ANOVA

變異數分析原理說明

檢定三個以上的獨立母體之平均值是否相等時，可採用變異數分析(The **An**alysis **of** **Va**riance, **ANOVA**)。因此，ANOVA是用來當成三個或三個以上的母群體平均數的差異顯著性考驗工具。變異數分析種類繁多，如下表：

依變數個數	自變數個數	名稱
1（單變數變異數分析）	1	單因子變異數分析
	2（以上）	多因子變異數分析
2（以上）（多變數變異數分析）	1	單因子多變量分析
	2（以上）	多因子多變量分析

變異數分析的例子

- 變異數分析是用來檢定兩個以上平均數是否相等或某個變數是否受某些因子所影響之統計方法。

  例如：

  (1)不同的行銷策略是否會影響產品之銷售量？(不同的行銷策略，其產品之平均銷售量是否相等？)

  (2)不同的教育程度與不同的性別對工作滿意度是否有影響？(不同的教育程度與不同的性別之員工，其平均之工作滿意度是否相等)

變異數分析常用之名詞

(1)實驗單位(experiment unit)：實驗所衡量的對象。例如：產品、員工為其實驗單位。

(2)因子(factor)：研究者所控制調整的因素。例如：行銷策略、教育程度為其因子。

變異數分析常用之名詞

(3)處理方法(treatment)：因子之各種水準或類別。例如：不同的行銷策略、不同的教育程度、不同的性別，如不同性別中的男、女為兩種不同的處理方法。

(4)依變數(dependent variable)：實驗單位對不同處理方法的反應變數。例如：銷售量、工作滿意度為其依變數。

- 在多樣本比較的情況下，我們可以進行一連串兩個樣本間平均數之t-test的檢定，如果有四個樣本（從另一個角度來說，是一個有四個類別之自變項，如宗教信仰，則每個類別自為一個sub-sample），則我們可進行六個不同之兩個樣本間的t-test。如果真是這樣做，除了非常麻煩外，最大的缺點是會增加犯Type I 錯誤之機率。如果每個t-test是定在α＝0.05之水準下進行檢定，進行一連串這樣的t-tests會使犯下至少一次Type I error的機會增加。換言之，即使每一個t-test是在α＝0.05之水準下進行檢定，其Type I error綜合起來事實上是大於0.05。換個角度來說，t-test做多了，總有一個t-test之結果會reject H_0，但此H_0可能為真。用ANOVA來分析就可以避免這樣的問題。

變異數分析的邏輯

✓ **ANOVA**之虛無假設H_0是$\mu_1 = \mu_2 = \mu_3 = ... = \mu_k$，也就是所有樣本均是來自同一母群體，或是各母群體的平均數之間無差異。更具體的說法是，每類別或項目之間在某一特質上並無差異（例如：不同宗教信仰者在支持死刑之態度上並無差異）。從這H_0之形式可看變異數分析是兩樣本之間**t-test**的延伸。至於說H_1則為「<u>至少</u>有一類別在某一特質上與其它類別有差異」。

- ✓ 如果上述之 H_0 為真，則每類別樣本平均數之差別應不大，且各樣本之標準差大小差不多。事實上ANOVA並不是問不同類別間是否有差異，而是問這些差異是否大到可以拒絕 H_0。
- ✓ 和 H_0 完全相反的情況是各類別之平均數相差極大，而各類別之標準差很小。換言之，各類別內之異質性很小，而類別間異質性很大。在這種情況下，如果我們將所有樣本合併，這個合併後之樣本的變異量(Variance)　，主要來自原來樣本和樣本間之差異。換言之，此合併後樣本之變異或離散之狀況主要源自原來各樣本間之差異。而 H_0 所假設的情況，是變異量主要是來自原各樣本（類別）內之差異，而非各樣本間之差異。

- 瞭解上面的敘述後，就很容易了解ANOVA之原理，ANOVA之檢定是建立在比較各類別（或樣本）間之變異量及各類別內之變異量。與類別內之變異量相比較下，當類別間之變異量愈大時，拒絕 H_0 的可能性愈大，反之，則愈小。
- ANOVA之公式，即在比較兩種對母群體之變異量（ σ^2 ）之估計值。其一估計值即是建立在各樣本內之變化，而另一則為各樣本間之變化。這即是ANOVA（ANalysis Of VAriance）之名稱的由來。

變異數分析的種類

- 單因子變異數分析**(One-way ANOVA)**

--旨在比較單一種自變項的不同處理方式對某依變項的影響。

- 雙因子變異數分析(Two-way ANOVA)

--考慮兩個因子對依變項可能產生的影響

Ex: 不同教育程度對男女薪資的影響：

第一類因子為教育程度，第二類是性別，因此不僅僅看教育程度對薪資的影響，更探討教育程度對男生，對女生的薪資影響。

	教育程度		
性別	國中	高中	大學
男	3.3	3.6	4.3
	2.8	3.2	3.9
	3.1	3.5	4.1
女	2.3	2.5	2.6
	2.2	2.3	3.1
	1.9	2.7	2.7

- 三因子變異數分析…等等

→ 凡是雙因子以上的變異數分析皆通稱為多因子變異數分析

- 以下，就以此例子來讓大家更了解ANOVA：

→某位老師想了解，講述法，啟發式教學法，電腦輔助教學法對五年級學生的英語成績是否有幫助？

1. 此為單因子變異數分析:只探討不同教學法對於單一群體的英語成績影響

→ 自變項為教學法，依變項為英語成績

- →某位老師想了解，講述法，啟發式教學法，電腦輔助教學法對五年級男女學生的英語成績是否有幫助？

- 1. 此為雙因子變異數分析：

→不僅僅可以比較不同教學法對學生英語成績的影響，亦加入了對男女生英語成績的比較。 在此，自變項有兩個因子(教學法與性別)，而依變項為學生英語成績。

- **完全隨機設計(completely randomized design)**
  是指研究者將不同的處理方法以隨機方式分派給實驗單位。例如，若研究不同的教學方法對學生學習成績是否有影響，則可由n個學生為實驗單位，然後將n個學生隨機分派於不同的教學方法，最後再記錄學生之學習成績。

定理1

$$\sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(x_{ij}-\bar{x}\right)^2=\sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(x_{ij}-\bar{x}_i\right)^2+\sum_{i=1}^{k}n_i\left(\bar{x}_i-\bar{x}\right)^2$$

單因子變異數分析—完全隨機設計

平方和

$$SST = \sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(x_{ij} - \bar{x}\right)^2$$ = 總變異(total sum of squares)

$$SSE = \sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(x_{ij} - \bar{x}_i\right)^2$$ = 組內變異(error sum of squares)

$$SSB = \sum_{i=1}^{k} n_i\left(\bar{x}_i - \bar{x}\right)^2$$ = 組間變異(sum of squares between)

- **SSW(SSE)(Sum of Squares Within)（組內離均差平方和）之公式是**

$$SSW = \sum (X_i - \overline{X}_k)^2$$

$\overline{X}_k$ **是每類別或組別之平均數**

- **因此，我們求SSW之方法是將各組每一分數減去此組之平均數，求其平方，然後加起來，每組都這麼做後，要全部加起來即得SSW。**

✓ 而**SSB (Sum of Squares Between)** 之公式為

$$SSB = \sum N_k (\bar{X}_k - \bar{X})^2$$

N_k 是各組之樣本數

$\bar{X}_k$ 是各組之平均

$\bar{X}$ 是合併樣本之平均數

- ✓ 知道了SSW及SSB，我們可以得到兩種母群體之 σ^2 的估計值

組內估計值＝SSW／dfw，dfw＝N－K，

組間估計值＝SSB／dfb， dfb＝K－1

N＝全部合併樣本數，K＝組數

單因子變異數分析—完全隨機設計

定理2

若獨立隨機樣本$X_{ij} \sim N(\mu_i, \sigma^2)$，$i = 1, 2, \cdots, k$，$j = 1, 2, \cdots, n_i$，則

$$(1)\ E(SSE) = (n-k)\sigma^2$$

$$(2)\ E(SSB) = (n-1)\sigma^2 + \sum_{i=1}^{k} n_i(\mu_i - \mu^2)$$

單因子變異數分析─完全隨機設計

定理3

若獨立隨機樣本$X_{ij} \sim N(\mu_i, \sigma^2)$, $i = 1, 2, \cdots, k$, $j = 1, 2, \cdots, n_i$, 則

在 H_0 成立條件下，

$$F_0 = \frac{MSB}{MSE} \sim F(k-1, n-k)$$

單因子變異數分析—完全隨機設計

■ 由定理3得知，進行變異數分析需滿足以下基本假設條件：

(1)常態母體：各組樣本需取自於常態母體。

(2)變異數具同質性：各組母體變異數需假設相等。而變異數是否具同質性，可利用樣本變異數檢定之。

(3)獨立性：各組樣本彼此獨立。

單因子變異數分析—完全隨機設計

單因子變異數分析表

變異來源	平方和	自由度	均方	f_0 值
處理方法 隨機誤差	*SSB* *SSE*	*k*-1 *n*-*k*	*MSB* *MSE*	$\frac{MSB}{MSE}$
總和	*SST*	*n*-1		

- ✓ 而ANOVA即在求，兩估計值間的相對大小，更具體說是求一個F ratio。

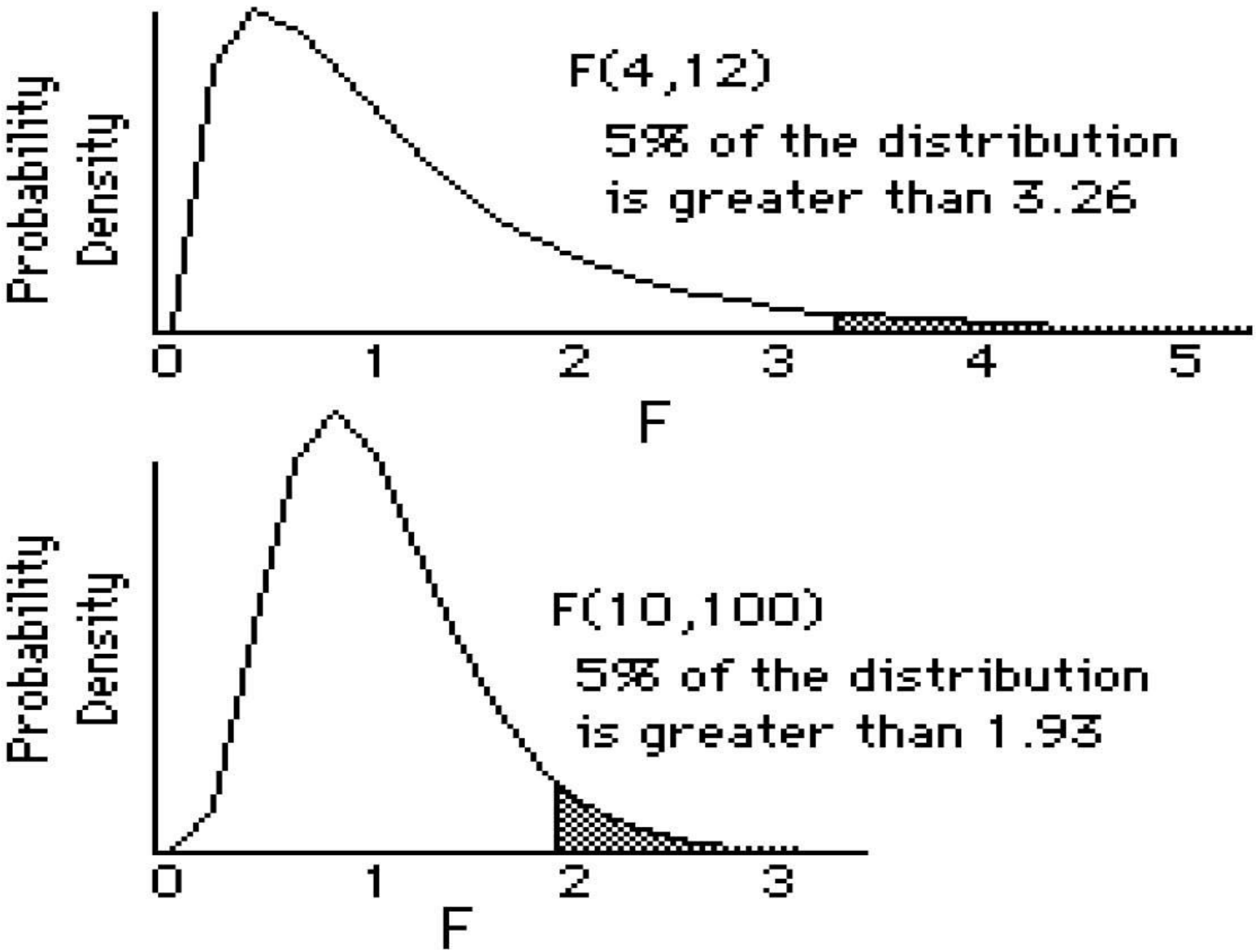
  F＝Mean square between／Mean square within

  ＝(SSB/dfb)／(SSW/dfw)

- ✓ 如果F＝0，即表示組間變異數為0，即各組平均數相同。

變異數分析時之F分配

- ✓ 這個F值之抽樣分配是隨dfb及dfw而變化，其分配之圖形如下：


F(4,12)
5% of the distribution
is greater than 3.26
Probability
Density
0
1
2
3
4
5
F
F(10,100)
5% of the distribution
is greater than 1.93
Probability
Density
0
1
2
3
F

變異數計算的捷徑

- 上述之計算公式為依原理所設計的，事實上我們有些捷徑可循，其中SST可用下式來算，

$$SST = \sum X^2 - N\overline{X}^2$$

- 此後用SSB之公式算出SSB後，以SST－SSB即得SSW。

$$SSW = SST - SSB$$

- 這樣計算可省不少事。

ANOVA檢定的各種侷限

- 此處所介紹之ANOVA，又叫做單因子ANOVA或簡單ANOVA(one-way ANOVA或Simple ANOVA)，這是因為我們只考慮一個自變項和一應變項之關係。ANOVA之應用可延伸到多個自變項與一個應變項之關係，在此暫不多說。
- ANOVA最大的限制是要用等距尺度及各類別之樣本數要接近。其次，ANOVA只能告訴我們樣本間之差異是否到了顯著水準，並不能告訴我們何類別或樣本與其它類別或樣本不同。

單因子變異數分析—完全隨機設計

例題 1

- 某市場調查公司欲調查市面上四種品牌之相同口味飲料之平均銷售量是否相同，於是由每一品牌隨機選定5個地區作調查，得其每個地區一個月之銷售量如下表(單位：千箱)

品牌			
A	B	C	D
26.5	29.0	26.9	30.5
28.7	27.6	28.3	31.2
25.2	25.4	27.8	29.9
29.3	28.3	26.2	28.1
25.3	29.7	25.8	30.3

單因子變異數分析—完全隨機設計

例題 1(續)

(1)請寫出此問題之假設。

(2)請寫出此問題之變異數分析表。

(3)請根據(2)之結果，以 $\alpha=0.05$ 檢定此四種品牌飲料之平均銷售量是否相等。

單因子變異數分析─完全隨機設計

解

(1)令 μ_i 表第 i 種品牌銷售量之平均數，則此問題之假設為

H_0: $\mu_A=\mu_B=\mu_C=\mu_D$ & H_1: μ_A、μ_B、μ_C、μ_D 不全相等

(2)每種品牌之樣本平均數 $\overline{X}_A$、$\overline{X}_B$、$\overline{X}_C$、$\overline{X}_D$ 及總樣本平均數 $\overline{X}$ 如下：

$$\overline{X}_A=\frac{1}{5}(26.5+28.7+25.2+29.3+25.3)=27$$

$$\overline{X}_B=\frac{1}{5}(29.0+27.6+25.4+28.3+29.7)=28$$

$$\overline{X}_C=\frac{1}{5}(26.9+28.3+27.8+26.2+25.3)=27$$

單因子變異數分析─完全隨機設計

解

$$\overline{X}_D = \frac{1}{5}(30.5+31.2+29.9+28.1+30.3)=30$$

$$\overline{X} = \frac{1}{20}(5\times 27+5\times 28+5\times 27+5\times 30)=28$$

經計算後可得

$$\begin{aligned} SST &= (26.5-28)^2+(28.7-28)^2+(25.2-28)^2+(29.3-28)^2 \\ &\quad +(25.3-28)^2+(29.0-28)^2+(27.6-28)^2+(25.4-28)^2 \\ &\quad +(28.3-28)^2+(29.7-28)^2+(26.9-28)^2+(28.3-28)^2 \\ &\quad +(27.8-28)^2+(26.2-28)^2+(25.8-28)^2+(30.5-28)^2 \\ &\quad +(31.2-28)^2+(29.9-28)^2+(28.1-28)^2+(30.3-28)^2 \\ &= 65.28 \end{aligned}$$

單因子變異數分析—完全隨機設計

解

$$SSB = 5\times(27-28)^2 + 5\times(28-28)^2 + 5\times(27-28)^2 + 5\times(30-28)^2 = 30$$

$$SSE = SST - SSB = 65.28 - 30 = 35.28$$

SSB之自由度為4-1=3，因此 $MSB = \dfrac{SSB}{k-1} = \dfrac{30}{3} = 10$

SSE之自由度為 $n-k$ = 20-4 = 16，因此

$$MSE = \frac{SSE}{n-k} = \frac{35.28}{16} = 2.205$$

由此可得， $f_0 = \dfrac{MSB}{MSE} = \dfrac{10}{2.205} = 4.535$

單因子變異數分析─完全隨機設計

解

其變異數分析表如下：

變異來源	平方和	自由度	均方	f_0 值
處理方法	30	3	10	4.535
隨機誤差	35.28	16	2.205	
總和	65.28	19		

(3)因為 $\dfrac{MSB}{MSE} \sim F(3,16)$，因此其拒絕域 $\{f_0 \geq f_{0.05}(3,16)\} = \{f_0 \geq 3.239\}$

而檢定值 $f_0 = 4.535 > 3.239$ 落在拒絕域中，因此拒絕 H_0，即四種不同品牌飲料之平均銷售量有顯著地差異。

單因子變異數分析—完全隨機設計

例題 2

某研究人員想瞭解*A*、*B*、*C*三種不同廠牌1800c.c.汽車之耗油率，於是此研究人員蒐集了資料，並以完全隨機設計方式蒐集資料並計算得到以下之變異數分析表，如下表所示。

變異來源	平方和	自由度	均方	f_0 值
處理方法	20	2	?	?
隨機誤差	50	27	?	
總和	70	29		

單因子變異數分析─完全隨機設計

例題 2(續)

(1)請完成此變異數分析表。

(2)請以 $\alpha=0.05$ 來檢定 $H_0: \mu_A=\mu_B=\mu_C$ (μ_i 表第 i 種品牌汽車平均每公升汽車可行駛之里程數)是否成立

單因子變異數分析─完全隨機設計

解

(1)因為 $MSB = \dfrac{SSB}{k-1} = \dfrac{20}{2} = 10$　　$MSE = \dfrac{SSE}{n-k} = \dfrac{50}{27} = 1.85$

所以 $f_0 = \dfrac{MSB}{MSE} = \dfrac{10}{1.85} = 5.41$

因此其完整變異數分析表如下所示：

變異來源	平方和	自由度	均方	f_0 值
處理方法	20	2	10	5.41
隨機誤差	50	27	1.85	
總和	70	29		

單因子變異數分析─完全隨機設計

解

(2)因為 $\dfrac{MSB}{MSE} \sim F(2,27)$，因此其拒絕域 $\{f_0 \geq f_{0.05}(2,27)=3.35\}$

而檢定值 $f_0 = 5.41 > 3.35$ 落在拒絕域中，因此拒絕 H_0，即三種品牌1800c.c.汽車之耗油率有顯著地差異。

變異數分析原理說明

- **範例：**

某校想要了解不同的教學方法對學生的學習成效是否有所差異，因而進行一項教學實驗。該校找來三組學生（每組各5位同學），施以不同的教學方法（民主式、專制式、放任式）。一段時日後施以測驗，測驗成績如右表所示。試問：此三種教學方法之成效是否有所差異？

教學方式 / 學生編號	民主式	專制式	放任式	
1	68	57	67	
2	69	60	67	
3	71	61	65	
4	67	62	64	
5	70	60	67	合計
行平均 ($\bar{X}_i$)	69	60	66	65($\bar{\bar{X}}$)
組間變異	80	125	5	210
組內變異	10	14	8	32
總變異	242			

變異數分析原理說明

總變異可分為兩部分，即組間變異（處理變異）與組內變異。以本範例來說明，每位學生測驗成績與總平均差異的來源，可分為兩大部分：一為來自教學方法所造成的差異（組間變異）；另一為來自學生個別差異（組內變異）。變異數分析的檢定統計量乃用*F*值來進行：

變異來源	公式 (SS)	平方和	自由度 (DF)		均方 $(MS) = SS/DF$	$F = MSB/MSW$
組間變異 (SSB)	$\sum_{i=1}^{k} n_i(\overline{X}_i - \overline{\overline{X}})^2$	210	$k-1$	2	$MSB = 105$	39.4
組內變異 (SSW)	$\sum_{i=1}^{k}\sum_{j=1}^{n}(X_{ij} - \overline{X}_i)^2$	32	$n-k$	12	$MSW = 2.7$	
總變異 (SST)	$\sum_{i=1}^{k}\sum_{j=1}^{n}(X_{ij} - \overline{\overline{X}})^2$	242	$n-1$	14	$MST = 17.3$	

變異數分析原理說明

- 假說與檢定
  - H_0：各教學法平均分數皆相等($\mu_1 = \mu_2 = \mu_3$)
  - H_1：各教學法平均分數不全相等

  檢定原理為若*F*值愈大，表示由於教學法不同所造成之變異值愈顯著，愈傾向拒絕 H_0。（範例一）
- 多重比較之探討

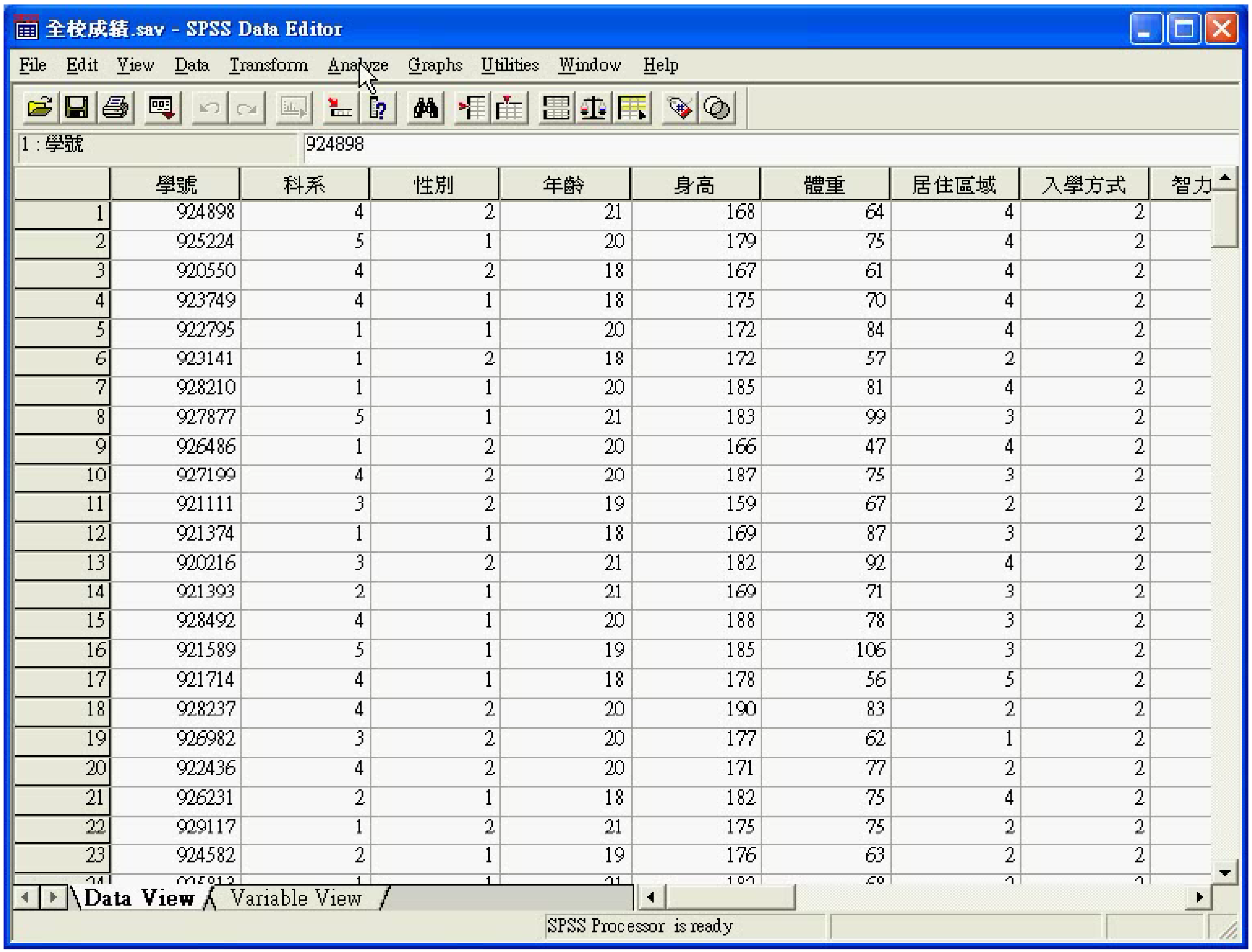
  在進行ANOVA比較之後，倘若結果顯示各組平均數間有顯著差異，則我們希望進一步了解哪一些平均數是不同的。此時各平均數間之比較組合不只一種，故稱多重比較(multiple comparision)。

範例一

以全校學生成績為例，探討不同科系之平均數學成績是否有差異。

- 操作：
  - 1.點選Analyze/Compare Means/One-Way ANOVA
  - 2.程式操作
  - 3.假說：
    - H_0：各科系數學平均分數皆相等($\mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$)
    - H_1：各科系數學平均分數不全相等
  - 4.分析結果

範例一—程式操作



	學號	科系	性別	年齡	身高	體重	居住區域	入學方式	智力
1	924898	4	2	21	168	64	4	2	
2	925224	5	1	20	179	75	4	2	
3	920550	4	2	18	167	61	4	2	
4	923749	4	1	18	175	70	4	2	
5	922795	1	1	20	172	84	4	2	
6	923141	1	2	18	172	57	2	2	
7	928210	1	1	20	185	81	4	2	
8	927877	5	1	21	183	99	3	2	
9	926486	1	2	20	166	47	4	2	
10	927199	4	2	20	187	75	3	2	
11	921111	3	2	19	159	67	2	2	
12	921374	1	1	18	169	87	3	2	
13	920216	3	2	21	182	92	4	2	
14	921393	2	1	21	169	71	3	2	
15	928492	4	1	20	188	78	3	2	
16	921589	5	1	19	185	106	3	2	
17	921714	4	1	18	178	56	5	2	
18	928237	4	2	20	190	83	2	2	
19	926982	3	2	20	177	62	1	2	
20	922436	4	2	20	171	77	2	2	
21	926231	2	1	18	182	75	4	2	
22	929117	1	2	21	175	75	2	2	
23	924582	2	1	19	176	63	2	2	

範例一—分析結果

- 判斷方法：p-value=Sig.=0.000<0.05，故拒絕 H_0
- 結論：科系間之平均數學分數有顯著差異 。

ANOVA

數學能力測驗

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	15703.89	4	3925.973	8.386	.000
Within Groups	105329.9	225	468.133		
Total	121033.8	229			

★多個平均數之多重比較

聯合信賴區間(變異數分析之多重比較)

在 k 個母體平均數之多重比較中，兩母體平均數差 $\mu_i-\mu_j$

之 $(1-\alpha)\times 100\%$ 聯合信賴區間為

$$\left[\ \bar{x}_i-\bar{x}_j-t_{\alpha/2a}(n-k)\sqrt{MSE\left(\frac{1}{n_i}+\frac{1}{n_j}\right)},\ \bar{x}_i-\bar{x}_j+t_{\alpha/2a}(n-k)\sqrt{MSE\left(\frac{1}{n_i}+\frac{1}{n_j}\right)}\ \right]$$

其中 $\bar{x}_i$、$\bar{x}_j$分別為兩母體之樣本平均數且其樣本個數分別為

n_i 及 n_j，且 $a=C_2^k$，n 表總樣本數。

★多個平均數之多重比較

例題 3

承例2，若A、B、C三種品牌1800c.c.汽車每公升汽油可行駛里程數之樣本平均數分別為 $\bar{x}_A=11$，$\bar{x}_B=12$，$\bar{x}_C=13$ 且樣本個數均為10，試求A、B、C三種品牌1800c.c.汽車平均每公升汽油可行駛之里程數差之94%聯合信賴區間。

★多個平均數之多重比較

解

由已知資料及前面計算得知，$\bar{x}_A = 11$，$\bar{x}_B = 12$，$\bar{x}_C = 13$；

$$n_A = n_B = n_C = 10 \text{；} k = 3 \text{；} a = C_2^k = C_3^k = 3 \text{；} MSE = 1.85$$

由此可知，$\mu_A - \mu_B$ 之信賴區間為

$$\bar{x}_A - \bar{x}_B \pm t_{\alpha/2a}(n-k)\sqrt{MSE\left(\frac{1}{n_A}+\frac{1}{n_B}\right)} = 11 - 12 \pm t_{0.01}(27)\sqrt{1.85\left(\frac{1}{10}+\frac{1}{10}\right)}$$

$$= -1 \pm 1.504$$

即 $\mu_A - \mu_B$ 之信賴區間為 $[-2.504,\ \ 0.504]$；

★多個平均數之多重比較

解

$\mu_B-\mu_C$ 之信賴區間為

$$\overline{x}_B-\overline{x}_C \pm t_{\alpha/2a}(n-k)\sqrt{MSE\left(\frac{1}{n_B}+\frac{1}{n_C}\right)}=11-12\pm 1.504=-1\pm 1.504$$

即 $\mu_B-\mu_C$ 之信賴區間為 $[-2.504,\ 0.504]$

最後，$\mu_A-\mu_C$ 之信賴區間為 $[-3.504,\ -0.496]$ 。

由此可知，發現 $\mu_A-\mu_B$ 及 $\mu_B-\mu_C$ 之信賴區間均包含0，而 $\mu_A-\mu_C$ 之信賴區間未包含0，因此 μ_A、μ_B 及 μ_B、μ_C 並無顯著地差異，而具有顯著地差異者僅有 μ_A　μ_C 。

變異數分析原理說明

- 各種多重比較方法之檢定整理如下表：

正交性	方法	統計量	各組樣本數	型I錯誤
正交	Fisher	t	可相同或不同	最高
非正交	Dunnett	t	可相同或不同	居中
	Bonferroni	t	可相同或不同	居中
	Scheffe	F	可相同或不同	最低
	HSD	q	需相同	較高
	Duncan's Method	q	需相同	居中

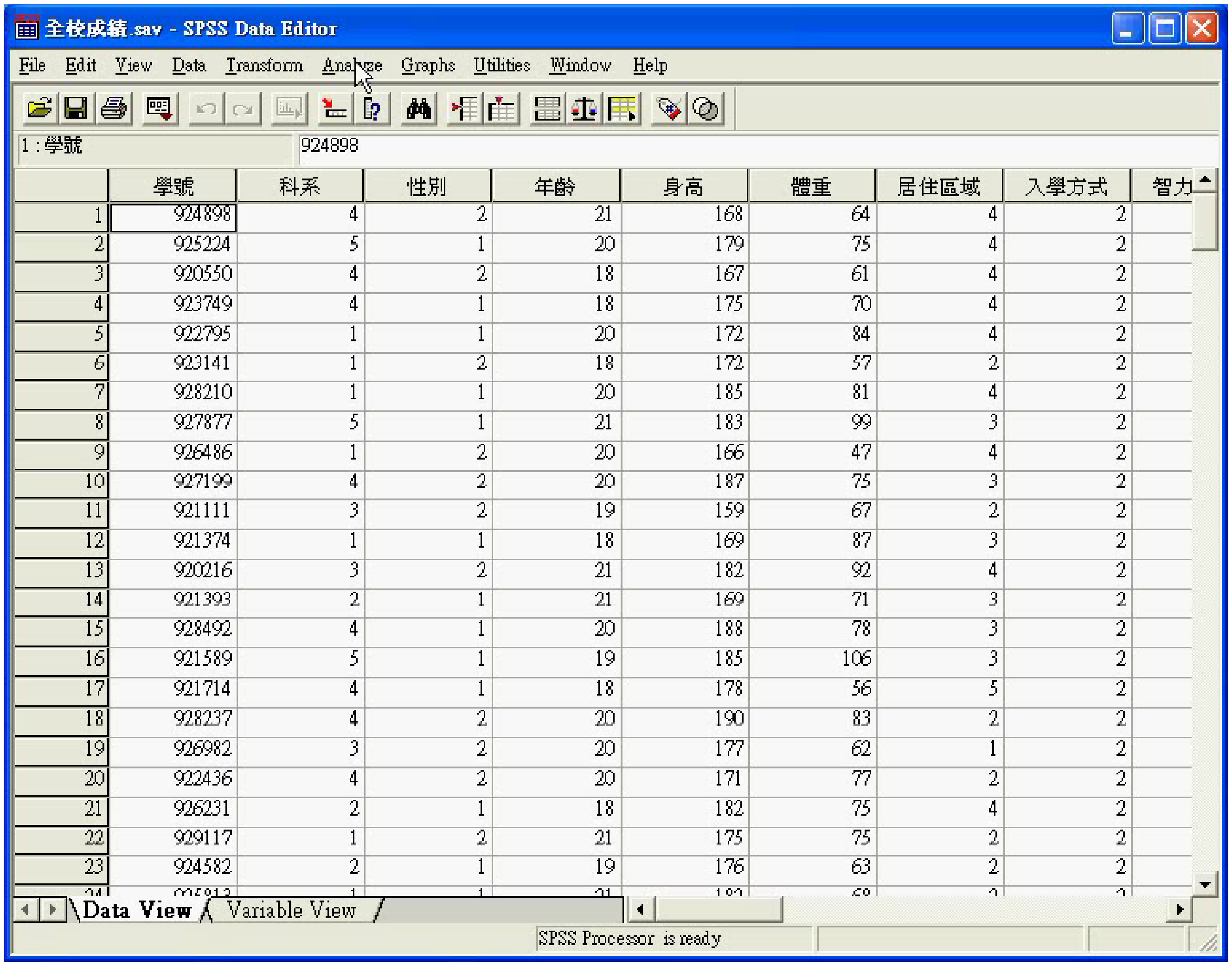
- 範例二

範例二

衔接範例一，分析各組間之差異情形。

- 操作：
  - 1.點選Analyze/Compare Means/ One-Way ANOVA/Post Hoc
  - 2.程式操作
  - 3.分析結果

範例二—程式操作



	學號	科系	性別	年齡	身高	體重	居住區域	入學方式	智力
1	924898	4	2	21	168	64	4	2	
2	925224	5	1	20	179	75	4	2	
3	920550	4	2	18	167	61	4	2	
4	923749	4	1	18	175	70	4	2	
5	922795	1	1	20	172	84	4	2	
6	923141	1	2	18	172	57	2	2	
7	928210	1	1	20	185	81	4	2	
8	927877	5	1	21	183	99	3	2	
9	926486	1	2	20	166	47	4	2	
10	927199	4	2	20	187	75	3	2	
11	921111	3	2	19	159	67	2	2	
12	921374	1	1	18	169	87	3	2	
13	920216	3	2	21	182	92	4	2	
14	921393	2	1	21	169	71	3	2	
15	928492	4	1	20	188	78	3	2	
16	921589	5	1	19	185	106	3	2	
17	921714	4	1	18	178	56	5	2	
18	928237	4	2	20	190	83	2	2	
19	926982	3	2	20	177	62	1	2	
20	922436	4	2	20	171	77	2	2	
21	926231	2	1	18	182	75	4	2	
22	929117	1	2	21	175	75	2	2	
23	924582	2	1	19	176	63	2	2	

範例二—分析結果

Multiple Comparisons

Dependent Variable: 數學能力測驗
Scheffe

(I) 科系	(J) 科系	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
電子系	資訊系	2.87	4.71	.985	-11.76	17.51
	企管系	8.55	4.64	.494	-5.84	22.95
	外文系	18.56*	4.80	.006	3.65	33.47
	法律系	21.58*	4.78	.001	6.74	36.42
資訊系	電子系	-2.87	4.71	.985	-17.51	11.76
	企管系	5.68	4.29	.780	-7.64	19.00
	外文系	15.69*	4.47	.017	1.81	29.56
	法律系	18.71*	4.44	.002	4.91	32.51
企管系	電子系	-8.55	4.64	.494	-22.95	5.84
	資訊系	-5.68	4.29	.780	-19.00	7.64
	外文系	10.01	4.39	.271	-3.62	23.63
	法律系	13.03	4.36	.066	-.51	26.57
外文系	電子系	-18.56*	4.80	.006	-33.47	-3.65
	資訊系	-15.69*	4.47	.017	-29.56	-1.81
	企管系	-10.01	4.39	.271	-23.63	3.62
	法律系	3.02	4.54	.979	-11.07	17.11
法律系	電子系	-21.58*	4.78	.001	-36.42	-6.74
	資訊系	-18.71*	4.44	.002	-32.51	-4.91
	企管系	-13.03	4.36	.066	-26.57	.51
	外文系	-3.02	4.54	.979	-17.11	11.07

*. The mean difference is significant at the .05 level.

結論：1. 法律系、外文系與企管系間平均數學成績無顯著差異。
2. 企管系、資訊系與電子系間平均數學成績無顯著差異。
3. 資訊系與電子系之平均數學成績，顯著高於法律系與外文系。

單因子變異數分析-隨機集區設計

- 當每種處方下的反應變數各有不同特性而非相同分配時，不能用上一節的作法。要將試驗單位再劃分成許多集區來討論。
- 隨機集區設計 (randomized block design): 先將試驗單位依其特質或屬性歸類於不同的集區 (block)，處方則隨機分派於同一集區內的各個試驗單位，而且一種處方只用於集區內的一個試驗單位。

★單因子變異數分析-隨機集區設計

定理4

若 x_{ijk} 為獨立之常態隨機變數，$i=1,2,\cdots,b$，$j=1,2,\cdots,k$，且變異數均相等，而其母體平均數 $\mu_{ijk}=\mu+\alpha_i+\beta_j+(\alpha\beta)_{ij}$ (α_i、β_j、$(\alpha\beta)_{ij}$ 分別表集區與處理方法所造成之離差)，則

(1)在 $H_0:\beta_1=\beta_2=\ldots=\beta_k=0$ 成立(即處理方法不影響依變數)之條件下，

$$F_1 \sim \frac{MSBW}{MSE} \sim F\left(k-1,(k-1)(b-1)\right)$$

★ 單因子變異數分析-隨機集區設計

定理4(續)

(2)在 $H_0' : \alpha_1 = \alpha_2 = \ldots = \alpha_a = 0$ 成立(即不同的集區不影響依變數)之條件下，

$$F_2 \sim \frac{MSBW}{MSE} \sim F\left(b-1, (k-1)(b-1)\right)$$

★ 單因子變異數分析-隨機集區設計

單因子變異數分析表(隨機集區設計)

變異來源	平方和	自由度	均方	f 值
處理方法 集　　區 隨機誤差	*SSBW* *SSBK* *SSE*	*k*-1 *b*-1 (*k*-1)(*b*-1)	*MSBW* *MSBK* *MSE*	$f_1 = \frac{MSBW}{MSE}$ $f_2 = \frac{MSBK}{MSE}$
總和	*SST*	*B*(*k*-1)		

★ 單因子變異數分析-隨機集區設計

例題 4

下表為10位員工操作三部不同機器生產一件相同產品所需之作業時間，請以 $\alpha=0.05$ 檢定：

(1)不同的機器之平均作業時間是否有顯著地差異？

(2)不同的員工操作機器之平均作業時間是否有顯著地差異？

★ 單因子變異數分析-隨機集區設計

例題 4(續)

員工(集區)	機器(不同處理方法)			總和
	1	2	3	
1	10.5	9.5	11.2	31.2
2	10	8.5	11	29.5
3	10.2	9.3	10.7	30.2
4	9.8	9.1	11.3	30.2
5	9.5	9	10.5	29
6	10.3	8.7	11.5	30.5
7	9.7	8.9	10.8	29.4
8	10	9.1	10.7	29.8
9	9.5	8.5	11.3	29.3
10	10.5	9.4	11	30.9
總和	100	90	110	300

★ 單因子變異數分析-隨機集區設計

解

假設一：H_0不同的機器之平均作業時間相等 & H_1: H_0不成立

假設二：H_0'不同的員工之平均作業時間相等 & H_1': H_0'不成立

欲檢定上述假設，需作以下計算：

員工(集區)作業時間之平均數：

$$\bar{x}_{1.}=\frac{31.2}{3}=10.4，\bar{x}_{2.}=\frac{29.5}{3}=9.83，\bar{x}_{3.}=\frac{30.2}{3}=10.07，$$

$$\bar{x}_{4.}=\frac{30.2}{3}=10.07，\bar{x}_{5.}=\frac{29}{3}=9.67，\bar{x}_{6.}=\frac{30.5}{3}=10.17，$$

$$\bar{x}_{7.}=\frac{29.4}{3}=9.8，\bar{x}_{8.}=\frac{29.8}{3}=9.93，\bar{x}_{1.}=\frac{29.3}{3}=9.77，\bar{x}_{10.}=\frac{30.9}{3}=10.3$$

★單因子變異數分析-隨機集區設計

解

機器(不同處理方法)作業時間之平均數：

$$\bar{x}_{.1}=\frac{100}{10}=10，\bar{x}_{.2}=\frac{90}{10}=9，\bar{x}_{.3}=\frac{110}{10}=11，$$

總平均數：

$$\bar{x}=\frac{300}{30}=10$$

由此可得：

$$SST=\sum_{i=1}^{10}\sum_{j=1}^{3}(x_{ij}-10)^2=23.32，\quad SSBW=10\sum_{j=1}^{3}(x_{.j}-10)^2=20$$

$$SSBK=3\sum_{i=1}^{10}(x_{i.}-10)^2=1.57，$$

$$SSE=SST-SSBW-SSBK=23.32-20-1.57=1.75$$

★單因子變異數分析-隨機集區設計

解

SST、SSBW、SSBK及SSE之自由度分別如下：

$$d.f.(SST)=bk-1=30-1=29$$

$$d.f.(SSBW)=k-1=3-1=2$$

$$d.f.(SSBK)=b-1=10-1=9$$

$$d.f.(SSE)=(k-1)(b-1)=18$$

由此可得

$$MSBW=\frac{SSBW}{k-1}=\frac{20}{3}=10$$ ，$$MSBK=\frac{SSBK}{b-1}=\frac{1.57}{9}=0.174$$

$$MSE=\frac{SSE}{(k-1)(b-1)}=\frac{1.75}{18}=0.097$$ ，$$f_1=\frac{MSBW}{MSE}=\frac{10}{0.097}=103.09$$

$$f_2=\frac{MSBK}{MSE}=\frac{0.174}{0.097}=1.79$$

★單因子變異數分析-隨機集區設計

解

其變異數分析表如下：

變異來源	平方和	自由度	均方	f 值
處理方法	20	2	10	103.09
集　　區	1.57	9	0.174	1.79
隨機誤差	1.75	18	0.097	
總和	23.32	29		

$$F_1 = \frac{MSBW}{MSE} \sim F(2,18)$$ ， $$F_2 = \frac{MSBK}{MSE} \sim F(9,18)$$

以顯著水準α=0.05查表可得其拒絕域分別為

$$\{f_1 > 3.55\} \text{及} \{f_2 > 2.46\}$$

★單因子變異數分析-隨機集區設計

解

然而檢定值 $f_1 = 103.09 > 3.55$，而 $f_2 = 1.79 < 2.46$

因此拒絕 H_0，但接受 H_0'，即「不同機器之平均作業時間有顯著地差異」，而「不同員工之平均作業時間無顯著地差異」。

★二因子變異數分析

定理5

若 x_{ijk} 為獨立之常態隨機變數，$i=1,2,\cdots,a$，$j=1,2,\cdots,b$，$k=1,2,\cdots,n$，且變異數均相等，其母體平均數 $\mu_{ijk}=\mu+\alpha_i+\beta_j+(\alpha\beta)_{ij}$ (α_i、β_j、$(\alpha\beta)_{ij}$ 分別表第一因子、第二因子及兩因子交互作用造成之離差)，則

(1) 在 $H_0:\alpha_1=\alpha_2=\ldots=\alpha_a=0$ 成立(即第一因子不影響依變數)之條件下，

$$\frac{MSA}{MSE}\sim F(a-1,ab(n-1))$$

★二因子變異數分析

定理5(續)

(2)在 $H_0':\beta_1=\beta_2=\ldots=\beta_b=0$ 成立(即第二因子不影響依變數) 之條件下，

$$\frac{MSB}{MSE}\sim F\left(b-1,ab(n-1)\right)$$

(3)在 $H_0'':(\alpha\beta)_{11}=(\alpha\beta)_{12}=\ldots=(\alpha\beta)_{1b}=\ldots=(\alpha\beta)_{a1}=(\alpha\beta)_{a2}=\ldots=(\alpha\beta)_{ab}$ 成立(即兩因子交互作用不影響依變數)之條件下，

$$\frac{MSAB}{MSE}\sim F\left((a-1)(b-1),ab(n-1)\right)$$

★二因子變異數分析

變異來源	平方和	自由度	均方	f 值
第一因子	SSA	$a-1$	MSA	$f_1=\dfrac{MSA}{MSE}$
第二因子	SSB	$b-1$	MSB	$f_2=\dfrac{MSB}{MSE}$
交互作用	$SSAB$	$(a-1)(b-1)$	$MSAB$	$f_3=\dfrac{MSAB}{MSE}$
隨機誤差	SSE	$ab(n-1)$	MSE	
總　　和	SST	$abn-1$		

★二因子變異數分析

例題 5

某研究者欲調查一特定品牌飲料之銷售量是否受其產品的口味及商店類型的不同影響，於是對四種不同口味的飲料及三種不同類型的商店作調查，隨機調查5天各不同類型商店之各口味飲料之銷售量，計算後得其二因子變異數分析表如下 (顯著水準為0.01)：

變異來源	平方和	自由度	均方	f 值
產品口味	5817	3	(c)	(e)
商店類型	406	2	(d)	(f)
交互作用	3618	6	603	6.88
隨機誤差	(a)	(b)	87.6	
總　和	14046	59		

★二因子變異數分析

例題 5(續)

(1)求上表之未知數(a)、(b)、(c)、(d)、(e)、(f) 。

(2)請問此資料是否顯示不同的產品口味有顯著地影響此品牌飲料之銷售量？

(3)請問此資料是否顯示不同的商店類型有顯著地影響此品牌飲料之銷售量？

(4)請問此資料是否顯示不同的產品口味及不同商店類型之交互作用有顯著地影響此品牌飲料之銷售量？

★二因子變異數分析

解

(1) $(a) = SSE = SST - SSA - SSB - SSAB = 14046 - 5817 - 406 - 3618 = 4205$

$$(b) = d.f(SSE) = d.f(SST) - d.f(SSA) - d.f(SSB) - d.f(SSAB)$$
$$= 59 - 3 - 2 - 6 = 48$$

$$(c) = MSA = \frac{SSA}{d.f.(SSA)} = \frac{5817}{3} = 1939$$

$$(d) = MSB = \frac{SSB}{d.f.(SSB)} = \frac{406}{2} = 203$$

$$(e) = f_1 = \frac{MSA}{MSE} = \frac{1939}{87.6} = 22.13$$

$$(f) = f_2 = \frac{MSB}{MSE} = \frac{203}{87.6} = 2.32$$

★二因子變異數分析

解

承上頁，因此完整之變異數分析表如下：

變異來源	平方和	自由度	均方	f 值
產品口味	5817	3	1939	$f_1=22.13$
商店類型	406	2	203	$f_2=2.32$
交互作用	3618	6	603	$f_3=6.88$
隨機誤差	4205	48	87.6	
總　　和	14046	59		

★二因子變異數分析

解

(2)因為 $F_1=\dfrac{MSA}{MSE}\sim f(3,48)$，經查表得知，其拒絕域為

$\{f_1>f_{0.01}(3,48)\}=\{f_1>4.24\}$。而 $f_1=22.13$ 落在拒絕域，因此

拒絕 H_0，即不同的產品口味對產品之銷售量有顯著地影響。

(3)因為 $F_2=\dfrac{MSB}{MSE}\sim f(2,48)$，經查表得知，其拒絕域為

$\{f_2>f_{0.01}(2,48)\}=\{f_2>5.113\}$。而 $f_2=2.32$ 不落在拒絕域，因此

接受H_0':不同類型之商店對產品之銷售量無顯著地影響。

★二因子變異數分析

解

(4)因為 $F_3 = \dfrac{MSAB}{MSE} \sim f(6,48)$，經查表得知，其拒絕域為 $\{f_3 > f_{0.01}(6,48)\} = \{f_3 > 3.22\}$。而 $f_3 = 6.88$ 落在拒絕域，因此拒絕 H_0''，即不同口味及不同類型商店之交互作用對產品之銷售量有顯著地影響。

雙因子變異數分析

- 當我們懷疑某一屬量的依變數可能同時受到兩個屬質自變數影響時，較正確的分析方法應是採用雙因子變異數分析，例如：如下表想要探討「居住區域」(α_i)與「性別」(β_j)對「存款」的影響。性別與居住區域對存款的影響效果稱為主效果；性別與區域是否同時對存款造成不同的效應則為交互效果。

性別 (B) / 居住區域 (A)	女性 ($j=1$)	男性 ($j=2$)	$\overline{X}_i$
東區 ($i=1$)	434.13 ($\overline{X}_{11}$)	464.21 ($\overline{X}_{12}$)	446.82
西區 ($i=2$)	433.57 ($\overline{X}_{21}$)	449.65 ($\overline{X}_{22}$)	441.51
中區 ($i=3$)	457.64 ($\overline{X}_{31}$)	343.04 ($\overline{X}_{32}$)	402.54
$\overline{X}_j$	439.9	425.22	433.04

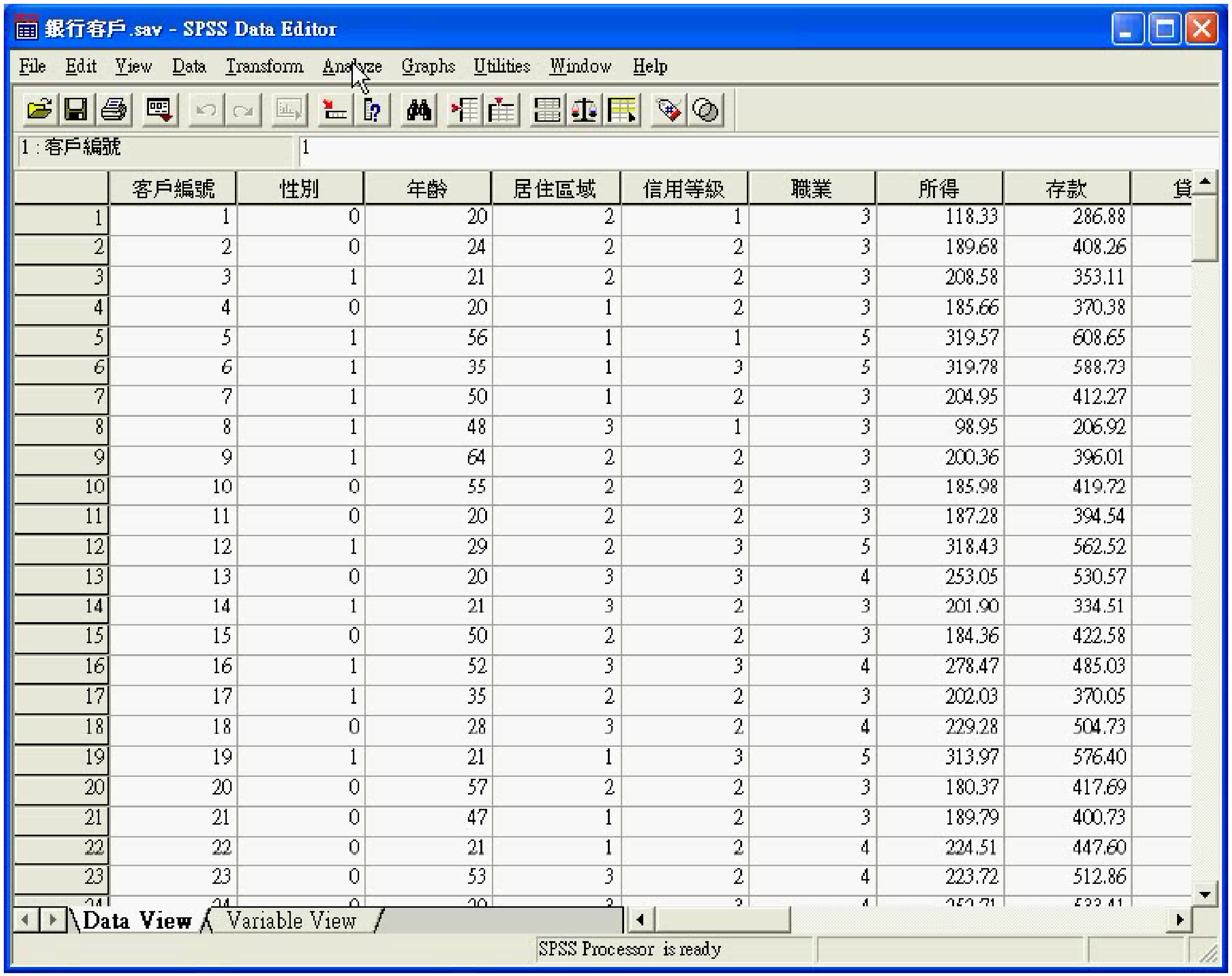
雙因子變異數分析

- 檢定內容

變異來源	假說一	假說二
A因子	H_0：各居住區域平均存款完全相等 H_1：各居住區域平均存款不完全相等	$H_0: \alpha_i = 0 \text{ for } i = 1,2,3$ $H_1: \alpha_i$ 不完全為 0
B因子	H_0：男女生平均存款完全相等 H_1：男女生平均存款不完全相等	$H_0: \beta_j = 0 \text{ for } j = 1,2$ $H_1: \beta_j$ 不完全為 0
交互作用	H_0：各交互作用皆為 0 H_1：各交互作用不完全為 0	$H_0: \gamma_{ij} = 0 \text{ for } i = 1,2,3$ $j = 1,2$ $H_1: \gamma_{ij}$ 不完全為 0

- 程式操作
- 分析說明

範例—程式操作



	客戶編號	性別	年齡	居住區域	信用等級	職業	所得	存款	貸
1	1	0	20	2	1	3	118.33	286.88	
2	2	0	24	2	2	3	189.68	408.26	
3	3	1	21	2	2	3	208.58	353.11	
4	4	0	20	1	2	3	185.66	370.38	
5	5	1	56	1	1	5	319.57	608.65	
6	6	1	35	1	3	5	319.78	588.73	
7	7	1	50	1	2	3	204.95	412.27	
8	8	1	48	3	1	3	98.95	206.92	
9	9	1	64	2	2	3	200.36	396.01	
10	10	0	55	2	2	3	185.98	419.72	
11	11	0	20	2	2	3	187.28	394.54	
12	12	1	29	2	3	5	318.43	562.52	
13	13	0	20	3	3	4	253.05	530.57	
14	14	1	21	3	2	3	201.90	334.51	
15	15	0	50	2	2	3	184.36	422.58	
16	16	1	52	3	3	4	278.47	485.03	
17	17	1	35	2	2	3	202.03	370.05	
18	18	0	28	3	2	4	229.28	504.73	
19	19	1	21	1	3	5	313.97	576.40	
20	20	0	57	2	2	3	180.37	417.69	
21	21	0	47	1	2	3	189.79	400.73	
22	22	0	21	1	2	4	224.51	447.60	
23	23	0	53	3	2	4	223.72	512.86	

範例—分析說明

- 由交互效果的檢定結果（下表）可知，性別與居住區域存在交互效果F = 5.886 (Sig.=0.003 < 0.05)。因而必須進一步控制某主要效果，檢定在該主要效果下之交互效果。

Tests of Between-Subjects Effects

Dependent Variable: 存款

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	256426.245[a]	5	51285.249	3.346	.006
Intercept	35224134.421	1	35224134.4	2297.884	.000
性別	24747.963	1	24747.963	1.614	.205
居住區域	77864.349	2	38932.175	2.540	.082
性別 * 居住區域	180448.318	2	90224.159	5.886	.003
Error	2958485.991	193	15328.943		
Total	40531610.984	199			
Corrected Total	3214912.236	198			

a. R Squared = .080 (Adjusted R Squared = .056)

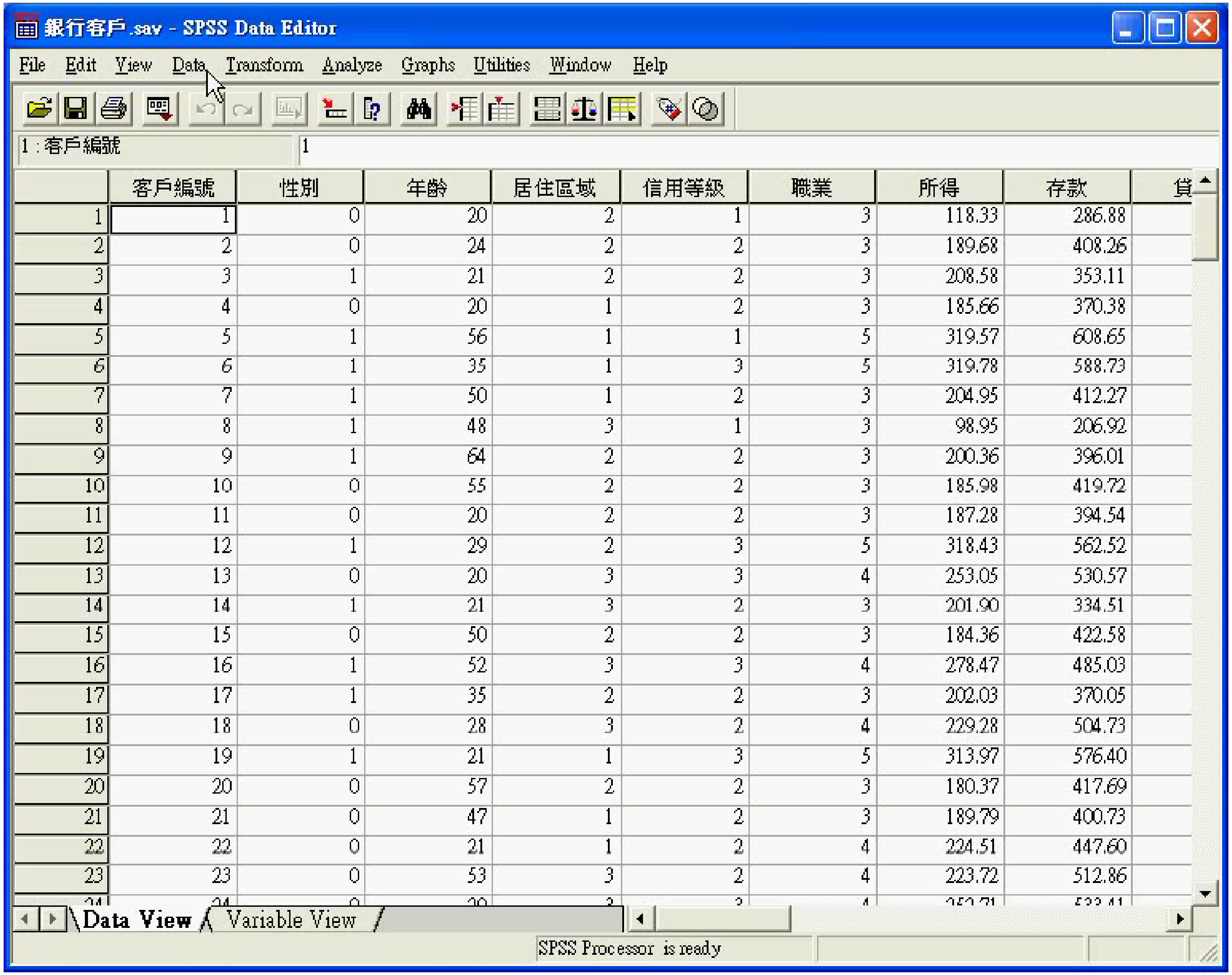
雙因子變異數分析

- 控制居住區域

  針對居住東區客戶(A1)，探討不同性別下之存款差異(A1B1與A1B2)。亦可使用單因子變異數分析，$F = t^2$，與 t 檢定之顯著值相同。

- 操作：
  - 1.點選Data/Select Cases/選居住區域 = 1進行兩樣本 t 檢定
  - 2.程式操作
  - 3.分析結果

範例—程式操作



	客戶編號	性別	年齡	居住區域	信用等級	職業	所得	存款	貸
1	1	0	20	2	1	3	118.33	286.88	
2	2	0	24	2	2	3	189.68	408.26	
3	3	1	21	2	2	3	208.58	353.11	
4	4	0	20	1	2	3	185.66	370.38	
5	5	1	56	1	1	5	319.57	608.65	
6	6	1	35	1	3	5	319.78	588.73	
7	7	1	50	1	2	3	204.95	412.27	
8	8	1	48	3	1	3	98.95	206.92	
9	9	1	64	2	2	3	200.36	396.01	
10	10	0	55	2	2	3	185.98	419.72	
11	11	0	20	2	2	3	187.28	394.54	
12	12	1	29	2	3	5	318.43	562.52	
13	13	0	20	3	3	4	253.05	530.57	
14	14	1	21	3	2	3	201.90	334.51	
15	15	0	50	2	2	3	184.36	422.58	
16	16	1	52	3	3	4	278.47	485.03	
17	17	1	35	2	2	3	202.03	370.05	
18	18	0	28	3	2	4	229.28	504.73	
19	19	1	21	1	3	5	313.97	576.40	
20	20	0	57	2	2	3	180.37	417.69	
21	21	0	47	1	2	3	189.79	400.73	
22	22	0	21	1	2	4	224.51	447.60	
23	23	0	53	3	2	4	223.72	512.86	

範例—分析結果

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means					95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
存款	Equal variances assume	.229	.634	-.906	62	.368	-30.08	33.18	-96.41	36.26
	Equal variances not assumed			-.892	52.719	.376	-30.08	33.72	-97.72	37.56

ANOVA

存款

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	14120.52	1	14120.524	.821	.368
Within Groups	1065728	62	17189.162		
Total	1079849	63			

Sig. = 0.368 > 0.05，無法拒絕H_0

結論：東區男性客戶之平均存款與東區女性客戶之平均存款無顯著差異。

雙因子變異數分析

- 控制性別

  針對女性客戶(B1)，探討不同居住區域下之存款差異（A1B1、A2B1與A3B1）。此時使用單因子變異數分析（居住區域分為3類）。

ANOVA

存款

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	11405.75	2	5702.875	.384	.682
Within Groups	1530732	103	14861.480		
Total	1542138	105			

Sig. = 0.682 > 0.05，無法拒絕H_0

結論：女性客戶在不同區域上，其平均存款無顯著差異

雙因子變異數分析

- 總結：經由上述的雙因子變異數分析，發現以下幾項現象
  - 平均存款在性別間無顯著差異（亦即，男性與女性之平均存款無顯著差異）。
  - 平均存款在居住區域間無顯著差異（亦即，東區、西區與中區之平均存款無顯著差異）。
  - 性別與居住區域對平均存款產生交互作用。對居住中區之客戶而言，女性客戶存款顯著大於男性；對男性客戶而言，居住在東區與西區客戶，其存款顯著大於居住中區住戶。

結論

- 變異數分析是用來檢定兩個以上平均數是否相等或某個變數是否受某些因子所影響之統計方法。而其主要概念是將資料之變異依其來源區分幾個不同的部份，然後再以兩樣本變異數比之抽樣分配──*F*-分配為基礎來進行檢定分析。

(1)在完全隨機設計之單因子變異數分析中，資料之總變異如下：

$$SST(\text{總變異}) = SSB(\text{組間變異}) + SSE(\text{組內變異})$$

結論

(2)在隨機集區設計之單因子變異數分析中，資料之總變異如下：

$$SST(\text{總變異}) = SSBW(\text{處理方法間之變異}) + SSBK(\text{集區間之變異}) + SSE(\text{隨機變異})$$

(3)在二因子變異數分析中，資料之總變異如下：

$$SST(\text{總變異}) = SSA(\text{第一因子間之變異}) + SSB(\text{第二因子間之變異}) + SSAB(\text{兩因子交互作用之變異}) + SSE(\text{隨機變異})$$