

第 8 章 屬性差異試驗(Attribute Difference Tests)：

樣品間屬性 x 如何不同?

(How Does Attribute X Differ Between Samples?)

- 8.1 前言：配對比較(paired comparison)設計
- 8.2 指向(directional)差異試驗：比較 2 個樣品
- 8.3 成對(pairwise)順位試驗：Friedman 分析 — 比較數個樣品在全部可能的配對
- 8.4 前言：多重樣品(multisample)差異試驗 — 區集設計(block design)
- 8.5 單純順位試驗：Friedman 分析 — 隨機完全區集設計
- 8.6 多重樣品差異試驗：評級(rating)方法 — 變異數分析(ANOVA)評估
- 8.7 多重樣品差異試驗：BIB 順位試驗(平衡不完全區集設計) — Friedman 分析
- 8.8 多重樣品差異試驗：BIB 評級試驗(平衡不完全區集設計) — ANOVA 評估

在 8.3 及 8.4 項，要求受試者將每一樣品和其餘的每個樣品都做比較，這樣的配對比較(paired comparisons)其優點是讓每樣品所興趣的屬性，都在有意義的尺度上得到好的測量，以及可能的對每配對其屬性的相對強度也測量，然而，出現的配對數目會隨著樣品數的增加而以指數關係急遽上升：

$$\begin{array}{l} \text{樣品數 } t \\ \text{可能配對數 } N = t(t-1)/2 \end{array} \quad \begin{array}{cccccccc} 3 & 4 & 5 & 6 & 7 & 8 & 9 \\ \hline 3 & 6 & 10 & 15 & 21 & 28 & 36 \end{array}$$

在 8.3 及 8.4 項，我們會提問：「那個樣品較甜 < 或者較新鮮、較喜歡? >」；這種作法是依據順位數據(rank data)（譬如，較甜的樣品指定為順位二而另一樣品為順位一），這導入某程度的人為；即直接得自每位回應者而差異程度不作測量。故其統計方法較簡單，相對的，若是評級數據(rating data)就需要專門的統計方法。

8.1 前言：配對比較設計(Paired Comparison Designs)

屬性差異試驗(attribute difference tests)：針對單項屬性譬如甜味，測量比較某樣品和另一或數個樣品之間的差異。

注意：樣品之間一項屬性無差異，並不意謂整體差異就不存在。

兩個樣品(two samples)(7.2 項)的屬性差異試驗，在試驗設計與統計處理上簡單，主要的難處在於決定試驗情況是單尾或雙尾的(one-sided or two-sided)（見範例 8.1 及 8.2）。

超過二個樣品時，一些設計可採用變異數分析，其餘則需要專門的統計學，複雜度隨樣品數目的增加而急速上升。其次是各種的多重配對試驗(multiple pair tests)的介紹，在 8.4 項討論多重樣品試驗(multisample tests)及其設計。

8.2 指向差異試驗(directional difference test)：比較兩個樣品

8.2.1 範圍與應用

當試驗目標是確定兩個樣品之間特定的感覺特性不同是何方式(例如，那個樣品是較甜的)。

在此模式，這方法又稱為配對比較試驗(Paired Comparison test)或二擇一迫選試驗(2-AFC; 2-alternative Forced Choice)，乃最簡單且最常用的感官試驗法之一，如有必要用到其它更複雜的測試，否則通常先採用此法測定。兩個樣品配對比較的其它形式為同/異試驗(Same/Different test)(參見 7.6 項)及配對喜好試驗(Paired Preference test)(參見 13.5.3.1 項)。

當使用配對比較試驗，須一開始就要區分是**雙尾的 (two-sided)(bilateral)**，最普遍)或是**單尾的(one-sided)(unilateral)**，當只一個回答是有興趣的或者只一個回答是正確的)應用(參見 14.3.2 項與範例 8.2 中的註解項)。

統一方法也應用於配對比較試驗，試驗所需的回應者數目受下列的因素影響：**(1)**是否試驗為單尾的(採用表 19.9)或雙尾的(採用表 19.11)，**(2)**試驗靈敏度參數(test sensitivity parameters)(α , β ,及 P_{max})的設定值。

- 在配對比較試驗，參數 P_{max} 取代來自第 7 章討論的整體差異試驗之參數 P_d 。

➤ P_{max} 是與同等強度(equal intensity)的**偏離** (即回應者的意見為 **50:50 分歧**)，對研究人員代表有意義的差異。例如研究人員認為回應者群體中的 **60:40** 分歧，與同等強度是**有意義地大的偏離**，然後，從 $P_{max} = 0.60$ 和研究人員從表 19.9 及表 19.11 內在選定的 α 及 β 值之對應項中，就可發現回應者的數目是多少。大致的原則：

- $P_{max} < 55\%$ 代表與同等強度的**小的偏離**(small departures)；
- $55\% \leq P_{max} \leq 65\%$ 代表與同等強度的**中等偏離** (medium departures)；
- $P_{max} > 65\%$ 代表與同等強度的**大的偏離**(large departures)。

8.2.2 原理

提供每位受試者 **2** 個編碼樣品，準備同數量的 **AB** 與 **BA** 組合，並隨機分配給受試者。要求受試者由左至右品嚐產品，填寫評分卷。清楚地告知受試者是否允許“無差異 no difference”的回答。

僅“**強迫選擇技術 forced choice technique**”是適合於正式的統計分析，然而某些情況下，受試者當感知不到有什麼時，他可能很反對去找出差異。這時感官分析專家就須決定是否：**(1)將分數平分給兩個樣品、或(2)忽略其分數**。步驟**(1)**可提高發現差異之機率，而步驟**(2)**則減低；因此，專家須面對影響結果的其一或另一方式的責任/誘惑。實際上，約一半的專家禁止“無差異”的回答，而另一半發現，快樂的品評小組是比較好的小組，最常採用步驟**(1)**。

8.2.3 試驗受試者

由於本試驗簡單，接受最低限訓練的受試者(讓受試者完全熟識試驗的屬性是足夠的)就可執行，或如果試驗有其特別的重要性(例如已上市產品的不良風味)，可選擇高度訓練的受試者，他們對該屬性顯示特別的敏感度。

由於用猜的機率是 **50%**，需要相當**大數量**的受試者。從表 19.12，以 **15** 次呈現為例，如預定達顯著水準 $\alpha = 0.01$ ，須有 **13** 次同意；如是 **50** 次呈現，同樣的顯著的顯著水準下要有 **35** 次同意的裁決。

8.2.4 試驗程序步驟

- ▶ 試驗控制(test controls)與產品控制(product controls)參照 3.2 及 3.3 項。
- ▶ 可能的話，同時提供樣品；或者先後依序提供。
- ▶ 準備相同數量的 A/B 及 B/A 組合，隨機分配給受試者。步驟細節參閱 7.3.4 項。
- ▶ 典型的評分卷，如圖 8.1 所示。注意不論是單尾的或雙尾的試驗，評分卷都一樣，但評分卷必須顯示是否允許“無差異”的回答(或受試者知道此事)。
- ▶ 在一份評分卷上可提供幾次連續配對比較的空間，但不可加入補充問題，因為會導入偏差/偏見。
- ▶ 計算有興趣的回應者數量，在單尾的試驗，計算正確回應數或感興趣方面的回應，參考表 19.10。而在雙尾的試驗，計算同意(agreeing)之回應數(更頻繁引用某一樣品)，參考表 19.12。

試驗設計 — 由於不同的人會有不同的鮮榨風味的想法，需要大的品評小組，但訓練不是強烈需要。採用 40 位受試者的配對比較試驗及 α 誤差 5%，即 $\alpha = 0.05$ ，認為是適當的。

虛無假設(null hypothesis) 為 H_0 : 鮮度 A = 鮮度 B，
另一假設為 H_a : 鮮度 A \neq 鮮度 B；

任一種結果都是感興趣的，故是雙尾的。將樣品編碼為 691 及 812，如圖 8.1 所示的評分卷用來收集數據。

Table 19.12 Critical Number of Correct Responses in a Two-Sided Directional Difference Test
(Entries are $x_{\alpha,n}$)
Entries are the minimum number of correct responses required for significance at the stated α -level (i.e., column) for the corresponding number of respondents, n (i.e., row). Reject the assumption of "no difference" if the number of correct responses is greater than or equal to the tabled value.

n	α							n	α						
	0.40	0.30	0.20	0.10	0.05	0.01	0.001		0.40	0.30	0.20	0.10	0.05	0.01	0.001
2	1	1	1	1	1	1	1	31	19	19	20	21	22	24	26
3	3	3	3	3	3	3	3	32	19	20	21	22	23	24	26
4	4	4	4	4	4	4	4	33	20	20	21	22	23	25	27
5	4	5	5	5	5	5	5	34	20	21	22	23	24	25	27
6	5	5	6	6	6	6	6	35	21	22	22	23	24	26	28
7	6	6	6	7	7	7	7	36	22	22	23	24	25	27	29
8	6	6	7	7	8	8	8	40	24	24	25	26	27	29	31
9	7	7	7	8	8	9	9	44	26	26	27	28	29	31	34
10	7	8	8	9	9	10	10	48	28	29	29	31	32	34	36
11	8	8	9	9	10	11	11	52	30	31	32	33	34	36	39
12	8	9	9	10	10	11	12	56	32	33	34	35	36	39	41
13	9	9	10	10	11	12	13	60	34	35	36	37	39	41	44
14	10	10	11	11	12	13	14	64	36	37	38	40	41	43	46
15	10	11	11	12	12	13	14	68	38	39	40	42	43	46	48
16	11	11	12	12	13	14	15	72	41	41	42	44	45	48	51
17	11	12	12	13	13	15	16	76	43	44	45	46	48	50	53
18	12	12	13	13	14	15	17	80	45	46	47	48	50	52	56
19	12	13	13	14	15	16	17	84	47	48	49	51	52	55	58
20	13	13	14	15	15	17	18	88	49	50	51	53	54	57	60
21	13	14	14	15	16	17	19	92	51	52	53	55	56	59	63
22	14	14	15	16	17	18	19	96	53	54	55	57	59	62	65
23	15	15	16	16	17	19	20	100	55	56	57	59	61	64	67
24	15	16	17	18	19	21	21	104	57	58	60	61	63	66	70
25	16	16	17	18	20	21	22	108	59	60	62	64	65	68	72
26	16	17	17	18	19	20	22	112	61	62	64	66	67	71	74
27	17	17	18	19	20	21	23	116	64	65	66	68	70	73	77
28	17	18	19	20	21	22	23	122	67	68	69	71	73	76	80
29	18	18	19	20	22	23	24	128	70	71	72	74	76	80	83
30	18	19	20	21	22	23	25	134	73	74	75	78	79	83	87
	18	19	20	21	23	25		140	76	77	79	81	83	86	90

Note: For values of n not in the table, compute $z = (k - 0.5n) / \sqrt{0.25n}$, where k is the number of correct responses. Compare the value of z to the $\alpha/2$ -critical value of a standard normal variable, that is, the values in the last row of Table 19.3 ($z_{\alpha/2} = t_{\alpha/2, \infty}$).

範例 8.1：指向差異(雙尾的)：Crystal Mix Lemonade

問題/情況 — 就檸檬水消費者的研究指出，消費者最在意檸檬/檸檬水風味是最像“剛榨的檸檬水”。該公司針對粉狀混合物已開發了兩種有潛力的風味系統；是否其中的一種比起另一種具有更鮮榨的檸檬特性，開發人員想要取得一些測量。

計畫目標 — 開發鮮榨檸檬特性高的一種產品。

試驗目標 — 測定兩種風味系統傳遞更像鮮榨檸檬水之相對能力。

Directional Difference Test	
Name: _____	Date: _____
Type of sample: _____	_____
Characteristic studied: _____	_____
Instructions: Taste each pair from left to right and enter your verdict below. If no difference is apparent, enter your best guess, however uncertain. "No difference" verdicts are permitted, but only as a last resort.	
Test pairs	Which sample is more _____
_____	_____
_____	_____
_____	_____
Comments: _____	

圖 8.1

指向差異試驗用評分卷範
例，呈現：配對比較，允許“無差異”的裁決。

篩選樣品 — 提前品嚐樣品以確認兩種樣品其檸檬風味的強度是相似的。

分析結果 — 樣品 812 被 26 位受試者選出具有更多的鮮榨檸檬風味，4 位受試者報告“無差異”，平分給兩個樣品。根據表 19.12，結論：40 人中的 28 位選出樣品 812，這人數已足以構成足夠的差異。

解釋結果 — 建議未來採用配方 812，具有顯著性地更多鮮榨檸檬特性。

範例 8.2：指向差異 (單尾的)：啤酒苦味

問題/情況 — 某釀酒商收到市場報告，指出其啤酒 A 的苦味不夠強，因此使用較多量的啤酒花做出試驗啤酒 B。

計畫目標 — 製造一款啤酒在感知上更苦些，但苦味也不能過強。

試驗目標 — 比較啤酒 A 與 B，確定是否得到小但顯著增加的苦味。

Table 19.12 Critical Number of Correct Responses in a Two-Sided Directional Difference Test (Entries are x_{α})

Entries are the minimum number of correct responses required for significance at the stated α -level (i.e., column) for the corresponding number of respondents, n (i.e., row). Reject the assumption of "no difference" if the number of correct responses is greater than or equal to the tabled value.

n	α							n	α						
	0.40	0.30	0.20	0.10	0.05	0.01	0.001		0.40	0.30	0.20	0.10	0.05	0.01	0.001
2	1	1	1	1	1	1	1	31	19	19	20	21	22	24	25
3	3	3	3	3	3	3	3	32	19	20	21	22	23	24	26
4	4	4	4	4	4	4	4	33	20	20	21	22	23	25	27
4	4	4	4	4	4	4	4	34	20	21	22	23	24	25	27
5	4	5	5	5	5	5	5	35	21	22	22	23	24	26	28
6	5	5	6	6	6	6	6	36	22	22	23	24	25	27	29
7	6	6	6	7	7	7	7	40	24	24	25	26	27	29	31
8	6	6	7	7	8	8	8	44	26	26	27	28	29	31	34
9	7	7	7	8	8	9	9	48	28	29	29	31	32	34	36
10	7	8	8	9	9	10	10	52	30	31	32	33	34	36	39
11	8	8	9	9	10	11	11	56	32	33	34	35	36	39	41
12	8	9	9	10	10	11	12	60	34	35	36	37	39	41	44
13	9	9	10	10	11	12	13	64	36	37	38	40	41	43	46
14	10	10	10	11	12	13	14	68	38	39	40	42	43	46	48
15	10	11	11	12	12	13	14	72	41	41	42	44	45	48	51
16	11	11	12	12	13	14	15	76	43	44	45	46	48	50	53
17	11	12	12	13	13	15	16	80	45	46	47	48	50	52	56
18	12	12	13	13	14	15	17	84	47	48	49	51	52	55	58
19	12	13	13	14	15	16	17	88	49	50	51	53	54	57	60
20	13	13	14	15	15	17	18	92	51	52	53	55	56	59	63
21	13	14	14	15	16	17	19	96	53	54	55	57	59	62	65
22	14	14	15	16	17	18	19	100	55	56	57	59	61	64	67
23	15	15	16	16	17	19	20	104	57	58	60	61	63	66	70
24	15	16	17	18	19	21	21	108	59	60	62	64	65	68	72
25	16	16	17	18	18	20	21	112	61	62	64	66	67	71	74
26	16	17	17	18	19	20	22	116	64	65	66	68	70	73	77
27	17	17	18	19	20	21	23	122	67	68	69	71	73	76	80
28	17	18	18	19	20	22	23	128	70	71	72	74	76	80	83
29	18	18	19	20	21	22	24	134	73	74	75	78	79	83	87
30	18	19	20	21	23	25	25	140	76	77	79	81	83	86	90

Note: For values of n not in the table, compute $z = (k - 0.5n) / \sqrt{0.25n}$, where k is the number of correct responses. Compare the value of z to the $\alpha/2$ -critical value of a standard normal variable, that is, the values in the last row of Table 19.3 ($z_{\alpha/2} = t_{\alpha/2, \infty}$).

試驗設計 — 選配對-比較指向差異試驗，因感興趣之處在於苦味的增加，沒有其它。計畫負責人選擇高的確定性，即 $\alpha = 0.01$ ，感官分析專家將啤酒編碼為 452 及 603，提供給 30 位受試者(已證實可測出苦味的小變化之能力的品評小組，評分卷上詢問“那樣品是較為苦的？”(而不是：“是 603 比 432 更苦些？”)，就不會引起受試者的偏見。

篩選樣品 — 樣品經 6 人的小品評小組先品嚐，確認苦味以外的差異都是最低的。

分析結果 — 有 22 位受試者選擇樣品 B。

虛無假設 H_0 ：苦味 A = 苦味 B，另一假設 H_a ：苦味 B > 苦味 A，這使得試驗為單尾的 (one-sided)。專家結論：根據表 19.10，苦味的差異可被感知在危險率 $\alpha = 0.01$ ，故試驗啤酒是成功的。

Table 19.10 Critical Number of Correct Responses in Duo-Trio and One-Sided Directional Difference Test (Entries are $x_{\alpha,n}$)
Entries are the minimum number of correct responses required for significance at the stated α -level (i.e., column) for the corresponding number of respondents, n (i.e., row). Reject the assumption of “no difference” if the number of correct responses is greater than or equal to the tabled value.

n	α							n	α						
	0.40	0.30	0.20	0.10	0.05	0.01	0.001		0.40	0.30	0.20	0.10	0.05	0.01	0.001
2	2							31	17	18	19	20	21	23	25
3	3	2						32	18	18	19	21	22	24	26
4	3	3	3					33	18	19	20	21	22	24	26
5	4	4	4	4				34	19	20	20	22	23	25	27
6	4	4	4	5	5			35	19	20	21	22	23	25	27
7	5	5	5	6	6			36	20	21	22	23	24	26	28
8	5	5	6	6	7	7		40	22	23	24	25	26	28	31
9	6	6	7	7	8	8		44	24	25	26	27	28	31	33
10	6	7	7	8	9	10		48	26	27	28	29	31	33	36
11	7	7	8	9	9	10	11	52	28	29	30	32	33	35	38
12	7	8	8	9	10	11	12	56	30	31	32	34	35	38	40
13	8	8	9	10	10	12	13	60	32	33	34	36	37	40	43
14	8	9	10	10	11	12	13	64	34	35	36	38	40	42	45
15	9	10	10	11	12	13	14	68	36	37	38	40	42	45	48
16	10	10	11	12	12	14	15	72	38	39	41	42	44	47	50
17	10	11	11	12	13	14	16	76	40	41	43	45	46	49	52
18	11	11	12	13	13	15	16	80	42	43	45	47	48	51	55
19	11	12	12	13	14	15	17	84	44	45	47	49	51	54	57
20	12	12	13	14	15	16	18	88	46	47	49	51	53	56	59
21	12	13	13	14	15	17	18	92	48	50	51	53	55	58	62
22	13	13	14	15	16	17	19	96	50	52	53	55	57	60	64
23	13	14	15	16	16	18	20	100	52	54	55	57	59	63	66
24	14	14	15	16	17	19	20	104	54	56	57	60	61	65	69
25	14	15	16	17	18	19	21	108	56	58	59	62	64	67	71
26	15	15	16	17	18	20	22	112	58	60	61	64	66	69	73
27	15	16	17	18	19	20	22	116	60	62	64	66	68	71	76
28	16	16	17	18	19	21	23	122	63	65	67	69	71	75	79
29	16	17	18	19	20	22	24	128	66	68	70	72	74	78	82
30	17	17	18	20	20	22	24	134	69	71	73	75	78	81	86
31	17	18	19	20	22	24		140	72	74	76	79	81	85	89

Note: For values of n not in the table, compute $z = (k - 0.5n) / \sqrt{0.25n}$ where k is the number of correct responses. Compare the value of z to the α -critical value of a standard normal variable, that is, the values in the last row of Table 19.3 ($z_{\alpha} = t_{\alpha,n}$).

註解：決定一個配對比較試驗到底是單尾的或雙尾的，其重點不在於詢問受試者的問題是一或二個回答。

單尾的試驗其發生情況，主要在試驗目標是去確認一個明確的“改良”或處理作用 (treatment effect) (也參見 14.3.1)。單尾的及雙尾的情況的一些範例為

單尾的

- 確認試驗啤酒是更苦的
- 確認試驗產品被喜好(因我們有預先的理由可期待)
- 在訓練嗜味人員：何者樣品是最有水果味的(使用 doctored samples)

雙尾的

- 決定那個試驗釀酒是更苦的
- 決定那個產品被喜好
- 大多數的其它試驗情況 — 每當另一假設為樣本是不同的，而不是“一個比另一個更多”

8.4 成對順位試驗 (Pairwise Ranking Test)：Friedman 分析 — 比較幾個樣品在全部可能的配對組

8.4.1 範圍與應用

當試驗目標是要比較幾個樣品的單一項屬性，例如甜味、鮮度或喜好，採用本方法。

本法特別有用於 3-6 個的樣品組，要用相當缺乏經驗的品評小組來品評。按所選定屬性的強度比例而排列樣品，並給樣品間的差異一個數值表示與該差異的顯著性。

8.4.2 試驗原理

同一時間給每位受試者呈現順序為隨機的 1 個配對 (one pair)，提問：“那個樣品是較甜的？”(較鮮的、喜好的等)，持續進行至每位受試者都品評所有可能的配對組。以 Friedman-type 統計分析評定結果。

8.4.3 試驗受試者

- 挑選、訓練與指導受試者，如 7.3.3 項中所述。
- 採用不低於 10 位的受試者；如 20 位或更多可用，可大為提高辨識的效率。
- **查明受試者能認清要探討的屬性**，例如，透過以屬性係已知強度差異的不同配對組來訓練。
- 取決於試驗目標，受試者可能需要是證實有測出屬性的小差異之能力。

8.4.4 試驗程序

- 試驗控制(test controls)與產品控制(product controls)，參見 3.2 及 3.3 項。
- 可能的話，樣品同時提供，或相繼提出。步驟的細節參照 7.3.4 項。
- 確認呈現的順序是真正地隨機；不該讓受試者期待一個固定的樣式，因這會影響裁決。

範例 8.4：玉米糖漿的口感

問題/情況 — 某混合食用糖漿的製造商想要將一款在給定的固形量仍低稠度的產品上市，已準備 4 種無調味玉米糖漿混合物 A、B、C、及 D 要品評(Carr, 1985)。

計畫目標 — 品評 4 種糖漿混合物的是合適性。

試驗目標 — 建立 4 種糖漿混合其感知口感的稠度在主觀尺度上的位置。

試驗設計 — 選擇成對**順位試驗**及 **Friedman** 分析：**(1)** 因為配對的呈現較少受這些樣品帶來的疲勞，及**(2)** 因為這試驗建立了一個有意義的尺度。

已證實能力的 12 位受試者品評 6 個可能的配對組 **AB、AC、AD、BC、BD、及 CD**。

工作表與評分卷分別顯示於圖 8.3 與圖 8.4。

- 配對組內、配對組間、及受試者之間都要隨機呈現。
- **只問一個問題：“那個樣品是更...？”**
- 不允許“無差異”的裁決；如果他們仍然發生，平均分配選票至樣品。

Date 11-6-98		WORKSHEET												No.	7B
CODE	SAMPLE	CODE												SAMPLE	
A	Blend 4238	C												CCSA Blend III	
B	Blend 133-88	D												Test Sample 11.3A	
Each Panelist receives the six possible pairs in balanced random order. Each sample is coded with a random number.															
Panelist														Order of presentation and serving code	
No.	1st	2nd	3rd	4th	5th	6th									
1	A D	B D	B C	C D	A C	A B							A B		
	119 634	128 824	316 967	242 659	978 643	224 681									
2	A C	B C	A D	A B	B D	C D							A B		
	293 781	637 945	661 153	837 131	442 839	659 718									
3	B C	C D	A C	B D	A D	A B							A D		
	926 563	873 611	194 228	798 478	184 278	478 924									
4	A D	B D	B C	C D	A C	A B							A D		
	455 857	764 452	975 815	523 824	556 982	737 539									
5	A B	A C	A D	B C	B D	C D							A B		
	834 245	285 259	782 679	114 966	713 561	393 495									
6	A D	B D	B C	C D	A C	A B							A B		
	662 196	516 777	843 581	375 313	327 415	881 242									
7	A C	B C	A D	A B	B D	C D							A B		
	341 918	949 188	428 742	486 585	635 154	545 363									
8	C D	A B	A C	B D	A D	A B							A C		
	787 479	491 563	259 396	659 797	899 727	112 157									
9	A B	A C	A D	B C	B D	C D							A C		
	578 322	352 336	537 434	961 242	261 396	966 876									
10	A C	B C	A D	A B	B D	C D							A B		
	814 952	378 381	148 297	848 383	679 165	448 781									
11	A D	B D	B C	C D	A C	A B							A B		
	498 383	131 919	466 866	794 898	526 851	721 122									
12	B C	C D	A C	B D	A D	A B							A B		
	675 536	495 778	622 159	263 751	953 779	296 956									

圖 8.3 成對順位試驗用工作表：Friedman 分析。範例 8.3：玉米糖漿的口感

Multiple Paired Comparisons Test			
Name:	Date:		
Type of sample:	Unflavored table syrup		
and difference:	thickness (mouthfeel)		
Instructions:			
1.	Receive the sample tray and note each sample code below according to its position on the tray.		
2.	Taste the first sample pair from left to right and note which sample is thicker (more viscous). Indicate by placing an X next to the code.		
3.	Continue until all 6 pairs have been evaluated. Rinse with water as needed to clear your palate.		
Pair no.	Left sample	Right sample	Remarks
6			
5			
4			
3			
2			
1			
If you perceive no difference, please make a best guess. Comments regarding reasons for your choice or the characteristics of the samples may be made under Remarks.			

圖 8.4 成對順位試驗用評分表：Friedman 分析。範例 8.3：玉米糖漿的口感

樣品 B： $(12 + 6 + 2) + 2(0 + 6 + 10) = 52$;

欄	樣品	(較稀薄 thinner)	A	B	C	D	
列	樣品	(thicker 較濃稠)	A	—	0	1	0
			B	12	—	6	2
			C	11	6	—	7
			D	12	10	5	—

樣品 A： $(0 + 1 + 0) + 2(12 + 11 + 12) = 71$;

欄	樣品	(較稠薄 thinner)	A	B	C	D
Row samples (thicker)	A	—	0	1	0	0
	B	12	—	6	2	2
	C	11	6	—	7	7
	D	12	10	5	—	—

樣品 C： $(11 + 6 + 7) + 2(1 + 6 + 5) = 48$;

欄	樣品	(thinner)	A	B	C	D
Row samples (thicker)	A	—	0	1	0	0
	B	12	—	6	2	2
	C	11	6	—	7	7
	D	12	10	5	—	—

分析結果 — 在下表，顯示每列的樣品被選為比起每欄的樣品口感更稠些之頻率次數(12 位受試者之中)。例如，當樣品 B 同樣品 D 一起呈現，有 12 位受試者中的 2 位感知樣品 B 較稠些。

Row Samples (Thicker)	Column Samples (Thinner)			
	A	B	C	D
A	—	0	1	0
B	12	—	6	2
C	11	6	—	7
D	12	10	5	—

Friedman 分析(Friedman, 1937; Hollander and Wolfe, 1973)的第一步驟是：計算每個樣品的順位和(rank sum)，目前的舉例，順位 1 指派給“較稠的 thicker”與順位 2 給“較稀的 thinner”樣品。將列頻率的和加上欄頻率的和乘以 2，即得順位和。例如

樣品 D： $(12+10+5) + 2(0+2+7) = 45$

欄	樣品	(thinner)	A	B	C	D
Row samples (thicker)	A	—	0	1	0	0
	B	12	—	6	2	2
	C	11	6	—	7	7
	D	12	10	5	—	—

樣品	A	B	C	D
順位和	71	52	48	45

試驗統計數值，Friedman's T 計算如下述：

$$T = (4/pt) \sum_{i=1}^t R^2 - (9p[t-1])^2$$

$$= [4(12)(4)][71^2 + 52^2 + 48^2 + 45^2] - [9(12)(3^2)] = 34.17$$

p = 基本設計被重複的次數(本例= 12)； t = 處理的數目(本例= 4)；
 R_i = 第 i 次處理的順位和； $\sum R_i^2$ = 所有 R 平方的總和，從 R_1 至 R_t 。

T 臨界值已作成表(Skillings and Mack, 1981)，針對 $t = 3, 4, 5$ 、與小的 p 值；表中未列的實驗設計， T 值與自由度 $(t-1)$ 的卡方 (χ^2) (參見表 19.5) 做比較，在目前的舉例，臨界的 T_s 值是

顯著水準 α 0.10 0.05 0.01
 臨界的 T 6.25 7.81 11.3

結果顯示於濃稠相對稀薄的順位和尺度上：

樣品 A B C D
 順位和 71 52 48 45



在同樣的尺度，比較兩個順位和之 HSD 值(Turkey's honestly significant difference; 參見 14.5.7 項) ($\alpha = 0.05$) 如下：

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

值 $Q_{\alpha,t,\infty}$ 是來自表 19.4。樣品 A 與 B 間的差異遠大於 12.6，即樣品 A 是顯著地較稀薄，因此比起樣品 B、C 及 D 這群，是更合意的。

表 19.5 卡方 χ^2 分布的上部機率點(輸入為 $\chi^2_{\alpha, \nu}$)



In the present case, the critical T_s are
 Level of significance, α 0.10 0.05 0.01
 Critical T 6.25 7.81 11.3

Instructions: (1) Enter the row of the table corresponding to the number of degrees of freedom (ν) for χ^2 .
 (2) Pick the value of χ^2 in that row, from the column that corresponds to the predetermined α -level.

ν	0.995	0.990	0.975	0.950	0.900	0.750	0.500	0.250	0.100	0.050	0.025	0.010	0.005
1	0.000393	0.000157	0.000982	0.00393	0.0158	0.102	0.455	1.32	2.71	3.84	5.02	6.63	7.88
2	0.0100	0.0201	0.0506	0.103	0.211	0.575	1.39	2.77	4.61	5.99	7.38	9.21	10.6
3	0.0717	0.115	0.216	0.352	0.584	1.21	2.37	4.11	6.25	7.81	9.35	11.3	12.8
4	0.207	0.297	0.484	0.711	1.06	1.92	3.36	5.63	9.24	11.1	12.8	15.1	16.7
5	0.412	0.554	0.831	1.24	1.64	2.67	4.35	6.63	9.24	11.1	12.8	15.1	16.7
6	0.676	0.872	1.24	1.64	2.20	3.45	5.35	7.84	10.6	12.6	14.4	16.8	18.5
7	0.989	1.24	1.69	2.17	2.83	4.25	6.35	9.04	12.0	14.1	16.0	18.5	20.3
8	1.34	1.65	2.18	2.73	3.49	5.07	7.34	10.2	13.4	15.5	17.5	20.1	22.0
9	1.73	2.09	2.70	3.33	4.17	5.78	8.14	11.4	14.7	16.9	19.0	21.7	23.6
10	2.15	2.54	3.17	3.88	4.74	6.36	9.34	12.6	15.9	18.3	20.5	23.2	25.2
11	2.60	3.05	3.82	4.57	5.58	7.58	10.3	13.7	17.3	19.7	21.9	24.7	26.8
12	3.07	3.57	4.40	5.23	6.30	8.44	11.3	14.8	18.5	21.0	23.3	26.2	28.3
13	3.57	4.11	5.01	5.89	7.04	9.30	12.3	16.0	19.8	22.4	24.7	27.7	29.8
14	4.07	4.66	5.63	6.57	7.79	10.2	13.3	17.1	21.1	23.3	26.1	29.1	31.3
15	4.60	5.23	6.26	7.26	8.55	11.0	4.3	18.2	22.3	25.0	27.5	30.6	32.8
16	5.14	5.81	6.91	7.96	9.31	11.9	15.3	19.4	23.5	26.3	28.8	31.3	34.3
17	5.70	6.41	7.56	8.67	10.1	12.8	16.3	20.5	24.8	27.6	30.2	33.4	35.7

列(V) 對應來自 ANOVA 的誤差的自由度

Table 19.4 Percentage Points of the Studentized Range: Upper- α Critical Values for Tukey's HSD Multiple Comparison

Procedure Instructions:
 (1) Enter the section of the table that corresponds to the predetermined α -level.
 (2) Enter the row that corresponds to the degrees of freedom for error from the ANOVA.
 (3) Pick the value of r in that row from the column that corresponds to the number of treatments being compared.

ν	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
The entries are $q_{\alpha, r, \nu}$, where $p(\beta < q_{\alpha, r, \nu}) = 0.99$																			
1	90.05	135.0	164.3	185.6	202.2	215.8	227.2	237.0	245.6	253.2	260.0	266.2	271.8	277.0	281.8	286.3	290.4	294.3	298.0
2	14.04	19.02	22.29	24.72	26.63	28.29	29.53	30.68	31.69	32.59	33.40	34.13	34.81	35.43	36.00	36.53	37.03	37.50	37.95
3	8.26	10.62	12.17	13.33	14.24	15.00	15.64	16.20	16.69	17.13	17.53	17.89	18.22	18.52	18.81	19.07	19.32	19.55	19.77
4	6.51	8.12	9.17	9.96	10.58	11.10	11.55	11.93	12.27	12.57	12.84	13.09	13.32	13.53	13.73	13.91	14.08	14.24	14.40
5	5.70	6.98	7.80	8.42	8.91	9.32	9.67	9.97	10.24	10.48	10.70	10.89	11.08	11.24	11.40	11.55	11.68	11.81	11.93
6	5.24	6.33	7.03	7.56	7.97	8.32	8.61	8.87	9.10	9.30	9.48	9.65	9.81	9.95	10.08	10.21	10.32	10.43	10.54
7	4.95	5.92	6.54	7.01	7.37	7.68	7.94	8.17	8.37	8.55	8.71	8.86	9.00	9.12	9.24	9.35	9.46	9.55	9.65
8	4.75	5.64	6.20	6.62	6.96	7.24	7.47	7.68	7.86	8.03	8.18	8.31	8.44	8.56	8.66	8.76	8.85	8.94	9.03
9	4.60	5.43	5.96	6.35	6.66	6.91	7.13	7.33	7.49	7.65	7.78	7.91	8.03	8.13	8.23	8.33	8.41	8.49	8.57
10	4.48	5.27	5.77	6.14	6.43	6.67	6.87	7.05	7.21	7.36	7.49	7.60	7.71	7.81	7.91	7.99	8.08	8.15	8.23
11	4.39	5.15	5.62	5.97	6.25	6.48	6.67	6.84	6.99	7.13	7.25	7.36	7.46	7.56	7.65	7.73	7.81	7.88	7.95
12	4.32	5.05	5.50	5.84	6.10	6.32	6.51	6.67	6.81	6.94	7.06	7.17	7.26	7.36	7.44	7.52	7.59	7.66	7.73
13	4.26	4.96	5.40	5.72	5.98	6.19	6.37	6.52	6.67	6.79	6.90	7.01	7.10	7.19	7.27	7.35	7.42	7.48	7.55
14	4.21	4.89	5.32	5.63	5.88	6.08	6.26	6.41	6.54	6.66	6.77	6.87	6.96	7.05	7.13	7.20	7.27	7.33	7.39
15	4.17	4.84	5.25	5.56	5.80	5.99	6.16	6.31	6.44	6.55	6.66	6.76	6.84	6.92	7.00	7.07	7.14	7.20	7.26
16	4.13	4.79	5.19	5.49	5.72	5.92	6.08	6.22	6.35	6.46	6.56	6.66	6.74	6.82	6.90	6.97	7.03	7.09	7.15
17	4.10	4.74	5.14	5.43	5.66	5.85	6.01	6.15	6.27	6.38	6.48	6.57	6.66	6.73	6.81	6.87	6.94	7.00	7.05
18	4.07	4.70	5.09	5.38	5.60	5.79	5.94	6.08	6.20	6.31	6.41	6.50	6.58	6.65	6.73	6.79	6.85	6.91	6.97
19	4.05	4.67	5.05	5.33	5.55	5.73	5.89	6.02	6.14	6.25	6.34	6.43	6.51	6.58	6.65	6.72	6.78	6.84	6.89
20	4.02	4.64	5.02	5.29	5.51	5.69	5.84	5.97	6.09	6.19	6.28	6.37	6.45	6.52	6.59	6.65	6.71	6.77	6.82
24	3.96	4.55	4.91	5.17	5.37	5.54	5.69	5.81	5.92	6.02	6.11	6.19	6.26	6.33	6.39	6.45	6.51	6.56	6.61
30	3.89	4.45	4.80	5.04	5.24	5.40	5.54	5.65	5.76	5.85	5.93	6.01	6.08	6.14	6.20	6.26	6.31	6.36	6.41
40	3.82	4.37	4.70	4.93	5.11	5.26	5.39	5.50	5.60	5.69	5.76	5.83	5.90	5.96	6.02	6.07	6.12	6.16	6.21
60	3.76	4.28	4.59	4.82	4.99	5.13	5.25	5.36	5.45	5.53	5.60	5.67	5.73	5.78	5.84	5.89	5.93	5.97	6.01
120	3.70	4.20	4.50	4.71	4.87	5.01	5.12	5.21	5.30	5.38	5.44	5.50	5.56	5.61	5.66	5.71	5.75	5.79	5.83
∞	3.64	4.12	4.40	4.60	4.76	4.88	4.99	5.08	5.16	5.23	5.29	5.35	5.40	5.45	5.49	5.54	5.57	5.61	5.65

The entries are $t_{\alpha, df}$ where $P(t < t_{\alpha, df}) = 0.95$

1	17.97	26.98	32.82	37.08	40.41	43.12	45.40	47.36	49.07	50.59	51.96	53.20	54.33	55.32	56.32	57.22	58.04	58.83	59.56
2	6.08	8.33	9.80	10.88	11.74	12.44	13.03	13.54	13.99	14.39	14.75	15.08	15.38	15.65	15.91	16.14	16.37	16.57	16.77
3	4.50	5.91	6.82	7.50	8.04	8.48	8.85	9.18	9.46	9.72	9.95	10.15	10.35	10.53	10.69	10.84	10.98	11.11	11.24
4	3.93	5.04	5.76	6.29	6.71	7.05	7.35	7.60	7.83	8.03	8.21	8.37	8.52	8.66	8.79	8.91	9.03	9.13	9.23
5	3.64	4.60	5.22	5.67	6.03	6.33	6.58	6.80	6.99	7.17	7.32	7.47	7.60	7.72	7.83	7.93	8.03	8.12	8.21
6	3.46	4.34	4.90	5.30	5.63	5.90	6.12	6.32	6.49	6.65	6.79	6.92	7.03	7.14	7.24	7.34	7.43	7.51	7.59
7	3.34	4.16	4.68	5.06	5.36	5.61	5.82	6.00	6.16	6.30	6.43	6.55	6.66	6.76	6.85	6.94	7.02	7.10	7.17
8	3.26	4.04	4.53	4.89	5.17	5.40	5.60	5.77	5.92	6.05	6.18	6.29	6.39	6.48	6.57	6.65	6.73	6.80	6.87
9	3.20	3.95	4.41	4.76	5.02	5.24	5.43	5.59	5.74	5.87	5.98	6.09	6.19	6.28	6.36	6.44	6.51	6.58	6.64
10	3.15	3.88	4.33	4.65	4.91	5.12	5.30	5.46	5.60	5.72	5.83	5.93	6.03	6.11	6.19	6.27	6.34	6.40	6.47
11	3.11	3.82	4.26	4.57	4.82	5.03	5.20	5.35	5.49	5.61	5.71	5.80	5.89	6.00	6.07	6.15	6.22	6.28	6.33
12	3.08	3.77	4.20	4.51	4.75	4.95	5.12	5.27	5.39	5.51	5.61	5.70	5.78	5.86	5.95	6.02	6.09	6.15	6.21
13	3.06	3.73	4.15	4.45	4.69	4.88	5.05	5.19	5.32	5.43	5.53	5.63	5.71	5.79	5.86	5.93	5.99	6.05	6.11
14	3.03	3.70	4.11	4.41	4.64	4.83	4.99	5.13	5.25	5.36	5.46	5.55	5.64	5.71	5.79	5.85	5.91	5.97	6.03
15	3.01	3.67	4.08	4.37	4.59	4.78	4.94	5.08	5.20	5.31	5.40	5.49	5.57	5.65	5.72	5.78	5.85	5.90	5.96
16	3.00	3.65	4.05	4.33	4.56	4.74	4.90	5.03	5.15	5.26	5.35	5.44	5.52	5.59	5.66	5.73	5.79	5.84	5.90
17	2.98	3.63	4.02	4.30	4.52	4.70	4.86	4.99	5.11	5.21	5.31	5.39	5.47	5.54	5.60	5.67	5.73	5.79	5.84
18	2.97	3.61	4.00	4.28	4.49	4.67	4.82	4.96	5.07	5.17	5.27	5.35	5.43	5.50	5.57	5.63	5.69	5.74	5.79
19	2.96	3.59	3.98	4.25	4.47	4.65	4.79	4.92	5.04	5.14	5.23	5.31	5.39	5.46	5.53	5.59	5.65	5.70	5.75
20	2.95	3.58	3.96	4.23	4.45	4.62	4.77	4.90	5.01	5.11	5.20	5.28	5.36	5.43	5.49	5.55	5.61	5.66	5.71
24	2.92	3.53	3.90	4.17	4.37	4.54	4.68	4.81	4.92	5.01	5.10	5.18	5.25	5.32	5.38	5.44	5.49	5.55	5.59
30	2.89	3.49	3.85	4.10	4.30	4.46	4.60	4.72	4.82	4.92	5.00	5.08	5.15	5.21	5.27	5.33	5.38	5.43	5.47
40	2.86	3.44	3.79	4.04	4.23	4.39	4.52	4.63	4.73	4.82	4.90	4.98	5.04	5.11	5.16	5.22	5.27	5.31	5.36
60	2.83	3.40	3.74	3.98	4.16	4.31	4.44	4.55	4.65	4.73	4.81	4.88	4.94	5.00	5.06	5.11	5.15	5.20	5.24
120	2.80	3.36	3.68	3.92	4.10	4.24	4.36	4.47	4.56	4.64	4.71	4.78	4.84	4.90	4.95	5.00	5.04	5.09	5.13
∞	2.77	3.31	3.63	3.86	4.03	4.17	4.29	4.39	4.47	4.55	4.62	4.68	4.74	4.80	4.85	4.89	4.93	4.97	5.01

8.5 前言：多重樣品 (Multiple) 差異試驗 - 區集設計 (Block designs)

在 8.1 至 8.4 項中所述的試驗，是根據一項選定的屬性來處理樣品的成對比較 (pairwise comparison)，在下列四項的試驗則以超過 2 個樣品的群為根據，同樣就選定的一項屬性 (譬如甜味、鮮度或喜好) 比較，並使用 14.5.2 項討論的區集設計。

8.5.1 完全區集設計 Complete block designs

- 最簡單的設計是同時排序 (rank) 樣品全部 (參見 8.6 項)，但這樣的結果不像更複雜的試驗那樣的精確或可行的。
- 其次最簡單的是 一起比較全部樣品，使用評級尺度 (rating scale)。
- 我們能比較所有的樣品在一個完全區集 (complete block) (8.7 項，多重樣品差異試驗 Multisample Difference test) 或限制味蕾的負擔 (或其它感官器官) 與品評員的短期記憶，透過 將比較分割成幾個小的區集 (平衡不完全區集 balanced incomplete block [BIB] 設計，8.8 及 8.9 項)。

Table 19.4 (Continued) Percentage Points of the Studentized Range: Upper- α Critical Values for Tukey's HSD Multiple Comparison Procedure Instructions:
 (1) Enter the section of the table that corresponds to the predetermined α -level.
 (2) Enter the row that corresponds to the degrees of freedom for error from the ANOVA.
 (3) Pick the value of q in that row from the column that corresponds to the number of treatments being compared.

The entries are $q_{\alpha, k, df}$ where $P(q < q_{\alpha, k, df}) = 0.95$

1	8.93	13.44	16.36	18.69	20.15	21.51	22.64	23.62	24.48	25.24	25.92	26.54	27.10	27.62	28.10	28.54	28.96	29.35	29.71
2	4.13	5.73	6.77	7.54	8.14	8.63	9.05	9.41	9.72	10.01	10.26	10.49	10.70	10.89	11.07	11.24	11.39	11.54	11.68
3	3.33	4.47	5.20	5.74	6.16	6.51	6.81	7.06	7.29	7.49	7.67	7.83	7.98	8.12	8.25	8.37	8.48	8.58	8.68
4	3.01	3.98	4.59	5.03	5.39	5.68	5.93	6.14	6.33	6.49	6.65	6.78	6.91	7.02	7.13	7.23	7.33	7.41	7.50
5	2.85	3.72	4.26	4.66	4.98	5.24	5.46	5.65	5.82	5.97	6.10	6.22	6.34	6.44	6.54	6.63	6.71	6.79	6.86
6	2.75	3.56	4.07	4.44	4.73	4.97	5.14	5.28	5.41	5.53	5.64	5.76	5.87	5.98	6.07	6.16	6.25	6.30	6.47
7	2.68	3.45	3.93	4.28	4.55	4.78	4.97	5.14	5.28	5.41	5.53	5.64	5.74	5.83	5.91	5.99	6.06	6.13	6.19
8	2.63	3.37	3.83	4.17	4.43	4.65	4.83	4.99	5.13	5.25	5.36	5.46	5.56	5.64	5.72	5.80	5.87	5.93	6.00
9	2.59	3.32	3.76	4.08	4.34	4.54	4.72	4.87	5.01	5.13	5.23	5.33	5.42	5.51	5.58	5.66	5.72	5.79	5.85
10	2.56	3.27	3.70	4.02	4.26	4.47	4.64	4.78	4.91	5.03	5.13	5.23	5.32	5.40	5.47	5.54	5.61	5.67	5.73
11	2.54	3.23	3.66	3.96	4.20	4.40	4.57	4.71	4.84	4.95	5.05	5.15	5.23	5.31	5.38	5.45	5.51	5.57	5.63
12	2.52	3.20	3.62	3.92	4.16	4.35	4.51	4.65	4.78	4.89	4.99	5.08	5.16	5.24	5.31	5.37	5.44	5.49	5.55
13	2.50	3.18	3.59	3.88	4.12	4.30	4.46	4.60	4.72	4.83	4.93	5.02	5.10	5.18	5.25	5.31	5.37	5.43	5.48
14	2.49	3.16	3.56	3.85	4.08	4.27	4.42	4.56	4.68	4.79	4.88	4.97	5.05	5.12	5.19	5.26	5.32	5.37	5.43
15	2.48	3.14	3.54	3.83	4.05	4.23	4.39	4.52	4.64	4.75	4.84	4.93	5.01	5.08	5.15	5.21	5.27	5.32	5.38
16	2.47	3.12	3.52	3.80	4.03	4.21	4.36	4.49	4.61	4.71	4.81	4.89	4.97	5.04	5.11	5.17	5.23	5.28	5.33
17	2.46	3.11	3.50	3.78	4.00	4.18	4.33	4.46	4.58	4.68	4.77	4.86	4.93	5.01	5.07	5.13	5.19	5.24	5.30
18	2.45	3.10	3.49	3.77	3.98	4.16	4.31	4.44	4.55	4.65	4.75	4.83	4.90	4.98	5.04	5.10	5.16	5.21	5.26
19	2.45	3.09	3.47	3.75	3.97	4.14	4.29	4.42	4.53	4.63	4.72	4.80	4.88	4.95	5.01	5.07	5.13	5.18	5.23
20	2.44	3.08	3.46	3.74	3.95	4.12	4.27	4.40	4.51	4.61	4.70	4.78	4.85	4.92	4.99	5.05	5.10	5.16	5.20
24	2.42	3.05	3.42	3.69	3.90	4.07	4.21	4.34	4.44	4.54	4.63	4.71	4.78	4.85	4.91	4.97	5.02	5.07	5.12
30	2.40	3.02	3.39	3.65	3.85	4.02	4.16	4.28	4.38	4.47	4.56	4.64	4.71	4.77	4.83	4.89	4.94	4.99	5.03
40	2.38	2.99	3.35	3.60	3.80	3.96	4.10	4.21	4.32	4.41	4.49	4.56	4.63	4.69	4.75	4.81	4.86	4.90	4.95
60	2.36	2.96	3.31	3.56	3.75	3.91	4.04	4.14	4.25	4.34	4.42	4.49	4.56	4.62	4.67	4.73	4.78	4.82	4.86
120	2.34	2.93	3.28	3.52	3.71	3.86	3.99	4.10	4.19	4.28	4.35	4.42	4.48	4.54	4.60	4.65	4.69	4.74	4.78
∞	2.33	2.90	3.24	3.48	3.66	3.81	3.93	4.04	4.13	4.21	4.28	4.35	4.41	4.47	4.52	4.57	4.61	4.65	4.69

8.5.2 平衡不完全區集設計 balanced incomplete block [BIB] designs

在完全區集設計，每區集 (列) 的大小等於探討的處理數量。目前的語意，一個區集被認定為提供給一位品評員的樣品組 (set of samples)，通常，品評員在一次會期中，不能品評超過 4 至 6 個樣品。如果比較的樣品 (處理 treatments) 數較大，例如 6 至 12 個，就可使用 BIB 設計。

取代呈現所有的 t 個樣品在一大區集，實驗者改呈現在 b 個小區集中，每個小區集含有 k 個樣品，而 $k < t$ 。

每個小區集的 k 個樣品須做挑選，因此所有樣品都被品評同樣的次數，以及所有的樣品配對組也同樣的次數一起出現在 b 個小區集中。

Cochran and Cox (1957) 提供大多數試驗情況下都能使用的各種 BIB 設計。

8.6 單純順位試驗 Simple Ranking Test：Friedman 分析

— 隨機(完全)區集設計

8.6.1 範圍與應用

當試驗目標是根據單項的屬性，譬如甜味、鮮度、喜好等，來比較幾個樣品。

做這樣的比較，**排序/順位(Ranking)**是最簡單的方式，但其數據僅是**順序的(ordinal)**，從每次回應得到的差異程度也無測量。連續的樣品都差異大，以及那些差異小的，也只用一格順位單位分隔。

順位是時間花費較少的，比起其它的方法，且特別適用於當樣品未了後面的分析，而要預先分類或篩選。

8.6.2 試驗原理

以平衡的隨機順序呈現給每位受試者**樣品組(set of samples)**，要求他們去**排序 rank** 樣品，根據感興趣的屬性。計算**順位和 rank sums**，以 **Friedman's test**，14.5.3.2 項介紹，做統計評估。

8.6.3 試驗受試者

挑選、訓練與指導受試者，如 3.2 及 3.3 項所述。使用不低於 8 位的受試者，如 16 位或更多可用，辨識的效率會大為提升。受試者可能需要特別的指導或訓練，以使他們能夠重複地識別感興趣的屬性 (參見 10.3.1.2 項)，取決於**試驗目標**，受試者可能要基於**驗證能力**做挑選，即**檢測屬性中的小差異**。

8.6.4 試驗程序步驟

- 試驗控制 Test controls 和產品控制 product controls，參照 7.3.3 項。
- 可能的話，同時呈現樣品，否則相繼提出。
- 給受試者平衡的隨機順序的 **t 個樣品的組**，重新安排樣品為 **排序/等級順序 rank order**。
- 樣品組可只呈現一次或幾次但其編碼不同。如果樣品組可呈現二或更多次，準確度可大幅提升。
- 在**喜好試驗(preference tests)**，指導受試者如何指派順位 1 給喜好的樣品，順位 2 給其次喜好的樣品，等等。針對強度試驗，指導受試者指派順位 1 給最低強度者，順位 2 給下一個最低者等。

建議：受試者根據每個樣品的第一次審理，安排樣品的臨時順序，然後根據進一步的測試來**驗證或更改順序**。

指導受試者對相鄰的樣品，做“**最好的猜測**”，即使它們似是相同的；但是，如果受試者拒絕猜測，他必須在“”註釋 Comments”處指出：認為樣品是同樣的。為了統計分析，指派相同的樣品的每個給予**平均順位(average rank)**，例如，在 4 個樣品的試驗，如一位品評員 無法區別兩個中間的樣品，指派 2.5 的平均順位給每個樣品，亦即 $(2 + 3)/2$ 。

如相同樣品組有需要其等級順序超過一項屬性，**每項屬性個別地進行程序**，使用**新的樣品編碼**，因而一次的品評不會影響下一次。評分卷列於圖 8.5。

範例 8.5：四種甜味劑的持續性比較

問題/情況 — 某心理學家的研究室想要比較 4 種人工甜味劑 A、B、C、及 D 的甜味持續性程度 (degree of persistence)。

計畫/試驗目標 — 確定有否甜味劑在吞嚥後口裡的甜味持續性是否有顯著的差異。

試驗設計 — 持續性的感覺可能因人而異而顯示大的變動，因此，需要人數較多的品評小組執行。順位試驗是適合的，因做起來簡單，不需要過多的訓練。這四種樣品以 48 位學生測試，每位給予 4 個樣品，是以三位數編碼並平衡的隨機順序提供。評分卷列於圖 8.5。

圖 8.5 單純順位試驗的評分表。範例 8.5：四種甜味料持續性的比較

Ranking Test			
Name: _____	Date: _____		
Type of sample: Artificial sweeteners			
Characteristic studied: Persistence of sweet taste			
Instructions			
1. Receive the sample tray and note each sample code below according to its position on the tray.			
2. Taste the samples from left to right and note the degree of persistence of the sweetness			
Wait at least 30 seconds between samples and rinse palate as required.			
3. Write "1" in the box of the sample which you find least persistent			
Write "2" for the next, "3" for the next, and "4" for the most persistent			
You may find it expedient to first arrange the samples in a provisional order, and then resolve the positions of adjacent samples by more careful tasting.			
4. If two samples appear the same, make a "best guess" as to their rank order.			
Code	_____	_____	_____
Rank	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Comments:	_____		

篩選樣品 — 此試驗需要很小心地準備，以確保 4 種化合物之間沒有超出哪些預期以外的差異，即不同的化學組成所導致的那些差異。四位有經驗的品嘗員品評與調整樣品，確使對平均的觀察者都是同等的甜味，以及在溫度、黏度、外觀(顏色、濁度與泡沫等)等的任何差異都不存在或已遮蔽，因而排除辨識是透過味覺與嗅覺以外的手段。

分析結果 — 表 8.1 顯示結果如何編排與順位和如何計算。

表 8.1 範例 8.5 的結果。四種甜味劑比較其持續性

Subject No.	Sample A	Sample B	Sample C	Sample D
1	3	1	4	2
2	3	2	4	1
3	3	1	4	1
4	3	1	4	2
5	1	3	2	4
—	—	—	—	—
—	—	—	—	—
44	4	2	3	1
45	3	4	1	2
46	3	4	1	2
47	4	1	2	3
48	4	2	3	1
Rank sum	135	103	137	105

公式 14.14 之試驗統計量 T 值為

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

參照表 19.5，看到自由度=3 的卡方(χ^2)其上部 5%臨界值為 7.81。由於 $T = 12.85$ 大於 7.81，結論：樣品是在 5%水準下，顯著地差異於甜味的持續性。

確定哪些樣品是顯著地不同，依公式 14.15 計算多重比較的臨界值：

$$\begin{aligned} LSD_{\text{rank}} &= 1.96\sqrt{48(4)(5)/6} \\ &= 24.8 \end{aligned}$$

任何兩個樣品其順位和的差距大於 $LSD_{\text{rank}} = 24.8$ ，即是顯著地不同在 5%水準下。因此，結論：樣品 B 與 D 兩者顯示顯著地較少甜味持續性比起樣品 A 及 C 兩者。而樣品 B 與 D 並無顯著地差異，樣品 A 與 C 之間也是。

Sample A	Sample B	Sample C	Sample D
135	103	137	105

表 19.5，卡方分佈(χ^2 -distribution) 的上部- α 機率點(輸入為 $\chi^2_{\alpha, v}$)



Instructions: (1) Enter the row of the table corresponding to the number of degrees of freedom (v) for χ^2 .
 (2) Pick the value of χ^2 in that row, from the column that corresponds to the predetermined α -level.

α	0.995	0.990	0.975	0.950	0.900	0.750	0.500	0.250	0.100	0.050	0.025	0.010	0.005
1	0.0000993	0.000157	0.000982	0.00393	0.0158	0.102	0.455	1.32	2.71	3.84	5.02	6.63	7.88
2	0.000157	0.000241	0.00158	0.00636	0.0271	0.216	0.711	2.00	3.00	3.84	4.61	5.41	6.33
3	0.000241	0.000357	0.00234	0.00935	0.0384	0.352	1.06	2.78	4.11	4.75	5.59	6.25	7.16
4	0.000357	0.000540	0.00393	0.0135	0.0538	0.554	1.32	3.36	4.61	5.14	5.99	6.58	7.38
5	0.000540	0.000811	0.00636	0.0200	0.0771	0.853	1.75	4.75	5.41	5.99	6.75	7.38	8.15
6	0.000811	0.00124	0.0104	0.0293	0.115	1.24	2.67	5.35	6.25	6.75	7.58	8.15	8.93
7	0.00124	0.00188	0.0158	0.0417	0.164	1.64	3.45	6.35	7.04	7.58	8.38	8.93	9.70
8	0.00188	0.00283	0.0234	0.0599	0.228	2.28	4.25	6.35	7.04	7.58	8.38	8.93	9.70
9	0.00283	0.00424	0.0357	0.0871	0.338	3.38	5.41	6.35	7.04	7.58	8.38	8.93	9.70
10	0.00424	0.00636	0.0540	0.124	0.500	5.00	6.35	6.35	7.04	7.58	8.38	8.93	9.70
11	0.00636	0.00935	0.0811	0.175	0.711	7.11	6.35	6.35	7.04	7.58	8.38	8.93	9.70
12	0.00935	0.0135	0.0135	0.241	1.00	10.0	6.35	6.35	7.04	7.58	8.38	8.93	9.70
13	0.0135	0.0200	0.0200	0.338	1.41	14.1	6.35	6.35	7.04	7.58	8.38	8.93	9.70
14	0.0200	0.0293	0.0293	0.475	1.94	19.4	6.35	6.35	7.04	7.58	8.38	8.93	9.70
15	0.0293	0.0417	0.0417	0.636	2.60	26.0	6.35	6.35	7.04	7.58	8.38	8.93	9.70
16	0.0417	0.0599	0.0599	0.853	3.57	35.7	6.35	6.35	7.04	7.58	8.38	8.93	9.70
17	0.0599	0.0811	0.0811	1.15	4.84	48.4	6.35	6.35	7.04	7.58	8.38	8.93	9.70

範例 8.6：啤酒苦味與分析不一致

問題/情況 — 某釀酒商的品管經理得知公司的品牌 P 和競爭者的啤酒花苦味物質，以標準分析方法測得的數值都一樣，然而他聽到公司啤酒嘗起來比較苦。在開始調查可能受非啤酒花的苦味物質污染之前，他希望確認存在感知的苦味有所差異。

計畫/試驗目標 — 試喝啤酒 P 的苦味，相對於競爭品牌 A、B、及 C。

試驗設計 — 以 12 位證實有測出苦味的小差異之能力的受試者排序四個樣品。

虛無假設為 H_0 ：苦味 P = 苦味 A、B、或 C；

另一假設為 H_a ：苦味 P \neq 苦味 A、B、或 C，關於 A、B 及 C 之間任何系統的差異都無事前的訊息。
 評分卷列在圖 8.5。

分析結果 — 參見表 8.2。注意：經驗型的品評員允許指派同樣的順位或“並列 ties”給樣品。

表 8.2 範例 8.6 結果表：啤酒苦味與分析不一致

Subject No.	Sample A	Sample B	Sample C	Sample P
1	1	2.5	2.5	4
2	2	1	4	3
3	1	3	3	3
4	2	1	3	4
5	2	3	1	4
6	1	1	4	3
7	3	1	2	4
8	1	2	3.5	3.5
9	2	3	4	1
10	2	1	3.5	3.5
11	2	3	1	4
12	2	1	4	3
Rank sum	22	22.5	35.5	40

當數據中出現並列(ties)，就必須使用另一形式的試驗統計量 T' (公式 14.16)。

計算 T' 值，首先要確定每區集(i)中的並列組數量(g_i)與每並列組的大小(t_{ij})。(每個非並列樣品被認為是一個大小 $t_{ij} = 1$ 的個別組)，只存在並列的區集需考慮，因為這些區集影響了 T' 的計算。

根據表 8.2，我們看到 並列 出現在區集 1、3、8 及 8。這些區集的 g_i 及 t_{ij} 值為

$$g_1 = 3, t_{1,1} = 1 \quad g_3 = 2, t_{3,1} = 1 \quad g_8 = 3, t_{8,1} = 1 \quad g_{10} = 3, t_{10,1} = 1$$

$$t_{1,2} = 2 \quad t_{3,2} = 3 \quad t_{8,2} = 1 \quad t_{10,2} = 1$$

$$t_{1,3} = 1 \quad t_{8,3} = 2 \quad t_{10,3} = 2$$

這些的值得用來計算公式 14.16 中 T' 的分母第二項為

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

T' 值 = 13.3，大於自由度 3 卡方(χ^2)上部 5%臨界值($\chi^2_{0.05,3} = 7.81$)，結論：樣品間有顯度的差異存在。

只有樣品 A、B 及 C 對比樣品 P 的比較是感興趣的，因此，採用適合於順位數據之比較試驗樣品與控制組或標準樣品之多重比較步驟 (multiple comparison procedure) (參見 Hollander and Wolfe, 1973)。

此多重比較的上部 5% (單尾的)臨界值是 13.1，樣品 P 的順位和比起樣品 A 及 B 的順位和大於 13.1。

試驗報告 — 品質保證經理結論：公司樣品 P 是比競爭者啤酒 A 與 B 顯著地更苦，因此，他開始著手一項調查，可能的汙染來自額外的苦味物質。

8.7 多重樣品差異試驗：評級(Rating)方法 — 以變異數分析 Analysis of Variance (ANOVA)評估

8.7.1 範圍與應用

當試驗目標要確定一項特定的感官屬性在 t 個樣品中如何變化，使用本方法， t 從 3 至 6 或至多 8，以及可能以一大組比較所有 t 個樣品。

註：在描述性分析(參見第 11 章)，當幾個樣品要比較，目前的方法可應用於每項屬性。

8.7.2 試驗原理

受試者評級(rate)所選定屬性的強度，在數字的強度尺度(numerical intensity scale)上，例如區間尺度(category scale) (參見 5.6.1 項)，並具體指定採用的尺度，以變異數分析評定結果。

8.7.3 試驗受試者

- 挑選、訓練及指導受試者如 7.3.3 項中所述。
- 採用不要低於 8 位的受試者；如果 16 位或更多，辨識效率大為提升。
- 受試者可能需要特別的指導，以使他們能夠再現地識別感興趣的屬性(參見 10.3 項)。
- 取決於試驗目標，受試者可能要挑選那些在屬性上具有高辨識能力的人。

8.7.4 試驗步驟

- 試驗控制(test controls)與產品(product controls)，參照 3.2 及 3.3 項。
- 可能的話，同時提出樣品，或先後依序地。
- 給受試者平衡隨機順序的一組 t 個樣品，使用具體指定的尺度來評級每樣品。
- 該組樣品可以只呈現一次，或呈現幾次但編碼不同。如能呈現二次或更多，準確度會大幅提升。

➤ 如果要評級一項屬性以上，理論上，樣品應針對每項屬性而分別呈現。在實際的描述性分析，因為要評級樣品的屬性的數量大抵 6 至 26 項，就不可能是上述的方式。

在沒需要個別評級每項的屬性，感官分析專家認為屬性之間會有一些相互依賴關係。例如，如果在貯藏壽命的研究，產品可能受微生物的作用而變陳化(stale)(譬如酸味)，或受氧化的作用(譬如油耗味 rancidity)，在某項的評級高時，也會提升在它項的評級，即使並不存在。這樣的影響必須透過讓受試者了解，以及強力的訓練使他們能獨立地認識每項屬性。

8.7.5 結果分析與解釋

結果以變異數分析之；參見 14.5.2.1 項。

範例 8.7：感官分析課程的普及性。隨機完全區集設計

問題/試驗目標 – 食品科學系照例在學期末要學生評級上課科目，採用 -3 至 +3 尺度，-3 = 很差，0 = 普通，+3 = 優良。30 位學生完成評分，結果列表 8.3，評鑑目的是要確認哪些科目需要改進。

分析結果 – 這數據適合隨機完全區集設計的變異數分析，學生視為“區集 blocks”，而評鑑科目是“處理 treatments”。

表 8.3 範例 8.7 的結果：多重差異試驗(評級)

Student No.	Courses Evaluated			Statistics
	Biology	Nutrition	Sensory	
1	2	-2	1	1
2	3	0	2	1
3	1	-3	0	0
4	1	0	1	0
5	0	1	0	0
6	-3	-3	-3	-3
7	-1	-3	-1	-1
8	-1	-2	-1	-1
9	2	-2	-1	-1
10	1	0	1	1
11	1	0	2	2
12	-1	-2	0	1
13	3	-3	3	3
14	0	0	0	0
15	-2	-2	-1	-1
16	1	-2	1	0
17	1	-1	0	0
18	0	-1	0	-1
19	3	3	3	3
20	1	-2	1	0
21	-2	-2	-2	-2
22	2	1	1	1
23	3	0	3	3
24	1	-1	1	1
25	1	1	1	1
26	0	-1	0	-1
27	1	0	2	0
28	2	-2	0	0
29	-2	-3	-1	-2
30	2	2	2	2

Note: Scale used: -3 to +3, where -3 = very poor, 0 = indifferent, +3 = excellent.

在表 8.4，評鑑科目的 F -統計量是高顯著的($F_{3,87} = 12.91, p < 0.0001$)，因此評鑑專家結論：這些課程的平均回應中是有所差異。專家並執行 LSD 多重比較程序以確定何者課程的平均值是彼此顯著不同(參見表 8.4 底部)，最後顯示：營養課程有顯著較低(較差)的平均評級分數比起其它三科，而其它三科的平均評級分數之中並無另外的顯著差異。這課程評鑑專家將結果同教授與系溝通，決定進一步的行動。

表 8.4 表 8.3 中結果的隨機完全區集設計：食品科學課程的普及性

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square	F	p
Total Students (blocks)	119	344.37			
Courses evaluated	29	188.87	15.97	12.91	<0.0001
Error	87	47.9	1.24		
Average ratings for the items evaluated with the 95% LSD multiple-comparison results					
Courses Evaluated	Biology	Nutrition	Sensory	Statistics	
Mean rating	0.80a	-0.83b	0.60a	0.30a	

Note: Mean ratings not followed by the same letter are significantly different at the 95% confidence level. $LSD_{0.05} = 0.57$

範例 8.8：五種啤酒的啤酒花特性 — 裂區設計(split block design)

問題/情況 — 某釀造商正生產一款啤酒新品牌，強調啤酒花特性，使用不同批次的啤酒花，其成本分別為美金 1.00、1.20、1.40、1.60、及 1.80 元/磅。

計畫目標 — 以成本考量，選擇最多啤酒花特性之批次。

試驗目標 — 比較最後 5 種啤酒的啤酒花特性程度；得到結果的可靠性之量度。

試驗設計 — 邏輯的方式是將五種啤酒排列在人數夠大的品嚐者面前，所以這是典型的多重差異試驗；20 位受試者在 0 至 9 的尺度上評定樣品，採用的評分卷列於圖 8.6。呈現的順序要隨機化，樣品以不同的編碼呈現在分開的三場次。

篩選樣品 — 兩位有經驗的品嚐者先品評樣品，確定是所要生產啤酒類型的代表，以及除了啤酒花特性，並無會干擾屬性的其它感官上的差異。

分析結果 — 品評結果列在表 8.5，對應的裂區 ANOVA 列在表 8.6。

表 8.5 範例 8.8 的結果。多重樣品差異試驗(評級)：五種啤酒的啤酒花特性

Sample No.	1	2	3	4	5
1	2.21	3.45	1.02	5.43	3.24
2	0.01	1.21	0.00	2.12	2.11
3	0.21	2.02	0.20	2.32	0.22
4	3.33	4.56	2.31	5.84	5.64
5	2.43	4.51	3.03	3.56	1.43
6	1.01	1.21	0.00	0.21	2.11
7	6.43	4.63	3.46	4.63	3.46
8	2.22	3.33	0.11	4.65	3.53
9	1.43	2.53	2.02	5.45	5.23
10	1.43	1.34	3.03	6.53	3.41
11	3.42	1.21	0.00	1.21	1.12
12	1.00	1.21	0.00	2.12	1.12
13	1.00	1.21	0.00	4.85	4.65
14	3.33	6.54	1.32	1.10	2.43
15	2.22	5.33	1.10	5.64	2.43
16	3.42	3.52	2.02	5.45	3.25
17	1.21	2.02	0.20	2.32	2.20
18	1.20	5.43	2.01	3.45	4.23
19	3.46	3.64	6.43	3.64	6.43
20	3.46	3.64	6.43	3.64	6.43

Explanation: e.g., Subject no. 20 rated sample no. 1 a "3" the first time, a "4" the second time, and a "6" the third time.

Multisample Comparisons Test	
Name:	Date:
Type of sample: Beer	
Characteristic studied: Hop character	
Instructions	
Taste the samples from left to right and note the intensity of the characteristic studied. Rate each sample on the following scale:	
0	Imperceptible
1	
2	Slightly perceptible
3	
4	Moderately perceptible
5	
6	Strongly perceptible
7	
8	Extremely perceptible
9	
Sample Code:	
Rating:	
Comments:	

圖 8.6 多重樣品差異試驗(評級)用評分卷。範例 8.8：五種啤酒的啤酒花特性

表 8.6 表 8.5 中結果的裂區 ANOVA：五種啤酒的啤酒花特性

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Squares	F
Total	299	975.64		
Replications	2	8.89		
Samples	4	221.52	55.38	41.88 ^a
Error(A)	8	10.58	1.32	
Subjects	19	412.30	21.70	17.79 ^a
Sample×Subject	76	89.81	1.18	0.97
Error(B)	190	232.53	1.22	

Note: Error(A) is calculated as would be the Rep×Sample interaction. Error(B) is calculated by subtraction.

^a Significant at the 1% level.

sample x subject 的交互作用是不顯著的：

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

sample 影響 與 subject 影響 兩者都高度顯著的：

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

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

由於交互作用不顯著，可以假定受試者在樣品的評級是一致的，但是，受試者影響的顯著性則表明受試者使用尺度的不同部分來表達他們的感知，這並不罕見。再者，當無交互作用，subject-to-subject 差異通常是其次的興趣所在。

樣本之間的差異是主要關心的問題，確定何者樣品顯著地不同於平均的啤酒花特性，使用 HSD 多重比較程序來比較樣品，如下述：

Sample	4	2	5	1	3
Mean	3.9	3.0	2.9	2.1	1.4

註：平均值無底線相連，代表是顯著地不同在 5%顯著水準。

$$HSD_{5\%} = q_{0.05,5,8} \sqrt{\frac{MS_{\text{Error}}}{n}} = 4.89 \sqrt{\frac{1.32}{60}} = 0.7 \text{ (q-value from Table 19.4)}$$

樣品 4 比起所有其它樣品都顯著地較大，樣品 2 與 5、平均評級幾乎相同，比起樣品 4 有顯著地較低的啤酒花特性，但比起樣品 1 及 3 則又顯著地較高，樣品 1 與 3 比起樣品 2、4 及 5 都顯著地較低的啤酒花特性。

解釋與報告結果 — 感官分析專家給釀酒商的報告包括樣品平均值表與 ANOVA 表，結論：試驗的五種啤酒中，樣品 4 產生顯著地較高的啤酒花特性，樣品 2 其次，也值得考慮。

8.8 多重樣品差異試驗：BIB 順位試驗(平衡不完全區集設計 **Balanced Incomplete Block Design**) — Friedman 分析

8.8.1 範圍與應用

當試驗目標是要確定某特定的感官屬性在多個樣品中以何種方式變化，以及任何一個時間都有過多的樣品要品評，使用本方法。常態上，這方法適用在要品評的樣品從 6 至 12 個或最多 16 個。

選擇目前的方法(順位)，是當品評員是對樣品類型是相當地位受訓練，以及/或偏好較簡單的統計分析，而採用 8.7 項(評級)的方法，當品評員可被訓練使用一項評級尺度是可行的。

8.8.2 試驗原理

取代所有 t 個樣品是以一大區集呈現，依據 Cochran and Cox (1957) 的實驗設計的其中一種，將它們呈現在一些較小的區集，或依據例如 Design Express (2003)。要求受試者按照所興趣的屬性來排序樣品。

8.8.3 試驗受試者

挑選、訓練及指導受試者如 7.3.3 項所述。查明受試者能夠認識所興趣的屬性，例如以已知屬性其強度不同的樣品組來訓練(參見 10.3 項與附錄 12.2)。

8.8.4 試驗程序步驟

- 試驗控制(test controls)與產品控制(product controls)，參見 3.2 及 3.3 項。
- 可以的話，樣品同時呈現，否則先後順序提出。
- 程序的細節參閱 7.3.4 項。
- 確定呈現的順序是真正隨機的；不讓受試者懷疑有規律的樣式，因這會影響及裁決。例如，只聲明“根據甜味排序樣品，最低甜味的樣品給順位 1，其次低的給順位 2 等等”。

範例 8.9：魚的種類

問題/情況 — 以往從 15 種魚製備了軍用口糧 XPQ-6 (魚條)，抱怨魚腥味的問題之後，已查明是出自某些魚種，因而想要指定限制的魚種，以便能夠權衡可利用性與成本相對於食物動亂的機率。

計畫目標 — 比較 15 種魚，取得魚腥味程度上的定量資料而應用於手中之問題。

試驗目標 — 比較 15 種魚所製成野外口糧的魚腥味程度。

表 8.7 多重樣品差異試驗：BIB 設計為範例 8.9 — 野外口糧 ($t = 15, k = 3, r = 7, b = 35, l = 1, E = 0.71$)

(1)	1	2	3
(2)	4	10	15
(3)	5	9	14
(4)	6	11	13
(5)	7	8	12
(6)	1	10	15
(7)	2	9	14
(8)	3	11	13
(9)	4	8	12
(10)	1	10	15
(11)	2	9	14
(12)	3	11	13
(13)	4	8	12
(14)	1	10	15
(15)	2	9	14
(16)	3	11	13
(17)	4	8	12
(18)	1	10	15
(19)	2	9	14
(20)	3	11	13
(21)	4	8	12
(22)	1	10	15
(23)	2	9	14
(24)	3	11	13
(25)	4	8	12
(26)	1	10	15
(27)	2	9	14
(28)	3	11	13
(29)	4	8	12
(30)	1	10	15
(31)	2	9	14
(32)	3	11	13
(33)	4	8	12
(34)	1	10	15
(35)	2	9	14

Source: From W.G. Cochran and G.M. Cox. 1957. *Experimental Designs*. New York: Wiley. With permission.

試驗設計 — 選擇平衡不完全設計 (balanced incomplete design) 的 **多重樣品差異試驗**，是因為 15 試驗產品的比較允許分散在 3 個組。隨機挑選的 105 位士官隨機分為 35 組，每組含 3 位受試者，按表 8.7 之設計，受試者每組隨機指派給 35 組樣品中的一組，在詢問卷中，要求受試者排序 (rank) 它的 3 個樣品，根據魚腥味從最少的 (= 1) 至最強的 (= 3)。

篩選樣品 — 在廚師的幫助下，就質地、外觀和風味上，盡可能一致地製備樣品，通過適當變化烹飪方法和其它組成分，讓歸因於魚種的差異最小化。每次提供的魚塊都篩選外觀，任何帶有大塊碎片或顯示其它看到偏差的都丟棄不用。

分析結果 — 讓結果較容易做分析，順位數據整理列於表 8.8。

表 8.8 範例 8.9 的結果，多重樣品差異試驗：用順位數據 — 野外的 BIB 設計

Block/Subject	Sample/Species														
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	1	2	3												
2				3			1			2					
3								3			2				2
4									3			1			
5										3					2
6		3			2	1					1				
7											3				1
8												3			2
9													2		1
10														1	3
—															
—															
—															
—															
101															
102		2													
103															
104															
105															
Rank sum	35	45	54	43	28	37	55	42	37	50	49	50	34	42	29

Note: Response 1, least fishy; 2, intermediate; 3, most fishy.

所給定魚種的**順位和** 就是對應該魚種的欄中所有數字的總和。
 計算 **Friedman's 試驗統計量 T 值**(參見公式 14.19)來確定如果魚種中魚腥味強度是有任何的差異。**T 值 = 68.53**，大於 χ^2 其自由度 $(t - 1) = 14$ 之上部 5%臨界值 $\chi^2_{14,0.05} = 23.69$ ，故結論：在數據組中，是的確顯著的種類是顯著地不同(見表 8.9)。

解釋與報告結果 — 從表 8.9 的結果，軍方領導結論：為價位與可利用性的考量，確定**樣品 5、15、13、1、6、及 9 之魚種**應該保留，因為這些產生最低程度的魚腥味，且彼此之間並無顯著地不同。

8.9 多重樣品比較試驗：BIB 評級試驗(平衡不完全區集設計)

— 以變異數分析評定

8.9.1 範圍與應用

當試驗目標是要確定某特定感官屬性在多量的樣品間是如何方式變化，且任何時候要品評太多的樣品，採用本方法。典型上會採用本方法是當要比較的樣品數是從 6 至 12 或最多 16 個。

選目前的方法(評級 **rating**)，是當品評員訓練使用評級尺度，是可行的且結果需要盡可能精確的與可行的。採用 8.8 項(順位)所述的方法是當品評員受較少的訓練及/或順位試驗已經可提供足夠的訊息。

註：在描述性分析(見第 11 章)，當要比較的樣品數量大時，可應用目前的方法去品評每項屬性。

樣品 14、8 及 4 之魚種臨時保留，以防由於高成本與供應困難而刪除過多第一群內的魚種，這樣做是認識到此事實，即樣品 14、8 及 4 其魚腥味強度的順位和是顯著地大於樣品 5 及 15，且和第一群內的其餘的樣品是並無顯著地不同，表 8.9 中的其餘樣品(2、11、10、12、3、及 7) 被排除用在野外口糧 XPQ-6。

表 8.9 表 8.8：魚條口糧其數據的統計分析與結果整理

Sample/Species	Rank Sum
5	28
15	29
13	34
1	35
6	37
9	37
14	42
8	42
4	43
11	45
10	49
12	50
3	54
7	55

Note: Means followed by the same letter are not significantly different at the 5% significance level (LSD_{rank} = 10.74).

8.9.2 試驗原理

取代所有 t 個樣品以一大區集呈現，依據 Cochran and Cox (1957) 的實驗設計的其中一種，改呈現在一些較小的區集，或依據例如 Design Express (2003)。要求受試者按照所興趣的屬性的**數字強度尺度**上(參見 5.6 項)評級樣品。明確說明要用的尺度，以變異數分析評定結果。

8.9.3 試驗受試者

挑選、訓練及指導受試者如同 7.3.3 項所述。確認受試者能夠認識所興趣的屬性，例如將屬性以已知的各種強度組來訓練，採用不低於 8 位受試者；如有 16 位或更多則辨識效率大幅提升。

受試者可能需要特別的指導，使得他們再現地認識所興趣的屬性(參見 10.3.1.2 項)，取決於試驗目標，受試者可能經挑選那些對興趣所在的屬性具有**高辨識能力者**。

8.9.4 試驗程序步驟

- 試驗控制(test controls)與產品控制(product controls)參照 3.2 及 3.3 項。
- 可以的話，樣品同時呈現，否則先後順序提出。
- 程序的細節請參閱 7.3.4 項。
- 確定呈現的順序是真正隨機的；不會讓受試者懷疑有規律的樣式，因這會影響及裁決。

註：如果超過一項的屬性要評級，不可避免的，這在所得的評級將會有些互相依存性(參見 4.3.5 項)。

8.9.5 結果的分析與解釋

結果採用變異數分析做解析(參見 14.5.4.1 項和範例 8.10)。

試驗設計 — 以 15 位受過良好訓練的品評 6 批次樣品的整體性不良風味，採用從 0 = 無不良風味至 9 = 非常明顯的不良風味之 10-點區間尺度(10-point category scale)。

- 品評員不能一次品評超過 4 個樣品，因而感官分析專家選擇來自 Cochran and Cox (1957)的 BIB 設計(參見表 8.10)。
- 15 位品評員的每位隨機指派一個從設計的 4 樣品區集。
- 每區集內樣品的呈現順序要隨機化。

範例 8.10 冰淇淋的參考樣品

問題/情況 — 屬進行中計畫的一部分，冰淇淋工廠品管經理定期篩選成品樣品，以便選擇一些批次將加入品質參考樣品(reference samples)之中，為了主要 QC 測試計畫中使用。定期都需要新的參考樣品，因較久的樣品會隨時間而變化，不再適合，且這程序也用來消除其中可能已變質的任何當前的參考樣品。

計畫目標 — 為 QC 測試之目的地，而維持成品冰淇淋之夠量的參考樣品庫存。

試驗目標 — 每天評級(rate) 6 批次的庫存，針對整體性不良風味，並丟棄不適合當作參考品的批次。

表 11 範例 8.10：冰淇淋參考樣品的 BIB 設計($t = 6, k = 4, r = 10, b = 15, l = 6, E = 0.90$)

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

Source: From W.G. Cochran and G.M. Cox, 1957. *Experimental Designs*, New York: Wiley. With permission.

分析結果：對整體性不良味道的評級數據，如表 8.11 所示。

表 11 範例 8.10：冰淇淋參考樣品的評級結果

Block/Subject	Sample					
	1	2	3	4	5	6
1	6	1	1	2	3	3
2	6	4	2	1	5	2
3	7	2	3	2	2	1
4	3	5	1	1	3	2
5	3	4	4	1	1	3
6	7	4	1	1	1	3
7	2	2	2	2	2	3
8	4	2	2	2	5	1
9	5	3	2	1	1	2
10	4	2	2	2	1	1
11	5	3	2	1	1	1
12	4	2	2	2	2	1
13	5	2	4	5	3	1
14	5.0	2.5	2.2	2.0	2.6	1.9
Adjusted means						

Note: 1: BIB design with rating; 2: Response—10-point category scale with 0 = no off-flavor, 9 = extreme off-flavor; 3: Adjusted means that are not connected by a common underscore are significantly different at the 5% significance level (LSD_{5%} = 1.1).

應用 **LSD** (最小顯著差異; Least Significant Difference) 多重比較程序 (multiple comparison procedure) 至樣品的平均評級，確定哪些樣品具有顯著不同的整體不良風味(參見表 8.11 註腳的註 3)。

解釋與報告結果 — 樣品 1 的平均不良味道評級是顯著地大於其餘的樣品，而其餘的樣品之間並無顯著的差異。感官分析專家向 QC 經理報告結果，附帶建議：樣品 1 的批次要取出丟棄。

數據採用能執行 **balanced-incomplete-block ANOVA** 的電腦程式 (參見 14.5.4.1 項)。得到的 BIB ANOVA 表，如表 8.12 所示。

表 8.12 範例 8.10：冰淇淋參考樣品之平衡不完全區集 ANOVA 表

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square	F	P
Total	59	150.98			
Judges (blocks)	14	39.73			
Samples (treatments, adjusted for blocks)	5	59.89	11.98	9.33	<0.0001
Error	40	51.36	1.28		

處理(treatments)的 F-統計量(即冰淇淋樣品)，當和 $(t - 1) = 5$ 及 $(tpr - t - pb + 1) = 40$ 自由度的 F-分佈(F-distribution)的上部 5%臨界值比較，發現是：顯著的 $(F = 9.33 > F_{0.05,5,40} = 2.45)$ 。