的蓬勃發展，以及國內對公共衛生方面的殷切需求，使得生物統計有很大的發展潛力，國內在這方面的重點研究方向，應把握理論與應用並重。

近年來，統計學在生物醫學以及流行病學研究上之應用已日漸展開，從與生醫學界及流病研究的合作過程中，亦可發掘新的統計研究課題，因此建立與其他研究領域良好的互動關係，是值得努力的。

與生物統計相關的領域生物資訊亦蓬勃發展，其研究在於運用數學、統計以及計算科學來解決分子生物學的問題，生物資訊研究的內容大多來自於大量快速增加的生物資料，例如不同物種的基因體序列資料或不同物種的基因表達資料，由實驗科學家所產生的資料需要經過註解及詳細的分析，以轉變為有用的知識，再進而增進醫藥衛生的發展，如新藥產生、基因治療等

目前生物統計的幾個主要應用領域，有臨床試驗、遺傳學、一般生物醫學等，此外國外新近的生統發展，尚有對計算生物學的研究，以及神經影像分析等。由這些實際問題所產生的統計方法研究，包括存活分析、長期追蹤資料分析、遺失資料分析方法之研究，均為生物統計發展的重點方向，茲分述如下：

I. 臨床試驗統計方法

臨床試驗自英人 B. Hill 將 R.A. Fisher 在農業試驗所發展出統計理論與方法應用在醫學與藥物評估上，發展出隨機双盲的臨床試驗之基礎，對人類健康福祉做出了重大貢獻。因其可將偏差 (bias) 降低至最低，所以臨床試驗所得結論較其他臨床研究方法更為客觀與嚴謹，成為臨床醫學最重要的方法學之一。而統計方法在臨床試驗之設計、執行、分析及結果詮釋上均扮演了極重要角色。

目前國內（外）臨床試驗研究現況可分下列重要領域：

(a) 存活分析 (Survival Analysis)(亦參考第 V.項 A 部分)

例如在臨床試驗中，由比較不同臨床處置（Treatments）所產生的檢定問題，由研究存活時間與共變因子（Convariates）之間的關係所產生的成比例風險模式(及其推廣)，以及臨床試驗中因試驗限期結束，或病人無法繼續追蹤等因素產生的設限資料（Censored Data），使得存活分析成為應臨床試驗分析上需要所產生的統計方法研究。

其中關於乘積極限估計（Product Limit Estimate）的研究、Cox 迴歸模型的研究以及一些 2-樣本、K-樣本檢定問題，已有大量的文獻產生。這些方法目前在醫學研究文獻上也已被經常地引用。

雖然在應用上這些方法似乎給出了合理的答案，但其複雜的機率結構，則需較深的理論架構來處理。若碰到較複雜的資料，如共變因子是隨著時間變化時，或是事件會隨時間再發生(Recurrence)、再演進時，將多維計數過程（Multivariate Counting Processes）作為複雜生命史資料的模式，使得存活分析得以採用近代隨機積分及鞅論做工具，獲得更豐富的結果。

目前國內在設限資料與存活分析領域如存活函數之估計，K-樣本檢定問題，K-族群排序存活函數檢定問題，廣義 Cox 迴歸模式，事件（疾病）史與共變因子關係模型之研究及選模方法，再發事件分析的理論與方法，交叉設計（Crossover Design）之存活分析及設限資料的相等性研究上已有良好的成果。

(b) 逐次設計及期中分析（Sequential Design and Interim Analysis）

逐次設計及期中分析的統計方法所處理的資料特性是，統計科學家在資料陸續收集時，可決定何時停止收集並做推論。逐次設計及期中分析應基於人道考量，當資料明白顯示一個臨床處置較另一個更為有效或有害時，就可考慮停止試驗。美國國家衛生研究院（U.S. National Institutes of Health）已規定由其所支助的臨床試驗均需採用逐次設計及期中分析。目前統計研究為當實際事件發生率較預期事件發生率為低時之設計與分析的方法學，另外主要性預防臨床試驗（Primary Prevention Trials）的試驗為對健康正常受試者（Healthy Normal Subjects）進行臨床處置之效益與風險的整體評估。所以必須同時考慮數個療效與安全性變數之逐次設計與期中分析的統計方法。

(c) 臨床試驗中生活品質（Quality of Life）之統計評估，多重問題（Problems of Multiplicity）

臨床試驗的目的不但包括對藥品的有效性與安全性之評估，而且對藥品改善生活品質也十分注重，故生活品質問卷設計與可靠度及其相關之資料分析均提供了重要的統計研究問題。當臨床試驗療效評估同時考慮數個指標時，許多相關統計方法就需重新探討其有效性，這是多重問題中的一種，國內在生活品質及多重問題之研究亦有不錯的成果。

(d) 遺失資料（Missing Values）分析方法之研究(亦參考第 V.項 C 部分)

臨床試驗幾乎都會碰到不完全資料（Incomplete Data）的狀況，例如試驗個體之資料是在不同時間點上收集，而在各時間點之遺失資料情形又不同時。因此遺失資料分析方法之研究與各種不同補入（Imputation）技巧之研究，目前亦是國內外熱門之研究題目。

(e) 生技產品的統計方法

全世界的臨床試驗大部份均為國際跨國性大藥廠所支助的藥物臨床試驗。由這些以查驗登記為目的藥物臨床試驗亦衍生及發展許多統計方法學。如樣本數再估算（Sample Size Re-estimation），固定劑量合併藥品（Fixed-dose Combination Drug）之統計設計與分析方法學，與非劣性臨床試驗（Non-inferiority Trials）之設計、分析，及非劣性界限（Non-inferiority Margin）之選用的方法。另外，新藥物在原開發國家如美國核准後，若要至新國家如臺灣申請查驗登記時，雖然不

必在新國家重複在原開發國家所有的各期臨床試驗，但仍然必須執行小型的銜接性臨床試驗（Bridging Studies）證明新藥物在新國家的療效與安全與原開發國家相同。銜接性臨床試驗之設計與分析是目前藥物產品統計方法學的主要研究題目。我國在非劣性臨床試驗及銜接性臨床試驗已有不錯的研究成果。

(f) 藥劑生體相等性（Bioequivalence）試驗統計設計與分析以及藥物動力學之研究（Pharmacokinetics）之研究

學名藥（Generic Drugs）之開發乃為生物製藥工業之起點。而藥劑生體相等性是世界各國學名藥審核批准上市所必要之條件。目前我國在族群相等性（Population Bioequivalence）及個體相等性（Individual Bioequivalence）與藥物動力學之研究方面亦有不錯之成果。

(g) 臨床試驗之實務研究

因為統計在臨床試驗的設計、執行與結果詮釋方面扮演非常重要的任務，所以國內已有許多統計研究人員積極參與各項臨床試驗的實務工作。如國家衛生研究院之臺灣癌症研究組織（Taiwan Cooperation Oncology Group, TCOG）的統計研究中心，衛生署優良藥品臨床試驗規範（Good Clinical Practice, GCP）之制定與執行，臨床試驗中心的建立，查驗登記臨床試驗之查核，與國家級臨床試驗中心的建置等工作。

II. 遺傳統計方法

國際間近十數年來遺傳學研究的一個基本目的在於尋找控制生物體性狀的基因結構及基因表現功能。完成這個目的作法有二個大方向：一是收集人類家族資料(在動、植物研究為設計特定交配系統)，藉由家族內個體遺傳特性，收集遺傳證據，建立檢定遺傳訊息存在的統計方法；一是針對基因功能表現(如，DNA轉錄或蛋白質體)，經由生物資訊研究方法，建立基因調控網絡。前者研究工具偏重於統計，學界一般歸之於遺傳統計研究範圍，且因人類家族資料為不可控制之交配系統，需要特別發展新的統計方法分析，故成為近年國際統計界研究的重點之一。後者，研究工具偏重於生物資訊平台的架設，建立完整可共用的資訊網路系統，才可經由資料探勘過程，逐步建構基因控制機制，屬於生物資訊整合性研究範圍。

近代遺傳統計研究發展的一些重要突破，有以下幾個環節。在 1970 年代最重要的工作是在解決複雜家族資料的概似函數建立，透過這個成果，後續之分離分析及參數連鎖分析得以結合電腦快速運算的進步，成功發展出具實用性的分析軟體，在遺傳疾病研究之基因定位工作上獲致許多實質成果。在 1980 年代遺傳統計的大部分工作在延續探討 70 年代研究問題，其中有別於以往研究的統計方向是開始注意到同源全等基因分析的重要，開始發展有關的無母數統計分析方法；另一個課題則是逐漸開始研究數量性狀的基因定位方法。1990 年代的研究工作起源於對基因在代與代間傳遞的特性運用，由此發展出的無母數統計方法具有簡單、易於實作及具韌性的優點，故受到研究者大量討論。90 年代至目前的研究，最重要的方向是在處理大量標識基因資料(如，SNP)的全基因體掃描方法相關問題。結合前三十年的研究，目前可行的全基因體基因搜尋工作大致上分成二階段

進行，第一階段可以使用參數連鎖分析，進行初期基因候選區的判定，第二階段則改以無母數分析方法進行細部定位。這樣的想法，使得分析受到多基因、環境及基因環境交互作用之複雜疾病數量性狀的研究工作難度，較有可能被解決。

III. 計算生物學統計方法

近年來在分子生物以及遺傳基因研究方面開啟了新的研究領域。如何瞭解在分子層次的生物醫學問題將是往後研究的挑戰。例如醫療個人化的趨勢將取決於對個人基因多型性的瞭解。

分子生物學的快速成長產生了大量的生物資料，包括基因與蛋白質的序列與微陣列資料，蛋白質結構、交互作用與 SNP 等。其關鍵在於高產量技術的商業化，資料的數量可能遠超過既有分析方式的處理能力。過去生物學家多半處理較小型、變數個數較少的資料，而現今計算生物學面臨的資料不僅數量上非常龐大，變數個數以千百計，且資料的格式或結構頗為複雜，加上高產量技術中伴隨的大量雜訊，即使僅對資料做簡單的描述，以期將資料中最重要的資訊摘要整理，以傳統的作法都極為不容易—無論是以圖形、數字、或模型呈現。若無適當的方法將巨量的資料轉化為知識，蒐集這些資料所投入的心力等同白費！

統計科學近年來應用於大型資料庫的經驗，正好為這些生物資料的描述與分析提供適當的工具。又如在蛋白質與基因的功能、蛋白質的交互作用、生化反應路徑、基因調控等課題上，生命的複雜運作形成資料的高度不確定性與多樣性，統計長久以來對不確定資料的研究，將是發現新的生物知識與進一步理解生物系統的基石。

(1) 國際與計算生物學相關的統計研究：

隨著人類與許多物種的基因體定序完成，現已邁入後基因體時代。先前計算生物學透過序列與結構的相似性，設法了解基因與蛋白質的作用(如 Guigo 等，2001)。目前的研究重點則進一步擴展到生理現象的生化過程(如 van Someren 等, 2002)、疾病的與基因功能間的關係(如 Giallourakis C 等, 2005)、蛋白質與生命系統的運作(如 Nooren 與 Thornton, 2003)。生物學家的研究將由過去深入了解各部分各單元的層次，提高到由整體的系統的角度觀察的層次。所需的統計分析也將由資料的單一描述延伸至成群地比較，由描繪個別分子的物理化學模型延伸至描繪交互作用的機率網路(如 Sachs 等, 2002)，甚至由直接實驗的數據延伸至文獻發現或資料庫的整合（即所謂後設分析—分析的分析，如 Sugiyama 等, 2003）。

由於生物現象的錯綜複雜，時常有新的發現或新的理論出現，補充過去的發現或提供更周延的解釋。在新的想法剛產生時，也需要使用電腦程式與演算機制來模擬或驗證這些發現。計算生物學中的問題許多牽涉到最佳化的求解(如 Handl 等, 2006)，如尋找最相似的序列、最吻合證據的生物演化樹，最適合的分子結構，最可能的反應路徑等。實際問題中大部分的最佳化無法使用窮舉法求解，必須借助統計理論尋求夠靠近最佳解的答案，諸如蒙地卡羅法、模擬退火法、基因演算法等(參見 Zhang 與 Shmulevich, 2002)，統計計算在這一方面的發展，也是受到電腦的長足進步所鼓舞。

而在相關的統計發展上，The R Project for Statistical Computing 可謂集其大成，其開放原始碼的架構吸引了統計各領域中與計算相關的專業人士，共同開發了易於使用的計算程式語言(http://cran.r-project.org/)。其他領域的科學計算可修改這些原始碼，快速地調整為特定學科的計算工具。其中應用於計算生物學且較知名者為 BioConductor. (http://www.bioconductor.org/)由於匯集眾多的力量，目前許多工具如平行計算、網路計算、資料庫銜接、最佳化求解、數值模擬等已臻成熟。而許多複雜的資料型態如生物晶片、分子結構、蛋白質體等也已累積了一些成果。

(2) 國內與計算生物學相關的統計研究：

國內從事相關研究學者很多，統計學專長者近來多半著重於建立生物晶片資料分析的方法論，以及由遺傳統計延伸的方法，而直接從事基因序列的比較或蛋白質體、基因體與演化關係者則較少。此外，中央研究院與國家衛生研究院也發展了生物晶片計算分析平台，少數研究員從事文獻的統計分析、生物醫學資料的視覺化系統。

另一方面，資訊電機專長的學者，研究範圍則較為廣闊，舉凡資料正則化、影像處理、資料庫整合、文獻分析、基因體序列、蛋白質結構、反應路徑等，均有學者從事相關研究，雖然多半著重在特定問題的處理而非完整的方法論，但仍累積了一些可供統計學者借鏡的經驗與成果。

IV. 一般生物醫學統計

生生物醫學的研究課題十分豐富，相應產生的統計問題及研究內容十分多樣性，例如傳染病相關的統計方法研究，包括傳染病散佈的隨機模型研究，以及傳染病時空分布的時間序列、空間統計或時空模型研究等。此外尚有診斷醫學以及神經影像醫學等研究，統計學亦扮演重要角色，以下將針對診斷醫學以及神經影像為例加以說明。

A.診斷醫學統計方法

疾病的精準分類才能確保治療或預防的效果，所以醫學診斷與篩檢方法的評估就愈形重要。近年來，美國食品藥物管理局草擬對醫學診斷檢測方法評估之統計規範，我國衛生署藥政處也積極辦理中。尤其，基因體醫學的研究促進了基因、蛋白質等微陣列晶片的量產；分子醫學診斷器材之檢測效度的評估，有賴統計研究方法的創新。另一方面，新穎的醫學診斷器材與工具常是價格昂貴，難利用大樣本的臨床研究來評估其效度。因此，診斷醫學統計方法之研究遂成為生物統計的新興課題。

B.神經影像統計分析

統計科學在神經影像中佔有重要的地位。神經影像研究領域起源於醫學及純科學，包括物理、化學、工程、數學及統計。除此基本研究領域之各別應用，目前神經影像實驗研究已走向高科技領域，採用 fMRI、腦磁波(MEG；magnetoencephalography)及腦電波(EEG； electroencephalography)等技術來探討認知心理學、人工智慧與智慧型學習輔助系統、現代科學教育之電腦適性測驗；

凡此，皆是倚仗統計科技的整合型研究。目前國內外採高科技的神經影像實驗室快速成立，並且結合醫學研究。國內建立的實驗室包含了中研院的 EEG 及 MEG 實驗室，台大、榮總、三總及長庚的 fMRI 實驗室，榮總是目前配備最完整的研究機構。國內參與影像及腦電波分析的統計專業人力，除中研院有神經影像分析團隊外，分散在其他大學的為數極少。事實上，統計的重要性已使得神經影像統計分析，發展為一個現代世界科技重要的研究領域。該領域的基本研究課題可分三方面：

(a) 空間分析

腦影像分析需要空間分析技術的協助，作空間領域的推論。研究的興趣在於探討負責實驗作業的大腦部位，及部位形狀隨著年齡成長的個別差異。空間分析主要為隨機場理論，用來解決影像中多重比較的爭議。隨機場理論的發展，為理論統計和實際腦科學應用結合的代表例子。這顯現統計神經影像，並非單純的將統計方法應用在腦影像，而是兩個領域的互動，成就了各自領域的快速成長。此外，其他方法如小波分析等也由於影像資料的特殊性，得以推展創新。

(b) 時間分析

時間分析也造就了統計及影像分析之間的互動。對從事時間序列分析的統計學家來講，腦電波實驗提供了豐富的真實資料，也刺激了非線性時間序列分析的研究。功能性磁振影像的分析，促使 frequentist 及 Bayesian 的時間序列模式的發展。

(c) 功能性連結和因果推論

連結及因果推論，是神經影像中快速發展的兩個課題；此方面課題，也是未來統計分析較具挑戰的研究領域。簡單的統計檢定實驗效果將非未來研究主流，新的統計方法必須能決定空間中不相連的兩塊腦區域，是否在不同實驗條件下產生功能性連結，彼此是否有因果關係。方法的主旨在於指認連結是否存在，並採信賴區間或假設檢定表示連結的不確定性。

V.生物統計中幾個重要的統計方法學研究

在 I.、II.、III.、IV.項中所提及的應用領域可刺激發展生物統計的重要方法，以下僅就國內可能獲致優秀成果的部分加以說明。

A. 存活分析

(在醫學研究、可靠度分析、勞動經濟學、社會學、精算學 … 中，往往可藉著分析“事件史”(event history) 資料而獲取寶貴的資訊。分析事件史資料的統計方法稱為存活分析 (survival analysis) ，目前仍是統計學門中一個蓬勃發展的領域，以生物醫學方向的應用為主流。藉此工具，科學家對“死亡”“發病”“疾病復發” … 等與人類健康生存息息相關的重要事件之演進得以有更充份的掌握。此外因生物問題相當複雜與多樣，分析方法必須更具彈性，這也刺激了統計方法和推論理論的發展。以下就國內外近年來在存活分析領域一些重要課題的發展概況整理歸納如下：

(a) 多變量存活分析 (multivariate survival analysis)：

主要探討數個存活變數的關聯性。例如遺傳學家可藉分析壽命間的關聯性以探討遺傳與共響之環境因子對於壽命之影響。以無母數的方法估計二維度的存活函數曾一度是個熱門的研究主題。學者欲將單維度最優的 Kaplan-Meier 估計量的許多美好性質延伸到更高維度時，卻遭遇到難以克服的問題，主要的障礙和高維空間缺乏絕對次序 (strict order) 有關。此外在單維度分析中被認為相當有幫助的鞅理論 (martingale theory) 到了高維度也因前述問題而有其侷限性。因此雖然文獻中已提出許多無母數估計方法，卻各有缺點，理論性質較好的估計量在計算上又相當複雜。於是學者漸漸的把推論的方向由無母數分析轉移到不失彈性的半母數模式，其中以 frailty models 與 copula models 最常見。 Frailty models 假設變數間的相關性可由無法直接觀測到的 frailty 變數所完全解釋，一旦此共享的變數值已知，所探討的多維變數即彼此獨立不具相關性。這種模式相當彈性，且“條件獨立” (conditional independence) 的假設因為具有科學上的解釋依據，而且分析時可借用貝氏的技巧，因此經常被用於分析關聯性的資料。Copula models 則比 frailty models 更廣義，其彈性在於邊際分配和聯合分配可以分開討論。推論上以半母數分析為主流，重點在於不對邊際分配做模式的假設而能估計關聯性參數。學者把模式套用於不同的資料結構，為因應更複雜資料也刺激的推論方法的發展。目前學者嘗試的方向包含“擬概似函數法”(pseudo-likelihood approach)，利用 U 統計量或是動差的性質以建構估計函數。此外模式選取(model selection)亦為一重要課題。

(b) 迴歸分析 (regression analysis)：

主要探討解釋變數對存活變數的影響。近年來有學者提出轉換模式(transformation model)，定義上更具彈性，可將過去常用的模式 (proportional hazard model, accelerated failure time …) 都涵蓋在其中。學者針對此廣義的模式所提出估計迴歸參數的方法，主要利用動差性質或是最小平方法(least squares)的概念以建構估計函數。雖然這些新方法在概念上仍脫離不了傳統推論理論的範疇，卻因為資料複雜且模式以另一種面貌呈現，藉此更能了解古典方法的深層意義。

(c) 競爭風險(competing risks)與多重事件分析 (Analysis of multiple events data):

個體身上所發生的事件往往不只一個，例如器官移植病人可能經歷各種併發事件，最後死亡。分析的方向在於處理事件發生的順序與彼此間的競爭關係。早期的研究方法利用隨機程序理論，假設事件的發生服從馬可夫模式、或是半馬可夫模式。另一方向是以競爭風險的角度分析最早發生的事件的風險函數。近年來則有統計學家則區分事件間不對等的競爭關係，針對所謂半競爭風險資料(semi-competing risks data)提出統計推論方法，研究方向包含關聯性的探討、邊際函數的估計或是迴歸分析。

(d) 存活分析與長期追蹤資料:

當存活變數為主要課題且解釋變數為長期追蹤的資料結構時，利用適當參數化模型假設在長期追蹤解釋變數上以便更有效且準確的預測存活時間。為使假設

條件更符合實際應用，非參數化模型假設應為當前可行方法。反之，當長期追蹤資料的迴歸問題為主要課題時，亦會發生資訊不完整的情形，此時存活分析所發展的技巧有助於處理因資料缺失所產生的估計偏誤。

(e) 重複事件 (repeated events) 或是復發資料 (recurrence data) 的分析:

有疾病往往會經歷多次的復發 (例如氣喘，精神病的發作...)。事件的間隔時間是互相關聯的,但其具有次序的結構又使這個問題和一般多維度存活分析不同，針對復發事件在獨立及非獨立截取之假設前提下，學者分別提出不同模式以描述發生率函數。此外，更針對發生率函數之估計及統計推論做深入之研究。

(f) 複雜資料結構之分析:

存活資料在實際應用上往往以相當複雜的面貌呈現，包含各樣的區間設限、截切 ... 或是具關聯性的家族資料。為了處理各種資料型態而發展的統計方法，不單只為了解決應用上的問題，同時也拓廣了統計推論的思維方向。因為基於單純資料型態所提出的方法往往包含不同統計概念在其中,只有當遭遇更複雜的資料型態時，其差異才會顯現。這些概念的澄清對統計學的整體發展是有助的。

(g) 包含長期倖存者的存活分析 (survival analysis with long-term survivors):

傳統的存活分析假設感興趣的事件遲早會發生，若在研究期間尚未發生，則該觀測值受到了設限。若是所謂的事件指"死亡"，這個隱含的假設是合理的，因為生物的壽命是有限的。但是若事件代表"患病"，此假設就有可議之處，因為有些人永遠不會得病。當存活分析應用的範圍不再限於死亡時，有學者開始把"長期倖存者"的存在納入分析中，他們是無論觀察多久都不會發生事件的個體。一個目前盛行的研究趨勢是在承認有長期倖存者的條件下,重新探討前述的推論問題。

B.長期追蹤資料分析（Longitudinal Data Analysis）

長期追蹤研究常見於許多研究領域，例如生物醫學與流行病學研究，人口學與生態學研究，環境與疾病之監測研究等。因此，長期追蹤資料分析方法，在生物統計研究中扮演著重要的角色。長期追蹤資料分析的統計方法研究，過去近二十年來，已有顯著的發展。儘管如此，由於長期追蹤資料型態的多樣性與資料蒐集特性，其統計分析方法仍然不斷地有新的進展。茲將國內外目前之研究現況整理歸納如下：(但相關之統計方法研究並不僅限於此)

(a) 迴歸模式之探討：

包括 (i)隨機係數(random coefficients)分析與隨機效應(random effects)之研究：針對連續型與離散型的資料，考慮個體潛在之隨機差異性，主要以線性與非線性的混和效應(mixed effects)模式為研究主題； (ii)時間變異係數(time-varying coefficients)之研究：探討時間變化對迴歸效應的影響，應用於各種迴歸模式，包括風險迴歸及倖存模式等。其他如動態(dynamic)模式、遞移(transition)模式等，也都被提出用於分析長期追蹤資料。此外，伴隨著上述模式的估計式與統計推論等，都是相關的研究問題。

(b) 估計函數與估計式之研究：

以 GEE 及 GLS 為基礎之估計式，及半參數迴歸分析等方式來處理長期追蹤資料，其推展出來的方法甚多，研究課題包括估計方程式及估計式有效性之探討與改進及其大樣本漸近性質之理論、共變異函數及相關函數之估計、隨機效應之模式與估計等。

(c) 函數型資料(functional data)分析方法：

將長期追蹤資料視為函數型態資料的情況很多，以函數型資料方法來做分析。這方面的研究多以無母數方法為主，結合平滑法(核迴歸、局部多項式、弧線平滑法等)、隨機過程理論、多變量分析的技巧等，發展出新的分析函數資料的方法。這方面的研究，在近十年來的發展尤其迅速，應用也相當廣泛。

C.遺失資料（Missing Values）分析方法之研究

生物醫學研究多屬觀察性研究，因此常見有遺失數據，即計畫中欲取得的數據因某種原因無法取得。近數十年來，遺失資料的統計方法學在統計學，特別是生物統計學研究領域中，一直是相當活躍的研究領域，由此反映出遺失資料相關問題在理論上的有趣性及在實務上的重要性。

目前文獻中處理遺失數據的方法學主要分為兩大類不同的方法：selection models 及 pattern-mixture models。此兩種方法的差異主要在於對資料遺失機制的建模方式：前者乃針對數據遺失狀態在給定數據值下的條件機率建模，而後者乃針對數據在給定數據遺失狀態下的條件機率建模。簡言之， selection models 方法基本為 likelihood-based 方法，一般可直接利用 likelihood 或 estimating equation 等傳統方法進行參數估計及推論；而 pattern-mixture models 基本上為插補（imputation）方法，其將遺失數據透過由某假設模型所得之預測值進行插補，但該假設模型一般不具 identifiability，因此需加上某些限制條件，並進而導致其參數估計及推論無法以傳統方法執行。長期以來文獻中以 selection models 相關方法的討論較多，但近年來關於 pattern-mixture models 的討論有逐漸增加的趨勢。

在 selection models 方法中的一個關鍵問題是數據遺失狀態在給定數據值下的條件機率是否與遺失數據本身有關。若此條件機率僅與觀測到的數據有關，而與遺失數據本身無關，則此時正確的統計推論可以僅由觀測數據的 likelihood 得到，不需考慮數據遺失機制（模型），亦即數據遺失機制是可忽略的；若此條件機率與遺失數據本身有關，此時僅由觀測數據之 likelihood 所得到的參數推論可能是偏誤的，必須將數據遺失模型納入考慮，亦即數據遺失機制是不可忽略的。近年來關於遺失資料的研究趨勢，已逐漸由數據遺失機制是可忽略的情形轉移至數據遺失機制是不可忽略的情形，因為後者較前者為更一般、且更符合臨床試驗及長期追蹤研究等實務中所面臨的資料遺失情況。

在長期追蹤資料與設限（censored）存活時間資料中，由於資料均需經過重複測量或長期追蹤始能取得完整數據，遺失資料是普遍發生的情形。又此兩類型資料的分析往往牽涉到較複雜的模型，因此其相關的遺失資料問題也極具挑戰性，並已成為近期遺失資料分析方法學研究的另一項重點。

從統計技術層面來看，遺失資料分析方法當前的主流已由參數化方法逐漸轉為半參數及非參數方法，以提高分析結論的穩健度。但另一方面，利用日漸成熟的 Markov Chain Monte Carlo 計算方法， 以 Bayesian 角度來分析遺失資料的方法學也逐漸增多。

目前國內統計學界有數名學者進行遺失資料方法學的研究，大多集中於迴歸分析的遺失資料研究。

D.視覺化統計方法

資料與資訊視覺化的目的是將資料中的數據以線條、圖形、顏色呈現，並經由適當的整理，以呈現隱含於資料中的結構與有用的資訊。近十年來的研究重點包含動態繪圖、互動式呈現技巧、著色與關連等軟體的開發。而後因應生物晶片分析的需求，矩陣式的呈現方式也成為研究重點。國內目前有少數人員從事相關研究，已有分析軟體供研究者使用。

資訊視覺化雖然不完全是新的研究領域，卻仍是一個尚待開發且頗具潛力的礦場，存在許多的研究課題與應用技術，值得有興趣之學者共同努力開發。當然此領域之研究工具除了常用之數理與統計方法外，尚須電腦繪圖之技巧（介面）。

在此所談之視覺化統計方法不含影像處理（image analysis）相關方法。

視覺化可以略分為統計資料之視覺化與統計模型之視覺化 - 統計資料之視覺化方法與技術用來觀察與萃取潛藏於資料中之訊息而統計模型之視覺化則著重於觀察模型配適之優劣與殘差之診斷，其統計意義與視覺化表現法有其本質與技巧上之根本差異。

(a)國際視覺化統計繪圖研究：

在視覺化統計方法或統計繪圖法之發展上，國際學界近年有幾個主要的研究方向：

- **Mosaic Display –**

由 Michael Friendly (Friendly M,1999)（http://www.math.yorku.ca/SCS/friendly.html）主導此領域之研究，歐洲亦有數個研究群如 Antony Unwin (Unwin AR, 1998, 1999)（http://stats.math.uni-augsburg.de/~unwin ）等。相關研究者雖宣稱 Mosaic Display 可以用來觀察高維度類別形資料，但實際應用到十維資料已經是極限。

- **Parallel Coordinate Plot –**

由 Alfred Inselberg (Alfred Inselberg, 2002)（http://www.math.tau.ac.il/~aiisreal/ ）提出之研究領域，歐美亞洲亦有數個研究群如 Edward J. Wegman (Edward J. Wegman , 1997) (http://www.galaxy.gmu.edu/stats/faculty/wegman.html)；Antony Unwin (Unwin AR, 1999)；Junji Nakano（http://jasp.ism.ac.jp/%7Enakanoj/ ）等。

- **Tree modeling visualization –**

由於樹形結構之統計方法（含分類迴歸樹 Classification and Regression Tree (CART) 與階層叢聚樹 Hierarchical Clustering Tree）使用方便，近年有一些研究工作針對樹形結構之模式進行模型選擇與結果之視覺化工作。如 Simon Urbanek (Urbanek, S, 2002, 2003) (http://simon.urbanek.info/)； Wei-Yin Loh (W.-Y. Loh,

1997, 2002) （http://www.stat.wisc.edu/%7Eloh/ ）等。

- **Diagnostic Plot** –

各種迴歸模型之診斷圖皆為此一領域之可能研究對象，一個較完整之診斷系統為 Chun-houh Chen 之 Interactive Diagnostic Plots for Multidimensional Scaling 。

- **Matrix Visualization** –

由於基因微陣列之普及，Michael Eisen(Eisen MB, 1998, 2001)（ http://rana.lbl.gov/EisenSoftware.htm ）將此領域之應用帶向高峰；而 Chun-houh Chen (Chen CH, 2002) (http://gap.stat.sinica.edu.tw/)則在此領域有全方位之研究。

- **Network Visualization with interactive statistical graphics and database linking for biomedical informatics and internet communication** –

基因體與蛋白質體醫學之大量資料帶來了基因與蛋白質交互作用網路視覺化之需求，網際網路之發達亦有著類是的需要。此一領域之研究有著無限的未來性與發展空間。

(b)軟體使用與開發

資料與資訊視覺化統計軟體之開發不易，既耗費人力電腦與時間又不易有共通認定之研究成果 – journal articles。必須制定遊戲規則加以鼓勵，否則永遠沒有統計同仁願意投入此研究方向。

(c)國內視覺化統計繪圖發展：

國內統計學界投入視覺化統計方法人力相當少，可能只有 1 至 2%之人力。相對的其研究成果在國際舞台上卻有相當突出的能見度與舉足輕重之地位，顯見這是一個值得投資開發的方向。

目前國內最主要的資料之視覺化統計方法研究團隊為中央研究院的廣義相關圖（Generalized Association Plots, GAP）研究。 GAP 在矩陣視覺化之研究雖然人單力薄，但是在統計相關領域人員之精神支持與其他應用領域學者之實際使用下，已經在國際統計學界與精神醫學與基因體研究應用領域逐漸展露身手，為國內統計同仁開拓出一塊新的活動空間。 GAP 研究人員近年來在許多重要國際會議與研究單位受邀發表了多場重要演講與課程講授，提升了不少台灣統計研究之國際舞台能見度，也應該是其他有興趣統計同仁可以進場開發新視覺化統計方法的時候了。

國內之統計模型之視覺化亦只有 GAP 相同團隊的多元尺度分析之交談式診斷圖（Interactive Diagnostic Plots for Multidimensional Scaling）。

五、 資訊科技相關之統計

計算方法在統計學領域中一直是很基本的課題。數學工具除了提供統計一個很好的理論架構外，也告訴我們計算的準則，因此隨著資料的複雜化與巨量

化，統計中的數學工具也變難的同時，我們對計算工具的要求也越來越多；不僅計算量加大，計算的難度也大幅度地加深。故雖然計算機已是普遍的工具，然而計算的難度大幅增加的結果，使得計算方法於資訊科學中更成為一門重要的學門。由於統計科學的多樣化特質，統計學者們或直接或間接地必須與計算及資料有密切的關係。美國統計學會會長 (1997) Jon Kettening 曾說：「我喜歡把統計想成一個是從資料中學習的科學，…」。 (I like to think of statistics as the science of learning from data …) 而很巧的是"從資料中學習" (learning from data) 這一段文字幾乎出現在所有資料探勘及機器學習的教科書上。統計與資訊工程的重疊由此可見。我們常常必須在服務者與被服務者(client and host) 的角色之間轉換，故如何與其他領域的學者合作是統計學者們的必修課。尊重專業是學門之間合作的唯一途徑，誠如統計學對其他學門的貢獻與重要性，計算科學對統計學門的發展也將扮演重要的角色。著名的機率學者鍾開萊教授 (Kai Lai Chung) 即曾經在他的書的前言中寫到：「某些人的技術細節是其他人的專業領域」*("One man's technicality is another's professionalism."-- from "Lectures from Markov Processes to Brownian Motion," 1980)*。為了讓統計科學發展地更穩健，統計與計算學門的合作，便得更緊密。除了培養新一代的統計學者使其具備相當的計算能力外，如何與計算領域的學者合作，應是統計學門尤其是統計計算相關之研究學人們所應努力的，絕不可輕忽了這些「技術細節」。

無論是在國內或國際間，統計、資訊科技和應用科學間的互相影響刺激了不同類型數據和科學調查的統計方法，使之得以迅速發展。以澳洲為例，統計與機器學習的合作密切，在他們發起的機器學習夏日學校每次都邀請統計學者參與。而在歐洲由義大利 University of Cagliari 電機系於 2000 年發起，至今仍經常於歐洲舉辦的"Multiple Classifier System Workshop"，即經常邀請統計學者為主要演講者(keynote speaker)，統計學受到的機器學習學者們的重視可見一斑。

隨著問題的發展與轉變，統計與計算的變化及涵蓋範圍甚廣，我們在有限的人力下只能就所學相關的範疇約略說明，掛一漏萬在所難免。由於領域的特性，我們時而以統計方法分類，時而以應用領域分類，難免重複之處甚多。也由於統計學門的特殊性，我們嘗試僅就方法的發展討論，至於其他相關應用(如生物資訊，工業統計，資料及資訊探勘(data and information mining)，醫學影像，視訊，資訊安全等等，不一而足)，將主要由其他學門應用領域的專家們來討論，在此將僅簡略陳述。本報告將重點集中在下列方法論上，不足之處請參考其他報告。在此討論的課題包括：統計計算(statistical computing)，分群方法(clustering)，分類方法(classification)，資料視覺化(data/information visualization)，隨機最佳化及變異貝氏方法(stochastic optimization and variational Bayes methods)，回饋式學習(reinforcement learning)，獨立成分分析法(independent component analysis)，資料探勘(data mining)，統計神經影像術(statistical neuroimaging)。為兼顧各別方法的特性及其共通性，我們保留每一個方法的特有發展需求，而把共同的需求留到最後。

I. 統計計算

在文獻中，統計學和計算科學的發展常常被放在一起探討。計算機科學的發展給予統計計算研究領域高度刺激的作用。統計的方法也為計算機科學和資訊技術提供其他的有價值的觀點。從擬似和類似的隨機亂數的產生開始，蒙第卡蘿方法 (Monte Carlo methods)在第二次世界大戰時核子反應的原子武器的發展就扮演著一個非常重要的角色。在第二次世界大戰之後，統計計算被運用於分析大範圍的頻譜和許多工業現實上所遭遇的問題。典型的案例如：the EM algorithm with related optimization methods (Dempster, Laird, and Rubin, 1977), Jackknife/bootstrap methods with resampling techniques (Efron 1979), Markov Chain Monte Carlo (MCMC) methods like the MH algorithm, Gibbs sampling, Hidden Markov models (在 Gilks 所著一書的回顧中, 1996, Liu, 2001)和其他許多方法 (在 Thisted 所著一書的回顧中, 1988, Kundu, 2004)。

世界上統計計算方法的密集研究可歸功於計算機科學的快速發展。以致有很多統計軟體能實現這些計算性的方法。其中一個較成功的案例如：由為統計計算的 R 計畫所開發的開放原始碼的軟體系統 R (http://www.r-project.org/). 有非常大量的新方法被放在這個整合型的系統，使得 R 變成在這世界上很受歡迎的軟體，任何人都可以透過網路找得到。

在亞洲/太平洋地區的統計計算研究正逐漸地提昇。因為在亞洲/太平洋地區高等教育的普及和資訊科技工業的高度發展，提供了大量的研究機會去發展統計計算更先進的方法。

在台灣除了許多研究人員的研究成果與軟體工具的發展之外，研究群也開始對於高維度與大量資料在科學與工業上的相關統計計算與應用進行共同合作。相似於發展方法學在統計方面的重要性，鼓勵軟體開發新的方法和特殊的應用亦同等重要。應該鼓勵個人和團體致力於提昇新的方法及在學術和工業上的統計計算科學應用。計算機科技近來的發展，像平行處理 (parallel processing)、格網運算 (grid computing) …等等，可以被進一步地整合統計方法，來發展最先進的統計計算方法。

(a) *分群方法*

最常見的分群方法其目的在於將資料分成數個較為相似部分，使得每一部份可以透過更為簡單的方式描述、摘要，且仍掌握整體資料的重要結構。

過去統計上發展的許多分群方法，大部分透過所有資料點的兩兩關係矩陣進行分群，或是給定所需的群體總數後直接對資料進行切割。但由於資料探勘與生物資訊所面臨的資料量異常龐大，計算兩兩關係矩陣的計算量常超過實際電腦可處理的大小，群體總數的約略數字也不容易估計，因此近來關於大資料的分群方法已成為研究重點。首先面臨的問題當然是資料庫的儲存管理與資料擷取技術，這部分的資料庫技術與軟體近來已臻成熟，但與統計相關的大量資料分群方法卻才在起步階段，目前約可分為兩大發展方向：一是發展不需計算兩兩關係的分群演算法，以減少計算時間但仍有與傳統方法相似的良好性質；二是發展視覺化的

呈現方式，讓資料分析者根據其經驗或專業判斷應將資料分為幾群或如何將資料分群。第一個方向以國外的學者研究較多，而第二個方向國內有部分統計學家參與，但更多是國內外的資訊工程學者提出視覺化的方法與應用軟體。

(b) 分類方法及樹狀法(Classification & CART)

在此談的分類問題(Classification)是指區別分析(discriminant analysis)。其主要目的是根據所觀察到的資料，建立一個分類規則(classification rule)或分類子(classifier)，再透過此規則對新的觀測值加以分類。所不同於群聚分析(clustering)的是資料包含反應變數的類別訊息。始於 1960 年代便已有許多樹狀結構分類法的提出及應用， 包括了 AID、ID3、CHAID 與 FACT。分類迴歸樹(CART)是 Breiman 等人在 1984 年所發展，其原理是使用二元遞迴(recursive partition)的方法將樣本分割。通常在使用 CART 進行資料分析時，可將反應變數屬性分為連續型與間斷型兩種，連續型資料可應用於預測，而間斷型資料則可應用於類別辨識。

分類樹主要的優點在於能透過樹狀圖的方式，將分析結果完整的呈現出來，具有視覺化與容易解釋的特色。將資料分析的結果以樹狀圖的方式呈現。由於此一呈現方式具有較為人接受的解釋結果的能力。使得此一方法成為機器學習（machine learning）領域中的一個基本工具。

分割的選取、終止分割的方法及區域模型的配適是此一方法的三個基本元素。統計界最早在樹狀法的研究為 Morgan and Sonquist (1963) 於 JASA 所提之 AID 方法，之後陸續有 THAID (Morgan and Messenger, 1973) 及 CHAID (Kass, 1980) 方法提出，Breiman 等人於 1984 年所提出之 CART 方法，將上述三個基本元素做有系統的探討，使得樹狀法的研究與應用進入了新的紀元。在資訊界，C4.5 (Quinlan, 1986, 1993) 是較為人所熟悉的樹狀分類法，而 M5 (Quinlan, 1992) 則為樹狀迴歸法。電腦的發展促進了相當多在這些基本元素上方法及理論的探討，其中有 FIRM (Hawkins, 1993)，QUEST (Loh and Shih, 1997)，RTREE (Zhang, 1999)，GUIDE (Loh, 2002)，LOTUS (Chan and Loh, 2004)，LMT (Landwehr et al., 2005)。而其應用面由早期的分類問題推廣到迴歸及存活資料分析。近來的研究是將方法延伸到多維度的反應變數、長期追蹤資料及廣義線性迴歸等問題上。

目前國外的研究中，已發展了針對不同資料型態的二元分類樹法則，例如存活樹(survival trees)、類別性資料、及時間序列資料等。另一方面，演算法的改進及加速，也有不少研究。近年，結合 bagging 和 boosting 等演算法來提高分類正確率的多重樹法，Breiman 於 2001 提出隨機森林(Random forests)，以 bootstrap 生成的 training data 和隨機選取的分割法則造出多重樹，在提高正確性外，也同時較不易 overfitting。而對資料分析更有助的是，隨機森林提供了兩個有用的指標：重要變數的選取與樣本間的相似程度，相關研究正方興未艾。而國內的研究，也累積了不少成果。CART 已經成功地應用在許多不同的領域中，包括工程計算、醫療診斷、經濟、社會與行銷研究等。包含 CART 在內的一些分類法則是資料探勘(data mining)中一項重要且熱門的研究議題。

(c) Sliced Inverse Regression (SIR,切片逆迴歸法) and PHD (principal Hessian direction)

在高維度非線性迴歸模型中，雖然有興趣於有效維度的縮減，然而可供使用的方法卻不多。至今維度縮減的研究，著重於提出新的方法論或修正已存在之方法，以解決高維度資料分析問題。

切片逆迴歸法(Sliced Inverse Regression)是維度縮減(dimension reduction)研究領域的一支，其發展過程已有 15 年的歷史。最早始於 Li 於 1991 所提出。自此開展了一個重要的統計研究問題。切片逆迴歸法的目的在於透過有效維度縮減子空間的估計，將解釋變數 x 的維度縮減後，可觀察反應變數 y 與低維度變數 b'x 的迴歸關係，其統計模型及演算法屬於無母數迴歸的範疇。切片逆迴歸法可視為加權的主成分分析法(PCA)，其加權的方法是考慮反應變數 y 的資訊。

近年來，國內外在切片逆迴歸法的理論發展及資料分析應用上已有相當多的研究。理論部份，包括不同資料型態的有效維度縮減子空間估計方法、有效維度縮減方向的極限性質，及資料維度數的估計等。一些以切片逆迴歸法為基礎的相關方法也陸續發展出來，例如 SAVE, SIR-II 和 PHD (principal Hessian direction)。在應用部份，包含多變量的時間序列資料分析，醫學影像的切割問題，與分類樹結合在分類問題上的應用，以及基因體學資料的分析皆已有不錯的發展。目前切片逆迴歸法的研究，較出色的研究群分佈於美國、日本及台灣。

(d) 馬可夫鍊蒙第卡蘿 (MCMC) 的計算方法

馬可夫鍊蒙第卡蘿 (MCMC) 的計算方法在很多的領域，如統計、生物、經濟、物理、計算機科學、數學，都很重要；這個方法以隨機過程的理論為基礎，提供統計推論的工具。

在統計的領域當中，一開始以貝氏統計受到 MCMC 方法的影響最深，過去在貝氏統計裏難以計算的後驗分配，如今可以藉著 MCMC 方法來找出後驗樣本，再藉著這些後驗樣本來進行推論；如今，MCMC 方法已經演變到可以處理更複雜的統計模式（如 generalized linear mixed models，longitudinal and survival models），解決更複雜的統計問題（如 density estimation，latent-class models，missing data、genetic statistics）；使用的人口也越來越廣，不僅是貝氏統計學家、還有頻率學派的統計學家，也逐漸開始使用 MCMC 方法。

未來在發展 MCMC 方法的挑戰裏包含了理論的深入以求計算方法的精進，來針對較複雜的統計模式建構出可用且快速的 MCMC 方法；另外則是標準化 MCMC 方法的使用，使得此類型的方法能更加友善的被各領域的人使用。

第一個挑戰首先需要培養熟悉隨機過程理論的貝氏統計學家或頻率學派統計學家，能轉化複雜的統計模式問題來合理使用複雜的 MCMC 方法，能具備高超的程式語言及計算能力創造出可用的方法。要培養同時具備這些能力而且運用自如的人才，一方面需要教育上長期的規劃以及有效的執行，另一方面可以讓不同專長的學者合作，這兩個方法必須同時進行；國科會可以整合國內散居各地的人力，設計從基礎到較高深的訓練課程，定期舉辦目標明確以及有目的導向的座

談會，來提供平台給現有的研究學者以及培養未來的統計學家。

第二個挑戰需要有能與其他科學領域溝通、能撰寫快速的計算方法、並提供給其他研究學者使用的人才。要面對這個挑戰需要培養統計學家們與別人溝通的能力，需要培養複雜程式的撰寫技巧，以及需要改進 MCMC 方法成比較標準化版本的人才。要達到這一部分的目標，有一些措施可以立即開始；例如，在上述的訓練課程中，應可以選擇部分內容並以此為導向來培養及選擇基礎人才，另外並輔以數名高階研究人員監控進度，再委託成員多元的委員會或小組導向目標；如此可以立即完成部分目標、看到成果、並進入較複雜或耗時的工作目標。只要人事費用穩定，軟硬體設施健全，這一些措施隨時可以在一至數個單位開始進行。相較於第一個挑戰，這反而是比較容易完成的目標。

(e) 隱藏式的馬可夫模型(Hidden Markov Models, HMMS)

隱藏式的馬可夫模型(hidden Markov models, HMMS)不僅僅對應用領域是很重要的，同時對理論研究者亦是非常有趣的議題。這些議題是屬於狀態空間(state space)模型的範疇，在狀態空間為有限的情況下仍存有令人滿意的特性。這在許多實際應用的問題上是特別有趣的，尤其當所考慮的系統的輸出被認為是以離散型式輸出時。

隱藏式的馬可夫模型確實被使用在許多的應用領域上。許多早期的理論研究都是起源於語音處理文獻(speech processing literature)[1]，當時隱藏式馬可夫模型仍然是分析語音資料的主要模型之一。即使系統的輸出可能是連續的，它仍被假定為是由一個潛在的分散式過程所支配。另外一個隱藏式馬可夫模型有顯著影響的研究領域為生物資訊，尤其是脫氧核糖核酸(DNA)序列分析[2,3]。鑑於生物序列分離的本性，隱藏式馬可夫模型是特別適合用來模型化的工具。隱藏式馬可夫模型也適用於財務分析上，尤其是當對於不同的社會制度管理下的信賴並根據一些馬可夫過程(Markovian process)接通時。這些工作已經整合了隱藏式馬可夫模型與時間序列分析的領域。而對於隱藏式馬可夫模型的應用有興趣的領域並不只侷限於這些，包括文件分類(text classification)、機器學習(machine learning)甚至可擴及天氣預測(weather forecasting)等。

然而，隱藏式馬可夫模型(HMMs)並非只是一個被拿來應用的模型，而沒有理論研究的價值。研究這些模型的估計時[4]，也顯示這些模型從純理論研究的觀點來看仍然相當有趣，且需要發展強而有力的方法論來評價模型估計的特性。還有一些與其性質相關的公開的問題，如：可識別性(identifiably)、模型選擇(model selection)、模型階層估計(model order estimation)等等。這些模型的計算特性也相當的有趣。而參數估計的高效率演算法也同時是理論與實務的一個有趣主題[5]。計算演算法對於計算生物學(computational biology)的使用者而言更具有莫大的興趣，尤其是當需要高效率的演算法來處理大量資料時。這是在資訊科學與統計學之間的一個高度重疊的重要領域。

台灣有許多在隱藏式馬可夫模型(HMMs)的領域內相當活躍的學者是同時從理論及應用的觀點來切入。全世界，特別是在計算生物學的研究領域內，因為

引進隱藏式馬可夫模型(HMMs)而再度活躍。在未來，一個重要的發展必是將兩組研究群帶進更緊密的合作中。有實用價值的方法與演算法常常需要有強力的理論基礎的支持來保證他們的表現，尤其是大樣本時(asymptotically)。另外，實務上的問題也能為理論的發展提供有趣的挑戰。

II. 資料視覺化

資料視覺化的目的是將資料中的數據以線條、圖形、顏色呈現，並經由適當的整理，以呈現隱含於資料中的結構與有用的資訊。

過去統計上發展的許多視覺化方法，大部分透過維度縮減技術，擷取眾多變數間最重要的幾個關係，而呈現在三維的圖形中。但當變數個數增至幾十或幾百時（如資料探勘與生物資訊面臨的問題），維度縮減已不敷使用。近十年來的研究重點包含動態繪圖、互動式呈現技巧、著色與關聯等軟體的開發，期望透過電腦的強大計算能力，呈現更多資料中的細節。而後因應生物晶片分析的需求，全矩陣式的呈現方式也逐漸成為研究重點。過去這部分的研究以國外學者為主，國內近來由於生物技術產業的需求，已有較多人員從事相關研究，並有分析軟體供研究者使用。

III.最佳化方法(Optimization Methods)

(a) 隨機最佳化及變異貝氏方法(Stochastic Optimization and Variational Bayes Methods)

資訊科技(information technology) 在近十年進步神速，在統計領域有重要的意涵。除了統計計算更為方便、迅速外，蒐集與儲存資料的技術與面向也有長足的突破。這發展同時也引發出相當多的重要新問題與不同的解決角度。其中固然有不少仍在現有統計理論架構內，然而有更多是不全在傳統統計領域，或屬於較非「核心」的問題。隨機最佳化(stochastic optimization)可以視為這類新問題的一個代表甚至是統合的主題(unifying theme)。而變異貝氏方法(variational Bayes methods)則是新崛起的重要解決角度。

最佳化(optimization)當然是一個古典的問題，與解根(root-finding)一體兩面可視為數值分析中的最重要的問題之一。過去的常用手法是以線性的方法處理，或是以較方便的函數逼近（如多項式、矩陣逼近）後，再輔以線性方法處理。然而，當這個最佳化(optimization)問題中的目標函數(objective function) 或設限(constraints)中包含隨機與不確定性時，這樣線性化的手法有其相當限制。

變異貝氏方法(variational Bayes method) 提供了一個適合統計學者切入隨機最佳化(stochastic optimization)的角度。這個角度將最佳化的問題轉化成積分計算問題，先以變異方法(variational method)找出一群可能的逼近函數，再以統計的方法(如經驗貝氏)選擇出一個最好的逼近。通常這個逼近仍無法直接有好的理論表現式而需以迭代或模擬的方式計算。在計算後驗機率期望值(posterior expectation)上，這個方法已有相當多的成功結果。參考 Variational-Bayes.org：URL http://www.variational-bayes.org/這樣一個以變異貝氏方法(variational Bayes

method)處理隨機最佳化(stochastic optimization)的手法與分類組合(ensemble classification)有相當密切的聯繫。以近年頗受矚目的 boosting 為例， Friedman, Hastie and Tibishirani (2000, Annals of Statistics)提出一個由 additive logistic regression via Newton-like updates for minimizing exponential criterion 的詮釋。雖然國際有十分活躍的後續研究，但對該遞迴(iteration)的收斂性及數值結果部分，仍有相當多的問題懸而未解。由於變異方法(variational method)在數學及數值計算上已有較成熟的理論與算則，再加上統計學界對貝氏方法及計算的瞭解，相信可以為這些問題提供一些新的觀點。

另一方面，這樣一個架構也提供分析一些複雜最佳化問題中，高階探索方法(***metaheuristics : a high-level strategy that guides other heuristics in a search for feasible solutions.***)的理論分析平台。許多實務上產生的最佳化或隨機最佳化(stochastic optimization)的問題無法有一個好用的方法，而高階探索方法(metaheuristics)也因應而生。舉凡早期的模擬退火法(simulated annealing)，近年的仿生學(如 ant colony system)，都在個別的問題中提供了可接受甚至不錯的解。然而，這些方法並非萬能，常為人攻擊的，其中經常有許的參數調變(tuning parameters)需要設定與選擇。如何去選取這些參數的理論引導之依據相對不足。在 boosting 研究上的成功，指示了一個可能的方向。而現有變異貝氏法(variational Bayes)的角度，也增加了詮釋及進一步修正這些高階探索式的可能。

由於隨機最佳化(stochastic optimization)，在我們所討論的意義下，是一個相當新的領域。高階探索式(metaheuristic)的結果很多，理論的結果(theoretical results)相對很有限，過去台灣統計學界也涉入不多。前瞻來看，這個問題牽涉的專長與人力可以很深入也可以很平廣，由尖端的理論學者到高計算量的算則 (algorithm)設計乃至程式撰寫；由一般問題探討到與實務界緊密結合的應用，都有其切入點。而與台灣目前的統計學者專長，也有相當好的連接，應是一個值得投入的領域。

(b) 分部最大化(Maximization by Parts)於概似推論之應用

分部最大化(maximization by parts, MBP)[1] 是當概似函數的二階微分過於複雜而難以計算時，發展出來的一種用來估計概似函數的方法。這是一種發表於 2005 年末的研究發展。因為當模型變得愈來愈複雜時，想要更精確地估計概似函數將變得更加困難，因此像這樣將概似函數分解成幾個不一樣的函數來表示，並保證在最佳化這些函數可得到一個漸近一致的解的辦法，是讓人覺得很有興趣的課題。這代表著另外一個與 Newton-Raphson、Fisher scoring 或者是 EM algorithm 之外的另一個解決方案。在這篇論文展示了這個方法對某些模型是有效的，而且提供了一個一般性的原則時，對於個別個案找出有用的分解仍是該被研究的課題。並且，與更多現行的方法作比較將有助於決定這個新方法在不同設定的概似最大化(likelihood maximization)問題中所扮演的角色。

IV. 統計學習(Statistical Learning)

(a) 回饋式學習 (Reinforcement Learning)

回饋式學習的主要歷史根源有二，其一為動物心理學中的試誤(trial and error)學習，其二則來自最佳控制(optimal control)理論及其相關的動態規劃(dynamic programming)問題。此二源頭在 1980 年左右匯集成今日回饋式學習的主要架構。

所謂回饋式學習是系統從環境到決策的遞迴學習過程，以使得累積報酬(累積強化信號函數)最大，回饋式學習不同於監督學習(supervised learning)之處，主要在於監督信號上。回饋式學習是由環境提供的信號對決策的好壞作評價(通常為量化信號)，而不是告知回饋式學習系統(reinforcement learning system, RLS)如何產生正確決策。由於外部環境提供的信息很少，RLS 必須靠自身的調適進行學習。通過這種方式，RLS 在決策－評價的環境中累積知識，改進行動方案以適應環境。

近年來，國際上對回饋式學習的理論有各式各樣的應用：例如網路安全、生物資訊、機器人、經濟預測、線上品管等等。國內這方面的研究大多集中在各大學及研究單位的電機及資訊系所，統計系所在這方面的研究，除了九○年左右在控制理論上稍有投入外，其餘著墨不多。其實以國內目前的統計研究水準來看，是很有機會在回饋式學習的理論上做出重要貢獻的。

(b) 獨立成分分析法(Independent Component Analysis, ICA)

主成分分析法(principal component analysis, PCA)是在多變量分析中相當受歡迎的統計工具。基本上，PCA 的目標是找到多個彼此之間並不相關的主成分，而這些主成分是經由特徵值分解法來分解被觀測數據的共變異數距陣所得。因此，在高斯(Gaussian)分佈的假設下，這些主成分之間具有統計獨立的特性。然而，當觀測值並非來自於常態分佈時，主成分分析法並不能協助我們取得這些彼此獨立的分量(component)。所以在過去的二十年中，發展出了另一種稱為獨立成分分析法的新工具。事實上，獨立成分分析法可以視為是主成分分析法的一種概括。但是，與主成分分析法不同的是獨立成分分析法不僅可以處理高斯分佈假設下的資料，對於非高斯(non-Gaussian)分佈或是具有更高階統計量的資料也可以處理。目前獨立成分分析法已成功的應用於不明資料源的區隔分析上，如統計模型化、自然影像的學習、自然的聲音、影像序列與醫學影像分析等未知分佈的資料上。

通常獨立成分分析法可以處理資料來源數量與感應器數量相等的情況。但是當資料來源的數量大於感應器的數量時，對於獨立成分分析法而言卻是一個艱鉅的問題。這稱之為過度完備的獨立成分分析法(overcomplete independent component analysis, overcomplete ICA)。在過去十年中，許多的統計學習演算法被提出來用以解決此過度完備的獨立成分分析(overcomplete ICA)問題。而且，過度完備的獨立成分分析(overcomplete ICA)也成功的被應用於過度完備(over-complete)的不明資料源區隔與具有過度完備基本資料集的自然影像的統計學習上。

而除了獨立成分分析法(ICA)與過度完備的獨立成分分析法(overcomplete ICA)之外，稀疏編碼(sparse coding)或是稀疏圖像(sparse representation)則是另一

種相似的方法，也成功的被應用於影像表現(image representation)與自然影像的統計學習上。

目前在台灣的一些計算機科學領域（如:電機工程、統計等）的研究學者已經在研究獨立成分分析(ICA)及其相關的問題，並且在其應用的問題上也獲得了具體的成果。

V. 應用(Applications)

(a) 統計神經影像術(Statistical Neuroimaging)

統計神經影像術的精神在於融和統計和資訊技術。統計學是可以被運用在大腦影像造影(brain imagining)的基本原理之一，在研究過程中所發展出的假設推論可透過統計的技術來加以解釋。更確切地來說，SPM—“參數影像功能映像區域之統計分析”(statistical parametric mapping)[1]在大腦影像造影的軟體中佔有領先的地位。另外，當影像程式變得愈精確、愈複雜和愈密集時，計算機科學和資訊處理將首當其衝。這類型的資料集合通常都含有數十到數百的 megabytes，甚至數十到數百的 gigabytes 的資料量。且為能對資料進行各式各樣的分析，軟體的功能和應用層面也成為關鍵性的重點。數以百萬計的個別統計分析，如迴歸分析(regressions)、主成分析(principal component analyses) 在處理多個資料集合時，往往因為有限的時間而受到侷限。漸漸地，將機器學習(machine learning)的一些構想結合神經影像術，以進行某些自動化的複雜性工作，或者透過資料探勘(data mining)的方法來實行探究型的分析。例如，自我組織網路(self organizing map)[2]技術可用於影像的分割，及獨立成分分析(independent component analysis)可說是應用在腦電掃描技術(electroencephalography, EEG) [3] 和功能性磁振造影(functional magnetic resonance imaging, fMRI)的一種領先方法論。更進一步來說，今年在一些具有領先地位的神經影像術會議(如 Human Brain Mapping)，特別設計成能結合統計和機器學習領域中的最佳構想的型式，這樣的技術將被廣泛認定為在大腦影像造影研究中的領先者。大腦影像造影技術的根源在於結合兩大群不同領域的學者像是計算機科學、統計學、物理學和工程學等擁有理工領域專門知識與素養的學者們，及藉由利用這些新興技術與想要探討大腦構造的醫學領域的學者來合作。透過在大腦影像造影研究而找到的構思和問題，不僅僅造福神經影像術本身，更可能造就上述的某些核心領域。當進行大腦影像造影所產生的許多問題的解決方法，需要多個核心領域研究中的創新想法時，對有興趣的學者而言，這將是一個持續富有成效的領域。

在台灣，統計神經影像術是一個持續發展的領域。在眾多的大學和研究機構中，都有對這個領域有興趣的研究團體和研究者，包含了統計和資訊科學，甚至遍及生物醫學工程(biomedical engineering)、神經科學(neuroscience)、心理學(psychology)和放射學(radiology)等應用學科。從最近跨大學、機構的研究合作中，可以發現到在台灣各個學門間的合作將更為緊密。台灣的研究團隊對統計神經影像術方法論所作出的貢獻不僅僅被應用於台灣，更為世界其他研究團隊所採用。台灣的學者和在美國、歐洲及亞洲其他研究機構的學者，也有一些正式的合

作計劃在進行著。

從世界的觀點來看，統計神經影像術是一個發展迅速的領域。那特別具有名望的神經影像術期刊，Neuroimage，也剛完成了將投稿的論文分成四個大類別的階段，其中一類即包含了方法論和模型的建構；而這個部分的編輯是統計學者，並非研究神經影像術的學者。這對統計學的方法論被納入研究神經影像術的體系中，提供了一個有效的佐證，這也將持續演變成較複雜的技術，而研究神經影像術的學者和統計學學者之間的互動將日趨重要。

台灣和世界性的統計神經影像術研究，未來的重心將持續引進各種不同學科以創造能激發更進一步創新的環境。我們期望先將這領域推廣至某個程度，接著再持續推動更深一層的共同研究。更可透過各種不同的方式來進行，包括教育的互相觀摩、對資料的處理角度和對結果的闡釋，但並不僅僅侷限於上述的層面。而合作的模式應可透過正式或非正式的形式，使資訊能更有效率的傳遞。

(b) 資料探勘

資料探勘的出現，一般咸認為源至西元 1987 年密西根大學的博士生 Fayyad ，暑期於通用汽車公司 (GM) 打工期間，為整合多個資料庫，以提供 GM 的技工快速查詢關於修理汽車的相關問題。進而發展出一套技術，此為資料探勘之濫觴。西元 2001 年麻省理工學院『科技評論』(Technology Review)元月號將資料探勘評為足以改變世界的前十大新興科技的第四名。在 2005 年回頭檢驗『科技評論』所做之預測，無論是在商業、工業、醫療、基因工程、氣象、安全產業甚至體育競賽，許多資料探勘的應用均如雨後春筍般地出現。足見麻省理工學院『科技評論』之權威性與客觀性。資料探勘（data mining）簡言之，是指由巨量資料集萃取出有隱藏其中具有價值之知識，亦即將資料轉換成知識的行為。拜現今的資訊科技之賜，資料的儲存已非昂貴的工作，且資料的蒐集也不僅限於傳統的抽樣調查或人工的收集方式。人造衛星每日可從遙遠的天際邊傳回地球上萬張的空照圖、7-11 便利商店的收銀機能自動地記錄每筆交易內容，並將其彙集至總部的資料庫中、網路伺服器上，完整記錄著個個連結的 IP，封包的傳送個數等資料、人類基因圖譜完成後，隨之而來的『後基因體時代』，產生出大量的生物及醫學資料等等。這些資料迅速累積，且日趨龐雜，其中也可能夾雜著許多錯誤，如何發展出一套科學且有效率的方法，以自動或是半自動的方式來探索和分析資料，將其去蕪存菁，將資料轉換成資訊，再由資訊結合專業，形成具有價值的知識。利用電腦快速的運算能力並結合統計、人工智慧、資料庫管理、機器學習等領域所發展的資料分析方法，對巨量繁雜的資料集進行整理歸納，希望能『看出』資料集的結構或隱含的規則，如此便能將一大堆雜亂無章的資料，轉換成有價值的知識寶庫。

資料探勘在程序上可分為：瞭解問題、資料蒐集、資料整理、模型選擇與建立、模型評估與結果的解釋、策略制訂，最後可以更進一步瞭解問題。為一個循環的過程。在每一環節，均可窺見統計的影子。

資料探勘的問題可概分為：預測（含分類(classification)與迴歸問題）、分群

(clustering)、資料視覺化(data visualization)、關聯分析（association rules）等等。解決這些問題的方法，在前述章節均有提及。

資料探勘(data mining)促進了各學門之間的合作。在各學門中，過去數十年來均各自成功且成熟地提出解決各類資料探勘問題的方法，例如：類神經網路、決策樹、支撐向量法、Fisher linear discriminate analysis、貝氏網路與各式分群演算法等等。某些方法具有堅固的統計學理，某些則是基於一些直覺與 heuristic。

在一些商業案例成功之後,現在有更多的研究學者願意試著將這種分析過程引入各自的研究領域中。舉例來說，資料探勘已經被製藥公司應用在藥品的研發上。另外，資料探勘在生物資訊的研究領域中，也同樣牽涉文字/文件(text/document)的分類技術。因為這種資料分析的過程中，同樣引用到大量的統計工具與資訊工程的方法論，而這些相關的領域，如：樣式辨認(pattern recognition)、機器學習(machine learning)，都比過去更受人歡迎。這對於統計學家來說是一個很好的機會,可以透過結合各式各樣的資料探勘計畫並將其引入許多有興趣的研究主題中，也能協助統計學家發展出更好的統計方法論。

(c) 統計在網路分析及資訊安全的研究

統計科學廣泛利用於各領域資料分析的工作。近代以來，電腦科學提供資料分析演算法運作(軟體)的計算平台(硬體)。電腦的運算由早期的超級電腦，演變為後來個人電腦的普及化。又由於網路技術的進步，個人電腦或單位計算單元之間由當初單機的作業連結成複雜的大規模網路架構。網路的連結使得資源與資訊的共享更為迅速方便。這樣的方便帶來了前所未有的現代化生活，譬如網路的搜尋儼然成為一個非正式的電子圖書館,或是擁有無線網路的攜帶式機制提供全球定位的服務。網路的服務帶來了方便，也讓非善意對計算機制的入侵有了更直接的管道。早期的病毒(virus)或駭客(hacker)的入侵以磁碟讀取的形式傳遞，現代的入侵大多仰賴網路的傳遞。這樣的情形使得資訊與計算安全的課題益發顯得重要。對於計算機制的入侵，由於方式的推陳出新與多樣性，往往讓阻擋者防不勝防。往往這時代精密嘔心瀝血的防禦設計在下一個時代新技術推出的同時失去功能。

資訊安全的防護方式可以簡單分為兩大方面：第一類的方式是在安全的防護上針對之前被破解的部分作設計上的更新或安全檢查的加強,彌補之前被攻擊的破洞，譬如加解密的技術運用於一些保密等級較高的資料防護，或是儲存相關資料系統的防護。這類防禦方式可以針對攻擊模式對症下藥，找出直接防禦的方法。但是由於攻擊技術的不斷進步，往往使得防禦者總是比攻擊者慢了一步，也使得計算系統暴露在不可預期攻擊的陰影之下。第二類的防禦方式是針對攻擊的動機來下手。一般的入侵或攻擊不外乎想要利用被攻擊者的系統資源，如計算資源或記憶體資源等，或不當獲取未經授權閱讀的資訊。這類的攻擊由於和正常系統使用者使用模式的不同，造成統計上的差異性，所以一般來說，我們可以利用統計上的分析做出入侵與否的判斷。統計分析在這方面提供一個優異的架構。

一般利用統計分析來判斷入侵多半是以非即時(off-line)的方式來進行，在一

般傳統安全防護以外，提供一個第二線的防護機制。這種處理方式容許較複雜、計算量較高的分析工具，也使得統計的複雜架構有發揮的空間。我們可以用兩個例子來說明統計在網路分析上及在資訊安全防護上所提供的協助。

針對一個或數個計算單元，我們可以測量其網路的使用量，譬如主機(server)端在單位時間內收發電子信件的頻率，傳送電子信件平均需要經過的路徑長度，網路在單位時間內傳送的封包(packet)數與平均封包的大小，網頁在單位時間內被閱讀的次數等。統計方法針對這些度量，提供一套模擬機制。這些模擬，可以消極的提供入侵偵防的判斷依據，也可以積極的提供系統效能優化的參考。以多個計算單元來說，我們可以測量諸個單元彼此網路流量等的相關程度，建成計算單元對多樣網路數據的矩陣，以判斷這些單元的關係，或是猜測這些單元中那一部分有否同時地被入侵。

第二個例子是針對一個系統，我們可以利用統計方法設計一個入侵偵防的防護軟體。譬如將某一固定遠端網路使用者針對己方主機運作的指令集(譬如所有system calls)看成一個時間序列，以隱藏式馬可夫模型的模擬，我們可以推測彼端正在進行的運算，或判斷是否為合法允許的運算類別。或利用彼端在單位時間內運算集的蒐集成一個特徵集，我們可以利用分群或分類的技術，譬如支援向量機、決策樹、類神經網路等來推估彼端是否為入侵者。

廣義來說，在網路分析、資訊安全、系統安全(system security)、軟體安全(software security)或實體安全(physical security)等相關領域下，統計架構可以提供的分析與防護還可包括垃圾信件的判讀，運用文件分類或分群技術來判斷是否為垃圾信件；或是監視系統(surveillance system)，譬如利用影像中的物體偵測與追蹤的技術來判斷是否為小偷等方面。這些項目都有統計分析極大的發展空間，相關的統計技術，諸如時間序列、MCMC、統計學習等都可以直接或間接的對這些問題的解決有所幫助。相關的研究可以參閱文件分析或影像分析等的相關論述，不在此詳細說明。

統計分析對於網路流量分析或是資訊安全的幫助在於，統計分析可以當作所有入侵偵防的最後一道防線[1]。大部分的入侵都會在己端主機資源的使用上透露出異常的狀況，所以統計分析可以提供這些異常狀況偵測的工具。統計分析架構的優點是，這些偵測一般並不會受到入侵技術演變的影響，所以一套機制可以長時間的擔任入侵偵防的工作而不會過時。缺點是統計為基礎的架構不免的都會有誤判(false positives 和 false negatives)的情形。這類誤判過高常造成使用者的困擾，最後造成安全管理人員經常性的移除此項防護，失去它的效果[2]。所以如何降低誤判率成為統計分析架構是否可以成功運用在入侵偵防上的一個關鍵。一般的安全管理公司常常搭配適當的人力編組，以應付各式各樣不同層級的安全警示狀況。無論如何，統計分析模式和一般安全防護的著眼點不同，所以可以雙方面

[1] 一般統計分析來擔任入侵偵防的工作多搭配其它入侵偵防的機制，特別是一般統計為基礎的入侵偵防多是以非即時的方式來進行，必須搭配一套即時的入侵偵防機制。

[2] 當然一般而言，入侵者或非入侵者，或是正當使用者或非正當使用者的角色常常不易清楚界定。

互相搭配使用，增加安全防護的深度與廣度。也因為如此，統計各類理論架構可以更實際的運用在網路分析和資訊安全相關研究上。

網路效能監控和資訊安全在近年來因為網路技術的進步，有益發重要的角色扮演。在各類實體或系統恐怖攻擊的事件之後，網路及資訊安全逐漸受到大家的重視。1992 年 Texas A&M 大學電腦中心遭受惡意攻擊，造成系統癱瘓；1999 年一位瑞典籍駭客攻入 Microsoft 的 Hotmail 的網站造成電子信件內容外洩；另外一位俄籍駭客入侵電子商務網站偷取三十萬筆信用卡卡號，如此例子不勝枚舉。一般網路的分析著重於利用統計的方式依照時間推估各類網路服務的需求，以辨別對於已端資源不當使用的狀況。積極的分析，可以進一步使得系統的效能最佳化。安全的議題方面則可以分為加解密技術、電子簽章、通訊安全、軟體安全、身分認證等。這些議題大部分都可以利用統計分析的方式，在各項安全防護的技術之後，提供另一層的保護。近年來由於一些新興入侵方式的出現，如木馬程式(Trojan horse)，分散式阻斷攻擊(DDoS 或 distributed denial of service)，側面頻道攻擊(side channel attack)，網路蠕蟲(worms)等，益發顯得統計角度切入問題的重要性[3]。目前一些較熱門的研究議題為一、觀看一個主機的正常系統或網路運作情形(可能也需要參考一天、一週或一年等的時間因素)，辨別主機是否遭受不當使用或非法入侵；二、利用多個主機的系統或網路運作狀況估計其是否為正常狀況，如果遭受到入侵，被入侵的範圍為何；三、嘗試提升統計分析的效率，以使得分析的模式可以即時(on-line)的運作在網路等的控管上，譬如調整網路路由器通行規則(rule-based approach)的順序可以提高監測的效能等；四、觀察網路架構嘗試提供較為迅捷的網路傳輸方式，譬如封包以最短路徑傳輸等。

在無線網路方面，無線網路科技在這幾年來迅速的發展，使得相關研究的重要性相對提高。無線網路與傳統網路的差別在於無線網路的「網路」本身是不固定的，同時也有較多潛在失去聯絡等的不可預期因素，所以一般的研究必須考量這兩方面的特殊性。無線網路安全方面，無線網路的迅速發展也帶來一些安全防護上的問題。早期的無線網路設備並沒有對於安全方面的問題有太多的著墨，大部分傳遞的訊息都是採取無加密的方式在傳送。但隨著無線網路使用者的增加和譬如無線感測網路(sensor networks)等廣泛應用於戰爭佈局、地震感測、地雷移除、交通流量分析、森林大火感測等重要領域，使得無線安全的問題逐漸被人所重視。對於無線網路的統計分析方面，研究者必須考量無線網路的幾何性質和網路架構不可預期性這兩大因素，作出對系統的有效分析，或是判別是否被入侵的有效估計。目前較重要的統計分析包含對 IEEE 802.11 協定等的感測器部署(deployment)、訊號及雜訊強度的分析、訊號雜訊比、節約的能源使用、傳輸速度、傳輸資料丟失比例等議題。至於安全方面的監測，統計分析的架構必須要考量無線網路計算單元一般擁有比較低的計算能力，所以統計的計算方面必須要適度的簡化，以達到低能量、低運算成本的客觀需求。

總而言之，在網路及資訊安全的領域仍需要相當技術上的突破，以應付未來

[3] 理由如前段所述。

資訊及網路時代的全面來臨。統計方法的角度一般仍並不為主要的資安領域所重視或採用，也因如此，統計分析在此領域不論在理論或實務上均有極大的發展空間。

VI. 合作研究

(a)從資料探勘，機器學習，統計學習理論到合作研究

資料探勘的名詞由資訊管理的專家們提出;在當時僅僅是希望從既有的資料庫中挖掘出有用的資訊。而由於資料量龐大，資料本身的前置處裡往往得藉助資訊工程師的幫助。然而資料探勘或後來的資訊及知識萃取並非指一個單一工具或方法，而是在資工或資管包裝下的統計資料分析過程，其中包括敘述性統計、統計模型建構、資訊的呈現。其所包含的資工技術大略是資料庫的技術，其所包含的分析方法則五花八門，從分類、分群、回歸分析、圖像辨識、資訊視覺化、多變量的各種方法到目前相當熱門的統計〔機器〕學習的方法。過程中需要統計學家與資訊工程師們充分合作，但往往彼此都認為對方的"專業"只是"技術細節"罷了。合作成功的例子並不多見，資訊管理人在商業包裝下，使得資料探勘成為話題，統計也因此更受重視，但跨領域的合作仍待加強，尤其是想把此一概念運用於"科學資料"時，我們更得融入該科學主題的專業知識，團隊合作的需求更高，而這正是我們〔統計〕教育中嚴重欠缺的部份。

我們以目前最受矚目的支持向量機(support vector machine)在資料探勘的應用為例。資料探勘的統計過程，有一大半是分類問題。在分類的工具中，核化的算則(kernelized algorithm)是常被用於資料探勘的問題,因為它有處理大量資料的演算能力。核化的另一好處是，對許多以"內積(inner product)"演算為主的傳統統計方法(如 PCA, FDA, CCA)，核化提供了一個相對容易的擴充平台(framework)，在幾乎不更改原有程式的情況下，達到非線性的性質。

這些研究大多是受支持向量機(support vector machine)興起的影響，而支持向量機(support vector machine)本身也成為是分類及迴歸於大量資料分析時最常使用的工具。而從發展的歷史來看，支持向量機(support vector machine)起源於 Rosenblatt (1957) 的 perceptron；而 perceptron 的發展更可回溯到更早的 McCulloch and Water Pitts 於 1943 年的論文 "A Logical Calculus of Ideas Immanent in Nervous Activity"。而此一發展的誘因，甚至可追溯到西班牙及英國的神經學學者 Dr. Cajal 和 Dr. Sherrington。其中，Dr. Cajal 更是一九○六年的諾貝爾獎得主。

資料探勘(data mining)促進了各學門之間的合作。在一些商業案例成功之後，現在有更多的研究學者願意試著將這種分析過程引入各自的研究領域中。舉例來說，資料探勘已經被製藥公司應用在藥品的研發上。另外，資料探勘在生物資訊的研究領域中，也同樣牽涉文字/文件(text/document)的分類技術。因為這種資料分析的過程中，同樣引用到大量的統計工具與資訊工程的方法論，而這些相關的領域，如：分類(classification)、分群(clustering)、資料視覺化(data visualization)、樣式辨認(pattern recognition)、機器學習(machine learning)，都比

過去更受人歡迎。這對於統計學家來說是一個很好的機會，可以透過結合各式各樣的資料探勘計畫並將其引入許多有興趣的研究主題中，也能協助統計學家發展出更好的統計方法論。

六、 地球科學及環境相關之統計

「地球科學與環境科學」涵蓋的範疇很廣，舉凡地球上對自然生態、環境和人類生存有關的活動都屬其範圍。與此相關的統計研究很難簡單且具體的描述其歷史發展，其涵蓋的領域亦分散在農業、生物、土木工程、大氣科學、生態學及其他許多行業中。最近數十年來，自然環境隨地球大環境的改變和工業急劇發展而產生巨大變化人類生存環境也因之受到影響。大環境的氣候變遷、地球暖化等使得全球各地地震、風災、水患等災禍頻傳；小環境空氣污染、電波干擾、暴露物質等直接對人類生命健康產生危害。因此有關研究亦刻不容緩的快速發展開來。由於相關的問題攸關國計民生，其所涵蓋的統計問題與資料包羅萬象，統計學者的介入與日俱增，並積極探討統計在複雜的地球科學與環境科學中的應用，同時也發展出許多新的理論與方法。

國外關於這方面最活躍的研究中心有賓州州立大學環境研究中心、芝加哥大學統計與環境科學整合中心、哈佛大學生物統計系之環境統計組、俄亥俄州立大學統計系的空間統計與環境科學研究群、美國國家中心大氣研究之地球物理統計計畫、及設於華盛頓大學之國家統計與研究中心等等。有關前述的各方面均有專屬單位研究，主題無法一一列舉，但綜合其研究領域大致有以下幾個方向：

(1) 決定論過程模型與隨機模型 (Deterministic Process Models and Stochastic Models)

此方面以專業的物理機制或理論等建立的決定論模型輸入，再藉由隨機模型的不確定性來分析並配適資料，例如空間-時間模型的建立。具體的應用範圍有大氣、地震、海洋、氣候等資料。

(2) 相關性資料與環境趨勢 (Correlated Data and Environmental Trends)

地球科學及環境科學資料大都是依時間或空間收集採樣而得，因此資料多具相關性，此方面研究以時間序列模型居多，且主要目的在探討模型是否隨時間或空間而改變並研究其改變的趨勢。應用範圍包括氣候變遷、環境變遷、生態影響等。

(3) 統計模型與科學認知 (Statistical Modeling and Scientific Conceptualization)

將資料的變化依其專業領域、認知觀念建立統計模型。例如藉由統計與生態學家的結合已推倒出一些紮實的科學理論；其他如環境法規、健康與風險分析、暴露物質與職業病等相關資料。

目前國內關於地球科學及環境相關之統計研究主題包括「地震活動與前兆之統計分析」、「颱風降雨量與風速之統計預測」、「空氣污染與健康危害之關係」、「人體測量值和環境評估之統計方法與應用」、「環境與健康相關風險評估」以及「生態統計」等。以下是各主題的詳細論述：

I. 地震活動與前兆之統計分析

地震活動是地球科學的主要研究項目之一，相關研究包含地震測量、地震震源物理、地震危險、地震前兆等。統計的應用主要在後兩者。傳統上，以非穩定波氏過程描述地震的發生率，藉以評估強震或餘震的風險，方法除涉最大概似估計，亦有貝氏分析的應用。近年來，地震統計除地震目錄外，亦輔以地震物理相關研究成果，因此，強震的風險評估需考量大地移動－地殼滑移率的影響；餘震風險可能與其主震引起的應力改變有關。這些在地科方面屬於熱門的研究，時空統計之應用是可期待的。

地震前兆的研究，傳統為地震寧靜的研究，包括地震發生率的降低及地震規模分布的改變等。與地震目錄無關的地下水位升降及氡氣的變化也是可能的地震前兆。希臘學者在 1980 年代發展地電相關的前兆，唯地電異常並無科學性的認證。日本在 1995 年神戶地震後進行電磁相關地震前兆研究，忙於測站的設立及數據的收集，卻無地震或統計學者參與進行數據系統化的分析。因為美國學者研究發現岩石碰撞產生正電子，台灣在 1999 年集集地震後亦展開相關的電磁地震前兆研究。此一領域的研究除涉前兆訊息或異常的統計鑑別，也涉及前兆訊息與地震發生耦合的相關性分析。目前研究著眼於驗證前兆訊息的相關物理機制及檢定前兆訊息助益地震預警的顯著性。

II. 颱風降雨量與風速之統計預測

台灣地區每年平均遭受三、四個颱風侵襲。颱風夾帶豪雨帶來人員的死亡、失蹤、房屋的毀損、土壤的流失，導致農漁業嚴重的損失，也直接影響到國計民生。近年來由於有較完整的觀測、預報及防範措施，因颱風侵襲所引致人員死亡與房舍毀壞的情形已較過去有許多改善，但颱風仍是造成台灣地區最主要的氣象災害，如 1996 年賀伯颱風帶來的豪雨，帶來近百人死亡和失蹤，光僅農業即達 199 億元。因此提升已有之雨量預報方法實為刻不容緩的。目前氣象局所收集的颱風資料包跨颱風之日序、經度、緯度、中心風速、時雨量、絕對距離、相對位置、測站風速、風向、表面壓力等，可謂十分完整。從統計學資料分析的角度來看，這是一筆非常豐富的資料，也牽涉到很多深淺不一的統計問題。

颱風作業預報，以往多以強度和路徑為主。雖也有不少的文獻針對颱風降水做探討，但國外對颱風降水作業預報之討論實則相當有限。反而台灣地區對這方面的問題，一直受到作業與研究單位的重視，並有許多學者嘗試發展更有效的預報方法以進行颱風降水量的預報。目前氣象局預報作業參考方法為首先對颱風的移動路徑作分類，再以過去通過固定地區降水量之平均作為預測雨量；另一方面亦有以可能影響颱風降水的八個預測因子用比擬法作颱風降水預測；或先對颱風移動路徑分西進與北進兩類，再以平均法做預報，並又考慮降水量與中心最大風速之比值的平均作降水預報，是為比值法。然而形成颱風的因素很多，其產生的降雨量亦為颱風本身的特徵之一，故降雨量和颱風的其他特徵或颱風的成因之間應存在著某些關係。目前國內關於此方面的研究仍以大氣科學學者就颱風發展之物理意義以決定論模擬方式進行雨量或風速的預測，因此投入大筆的經費與人力

在提升資料收集的準確性，如衛星雷達、無人飛機偵測等更新設備方面。雖然有大量豐富的資料，反而與統計界互動非常有限。國內統計學者投入此領域的研究很少，仍利用敘述統計量、迴歸模式、多變量分析等傳統統計方法對資料進行分析並對雨量與風速分別進行預測；近來也有一些嘗試如貝氏模型選擇、資料探勘、長時期模型與空間統計等等。

III. 空氣污染與健康危害之關係

目前空氣污染與健康危害的熱門研究重點包括多種空氣污染物混合物的聯合健康效應、尋找危害健康的主要污染成分及污染源和低污染物濃度對易感受族群的危害等。國外從事這方面的研究都是由流行病學、毒理學、環境科學及統計學等不同領域的專家所組成的團隊來執行。研究方式主要包括流行病學研究、動物實驗、細胞毒性反應及尋找關鍵基因等。常面臨的統計問題有實驗的設計與規劃、大量時空資料的分析及模式的建立等。這方面的研究可以提供我們進一步了解各種空氣污染物對健康的危害程度，同時提供政府環境主管部門訂定較正確的空氣污染物管制標準的依據及管制策略以維護大眾健康。

IV. 人體測量值和環境評估之統計方法與應用

環境暴露評估為環境風險評估（另見學門規劃生物統計的「環境與健康相關風險評估」）中極為重要的一環，風險分析與評估的正確性有賴於良好完善的環境暴露評估。環境暴露的健康風險評估對象，主要可區分為受污染地區的社區居民，以及生產線上職業暴露的勞工。前者由於污染物的擴散有時間與空間分布的特性，時空統計（spatial-temporal）方法的發展與應用相對重要。至於職業暴露評估，由於作業環境相對集中，較不需考慮地理空間的分布，因而重點在於作業流程中暴露濃度時間性的監測數據是否超過標準，與工人的體內累積濃度是否對健康造成危害，而傳統統計方法也較偏重時間數列或極值理論（extreme value theory）。

由於生產線工人長期身處第一線高風險暴露，在工業衛生安全規範上，必須制定安全暴露閥極值（threshold limit value, TLV），因此作業環境暴露量的監測也愈顯重要。一般環境暴露監測器的設置點為固定位置，但工作場所工人為四處走動，加以個體之間對毒性物質代謝能力的差異性極大，因此以環境測量值作為暴露規範可能無法實質保護多數勞工。近年來國際一些相關研究學會如 American Conference of Governmental Industrial Hygienists (ACGIH), the Deutsche Forschungsgemeinschaft (DFG), the Japan Society for Occupational Health (JSOH) 等倡議以人體內該化學物質或其代謝物的濃度作為生物指標值（biological exposure indices, BEI），用以取代相對應的外在暴露閥極值 TLV，其做法通常為量取工人血液或尿液中適當的化合物濃度，稱為生物標記（biomarker），用以正確反應工人本身的實際暴露量與身體負荷（body burden）。在統計分析方面，相關文獻主要在探討生物標記與外在暴露濃度之間的關聯性，例如美國北卡大學工業衛生學者 S. Rappaport 與生物統計學者 L. Kupper 一系列利用 mixed model 建

立兩者之間關係的文章。但由於簡單的迴歸模式不足以描述人體複雜的代謝機制，以及動態的體內濃度，欲正確反應兩者之間的關聯性，模擬實際生理機制的人體毒物動力學模式（physiologically-based toxico-kinetic model）的應用也日益受到重視。

人體毒物動力學模式主要將人體依組織分為肝臟、脂肪、血液、肌肉等空腔，並模擬化學物質進入人體後，在各空腔傳送途徑與代謝機制，用以描述化學物質或其化合物在體內特定組織的動態濃度。由於此模式牽涉許多因人而異的未知參數，且樣本人數通常少於 20 人，嚴謹的統計分析方法直到 Bois & Gelman (1996) 在貝氏統計的架構下，利用 MCMC 方法才得以解決。但 Bois 等人將模式理論值與人體測量值的誤差視為觀測誤差（measurement error），並未考慮不同時間點的濃度值所牽涉的過程誤差（process error），同時考慮此兩種誤差的模式結構應為 state-space model。最近丹麥學者 Overgaard, Kristensen (2005) 等人在 state-space model 架構下提出模式參數估計的方法，但其統計假設與結果存在若干缺失，且缺少實例說明，相關研究的統計方法仍有相當的改進空間。

藉由人體毒物動力學模式參數估計的確立，人體測量值或生物標記的實際應用，在職業暴露法定規範方面，可以用變異性較小的生物指標值 BEI，取代外在暴露閥極值 TLV，且由於可描述 BEI 的動態行為與體內實際暴露值，對於勞工整體與個人，均能發揮安全防護的最大功效。在環境污染評估方面，由於污染廢棄物的確實場址或污染範圍不易偵測，但可藉由受影響的社區居民體內測量值，量取其隨時間累積的暴露量，間接評估環境污染程度。此處統計方法的應用，除了毒物動力學模式之外，也與時空統計與抽樣理論密切相關。國外目前美國俄亥俄州立大學統計系由 Cressie 領導的 SSES 研究團隊，也有將時空統計方法應用在相關的研究課題。

國內目前有關於人體測量值在環境暴露評估的研究，主要為環保署或勞委會委託工業衛生或環境學者所作的若干勞工職業暴露的研究計畫，統計學者的參與，有待積極努力。

V. 環境與健康相關風險評估

環境風險評估（environmental risk assessment）包括健康與生態風險評估（health and ecological risk assessment），是一門典型的需要跨領域整合的新興科學。其牽涉的領域涵蓋環境科學、統計學、毒理學、工業衛生、流行病學、生物學與生態學等，影響的層面更上至環境與衛生政策的制定規範，並廣及社會大眾的風險認知與行為。而統計學家在其中所扮演的角色則為根據蒐集所得的相關資訊與數據，在與各不同領域專家學者密切合作討論的過程中，釐清屬於統計範疇的隨機變異，以建立正確的統計假設與模式，將相關因素的不確定性降到最低，並據以推估相關可承受風險與建議界定理論值。

以美國環保署現行的風險評估標準程序為例，包括有危害物質鑑定（hazard identification）、暴露評估（exposure assessment）、劑量反應評估（dose-response assessment）、以及風險特性化（risk characterization），除了危害物質鑑定為環境

與毒理專業外，其他部份都需要有統計學家的密切參與，以下分述各程序中與統計相關的部份與需要的統計工具，以及可能的未來發展與建議。

在暴露評估方面，主要為污染物的傳輸與沉降（pollutant transport and fate）評估，其中有些污染物由人體直接攝取（如汽機車與工廠排放的廢氣、受污染的飲用水等），有些則經由食物鏈進入人體（如工業製程與焚化爐燃燒所產生的戴奧辛沉降到土壤、河流、地下水，經由養殖魚貝類、畜牧動物、蛋奶製品、農作物等食物進入人體）。國內目前統計應用成功的範例主要集中在空氣污染與人體健康關聯的環境流行病學研究，其他本土亟需統計學者投入解決的問題如：食品衛生的戴奧辛毒鴨蛋事件污染擴散程度與污染源的確定、廢五金與石化工業的污染、垃圾焚化爐燃燒所產生的戴奧辛與重金屬污染、環評問題中交通流量對周圍生態水質污染的影響等，皆屬於環境暴露評估研究的範疇。

傳統上環境學者會利用一些簡單的敘述統計量與迴歸模式對所採樣的數據進行分析，並得出結論。但由於此類數據一般而言都具有時間與地理分布的特性（例如空氣品質監測站資料），很自然地會使用到時空統計（spatial-temporal statistics）或地理統計（geostatistics），目前層次貝氏時空統計（hierarchical Bayesian spatial statistics）利用 MCMC 計算模擬被廣泛應用，結合衛星定位與 GIS 資料，在理論與應用研究上都有極大的發揮空間。在領域知識（domain knowledge）方面，污染物本身有其化學特性，而擴散模式則必須遵守物理定律，其動態行為預測有賴於複雜的偏微分方程解，單純依據有限觀測數據所建構的統計模式與假設可能與事實不符，因此現行時空統計方法可能有其缺失與盲點，這也是統計學者不容易參與相關研究原因與面臨的挑戰。至於經由食物鏈的暴露，則必須在接著污染物在空氣、水、土壤各介質的暴露評估後，結合生態風險評估，以估算受影響生態系統（ecosystem）中生物體的攝取量，才能進一步作相關的健康風險評估。其他研究課題如工業衛生的勞工職業暴露評估，此類問題較偏向時間性暴露值指標與體內累積濃度，時間數列與利用人體毒物（藥物）動力學模式（physiologically-based toxico-kinetic or PBTK model）估算動態體內濃度，為相關可使用的統計方法。

在劑量反應評估方面，傳統統計方法的應用，主要偏重在根據動物毒性實驗數據，建立可行的劑量反應模式。其做法通常對實驗動物施以極高的劑量，再根據所建立反應模式，外插到極低的劑量風險，乘以不確定因子（uncertainty factor），以制定致癌物（如戴奧辛）或非致癌物（如食品添加物）的每日可攝取量標準。但由於其中牽涉到物種外插（動物到人）、高劑量外插低劑量毒理機制與模式等諸多不確定因素，統計推估的合理範圍有其局限。近年來 benchmark dose (BMD)計算逐漸取代傳統的最高無害劑量（NOAEL）計算，相關的統計方法發展值得注意。一個可行的替代做法為環境流行病學研究，在已知的高危險暴露族群，根據其環境暴露評估與疾病盛行率，找出直接的人體劑量反應的關聯性。另外，此處正確劑量應為體內劑量（internal dose），而環境測量值與實際人體攝取量或體內劑量之間仍有相當的落差，兩者之間關係的確立有賴於 PBTK

模式的引進與模式參數的估計，相關統計方法的研究與應用也是方興未艾。

作為風險評估的核心課題，風險特性化結合暴露評估與劑量反應評估，計算相關風險，並綜合考慮參數與模式等不確定性因素進行不確定性分析(uncertainty analysis)，以提供決策者風險管理的參考。一般環境學者在做不確定性分析，主要利用 Monte Carlo simulation 的方法，但專業知識的不確定性難以電腦模擬評估，如何將專家判斷（expert judgment）以機率分布描述，引進不確定性分析，為統計學者一大挑戰，貝氏統計提供一個可能可行的方向，最近 Imprecise Probabilities 相關的研究方法也值得參考。

有別於環境污染物的健康風險評估，食品安全衛生如禽流感、狂牛症、口蹄疫等傳染性病毒，或抗生素使用等所造成食物的污染與食用疑慮，均為微生物風險評估的重要課題。其評估程序與環境風險評估相同，但做法與內容則和流行病學、微生物學等領域密切相關。其中衍生的許多統計問題，需要更多的統計學者投入參與，共同努力解決。

VI. 生態統計

生態統計主要研究動物和植物之族群（或群落）數目、遷徙、生存和死亡之動態發展以及與環境空間分布的關係。可以看成「生物多樣性」研究的一部份。由於國際上有 188 國簽署「生物多樣性公約」，人類正積極挽救快速消失的物種及基因，同時亦防止生態體系劣化。因此生態統計及其相關的環境統計在生物多樣性的研究中扮演重要的角色，在此主題下，主要研究包含

生物族群的抽樣方法：傳統的抽樣方法無法應用在動植物族群上，因此必須發展特殊的抽樣方法（重複捕取、繫放回收、線上穿越、捕取力抽樣、除去法抽樣、距離抽樣、逐次抽樣、網格抽樣、穩健抽樣與 adaptive 抽樣等）。

生物多樣性指標的研究及統計估計：生物族群的描述往往無法由一、二個指標來刻畫，必須發展不同角度的多樣性指標（生態學家有α指標、β指標與γ指標，也有地方性、區域性與全球性指標）。

估計生物族群參數的統計推論研究：由於生物抽樣中個體（或種類）有部份無法抽樣到，因此相關的統計推論、模式選擇方法會隨抽樣方法的不同而有所差異。由於參數數目往往很多，因此多參數模式估計與選擇亦為一主要的研究主題。

七、 社會統計

I. 財務統計

(a) 財務時間序列模型

資產價格或其報酬率為財務領域最常見的時間序列資料。 經 70~80 年代研究， 主要性質如前後期相關不顯著， 波動率隨時間緩慢改變且向均值回復(mean reversion)， 報酬率分佈厚尾(heavy tail)等現象， 均已獲得共識。 早期的模型多以連續時間的擴散過程(diffusions)表現， 但這類模型通常較不易進行統計推論。 ARCH 模型則打開了一扇門， 讓統計方法在此領域有一著力點， 而近年來 ARCH 的各種延伸型式也已被廣泛探討及應用。 除此之外， 資產價格的長

期記憶(long memory)已是熱門的課題， 而為解決前述厚尾現象， 多種分配及相對應的隨機過程也吸引相當多的學者投入。同時，利用現有的模型(如 ARCH)進行區域平滑(local smoothing)的半母數方法亦日漸受到重視 。 簡而言之， 未來的財務時間序列模型將朝向更貼近資料特性發展。

(b) 衍生性商品定價

自 1973 年三位諾貝爾獎級的學者 Black, Scholes 及 Merton 發表了計算歐氏選擇權(European Option)價格的文章以來，衍生性商品定價這個新興領域即以驚人的速度發展。近十幾年來，不少機率、統計、數學及科學計算的學者專家陸續地加入數理財務這個領域之研究。在這個領域較常用到的數學工具有統計、隨機分析及對局論。其中對局論主要是建構在奈許的平衡(Nash equilibrium)理論，討論市場在平衡狀態下的運作情形,常被應用來討論市場機制與市場資訊的情況以及債券，股票與期貨等金融商品的價格變化與投資者的投資策略等。進入 90 年代後， 由於計算機的快速進展，文獻許多快速的演算法被提出用來計算各種制式及奇異選擇權的定價相關問題。選擇權可分為美式與歐式選擇權，在實務上，現今大多數交易所的股票選擇權都是美式選擇權,股票指數選擇權則有歐式和美式，外匯選擇權以美式為主，期貨選擇權也多為美式，因此美式選擇權的定價為一實際且重要的課題。由於美式選擇權具有提早履約的特性，因此價格比相對應的歐式選擇權高，而提早履約的特性也是影響美式選擇權定價的重要因素。如何掌握最佳提早履約的時間點一直是許多學者持續研究的問題，如 Lai and AitSahlia (2001), Longstaff and Schwartz (2001)等人。此外如何利用 Black-Scholes 公式導出的隱含波動率來推論市場動態在近年也有不少的發展， 其中無母數方法扮演重要角色。 另一重要的趨勢則是以天氣、經濟數據作為標的的衍生性商品的出現， 前者如溫度， 後者如通膨及國民生產毛額， 這意味著對於更複雜的動態及迴歸模型有強烈的需求。目前文獻中對於定價的探討，大多是在完整市場(complete market)的假設下利用鞅論推導衍生性商品的無套利價格。然而真實市場大多是不完整市場，因此近年來已有不少研究開始探討在不完整市場假設下，衍生性商品的定價與避險問題。

(c)風險管理

近年來,由於大環境持續不景氣國內外許多知名公司發生嚴重的財務危機甚至於造成公司倒閉，這個現象也直接影響到衍生性金融產品（derivative securities）的定價。因此不少國內外的學者, 如 Jarrow and Turnbull(1995), Duffie and Singleton (2003) 等人均投入包含信用風險（credit risk）的債券（bonds）和選擇權（options）定價問題的研究，。信用風險模型可供銀行作為風險管理和績效評量之參考依據，亦可為投資人帶來更多的資產保障，另外，也可作為主管機關金融監理與檢查之參考基準。因此，考慮信用風險的因素後，如何推導合理的定價模型乃成為一個相當重要的課題。信用風險包含二個部分：標的公司本身倒閉的風險和衍生性產品賣方不履約的風險。目前文獻上較常採用包含信用風險的定價模型為: (a)結構式(structural form model) (b) 縮減式(reduced form model)兩種模型 。

(d) 市場的微結構與高頻資料研究

交易市場由其交易方式可分為 (a)委託單驅動市場（order-driven market）：委託單經由經紀商傳到交易所，透過 auction 方式與其他人的委託單撮合之市場，例如：台灣證券交易所。以及 (b)報價驅動市場（quote-driven market）：非

由 auction 方式撮合，而是由造市者（market maker）提供報價，與大眾之委託成交，例如：倫敦證券交易所。重要的研究課題包含買賣價差（bid-ask spread），市場的流動性（liquidity），價格的效率（price discovery）以及資訊傳遞的速度。其中買賣價差的研究，主要是探討資訊不對稱成本、庫存成本及訂單成本在價差中所佔的比例。資料的型態已由早期的日資料演變成每日每筆的交易資料，因此高頻資料的分析對是日後研究發展將佔有一個重要的地位。

II. 教育統計

教育統計實為教育與統計兩個學門的綜合，也可視為將統計應用在教育情境，故本領域的學者也多數申請教育學門的計畫。在人才方面，不論是心理計量或教育統計的研究人力國內並不多，且本領域為跨學科整合，背景知識需要理科與文科兼具，故在傳統的教育學院研究人力的培育較為緩慢。

本領域發展出三個重要的方向：

(a) 教育研究或實驗中的各種統計方法

國內在教育實驗方面，並沒有太多的發展，大多依循國外所發展的方法及電腦軟體，以 SAS, SPSS, LISREL 為主要工具。研究集中在統合分析、樣本數估算、違反假設的強健統計量等。

(b) 教育測驗裡的評量理論及技術

評量理論及技術則是國內教育統計學者的主力，包含有 IRT (Item Response Theory) 理論、國中基測的實施、Fuzzy 在量表上的應用等。有關電腦化測驗、電腦模擬實驗等，都需要數學、統計、資訊、及學科專家的投入與整合。有鑑於大型測驗如國中基測的重要與影響深遠，以及資料庫的興起、國際間比較等議題不斷出現，本領域實有進一步扶持之必要。

(c) 統計推論中校正測量特有的誤差

國內統計界有少數研究測量誤差校正，但並未和實際應用結合。

III. 認知科學

(a) 分析資料為高科技所產生的儀器訊號

以往認知實驗研究所探討的注意力或知覺等，多根據受試者反應時間長短來判斷效果的顯著性。統計多變量分析及顯著性考驗等可以協助實驗者對效果作客觀的判斷。目前實驗研究已逐漸走向高科技領域採功能性磁振造影(fMRI；functional magnetic resonance imaging)、腦磁波(MEG； magnetoencephalography)及腦電波(EEG； electroencephalography)等技術來探討知覺效應、及注意力等問題。新技術所產生的分析資料為儀器的訊號，不同於以往反應時間長短；傳統分析方法僅極小部份適用於分析訊號。

(b) 網路圖、貝氏網路系統之建立及貝氏網路推論

認知科學中對行為的預測及診斷可能牽涉上千個變項。變項之間的因果關聯可以透過學習或認知理論以網路圖(graphical models; acyclic directed graph)來表示。網路圖也可視為一個較複雜的徑路圖。貝氏網路推論又稱機率導向的推論(probability-based inference; Mislevy & Gitomer, 1995)也是認知心理學結合人工智慧及統計科技的整合型研究例子。貝氏網路系統建立的重要工作是設計複雜的學習者模式(student model)。模式中應包含學習者成功的達成目標行為所具備的知識及相關策略。系統中設計了許多試題或任務來測量模式中的知識及相關策略；在執行中系統根據學習者的表現決定教學的介入或分配下一個工作。目前貝

氏網路在智慧型學習輔助系統(intelligent tutoring system; Mislevy & Gitomer, 1995)及電腦適性測驗(computerized adaptive testing; Almond & Mislevy, 1998)的建立上已逐漸受到重視。

(c) 認知實驗室的快速成立

目前國內外採高科技的認知實驗室快速成立，並結合醫學研究。國內建立的實驗室包含了中研院的 EEG 及 MEG 實驗室，台大、榮總、三總及長庚的功能性磁振造影(fMRI)實驗室，榮總是目前配備最完整的研究機構。國內參與影像及腦電波分析的統計專業人力，主要集中在中研院的神經影像分析團隊。

八、 人才培育

人才為學術研究所需最重要的資源，沒有優秀的研究人員，縱有充分的研究經費與一流的設備，也無法提升研究的質量。非常值得憂心的是國內統計界已面臨人才斷層的嚴重現象，學生的論證訓練日益不足，優秀的學生願意讀博士班而致力研究者愈來愈少，且出國留學者無論質與量均不如過去。台灣統計界的新陳代謝與研究水準的提升有賴年青人的加入與專注研究。

肆、規劃重點之參考指標

1. 在統計理論與應用上具有重要性或發展潛力：

統計研究之理論眾多，且應用範圍廣泛。以目前的資源，將選擇著重在數個具學術與應用價值，或具發展潛力且值得長期研究之領域。

2. 國內現有人力資源之配合性或國際趨勢之相關性：

國內現有人力資源所屬之研究領域，雖為較具個人興趣之研究，但在原有之架構上較容易獲得實際執行的助力，且現有人員將具有較高之參與及配合意願。而順應國際趨勢之研究，有助國內研究之國際化及提昇國內學術水準。

3. 具提昇統計應用水準之跨領域性：

目前許多其他的研究領域，都必須應用統計模式或資料分析以解決實際問題。而統計本為一應用之科學，跨領域之研究不僅能推廣統計應用，且能累積本土化之研究動機。

4.具研究群之整合性：

研究群的集結經常能加速統計領域及跨領域之進步與發展，且同時較具國際競爭能力。因此已具研究群規模之領域，或較易整合形成研究群之領域將列為規劃重點。

5. 配合國內其他各界發展之需要：

為了維持我國在國際上之競爭能力，促使科技升級及經濟持續發展，如何將統計應用在各領域中已是一重要課題。因此，能配合其他各界發展需求者，將列入規劃項目。

伍、重點研究方向

一、統計方法與理論

I. 迴歸分析理論與方法

未來數年內，下列研究方向應是迴歸分析領域主要的研究課題：

(a) 在大型資料的應用需求日益增加的現在及未來，大型資料下的模式選取相關理論及方法學已成為重要的新興領域之一。例如近來發展的懲罰迴歸（penalized regression）及錯誤發現率（false discovery rate; FDR）方法學，均已逐漸形成了研究的風潮。國內統計界目前在此領域的工作尚不多，但在國際的熱潮帶動下，預期將會有愈來愈多國內學者投入此一領域的研究。

(b) 半參數及其他更具建模彈性的迴歸模式將成為主流迴歸模式。平滑方法、Bayesian 計算等技巧將更多地被應用於回歸分析中。

(c) 高維度資料的迴歸分析將成為新興的研究領域；其中包括高維度資料的降維、視覺化均是相關的研究課題。

II. 非參數及半參數模型與方法

這包含了參數模型檢定、定性平滑函數估計與檢定、斷點或轉折點的估計與檢定、影像處理、半參數模型、泛函資料與長期追蹤資料分析、以非參數模型為基礎的分類法及群聚法。

III. 時間序列與空間統計

目前時空統計的發展相較於純時間序列或純空間統計方法，尚未成熟且仍有很大的發展空間，在應用上時空統計分析的軟體更是極為缺乏。值得研究的方向如下：

(a) 建構更具彈性的時空統計模式

時空資料經常存在複雜的統計結構，如資料在空間中常是非平穩的，時空互變異關係常是不可分離的且不對稱的，當前的時空統計模式對時空資料的配適仍有相當的限度。如何建構一組有足夠彈性、具解釋性 (例如模式參數具有某種物理意義)、且能適當配適資料的時空模型，依然是一項挑戰。

(b) 大量資料的計算問題及時空資料視覺化方法

複雜的模式常伴隨嚴重的計算問題，事實上目前大多時空模式皆無法處理大量資料，這也是時空資料分析軟體不易發展的主要原因。尤其當前許多時空資料 (如大氣及遙測資料) 的量均非常大，如何解決計算問題、如何以圖形等方法呈現複雜的資料、及如何從大量資料中發掘資料中之訊息，仍為未來發展的一大方向。

(c) 模式選擇、評估、及診斷方法

各種時空統計模式如雨後春筍般地被提出，不同的模式著重在不同的時空統計結構，因而對不同的資料有不同的配適能力。在分析時空資料時，如何選擇一個較適當的時空模式，如何評估一個模式配適度的好壞，及如何診斷配適不足之

處，皆有待進一步研究。

IV.實驗設計

以下我們針對最適設計與因子設計的未來發展，提出一些研究方向。

(一) 最適設計

(a) 更多樣的模型選擇

探討最適設計的文獻以往大部份集中於線性模型(linear model)。近年來由於最適化技巧的進步和電腦運算能力的提高，我們可以對更複雜且更接近真實數據的模型做深入的探討，例如近年來在工業統計應用上常出現的非線性模型(non-linear model)、廣義線性模型(generalized linear model)、多反應迴歸模型(multiple response regression model)和反應曲面模型(response surface model)等。在這些模型下，當實驗區域為正方體或某受限區域 (restricted region, 如球體或混合配方實驗(mixture experiment)裡的單型(simplex))時，關於參數估計、預測、校準估計(calibration)以及最適設計之尋找與建構的研究課題，皆具有相當程度的發展空間。

(b) 穩健設計 (Robust design)

針對上述之各類模型，考量模型穩健 (model robustness) 、誤差穩健 (error robustness) 或存在噪音變數 (noise variables)時的各種最適設計問題。

(c) 計算機輔助設計 (Computer aided design)

在複雜的模型假設下，如何利用統計模擬或適當的演算法，來快速且有效率的找到最適設計，亦是一個重要課題。這裡所謂的複雜模型，包括非線性模型、無母數迴歸模型 (nonparametric regression model)、部分線性模型 (partially linear model) 或滿足某些微分方程組的隱藏式模型 (implicit model) 等。

(d) 具有不等變異數之模型

同質變異(homogeneity of variance)雖是實驗設計模型最常見的基本假設之一，但在真實實驗裡，反應變數的變異數會隨著因子的調整而改變的情形，並非不常見。針對這種實驗，如何找出分散因子(dispersion factor)，以及如何建構最適設計，亦是值得研究的方向。

(e) 模型鑑定(model discrimination)

一般而言，最適設計所採用之準則常與模型的參數估計與檢定有關。但是一個實驗除了估計和檢定參數之外，尚有另一個重要目標—選取一個最能夠描述資料的模型。所以如何將模型鑑定和參數估計及檢定的準則相結合，找出能同時兼顧此二目標之設計，亦是未來研究之重要課題。

(f) 序貫與貝氏設計 (sequential and Bayesian design)

序貫與貝氏方法為國內實驗設計方面尚無人從事的重要方向。如前述非線性模型與廣義線性模型下的最適設計通常與未知參數有關，此時序貫與貝氏方法為重要的工具。貝氏方法在實驗設計有廣泛的應用。

(二) 因子設計(factorial design)

(a) 多層誤差結構下的實驗設計(multi-stratum factorial design)

隨機化是實驗設計的三大基本原則之一，但在現實的實驗情況下，有時完全隨機(complete randomization)未必是可行的，或者它不再是最好的策略。針對此類問題，實驗設計文獻中提出了許多不同隨機限制下的實驗設計型態，比如區集設計(block designs)、多層區集設計(split-plot designs)、巢狀設計(nested designs)等。由於不同的隨機限制，這類實驗設計具有多層的誤差結構。以往這類問題的文獻多著重於實驗次數較多且滿足某些平衡條件下的實驗的分析與設計，一般多採用標準的變異數分析，或將某些變數的效應視為隨機效應(random effects)，在適當的混合效應模型下以 Restricted MLE 的方法來估計各個層級的誤差。然而如何建立一有系統的準則來探討這類問題的實驗設計，到目前為止尚無一致的結論。尤其是在實驗次數遠小於所有因子可能的組合時，如何定義一個有效的最小偏差或其他準則來選取多層誤差結構下的因子設計，應是未來重要的研究方向之一。

(b) 複雜型反應變數的因子設計(Designing experiments for complex responses)

以往因子設計所探討的問題，大多集中於連續型反應變數所適用的線性模型(linear model)下的因子設計問題。然而有時反應變數具有其他特性，例如當反應變數為離散型變數、據有空間特性，或為圖像變數等複雜的變數時，所適用的分析模型多為廣義線性模型或廣義線性混合模型 (generalized linear mixed model)。未來因子設計的研究重點之一為探討複雜型反應變數所適用的廣義模型下的因子設計問題。

(c) 因子篩選(factor screening)或小樣本多因子的設計及分析

在以往的實驗設計裡，因子篩選實驗一向被視為主效應設計(main effect design)，而所謂的重要因子即被定義成具有顯著主效應的因子。近幾年來，因為在因子篩選實驗分析上的進步，重要因子不再與顯著主效應劃上等號。在這個新觀點下，如何定義重要因子與設計實驗，便成為一個重要的問題。除此之外，因子篩選更進一步與投影設計(projected design)的概念相結合，產生了使用一個設計來同時達成因子篩選與反應區面探索兩個目標的實驗方法。而觀察這方面的近期研究，亦可發現這個課題正逐漸與模型選取(model selection)的分析方法產生關連。隨著小樣本多因子的實驗越來越多，因子篩選設計與分析的重要性也日漸增加。由於貝氏資料分析在應用上越來越受重視，近年來也不斷有推陳出新的貝氏方法被應用在因子篩選的分析上，因此探討貝氏資料分析下的因子篩選實驗設計也將會是未來的研究重點之一。

(d) 不正規因子設計之建構

隨著因子設計的研究越來越趨成熟，因子設計的目標也由正規設計逐漸擴展到不正規設計上。近年來因為電腦運算能力的提昇，已有一些研究試圖藉由計算機對給定的批量(run size)窮舉出所有的直交表，但因運算量過大，只能

完成一些批量較小的設計，而藉由理論發展來建構此完整列表，則正在起步的階段。因為正規設計的理論架構(如群論或有限投影幾何學)，無法直接套用到不正規設計上，使得關於不正規設計的完整列表與正規設計的完整列表之研究，其難度差異有如天壤之別。另一方面，要有完整的列表，就必須能有效地判定兩個設計是否同構(isomorphic)。目前檢測因子設計是否同構的方法，除了有由編碼學或 space-filling 的角度提出的一些方法，另亦有藉 J-characteristics 及指標函數(indicator function)區分非同構設計的檢測法。可惜這些方法有些雖可以完全區分非同構的設計，但卻需要極大的計算量;有些雖可節省計算量，但無法完全區分非同構的設計，或是只能適用於滿足某些嚴格條件限制下的設計。因此如何發展出能適用各種設計、計算量小、且能完全區分非同構設計的檢測方法應是未來的發展方向一。

(e) 指標函數(indicator function)

指標函數是一個近期出現的理論架構。其使用多項式來定義因子設計，而該多項式的係數則攜帶著因子效應之間彼此別名(aliasing)程度的訊息。指標函數可定義在正規及不正規設計上，而在正規設計中，指標函數與定義對比子群(defining contrast subgroup)具有相同的結構。因為這個性質，許多原本在正規設計上存在的性質可藉由指標函數推廣到不正規設計上。目前指標函數的研究正在起步的階段，在未來應有更多的理論與應用發展。

因為實驗設計的理論發展，常隨著不同型態實驗的出現，而產生新的研究課題及方向，故我們在以下討論兩種在未來值得注意的新型態實驗及其研究課題。

(一)生物技術實驗設計(designs for biotechnology)

隨著生物上各種與基因相關的研究課題越來越熱門,各類與生物技術相關的實驗也如雨後春筍般地層出不窮。有鑑於生物科技將會是未來最熱門的研究學門之一，我們在此提出關於生物技術實驗的一些研究方向。

(a) 以多標的微陣列(multiplex microarray)為平台之生物晶片診斷試劑的確效實驗(validation experiment)設計

傳統的診斷試劑是根據單一分析物(analyte)來判定疾病存在與否。但是以多標的微陣列為平台之生物晶片診斷試劑的分析物為基因,所以分析物的數目可多達數萬個，這使得其確效實驗的設計更困難，也更具挑戰性，我們預期這將是未來重要的研究領域。

(b) 單核甘酸多樣性的研究設計

單核甘酸多樣性(single nucleotide polymorphisms, SNPs)是人類或生物最常見的基因變異。這些 SNP 單獨基因突變雖是低透視低風險(效應)，但整群 SNP 的累積效應則可能增加癌症發生的風險或改變作物的產量。以整個基因體進行 SNP 篩選之實驗設計，將會是未來重要的研究議題。

(c) 中醫藥臨床試驗之設計

中醫藥臨床試驗的統計方法必須將中草藥個人化變動劑量、複方與多重療效、以及安全性指標的特性納入考慮。因此最適中醫藥臨床試驗設計極具挑戰

性，同時也是迫切需要發展的議題。

(d) 前瞻性的目標臨床試驗之設計

i. 傳統的臨床試驗是以病人族群(patient population)為推論目標。但目前因藥物基因體學及微陣列的發展，疾病治療已可用分子為目標進行個人化醫療處置。所以發展目標臨床試驗之設計亦是試驗設計未來最重要的研究領域之一。

ii. 當具有分子目標的病人盛行率很低時，目標臨床試驗會因收納病人不易而使其可行性大幅降低。此時該如何在族群與個人化之間取捨，亦是未來統計的挑戰。

iii. 生物晶片在鑑別分子目標上的準確度，會對目標臨床試驗療效評估造成影響。所以生物晶片的臨床用途試驗與藥物臨床效用試驗（clinical effectiveness trials）之合併設計也是將來發展重點。

(二) 電腦模擬實驗(experiments for computer simulation)

在近二十年來，隨著電腦科技的快速發展，一種在電腦虛擬的世界裡執行的新型態實驗，正方興未艾。許多以往礙於經費、實驗技術或現實世界的種種限制，而無法在真實世界裡被執行或觀測的現象，在藉由電腦程式所架構出來的虛擬世界裡，可以隨意地修改輸入變數以觀察其所造成的變化。

若比較電腦模擬實驗與傳統的自然實驗(physical experiment)，兩者之間有幾個重要的差異：

(i) 前者重覆執行同一組水平組合時，會得到相同的結果。亦即沒有實驗誤差的存在。

(ii) 前者實驗中沒有實驗單位(experimental unit)的概念。故無需考慮實驗單位是否同質(homogeneous)的問題。

(iii)在前者實驗中，因子水平的調整與設定非常方便且精確。

因為(i)和(ii)，傳統自然實驗裡非常重要的三大原則－replication、randomization 及 blocking－所處理的問題，在電腦模擬實驗裡皆不用考慮。而因為(iii)，電腦模擬實驗的水平數，亦可由傳統自然實驗裡常用的兩至三水平，大幅提昇至非常高的水平數。因為這些差異，電腦模擬實驗為實驗設計這個學門帶來了新的挑戰及研究課題。以下是一些電腦模擬實驗所帶來的新研究方向：

(a) 統計建模(statistical modeling): 比如利用 random field model 來處理具有(i)性質的數據。

(b) 極大量的因子數(large number of factors): 因為實驗是藉由電腦程式來執行，且電腦程式常有許多的變數，這使得電腦模擬實驗經常必須處理較多的因子。

(c) 非常多的水平數(large number of levels): 比如 Latin-hypercube 設計的出現，使得水平數大幅增加。

(d) 新的設計準則(design criteria)及最適設計: 比如要求設計具有 space-filling 的性質，或是針對不同的實驗目的，如函數積分或是預測，來訂定設計準則，

並獲得其最佳設計。

証實與確認(confirmation and validation): 電腦模擬實驗的結果未必會與自然實驗一致，若同時有自然實驗及電腦模擬實驗的結果，便可以利用前者來驗證電腦模擬實驗的正確性。

二、應用機率

I. 演算法隨機分析

拜計算機計算能力之大幅提升與網際網路所帶來的便利性之賜，大量資料的處理在現今各科學領域中愈形重要，特別是在一些熱門學科，如生物資訊、資料採礦、複雜網路、機器學習等。如何設計更有效的演算法處理更大量的資料，及其伴隨而來的演算法效率漸近分析問題，皆是目前的重要研究方向。從小樣本到大樣本，從精確分析到漸近分析，從 worst-case analysis(可能是 NP-complete)到 average-case analysis，約略指出未來大致的主要研究方向。

三、工程及工業統計

I. 線上品質管制與監控未來研究重點:

(a) 機台缺失偵測與分類 (Fault Detection and Classification)

由於監控設備 (sensor equipment) 與資訊科技的急速發展，機台設備狀態資料得以即時擷取並儲存，而這些資料可用於即時的機台缺失偵測。這些即時的機台資料大多互相關聯且具自我相關性，因此以往的研究主要以多變量 SPC 與時間序列分析來發展偵測工具。除了偵測機台缺失外，最近的研究趨勢更包括了：

(i) 機台缺失分類 (Fault Classification)

缺失分類將有助於缺失的診斷 (fault diagnosis) 及快速的機台改善 或維修。分類的方法主要是利用特有的統計形態 (statistical pattern) 將缺失作初步的分類，並與物理上造成缺失的原因聯結，而得到最後的缺失分類。主要具挑戰性之研究課題為如何有效率的發現每一類缺失的統計形態，並與機台失效模式與原因聯結在一起。

(ii) 虛擬量測 (Virtual Metrology)

機台製程的狀態如果可以即時且精確的獲得，則品質特性便可從這些即時狀態資料之迴歸模型獲得; 此由機台狀態資料預測所得到的品質特性值就被稱為虛擬量測值。這些預測量測值雖無法百分之百取代現有的量測，卻可作為決定量測取樣點與量測取樣頻率的根據。主要的研究挑戰在於如何在動態的製程變動下，建構夠穩健或是具調適性的品質特性迴歸模型。

(iii) 機台健康指標 (Equipment Health Index)

機台的狀態如果能夠以單一指標來顯示，不但有助於製程工程師對製程的掌握，更可被設備工程師用來作為機台維修保養的根據，整合即時機台狀態資料來描述機台健康狀況，便成為最新的研究趨勢。主要的研究挑戰在於如何選取最具代表性的統計量來整合各種機台狀態資料，此統計量不但需有效反應真正機台的

健康狀況，更必須在不同製程配方下維持其一致性。

(b) 統計製程管制 (Statistical Process Control)

國內外目前在製程設備監控方面，較熱門的研究主題有:

(i) 多變量製程之管制

(ii) 量測誤差 (measurement errors) 對管制圖之影響及其效應

(iii) 少量多樣製程之管制問題

(iv) 自我相關 (auto-correlated)資料之製程管制問題

(v) 改變點 (change-point) 模型之研究

(vi) SPC 及 EPC 之整合

(vii) 無母數製程管制

(c) 批次控制器 (Run-to-Run Control)

針對製程 I-O 模型為 MIMO 系統的批次回饋控制，當製程有漂移或老化現象且製程干擾為 vector ARIMA (p, d, q) 時間序列下，李水彬 (2001) 提出 double MEWMA 回饋控制器，唯此控制器之最適折扣矩陣之解析解並不易獲得。除此之外，欲設計穩定的 MIMO 批次回饋控制，在線外 (off-line) 階段，如何決定適當樣本數以確保製程之穩定收斂機率能達到給定水準，亦是值得深入研究之課題。

過去的批次控制器主要是以單一製程產出後的量測資料作為回饋或前 饋調控製程配方的根據，然而在複雜的高科技產業中，一個產品往往需要經過上百道製程方能完成。以 12 吋晶圓廠為例，一片晶圓需要經過 200 至 400 個精密奈米製程步驟，單一步驟的監控通常無法確保最後晶圓的良率。因此，如何透過最後的良率來整合監督整廠的批次控制器，以達到最終高良率的目標，便成為最新的研究趨勢。

(d) 剖面資料製程監控 (Profile Monitoring)

製程監控在品管上是很重要的一環。近年來統計品質管制上的研究範疇愈來愈廣，由單變量管制問題推廣至多變量管制問題及由獨立性資料推廣至自我相關 (auto-correlated) 資料等，而最近一個新的研究主題為如何監控具有剖面 (profile) 資料之生產製程。

一般而言，一個產品之品質特性，通常為單變量或多變量隨機變數。然而對某些製程而言，產品之品質特性是由一些反應變數和解釋變數之間的關係來界定，因此品質特性乃以一條曲線 (或曲面) 或其離散化 (discretized) 後之資料型式來呈現，稱之為剖面資料。例如苯丙氨酸甲脂 (一種低熱量人工甜味劑) 在每公升水中之溶解量與水溫度的關係，可用一條曲線來描述 (Kang & Albin, 2000)，此種形式之資料又稱為曲線資料 (curve data)。函數資料分析 (functional data analysis) 為處理曲線資料的一種分析方法，目前為無母數迴歸研究領域內相當受到重視的研究課題。相關文獻可參考 Ramsey & Silverman (2002, 2005)。

目前文獻上著墨較多的是線性剖面資料之監控，最近亦開始探討非線性資料的情形。大部分的研究假設所有的剖面資料都是來自同一個函數加上隨機誤差的

模型，其作法為先用參數迴歸模型來配適剖面資料，再針對參數估計量來建構管制圖。然而，有許多製程其產品的剖面資料為一群並不相同但具有類似形式的函數，上述固定效應 (fixed-effect) 參數模型在此狀況下並不合適。這種函數的差異性可用隨機效應參數模型來描述或是用函數資料分析的方法來處理。目前國內外都已有研究開始朝這個方向進行。

如何有效地監控剖面資料的製程是一個兼具學術性與實用性的問題，Woodall *et al.* (2004) 提供了一個不錯的文獻回顧，也提出未來研究的展望與方向。值得一提的是，他們視此領域為 SPC 中最具前途的研究領域。而如何將函數資料分析應用於 SPC 製程監控上，目前仍在起步中。未來亦可推廣至高維度剖面資料監控問題之研究。

II. 可靠度分析未來研究重點:

(a) 高可靠度統計推論 (不可維修產品)

在加速壽命模型部份，未來可考慮的研究重點如下：

(i) 加速壽命模型的無母數統計推論問題

(ii) 較一般化的壽命-應力統計模型如 GG3-Generalized Eyring 模型之相關統計推論問題

(iii)較複雜的逐步應力 (step-stress) 或連續應力 (progressive-stress) 加速壽命模型之統計推論問題

(iv) 加速壽命模型挑選 (model selection) 問題

而在衰變及加速衰變模型部份，未來可考慮：

(i) 如何將混合效應模型及隨機擴散過程等模型之優點加以整合，應是較具挑戰的工作。此類型的統計推論工作大多十分複雜，除了模型中的參數估計量沒有簡便的解析解之外，通常其壽命分配亦無簡便的公式可供使用，因此如何提供較精確的參數近似估計方法及產品平均壽命之近似估計公式將是一大挑戰。

(ii) 除了衰變模型外，較複雜的加速衰變及逐步應力加速衰變等模型之相關統計推論問題亦是值得深入研究的課題。

(iii) 加速衰變模型挑選問題。

(b) 奈米元件可靠度分析

民國九十四年，由國科會、中研院、工研院、教育部等機構共同推動跨部會之「奈米國家型科技計畫」。其計畫之總目標為整合我國奈米科技相關研究上、中、下游之研究人力與技術資源，以加速奈米科技之發展與應用。除此之外，國內學術界及產業界亦紛紛成立奈米研究中心進行奈米科技相關產品之研發工作。

有關奈米元件可靠度方面，瑞士 EMPA 研究機構早在 2001 年即開始以 reliability of nano structured materials and devices 為主題進行研究，並在 2003 年開始執行 reliability and degradation physics of ultra-thin dielectrics (Nanoxide) 之研究計畫。美國國家標準局 (NIST) 亦在 2004 年舉辦 Workshop on Reliability Issues in Nano-materials。可靠度重要期刊 *IEEE Transactions on Reliability* 也將在 2007 年出刊奈米可靠度分析之專刊。以上皆顯示奈米元件可靠度之重要性。

奈米科技近年來漸趨成熟，奈米材料、產品和設備可靠度的統計方法逐漸受到重視。配合國家重點計畫與奈米產業發展，有關奈米材料、產品和設備的前瞻性研究主題如下：

(i) 壽命分佈的模型建立與確認
(ii) 可靠度測試 (尤其在加速壽命) 之設計
(iii) 老化、衰變及失效率之建模
(iv) 可靠度標準之建立
(v) 可靠度預測和保證

(c) 可維修模型統計推論 (不可維修產品)

i. 可靠度專家 Nelson 和 Lawless 於 2005 年在美國統計年會，針對可維修模型之相關統計推論問題進行兩場演講，其演講內容可做為此領域未來研究之參考。傳統上，可維修模型大多著重在觀察失效事件上進行分析，Nelson 則認為需將成本及其相關資料一同列入考慮。另外，Lawless 則強調重現數據之迴歸分析，尤其需重視穩健方法在重現率和平均累積函數上之研究。
ii. 有關具有截尾 (truncated) 重現數據之參數及無母數的研究，可參閱 Robinson & Chukova (2004)。
iii. 有關 window observation data 之研究，可參閱 Zou, Meeker, & Wu (2006)。

(d) 軟體可靠度 (可維修產品)

IEEE 軟體可靠度學會曾於 1988 年定義「軟體可靠度管理」是強調如何避免發生人為錯誤、有效偵測及移除程式中失敗的原因 (fault detection and removal)，並且在有限的測試時間下，使用適當測試方法以達到軟體可靠度的最大化。因此，Pan (1999) 建議軟體可靠度未來研究方向大致應分為：

i. 軟體可靠度之預測、估計及建模。根據觀察和累計的失敗或錯誤之訊息所做的統計推論以期選擇最適當的預測模型。
ii. 有關軟體測試之最適發行時間 (optimal release time) 之相關統計推論問題。
iii. 有關如何利用失敗或錯誤之訊息來預測及降低程式失敗率，並提昇軟體之品質。

III. 實驗設計與品質改善未來研究重點:

(a) 不規則實驗區間 (Irregular Experimental Region)

以往的實驗設計皆假設實驗區間為一正方體 (或長方體)，我們稱其為正規實驗區間。但是當因子之間彼此相關 (related) 時，一個合適的實驗區間常是不規則的幾何體。比如在汽車工業的焊接實驗中，常可見到電流強度與焊接時間兩個因子，當電流強度調高 (或調低) 時，焊接時間的實驗區間也必須隨之調低 (或調高)，才能產生好的焊接點，此時一個合適的實驗區間便已不再是正方體。對於不規則的實驗區間，我們可使用滑動水平 (sliding level) 的實驗方法來分布實驗點，但在文獻上關於這個課題的討論並不多。我們相信針對不規則實驗區間的設計、建模、以及資料分析問題，未來將會有許多可發展的空間。

(b) 貝氏方法 (Bayesian Method)

因為近十幾年來貝氏資料分析與統計計算上的各種長足進步，使得貝氏方法在應用上越來越受重視。在實驗數據的分析上，近年來也不斷有推陳出新的貝氏方法被應用在因子篩選分析、模型選取 (model selection) 以及超飽和設計 (supersaturated designs) 的分析上。若仔細觀察這些貝氏分析法，可以發現它們皆是在處理實驗數據自由度不足的問題。所謂的自由度不足，是指實驗模型裡的效應個數遠多於自由度，導致所有的效應無法被同時估計，遑論要檢定其是否顯著。尤有甚者，自由度不足的情況還經常伴隨著因實驗無重覆測試而使得連模型的變異數也無法估計的情形。這讓迴歸分析裡的各種變數選取，諸如 C_p, AIC 之類的分析方法，也無法使用。由目前這方面的研究看來，雖然有 frequentist approach 可以解決這問題 (如 Cheng & Wu, 2001)，但是藉由貝氏模型選取 (Bayesian model selection) 來分析這一類的數據，應可提供更靈活更恰當的解答。我們相信隨著工業上小樣本多因子的實驗越來越多，這方面的研究在未來將會更受重視。

(c) 多階段程序之實驗 (Experiments for Multi-stage Process)

在以往的實驗設計裡，我們常將整個實驗過程描述成一個程序。在這個程序一開始時，我們改變因子的輸入值，而在這個程序結束後，我們將會得到反應變數的輸出值。可是隨著工業製程上的實驗越來越複雜 (比如晶圓的製造常需經過上百道步驟)，單一程序的描述法已不能滿足這類型實驗的需求，而必須將整個實驗過程視為多個不同階段的程序的組合。在多階段程序的實驗裡，每一個程序都有各自的因子可供調整，也可能有各自不同的反應變數輸出。而上一個程序的反應變數，在接下來的程序中，必須被當成輸入變數來分析。針對多階段程序的實驗法、設計選取、以及資料分析方法的研究，目前在文獻上寥寥可數，但是我們相信其在未來應是一個重要的研究領域。

(d) 未來實驗設計的新領域

我們注意到近來統計的實驗設計與分析開始被應用在幾個截然不同的領域上。這些領域包括市場調查 (marketing)、藥物開發以及奈米科技 (nano-technology)。正如我們在前面所提到的，新領域的實驗常會給實驗設計帶來新的研究課題，我們希望在未來能多注意這些新興領域的實驗設計研究進展。

參考文獻

1. Balakrishnan, N. (2006), "Progressive censoring methodology: An appraisal (with discussion)," *Test*, To appear.
2. Barlow, R.E. and Proschan, F. (1981), *Statistical theory of reliability and life testing: Probability models*. Sliver Spring, MD
3. Boulanger, M. and Escobar, L. A. (1994), "Experimental design for a class of accelerated degradation tests," *Technometrics*, 36, 260-272.
4. Box, G. E. P. and Kramer, T. (1992), "Statistical Process Monitoring and Feedback Adjustment – A Discussion," *Technometrics*, 34, 251-285.
5. Cheng, S.-W. and Wu, C.F.J. (2001), "Factor Screening and Response Surface

Exploration (with discussion)," *Statistica Sinica*, 11, pp. 553-604.
6. EMPA, http://www.empa.ch/plugin/template/empa/681/*/---/l=2.
7. Ingolfsson, A. and Sachs, E. (1993), "Stability and Sensitivity of an EWMA Controller," *Journal of Quality Technology*, 25, 271-287.
8. Kang, L. and Albin, S. L. (2000), "On-line Monitoring When the Process Yields a Linera Profile," *Journal of Quality Technology*, 32, 418-426.
9. Kotz, S. and Johnson, N. L. (1993), *Process Capability Indices*, Chapman and Hall, London.
10. Kotz, S. and Lovelace, C. R. (1998), *Process Capability Indices in Theory and Practice*, Arnold, London.
11. Lawless, J. and Crowder, M. (2004), "Covariates and Random Effects in a Gamma Process Model with Application to Degradation and Failure," *Lifetime Data Analysis*, 10, 213-227.
12. Lawless, J. (1983), "Statistical Methods in Reliability," *Technometrics*, 25, 305-335.
13. Lawless, J. F. and Nadeau, C. (1995), "Some Simple Robust Methods for the Analysis of Recurrent Events," *Technometrics*, 37, 158-168.
14. Lawless, J. (2005), "Regression Analysis of Recurrent Events Data," *Joint Statistical Meetings*, Minneapolis, Minnesota, USA.
15. Lu, C. J. and Meeker, W. Q. (1993), "Using degradation measures to estimate a time-to-failure distribution," *Technometerics*, 35, 161-174.
16. Lisnianski A. and Levitin G. (2003), *Multi-state system reliability*. World Scientific.
17. Lindqvist, B. (2003), "Heterogeneity Modelling for Recurrent Events," *International Conference on Reliability and Survival Analysis* (ICRSA), 2003.
18. Myers R. H. and Montgomery, D.C. (1995), *Response Surface Methodology*, New York: John Wiley & Sons.
19. Nelson, W. (1988), "Graphical Analysis of System Repair Data," *Journal of Quality Technology*, 20, 24-35.
20. Nelson, W. (1990), *Accelerated Testing: Statistical Models, Test Plans, and Data Analysis*, New York: John Wiley & Sons.
21. Nelson, W. (1995), "Confidence Limits for Recurrence Data – Applied to Cost or Number of Product Repairs," *Technometrics*, 37, 147-157.
22. Nelson, W. (2003), *Recurrent Events Data Analysis for Product Repairs, Disease Recurrences, and Other Applications*, Philadelphia: ASA-SIAM
23. Nelson, W. (2005), "Graphical Analysis of Recurrent Events Data," *Joint Statistical Meetings*, Minneapolis, Minnesota, USA.
24. NIST, http://www.nsti.org/Nanotech2004/.

25. MacGregor, J. F. (1988), "On-line Statistical Process Control," *Chemical Engineering Progress*, 84, 21-31.
26. Montgomery, D. C. (2005), *Introduction to Statistical Quality Control*, 5th Edition, New York: John Wiley & Sons.
27. Meeker, W. Q. and Escobar, L. A. (1998), *Statistical methods for reliability data*. New York: John Wiley & Sons.
28. Pan, J. (1999), Software Reliability. Carnegie Mellon University. http://www.ece.cmu.edu/~koopman/des_s99/sw_reliability/
29. Pearn, W. L. and Kotz, S. (2006), *Encyclopedia and Handbook for Process Capability Indices*, World Scientific, Singapore.
30. Ramsey, J. O. and Silverman, B. W. (2002), *Applied Functional Data Analysis*, New York, Springer.
31. Ramsey, J. O. and Silverman, B. W. (2005), *Functional Data Analysis*. 2nd Edition. New York, Springer.
32. Rausand and Hoyland A. (2004), *System reliability theory models, Statistical methods, and applications*. New York: John Wiley & Sons.
33. Robinson, J. and Chukova, S. (2004), "Estimating Mean Cumulative Functions from Truncated Automotive Warranty Data", *Communications of the Fourth International Conference on Mathematical Methods in Reliability, Methodology and Practice* (in CD, 4 pages), Santa Fe, New Mexico, USA.
34. Woodall, W. H., Spitzner, W. H., Montgomery, D. C. and Gupta, S. (2004), "Using Control Charts to Monitor Process and Product Quality Profiles," *Journal of Quality Technology*, 36, 309-320.
35. Wu, C. F. J. and Hamada, M. (2000), *Experiments: Planning, Analysis, and Parameter Design Optimization*, New York: John Wiley & Sons.
36. Yu, H. F. and Tseng, S. T. (1999), "Designing a degradation experiment," *Naval Research Logistics*, 46, 689-706.
37. Zuo, J., Meeker, W. Q. and Wu, H. (2006), "Analysis of Window-Observation Recurrence Data", Preprint of the Department of Statistics, Iowa State University, USA.
38. 李水彬, (2001), 批次控制之設計與其穩定性分析, 國立清華大學博士論文.
39. 奈米國家型科技計畫, http://nano-taiwan.sinica.edu.tw/newsbig5.asp.

四、生物統計與生物資訊

I. 臨床試驗統計方法

藥物基因體學（Pharmacogenomics）及微陣列（Microarray）為過去十年最重要的基礎科學發展及科技突破。它們在疾病的診斷與治療上均可達到以分子目

標（Molecular Target）進行個人化醫療處置（Individualized Medicine）。所以臨床試驗統計方法未來規劃重點將以發展以分子層次之前瞻性臨床試驗設計與分析方法為主：

(a) 評估多標的微陣列為平台之生物晶片診斷試劑的統計方法

個人化醫療的先決條件必須鑑別病人是否具有藥物治療的分子目標。所以微陣列基因晶片與蛋白晶片等診斷試劑必須達到適當品質，其準確性（Accuracy）、精確性（Precision）及再現性（Reproducibility）均是品質控管的重要目標。而生物晶片是多標的（Multiplex），所以應發展出評估鑑別分子標的生物晶片品質與其相對臨床用途試驗（Clinical Utility）Trial 設計與分析的統計方法。

(b) 前瞻性的目標臨床試驗

傳統的臨床試驗是以病人族群（Patient Population）為推論目標。但目前因藥物基因體學及微陣列的發展，使疾病治療可以分子為目標進行個人化醫療處置。所以發展目標臨床試驗之設計與分析統計方法將是未來臨床試驗最重要的研究領域。另外，若具有分子目標的病人盛行率很低，目標臨床試驗因收納病人不易且可行性低，所以如何在族群及個人化之間的取捨亦是未來統計的挑戰。而生物晶片之鑑別分子目標的準確度對目標臨床試驗療效評估亦發生影響。所以生物晶片的臨床用途試驗與藥物臨床效用試驗（Clinical Effectiveness Trials）試驗合併設計與分析的統計方法亦是將來發展重點。

(c) 臨床試驗的調整設計（Adaptive Design）統計方法

逐次設計，期中分析與樣本數再估算已成為目前臨床試驗最常用的調整設計統計方法。但為了面對臨床試驗執行時的各種不確定性如改變病人族群，修正臨床假說或調整藥物劑量，發展出相對應的統計方法亦是十分重要而極具挑戰性的研究題目。

(d) 生物製劑（Biological Products）的藥劑生物對等性（Bioequivalence）評估的統計方法

生物製劑為由生物方法所製造的大分子藥物。生物製劑大部份為蛋白質，其特性與傳統由化學方法製造之小分子藥物不同。所以傳統生物對等性試驗並不完全適用評估生物製劑的學名藥（又稱 Biosimilar Drug Products）。所以評估生物製劑學名藥的試驗設計與分析統計方法亦是臨床試驗的重要研究領域。

(e) 銜接性試驗（Bridging Studies）的統計方法

目前銜接性試驗統計方法仍然為試驗間比較（Between Study Comparison），其療效評估仍可能有偏差。而銜接性試驗不限於族群間的銜接，亦可推廣至成人族群到兒童族群之銜接或替代指標（Surrogate Endpoints）與真實療效指標（True Efficacy Endpoints）的銜接。所以銜接性試驗的統計方法不但是未來發展重點，而且可以降低臨床試驗的成本。

(f) 健康相關生活品質的研究

統計方法的相關研究包括（i）由於健康相關生活品質是一個多維度的概念，如何處理生活品質量測的多重指標是一個值得研究的問題。如何利用題組反應理

論如 Rasch 等模式來將多重題組的回答抽出有用的量測尺度也仍是一個熱門的研究問題。(ii) 同時分析存活時間與重複量測生活品質的聯合模型有它的重要性，尤其在研究存活是主要結果的嚴重疾病的臨床試驗上。目前已經有很多相關文獻，發展出的統計模型在技術上也很複雜了。(iii) 評估健康相關生活品質是相對重要的慢性病治療或醫療資源分配時，生活品質調整存活的估計就顯得重要。國內目前有研究團隊發展一系列生活品質調整存活的估計方法，正用來估計台灣各種主要疾病的健康餘命及包括臨床治療比較與中醫藥療效的評估等其他應用。這方面的理論與應用值得進一步研究。

(g) 中醫藥臨床試驗統計方法

中醫藥臨床試驗的統計方法必須考慮中草藥個人化變動劑量複方與多重療效與安全性指標的特性。這方面的統計方法是極具挑戰性也是迫切需要發展的部份。

抗生命威脅疾病(如癌症)藥物的研發一直是製藥界研究的重點。近年來，許多藥廠的注意力已漸漸轉移至中醫藥的研發。針對西藥的研發，統計方法早已扮演了極重要角色。然而中醫藥在效能性與安全性的評估方式，明顯與西藥的評估方式不盡相同。因此針對西藥研發所發展的統計方法，也不一定適合用來解決中醫藥的研發所產生的問題。尤其是中醫藥臨床試驗的統計方法必須考慮中草藥個人化變動劑量複方與多重療效與安全性指標的特性。這方面的統計方法是極具挑戰性也是迫切需要發展的部份。關於中醫藥研發未來規劃重點簡述如下：

(i) 中草藥的品質管制：

中草藥並不像西藥，大多僅包含單一活動性成分，大多數中藥是由許多不同成份組合而成。一般而言，這些成份的藥理學活動性、成份間的交互作用、以及各成份間所需的相對比例，通常是未知的。如何維持中藥複方成份的品質之一致性或重製性，便是一大課題。因此需藉由發展有效的統計品質控管程序，以確保中草藥的種源或是複方成品的一致性或重製性。另一方面，並可藉由中藥的基因體計畫，協助藥材鑑定。

(ii) 中草藥的穩定性分析：

藥物產品上市前，相關藥政法規均會要求藥品需標明使用的安全有效期限。為了符合這些規範，便需藉由穩定研究的執行，來描繪藥物的退化。目前，針對單一成份藥品，已經有現成的統計方法可以協助決定藥品的有效期限。然而對於中草藥的複方成份，便需依賴新的統計方法來協助估計藥物的有效期限。

(iii) 中醫診斷方式的科學化：

中醫的診斷模式主要包括望、聞、問、切等四個項目。然後中醫師再根據八綱辨證、五行、五臟六腑、以及經絡傳遞的訊息，來卻認診斷。這樣的診斷模式，會因中醫師的不同,而有所差異。再者，如何利用這樣的診斷，來判斷中醫治療的有效性？因此，便需發展統計方法，以建立中西醫的療效評估的相關性。此外，尚需發展統計方法，來判斷中醫診斷工具的信度與校度。其他如脈診儀、舌診儀均可由統計方法來幫忙建立。

(h) 臨床試驗的實務研究

繼續參予國內各項臨床試驗體制建置、運作、執行與評估等工作。與產業界合作培育臨床試驗所須的統計人才。積極加入國際性臨床試驗研究學會或組織並擴大國際交流。另外，目前國內已有衛生署優良藥品臨床試驗規範（Good Clinical Practice, GCP），以規範臨床試驗臨床與資料處實務之執行。在未來，亦應制定優良藥品臨床試驗統計規範，以規範臨床試驗統計設計與分析準則，俾使所執行之臨床試驗的統計設計與分析能正確評估藥物作用與試驗結果。

II. 遺傳統計方法

國內近年從事統計遺傳研究人員較十年前增加許多，但因國內遺傳學研究相對國外仍難進入尖端領先，統計領域對遺傳研究方向重點的選擇，必須依據國際間在這方面的發展而定，以下是國際間遺傳統計近年與未來的重點研究方向：

(a) 全基因體基因有效搜尋策略的發展

複雜疾病成因可能受多基因影響，由於基因可能散布於不同染色體上，故如何透過連鎖分析原理，能在龐大基因庫中設計出節省時間、經濟成本的搜尋策略，成為重要課題。這方面統計方法發展包括：分段、分區的計算法則設計，粗略連鎖分析方法和精細連鎖分析方法的建立與搭配，及多個基因的聯合搜尋、共同影響、回饋分析。

全基因體的關聯分析亦是目前重要的課題，受試者的全基因體同時有幾十萬個 SNP 資料，如何找出可靠有效的搜尋方法將是重要且具挑戰的課題；如多點連鎖不平衡之基因定位分析，可考慮不同設計下的分析方法，如病例-對照設計或其他設計。

(b) 多標識基因聯合分析的混淆基因型判定方法

當利用多個標識基因進行連鎖分析，由聯合基因型(或外表型)推論單套基因型的過程，產生統計方法學上所謂的訊息缺失研究問題，如何在統計缺失訊息研究基礎上，發展出適合遺傳研究的大量訊息缺失問題的統計方法，值得期待。

利用多個標識基因進行連鎖分析，亦需找出有效區塊，再予以推論單套基因型，由於家族資料的收集通常是不完整的，在這些情況下如何建立有效可靠的方法，仍有相當的改進空間。其中包括染色體減數分裂時交錯點分布的估計及其對後續分析的影響都值得注意。

(c) 多標識基因聯合分析基因型訊息缺失、定型錯誤之統計分析方法

多標識基因聯合分析過程，部分基因型因實驗室各種人為操作因素產生缺失，或定型錯誤時，發展新的統計校正方法及合併錯誤訊息的連鎖分析方法，有待更多研究。

(d) 有效的多標識基因的統計方法發展

現有的多標識基因分析方法，仍為有限數目的多標識基因。未來對發展能處理更大量的多標識基因的連鎖分析方法，及相關延伸統計分析障礙，仍有待突破。

(e) 可處理複雜家庭資料結構的數量性狀概似函數分析方法

現有的以概似函數為基礎的數量性狀變異數成分分析方法，受限於統計計算複雜性，仍僅能處理有限的家庭結構。發展新穎的概似函數計算方法，能同時處理多基因變異成分、環境成分，為有挑戰性的課題。

(f) 可處理複雜家庭資料結構的數量性狀無母數統計分析方法

概似函數分析方法受限於常態假設窄度，常難於實際分析時得到正確推論，

發展無母數分析方法能處理數量性狀的複雜性，是必要且有難度的研究。當數量性狀為時間變項，如存活或發病年齡時，如何結合存活分析於遺傳統計，亦是一重要研究方向。

(g) 合併多基因、環境因素，基因環境交互作用的數量性狀分析模式發展

現有分析數量性狀方法尚僅能及於處理有限多基因，環境之處理大多以隨機化假設帶過，對基因與環境之交互作用之模式化作法，尚未有明確的討論，故人類有關數量性狀疾病研究，一直未有重大突破。因此有關基因-基因以及基因-環境交互作用的模型及分析方法，將是研究複雜疾病的重要課題。

(h)多反應變數的數量性狀分析方法發展

複雜疾病分析困難度靠單一數量性狀模式分析並不容易解決，設定多個數量性狀同時進行分析是一條解決之路，此牽涉到多變數統計用於家庭資料之方法學發展。目前一重要的發展方向為如何利用內生表徵型(endophenotype)來增強連鎖分析或關聯分析的檢定力。

(i)同家庭資料結構，多方式萃取遺傳訊息的合併分析方法

家庭資料結構因萃取遺傳訊息方法不同，可以設計不同的分析方法，發展結合同(結)構異(方)法的分析方法，可以更有效檢定出複雜疾病基因。在設計上同一家庭的成員應如何選取同質(concordant)或異質(discordant)的手足以達最佳效果，亦是重要的課題。

(j) 貝氏觀點之遺傳分析方法發展

貝氏觀點的遺傳統計方法一直未被充分發展，此方面研究極為重要，如何定義遺傳事前訊息，處理家庭資料複雜性的貝氏統計計算，並非易事。

(k) 鑑取遺傳資料的各種統計分析校正方法發展

家庭資料因取樣之有偏性，而延伸參數估計的有偏性校正，已有一些相關研究，但隨遺傳研究的進展，針對不同家庭資料結構，設計合宜的鑑取校正方法，仍有極大空間有待填補。

(l) 結合遺傳統計方法與生物資訊研究方法之基因搜尋、功能調控方法發展

目前遺傳統計與生物資訊方法的研究仍處於一體兩面，各自發展。未來針對同一問題，結合二方面的想法與技術，發展出一套整體分析的有效方法，尚屬混沌、有待拓展的領域。

III. 計算生物學統計方法

(a) 專門適用於生物資料的統計方法－生物資料可以分為文獻資料、序列資料與數據資料。文獻資料如各期刊所刊登的論文、Gene Ontology 等較複雜的格式，序列資料則是基因序列、蛋白質需列等較規律、統一的格式。長久以來統計處理的幾乎全為數據資料，近來有許多方式將前兩者的問題轉換為數據資料的問題。然而三者的特性畢竟有許多差異，透過轉換解決只是暫時的權宜之計，統計學家應發展更恰當的分析工具，以免面對這類生物資料有削足適履或喝靴搔癢之憾。

(b) 將新的成果納入遺傳統計分析－目前對致病基因的探討已從尋找疾病基因延伸至致病機制的瞭解，例如癌症並非單一致癌因子引起，而是許多因子與許多步驟共同作用的結果。在疾病的治療與預防上，須對疾病相關的基因與其產物進行整體的研究。而致病基因的研究未來仍將要與遺傳統計結合。致病的可能

原因非常多，尤其非單一基因病變，例如高血壓與精神病等，在所有基因的功能未完全明瞭前，遺傳統計的方法仍然有其價值。隨著對基因的功能了解越來越多，遺傳統計的方法也應將這些資訊結合於典型的分析之中，才能有效地利用人類合作的成果。

(c) 發展大量資料的處理方法－過去由於可取得的資料太少，統計必須引進一些無法檢驗的假設，才能順利地分析資料。而生物學的資料爆炸，使得統計有機會 去除大部分的假設，「讓資料自己說話」，資料挖礦、機器學習、資訊視覺化等技術便是因應的產物。雖然統計科學應用於大型資料的成果斐然，但資料量成長的速度仍遠大於電腦改進的速度，尤其是兩兩比較與交互作用的計算量，更是資料量的平方或立方倍，因此大量資料的分析方法（或演算法），以近似、估計或重新設計的方式減少計算的時間，仍是值得持續關注的課題。

(d) 計算工具結合互動的使用者介面－另一方面，計算生物學需要的資料分析與呈現方式，必須有即時的、互動的圖形化使用者介面配合，然而 R Project 的使用者介面尚處於初步階段，仍無較符合需求的成果，分析者必須將就介面，而非選擇切合目的的介面。目前計算生物學的分析軟體多半是研究者或某些團隊自行開發以因應個別需求。由於互動的介面較耗費電腦資源，這些軟體中資料分析的運算往往會大幅簡化，而未使用目前已有的統計資源，殊屬可惜。現已有不同團隊著手進行整合，主要可分為兩大類：一是建構在 BioConductor 的基礎上開發更實用的使用者介面，一是將眾多的統計方法整合為現有的計算生物軟體所用。無論如何，統計計算的結果與圖形化的互動呈現若能緊密結合，將促使計算生物學研究的進展更加快速。

(e) 統計學可用來解決序列片段的連結、DNA, RNA 及蛋白質之序列分析、蛋白質結構及功能的預測、新陳代謝功能之分析及模擬、基因調控網絡及生物反應路徑之探討等。

統計科學應用於生物資料之分析，包括模型建立、模擬、資料採礦、資料處理以及型態之發現等。統計學者可發揮的部分在於這些問題的機率模型、p 值以及檢定力的計算、估計方法、檢定方法閥值的建立、計分方式以及決策法則之建立以及算則的最佳化。例如在模式或計算上，Hidden Markov Models 以及 Markov Chain Monte Carlo 提供了重要的計算法則。

(f) 在生物資訊研究中，有許多大量重複比較的問題，例如數萬基因的微陣列基因表達資料，或多人在數十萬生物標記(SNP)的基因多型性資料均會衍生多重比較的問題，是十分重要的研究方向。

(g) 計算生物學的問題十分豐富，統計方法可扮演重要角色，例如：

(i) 蛋白質交互作用之預測，可採用估計 domain-domain 交互作用的機率，用概似函數或貝式統計的方法，目前仍有很大的改進空間。其他如蛋白體資料之分類，仍沒有很好的方法。

(ii)基因表達資料分析中一個重要問題是辨識有顯著差異表達的基因，並選出標記基因以作腫瘤分類；此外依時間不同的基因表達分析亦是重要的研究方向。其

他如以 array-CGH 的資料來瞭解 DNA copy number 的改變，進而瞭解腫瘤的成因。貝式方法，MCMC 均是重要可採用的方法。

(iii)用序列資料做分子分型的工作，特別是同類病毒中的分型，可利用生物資訊的辦法，如 Hidden Markov Model 來處理，也可發展其他有效的做法。

(iv)分子演化樹分析亦是統計學可發揮的課題。族群遺傳學的研究方法漸被用來瞭解細菌或病毒的序列演化，這些知識在疫苗或藥物研發上亦扮演重要角色。

IV. 一般生物醫學統計

A.診斷醫學統計方法

「收方操作特徵」(Receiver Operating Characteristic)曲線，係同時評估醫學診斷工具的靈敏度(Sensitivity)與特異度(Specificity)，其分析研究已成為醫學診斷統計方法之基礎。但為配合實際應用之情況，有關的統計方法之研究方向仍頗有發展。醫療器材臨床試驗的統計方法，應如何避免檢測的誤差，如產生自病例的確診問題、欠缺無瑕診斷工具的參考等。另外，統計分析應如何調整受測者的共變項，多重檢測工具使用於相同受測者之效度評估，DNA 及基因表現類多標的陣列平台診斷測試中之統計推論，皆是未來的研究重點。無母數統計方法可能是較受重視的研究方式。最近倡議的「藥物與診斷組合之臨床試驗」(Clinical Trials of Drug-Diagnostic Test Combinations)，將基因診斷效度與藥物療效在同一試驗中評估，也將帶動研究的興趣。

B.神經影像統計分析

傳統統計方法多偏重以模式化方法來分析研究資料，這些方法無法有效處理現代科技的複雜訊號與資料。統計及資訊科技持續在進步，資料分析所受的限制將越來越少，且仰賴快速的計算。”神經影像統計分析”規劃的重點為：

(a) 結合資訊、醫工及認知科學的團隊研究、及整合型計劃。

藉助先進的技術如功能性磁振造影(fMRI; functional magnetic resonance imaging)等來收集及分析資料。新科技所帶來的衝擊不僅影響到研究方法也影響到研究生態。不論科學或技術層面的研究必須整合各方面專業人才。例如，採 fMRI 來探討知覺效應必須結合神經、實驗、醫學工程及分析方法等各方面專業人員。統計方法的支援不同於以往僅重在實驗設計及資料分析，而是整個實驗過程的監控。統計學家必須和其他專業充分配合，在過程中評估可能發生的狀況，並尋求解決之道；此外，在實驗過程中必須隨時調整方法以適應新問題的產生。統計研究必須脫離靜態的工作環境並實際參與團隊合作；以解決問題為手段，以推動”重要研究發現”為最終目的。

(b) 大型資料分析方法的專業培育。

國內統計研究所應設計核心課程，培養專業的統計計算人力，投入國內相關科學的團隊研究。課程除包含統計專業訓練外，並加強“統計資訊與機器學習”、“訊號處理及影像分析”、“認知與神經生理”、“醫學工程”等專業課程。自然處應鼓勵國內統計學家參與國外神經影像及認知實驗室研究，及彼此的往訪交流； 或與國外實驗室合辦研討會。自然處(統計組)可推動整合型計劃，結合國內相關人

力研究特定的重要課題。

(c) 鼓勵研發適合神經影像及腦電波訊號分析的統計科技。

結合資訊工程技術以改進資料分析方法是未來發展趨勢。例如目前廣泛運用在分析 EEG 資料的獨立成份分析法(independent component analysis)為以往工程界作控制使用的技術。例如成份分析技術的應用使得實驗心理採 EEG 技術，有不同於以往的重要發現。同時，貝氏網路的估計工作主要是計算行為結果在其他相關知識及策略變項已知情況下的條件分佈(conditional distribution)。當此條件分佈獲知後，便可預測受試者在不同階段可能發生的困難及其所缺乏的相關知識等。由於網路推論中相關的變項過多，所以一般多變量徑路分析的方法不敷使用。此外，在機器學習和統計的連結逐漸脫離之際，彼此可藉神經影像分析重新建立關係。神經影像的資料量龐大且複雜，需要資料採礦技術的協助，以將所有重要的變項完整呈現。換言之，神經影像研究也同時推動資料採礦技術在統計與計算機領域的發展。

V.生物統計中幾個重要的統計方法學研究

A. 存活分析

目前在台灣從事存活分析研究的學者，其研究方向和國際的主流能吻合，在個人專長的問題上，也能掌握新的趨勢。茲將未來存活分析可行研究規劃簡述如下：

(a)加強與生物、醫學跨領域間之合作：許多研究課題在方法論的發展上雖已漸成熟，然單憑創意以發表在一流期刊的難度增大。許多發表在一流期刊的文章，除了提出有新意的方法論外，尚伴隨第一手的資料分析。因此，台灣從事存活分析領域的學者應參與實際的醫學研究計劃，突迫學術上的瓶頸。

(b)針對時間函數之反應值、存活時間及遞迴性事件資料結構，建立更合理且具解釋性的模型。並藉由更廣泛之非參數化潛藏因子模式解釋並建立反應值內部相關，反應值、存活時間與遞迴性事件之相關性及觀測個體非齊一性質被。

B. 長期追蹤資料分析（Longitudinal Data Analysis）

(a)長期追蹤資料迴歸模式化(參數及半參數模式)與估計式問題之研究：由於長期追蹤資料的特性，使得資料分析及模式化的過程有相當大的彈性，加上計算工具的不斷發展，使得較複雜的模式得以加入發展應用，伴隨而來的則是相關估計式的問題。經由統計模式化之研究，發展長期追蹤資料分析之工具，藉以解答各種不同的科學問題，將使長期追蹤資料分析方法更加豐富與精進。

(b)函數型資料分析之應用與方法研究：函數型態的長期追蹤資料，其資料量通常很大，且其資料特性與統計分析方法與傳統的長期追蹤資料分析有許多不同。它可以結合多變量分析、隨機過程、時間序列分析、無母數迴歸、統計計算等方法，其統計分析方法與理論都在持續發展當中，故其統計應用與方法之研究都值得深入探討。

C.遺失資料（Missing Values）分析方法之研究

(a) 遺傳統計及生物資訊學在生物統計領域中將是非常重要且蓬勃發展的

領域,其所牽涉的遺失資料問題預期也將是遺失資料分析方法學領域的重要課題之一。

(b) 長期追蹤資料及存活時間資料的遺失資料分析方法尚未發展成熟，應是值得投入更多心血的研究課題之一。

(c) 隨著高維度資料的日益普遍，高維度資料相關之遺失資料問題將是極具挑戰性與重要性的領域。

(d) 過去關於遺失資料問題的探討多著重於估計與推論問題。未來應可近一步探討其相關之模式選取與模式診斷問題。

D.視覺化統計方法

一般統計分析要處理的資料都是資訊視覺化的可能對象,以下與其它統計研究領域相關之視覺化可能發展課題，都是未來發展的方向：

- **統計資料之視覺化方法**

(a) 類別型 (categorical) 資料之視覺化

當資料型態為類別性，尤其是名目 (nominal) 資料時，視覺化出現二個困難：(i) 連續型資料可以很容易使用單向或雙向的顏色、大小、長度等呈現資料結構,名目型資料的編碼並不具量度的意義,需經過某種尺度化 (scaling) 轉換。(ii) 變數與個體關係之計算不易—欲將名目型資料以視覺呈現，必須求算變數或個體之關係，但相關的文獻中可供選擇的關係不多，大多是列聯表形式，可以採尺度化轉換或對數線性模式 (log-linear model) 等統計方法。

(b) 多時點（相同變項）資料之資訊視覺化

例如某資料為患者入院時數個症狀的診斷結果,資料也有可能在其它時間點取得，如出院及追蹤資料。如何將多時點資料在同一張圖中呈現，或多張並列以同時探索患者、症狀與時間三個因素之交互結構是一個相當具挑戰性的問題，長期追蹤 (longitudinal) 統計模型可能需要此類的視覺化方法。

(c) 多條件（不同變項）資料之資訊視覺化

基因型、表現型、疾病診斷可以視為患者的三個資料集，如何將不同資料、不同變數以及患者三個因素的結構透過視覺化呈現？此問題類似多時點資料之問題,其共通處是二者皆有相同的個體,但是多時點資料每個時點測量相同的一套變數，而多條件資料每個條件下測量不同的變數群，正交相關 (canonical correlation) 類之統計理論應該可以派上用場。

(d) 條件式（變項校正）資訊視覺化

資訊視覺化與其它統計分析模型都可能有需要做變數校正 (covariate adjustment) 之場合，例如性別，年齡對模式選擇之影響。若該欲校正變數為類別型如性別，則可能將變數之關係矩陣分解成群內 (within group) 與群間 (between group) 二關係矩陣之線性組合以探討該變數對其它變項之影響，以及視覺化呈現的差異。

(e) 相依 (dependent) 或群集 (clustered) 資料之資訊視覺化

當資料存在相依結構時，例如以家庭為單位蒐集之社會調查或遺傳統計資

料，資料之資訊視覺化亦出現二個困難：(i) 個體間的關係計算不易—此處之關係存在二個層次，即群集（家庭）間與群集內。群集間之關係如何計算？是否保留群集內之結構？都不是容易解決的問題。(ii) 個體間的關係與群集的結構如何同時呈現？一般之相依資料統計理論或有用武之處。

(f) 巨量資料之資訊視覺化

在一般的電腦顯示器上可以輕易呈現 1000 個變項與 1000 個個體資料的視覺化圖形。1000 個變項以一般的資料處理而言不算少，但是 1000 個樣本點卻不算多。處理上萬甚至百萬筆資料時，計算速度、記憶容量與圖檔顯示皆造成電腦負荷。此時抽樣 (sampling) 方法、序貫 (sequential) 分析、修勻 (smoothing) 技術與影像 (image) 處理等多種理論與方法將可達到合理的近似結果。

(g) 資訊視覺化之遺漏值 (missing value) 處理

資料出現遺漏值時將影響資訊視覺化的結果，如變數與個體的關係求算需要注意，其方法可能與 EM 演算法精神類似。遺漏值如何呈現才不會嚴重影響視覺又能註明遺漏值？另一方面，遺漏值也可能可以用資訊視覺化之特性插補，例如用變數與個體二維的鄰近點進行估計。

- **統計模型之視覺化方法：**

大部分之統計模型皆為配適資料而發展，一個正確的選模過程應該是先經過合適之資料之視覺化，對欲配適之資料有相當程度了解後再進行模式之選取。研究或分析人員經常忽略的工作是對所配適之資料與模式進行更進一步之視覺化工作 – 觀察配適之結果與可能之殘差分布，當然這可能歸因於統計模型視覺化工具之匱乏；也因如此統計模型視覺化尚有相當之研發空間。各個統計領域之模型建構皆有其共通性與特異性，因此統計模型視覺化之發展有其延續性優點與發展空間。

- **參與國際會議：**

國內統計同仁較常參加北美的相關統計會議而少參與歐亞相關活動，應該鼓勵統計界同仁多參加歐洲的 COMPSTAT (by The European Regional Section the International Association for Statistical Computing (IASC-ERS, http://www.iasc-isi.org/)) 與亞洲的 IASC Asian Conference on Statistical Computing (by The Asian Regional Section of the International Association for Statistical Computing (IASC-ARS), http://www.iasc-ars.org/). 北美則有 INTERFACE Symposium on the interface of statistics, computing science, and applications ， by The Interface Foundation of North America, http://www.galaxy.gmu.edu/Interface2006/i2006webpage.html) 等相關會議，

五、資訊科技相關之統計

I. 統計計算

(a) 分群方法

由於實際應用的需要，大資料的分群法理應持續發展。除此之外，也應將分群方法擴展到更廣泛的資料型態上，其中最重要者應屬類別性資料與相關性資料。無論是過去的方法或大資料的演算法，多半只考慮量化資料的分群。然而問卷調查、消費交易資料或生物醫學現象，經常都無法避免類別性資料。類別性資料的分群方式可能更為困難，因不同類別的差異不容易量化或定義，更遑論將視覺化呈現應用在類別性資料的分群上。另一方面，個體間彼此獨立的假設也往往不切實際，例如個體來自同一家庭或同一班級將存在某些相關性，如何在分群時將相關性納入考慮雖是頗困難的問題，卻有其實用性與急迫性。

(b) 分類方法及樹狀法(Classification & CART)

1. 找出最佳樹狀結構以降低誤判率一直是分類樹研究中一項主題，其中包含變數選取，分割點的決定、停止規則(stopping rule)與不平衡資料 (unbalanced data) 分割原則的探討。
2. 大量且高維度的資料，將使樹狀圖的呈現方式受限制，針對大量資料發展其它的分類視覺化方法，與如何顯示分割資訊也是未來的研究方向之一。
3. 分類法則的優劣，需要一些標準資料集來測試，雖有國外常用的資料庫(例如 UCI)可供驗証，但建立一套國內分類法則專用的資料庫，並提供國際使用，讓其成為標準，也是提昇國內學術表現的方法。

(c) Sliced Inverse Regression (SIR,切片逆迴歸法) and PHD (principal Hessian direction)

切片逆迴歸法的優點在於其簡單的演算法及其廣泛的應用層面。推動重點除了一般研究人員在統計模型理論上的再鑽研及與不同的研究領域相結合外，各學校之間也應運用資源、整合人力，規劃專題課程，並重理論與實務，讓學生選修。同時學生們應培養撰寫程式的能力，並能應用於不同的資料型態。

1. 針對具類別預測變數之有效維度縮減之研究(因具類別型態之預測變數，其線性組合不易解釋)，考慮架構新的加權核矩陣去估計 CPS/CPMS 維度縮減空間；參考 Li, Cook and Chiaromonte (2002, 2003), Ni and Cook (2006)。
2. 針對多變項反應變數之有效維度縮減研究，迄今仍少有分析方法可用。此研究強調尋找最具訊息之反應變數與預測變數間的線性組合，以達到有效維度縮減之目標；參考 Li, Aragon, Shedden and Agnan (2003), Setodji and Cook (2004)。
3. 針對測量誤差之有效維度縮減研究；參考 Carroll and Li (1992), Lue (2004)。
4. 針對非線性迴歸設限資料之有效維度縮減研究，可採取設限誤差之條件期望估計值修正，並搭配平滑配適法，以達有效維度縮減；參考 Li, Wang and Chen (1999)。

5. 近期 Cook and Li (2004)提出 iterative Hessian transformation (IHT)之維度縮減方法，用以估計維度縮減空間 CMS，此提供了除了 OLS 與 PHD 之外的估計 CMS 方法；另外，Chiaromonte, Li and Zha (2005) 提出 contour regression 維度縮減估計方法。
6. 可將維度縮減方法(SIR 或 pHD)，推廣至曲線反應變數資料之研究、泛函資料分析等方面。概括而言，對於高維度非線性迴歸問題，SIR 與 pHD 方法是深具發展性的分析方法，可廣泛地應用於各種統計領域，如實驗設計、區別分析、形態辨識等方面。經由樹形迴歸，測量誤差迴歸，設限迴歸與多變項反應變數迴歸等的探討,將使得此兩種方法成為較可行且具幾何觀點分析的絕佳工具，相信研究結果將可增進迴歸分析方法的研究方向。
7. 高維度資料分析-維度縮減方法研究、非線性迴歸分析、多變量分析及統計計算與有效維度縮減方向估計方法。
8. 樹形迴歸、敏感度分析、測量誤差迴歸。
9. 多變項反應變數之有效維度縮減研究。
10. 泛函型資料分析(functional data analysis)，長期追蹤資料(longitudinaldata analysis)，時間序列(times series)。
11. 希望國科會能建立資料庫，讓研究者能將新的方法論運用於收集之資料，便於分析。
12. 統計模型及演算法的改進，使其能適用於不同的資料型態。
13. 演算法核化(kernelization)及與其它統計方法的比較及結合。
14. 針對大量且高維度資料的理論發展及分析應用並結合現代計算機及圖學迴歸法，發展互動資料分析工具。

II.資料視覺化

高維度的資料視覺化方法應是未來積極發展的重點，隨著科技的發展，蒐集資料的成本大幅降低，越來越多的資料包含許多變數與大量樣本，傳統的統計方法是否仍適用如此複雜的資料漸受質疑。而欲檢驗統計方法的合理性，視覺化的呈現是最直接也最快速的途徑。這些資料的複雜性，除了樣本數量大與高維度外，還有不符合獨立假設、遺漏值普遍存在、變數尺度不一等問題。就研究前景而言，視覺化方法是一個頗具潛力、仍有許多課題未解決的領域；就實際應用而言，除了檢視其他統計方法是否有效外，也幫助各類科學研究或社會現象直觀地了解蒐集到的資料。龐大資料的視覺化，往往不適合直接應用既有方法，必須另外設計恰當的方法，以因應資料膨脹衍生的特性。所以此領域需要更多的研究人力投入，並有合作與競爭以加速發展的速度。

III. 統計學習(Statistical Learning)

(a) 回饋式學習 (Reinforcement Learning)

1. 貝式控制理論的研究。
2. 控制、預測及選模三合一問題。

3. 多個學習器(監督學習，非監督學習，及回饋式學習)的整合問題。
4. 生物資訊的應用。
5. 網路安全的應用。
6. 經濟及財務預測上的應用。

(b) 獨立成分分析法(Independent Component Analysis, ICA)

1. 專注於過度完備的獨立成分分析法(overcomplete ICA)：這是一個相當困難的問題，因為我們只能使用少量的資訊來預測(或者估計)原先的來源。過度完備的獨立成分分析法的特性至今仍然不甚清楚，並且對於統計學家來說也有深入研究的價值。
2. 連續且有相互關聯的來源：在獨立成分分析法與過度完備的獨立成分分析法中，資料來源的特性是獨立的邊際稀疏分佈。然而，對於不明來源的分離問題，時間序列模型和一些其他的統計模型能用來找出相對應的信號。至於如何來調整獨立成分分析法與過度完備的獨立成分分析法的演算法以符合不同的模型假設則是另一個有趣的問題。
3. 應用於過度完備的不明來源分離(overcomplete blind sources separation)、醫學影像…等相關的問題。

IV. 應用(Applications)

(a) 資料探勘

1. 由二元分類到多元分類問題。
2. 由分類問題到排序（ranking）問題。在許多機器學習的問題與實際應用中，目標不再是簡單的將事或物做分類，而是將事或物依照需求或正確的可能性由大到小排序（ranking），且近期關於排序及其相關議題的理論也正在逐漸發展中。
3. 由特徵選取（feature selection）延伸至特徵排序。
4. 如何依問題特性，選取較適合的分類演算法。
5. 如何評估檢定不同的分類演算法結果的差異性。
6. 如何有效地組合不同的分類演算法 （ensemble learning）。
7. 如何評估分群演算法的結果。
8. 構建提供自動化資料分析的平台。
9. 為現有的方法提供較嚴謹的統計詮釋。

(b) 統計在網路分析及資訊安全的研究

統計方法在網路評估及資訊安全上的研究仍然有很多熱門的議題亟待統計專業研究人員的參與。大體來說，未來相關的研究重點可以劃分為以下幾類：

1. 網路效能分析：利用統計方法分析系統或網路於一天或一週等的運作模式，嘗試對系統或網路效能最佳化的工作。
2. 入侵或不當使用之偵測:對於計算或網路資源的一般正常使用與不正當使用界定出模式，以其對不當使用(可能為入侵或非入侵)狀況發生時採取立即的圍堵措施。譬如對盜刷信用卡者行為模式的關聯性法則分析(association rule)

可以適當的建立信用卡盜刷的行為模式，在第一時間停止其使用權。此方法的研究可以以即時和非即時的方式來運作。

3. 身分驗證：針對某特定人士建立其資源使用行為模式，以其對可能盜用的情形作第一時間的暫時停權處置。此法雖不可能得到百分之百的正確率，但是對特殊未知盜用方式卻提供一套另一層的安全防護。
4. 智慧卡或 RFID 等的相關技術：智慧卡等的廣泛使用將是現代科技進步的指標。統計方法可以對大量卡片提供的資訊做分析，使得效能上或安全上可以更達到最佳化的需求。
5. 無線網路及其安全：無線網路將逐漸某種程度的取代一般網路。統計方法的分析需要考量無線網路不固定的二維或三維的幾何性質。另外也可以嘗試在無線網路機制低能量、低運算能力、低成本的考量上提供適當的分析。

六、地球科學及環境相關之統計

I. 地震活動與前兆之統計分析

1. 地震危險研究：建立地震目錄分析方法，加強相關地球物理知識，創新地震活動統計研究。此外，增進建物結構了解，研究地震危險分布，進行地震保險精算工作。
2. 地震前兆研究：針對具有完整測量之電磁變化數據，探求與地震活動的相關性，並且評估前兆訊息預警強震效益。此外，利用主震引起的應力改變或相關地質結構訊息，亦可用於預警強餘震之可能發生區域。

II. 颱風降雨量與風速之統計預測

由於颱風夾帶的強風豪雨對人類生命財產或大自然生態保育直接造成傷害，因此防災為急迫的需要，其所衍生的關於政令、法規、風險評估等問題都需要統計方法的介入。一切尚在起步，因此未來發展空間很大。未來規劃重點為

1. 雨量與風速預測
2. 颱風與生態發展
3. 土石流的預警與防範
4. 風險評估
5. 農漁業政令宣導等

III. 空氣污染與健康危害之關係

國內有不少流行病學、環境科學與大氣科學領域的專家學者從事空氣污染與健康效應的個別研究，但缺少整合型的大計畫。主要原因可能包括研究經費的限制、難有多年期的大計畫補助和缺乏執行大計畫的領導人才。統計學者因而很難有參與這類大型計畫的機會。等待機會不如創造機會，國內統計學者可以考慮學習美國幾個大學環境統計研究中心的作法，成立研究群來主導研擬空氣污染健康效應的整合型計畫。以發展研究方法為主軸的研究群也可以同時與國內外空氣污

染研究中心合作來開拓這個領域。

IV. 人體測量值和環境評估之統計方法與應用

目前國內人體測量值的資料型態多數為勞工尿液的生物標記檢測值,血液檢測值的數據極為有限,現階段可先發展以血液或尿液的生物標記為主,人體毒物動力學模式的統計方法與應用，統計學者可積極參與環保署或勞委會的研究計畫，以取得研究數據，並爭取合作機會。至於在環境污染物暴露評估，以及進一步在健康風險的統計方法與應用，可列為中長期努力的目標。

V. 環境與健康相關風險評估

由於風險評估牽涉領域甚廣，而所投入統計學者研究人力極為有限，以下僅就統計方法在環境風險評估中的暴露評估、劑量反應關係、以及風險特性化三方面，列出可以或正在進行中的研究子題，俟研究成果與資源人力逐漸豐碩，再尋求跨領域整合型研究計畫與建立研究群。

A.在暴露評估方面

1. 多介質動態傳輸與轉換模式（multimedia dynamic transportation and transformation model）的統計模式的建立與應用
2. 垃圾焚化爐所產生戴奧辛與重金屬環境暴露評估
3. 人體毒（藥）物動力學模式的參數估計與在職業暴露評估的應用
4. 石化與冶煉工業環境暴露與流行病學相關研究在劑量反應關係方面
5. 高低劑量的外插（如 Benchmark Dose 的計算）相關統計方法研究
6. 高危險暴露族群環境流行病學相關研究
7. 環境暴露與內在劑量的統計關聯

B.在風險特性化方面

1. 廣義風險指標（包括地域、時間、與族群等）的建立
2. 不確定性分析（如食品環境污染物暴露變異量）相關統計方法研究

至於微生物風險評估，由於其評估程序相同，但所需專業知識迥異，且國內目前缺乏相關的研究數據，在現階段規畫重點上，暫時以環境風險評估為主，一旦有相關領域專家學者的合作需求與機會,將可藉助移植在環境風險評估研究所累積的經驗，進行跨領域合作。

VI. 生態統計

我國在 2004 年中研院成立了「生物多樣性研究中心」其後東海大學也成立的「熱帶生態學與生物多樣性研究中心」，然而統計界與生物界的交互作用有待加強與開展。未來的規劃重點應著重在本土動植物數據的統計分析之應用。各國環境與物種不盡相同，無法完全引用他國知識，必須各國展開其各自特有種的研究。我國在此方面有待統計學家的加入與努力。

七、社會及經濟統計

I. 財務統計領域報告

建議未來規劃重點可以包含下面幾個方向:

(a) 模型的推廣

財務時間序列模型應朝向更貼近資料特性發展，諸如應包含非均質變異、均數回歸、厚尾分佈等特性。例如考慮修改 Black-Scholes-Merton 模型中，波動σ (volatility) 為常數的假設。因為在現實的狀況中，σ為 t的凸函數，而不是常數,σ的計算對於選擇權評價及一般衍生性金融商品的研究佔有著舉足輕重的地位。

(b) 衍生性商品定價的研究

可朝向美式選擇權及奇異選擇權在不完整市場的探討。此外由於隨機分析所能運用的工具有限，有些相當好的模型不能被很完整地解出來。此時，統計對於隨機分析相關的計算問題提供一個很好的工具。例如利率的模型真正被廣為接受的並不多,其中的原因是模型必須滿足利率的正值及均數回歸等性質。例如 Cox-Ingersoll-Ross 模型，利率(r_t)滿足隨機方程$dr_t = (b - a\, r_t)dt + \sigma\sqrt{r_t}\; dW_t$，但此方程式極難解。因此這時便需靠一些統計與數值方面的方法來解決這個問題。

(c) 信用風險的定價及信用及風險的評估

對於信用風險的定價模型除了結構與縮減式模型,可以考慮提出一些更具實用性的模型。此外信用及風險的評估在風險管理也是一個重要的課題。

(d) 市場的微結構與高頻資料研究

市場微結構的研究，例如買賣價差，市場的流動性，價格的效率以及資訊傳遞的速度都是財務統計可以發揮的領域，並須培養處理高頻資料庫、高頻資料的分析及具備程式能力的人才。

(e) 應用機率與統計的方法求最佳解及隨機微分方程求數值解

此外如何應用機率與統計的方法探討離散時間模型求最佳解,及隨機微分方程求數值解等問題，也是一個值得研究發展的方向。

(f) 長記憶過程、馬可夫鏈過程及 Copula 的研究

此外如長記憶過程、馬可夫鏈過程的研究、及 Copula 的理論與應用也都建議規劃為未來財務統計的研究發展重點。

II. 教育統計領域

未來發展方向與重點：

(a) 培養測驗統計專業人才，及研發測驗統計相關軟體。

(b) 編製適用於國內的智力、性向、成就、人格、興趣等領域之測驗或評定量表。

(c) 培育未來教育界測驗之實施、管理者。

(d) 因應特殊教育發展的需求,培育特殊兒童之鑑定、診斷與評量之專業人才。

(e) 協助各級學校改進教學評量技術，以提昇教學成效。

(f) 提供國中小教育相關人員進階的實務及系統的方法論學程，使其具備教育測驗與評量及研究方法的專業素養，強化國民教育師資結構中的教學評量與研究方法的人力資源。

(g) 進行測驗評量與科技整合議題的探討，如科技與評量、電腦化動態學習診斷、及認知科學與評量等前瞻性研究，以利學生學習進展的積極經營。

III. 認知科學

統計及資訊科技持續在進步，資料分析所受的限制將越來越少，且仰賴快速的計算。此外，認知實驗研究開始藉助先進的技術如功能性磁振造影(fMRI)等來收集及分析資料。新科技所帶來的衝擊不僅影響到研究方法也影響到研究生態。不論科學或技術層面的研究必須整合各方面專業人才。例如，採 fMRI 來探討知覺效應必須結合神經、實驗、醫學工程及分析方法等各方面專業人員。統計方法的支援不同於以往僅重在實驗設計及資料分析，而是整個實驗過程的監控。統計學家必須和其他專業充分配合，在過程中評估可能發生的狀況，並尋求解決之道；此外，在實驗過程中必須隨時調整方法以適應新問題的產生。統計研究必須脫離靜態的工作環境並實際參與團隊合作；以解決問題為手段，以推動重要研究發現為最終目的。傳統統計方法多偏重以模式化方法來分析研究資料，這些方法無法有效的處理複雜的訊號。例如目前廣泛運用在分析 EEG 資料的獨立成份分析法(independent component analysis)為以往工程界作控制使用的技術。成份分析技術的應用使得實驗心理採 EEG 技術，有不同於以往的重要發現。所以結合資訊工程技術以改進資料分析方法是未來發展趨勢。貝氏網路的估計工作主要是計算行為結果在其他相關知識及策略變項已知情況下的條件分佈(conditional distribution)。當此條件分佈獲知後，便可預測受試者在不同階段可能發生的困難及其所缺乏的相關知識等。由於網路推論中相關的變項過多，所以一般多變量徑路分析的方法不敷使用。統計規劃的重點為：

(a) 結合資訊、醫工及認知科學的團隊研究、及整合型計劃。
(b) 大型資料分析方法的專業培育。
(c) 鼓勵研發適合神經影像及腦電波訊號分析的統計科技。

八、人才培育

人才培育是一個非常急迫的問題，亟待在下述幾個方面努力：

1. 增進年輕人對統計科學的認識，激發他們對統計的興趣，進而吸引他們進入統計的領域。
2. 改善目前統計的課程與教學。
3. 鼓勵年輕學生出國攻讀博士學位。
4. 提供國內博士生及國內培養的博士出國進修的機會；鼓勵他們在畢業後從事博士後研究以拓展視野，提升研究能力。

陸、資源需求與配合情形

(一) 現況分析

年度	經費(萬)	件數	年度	經費(萬)	件數
78	701	34	87	4374	138
79	787	39	88	4743	143
80	1267	50	89-1	5294	152
81	1696	61	89-2	5788	152
82	2481	85	90	6368	170
83	2569	85	91	7755	151
84	2598	88	92	8248	160
85	3750	126	93	9069	166
86	3899	146	94	9420	159

[註] 89 年會計年度改成日曆年所以 89-1 指 88.8.1-89.7.31，89-2 指 89.8.1-90.7.31。

上表顯示國科會統計計畫 (以自然處為限) 的件數在近幾年快速成長，預料此成長仍會維持數年。除原有之統計系及數學系或應數系之統計研究人員，現在已有十一個統計研究所 (中研院、政大、中興、中央、清大、交大、成大、中正、逢甲、東海和輔仁)[4]，比三年前多了四所，預計仍會再增加。各校相繼成立統計研究所，一方面是此學門重要，一方面是國內環境與國人才至均適合發展統計。{

國內統計界的學者人數雖與其他學門相比仍較少，但在國際上表現相當優異。例如，有數十位國際統計學院 (International Statistical Indtitute) 的 elected member，數理統計協會 (Institute of Mathematical Statistics) 的 fellow 也有好幾位，其他如：美國統計協會 (American Statistical Association) 及第三世界科學院 (The Third World Academy of Sciences)，在國內統計界也均有 fellow。另外，數份重要的國際統計期刊，國內均有副編輯 (associate editor)，例如：American Statistician，annals of Statistics，IEEE Transactions on Reliability，Biometrics，Journal of Statistical Planning and Inference，Journal of Multivariate Analysis，Statistics and Probability Letters。尤有進者，由中研院統計所與泛華統計協會合辦而在台灣發行的統計期刊 Statistica Sinica，創刊於西元 1991 年，自 1992 年起便收錄進 SCI Journal Citation Reports 中，算是以相當快的速度被國際學術界所肯定。在 Statistics and Probability 項下，按影響係數 (inmpact factor)，1992 年排在第 22

[4]：現況--東海大學、真理大學、中央大學、台北大學、交通大學、成功大學、政治大學、高雄大學、清華大學、彰化師範大學、淡江大學、逢甲大學、輔仁大學、銘傳大學。

名 (共選入 44 份)，1993 年排名 19 名 (共選入 45 份)，此實為中國人的一項殊榮。才發行五年便超越許多已發行數十載的統計期刊 (如：日本、北歐及加拿大等學術先進國家統計方面的代表性期刊)，且在所有新期刊評價是最好的。此期刊並將自 1996 年起，由目前的每年發行兩期，改為每年發行四期。而國內統計界的代表性刊物--中國統計學報，也自 1994 年起，由原先的每年發行兩期，改為每年發行四期，此亦反應出國內統計研究的欣欣向榮。又值得一提的是，近五年來且有不少位在國際上頗富聲明的華人統計學者回國定居，貢獻其研究經驗，對國內統計研究水準之提昇增加不少助力。

由於研究成果的突出，，帶動了統計活動的蓬勃發展，使各種歷史悠久的國際統計組織之下的研討會相繼在國內舉辦，，許多著名的統計學者，也都樂意來台講學。另一方面，有鑒於統計畢竟是一門應用科學，除了專注於理論的探討，已持續追求國際上較高的學術地位,統計界更已開始注意到積極將統計方法應用至各領域中，以促使國內科技提昇及使經濟持續發展。

另外，由於踏進應用的領域，統計對資源的需求遂與日增加。統計在許多領域均需要跟社會結合，於是對於人力有相當的需求。如做調查、做數據分析及與外界接觸均需要充分的人手。

事實上，在某些應用領域需要專人來協助收集和分析資料，如此計畫才易進行。現階段因此項人力取得困難，再加上應用方面的研究成果不易被肯定，致使很多統計學者猶豫做這方面的計畫。至於儀器設備的需求和合法軟體的採購，在現有侷限的經費下，往往無法達到令人滿意的供應。

(二)未來之需求與展望

(1)、成立研究中心

如何使發展出來的統計方法實際可用是一重要課題,建議在推動統計學門重點方向時，能陸續在各主要的學校成立統計中心，分別給行政與技術的專任助理、提供必要的設備、常用統計套裝軟體的經費及一些行政雜支，使其能充分運轉。主要工作為收集資料，然後加以分析，並提供其他領域人員諮詢服務。將來可採取由本會支援部份經費，而該中心要有配合款的方式進行，而視其績效作為往後支持之依據。

(2)、提供充分的人力設備

在重點支援的領域，視其需要，給予充分的人力配備，包括增加博士後研究的名額，博士、碩士研究生的名額能更寬裕，以及支持專任助理。本會應按期評估其成果而決定是否繼續給予如此充沛的人力。

(3)、增加合作聯繫費用

增加舉辦國際或國內學術研討會的經費，計畫中可編列旅運費，包括邀請國內外相關領域的研究人員來共同合作，及主持人之出國開會費用等 目前只有整合型計畫才能編列)。另外，並可有舉辦座談會的經費，以使統計學者能與其他領域的研究人員有交換心得的機會。

(4)、儀器設備的支援

有些計畫需要大量借助計算機才能進行,此時本會宜對儀器設備方面盡量支持。

(5)、加強統計在各領域的應用

為能充分掌握我國經濟發展的情形，需做許多統計調查，且需用所得之統計資料，對未來做預測，並作為制定政策之依據等。統計工作者在其中均扮演重要角色。國科會若能有適當的支援及鼓勵，將使統計學者在理論及應用方面均能有傑出的貢獻。

柒、規劃重點的推動

一、 統計方法與理論

I. 非參數及半參數模型與方法

這些前瞻性研究課題在國內已有相當的基礎，也有極好的成果，但更長程的發展，須重視人才培育、鼓勵研究群形成、及加強與國際學術交流。

II.模型選取與識別

1. 鼓勵及支助國內學者參加國外研討會及短期訪問。
2. 支助國內學術機構舉辦國際研討會並邀請知名學者參與。

III. 時間序列與空間統計

(a) 鼓勵對時空統計問題或對分析複雜資料有興趣的學生或學者，投入時空統計領域。

(b) 由於時空資料源至大氣、水文、環境、流行病及生態等領域，此類資料的分析有時由非統計專業人士執行，因而可能在採用過度簡化的統計模式下產生較不精確的估計或推論。我們鼓勵統計學者，與資料來源的領域專家合作，並逐漸建立互惠關係。一方面統計學者可以幫助這些領域提供嚴謹且有效率的統計方法，另一方面這些領域如能將一些珍貴的時空資料提供給統計學者，統計學者也可藉由分析一組複雜的時空資料，得到一些啟發甚至發展新的統計方法。

IV. 模型選取與識別

（a） 非條件均質變異時間數列的選模。
（b） Boosting 最佳停止法則的研究。
（c） 潛在變數(Latent variable)模型及測量誤差模型的選取。
（d） 高維度資料的分類及選模。
（e） 非穩定時間數列之分類。
（f） 長期相關時間數列的預測及選模。
（g） 隱藏馬可夫模型的預測及選模。
（h） 半或非參數迴歸模型的選取。
（i） 貝氏選模及模型平均(Model averaging)。
（j） 時空數列的預測及選模。
（k） 選模的非大樣本理論之建立。
（l） 離散資料的選模。
（m） 選模後的統計推論問題。

V. 模型選取與識別

(a) 鼓勵及支助國內學者參加國外研討會及短期訪問。
(b) 支助國內學術機構舉辦國際研討會並邀請知名學者參與。

二、 應用機率

雖然國際上應用機率的研究日益成長,但近年來國內這方面研究人才的成長則極有限。唸統計的學生，多半傾向走入較統計的領域，而應用機率會用到較多的機率，使有些學生不易親近。為吸引更多年輕的學子投入此領域，可考慮每年在適當時機舉辦應用機率科學營，鼓勵大學相關學系的學生參加。邀請適當講員，以生動活潑的題材，讓他們了解應用機率各領域之內容、重要性，及值得探討的方向，在各研討會也應儘量設置應用機率組，甚至以應用機率為主題舉辦研討會。以國人才智，相信應用機率的領域，是很適合國人投入研究的。

三、 生物統計與生物資訊

(a) 期望國科會提供適當經費推動，鼓勵整合型計畫，並能與國內生物醫學需求配合，整合型計畫的形成可由統計的研究者與其他領域的研究者共同組成，也可參與其他學門的計畫，儘量瞭解其領域知識(domain knowledge)，由接觸瞭解的科學問題，轉會為統計問題。對生物統計成果的評估，應對方法論的研究以及應用型的論文給予認可，在 SCI 國際期刊發表可作為評估的指標之一。

(b) 注意人才培養：

如何讓年青學者接觸到新的問題,可由參與有資料來源的計畫著手,這些資料可提供為研究材料，如此也可刺激一些新的方法研究。

在國內訓練學生方面，可設計一些針對特別專題的短期研習課程或科普演講，使其對新領域產生興趣與瞭解。可舉辦科學營之類的活動，也可推動一些如千里馬計畫，選優秀的年輕人員出國訓練。

在培養人才方面,我們能培養多少年輕研究者仍留在研究社群,且維持其研究的熱誠，將是一大挑戰。

在就業方面，因是市場取向，受到外在環境影響，例如大多藥廠規模小，較無研發環境,CRO(Contract Research Organizations)的市場亦小等，如何建立像美國 FDA、NIH 那樣的工作機會將是一努力目標。

(c) 建議政府如衛生署、國科會等機構訂定措施獎勵統計研究人員從事臨床試驗相關工作，國科會支助之臨床試驗計畫應要求有專業統計研究人員參與，以提高品質。

(d) 推動建立”統計師”制度，生物統計與生物資訊專業領域人才缺乏，鼓勵相關科系學生或老師出國進修相關領域之研發。

(e) 成立“統計中心”推動發展以及與別的領域整合研究，使能發揮最大功能。

四、 資訊科技相關之統計

I. 資訊科技暨物理科學相關之統計規劃重點的推動

我們認為現在是一個團隊合作的時代。沒有一個單一門學科能夠解決現今愈趨複雜的研究問題；尤其在統計與資訊科學所共同鑽研的範疇；像是在生物資訊(bioinformatics)，文件探勘(text mining)，(醫學)影像處理，聲音、多媒體的分群分類等研究的課題上，往往涉及不只一、兩個領域專長。因此重要的是如何培養我們自己能在這些關鍵議題的研究討論中，成為一個良好的溝通者才是最重要的。因為唯有如此，才能真正完整的了解問題，也才可以進一步討論如何解決問題。

我們認為成為一個良好溝通者的首要必備條件為傾聽與尊重，並了解彼此的領域特質，並且學習更多相關的基本"技術細節"是達到目標的最好方法。所以，我們建議在統計學程中替學生設計涵蓋適合各學門的課程。鼓勵並促進各學門的研究學者多進行團隊合作，包括研究預算在內，並表彰各學科達成的研究成果。而不要只是專注於目前研究的議題上，這些鼓勵辦法應該同等地擴及到各領域間的彼此合作，如此，統計學門在未來就能發展的更好。

最後我們從 Carnegie Mellon 大學將成立美國第一個機器學習系來看，跨領域的研究將愈受重視，資料探勘僅是一個學門合作的例子。Carnegie Mellon Provost and Senior Vice President Mark S. Kamlet 於新聞稿中說: *"Drawing on both computer science and statistics, machine learning is another example of how the university's tradition of multi-disciplinary research produces important scientific insights."* 統計與資訊學門的特性，使得我們"能夠"有機會參與各領域的研究。另外，從合作研究的觀點，我們希望在統計學程的設計上，應讓學生多涉獵其他相關學門的知識，各個學校，系所應發展各自的特色，並配合各校及區域性的發展，來設計學程，方可使統計穩健發展。

推動重點除了一般研究人員在統計模型理論上的再鑽研及與不同的研究領域相結合外，各學校之間也應運用資源、整合人力，規劃專題課程，並重理論與實務，讓學生選修。同時學生們應培養撰寫程式的能力，並能應用於不同的資料型態。這方面的訓練，可從兩方面來進行。培養統計科系學生的程式開發能力，及加強資訊科系學生的統計概念。比較詳細的項目則列舉如下：

1. 提供相關跨領域專題研究計畫補助，並鼓勵產學合作以增加經費來源，並減輕對國科會的依賴。
2. 透過整合性計畫，使資訊工程、生物、統計的研究人員參與合作研究，在解決實際問題的過程中，產生新的統計方法。
3. 提攜年輕有潛力的學者，鼓勵基礎與應用並行的研究。
4. 鼓勵及支助國內學者多參加國外研討會，尤其是 Multiple Classification System (MCS), International Conference of Machine Learning (ICML)等等俱有跨領域特性的研討會；此類有類審稿制度會議論文(refereed conference

paper) 通過率往往只略高於 20% (以 ICML 2004, 2005 為例)，在資訊工程甚至於比一般的期刊論文(journal paper) 還更具有貢獻或影響力，甚至在資訊工程系所被視為升等之依據，故我們應對於此類有審稿制度的會議論文(refereed conference paper)給予一定的評價。

5. 鼓勵團隊合作，鼓勵聘僱專任研究助理、博士後研究員或技術人員協助研究，並多舉辦增加國內學術交流的研討會。
6. 舉辦資料分析競賽，鼓勵統計及非統計之學生參與，並提供獎助學金。
7. 支助國內學術機構舉辦國際研討會並邀請知名學者參與，尤其是能舉辦序列式的研討會。每年應開辦專題研討會(workshop)，鼓勵各界參與。
8. 加強各研究單位間的互動，如中央研究院與各大學的師生交換、資料共享等，讓學生有機會接觸更多的統計問題，研究人員也能獲得最有興趣的人員協助研究。
9. 增設大學中的統計學系，使得大部分的大學都能有統計系所與統計諮詢中心，協助其他系所和地區性產業解決統計分析問題，同時增加統計專長的就業機會。鼓勵移地研究及短期訪問。

五、 地球科學及環境相關之統計

目前國內已有一些統計人員投入上述研究，然而多半是因個人關係取得與其他領域學者合作的機會。綜觀上述研究主題，攸關國計民生，其所涵蓋的統計問題與資料包羅萬象，是非常值得深入挖掘的一塊寶地，目前雖有一些嘗試性的研究，但礙於人力、經費的限制，在發展新統計方法與解決實務問題兩目標中常有顧此失彼之憾。欲發展此一領域之統計研究，統計學者應化被動為主動，從過去配角的身分晉升為主角，主動整合成立研究群，執行以統計人員為主的整合型研究計劃，設定研究方向，再邀請其他領域學者的加入，建立溝通協調的互動平台，相輔相成，在合作中居於領導地位，方能積極發揮統計科學的影響力，並吸引其他有興趣人員的參與；而上述幾個研究主題，除生態統計較臻成熟外，仍屬統計學者單打獨鬥的狀態，但其實應可廣邀統計學者就上述規劃重點，以整合型計畫方式主導執行。另外仍應鼓勵統計學者去參與地球科學及環境科學研究團隊，有更多機會瞭解彼此專業，提升合作品質，自然有機會主導。

至於整體推動則希望藉由國科會的協助，在開始時以鼓勵的立場提供充裕的研究經費，撥允上述整合型合作計劃，在成果的衡量上，對非統計主流的期刊也能給予應有的重視，同時並鼓勵團隊合作，經費的使用更加彈性，如聘任專任助理等，使得有興趣、有能力的學者願意積極投入，也讓其他領域的合作者深入了解統計人員的工作，並體認到統計的重要性。另一方面，亦應思考培育未來年輕研究學者投入參與，以及提供其生涯規畫的願景(學術或非學術)，以吸引更多年輕優秀學子加入研究團隊。例如主動到地科或環境相關系所開設統計分析或諮詢的課，當這些相關領域的學生未來投入學術研究，由於對統計有較深的認識，自然會前來尋求協助，而創造出合作的空間。

六、 社會及經濟統計

I. 財務統計

隨著台灣加入 WTO，經濟全球化發展的趨勢，各種新型金融性衍生商品不斷的推陳出新，蓬勃發展。在風險管理方面由於巴塞爾新資本協定將於 2006 年實施，各國政府勢將參採巴塞爾資本協定與相關公報為準則，建立各類風險定義，並將風險控管精神納入金融相關法規內。如何配合巴塞爾新資本協定，建構完整且適合我國金融環境之風險模型架構，提供評估與強化銀行風險管理之準則依據，並健全銀行業的風險管理機制，將成為啟動我國金融業永續發展之鎖鑰。

為了提升台灣在國際上的競爭力，台灣的金融界勢必仿照歐美一般吸收大量機率與統計人才，來從事研發工作。對未來有興趣進入金融界的學生，學習統計與數理財務的知識，將會是必要的訓練。由於統計機率、數學與電腦是數理財務研究的基礎訓練與主要知識，研究者必須具備嚴謹的數理背景，才能充分利用過去資料與數據加以分析，進而推斷並建立模型，並對未來進行預測與分析。因此未來規劃重點的推動如下：

(a) 強化大學部及研究所統計教學
加強在大學部及研究所的統計理論與計算及數理財務教學。

(b) 加強跨領域的合作與交流
鼓勵經濟、財務、數學及統計等學門跨領域的合作與交流，成立跨學門的數理財務及財務工程學程，幫助並培養年輕學子學習這方面的知識。

(c) 舉辦學術研討會
在學術研究方面也要鼓勵舉辦研討會以加強國際間的學術研究與交流。

II. 教育統計領域

由於傳統的統計系很少老師涉入教育領域，所以目前多數的教育統計學者是由教育學院或師範大學培養，且在教育學門提出計畫申請。未來本領域的規劃與推動，應結合自然處統計學門有興趣的學者一同推動：

(a) 加強教育統計學者之電腦、網路及數理統計訓練
由於電腦及網路的興起，對教育統計學者有極大的影響，目前電腦模擬、網路測驗軟體都方興未艾。除了電腦化技術外，研究者若有堅實的數理統計基礎，才能處理在教育情境中的資料型態非常態、違反假設等問題，並發展新的模型。

(b) 鼓勵統計學者投入資料庫、抽樣、及國際比較等議題
資料庫、抽樣、及國際比較等議題，也是未來重要的推動方向，或可結合社會學家等一起來做。國內目前有的高等教育資料庫、台灣教育長期追蹤資料庫、台灣社會變遷調查等，都值得統計學者投入。

III. 認知科學

(a) 國內統計研究所應設計認知科學核心課程
國內統計研究所應設計核心課程，培養專業的統計計算人力，投入國內認知科學的團隊研究。課程除包含統計專業訓練外，並加強認知、神經生理、訊號處理、醫學工程、及影像分析等專業課程。

(b) 鼓勵國內統計學家與國外認知實驗室學術交流
自然處應鼓勵國內統計學家參與國外認知實驗室研究，及彼此的往訪交流；

或與國外實驗室合辦研討會。

(c) 推動整合型計劃

統計學門可推動整合型計劃，結合國內相關人力研究特定的重要課題。

七、 人才培育

目前國內統計所(或出國進修統計科學)學生的主要來源為數學或應用數學系及統計系的學生。因統計非數學系的主流課程，要吸引數學系的學生在眾多不同的方向中選擇統計為其志業，需要提供機會增進他們對統計的接觸與了解。因此雖然各統計研究所的教學對象以研究生為主，亦應積極開設大學部的統計與機率課程。而國內各校的統計系隸屬管理學院或商學院，為配合學生出路及提升學生學習動機，須與他系合作跨領域學程，如精算學程、計量財務學程、行銷學程，其優點為擴大統計之影響及使其他領域用錯統計方法之機會降低，但缺點為優秀之學生有第二專長後，可能會轉入其他領域 ，因此亦存在吸引他們留在統計領域繼續深造的問題。此外一般統計系學生數學與理論方法的基礎較為欠缺。針對這些，可鼓勵老師帶大學部學生參與『大專生之專題計畫』，可適當引導學生進入統計方法之研究，並加強數理基礎, 提高學生繼續進修統計科學的意願。由於統計科學在跨領域研究扮演的角色日益重要，應鼓勵非數學與統計系的學生進入統計的領域，譬如開放統計副修給外系的學生。近年來中央研究院統計科學研究所利用暑假舉辦統計科學營的活動，敦請國內外知名學者對大學生(不拘科系)就統計科學的各個層面做深入淺出、生動的介紹，亦是有效的做法。已辦過的統計營以普及性的介紹為主，但亦可由各系推薦優秀的大學生，集中由老師帶領對實際問題做較深入的學習、探討與操作。2005 年 6 月在 UCLA 舉辦的 Summer Program in Statistics for Undergraduates 提供了一個可供參考的模式（參見 http://summer.stat.ucla.edu）。為配合近年來統計科學受各應用學科之衝擊所產生的變化與帶來的機會，大學統計及相關課程的重新規劃與設計也是刻不容緩的。

老師們對教學的重視與熱忱為人才培育重要的一環。關於學生的訓練，普遍的意見是必須加強論證與計算能力的訓練，因為不論統計研究或實務，都需要進行推理及論證，而計算在統計科學裡更扮演了非常重要的角色。目前國內各校的統計研究所尤其是博士班規模皆小，相當程度限制了學生選課及論文題材的選擇，影響到學生未來在學術上的發展。我們鼓勵跨校統計系所之間的合作，互補不足之處。在區域性可進行如目前清華與交通大學統計所博士班的合作，在全國性的層次可集各統計系所與中央研究院統計科學研究所之力開辦暑期課程。中研院統計所充足整齊的研究陣容可在國內統計人才的培育上發揮很大的力量。

不可否認目前出國留學進修統計科學的年輕人已比過去少很多，除了意願降低以外，獲得國外大學給予獎助學金也比過去困難。不過即使第一年自費，若表現不錯，自第二年起仍有獲得獎助的機會。因此我們希望除了增加全額公費留學的名額外，政府可提供部分(譬如一年的)資助供更多有志深造的年輕學生出國攻讀博士學位。此外加強學生英文的表達與寫作能力非常重要，老師們應灌輸學生

們這方面的認知。

對留在國內攻讀博士學位者,亦應提供充分的資源使他們能安心學習並有進修的機會。研究計劃中有關博士生的獎助學金名額應放寬,以免自己限制學門的發展;亦可另以獎學金的方式頒予優秀的研究生,而非僅附於研究計畫下。國科會近年有「千里馬」計劃,使得博士生有機會出國一年進修學習,這部份的名額應增加,讓更多的學生有機會接觸不同文化、統計研究的不同層面及國外一流的學者,拓展視野。我們並鼓勵研究計劃中「博士後研究」的申請,因博士一畢業就到學校教書或其他機構工作,往往教書或做計劃的壓力使得沒有充份時間作研究,「博士後」的訓練至少有一個充分的緩衝期可獨立發展研究能力。

捌、附錄

一、統計學門規劃修訂現況調查問卷報告

問卷調查寄給 44 個研究單位，這些單位均有研究人員執行 85 年度國科會自然處之計畫，經過催繳，問卷全部回收。以下就問卷調查資料統計如下：

1. 統計科學專任師資/研究人員在 個學校(機構)(含 個單位)之分佈狀況。以下數字僅包括助理教授/助理研究員(含)以上。(見圖一)--???
2. 各校在學統計學門學生人數(若數學系或應數系中無分組者暫不計算)(見表一)
3. 94 年度自然處統計科學專題研究計畫各單位分布情形(見圖二、表二)
4. 近三年舉辦之統計研討會(見表六)

二、近年來統計學門專題研究計畫狀況

1. 87-94 年自然處與數學、統計學門專題計畫件數及經費統計表(見表三)
2. 87-94 年自然處統計學門專題計畫核定件數及經費(見圖三)

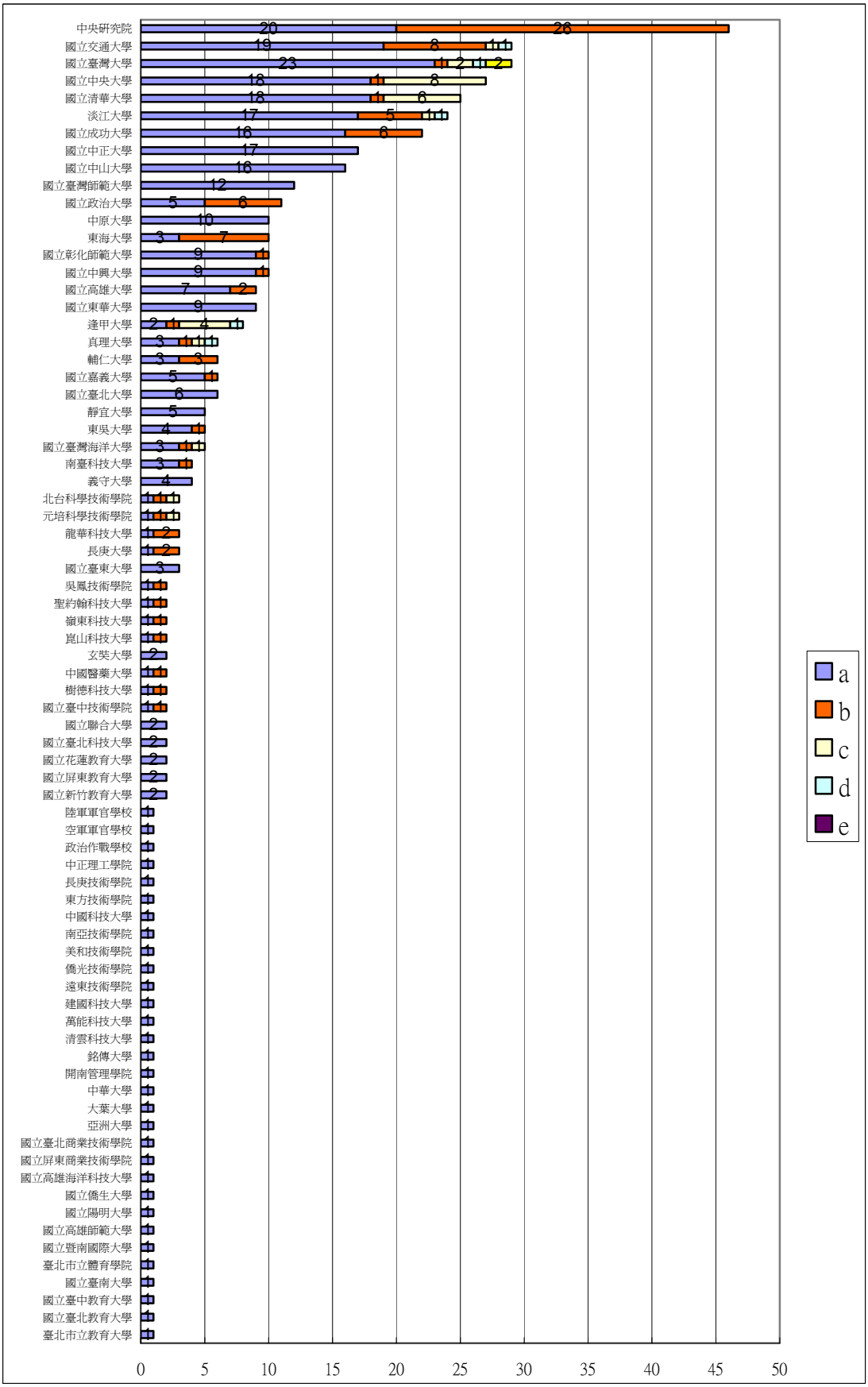
三、近年來統計學門研究具體成果

1. 1993 – 2005 年亞洲四國統計學門論文被 SCI 收錄篇數比較(表四)
2. 2000 – 2005 各統計系所發表論文狀況(表五)
3. 近三年發表於有審稿出版品之論文篇數(見表六)

表一、各校在學統計學門學生人數(若數學系或應數系中無分組者暫不計算)

系所名稱	大學部(在職)	碩士班(在職)	博士班
中原大學應用數學系		18	7
中華大學應用數學系		40	1
東海大學統計系	405	29	3
真理大學數理科學研究所		16	
真理大學數理統計與精算學系	166		
國立中山大學應用數學系	227	71	26
國立中央大學統計研究所		42	25
國立中正大學數學系、應數所、數統所		17	
國立中興大學應用數學系		41	4
國立台北大學統計學系	317(202)	47	
國立台灣大學流行病學研究所 生物醫學統計組		12	11
國立台灣大學數學系		17	3
國立台灣師範大學數學系		2	0
國立交通大學統計學研究所		48	22
國立成功大學統計系、所	211	54	15
國立政治大學統計學系	198	60	17
國立高雄大學統計學研究所		27	0
國立清華大學統計學研究所		49	19
國立嘉義大學應用數學系		3	
國立彰化師範大學統計資訊研究所		5	
淡江大學統計學系	908	33	2
淡江大學數學系、所	208	21	9
逢甲大學統計系、統計與精算研究所(碩士班)、應用統計研究所(博士班)		52	6
輔仁大學統計資訊學系	483		
輔仁大學應用統計研究所		44(59)	
銘傳大學應用統計資訊學系	437	25	
靜宜大學應用數學系	137		
總數	3691	808	161

圖二、94 年度自然處統計科學專題研究計畫各單位分佈情形 (總件數 443 件)



表二、94 年度自然處統計科學專題研究計劃各單位分佈情形 (總件數 443 件)

機關	件數	系所
中央研究院	46	數學研究所 (20)、統計科學研究所 (26)、
國立臺灣大學	29	數學系暨研究所 (23)、財務金融學系暨研究所 (1)、農藝學系暨研究所 (2) 、公共衛生學院公共衛生學系 (1)
國立交通大學	29	應用數學系(所) (19)、統計學研究所 (8)、管理科學系(所) (1)、財務金融研究所 (1)
國立中央大學	27	數學系 (18)、工業管理研究所 (1)、統計研究所 (8)
國立清華大學	25	數學系(所) (18)、計量財務金融學系 (1)、統計研究所 (6)
淡江大學	24	數學系 (17)、統計學系 (5)、經營決策學系 (1)、生命科學研究所 (1)
國立成功大學	22	數學系暨應用數學所 (16)、統計學系（所） (6)、
國立中正大學	17	數學系 (17)、
國立中山大學	16	應用數學系(所)、 (16)、
國立臺灣師範大學	12	數學系(所)、 (12)、
國立政治大學	11	應用數學學系 (5)、統計學系 (6)、
國立中興大學	10	應用數學系(所) (9)、資訊科學系(所) (1)、
國立彰化師範大學	10	數學系暨研究所 (9)、統計資訊研究所 (1)、
東海大學	10	數學系 (3)、統計學系 (7)、
中原大學	10	應用數學系 (10)、
國立東華大學	9	應用數學系暨研究所 (9)、
國立高雄大學	9	應用數學系 (7)、統計研究所 (2)、
逢甲大學	8	應用數學系 (2)、應用數學研究所 (1)、統計學系 (4)、統計與精算研究所 (1)
國立臺北大學	6	統計學系 (6)、
國立嘉義大學	6	應用數學系 (5)、農藝學系 (1)、
輔仁大學	6	數學系(所)、 (3)、統計資訊學系 (3)、
真理大學	6	數學系 (3)、數理科學研究所 (1)、會計學系 (1)、數理統計與精算學系 (1)
國立臺灣海洋大學	5	資訊工程學系暨研究所 (3)、河海工程學系暨研究所 (1)、生物科技研究所 (1)
東吳大學	5	數學系(所)、 (4)、商用數學系(所)、 (1)、
靜宜大學	5	應用數學系(所)、 (5)、
義守大學	4	應用數學系 (4)、
南臺科技大學	4	企業管理系暨研究所 (3)、通識教育中心 (1)、
國立臺東大學	3	數學教育學系 (3)、
長庚大學	3	公共衛生學科 (1)、通識教育中心 (2)、
龍華科技大學	3	機械工程系 (1)、資訊管理系 (2)、
元培科學技術學院	3	資訊工程系 (1)、訊管理系 (1)、財務金融系 (1)
北台科學技術學院	3	資訊管理系 (1)、國際貿易系 (1)、通識教育中心 (1)

機關	件數	系所
國立新竹教育大學	2	應用數學系 (2)、
國立屏東教育大學	2	數學教育學系(2)、
國立花蓮教育大學	2	數學教育學系(2)、
國立臺北科技大學	2	通識教育中心(2)、
國立聯合大學	2	通識教育中心 (2)、
國立臺中技術學院	2	企業管理科 (1)、銀行保險系 (1)、
樹德科技大學	2	資訊管理研究所 (1)、休閒事業管理系 (1)、
中國醫藥大學	2	公共衛生學系 (1)、通識教育中心 (1)、
玄奘大學	2	應用數學系 (2)、
崑山科技大學	2	資訊管理系 (1)、會計系 (1)、
嶺東科技大學	2	應用外語系 (1)、財務金融系 (1)、
聖約翰科技大學	2	國際貿易系 (1)、全人教育中心 (1)、
吳鳳技術學院	2	資訊管理系 (1)、國際企業管理系 (1)、
臺北市立教育大學	1	數學資訊教育學系(1)、
國立臺北教育大學	1	數學暨資訊教育學系(所)、 (1)、
國立臺中教育大學	1	數學教育學系(所)、 (1)、
國立臺南大學	1	數學教育學系 (1)、
臺北市立體育學院	1	通識教育中心 (1)、
國立暨南國際大學	1	國際企業學系(所)、 (1)、
國立高雄師範大學	1	數學系(所)、 (1)、
國立陽明大學	1	公共衛生研究所 (1)、
國立僑生大學	1	先修班 (1)、
國立高雄海洋科技大學	1	資訊管理系 (1)、
國立屏東商業技術學院	1	會計系〈科〉(1)、
國立臺北商業技術學院	1	資訊管理系〈科〉(1)、
亞洲大學	1	醫務管理學系(1)、
大葉大學	1	共同教學中心(1)、
中華大學	1	企業管理學系(1)、
開南管理學院	1	財務金融學系(1)、
銘傳大學	1	應用統計資訊學系(1)、
清雲科技大學	1	國際企業經營系所(1)、
萬能科技大學	1	工業管理系(1)、
建國科技大學	1	通識教育中心(1)、
遠東技術學院	1	資訊管理系(科)、 (1)、

機關	件數	系所
僑光技術學院	1	國際貿易系(1)、
美和技術學院	1	資訊科技系(1)、
南亞技術學院	1	應用外語系(1)、
中國科技大學	1	通識中心(1)、
東方技術學院	1	電子與資訊系(1)、
長庚技術學院	1	資訊管理系(1)、
中正理工學院	1	資訊科學系(1)、
政治作戰學校	1	心理學系(1)、
空軍軍官學校	1	教學部通識中心(1)、
陸軍軍官學校	1	管理科學系(1)、

表三、87-94 年自然處與數學、統計學門專題計畫件數及經費統計表

年度	件數				
	自然處	數學	百分比(數)	統計	百分比(統)
90	1633	263	16.11%	172	10.53%
91	1562	258	16.52%	152	9.73%
92	1693	274	16.18%	160	9.45%
93	1807	307	16.99%	166	9.19%
94	1873	285	15.22%	159	8.49%

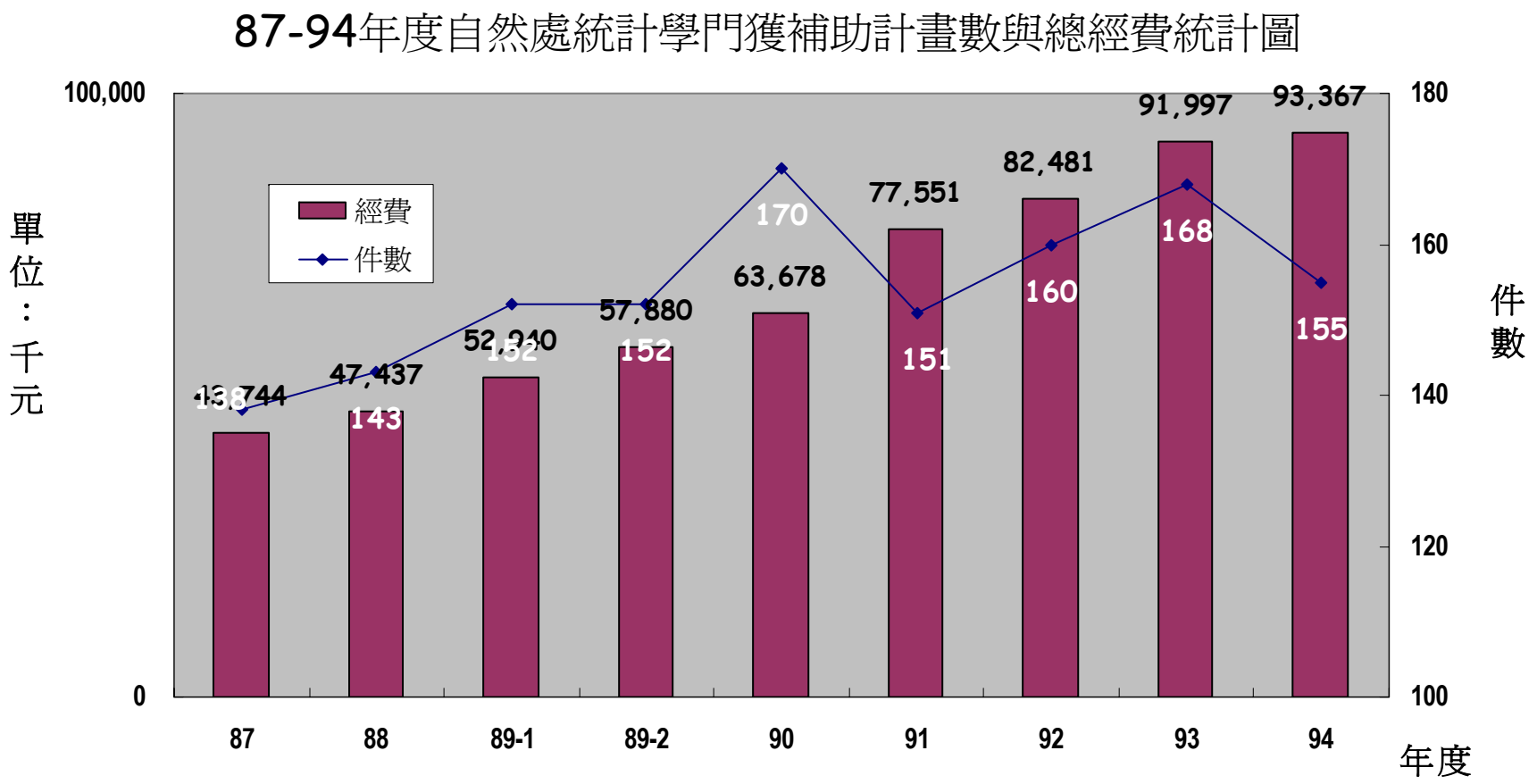
年度	總經費 (佰萬元)				
	自然處	數學	百分比(數)	統計	百分比(統)
90	1824	114	6.25%	65.65	3.60%
91	2174	144	6.62%	78.99	3.63%
92	2646	136	5.14%	82.84	3.13%
93	2768	154	5.56%	90.69	3.28%
94	3350	149	4.45%	94.2	2.81%

年度	計劃平均經費		
	自然處	數學	統計
90	111.7	43.35	38.17
91	139.18	55.81	51.97
92	156.29	49.64	51.55
93	153.18	50.16	54.63
94	178.86	52.28	59.25

表四、1993 – 2005 年亞洲四國統計學門論文被 SCI 收錄篇數比較

年度	台灣	南韓	大陸	日本
1993	54	35	113	113
1994	73	59	85	95
1995	75	55	95	93
1996	100	71	105	108
1997	107	74	125	127
1998	107	94	148	140
1999	146	90	151	135
2000	165	97	198	130
2001	162	104	265	124
2002	123	93	243	147
2003	104	88	246	145
2004	157	112	304	155
2005	148	118	327	177

圖三、87-94 年自然處統計學門專題計畫核定件數及經費



1.94年度之數據統計至94.10.20.止。

2.資料來源：國科會學術系統 PR81220程式。

表五、2000－2005 各統計系所發表論文狀況

年度	2000		2001		2002		2003		2004		2005	
單位	總篇數	SCI 數	總篇數	SCI 數	總篇數	SCI 數	總篇數	SCI 數	總篇數	SCI 數	總篇數	SCI 數
國立台灣大學流病所生統組	8	3	6	0	5	2	6	2	9	0	4	0
國立清華大學統計學研究所	22	9	14	8	17	12	25	8	18	7	10	5
國立交通大學統計學研究所	14		11		13		14		14		10	
國立成功大學統計系、所	27	11	16	10	23	3	23	0	18	6	21	7
國立政治大學統計學系	24	4	26	4	32	2	26	10	32	4	14	3
國立中央大學統計研究所	10		8		16		16		33		25	
國立中正大學數學系、應數所、數統所												
國立台北大學統計學系	24	0	30	17	17	6	30	9	20	5	14	3
國立高雄大學統計學研究所							2	0	2	1	2	1
國立彰化師範大學統計資訊研究所									1	0	1	1
東海大學統計系	13	6	11	4	8	1	8	6	9	6	7	6
淡江大學統計學系	15	5	13	6	16	10	14	7	25	14	21	12
逢甲大學統計系、統計與精算研究所、應用統計研究所	10	6	6	3	17	6	16	9	24	14	23	11
輔仁大學統計資訊學系	38	4	41	2	54	1	41	2	35	2	13	3
輔仁大學應用統計研究所	2	0	6	1	2	0	0	0	7	0	3	0
真理大學數理統計與精算學系	2		5		7		3		12		6	
銘傳大學應用統計資訊學系	18		17		12		30		41		40	
合計	227	48	210	55	239	43	254	53	300	59	214	52

表六、近三年舉辦之統計研討會

時　　間	名　　稱	性質	主　辦　單　位	經 費 來 源
95.4.3 – 95.4.7	分析和幾何學中的約當結構研討會 2006 (The third workshop on Jordan structures in analysis and geometry)	國際	國立中山大學應用數學系	國科會、數學中心、南區理論中心
94.9.29 - 94.10.3	International Conference on Nonlinear Analysis and Optimization with its Applications	國際	中原大學應用數學系	數學理論中心、中原大學
94.6.25 - 94.6.26	2005 中華機率統計學會年會及學術研討會、第十四屆南區統計研討會暨 2005 中華資料採礦協會年會	國際	國立成功大學統計系、所	財團法人張文豹文教基金會
94.10.21 - 94.10.24	第六屆台菲分析研討會 (6th Taiwan-Philippine Symposium on Analysis)	國際	國立中山大學應用數學系	國科會、南區理論中心、中山大學
93.8.23	2004 年真理大學數理科學國際學術研討會	國際	真理大學數理統計與精算學系	教育部 5 萬及自籌
93.6.15 - 93.6.17	2004 台北、上海、香港精算研討會	國際	東吳大學商用數學系	精算學會與各保險公司贊助
93.12.25 - 93.12.28	2004 年泛函分析研討會 (2004 NCTS Workshop on Functional Analysis)	國際	國立中山大學應用數學系	南區理論中心
92.12.15 - 92.12.16	理論統計暨應用統計之最新發展研討會	國際	淡江大學數學系、所	淡江大學
每年 3~4 次	人才培育計畫	國內	大同大學應用數學系	教育部顧問室

時　　間	名　　稱	性質	主　辦　單　位	經 費 來 源
94.8.5 – 94.8.6	2005 著色研討會-邊著色與圖的剖分	國內	國立中山大學應用數學系	南區理論中心
94.4.29	應用統計資訊研討會	國內	輔仁大學應用統計研究所	銘傳大學
94.1.13 - 94.1.15	第二屆統計與機器學習研討會	國內	靜宜大學應用數學系	數學中心,中研院統計所,靜宜大學
93.8.23	2005 量子演算法研討會	國內	國立中山大學應用數學系	南區理論中心
93.8.1 – 94.2.28	微分方程與反問題研討會	國內	國立中山大學應用數學系	南區理論中心、中山大學
93.7.24 - 93.7.26	著色研討會	國內	國立中山大學應用數學系	南區理論中心
93.7.2	統計研討會	國內	靜宜大學應用數學系	靜宜大學
93.7.2	From Data to Knowlede	國內	靜宜大學應用數學系	靜宜大學
93.6.19 - 93.6.20	第二屆台灣智慧科技與應用統計研討	國內	國立中正大學數學系、應數所、數統所	台灣智慧科技與應用統計學會、中正大學數學系、國科會數學研究推動中心
93.5.29	第一屆統計薪傳研討會	國內	輔仁大學統計資訊學系	數學推動中心、系預算

時　　間	名　　稱	性質	主　辦　單　位	經 費 來 源
93.5.28	第二屆保險與精算	國內	真理大學數理科學研究所	真理大學
93.3.20 - 93.3.21	應用機率統計研討會	國內	國立東華大學應用數學系	國科會數學推廣中心
93.2.5 – 93.2.6	統計機械學習研討會	國內	國立東華大學應用數學系	數學推廣中心、中研院
93.2.27	應用機率統計研討會	國內	國立東華大學應用數學系	國科會數學推廣中心
93.12.28	2004 清大統計論壇	國內	國立清華大學統計學研究所	國科會數學中心、清大理學院、清大統計所
93.11.13 - 93.11.14	93 年統計學術研討會	國內	國立中正大學數學系、應數所、數統所	中國統計學社、行政院主計處、中國主計協進會、中正大學數學系、中研院統計所
92.9.20 - 92.9.21	台灣智慧科技與應用統計研討會	國內	輔仁大學應用統計研究所	國科會數學中心；輔仁大學；台灣智慧科技與應用統計學會
92.6.4	統計研討會	國內	靜宜大學應用數學系	靜宜大學
92.6.26 - 92.6.27	第十二屆南區統計研討會暨九十二年度中華機率統計學會年會及學術研討會	國內	國立高雄大學統計學研究所	
92.3.17 - 92.5.12	2003 年南區組合數學系列研討會	國內	國立中山大學應用數學系	

時　　間	名　　稱	性質	主　辦　單　位	經 費 來 源
92.11.22	92 統計年會暨學術研討會	國內	輔仁大學統計資訊學系	中國統計學社、主計處、數學推動中心、系預算
92.11.21	第二屆統計薪傳研討會	國內	輔仁大學統計資訊學系	數學推動中心、系預算
92.1.1 - 92.12.31	2003 年分析及其應用研討會	國內	國立中山大學應用數學系	
91.6.27- 91.6.28	第十一屆南區統計研討會暨九十一年度中華機率統計學會學術研討會及年會	國內	國立中山大學應用數學系	
91.6.10 - 91.10.12	2002 年南區組合數學系列研討會	國內	國立中山大學應用數學系	
91.3.23 - 91.12.31	分析及其應用研討會	國內	國立中山大學應用數學系	
91.11.30	中國統計學社九十一年社員大會暨統計學術研討會	國內	東海大學統計系	中國統計學社、行政院主計處、中央研究院統計科學研究所、東海大學學術發展文教基金會
91.11.30	2002 年真理大學數理科學學術研討會	國內	真理大學數理統計與精算學系	自籌
91.1.12	第三屆全國管理學院統計系(所)學術暨教育行政研討會	國內	東海大學統計系	東海大學學術發展文教基金會
90.9.15 - 90.12.31	泛函分析研討會	國內	國立中山大學應用數學系	