

第 4 章

實驗設計 (*Experimental Design*)



4.1 實驗設計的概念

4.1.1 實驗設計中的變項

4.1.2 實驗組與控制組

4.1.3 兩個以上的獨變項

4.1.4 兩個以上的依變項

4.2 不同的實驗設計

4.2.1 前實驗設計

4.2.2 真實驗設計

4.2.3 準實驗設計

4.3 實驗受試的選擇與分派

4.3.1 受試者的選取

4.3.2 受試者的分派

4.4 實驗設計的計量原理

4.4.1 獨變項水準數與統計分析

4.4.2 變項數目與統計分析

4.5 實驗設計的混淆與干擾

4.6 結語

實驗研究可以說是科學領域最重要的一種研究方法。尤其當心理學邁入科學心理學的時代之後，實驗設計即成為心理學研究的主流方法。事實上，實驗法不僅是心理學的主要研究取向，同時也是其他學科研究者用來探討某特定方法的效果或是進行不同操作歷程比較重要的研究策略。

所謂**實驗**（experiment），顧名思義是指實地的探索與檢驗，或稱為**嘗試**（trial）。一個實驗是由一個單一的嘗試或是一連串的嘗試所組成，目的在**確認**（confirmation）某些研究者所提出的問題，例如找出最好的工作方式，或是去**解釋**（exploration）某些現實世界中事物的關係與各種現象，例如探討是什麼原因造成飛機駕駛員的疏忽。一個公正客觀的實驗程序所得到的結果，可以充實或修正人類世界既有的知識與經驗，協助人們去解決問題或進行決策。更具體的來說，實驗是藉由研究者所進行的科學性的嘗試，來找到人類事物的因果關係的解釋證據（Cook & Campbell, 1979），也就是一套科學的「採取行動」與「觀察結果」的客觀程序（Babbie, 2004）。

所謂「採取行動」，是指一個實驗若要能夠有效率的進行，必須經過縝密的規劃，發展出一套客觀的程序，並在嚴謹的條件下來執行研究者所關心的作用因素（獨變項）。而在行動過程中，研究者必須適時的就行動產生的作用進行現象的觀察與記錄，留下研究數據，一旦行動順利完成之後，即進行結果的分析與詮釋。在此一過程中，心理計量的概念與方法，扮演著一個相當重要的角色。尤其從一個科學發現的觀點來看，實驗設計是否符合**統計設計**（statistical design of experiments）的要求，是實驗研究能否成功、產生意義的基本要件（Montgomery, 1984）。這裡所指的統計的角色，並不僅是單純的哪一種分析方法的選擇，而是包括了數據的獲得、觀察紀錄的方法、數據品質的評估、結果的分析與關係的解釋等一連串計量的過程。因此，實驗設計必須與**統計的方法學**（statistical methodology）或是心理計量學的知識內涵一併考量，如此才能協助研究者找到有意義的結論。

4.1 實驗設計的概念

實驗的進行端賴一套嚴謹的實驗設計，而**實驗設計**（experimental design）是指一套將受試者安排入實驗情境與進行統計分析的計劃（Kirk, 1995）。更具體來說，實驗設計是由一套用以檢驗科學假設的活動所組成，這些活動包括**統計假設**（statistical hypothesis）的建立、實驗的情境與條件的設定（獨變項的決定）、測量以及實驗控制的方式的決定（依變項與控制變項的決定）、受試者的選取條件的設計（抽樣設計）、以及統計分析方式的決定等步驟。從這些具體的操作步驟中，我們可以看出實驗設計與統計分析當中具有相當緊密的關係，甚至在一般的實驗設計教科書中，絕大部分的篇幅是在探討不同的實驗設計與統計分析原理間的關聯性（eg, Kirk, 1995; Lindman, 1992）。甚至可以說，實驗設計就是一門以統計觀念為核心的研究方法學。

基本上，為了維持科學研究的客觀性，一個實驗必須具備三個基本的原則：**可複製性**（replication）、**隨機性**（randomization）、**區組性**（blocking）。可複製性表示一個實驗可以在相同的條件下，被重複操作獲得相同的結果，即使不是完全的相同，所存在的差異（即實驗的誤差），也必須在一定合理的範圍內。其次，隨機化是實驗設計能夠符合統計理論的重要程序。藉由隨機化，我們可以確保實驗的進行是在一個客觀的基礎上，不同的嘗試之間，除了研究者的實驗操弄之外，並沒有特定因素影響我們所關心的效果變動。最後，所謂區組性是實驗當中用來增加客觀精確性的技術。一個區組指的是實驗材料當中同質的一部份，區組化可以協助研究者分離不同的操作程序，以便進行比較。

上述三個實驗設計的基本特性，一方面說明實驗研究所關心的核心問題，同時也構成了實驗設計當中，不同的統計分析方法的差異所在，例如實驗誤差的估計、隨機化的統計分配原理、區組設計的分析方式等等。正也說明了一個成功的實驗，反應在一套嚴謹的測量－統計程序當中。

4.1.1 實驗設計中的變項

獨變項（independent variable；簡稱 IV）或**自變項**，是研究者所操弄的變項，也就是實驗差異來源之所在，例如燈光的明亮度、噪音的大小等等。**依變項**（dependent variable；簡稱 DV）則是實驗當中被觀察、記錄的變項，例如員工的生產力、人們對於刺激的反應速度、駕駛的注意力等等。通常一個研究的依變項，就是研究者最關心的概念或現象之所在，而獨變項則是造成該依變項變化的來源。例如酒後駕車容易造成車禍，如果要證明酒精對於人們駕車的影響，我們就必須設計一個實驗，看看喝了多少杯酒會造成駕駛注意力的遲緩或喪失。此時，注意力的遲緩就是依變項，而造成注意力遲緩的原因是喝酒的量的多寡，喝酒的量也就是獨變項。

某一個變項會被納入一個實驗中作為獨變項，主要是因為研究者認為該變項是影響依變項的主要因素而被研究者選用。而且一個實驗當中，獨變項的強度或數量大小，可以由實驗者決定，因此又被稱為自變項或「因」變項，表示該變項的選擇與變化是由實驗以外的因素決定，而非內部的因素來決定。而前面提及的注意力遲緩之所以被稱為「依」變項，則是因為此一變項的數量變動，是由獨變項所造成，亦即由實驗的內部因素來決定，因此又被稱為「果」變項。

依變項的變化如果要完全歸因於獨變項的影響結果，有一個很重要的先決條件，也就是當我們在操弄獨變項的不同強度下去觀察依變項的數量的變動時，其他的條件必須被控制在一定的恆定狀況下，如此一來依變項的數量的高低變動，便不能歸因於其他的因素。例如每一個接受喝酒實驗的人，事前都沒有吃過其他食物，如果某些受試者喝酒前吃了東西，有些則沒有，再去測量他們的注意力，我們便不敢篤定的說，注意力的變化一定是由於他們喝了多少量的酒。另外一個可能的情形是，除了喝酒的量不同以外，每一個人喝的酒的種類也不同，如此一來，造成注意力變化的因素除了酒量還有酒的種類，因此我們便不敢斷言注意力的變化是因為酒量的因素來影響。

要維持實驗的客觀性，研究者在進行實驗操作時，都會儘可能的使實驗的環境保持一致，除了獨變項與依變項，其他的變項都應維持一致，但是有些因素則無法維持一致，例如喝酒實驗當中，參加實驗的人距離前一餐飯的時間可能都不一樣，因此飯後多久的效果無法被實驗者在事前進行控制，只好將每一個受試者距離吃完飯的時間的長短加以記錄，以便分析時將它納入分析，此時，離開前一餐的時間就是實驗當中特別用來進行控制的變項，稱為**控制變項**（control variable）。

4.1.2 實驗組與控制組

一個實驗的操作，除了牽涉到獨變項、依變項、與控制變項的決定之外，另一個區分是依據受測者本身的狀況的不同來看，而分為**實驗組**（experimental group）與**控制組**（control group）。被視為是實驗組的受測者必定是接受了實驗操弄的影響，而在獨變項上屬於不同狀況（不同的獨變項水準）。而控制組則是指沒有接受實驗操弄的影響，他們在獨變項上應是處於最基本的一種狀況。

例如圖 4.1 研究者想看看喝一杯啤酒、五杯啤酒和十杯啤酒的受試者，他們注意力遲緩的程度，因此找了三群受測者，分別喝下一杯啤酒、五杯啤酒和十杯啤酒，這三群人，就是三個接受不同實驗操弄的實驗組，他們分別屬於獨變項的三個水準：微量（一杯）、中量（五杯）、大量（十杯）。另外，為了看出喝酒的影響，我們還需要一組受測者完全沒有喝啤酒（無量），總共四群受測者（四個區組），來比較他們注意力的強弱。此時獨變項是喝酒的量（杯），總共有四個水準：無量、微量、中量、大量，其中第一個水準是為控制組，其他三個水準屬於不同實驗操弄組。實驗的分析，就是比較這四組受測者在注意力上的高低。

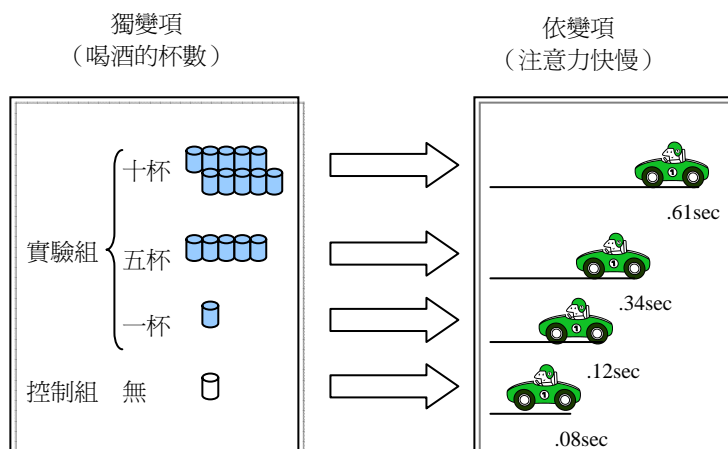


圖 4.1 實驗設計的概念圖

4.1.3 兩個以上的獨變項

■多因子設計的特性

當一個實驗不只關心一種實驗操弄對於依變項的影響，而是同時處理多種實驗的操弄時，稱為**多重因子設計**（factorial design），也就是一個研究中包含了多重的獨立變項。此時不僅實驗的程序增添了相當的複雜性，在計量與分析的層次上也必須以不同的方法來處理。例如以前述酒精對於駕駛注意力的影響的例子，如果研究者想要探討喝酒的量（幾杯）與酒的種類兩種不同的因素對於駕駛注意力的影響，就是一個典型的二因子實驗設計。第一個獨變項（A 因子）是喝酒的量，第二個獨變項（B 因子）是酒的種類，這兩個因素被研究者認為是造成注意力變差的原因，因此被納入實驗研究當中。

■多因子設計的優點

一般而言，實驗研究通常不只有一個獨立變項，而可能有二到多個獨立變項同時作用在依變項上。使用這種多因子設計有下列幾個好處，第一，同時處理多個獨變項可以提高研究的效能。如果研究者認為有三個獨變項（三種不同的實驗處理）對於依變項都有相當的影響，執行三次實驗，每一次只處理一個獨變項的效率遠不如一次處理三個變項。

第二，多個獨變項的實驗研究可以提昇研究的實驗控制效果。當我們在觀察依變項的變化時，因為同時存在不同的影響來源，無形中每一個獨變項都兼顧到其他獨變項的作用，避免混淆效果，進而提高研究結果的客觀與正確性。

第三，多個獨變項可以讓我們得到更豐富的研究發現。當只有一個獨變項（A 因子）時，依變項的變化只能歸因於該獨變項的影響，但是當實驗設計當中安排兩個獨變項（A 與 B 因子）時，影響依變項的來源除了 A 因子與 B 因子的本身之外，還增加一項同時考慮 A 與 B 因子影響力的交互效果（interaction effect）。

例如，如果一個實驗設計包含了三個因子 A、B 與 C 因子，此時除了三個因子分別對於依變項具有的影響力（稱為主要效果）之外，交互效果可以再區分為二因子的交互效果（包括 A 與 B 因子、A 與 C 因子、B 與 C 因子的交互效果）以及 $A \times B \times C$ 三因子對於依變項的交互效果。當獨立因子越多，實驗處理越多，我們可以在依變項身上看到更多元的變化。

第四，多重的獨變項可以提高實驗研究的類化程度。因為一個實驗同時處理多個影響依變項的可能因素，對於依變項的觀察跨越了不同的情境或受試者，因此研究的結果可以推及更為廣泛的層面，對於人類行為的瞭解更加完整，所建立的理論知識與理論模型也就更加周延，提高實驗研究的外在效度。

4.1.4 兩個以上的依變項

■多變量設計的特性

依變項是一個研究最重要的部分，他反映了研究者所關心的行為內涵，也是一個實驗可以得到的因果結論當中的「果」的部分。研究者操弄再多的獨變項，實際觀察的還是依變項的變化情形。通常一個實驗可能包含多個獨變項，但僅處理一個單一的依變項，也就是利用單一的計量指標來反應依變項的變動，以瞭解不同的「因」如何來影響這一個「果」。但是有時依變項本身可能無法從單一的計量指標來反應依變項所欲反應的內容，必須使用多重的計量指標，或是研究者想要同時包含多種依變項的變化情形，如此一來在實驗進行之時，便須紀錄多重依變項的數據。當一個實驗設計中包含多重依變項時，必須使用更高階的統計分析方法，例如**多變量變異數分析**（MANOVA）或**多變量共變數分析**（MANCOVA），才能有效處理多個依變項所存在的特殊數學與統計的問題。

以先前喝酒對於駕駛注意力的影響的例子來看，駕駛的注意力是以單一的指標來處理（看到交通標誌燈號到車輛停止的時間），但是注意力也可以區分成不同的層面，各以不同的指標來測量，例如研究者認為喝酒後對於駕駛的視覺與運動神經反應的影響各有不同，因此對於注意力區分成視覺的部分與動作反應的部分，分別測量受試者看到交通標誌燈號的時間（Y1），以及採取煞車動作直到車輛停止的時間（Y2），共有兩個依變項。

■多變量設計的優劣

在實際的研究設計當中，使用多重獨變項的例子遠比使用多重依變項的研究來得多，主要的原因是多重依變項的處理使得研究結果的解釋困難。從計量的觀點來看，解釋上的難度包括概念上的困難與分析技術上的困難。一般而言，依變項比較容易牽涉到定義的問題，相

對於獨變項，依變項多屬於潛在的心理特質或是不可直接觀察的心理特質或屬性，需要倚賴操作型定義來限定測量的內容，測量時還有信度與效度的問題。相對之下，獨變項多屬實際事實性的變項，且由實驗者決定其內容與水準數，例如喝哪一種酒，喝的量是多還是少，是可以直接加以操作與觀察的變項，因此多重獨變項的使用遠比多重依變項的使用可以避免涉入心理計量的問題。

從計量分析技術的觀點來看，獨變項屬於類別變項，依變項多屬連續變項。多重獨變項僅是測量水準數目的增加，但是多重依變項則涉及複雜的共變關係的釐清，在計量分析上，多重依變項涉及的統計技術不僅從難度到複雜度都遠勝於多重獨變項的影響。更重要的是依變項之間的變化應如何解釋，與「因」之間的關係為何，也增加了研究者的困擾。如果說實驗研究的優點是透過嚴謹的實驗控制來獲得一個明確的因果關係的證據，那麼依變項之間的關係則相當可能造成因果關係的混淆，因為依變項彼此之間的影響不受研究者的操弄控制，某一個依變項可能受到其他依變項的影響，造成該依變項的結果無法歸因於獨變項的影響。

雖然在統計學上，已經針對多重依變項的實驗設計發展出完整的分析技術，但是使用統計工具這種統計控制的手段來解決實驗設計控制不足的問題，一方面是把實驗問題轉換成統計問題，使得研究者必須先去瞭解這些具有相當難度的統計方法的運用原理以及可能存在的分析問題，根本上也失去了使用實驗研究來探討人類行為因果現象的原始初衷。

到了科技發達的今天，雖然電腦可以代勞進行分析的工作，但是統計技術的難度也相對增加，實驗設計的混淆問題仍無法獲得根本的解決，因此除非萬不得已，或是依變項之間的複雜關係已有明確的理論概念或先前研究做了清楚的處理與說明，否則一般並不鼓勵採用多重依變項的實驗設計，即使研究中包含多個依變項要處理，也把每一個依變項獨立出來討論，進行幾次單一依變項的統計分析，以避免涉及先前所提到的依變項共變關係難以說明的尷尬處境。

4.2 不同的實驗設計

學者 Donald Campbell 與 Julian Stanley 在 1963 年發表了一本探討實驗設計的著作《實驗與準實驗研究》(*Experimental and Quasi-Experimental Design for Research*)，書中詳細介紹了十六種不同的實驗研究設計，並探討影響實驗研究效度的重要因素，成為討論實驗設計的經典著作。本節僅摘要該書中十種常見的實驗研究來加以介紹，這十種實驗設計的概念圖示請見表 4.1。

4.2.1 前實驗設計 (pre-experimental design)

Campbell 與 Stanley 首先介紹了三種常見、但是未能完全符合實驗設計要義的研究設計，稱之為前實驗設計，分別為表 4.1 中的前三種設計：單組後測設計 (one-shot case study)、單組前後測設計 (one-group pretest-posttest design)、與靜態組間比較 (static-group comparison)。這三種研究設計雖然未能符合實驗設計的嚴格要求，但是卻有簡單、易於實施的優點，常用於非正式、非專業的場合中，例如學校課程的評估、工商管理活動的評鑑、市場研究等等。

第一種單組後測設計，僅是在一群受試者身上施以某種實驗處理（例如接受某種自我成長課程訓練），然後測量他們的反應（例如自我肯定程度），第二種研究設計：單組前後測設計，僅是多出在事前先測一次反應（自我肯定前測），然後得以計算出前後測的改變量，這兩種方法的共同問題是缺乏比較的對照點，第一種設計僅能瞭解學員自我肯定程度，分數的高低無法反應課程的效果，第二種設計即使增加了前測，得到兩次自我肯定的分數差異，但是我們仍然沒有充分的信心證明這個差異是來自於課程的效果。

表 4.1：不同的實驗設計圖示

實驗設計名稱	實驗處理模式	實驗對照	前測控制	隨機分派
前實驗研究設計 (pre-experimental design)				
Type 1. 單組後測設計 one-shot case study	$X \rightarrow T_2$	×	×	×
Type 2. 單組前後測設計 one-group pretest-posttest design	$T_1 \rightarrow X \rightarrow T_2$	×	√	×
Type 3. 靜態組間比較 static-group comparison	E $X \rightarrow T_2$ C $\rightarrow T_2$	√	×	×
真實驗研究設計 (true-experimental design)				
Type 4. 隨機化實驗控制組前後測設計 randomized control-group pretest-posttest design	E _r $T_1 \rightarrow X \rightarrow T_2$ C _r $T_1 \rightarrow \rightarrow T_2$	√	√	√
Type 5. 隨機化實驗控制組後測設計 randomized control-group posttest design	E _r $X \rightarrow T_2$ C _r $\rightarrow T_2$	√	×	√
Type 6. 所羅門四組設計 Solomon four-group design	E _r $T_1 \rightarrow X \rightarrow T_2$ C _r $T_1 \rightarrow \rightarrow T_2$ E _r $X \rightarrow T_2$ C _r $\rightarrow T_2$	√	√	√
準實驗研究設計 (quasi-experimental design)				
Type 7. 非隨機實驗控制組前後測設計 non-randomized control-group pretest-posttest design	E $T_1 \rightarrow X \rightarrow T_2$ C $T_1 \rightarrow \rightarrow T_2$	√	√	×
Type 8. 對抗平衡設計 counterbalanced design	1 A B C D 2 B D A C 3 C A D B 4 D C B A	√	-	√
Type 9. 單組時間序列分析 one-group time-series	$T_1 T_2 T_3 T_4 X T_5 T_6 T_7 T_8$	×	√	×
Type 10. 實驗控制組時間序列分析 control-group time-series	$T_1 T_2 T_3 T_4 X T_5 T_6 T_7 T_8$ $T_1 T_2 T_3 T_4 T_5 T_6 T_7 T_8$	√	√	-

第三種設計：靜態組間比較，雖然比第一種設計增加了一個對照組，使得研究者可以比較有接受課程訓練的學員與沒有接受課程訓練的受試者，在自我肯定得分上的差異，然而實驗組與控制組這兩群樣本是如何獲得的，並沒有進行隨機化的分配，也沒有先施以前測來觀察實驗前後的改變量，因此，即使可以針對有無參加課程的結果進行比較，但是仍然缺乏一個客觀的基礎，兩組後測的差異比較，充滿了各種可能的干擾與混淆因素，例如這兩群人是背景十分不相似的人，我們也不知道後測得分高的人（自我肯定強的人），是不是在事前就已經比別人有較高的自我肯定？

這三種方法，不是缺乏實驗控制對照（第一與第二類設計），就是缺乏前測作為改變分數的基準參照（第一與第三類設計），用這種研究的數據來證明因果關係的存在，立論基礎過於薄弱，因此不被視為是正式的實驗研究設計。但是如果作為一個概況的瞭解，或是進行初探性質的研究，卻不失簡易便利，也能提供相當程度的資訊。

4.2.2 真實驗設計（true-experimental design）

上述三種研究設計雖有實驗設計的雛形，但是仍然有著相當明顯的缺憾，一個良好的實驗設計，也就是 Campbell 與 Stanley 所謂的真實驗設計，最重要的是要具備實驗組與控制組的對照、前測與後測的對照、以及實驗組與對照組的隨機分派三個要件，如果同時符合這三個要件，可以視為一個標準的實驗研究，但是如果符合兩者，仍可勉強接受屬於實驗設計，但是卻必須特別注意缺少第三個要件所可能付出的代價與錯誤。

表 4.1 中第四種設計，隨機化實驗控制組前後測設計（randomized control-group pretest-posttest design），同時具有上述三個要件，是最標準的實驗研究。以喝酒的量對於駕駛注意力的影響為例，我們找到一群受試者，將他們隨機分配到不同的組別（實驗組 Er 與控制組 Cr）

當中，然後在實驗之前測量他們的注意力（反應時間的快慢; T1），然後對於每一個實驗狀況施以適當的處理（給他們喝下不同份量的酒; X），然後再測量這些受試者的注意力（反應時間的快慢; T2）。如果在實驗之前，最後將四次測量資料進行統計分析，所得到的研究結果即可以作為飲酒對於駕駛注意力影響的因果關係的具體論證。

第四種設計完全滿足實驗設計的三個要件，是實驗方法當中最為典型的實驗設計。與第四種設計類似的為第五與第七種設計，但是他們各缺乏了一個條件，第七種設計因為缺乏隨機分派，因此被歸類為準實驗設計。至於第五種設計隨機化實驗控制組後測設計，雖然具有實驗與對照組，分組的進行也有隨機分派，但是缺乏前測的數據，使得實驗效果的後測得分缺乏前測的分數來進行參照調整，或是進行前後測改變分數的檢定分析，造成實驗的因果論證存在一個前測立足點不明確的威脅。然而，第五種設計的結果仍然相當程度被大家接受，主要是因為隨機化可以去除立足點不確定的威脅。一般而言，前測的存在是在獲得實驗效果的參照點，後測得分高者不一定是因為實驗效果強，而是因為前測分數就比較高，但是，如果實驗進行之前對於實驗組與控制組的隨機分派有確實達成隨機化的目的，那麼實驗前的立足點即趨於一致，而減少了個別差異的威脅。

■ 所羅門設計（Solomon four-group design）

所羅門（Solomon, 1949）發展了一種特殊的實驗設計，同時包含了第四種典型實驗設計以及第五種捨棄前測測量的實驗設計，除此之外，在操作程序以及實驗條件與原來的設計並無顯著的不同。此一設計的最大目的，在檢驗前測對於後測的練習與記憶效果，因為雖然前測分數可以作為實驗的統計控制，也可以用來作為實驗效果的改變分數的計算依據等優點，但是前測的使用卻使得受試者獲得先期練習後測測量的機會，後測的分數可能會受到前測分數的影響。

所羅門設計中，兩個實驗組的受試者在後測的得分上應該沒有顯著的差異，因為他們都經過了隨機分派與相同的實驗處理，但是第二

個實驗組並不受到前測 T_1 的影響，如果檢驗兩個實驗組的後測 T_2 分數，可以得到前測 (T_1) 的影響。除了兩個實驗組可以相互比較之外，兩個控制組也可以進行相似的比較，第一個控制組的 T_2 可能受到前測 T_1 測量的影響，但是第二個控制組的 T_2 則未受到前測分數的影響，兩個控制組的後測 T_2 分數的比較反應出前測的影響。

值得注意的是，兩個實驗組的 T_2 比較與兩個控制組的 T_2 比較，這兩者具有不同的意涵。兩個實驗組的 T_2 比較雖然反應了前測的影響但是此時前測的影響可能與實驗處理 X 產生交互作用的影響，因為兩個實驗組的 T_2 比較參雜了實驗效果；相對的，兩個控制組的 T_2 比較則未包含實驗效果的作用，因此，可以再去比較兩個實驗組的 T_2 平均值與兩個控制組的 T_2 平均值，所得到的即是前測 T_1 與實驗處理 X 交互作用 ($T_1 \times X$) 的影響。

所羅門設計雖然提供了更充分的檢驗證據，可以有效的提升研究的內在效度，但是需要更多的樣本，增加了研究人員的負擔與成本，所以並不常見於實際的研究中。但是所羅門設計所接觸的前測干擾問題值得研究人員特別注意。如果一個研究的前測的實施對於後測的影響非常明顯，所羅門設計可能就是必要的選擇，否則，研究者選用傳統的古典實驗設計（設計四）即相當足夠。

4.2.3 準實驗設計 (quasi-experimental design)

實驗研究的進行，最好的地點是在一個不受外界干擾、配備齊全的實驗室。在實驗室中，研究者可以專注於實驗的操弄，並將實驗過程控制在一定的條件下，避免外在因素的干擾，稱為**實驗室實驗法** (laboratory experiment)。然而，有些實驗並無法在實驗室進行，必須到實際的真實環境中操作，例如課程教學的研究，實驗處理（不同的教學課程）必須在教室中進行，這類實驗稱為**田野實驗法** (field experiment)。田野實驗最大的問題，是無法對於環境進行精密嚴謹的

控制，對於受試者也無法進行隨機選擇與分派，因此，不論研究者多麼努力，還是無法達到傳統實驗的隨機化的嚴格要求。

另外一種無法進行隨機處理的情況是實驗研究的獨變項無法操弄，必須採用測量的方法，透過抽樣程序來達成操弄的目的，例如智力高低、社經地位會影響學習成果，此時智力水準、社經地位作為獨變項，僅可以「取樣」，而無法「操弄」。這些無法操弄的獨變項，必須藉由測量與抽樣的手段來處理，伴隨而來的一些混淆因素，例如受試者是否具有某些獨特特質或個別差異，可能會影響實驗的進行。上述這些具有實驗研究的基本規格，但是缺乏嚴謹的實驗控制的實驗，尤其是沒有隨機化的處理的實驗研究，稱之為**準實驗研究設計**（quasi-experiment design）。表 4.1 中的研究設計七至十，共同的特徵是缺乏隨機化的處理，因此皆屬於準實驗研究設計。

準實驗設計與真實驗設計設計的差異，除了不同組別的受試者未使用隨機分派的程序來分發之外，其他均相同。以第七種設計是非隨機化實驗控制組前後測設計為例，此一設計與第四種古典實驗設計的實驗程序幾乎完全相同，僅缺乏了隨機化的處理。此類研究設計經常可見，因為在實際研究過程中，往往受限於某些因素，使得受試者無法進行隨機化分派，不同實驗組之間在實驗之前無法假設是等同的狀態，但是因為具有了前測，在實驗之前所存在的個別差異威脅，可以配合統計分析（例如共變數分析），將前測分數作為控制變項，排除其影響，因此所獲得的因果論證仍然具有相當程度的效力。

■ 拉丁方格設計（Latin square design）

當研究者採用受試者內設計，讓同一組受試者接受不同的實驗處理時，因為只有一群受試者，因此不同的實驗狀況之間並不需要進行樣本隨機分派處理，但是受試者的反應卻有可能受到實驗順序的影響，造成實驗效果的混淆。第八種實驗設計，即是以**對抗平衡**（counterbalancing）原理來進行處理實驗順序的問題，由於缺乏隨機化設計，因此也是屬於一種準實驗設計。

假設今天有 A、B、C、D 四種實驗狀況，如果採用完全對抗平衡設計，總計可以產生 $4!$ 種（24 種）不同的實驗順序組合，研究者要重複操弄 A、B、C、D 四種實驗處理 24 次，共計 96 次的實驗處理，相當耗費人力。若以圖 4.2 的拉丁方格來處理，24 種實驗順序被大幅簡化成四組程序，每一個實驗狀況至少一次會出現在另三種實驗條件之後，且每一個實驗狀況與前一個實驗狀況是固定的。對於某一個特定的實驗狀況而言，四組實驗設計代表四種痕跡效應（carryover effect），以 D 為例，組合一是 ABC 三種效果的痕跡效應，組合二是 BC 兩種效果的痕跡效應，組合三是 C 效果的痕跡效應，組合四是無實驗痕跡效應。對於 D 而言，其他三種實驗狀況的痕跡效果都被考慮進去了，但是三個實驗狀況的相互順序則不考慮變化，以簡化實驗操弄程序。在這種情況下，每一個受試者僅需接受一套實驗順序即可，可以減少受試者的負擔。

拉丁方格的使用上，受試者人數除了等於實驗設計的數目，也就是每一組實驗設計安排一個受試者（以本例子而言需要四位受試者），也可能是實驗設計數目的倍數（4、8、12...人），使每一組實驗設計有多個受試者，而每一個受試者仍然僅參與一組實驗設計。每一個實驗處理所累積的總人數越多，統計學的一些假設（例如常態性假設）就越能夠達到，研究者即可以進行一些統計檢定來檢驗拉丁方格的適切性（見 Kirk, 1995）。

表 4.2 拉丁方格設計

		實驗順序			
		1	2	3	4
實驗設計	1	A	B	C	D
	2	B	C	D	A
	3	C	D	A	B
	4	D	A	B	C

以表 4.2 為例，拉丁方格的四組實驗設計執行完畢之後，且每一組實驗設計累積了相當的受試者，研究者可以利用一般的單因子變異數分析來檢驗不同的四組實驗設計在依變數的得分上是否有無顯著的差異，如果發現有顯著的差異，代表某些組合可能有著特殊的痕跡效應的影響，該組合可能必須排除在研究的分析之外。此外，研究者也可以檢驗不同順位的得分的差異，檢驗不同的順位是否具有特定的得分情形，作為瞭解痕跡效應的補充資訊。如果這兩個檢定均達顯著差異，研究者可能必須另外設計一套拉丁方格，來除去痕跡效應的影響。

另一個類似的作法，是當研究者有額外的受試者時，可以另外設計一個拉丁方格，來安排給這些受試者使用。例如前四個受試者以表 4.2 的方格來進行實驗，另外四個受試者以表 4.3 的拉丁方格來進行實驗。表 4.3 中的第 2、3、4 三組實驗設計即是表 4.2 的拉丁方格中所沒有的順序組合。如此一來，可以涵蓋更多的可能實驗順序，但是每一組實驗設計的人數可能就無法繼續擴充。值得注意的是，表 4.3 的方格與表 4.2 的一個最大不同，是每一個實驗順序組合中，某一個特定的實驗狀況前面所跟隨的狀況都不一樣，又稱為平衡性拉丁方格（balanced Latin square）。此種設計較表 4.2 的拉丁方格更能夠平衡實驗順序的影響。

表 4.3 平衡性拉丁方格設計

		實驗順序			
		1	2	3	4
實驗設計	1	A	B	C	D
	2	B	D	A	C
	3	C	A	D	B
	4	D	C	B	A

■時間序列分析 (time-series analysis)

最後兩種設計，單組時間序列分析 (one-group time-series) 與實驗控制組時間序列分析 (control-group time-series)，明顯的與其他八種設計不同，主要是運用在縱貫研究 (longitudinal study) 上。時間序列設計 (time-series design) 的特色是對於某個測量指標進行週期性的追蹤測量，以了解該指標所反應的行為特質變動情形。如果在一個完整的時間序列中，插入一個實驗處理 X，再觀察整個時間趨勢的變化情形，又稱為中斷性時間序列設計 (interrupted time-series design)，如表 4.1 的最後兩種實驗設計所示。此種研究設計不但可以用以瞭解時間變動的趨勢，提供預測的功能，也可以用來檢驗實驗處理的影響，是一種實用性相當高的研究設計。

時間序列分析因為採用持續性的測量，因此必定使用受試者內設計，即同一組受試者必須接受多次的測量。表 4.1 的第九種設計中，前面四次測量 (T_1 到 T_4) 可以視為是實驗處理前的前測，後面四次測量 (T_5 到 T_8) 則是實驗後的後測。其中緊鄰實驗處理 X 的前後兩次測量 T_4 與 T_5 ，則可以視為一般橫斷研究的前測與後測，其他的六個測量 $T_1T_2T_3$ 與 $T_6T_7T_8$ 則是為了建立時間發展序列的模式而進行的測量。利用迴歸技術， $T_1T_2T_3T_4$ 四個測量可以建立一套實驗前的時間序列模型，而 T_5 到 T_8 四個測量則可以建立實驗後的時間序列模型，實驗效果的大小，除了反映在 T_4 與 T_5 的對比上，也反映在兩套迴歸模型的差異上，研究者可以比較兩者的斜率與截距，來比較實驗前與實驗後的行為模式的變化。

單組時間序列分析最大的問題是缺乏對照的比較，同一群受試者接受多次的測量，在所列舉的十次測量過程當中，受試者除了受到實驗處理的影響之外，還受到其他外在因素與個人不同的生活事件的影響，也就是歷史因素的混淆，因此可以增加一組受試者作為控制組，控制組的操作程序與實驗組的程序完全相同，唯一的差別是實驗操弄的有無，如表 4.1 的設計十之實驗控制組時間序列分析所示。

4.3 實驗受試的選擇與分派

研究者在設計一個實驗之時，必須考慮多方的因素，進行一連串的決策，以達到最佳的實驗控制。首先，研究者必須考量到如何獲得適當的一群受試者，然後考慮如何將這一群參與研究的受試者安排到不同的實驗情境中。前者是受試者選擇問題，後者是受試者分派問題，兩者都是實驗研究的重要議題。

4.3.1 受試者的選取

一般的想法會以為，相對於非實驗設計，實驗研究在處理受試者時，不論在抽樣程序上或樣本規模的要求上都較為單純，重要的是如何將受試者分派到適當的組別當中。因此，在一般實驗設計的教學上與實際作業時，都會強調受試者分派（subject assignment）的操作，而非強調受試者選擇（subject selection）的問題，著重於實驗組與對照組的可比較性（comparability），而非瞭解受試者的個別差異。但是這並非是正確的認知。如果我們將受試者選擇的問題區分為實驗前、實驗中、與實驗後三個階段來看，我們可以發現受試者的狀態在不同的階段都會影響著整個實驗的進行。

■實驗前受試選擇的影響

首先，受試者樣本怎麼來的？他們的參與實驗的目的、動機、身心狀態、組成結構、背景知識為何？這些因素都可能左右著受試者參與實驗的行為表現。雖然這些因素對於實驗結果的影響可以被視為是一種隨機變異，但是如果隨機變異的大小可以在受試者選擇時加以有效控制的話，那麼對於研究結果的價值將有相當大的貢獻。那麼為什麼無法有效的控制呢？主要的問題在於與成本效率與機率問題。由於

實驗研究著重於「操弄」，因此樣本規模不能太大，在一個小樣本上，樣本的代表性就受到限制，因此，對於各種干擾變項就不易控制。這也就是實驗設計著眼於將受試者「隨機分配」到各實驗組與對照組，而非強調樣本的「隨機選擇」，因為小樣本的機率理論是不穩定的。

為了解決樣本選擇的隨機性與代表性不足的問題，實驗研究採取三種策略來因應：第一是統計的處理與控制，第二是樣本特性的檢驗程序，第三是配對樣本或區組設計（block design）的使用。首先，讀者如果略具統計知識，都理解當樣本數小於 30 時，數據的常態性即無法維繫，此時宜使用 t-test 等可以用來處理小樣本的統計技術；另一方面，如果干擾變項存在，無法利用隨機分派來排除干擾作用，則可以使用共變數分析（ANCOVA）；這些統計方法，其實就是為了實驗研究所發展出來的技術。

第二是樣本特性的檢驗與校正技術，例如進行 ANOVA 與 t 檢定時，有一個重要的檢驗是樣本的同質性檢驗，目的就是在檢查樣本是否具有本質上的差異。為了積極改善這些個別差異的影響，第三種方法，也就是配對樣本與區組設計就隨之發展，透過樣本重複使用或配對分組，來降低干擾變數的影響。這幾種因應策略的使用，使得統計技術變得更為繁複，間接促成現代統計的發展（Wilcox, 2003），但是問題的根本似乎沒有獲得解決。

■實驗中受試選擇的影響

其次，一旦樣本選定之後，在實驗執行的過程中，隨著時間的進展，樣本特性也會與不同因素產生交互作用。比較重要的包括選擇與成熟交互作用（selection-maturation interaction）與選擇與實驗交互作用（selection-treatment interaction），這兩種現象都被視為是影響實驗的內在效度的主要原因。

其中，選擇與成熟交互作用是指受試者的選取與分派，隨著時間的累積所發生的交互作用，例如某些年紀小的受測者隨著時間的延

伸，有較為明顯的改變；或是某些受試者容易因為實驗過程冗長發生的疲勞與抗拒效應。另外，選擇與實驗交互作用則是指不同的受試者對於實驗處理有不同的反應，造成觀察值的變動並非因為實驗操弄所造成，而是受試者的特殊反應，例如不同專業知識背景者，對於特殊的刺激有較佳的反應。

這些基於實驗執行過程中所發生的細微變動不易被客觀的紀錄與察覺，而有賴研究者敏銳的觀察與經驗累積的反應，或是透過細膩的實驗設計來避免問題的發生，若再加上受試者流失之類問題的困擾，都在考驗著研究者的智慧與問題解決能力。

■實驗後受試選擇的影響

一旦實驗完成之後，研究數據陸續蒐集完成，受試者也隨之解散。此時雖然受試者選擇的影響不再發生作用，但是由於樣本的特性會反應在研究數據的內容上，使得研究數據的判讀必須兼顧樣本的特性。

最重要的一個問題是如何下結論。前面已經說明了，一般實驗研究的樣本數不多，因此在母群體的代表性上就不會有太好的表現。更明顯的一個問題是，許多研究的母體顯然與其所欲關心的母體有別，例如許多實驗以大學生為受試者，母體自然是大學就讀中的學生，而非一般老百姓。從科學探索的角度來說，無論以誰來進行研究，所建立的知識都具有一定的解釋力，但是從實務的角度來看，研究者必須忠實於他的工作，翔實的報告研究實施的方法與相關的問題，這一個要求，也是每一個領域中的學術專業倫理的要項之一。也就是說，實驗研究的結論，必須忠實於他的發現過程，用哪一種樣本，下哪一種結論，在推論時，更是要小心為之，不做過度推論。

儘管如此，Babbie (2004) 仍指出，這種受試者選擇的偏離特性，是實證科學的一種危機與缺陷。雖然這個問題對於實驗研究這一種以假設考驗及「行動採取」的解釋性研究 (explanatory research) 的影

響，不如對於非實驗研究這一類以現象的描述為主要目的的**描述性研究**（descriptive research）來得嚴重，但是也挑戰著社會大眾對科學的信心。

4.3.2 受試者的分派

當受試者選定之後，實驗研究所遇到的問題，是如何將這一群參與研究的受試者安排到不同的實驗情境中，也就是決定獨變項當中每一個水準下的實際受試者。如果一個實驗有一個獨變項 A 因子，該獨變項有三個實驗處理與一個控制組，共計四個水準，那麼是不是要將所有的受試者拆散，隨機分配到四個水準當中呢？還是讓每一個受試者都能夠經歷這四種狀況呢？前者將受試者打散，分別接受不同的實驗處理或作為控制組的實驗設計，稱為**獨立樣本設計**（independent sample design）或**受試者間設計**（between-subjects design），也就是說，獨變項當中的每一個水準所使用的受試者是完全獨立無關的，不同實驗效果造成的差異是受試者之間的差異。後者每一個受試者都經歷每一種實驗狀況，稱為**重複量數設計**（repeated measure design）或**受試者內設計**（within-subjects design），也就是說，同一批人重複的接受不同的實驗處理，參與獨變項當中的每一個水準的情境，不同實驗效果造成的影響主要是反應在受試者參與不同實驗處理所產生的受試者內的變異。

（一）受試者間設計

受試者間設計是一種較為單純、安全的實驗設計，所謂安全是因為每一個受試者只接受一種特定的實驗處理，實驗處理的效果不會參雜其他操作性因素的影響（例如受試者內設計即要求同一個受試者也接受其他的實驗處理），研究者主要的工作是讓不同實驗組別的受試者，在接受了實驗處理後，會在依變項上產生最大的變化。而且為了

避免不同的實驗情境的受試者的組成有系統性的差異，（例如某一個實驗情境的女生較多，男生較少），受試者的分配必須做到隨機化，這也就是受試者間實驗設計的**隨機分派**（random assignment）的重要概念。透過隨機化，使不同組別的受試者在進入實驗室之前具有最小的差異，以確保實驗效果的比較是在沒有其他系統性因素的干擾下進行，此時，經過隨機化處理的不同實驗組，稱為**相等組**（equivalent groups），有經過隨機化處理的實驗設計，稱為相等組實驗設計。相等組的意義表示不同的組別的受試者，具有相同的特性與條件，也就是不同組的受試者接受相同的實驗處理，他們的行為結果在統計上應該是相等的，而如果接受不同的實驗處理，則得到不同的行為產出，此時，不同實驗處理下的不同的行為產出，即可歸因於實驗處理效果，而可以作為因果關係的證據。

隨機分派的隨機化，是一種統計的隨機化，利用亂數表或某一種方法將受試者分配到不同的組別，但是分配的結果不一定會有完全平衡的結果，組與組之間存有某種系統性差異的機率仍是存在的。如果在研究進行之前，研究者即預知某一個因素是影響實驗進行的重要混淆因子，在進行隨機分派之時，可以將該因子做特別的處理之後，再進行隨機化，以人為的方法確保該因子在不同的組別是完全一致的。例如喝酒對於駕駛的注意力的影響，可能受到駕駛的性別的影響，而有不同的結果。因此，研究者可以強迫兩組不同組別的受試者有相同比例的男性與女性受試者，然後分別針對不同性別進行隨機化分派，此時不同組別的受試者結構，在性別因子上是完全相同的。

另一種策略來落實隨機化，是利用**配對**（matching）的方式，將不同組別的受試者依照某一個因子來加以配對，使得不同組別的受試者，在某一個因子上具有相類似的分佈條件，以形成一個**區組**（block）。例如喝酒對於駕駛的影響，可能受到開車的資歷的干擾，經驗老到的駕駛與資淺的駕駛的反應情形可能就不同。因此研究者可以讓每一組的受試者依序是屬於資深或資淺的來加以配對，程序上是先蒐集受試者的開車資歷，然後依序排列，如果要將受試者分成四組，那麼最資深的四個人挑出後，以隨機化分配到四個組，形成一個

區組，然後次資深的四個人再隨機分派到四個組（第二個區組），直到所有的受試者都分派完畢，此時，每一個組的受試者，在資歷因子上是具有配對關係的，每一個實驗組的第一位受試者是最資深的，其他受試者的資歷排序都是一樣的，且四個實驗組的受試者的平均年資是相同的，沒有顯著的資歷差異存在。如此一來也可以達成去除受試者系統性因素的干擾。

（二）受試者內設計

相對於受試者間設計的「簡單」與「安全」，受試者內設計則是一個較為複雜、不安全的實驗設計。因為獨變項的每一個水準下的受試者都是同一組人，不同的實驗處理後產生的依變項的變化，除了實驗效果的作用之外，都同時伴隨著同一個人參與多次實驗情境的混淆因素，更具體來說，實驗處理的效果是從每一個受試者經歷不同的實驗情境後，在受試者內部所比較出來的。有趣的是，研究者反而比較偏好使用「不安全」的受試者內設計。主要的原因是效率與經濟因素，因為受試者內設計使用較少的受試者，操作上相對簡單。

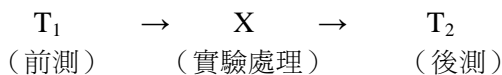
其次，使用受試者內設計因為受試者人數較少，而且每一個實驗情境的實驗效果，都是由同一組人測得的，因此在統計上可以排除較多的測量誤差，同樣的實驗強度下，因為誤差較小而較容易得到顯著的統計結果，**統計檢定力**（statistical power）較受試者間設計為高，因此獲得研究者的喜愛。在受試者內設計中，測量的誤差當中包含受試者在每一個實驗情況下的隨機波動（隨機誤差），以及受試者在不同實驗處理對於依變項測量的穩定差異，此一部分被視為是可被估計系統性誤差，可以從實驗效果中分離出來，稱為實驗處理中的**受試者間變異**（between-subject differences），如果將此一部份獨立估計，可以降低測量誤差，造成統計檢定力的提升。

值得注意的是，以配對組方法來將受試者安排到不同的區組的實驗設計，雖然是一種受試者間設計的作法，但是在統計上，則是以受試者內設計的原理來進行分析（Kirk, 1995），因為同一個區組的受試

者雖然是不同的個體，但是區組內的個體之間具有同質性，原理也如同受試者內設計一樣，同一個人的重複測量會產生重複計算抽樣誤差的現象，必須加以扣減。因此，配對設計有時候也會被視為受試者內設計的一種應用方法，端看是以實驗設計的安排或是統計分析的觀點來看。

(三)前後測設計

受試者內設計除了出現在重複量數設計之中，即不同實驗組使用相同的受試者來進行實驗，另一種更常出現的狀況是**前後測**（pre-post test）的實驗設計。一般而言，實驗研究的基本程序是界定出獨變項與依變項，抽取樣本後，依照獨立樣本設計的原理進行隨機分派，將受試者分配到不同的組別後，操弄獨變項來觀察依變項的變化，但是，為了確保依變項的變化是受到獨變項的影響，研究者通常會在研究之前，先測量一次依變項的得分，是為前測，然後在實驗操弄執行完成之後，再測一次依變項，是為後測，兩次測量的變化，即可歸因於實驗的效果。



前後測的實施，使得研究者可以利用兩次分數的差距來分析實驗的變動分數，也就是實驗前後的依變項的**差異分數**（different scores）分析。在前後測設計中，雖然前後測分數的測量是一種受試者內設計，但是前後測分數並不在反應獨變項不同水準的差異，也就是說，不同實驗狀況與控制組的差異無法在單一的一組前後測差異分數中獲知，因此前後測設計通常還要搭配獨立樣本設計，也就是要納入一個獨變項，來反應實驗效果的大小，屬於混合設計的範圍，我們將在下面的小節中介紹。

前測分數有一個重要的功用，是作為控制的用途。因為在實驗之前，每一位受試者在依變項的得分上應無顯著差異，如果實驗組與控制組在前測的分數達顯著的差異，顯示研究者的分組工作沒有隨機化，研究者應該重新進行分組，以確保研究的品質。如果不同實驗情況的受試者在前測分數僅有微弱的差異，並未達到顯著的水準，我們可以利用統計的方法，將前測分數當作共變數，進行共變數分析，使得研究的數據可以在更公平的情況下進行。

使用受試者內設計雖然較為節省資源，有較高的統計檢定力，但是也要付出相當的代價，最重要的是**痕跡效應**（carryover effect）的威脅，也就是受試者在經歷不同的實驗處理時，前一個實驗處理對於下一個實驗處理的干擾。痕跡效應中，造成干擾與混淆的原因可能是記憶效果、學習效果、成熟效果等等皆有可能，而且對於依變項測量的影響方向也不一定是遞增或遞減，無法預估，因此無法單純的以統計的控制手段來排除痕跡效應，研究者可以運用對抗平衡的原理來處理實驗順序的影響，或是採用**拉丁方格**（Latin-square design）的設計，來檢驗評估實驗順序的影響。

對抗平衡的原理在前面第二節已經做過說明，基本上是利用多組平行的實驗設計，讓受試者在接受多次不同的實驗操弄時，可以接觸到不同的實驗順序，藉使痕跡效應可以被平衡掉。研究者為了兼顧成本因素與對抗平衡的精神，通常會讓每一位受試者隨機接受一種實驗順序，而不需要接受所有的實驗順序，此一作法是假設實驗順序與受試者因素是獨立的，不同的人在接受不同的實驗處理順序的反應是類似的，不同的受試者不會對特定的順序有特別的反應模式，因此可以將順序的效果，以不同受試者的得分求取平均值。例如有三個實驗狀況 A、B、C，受試者必須接受這三種實驗處理時，可以產生六種不同的排列：A→B→C、A→C→B、B→A→C、B→C→A、C→A→B、C→B→A。如果每一位受試者都接受過這六種程序，再求得 A、B、C 各個實驗處理在依變項上的平均值，即可相當程度的平衡掉痕跡效應的威脅。

以上述方法可以使得受試者的負擔減輕，但是對於實驗者而言，總計仍要操弄 18 次的實驗操作，成本仍高。因此研究者多以拉丁方格設計，使得不同的實驗處理的順序效果可以被考量，省略重複性的比較。

(四)混合設計

實驗設計可以區分為受試者間設計與受試者內設計兩大類，這兩種設計的選用取決於研究者要如何進行他們的實驗，如何操弄實驗處理。當研究者僅關心的一個獨變項對於依變項的影響，這一個獨變項可以是受試者間設計或受試者內設計兩者任何一種，但是如果有兩個以上的獨變項，極可能出現其中一個獨變項是以受試者間設計來進行操弄，而另一個獨變項是以受試者內設計來操作，如果一個實驗當中同時包含受試者間設計與受試者內設計，稱為**混合設計**（mixed design）。

不論是混合設計的實驗研究，或是完全採用受試者間設計或受試者內設計的實驗研究，從研究設計的立場來看，目的均在提供研究者探討他所關心的因果關係的研究架構，而沒有絕對的優劣可言。然而，從計量分析的立場來看，混合設計的實驗研究，因為同時具有兩套測量程序，在分析上也就相對複雜。混合設計、受試者間設計與受試者內設計雖然都是使用**變異數分析**（analysis of variance）來進行統計檢定，但是在分析過程中必須考慮不同的測量誤差的估計方法，在變異來源上進行不同方式的拆解分割，統計檢定值（例如 F ratio）的基礎不盡相同，如果又納入拉丁方格等複雜的研究設計，整個統計分析的策略上顯得異常複雜（讀者可參考相關的統計書籍，例如 Kirk, 1995; Berger & Maurer, 2002）。

4.4 實驗設計的計量原理

在各種實驗設計中，不論是否符合實驗設計的嚴格要求，所得到的研究數據都仍然具有一定的分析價值，如果搭配適當的統計分析技術，有些干擾與混淆因素可以獲得相當的控制。

實驗設計的基本精神是操弄獨變項、觀察依變項，獨變項反映的是不同的實驗處理，是為類別變項；依變項則涉及受試者的反應或行為頻率測量，多屬連續變項。在不同實驗狀況下，受試者在依變項上的得分可以求出一個平均值，反應出實驗處理後依變項的強度大小，如果這一個平均數顯著的不同於控制組的平均數，我們可以據以證明實驗效果的存在，相反的，如果實驗組受試者在接受完實驗處理後，依變項測量的平均數沒有顯著的不同於控制組的平均數，則反應實驗效果並不存在，在統計學上，這種推論程序可以利用平均數檢定來進行驗證。

4.4.1 獨變項水準數與統計分析

一個實驗應使用上述何種統計分析技術來分析實驗資料，主要取決於實驗當中獨變項的數目與每一個獨變項的水準數（ K ）。 $K=1$ 、 2 、 3 代表獨變項包含一個、兩個、三個水準。通常獨變項的水準數至少為 2 ，一個水準為實驗處理，另一個水準則為控制組（作為對照）。多數的研究獨變項的水準數會大於三（ $K \geq 3$ ），因為該實驗處理可能包含不同的強度，例如強、中、弱的實驗處理，若再包含控制組，水準數為 4 。

■ $K=1$

一個獨變項如果水準數為 1 （ $K=1$ ），表示獨變項僅有一個實驗組而沒有控制組，屬於單組設計（如表 4.1 的 Type 1、Type 2 與 Type 9 三種實驗設計）。從統計的觀點來看， $K=1$ 時，獨變項便沒有存在的意義，

因為獨變項已成為常數，在實務上，這種設計因為缺乏控制組作為對照，實驗效果也無法作為支持因果關係的論證。儘管如此，當 $K=1$ 時，我們仍然可以使用基本的描述統計（平均數與標準差等等），去呈現依變項測量的結果，例如喝下一杯酒的駕駛，平均反應時間為.12 秒。這個平均數因為缺乏其他平均數的比較，而無法顯示他的意義。到底.12 秒的反應時間是否受到酒精的影響呢？

雖然 $K=1$ 的實驗，僅產生一個平均數。一個平均數缺乏比較而無法顯示其意義，但是，如果利用單樣本平均數統計，我們可以檢測此一平均數是否顯著的不同於某一個特定數目。如此一來，多少也反映了這個平均數的大小意義。例如拿.12 與.10 的反應時間來比較，可以檢定出是否.12 這個平均數是否不同於.10。值得注意的是，這種比較只能用來檢驗這個平均數與其他數目的相對意義，檢定的結果與實驗的操作並沒有任何的關連，因為這個實驗僅僅產生一個平均數。

■ $K=2$

當一個獨變項有兩個水準時，表示實驗處理只有一種狀況（或型態），另一組是為控制組。如果該實驗採獨立組設計（受試者間設計），兩個組使用兩群不同的受試者，實驗完畢之後會得到兩個依變項的平均數， \bar{X}_E 代表實驗組平均數， \bar{X}_C 代表控制組平均數，使用獨立樣本 t 檢定就可以檢驗實驗效果的顯著性。

如果實驗採用的是受試者內設計，兩組受試者是同一群人（重複量數設計），實驗完畢之後也會產生兩個平均數， \bar{X}_E 代表實驗組平均數， \bar{X}_C 代表控制組平均數，唯一的不同是這兩個平均數是由同一組人所計算得出，在計算統計檢定值，可以特別針對每一個受試者在重複測量間的平均值的變異量加以估計，作為系統性誤差，從實驗誤差中予以扣除（詳細的統計原理請參考 t 考驗）。此時，統計技術應選用相依樣本 t 檢定來檢驗實驗效果的顯著性。相依樣本 t 檢定與獨立樣本 t 檢定基本原理是相似的，只有在誤差項的處理上有所不同。

■ $K \geq 3$

當一個獨變項的水準數超過兩個，表示實驗處理有多種不同的狀況（或型態），實驗完畢後，依變項的測量會產生超過兩個以上的平均數，其中一個是控制組平均數 \bar{X}_C ，其他兩個或以上的平均數是該實驗處理不同強度下，受試者在依變項上得分的平均數 \bar{X}_{E1} 、 \bar{X}_{E2} ...。此時 t 考驗即不敷使用，需採用單因子變異數分析（one-way ANOVA），來檢驗這三個以上的平均數是否具有顯著的差異。

如果該實驗採獨立組設計（受試者間設計），不同組的受試者不同，採用獨立樣本單因子變異數分析即可，如果是重複量數設計（受試者內設計），即需改用相依樣本單因子變異數分析來檢驗平均數的差異性。相依樣本變異數分析與獨立樣本變異數分析基本原理也是相似的，也是在誤差項的處理上有所不同。

4.4.2 變項數目與統計分析

一個獨變項代表一種實驗處理，有多少種不同強度（或不同型態）的實驗處理，會反應在該變項的水準數的多寡上。水準數越多，統計分析的方法雖會較為複雜，但是影響並不大，因為僅是組數的增加。但是，如果是獨變項的數目增加，那麼統計分析的難度就會大幅度的提高，因為不同的獨變項會有交互效果。

一般而言，簡單的實驗設計只有一個獨變項，稱為單因子實驗，如果 $K=2$ ，則使用 t 考驗；若 $K \geq 3$ ，則使用單因子變異數分析。假設今天某一個實驗有兩個獨變項，表示有兩種實驗處理，稱為二因子實驗，三個獨變項的實驗表示有三種實驗處理，稱為三因子實驗，因子數越多，統計分析的方法越複雜，因為在實驗處理實施後，每一個因子在依變項上都會產生多個平均數，多個因子產生多個平均數；各個獨變項之間又存在有交互效果，反應在細格平均數（ \bar{X}_{Cell} ）上。

對於一個二因子實驗，兩個獨變項若各有 3 個水準，此時實驗處理

實施完畢後，除了兩個獨變項各自產生的三個平均數以外，交互作用還會產生九個細格平均數，如此一來，這 15 個平均數的比較就不能用單因子變異數分析來處理，而需使用**多因子變異數分析**(factorial analysis of variance)來檢驗實驗效果。如果每一種實驗處理的強度又有多種類型，每一個獨變項水準數增加，統計的方法益形複雜。

除了平均數的數目增加以外，多因子實驗設計下，每一個獨變項是使用受試者內設計或受試者間設計，又會影響統計分析的方法。如果所有的獨變項是獨立組設計（受試者間設計），可使用完全獨立樣本多因子變異數分析來檢定實驗效果，如果所有的獨變項都是採用相依原理分派受試者到不同實驗情境，例如配對組設計與重複量數設計（受試者內設計），可使用完全相依樣本多因子變異數分析來檢定實驗效果，如果有的獨變項是受試者間設計、有的獨變項是受試者內設計（混合設計實驗），則需使用混合設計多因子變異數分析來進行分析。基本上，凡是涉及受試者內設計的統計方法，都與區組設計變異數分析有關，這些內容我們將在統計原理的專書中介紹。

至於依變項的數目，對於統計分析的影響更為明顯。多數的實驗研究，依變項只有一個，使用一般的變異數分析即可。如果有多個依變項，傳統的變異數分析即不敷使用。多重依變項的實驗設計，必須使用**多變量變異數分析**(multivariate analysis of variance)來檢驗實驗效果，多變量變異數分析屬於高等統計的範圍，分析的程序也更為複雜，將在本列叢書的多變量統計篇介紹。

4.5 實驗設計的混淆與干擾

一個實驗的好壞，受到兩大因素的影響，第一是研究者本身的設計或執行實驗的能力，包括背景知識的訓練、個人的創意、做事的嚴謹態度、分析技術的配合，第二是實驗本身導致實驗無效的各種威脅，這些對於實驗的潛在影響因素，亦即 Campbell 與 Stanley（1963）以及 Cook 與 Campbell（1979）所指出的實驗的內在效度與外在效度的威脅因素，雖然可以透過實驗者事前在研究設計當中以及過程當中予以防範，但是隨著實驗的進行仍然相當可能發生。這些影響因素經過摘要整理後，列於表 4.4。

從實驗者本身來說，傑出的實驗研究與研究者本身對於所要研究的問題的掌握，以及研究者設計整個實驗的個人「創意」或「巧思」有著絕對的關聯。包括了所設計的實驗是否能夠切合研究問題與假設，以及是否獲得充分的實驗工具與器材的支援兩大因素。如果研究者事前對於所欲探討的問題有著充分的研究，掌握了所有關鍵性的文獻，瞭解前人研究所不能解決的問題之所在，並能夠整合其他來源的知識與學習成果來解決手上的問題，如此專業而踏實的訓練與研究計畫發展過程，可以協助研究者提出獨特、適切而有研究價值的研究假設。進一步的，如果研究者在研究方法上有著豐富的經驗，研究的環境也能夠充分的提供資源，使得研究者可以扮演有米得炊的巧婦，將實驗設計中的巧思能夠確實的放入實驗過程來充分發揮，如此一來，實驗的進行就已經有了一個良好的基礎條件。

有了好的研究設計與外在支援，要獲得滿意的研究成果，還與研究者的實際執行能力有關。一個具有創意巧思的實驗設計，在一個配備齊全的實驗室中，如果操作人員的執行能力不足，實驗處理無法有效的操弄，獨變項效果無法充分的反應在受試者身上，或是研究人員無法有效的控制研究環境與混淆干擾因素，使得各種條件可以保持一致，最後的研究成效必定不彰。因此，研究人員的實際研究操作能力也是必須事前加以培養訓練的重要能力。

表 4.4 實驗效度的混淆干擾因素

混淆/干擾來源	說 明	範 例
歷史 history	發生在實驗過程中的特殊事件	受試者在實驗過程中經歷到的不同事情，例如一本書的影響
成熟 maturation	隨著時間的進展在受試者身上發生的自然改變	受試者的疲勞、飢餓、注意力喪失、態度改變
測驗 testing	多次測量的干擾效應、練習效果或特殊反應心向	受試者熟悉了測量的內容、討好的作答傾向
工具 instrumentation	工具本身的使用造成的混淆	評分者的差異、工具信度偏低
統計迴歸 statistical regression	多次測量下極端分數朝向中間化的計量傾向	前測高分者在後測有向下修正的傾向
選擇偏誤 selection biases	由於受試者選取與分派過程發生的偏誤	某一組的受試者普遍有較高的參與動機
實驗受試者流失 experimental mortality	受試者的流失造成的影響	受試者因為特殊緣故無法繼續參與研究
因果時序 causal time-order	不正確或模糊、不明確的因果關聯造成的影響	受試者的焦慮影響學習，而學習也影響了焦慮
散佈模仿 diffusion/imitation	實驗過程中，不同組別的互動與訊息洩漏	實驗者無意間透露其他實驗組的動態
補償 compensation	實驗者對於不同實驗處理組別的特殊對待	實驗者怕控制組的受試者無聊而有進一步的討論與互動
努力/放棄 compensatory rivalry or demoralization	受試者處於不同實驗處理下的特殊反應	受試者被分配到較差的組別導致想要力爭上游的動機或退縮
前測的反應與互動 reactive/interaction effects of pretest	前測的實施對於受試者的特殊影響	前測的實施使得受試者猜測實驗的內容而試圖討好實驗者
實驗的反應與互動 reactive/interaction effects of experimental	實驗過程對於受試者的特殊影響	受試者覺得實驗過於敏感、激烈造成的影響
選擇與成熟交互作用 selection-maturation interaction	受試者選擇與分派的問題加上時間的累積造成的合併影響	某一組受試者比較不想參加研究，加上實驗過程冗長發生的疲勞與抗拒效應
選擇與實驗交互作用 selection-treatment interaction	受試者選擇與分派的問題加上實驗處理的合併影響	某一組受試者比較不想參加研究，加上實驗是操作無聊的活動
多重實驗處理的干擾 multiple treatment interference	使用多次實驗處理的干擾	受試者被要求參加不同的活動，前一個活動影響下一次的活動

這些影響實驗的內外效度的各種威脅因素上，表 4.4 列出了十六種較為常見的現象，這些內容在第二章中已有初步的介紹。在這十六種威脅因子當中，多數影響到的是一個實驗的內在效度。例如歷史、成熟、選擇偏誤、測量、工具、統計迴歸、受試者流失、實驗效果的散佈模仿、補償作用等等，這些現象的發生，與一個實驗設計的基本要件的落實與否具有相當密切的關係。如同前面所說的，一個典型的實驗設計，必須做到隨機分派、實驗與控制對照、以及前測後測的比較三個要件，如果研究者確實實踐這三個要件的要求，表 4.4 的問題有絕大部分可以避免，因此，一個實驗是否遵循實驗的設計原理，是討論內外效度的重要前提。

在實驗設計的基本要件中，實驗與控制組的對照設計主要就是在協助避免特殊歷史事件的威脅，因為不論是實驗組或是控制組的受試者，都有可能遇到在實驗考量之外的特殊經驗，對於實驗組與控制組，特別的經驗被以隨機誤差的處理，歷史的威脅即可有相當程度的降低；另外，隨機分派則可以處理選擇偏誤的問題，前後測設計可以將測驗效果加以控制，所羅門設計可以特別去處理前測的反應效果。但是，也有部分的混淆因素無法透過實驗設計處理來排除，例如受試者的流失、統計迴歸效應、因果時序的模糊性、受試者反應等等。也就是說，儘管一個實驗設計符合所有的實驗設計要件，內在效度的威脅只能部分的消滅，如果要排除其他的一些威脅因素，但是還要有其他條件的配合，例如實驗程序的掌控，以及適當的統計計量策略的使用等等。有關計量方法與統計策略的使用，本人將在另一本此系列叢書中做詳細說明。

另外，有些影響因素則可能削弱了實驗的外在效度，減低了實驗結果推論到現實情境與不同場合的能力。例如樣本的選用與實驗處理的交互作用、受試者對於實驗程序反應、受試者對於前測的特殊反應、多重實驗處理的運用等等。都使得研究的實驗處理受到了扭曲，與真實環境下的作用有所不同，未來要應用到現實問題的解決時，產生不同於當初研究所獲得的結果的效應。如果實驗成果無法推及到真實世界，那麼學術的成就也就失去了意義。

4.6 結語

自從實證典範在學術領域獲得了主導地位後，實驗方法即成為科學研究的主要策略，即是因為實驗方法的蓬勃流行，因此統計方法也就獲得相對的發展。本章的主要目的在介紹不同的實驗設計，但更重要的目的是在說明實驗設計與統計分析之間的關係，使得研究者可以在確立了實驗設計的基本概念之後，可以順利的銜接分析的方法與技術。

事實上，曾經操作過實驗方法的研究者都有一個共通的心得，也就是實驗設計本身或許沒有那麼複雜，實驗的操弄也可以照表操課，而最困難的部分就是如何把辛苦蒐集得到的數據加以分析，以得到即符合實驗要求，也能凸顯研究設計理念的研究結果。另一個常遇到的挑戰，是干擾因素何其多，我們到底要在實驗程序中控制哪些因素，而利用統計方法來控制哪些因素，這些問題通常並沒有一個標準答案，必須靠研究者自行判斷與取捨。因此，除了熟悉實驗設計原理之外，好的實驗研究必須要有嚴謹的測量與統計程序相配合，這也是為什麼統計分析的要求與嚴格訓練是社會與行為科學領域的重要傳統。

本章重要名詞

實驗 (experiment)	控制組 (control group)
實驗設計 (experiment design)	多因子設計 (factorial design)
統計假設 (statistical hypothesis)	多變量變異數分析 (multivariate analysis of variance)
可複製性 (replication)	多變量共變數分析 (multivariate analysis of covariance)
隨機性 (randomization)	交互效果 (interaction effect)
區組性 (blocking)	前實驗設計 (pre-experimental design)
獨變項 (independent variable)	單組後測設計 (one-shot case study)
依變項 (dependent variable)	
控制變項 (control variable)	
實驗組 (experimental group)	

單組前後測設計 (one-group pretest-posttest design)
靜態組間比較 (static-group comparison)
真實驗設計 (true-experimental design)
隨機化實驗控制組前後測設計 (randomized control-group pretest-posttest design)
所羅門設計 (Solomon four-group design)
準實驗設計 (quasi-experimental design)
實驗室實驗法 (laboratory experiment)
田野實驗法 (field experiment)
對抗平衡 (counterbalancing)
拉丁方格設計 (Latin square design)
痕跡效應 (carryover effect)
平衡性拉丁方格 (balanced Latin square)
單組時間序列分析 (one-group time-series)
實驗控制組時間序列分析 (control-group time-series)
縱貫研究 (longitudinal study)
時間序列設計 (time-series design)
中斷性時間序列設計 (interrupted time-series design)
受試者分派 (subject assignment)
受試者選擇 (subject selection)
區組設計 (blocking design)
選擇與成熟交互作用 (selection-maturation interaction)
選擇與實驗交互作用 (selection-treatment interaction)
解釋性研究 (explanatory research)
描述性研究 (descriptive research)
獨立樣本設計 (independent sample design)
受試者間設計 (between-subjects design)
重複量數設計 (repeated measure design)
受試者內設計 (within-subjects design)
隨機分派 (random assignment)
相等組 (equivalent groups)
配對 (matching)
統計檢定力 (statistical power)
前後測 (pre-post test)
差異分數 (different scores)
混合設計 (mixed design)
變異數分析 (analysis of variance)
單因子變異數分析 (one-way analysis of variance)