第 14 章　變異數分析

- 變異數分析(analysis of variance; ANOVA)：檢定三個或三個以上的母體平均數是否相等的統計方法，或檢定因子對依變數是否有影響的統計方法。
- 為什麼需要變異數分析？
  - ➢ 假設我們要檢定三個母體的平均數相等與否：$\mu_1=\mu_2=\mu_3$
  - ➢ 依照 13 章兩母體平均數檢定的方法，我們應該可以分別檢定底下三組假設

$$H_0:\mu_1=\mu_2 \qquad H_0:\mu_2=\mu_3 \qquad H_0:\mu_1=\mu_3$$
$$H_1:\mu_1\neq\mu_2 \qquad H_1:\mu_2\neq\mu_3 \qquad H_1:\mu_1\neq\mu_3$$

若三個檢定均無法拒絕虛無假設，即代表 $\mu_1=\mu_2=\mu_3$ 是正確的；若其中至少有一個檢定拒絕了虛無假設，則代表 $\mu_1=\mu_2=\mu_3$ 是不正確的(至少有一個等式不成立)。

- ➢ 然而上述作法是有問題的：若把顯著水準設定為 $\alpha=0.05$；則代表在 $\alpha=0.05$ 下檢定 $\mu_1=\mu_2=\mu_3$ 的虛無假設時，型 I 錯誤為 $\alpha=0.05$ [虛無假設正確，但卻錯誤地拒絕虛無假設的機率]。

  然而當我們在進行以上三組檢定時，若亦設定顯著水準為 $\alpha=0.05$，則當虛無假設 $\mu_1=\mu_2=\mu_3$ 正確時，真正的型 I 錯誤並非 $\alpha=0.05$，而是 $1-(1-0.05)^3=0.142625$，亦即，錯誤地拒絕虛無假設的機率會大於顯著水準 $\alpha=0.05$。
- ➢ 若我們要檢定 k 個母體平均數是否相等，則需進行 $m=C_2^k$ 次兩母體平均數檢定，因為 $k\uparrow\Rightarrow m\uparrow$，這 m 次檢定所造成的型 I 錯誤為 $1-(1-\alpha)^m\rightarrow 1$；亦即，儘管虛無假設正確(所有的母體平均數都相等)，但我們總是會拒絕虛無假設。
- ➢ 這就是為什麼需要變異數分析，而不可以進行多次兩母體平均數檢定的原因。

檢定多個母體平均數是否相等

- 假設黑松汽水公司計畫購買新機器，以便將可樂裝入 600ml 的瓶子。該公司目前正在考慮三種廠牌的裝填機，在決定之前，該公司從每種廠牌的機器各隨機選取 5 部，測試 1 小時的裝填量，測試結果如下。

	A	B	C	D
1	樣本	廠牌A	廠牌B	廠牌C
2	1	106	107	100
3	2	80	112	104
4	3	104	115	95
5	4	110	104	100
6	5	90	117	106
7	樣本平均數	98	111	101
8	樣本變異數	158	29.5	18

- 符號：假設我們現在要檢定 k 個獨立母體 $Y_i, i=1,\ldots,k$ 的平均數是否相等，令 μ_i，$i=1,\ldots,k$ 代表其母體平均數($\mu_1,\mu_2,\ldots,\mu_k$)，σ_i^2，$i=1,\ldots,k$ 代表其母體變異數($\sigma_1^2,\sigma_2^2,\ldots,\sigma_k^2$)。
  - ➢ 從 k 個獨立母體中抽出底下 k 組隨機樣本

樣本組號	樣本個數	隨機樣本
1	n_1	$Y_{11},Y_{12},Y_{13},\ldots,Y_{1n_1}$
2	n_2	$Y_{21},Y_{22},Y_{23},\ldots,Y_{2n_2}$
⋮	⋮	⋮
k	n_k	$Y_{k1},Y_{k2},Y_{k3},\ldots,Y_{kn_k}$

  - ➢ Y_{ij} 代表第 i 組中的第 j 觀察值。
  - ➢ 第 i 組的樣本平均數：$\bar{Y}_i=\frac{1}{n_i}\sum_{j=1}^{n_i}Y_{ij}$

    第 i 組的樣本變異數：$S_i^2=\frac{1}{n_i-1}\sum_{j=1}^{n_i}(Y_{ij}-\bar{Y}_i)^2$
  - ➢ 總樣本平均數：$\bar{\bar{Y}}=\frac{1}{n}\sum_{i=1}^{k}\sum_{j=1}^{n_i}Y_{ij}$，式中 $n=n_1+\cdots+n_k$

- **變異數分析的假設：**
  - 假設因子對依變數的影響效果是固定的；亦即，$\mu_i - \mu$ 為一常數而不是隨機變數。
  - 每個小母體均為常態分配：$Y_i \sim N(\mu_i, \sigma_i^2), i = 1, \cdots, k$。
  - <u>變異數齊一性</u>(Homogeneity)：每個<u>小母體</u>(subpopulation)的變異數均相等，亦即 $\sigma_i^2 = \sigma^2$。
  - 抽樣方法為獨立簡單隨機抽樣，亦即自 k 個小母體分別抽取獨立之隨機樣本。
- **變異數分析的步驟：**

  (1) 設立兩個假設。

  (2) 選取 $\dfrac{MSF}{MSE}$ 為檢定統計量。

  (3) 決定決策法則。

  (4) 下結論。

建立兩個假設

- **虛無假設**：k 個母體平均數相等

  $H_0: \mu_1 = \mu_2 = \cdots = \mu_k$

  - 這種虛無假設稱為『**聯合假設**(joint hypothesis)』：由數個『單一假設(single hypothesis)』所組成[$H_0: \mu = 0$ 這種僅由一個等式所組成的虛無假設即為單一假設]。

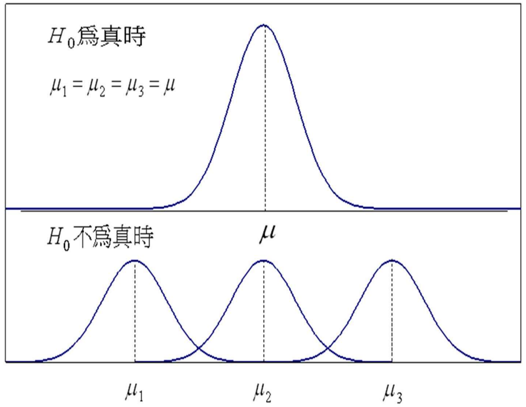
  對立假設：虛無假設不成立

  $H_1: \mu_i$ 不全相等

  - 在 ANOVA 分析(任何聯合假設的檢定)中，對立假設僅宣告虛無假設不成立；因為只要虛無假設中的等號有一個不成立即代表虛無假設錯誤，而這種情況有太多種了，難以明確寫出所有情況。
- 例子：在填裝機的例子中，令 μ_1、μ_2、μ_3 分別代表 3 種廠牌填裝機每小時的平均裝填量。則虛無假設與對立假設為：

$H_0: \mu_1 = \mu_2 = \mu_3$ [3 種機器的平均裝填量相同]

$H_1: \mu_i$ 不全相等 [3 種機器的平均裝填量不全相同]



- 若虛無假設 $\mu_1 = \mu_2 = \mu_3$ 為真，則我們可預期 3 個子樣本平均數應該很接近。因此，子<u>樣本平均數間的差異越小，越可支持虛無假設正確</u>；反之，<u>若子樣本平均數間的差異越大，即代表越可支持虛無假設是錯誤的</u>。

選取 $\dfrac{MSF}{MSE}$ 為檢定統計量

- 變異數分析的想法從這裡開始：第 i 組中的第 j 個觀察值 Y_{ij} 與總樣本平均數 $\overline{\overline{Y}} = \frac{1}{n}\sum_{i=1}^{k}\sum_{j=1}^{n_i} Y_{ij}$ 的差異(<u>總差異</u>)可分解為

  $$Y_{ij} - \overline{\overline{Y}} = \underbrace{(\overline{Y}_i - \overline{\overline{Y}})}_{\text{組間差異}} + \underbrace{(Y_{ij} - \overline{Y}_i)}_{\text{組內差異}}$$

  - Y_{ij} 為第 i 組中的隨機樣本，所以一定會與總樣本平均數 $\overline{\overline{Y}}$ 有所差異，而這個差異的來源可能有二：(1)每組的平均數本來就與總平均數不同；(2)每組內樣本的隨機誤差。
  - **組間差異**(difference between subpopulations)：第 i 組的樣本平均數 $\overline{Y}_i$ 與總樣本平均數 $\overline{\overline{Y}}$ 的差異；因為 Y_{ij} 為來自於第 i 組母體的樣本，若子樣本的平均數與總平均數不同即會造成該項差異。由於子樣本的平均數不同是因子所因起的，故又稱為<u>因子的差異</u>(difference due to factor)。

- **組內差異**(difference within subpopulations)：來自於第 i 組的樣本觀察值 Y_{ij} 與該組之樣本平均數 $\bar{Y}_i$ 的差異；這些差異是抽樣所造成的誤差，故為隨機差異(difference due to error)
- 例子：在可樂裝填機的例子中。從第 i 品牌中所抽出的第 j 台機器之填裝量 Y_{ij}，與總樣本平均數 $\bar{\bar{Y}}$ 之差異可能來自於

$$Y_{ij}-\bar{\bar{Y}}=(\bar{Y}_i-\bar{\bar{Y}})+(Y_{ij}-\bar{Y}_i)$$

(1) 各廠牌裝填機每小時平均裝填數量不同[組間差異]。

(2) 每個廠牌的裝填機每小時裝填數量不同[組內差異]。

若 3 種廠牌裝填機的平均裝填量相等($\mu_1=\mu_2=\mu_3=\mu$)，則各廠牌裝填機的樣本平均裝填量 $\bar{Y}_i$，應與所有廠牌裝填機的總樣本平均裝填量 $\bar{\bar{Y}}$ 差不多，因此 $\bar{Y}_i-\bar{\bar{Y}}\approx 0$，故總差異 $Y_{ij}-\bar{\bar{Y}}$ 都是來自於抽樣時的隨機誤差 $Y_{ij}-\bar{Y}_i$。反之，若 3 廠牌的平均裝填量不相等，總差異 $Y_{ij}-\bar{\bar{Y}}$ 中應有很大部分可歸因於因子差異 $\bar{Y}_i-\bar{\bar{Y}}$。

- **ANOVA 背後的想法**：對 $Y_{ij}-\bar{\bar{Y}}$ 進行拆解，將總差異的來源區分為組間差異與組內差異。若虛無假設(每組的母體平均數相等)成立，則 $\bar{Y}_i$ 與 $\bar{\bar{Y}}$ 應該很接近[$\bar{Y}_i-\bar{\bar{Y}}\approx 0$]，$Y_{ij}$ 與 $\bar{Y}$ 的差異應該都是隨機差異所造成；因此，組間差異會很小，組內差異會較大。若虛無假設不成立，則 $\bar{Y}_i$ 與 $\bar{\bar{Y}}$ 應該有較大差異[$\bar{Y}_i-\bar{\bar{Y}}$ 不接近 0]，Y_{ij} 與 $\bar{Y}$ 的差異應該有很大的部份是因子的差異所造成；因此，組間差異會較大。
- **ANOVA 就是在比較組間差異與組內差異的相對大小；若組間差異相對較小，代表虛無假設正確(每組的母體平均數相等)；若組間差異相對較大，代表虛無假設錯誤(每組的母體平均數不全相等)**。但是在比較時必須將所有樣本都納入考量，故須加總所有樣本；加總時為了避免正負抵銷，故將誤差取平方。

- 將 $Y_{ij}-\bar{\bar{Y}}=(\bar{Y}_i-\bar{\bar{Y}})+(Y_{ij}-\bar{Y}_i)$ 左右兩邊平方可得

$$(Y_{ij}-\bar{\bar{Y}})^2=(\bar{Y}_i-\bar{\bar{Y}})^2+(Y_{ij}-\bar{Y}_i)^2+2(\bar{Y}_i-\bar{\bar{Y}})(Y_{ij}-\bar{Y}_i)$$

對所有的 $i=1,...,k$ 、 $j=1,...,n_i$ 加總可得

$$\sum_{i=1}^{k}\sum_{j=1}^{n_i}(Y_{ij}-\bar{\bar{Y}})^2=\sum_{i=1}^{k}\sum_{j=1}^{n_i}(\bar{Y}_i-\bar{\bar{Y}})^2+\sum_{i=1}^{k}\sum_{j=1}^{n_i}(Y_{ij}-\bar{Y}_i)^2+2\sum_{i=1}^{k}\sum_{j=1}^{n_i}(\bar{Y}_i-\bar{\bar{Y}})(Y_{ij}-\bar{Y}_i)$$

由於

$$\sum_{i=1}^{k}\sum_{j=1}^{n_i}(\bar{Y}_i-\bar{\bar{Y}})(Y_{ij}-\bar{Y}_i)=\sum_{i=1}^{k}(\bar{Y}_i-\bar{\bar{Y}})\sum_{j=1}^{n_i}(Y_{ij}-\bar{Y}_i)=\sum_{i=1}^{k}(\bar{Y}_i-\bar{\bar{Y}})[n_i\bar{Y}_i-n_i\bar{Y}_i]=0$$

因此

$$\sum_{i=1}^{k}\sum_{j=1}^{n_i}(Y_{ij}-\bar{\bar{Y}})^2=\sum_{i=1}^{k}\sum_{j=1}^{n_i}(\bar{Y}_i-\bar{\bar{Y}})^2+\sum_{i=1}^{k}\sum_{j=1}^{n_i}(Y_{ij}-\bar{Y}_i)^2$$

- **總變異**(total sum of squares; **SST**)

$$SST=\sum_{i=1}^{k}\sum_{j=1}^{n_i}(Y_{ij}-\bar{\bar{Y}})^2$$

- **因子變異**(sum of squares due to factor; **SSF**)

$$SSF=\sum_{i=1}^{k}\sum_{j=1}^{n_i}(\bar{Y}_i-\bar{\bar{Y}})^2=\sum_{i=1}^{k}n_i(\bar{Y}_i-\bar{\bar{Y}})^2$$

- **隨機變異**(sum of squares due to error; **SSE**)

$$SSE=\sum_{i=1}^{k}\sum_{j=1}^{n_i}(Y_{ij}-\bar{Y}_i)^2=\sum_{i=1}^{k}(n_i-1)S_i^2$$

- 總變異(總差異的平方和)可分解成：『因子所造成的差異之平方和』與『隨機差異的平方和』

$SST=SSF+SSE$

總變異＝組間變異＋組內變異

$\boxed{\text{總變異＝因子變異＋隨機變異}}$

- ANOVA 判斷虛無假設正確與否的直覺基礎：$\dfrac{\text{因子變異}}{\text{隨機變異}}$

虛無假設正確：$\dfrac{\text{因子變異}}{\text{隨機變異}}$ 較小

虛無假設錯誤：$\dfrac{\text{因子變異}}{\text{隨機變異}}$ 較大

- 但以『因子變異與隨機變異的比率』來做決策可能出問題：

$$SST = SSF + SSE$$

自由度 $(n-1) = (k-1) + (n-k)$

式中，總樣本數 $n = \sum_{i=1}^{k} n_i$ 。

$SST = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(Y_{ij} - \overline{\overline{Y}})^2$ ：加總了 n 項，但計算變異時須先估計 $\overline{\overline{Y}}$ ，故自由度為 $n-1$ 。

$SSF = \sum_{i=1}^{k} n_i(\overline{Y}_i - \overline{\overline{Y}})^2$ ：加總了 k 項，但計算變異時須先估計 $\overline{\overline{Y}}$ ，故自由度為 $k-1$ 。

$SSE = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(Y_{ij} - \overline{Y}_i)^2$ ：加總了 n 項，但計算變異時須先估計 k 個 $\overline{Y}_i$ ，故自由度為 $n-k$ 。

- 當我們增加每組的樣本個數 n_i 時，TSS 一定會增加，但有可能 SSF 並未增加，僅 SSE 增加，因而 SSF/SSE 變小；此時若虛無假設是錯誤的，則我們就會錯誤地接受虛無假設

- **平均變異(即變異數)**：定義因子平均變異與隨機平均變異如下
  - **因子平均變異(因子變異數)**：

$$MSF = \frac{SSF}{k-1}$$

  - **隨機平均變異(隨機變異數)**：

$$MSE = \frac{SSE}{n-k}$$

  - ANOVA 決策的依據：$\dfrac{MSF}{MSE}$

    虛無假設正確：$\dfrac{MSF}{MSE}$ 較小

    虛無假設錯誤：$\dfrac{MSF}{MSE}$ 較大

  - 然而到底多小才叫小，多大才叫大呢？底下依序解釋。

- 令 μ_i 表示各子母體($i = 1,...k$)的母體平均數，σ^2 表示各子母體的母體變異數。則

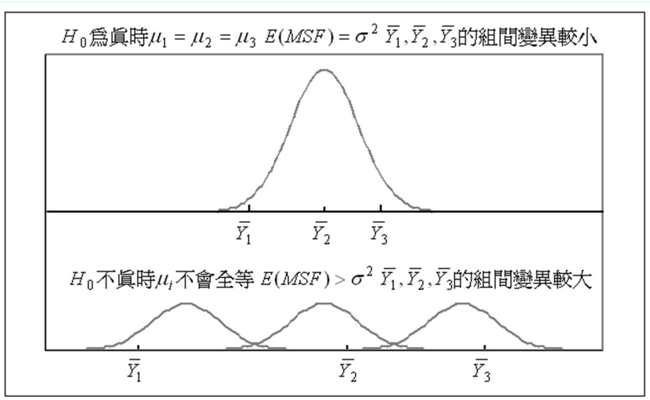
$$E(MSF) = \sigma^2 + \tfrac{1}{k-1}\sum_{i=1}^{k} n_i(\mu_i - \mu_0)^2$$

式中 [總平均數] $\mu_0 = \frac{1}{n}\sum_{i=1}^{k} n_i\mu_i$

$$E(MSE) = \sigma^2$$

- 上述結果的涵義：
  - 若虛無假設 $H_0: \mu_1 = \mu_2 = \cdots = \mu_k$ 正確[令子母體的平均數均為 μ^*]，則整體母體平均數 $\mu_0 = \frac{1}{n}\sum_{i=1}^{k} n_i\mu_i = \frac{1}{n}\sum_{i=1}^{k} n_i\mu^* = \mu^*$ 。此時 $\frac{1}{k-1}\sum_{i=1}^{k} n_i(\mu_i - \mu_0)^2 = \frac{1}{k-1}\sum_{i=1}^{k} n_i(\mu^* - \mu^*)^2 = 0$ ，因此 $\frac{E(MSF)}{E(MSE)} = \frac{\sigma^2}{\sigma^2} = 1$；亦即，當所有子母體的平均數均相等的虛無假設 $H_0: \mu_1 = \mu_2 = \cdots = \mu_k$ 成立時，我們應可預期 $\frac{MSF}{MSE} = 1$ 。

- 若虛無假設 $H_0: \mu_1 = \mu_2 = \cdots = \mu_k$ 不正確，則整體母體平均數 μ_0 至少與部分子母體的平均數 μ_i 不相等，此時 $\frac{1}{k-1}\sum_{i=1}^{k} n_i(\mu_i - \mu_0)^2 > 0$ ，故 $E(MSF) > \sigma^2$ ，因此 $\frac{E(MSF)}{E(MSE)} > 1$ ；亦即，當子母體的平均數不全相等時(虛無假設錯誤)，我們應可預期 $\frac{MSF}{MSE} > 1$ [$\frac{MSF}{MSE}$ 會較大]。



- 由於樣本資料必然存在抽樣誤差，所以我們當然不能僅根據 $\frac{MSF}{MSE}$ 是否等於 1 或大於 1 來判斷虛無假設正確與否，而是要根據 $\frac{MSF}{MSE}$ 的抽樣分配界定出拒絕域與接受域;但上述分析至少告訴我們：要拒絕虛無假設一定是 $\frac{MSF}{MSE}$ 大於 1 時，亦即拒絕域在右尾。
- $\frac{MSF}{MSE}$ 的抽樣分配：當虛無假設 $H_0:\mu_1=\mu_2=\cdots=\mu_k\ \ (=\mu)$ 成立時

$$\frac{MSF}{MSE}\sim F(k-1,n-k)$$

式中 $n=\sum_{i=1}^{k} n_i$ 為所有子樣本的樣本個數總和，k 為子樣本組數

- ANOVA 的檢定統計量與決策法則：

***F* 統計量**：

$$F=\frac{MSF}{MSE}\sim F_{k-1,n-k}$$

<u>**樣本 *F* 統計量的計算步驟**</u>：ANOVA 的題目通常會給定 k 組子樣本的樣本個數 n_i 、子樣本平均數 $\overline{Y}_i$ 、子樣本變異數 S_i^2 ；換句話說，必須有這些數據(或擁有可以算出這些數據的資料)，才可以進行變異數分析。

(1) 計算組間變異(因子變異)SSF 與組內變異(隨機變異)SSE

$$\boxed{SSF=\sum_{i=1}^{k} n_i(\overline{Y}_i-\overline{\overline{Y}})^2}$$

其中 $\overline{\overline{Y}}=\frac{1}{n}\sum_{i=1}^{k}\sum_{j=1}^{n_i} Y_{ij}=\boxed{\frac{1}{n}\sum_{i=1}^{k} n_i\overline{Y}_i}$

$$\boxed{SSE=\sum_{i=1}^{k}(n_i-1)S_i^2}$$

(2) 算出 SSF 與 SSE 的自由度 $(k-1)$ 、$(n-k)$

(3) 計算因子平均變異 MSF 與隨機平均變異 MSE

$$\boxed{MSF=\frac{SSF}{k-1}}\text{、}\boxed{MSE=\frac{SSE}{n-k}}$$

(4) 計算樣本 F 統計量

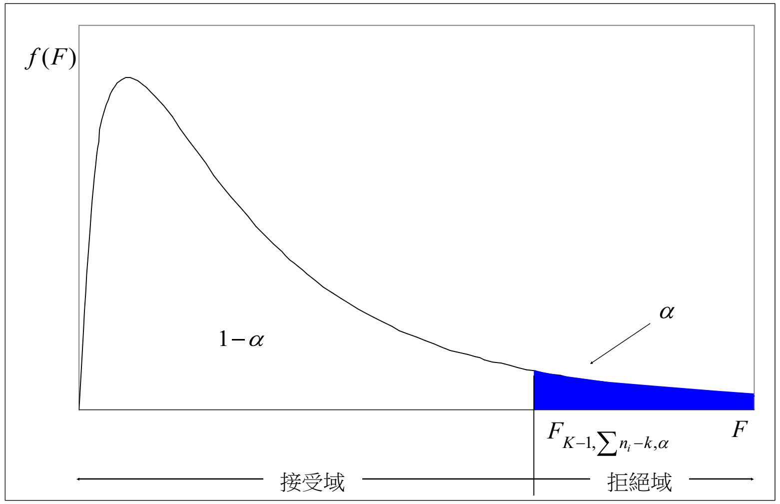
$$\boxed{F=\frac{MSF}{MSE}}$$

<u>**變異數分析表**</u>(table of ANOVA)：變異數分析的結果通常會按照計算的步驟整理成底下表格。

變異來源	平方和(SS)	自由度(df)	平均平方和(MS)	F
因子(組間)	SSF	$k-1$	$MSF=\frac{SSF}{k-1}$	$\frac{MSF}{MSE}$
隨機(組內)	SSE	$n-k$	$MSE=\frac{SSE}{n-k}$	
總和	SST	$n-1$		

決策法則：假設 $F_{k-1,n-k,\alpha}$ 代表在顯著水準 α 下，根據 $F_{k-1,n-k}$ 分配所找出的右尾臨界值，則[F 代表<u>樣本 F 統計量</u>]

- 若 $F>F_{k-1,n-k,\alpha}$ ，則<u>**拒絕**</u>虛無假設 $H_0:\mu_1=\mu_2=\cdots=\mu_k$ 。
- 若 $F\le F_{k-1,n-k,\alpha}$ ，則<u>**接受**</u>虛無假設 $H_0:\mu_1=\mu_2=\cdots=\mu_k$ 。



● 例子：3 個廠牌的可樂填裝機每小時平均填充量是否相等。根據之前例題的資料，我們整理出底下的數據

廠牌 A	廠牌 B	廠牌 C
$n_1 = 5$	$n_2 = 5$	$n_3 = 5$
$\overline{Y}_1 = 98$	$\overline{Y}_2 = 111$	$\overline{Y}_2 = 101$
$S_1^2 = 168$	$S_2^2 = 29.5$	$S_3^2 = 18$

請在顯著水準 $\alpha = 0.05$ 下，檢定三廠牌填裝機每小時平均填充量是否相等？

➢ 虛無假設與對立假設

$H_0 : \mu_1 = \mu_2 = \mu_3$ [3 廠牌填裝機的平均裝填量相同]

$H_1 : \mu_i$ 不全相等 [3 廠牌填裝機的平均裝填量不全相同]

➢ 總樣本個數：$n = \sum_{i=1}^{k} n_i = 5+5+5 = 15$

➢ 總樣本平均數：

$$\overline{\overline{Y}} = \tfrac{1}{n}\sum_{i=1}^{k} n_i \overline{Y}_i = \tfrac{1}{15}(5\cdot 98 + 5\cdot 111 + 5\cdot 101) = 103.33$$

➢ 因子平方和(SSF)：

$$SSF = \sum_{i=1}^{k} n_i (\overline{Y}_i - \overline{\overline{Y}})^2$$

$$= 5(98-103.33)^2 + 5(111-103.33)^2 + 5(101-103.33)^2 = 463.33$$

SSF 之自由度為 $k-1 = 3-1 = 2$

➢ 隨機平方和(SSE)：

$$SSE = \sum_{i=1}^{k} (n_i - 1) S_i^2$$

$$= (5-1)158 + (5-1)29.5 + (5-1)18 = 822$$

SSE 之自由度為 $n-k = 15-3 = 12$

➢ 因子平均變異 MSF 與隨機平均變異 MSE

$$MSF = \tfrac{SSF}{k-1} = \tfrac{463.33}{2} = 231.66$$

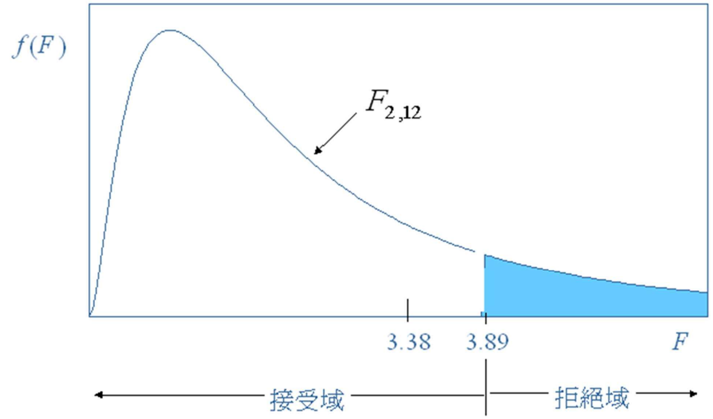
$$MSE = \tfrac{SSE}{n-k} = \tfrac{822}{12} = 68.5$$

➢ 樣本 F 統計量：

$$F = \tfrac{MSF}{MSE} = \tfrac{231.66}{68.5} = 3.38$$

➢ 變異數分析表：以上結果可整理成下表

變異來源	平方和（SS）	自由度（df）	平均平方和（MS）	F 值
因子	463.33	2	231.66	3.38
隨機	822	12	68.5	
總和	1285.33	14		



➢ 臨界值：根據 $F_{k-1,n-k} = F_{2,12}$ 分配所找出之右尾 $\alpha = 0.05$ 臨界值為 3.89。

➢ 結論：由於樣本 F 統計量(3.38)小於臨界值 3.89，故無法拒絕虛無假設；亦即，『在 5%的顯著水準下，三廠牌填裝機每小時平均填充量無顯著差異』。

● 例子：全家便利商店的店長想了解飲料陳列位置與銷售量間的關係，將同一飲料陳列於冷飲冰櫃中的高、中、低區，各陳列 5 日，5 日的銷售結果如下表。請在顯著水準 $\alpha = 0.05$ 下，檢定 3 個陳列位置的平均日銷售量是否有差異？

1	樣本觀察值	陳列位置		
2		低	中	高
3	1	85	105	87
4	2	89	91	90
5	3	91	118	92
6	4	90	116	97
7	5	88	112	95
8	樣本平均數	88.6	113.2	92.2
9	樣本變異數	5.3	25.7	15.7

- 令 μ_1、μ_2、μ_3 分別為 3 個陳列位置(低、中、高)的平均日銷售量。根據上表可以整理出底下資訊

低	中	高
$n_1 = 5$	$n_2 = 5$	$n_3 = 5$
$\overline{Y}_1 = 88.6$	$\overline{Y}_2 = 113.2$	$\overline{Y}_2 = 92.2$
$S_1^2 = 5.3$	$S_2^2 = 25.7$	$S_3^2 = 15.7$

- 虛無假設與對立假設

  $H_0 : \mu_1 = \mu_2 = \mu_3$ [3 個陳列位置的平均日銷售量相同]

  $H_1 : \mu_i$ 不全相等 [3 個陳列位置的平均日銷售量不全相同]

- 總樣本個數：$n = \sum_{i=1}^{k} n_i = 5 + 5 + 5 = 15$
- 總樣本平均數：

$$\overline{\overline{Y}} = \tfrac{1}{n}\sum_{i=1}^{k} n_i \overline{Y}_i = \tfrac{1}{15}(5 \cdot 88.6 + 5 \cdot 113.2 + 5 \cdot 92.2) = 98$$

- 因子平方和(SSF)：

$$SSF = \sum_{i=1}^{k} n_i (\overline{Y}_i - \overline{\overline{Y}})^2$$

$$= 5(88.6 - 98)^2 + 5(113.2 - 98)^2 + 5(92.2 - 98)^2 = 1765.2$$

  SSF 之自由度為 $k - 1 = 3 - 1 = 2$

- 隨機平方和(SSE)：

$$SSE = \sum_{i=1}^{k} (n_i - 1) S_i^2$$

$$= (5-1)5.3 + (5-1)25.7 + (5-1)15.7 = 186.8$$

  SSE 之自由度為 $n - k = 15 - 3 = 12$

- 因子平均變異 MSF 與隨機平均變異 MSE

$$MSF = \tfrac{SSF}{k-1} = \tfrac{1765.2}{2} = 882.6$$

$$MSE = \tfrac{SSE}{n-k} = \tfrac{186.8}{12} = 15.57$$

- 樣本 F 統計量：

$$F = \tfrac{MSF}{MSE} = \tfrac{882.6}{15.57} = 56.70$$

- 變異數分析表：以上結果可整理成下表

變異來源	平方和(SS)	自由度(df)	平均平方和(MS)	F值
因子	1,765.20	2	882.6	56.7
隨機	186.8	12	15.57	
總和	1,952.00	14		

- 臨界值：根據 $F_{k-1,n-k} = F_{2,12}$ 分配所找出之右尾 $\alpha = 0.05$ 臨界值為 3.89。
- 結論：由於樣本 F 統計量(56.7)大於臨界值 3.89，故拒絕虛無假設，接受對立假設；亦即，『在 5%的顯著水準下，3 個陳列位置的平均日銷售量具有顯著差異』。