

實驗設計淺談

羅夢娜

國立中山大學應用數學系

一家專門做西裝褲的服裝製造公司，想要比較四種不同布料，如麻紗、棉質、絲質和毛料做出來的西裝褲中，那一種布料做的最耐穿？於是每種布料做 10 條西裝褲，提供給 40 位志願試穿的人各穿六個月，試穿期間每週穿 4 天。然後再拿回來比較褲子破損的情形。但這裡面有一個問題是，即使同種布料做的西裝褲，給不同的人試穿，其破損的程度都不盡相同，何況不同種布料做的呢？換句話說，我們如何分辨那些破損是由於人為的因素？那些是因為布料本身的耐磨性？還是一些其他因素的影響？

這類比較實驗 (comparative experiments) 在日常生活中到處可見。如幾種不同的肥料，那種能使稻米產量增加最多；幾種不同的感冒藥，何者療效最佳？幾種不同的教學方式，那種學習效果最好？通常每做完一個實驗，都會有些數據，這些數據經過分析後會得到一些結論，但這些結論究竟有多少可信度？我們是否可以安全地引用它？這類實驗，若事前沒有經過周詳的考慮及計畫，即使做完後，所得的結果可

能會與事實有極大的偏差。

費雪及實驗設計

很早以前，人們即知道實驗需要好好設計，只是早期很少有人研究它。現代有關實驗設計的一些基本想法，主要來自於著名的英國統計學家費雪 (R.A. Fisher)。費雪是個興趣廣泛，天才橫溢的人。他在高中時即因數學資優獲得獎學金進入劍橋大學，並按照當時英國傳統，研讀自然哲學，主修物理。他在大學時代即從研讀當時著名的統計學家，皮爾森 (K. Pearson) 所著「數學對進化理論之貢獻」(Mathematical Contributions to the Theory of Evolution) 一書，而注意到數學在生物上的應用。他從 1919 ~ 1933 年，在坐落於英國賀福雪郡的賀平登 (Harpenden, Hertfordshire) 一個農業研究所，若森史特實驗工作站 (Rothamsted Experimental Station) 工作。在那良好的研究環境中，他與一羣生物學家一起工作。他對生物的廣泛興趣，使得他能了解他的同僚們所提出的問題，並能提供他們許多有關統計方面的建設性看法；同時在那段期間中發展出許多用在農場上田間試驗的實驗設計。他最有名的一本書「研究用的統計方法」(Statistical Methods for Research Workers, 初版發行於 1925 年)，曾經出版到第十四版，並翻譯成多國文字。他不僅對統計學的理論及應用有深遠的影響，對遺傳學亦有相當貢獻。他對實驗設計的主要看法有以下數點：

一 在比較不同處理 (treatments) 的效果時，實驗應能對受驗者所產生的反應值 (responses) 的變異情形，提供有意義的估計。

二 應利用隨機化 (randomization) 方式進行實驗，以便得到有關變異性的估計。

三 應利用區集法 (blocking) 來消去已知存在的外來變異的影響。

四 應根據實驗進行的方式，來決定當用何種統計方法分析實驗的結果。

五 建立複因子實驗 (factorial experiments)，如此我們可在一單獨而複雜的實驗中，同時檢驗不同因子的效應，而不需對每一因子分別做實驗。

前面所提到的這些名詞，如處理、隨機化、區集

、複因子實驗等，我們在後面會有進一步的說明。

名詞之演進

早期令統計學家感興趣的許多實驗多半來自於農業上的研究。從實驗設計的許多名詞裡，我們仍可看到因此而留下的痕跡。例如用來接受「處理」的實驗單位，我們稱為一個「區」(plot)。一個區在它的原意可能是某種農作物生長的某塊地，但現在則可能為一醫院中的病人，一片動物的組織，某種動物接受注射的身體部位，或是一批同型機器中的某一台，或是某位接受調查的人等。實驗的目的是要比較一些不同處理的效果；因此對於每一種處理，可能會有一個或數個區接受此處理，然後我們再根據這些由不同區上所得的觀察值來比較。而由一區上所得數量化的觀察值，常被稱為其產量(yield)。在實驗設計的結構中，若有一組區具備某些共同的特性，則稱此組區構成一個區集。

有時我們想測試的處理為由好幾個因子所組成；如我們想知道一生產過程中的最佳條件為何？而其中牽涉到 24 種處理方式。這 24 種處理是由 3 種溫度，4 種使用材料以及 2 種加熱時間，總共有 3 個因子(溫度、材料使用量、加熱時間)所組成。每一因子的不同狀態(如溫度為 20°C、30°C 及 40°C)，稱為該因子的水準(levels)。如上所述溫度有 3 個水準，材料使用量有 4 個，加熱時間有 2 個。這類實驗通常採用複因子實驗方式進行。

隨機化及區集

實驗設計裡的一個主要問題，是實驗單位在接受處理後所產生的反應，不僅僅受處理的影響，往往亦受許多其他外來因素的影響。如本文一開始所提到西裝褲破損程度與試穿的人的習慣也有關係，便是一種外來因素。如果採用隨機化及區集方式來做實驗，只要設計得當，可以消除許多外來的偏差，而比較出因處理不同所產生的差異。

隨機化

採用隨機化方式做實驗包含了兩方面：一為隨機選擇實驗用的「區」；二為隨機指定那些區應該接受

何種處理以及實驗進行之次序。有些簡單方法如丟銅板、抽籤，或使用亂數表等，詳細過程我們在此不多提。更重要的問題是：究竟有無必要採取隨機化？

首先如前所述，以隨機化方式進行實驗可消除外來的偏差。因為無論實驗者多麼誠實，當他在實驗限制範圍內有自由選擇的權利時，非常容易受所選擇「區」的影響，而對處理產生偏見。當然任何太明顯偏愛某種處理的情形，我們會發現而摒棄，但因下意識的偏愛而造成實驗結果不正確的例子到處可見。

其次，實驗者除了要使自己相信實驗結果，還要設法使其他人相信他是在非常公正的情況下得到結果，使用隨機化的方式進行實驗可達到公正、公開的目的。在著名統計學家寇克恆及寇克斯(Cochran & Cox)二位所寫的書中，建議將「隨機化」看成一種保險。雖然許多實驗無論是否採取「隨機化」可能結果都差不多，問題是我們不知道那些不採取隨機化不會出錯，那些卻會出錯。如果某實驗者就是不喜歡採用「隨機化」，那有時他可能很幸運沒有問題，有時則可能出差錯而導入歧途。

一種最簡單的完全隨機化設計(completely randomized design, CRD)，即那些「區」應接受何種處理是完全隨機化決定。如前述有 24 個處理的例子，若我們已隨機地選出 24 個區可做實驗，則每一處理恰有一區對應。我們將處理及區分別由 1 排至 24，再做 24 個籤，上面有 1 ~ 24 的號碼，然後抽籤；第一個抽中的號碼若為 3，則第 1 區接受第 3 種處理，依此類推，以抽籤方式決定處理及區的對應關係。因此法將所有區及處理都拿出來一起排列，所以稱為「完全隨機化」。當然我們亦可用亂數表取代抽籤方式。

區集

除了實驗應採取隨機化外，還有那些值得我們注意的事？做過實驗的人都能體會到區的條件相當一致的好處。但區很不容易條件一致。例如在前面的例子中，若找到 40 位試穿西裝褲的人當中，有 8 位是農夫，8 位是坐辦公桌的職員，12 位是工廠的工人，8 位是統計學家等等，由於他們工作性質不同，就造成「區」的條件不一致的情形。而若在完全隨機化地指定他們試穿的西裝褲種類時，恰好產生穿麻紗料的有 8 位都是農夫，穿棉質料的有 7 位是職員，穿絲質

料的有 9 位是工廠的工人，穿毛料的有 8 位是統計學家的情形。我們可見此種實驗對那些料子做的西裝褲較有利。雖然經過完全隨機化方式安排，此種「區」具有強烈不均勻分布的極端情形不常出現，我們仍應設法避免這類可能出現的設計。實驗設計許多理論論都針對著區與區間條件差異相當大的情形而探討。完全隨機化設計是一種可能的設計方式，但往往不是最好的。

一種可能的改進方式，即限制試穿者屬於某一行業，但這樣使結果適用的範圍受限制。所以我們還可以另一方式，即先將試穿者按所從事之行業分組，再採取有限制的隨機化方式，使每一布料試穿的人當中都有兩位農夫，三位工廠的工人等等。如此在比較布料耐磨性時，可將因工作性質不同對布料的影響平衡一下，且能適用較大範圍。對於這種有計畫性的隨機化實驗，最後在做統計分析時，要注意使用正確的分析方法，方能真正比較出處理效果的不同。

隨機化區集設計

下面我們介紹最常使用的一種實驗設計，可限制隨機化方式，即隨機化區集設計 (randomized block design)。一方面是由於它適用於很多種情況，一方面是因它很容易修訂。使用區集是增加實驗精確度的一種技巧。如前所述，我們希望區的條件越一致越好，但很不易做到，所以將條件相當一致的區聚在一起成為一個區集。如此將所有區分成好幾個區集，再隨機化進行實驗。

隨機化區集設計即在比較 t 種處理時，每一區集中恰有 t 個區，並在每一區集內將 t 種處理完全隨機化地分配於區，各做一次。所以若一隨機化區集設計中有 r 個區集，總共有 $N = rt$ 個區，就好像是由 r

表：一隨機化區集設計 (四種處理，三個區集)。

區集內之區	區 集		
	I	II	III
1	D	A	D
2	C	C	B
3	A	B	C
4	B	D	A

註：四種布料分別以 A、B、C、D 代表。

個完全隨機化設計所組成。例如在比較四種不同布料的西裝褲時，若我們選擇三種不同行業 (I、II、III) 的人來試穿，每一行業中隨機地挑選四位，再經過完全隨機化方式安排，一種可能的隨機化區集設計如附表。

只要我們事前按照隨機化區集原則安排好次序，再照著進行，就是一隨機化區集實驗。但還要注意的是只有在引進一區集劃分後，能使區集內區與區間的變異比所有區擺在一起的變異小很多時，使用區集方式才會真正有益。

通常農業的田間試驗中，鄰近的地區大致上土壤肥沃程度、氣候等都是一致的，所以可將鄰近的土地當作一區集。此外例如同胎生的動物（做營養試驗）、同一輛車子上的輪胎（比較輪胎的耐久力）、同一工廠內的相同機器（比較不同作業標準的實施）、同一天內所做的零件（比較機器的性能）等等，這些區都是適合聚在一起構成區集的好例子。

隨機化區集設計無論 r 或 t 為何，都相當容易建立，它本身的結構沒有特別的學問，但它卻是許多其他複雜設計的基礎。有時一系統中，同時引入兩個或更多區集因子可使結果更精確。除了隨機化及區集二原則外，重複實驗 (replication) 亦相當重要，因為它能提供有關實驗誤差的估計值，並能增加實驗的精確度。只是重複實驗有時基於經濟上的考慮，無法做到。在有些情況下，已有新的統計方法可分析沒有重複實驗的數據。

複因子實驗設計

在一個實驗中，同時檢驗好幾個不同因子的複因子實驗有許多好處並且實用。無論一研究者在做理論或應用研究時，他多半會同時對好幾個因子感興趣。例如，一生理學家想知道人體在不同溫度下工作的反應，他可能以所穿著衣服的質料及多寡代表人體所處的不同溫度，「工作」則為在一定時間內執行某些任務（如爬樓梯、跳繩，或騎腳踏車等）。那麼他的反應可能為脈搏、體溫、流汗程度，或一些由主觀決定的疲勞程度中之任一種。他可能不僅對比較穿不同衣服時的平均反應有興趣，還會想知道這些反應與其他因子，諸如年齡、性別、室溫、濕度及所執行任務間

之關係。很明顯地他需要事前計劃，將有許多不同因子，每個因子有好幾個不同水準，形成各種組合的狀況考慮進去，再做實驗。如果他一次實驗只探討一種因子的反應，那想要將許多因子各研究一次，不但要花許多時間精力，且分別所得到各因子對反應最有利條件的組合，未必即為最佳組合。

任何一實驗，其處理為由不同因子的不同水準組合形成時，即為一複因子實驗，其進行方式可以完全隨機化，或隨機化區集等設計來決定。當然若採用複因子設計，即使只有少數因子的各種組合亦會相當多，有時基於經濟或實際困難，可先將每一因子各取二水準，做一次探討性的研究，先找出影響最大的幾個主要因子，再深入詳細研究；甚至可採取所謂部分複因子設計 (fractional factorial design)。

曾經有人因為對騎自行車很有興趣，想知道應怎樣調整他的自行車，可使他騎車爬上一山坡的時間最省？可能有影響的因子有座位的高低、手把的位置、齒輪的速度、輪胎的軟硬等；他甚至把有沒有吃早餐、是否穿雨衣，都當作因子一起考慮。即使每個因子都只取兩個水準，全部都作一次要作 $2^6 = 64$ 次，而請人幫忙又不容易且會產生其他問題。所以決定採用部分複因子設計，只做全部的 $\frac{1}{8}$ 共 8 次；再重複一次，變成總共 16 次，以便能夠估計誤差值。他將得到的數據，經過統計分析，發現座位調到「高」的位置，輪胎的氣要足及齒輪速度要快些，可節省上山的時間。事實上，只要我們有興趣，日常生活中到處都是做實驗的好題材。媽媽做蛋糕應用那種麵粉（高筋、低筋）、糖應放多少湯匙、烘烤時溫度的高低、時間的長短等等，都可能成為有趣的研究項目。

遠景看好

近二十年來，日本的許多產品如汽車、彩色電視機、音響、照相機、輪胎、機器人、電子或機械工具等等吸引人的高科技產品，都在世界市場上占有一席之地。一向以世界科技領導地位自居的美國，也感受到日本強大的競爭，而開始檢討日本成功的因素。其中一項主要因素即日本工業界引用了大量的統計方法，而且並不只是簡單的統計方法。美國統計學家寇克斯估計，目前在日本工廠中一年內，大約做了超過

100 萬個經過設計的大小實驗，並且他估計這數目比西方國家從過去到現在所做過的還要多。一位美國電話電報公司貝爾實驗室的高級主管高福瑞 (A. B. Godfrey)，在一次訪問日本富士全錄公司時，發現當天他們所做的實驗中，有的包含高達 160 個因子。

實驗設計不僅可應用在農業上，在工業上及經濟上亦有廣泛的應用，其重要性自不在話下。本文所介紹為實驗設計中基本但卻重要的觀念及方法。許多較複雜而有用的設計，包括由日本田口玄一 (Taguchi) 所倡導使用且有相當成效的直交表實驗計畫法，都是由這些方法所衍生出來。此外在實驗設計中，根據設計而能做正確的統計分析，對實驗結果的可靠性亦扮演極重要的角色。限於篇幅，在此不多提，有興趣讀者可參考一般實驗設計的書。

參考資料

- 1 G.E.P. Box, W.G. Hunter and J.S. Hunter, *Statistics for Experimenters*, Wiley, New York, 1978.
- 2 W.G. Cochran and G.M. Cox, *Experimental Designs*, 2nd edition, Wiley, New York, 1957.
- 3 D.J. Finney, *An Introduction to the Theory of Experimental Design*, The University of Chicago Press, Chicago & London, 1960.
- 4 R.A. Fisher, *Statistical Methods for Research Workers*, 14th edition, Hafner Press, New York, 1970.
- 5 A.B. Godfrey, "Training & education in quality & reliability — a modern approach", *Commun. Statist. — THEOR. METH.*, 14(11): 2621 ~ 2638, 1985.
- 6 W.H. Kruskal and J.M. Tanur edited, *International Encyclopedia of Statistics*, Volume 1, The Free Press, New York, 1978.