# One-Way ANOVA

## ~ Analysis of Variance (ANOVA) ~

單因子的變異數分析

# 單因子的變異數分析
# ~完全隨機

基本變異數分析表

# 變異數分析的步驟建議

**第1步**

設立兩個假設：

$$\begin{cases} \mathrm{H_0}: \mu_1 = \mu_2 = \cdots = \mu_\mathrm{k} = \mu \\ \mathrm{H_1}: \mu_1, \mu_2, \cdots\cdots \mu_\mathrm{k} \end{cases}$$

**第2步**

建立變異數分析表（ANOVA table），並利用「F檢定」做檢定：

$$F^* = \frac{MSA}{MSE}$$

**第3步**

下結論並做決策或推論：

$$F^* > F_{k-1, n_T-k, \alpha} \rightarrow \text{拒絕} \mathrm{H_0}$$

$$F^* \leq F_{k-1, n_T-k, \alpha} \rightarrow \text{不拒絕} \mathrm{H_0}$$

# 總變異 = 組間變異 + 組內變異( SST = SSA + SSE)

## SST (總變異，sum of squares total)

$$\sum_{j=1}^{k}\sum_{i=1}^{n_j}\left(x_{ij}-\bar{\bar{x}}\right)^2 = n_T \cdot \left[\frac{1}{n_T}\sum_{j=1}^{k}\sum_{i=1}^{n_j}\left(x_{ij}-\bar{\bar{x}}\right)^2\right] = n_T \cdot \boldsymbol{\sigma}_T^2 = (n_T-1)S_T^2$$

## SSA(組間變異，sum of squares among groups)

$$\sum_{j=1}^{k}\sum_{i=1}^{n_j}\left(\bar{x}_j-\bar{\bar{x}}\right)^2 = \sum_{j=1}^{k} n_j\bar{x}_j^2 - n_T\bar{\bar{x}}^2 = n \times k \times \left[\frac{1}{k}\sum_{j=1}^{k}\bar{x}_j^2-\bar{\bar{x}}^2\right] = n_T \cdot \boldsymbol{\sigma}_{\bar{x}}^2$$

(假設每個抽樣樣本數都相同，$n=n_j$ )

## SSE (組內變異，sum of squares error)

$$\sum_{j=1}^{k}\sum_{i=1}^{n_j}\left(x_{ij}-\bar{x}_j\right)^2 = \sum_{j=1}^{k} n_j \left[\frac{1}{n_j}\sum_{i=1}^{n_j}\left(x_{ij}-\bar{x}_j\right)^2\right] = \sum_{j=1}^{k} n_j \cdot \boldsymbol{\sigma}_j^2 = \sum_{j=1}^{k}(n_j-1)\cdot S_j^2$$

# 變異數分析表(ANOVA Table)

**MSA**
(樣本間均方)
(mean square between)

$$MSA=\frac{\sum_{j=1}^{k}\sum_{i=1}^{n_j}(\bar{x}_j-\bar{\bar{x}})^2}{k-1}=\frac{SSA}{k-1}$$

**MSE**
(樣本內均方)
(mean square within)

$$MSE=\frac{\sum_{j=1}^{k}\sum_{i=1}^{n_j}(x_{ij}-\bar{x}_j)^2}{n_T-k}=\frac{SSE}{n_T-k}$$

$$F^*=\frac{MSA}{MSE}\sim F_{k-1,\,n_T-k}$$

| 變異來源 | 平方和 | 自由度 | 平均平方和 | F檢定 |
|---|---|---|---|---|
| 因子(組間) | SSA | k-1 | MSA | $F^*=\frac{MSA}{MSE}$ |
| 隨機(組內) | SSE | $n_T$-k | MSE | |
| 總和 | SST | $n_T$-1 | | |

# 案例1

為瞭解台灣地區稻米的產量是否會因北、中、南產生差異，於是自北、中、南隨機抽取若干樣本，得到資料如下所示：（單位：公噸/公畝）

| 地區 | 產量 | | | |
|---|---|---|---|---|
| 北部 | 1 | 3 | 2 | |
| 中部 | 3 | 2 | 4 | 2 |
| 南部 | 7 | 6 | 4 | |

若母體滿足變異數分析之假設，請你利用上述資料檢定，在顯著水準0.05下，台灣區稻米的產量是否會因北、中、南產生差異？

「應用統計學 二版」p349，李德治、童惠玲，博碩文化

# 案例1 解說(a)

(1)假設1,2,3分別代表北部、中部、南部，所以假設檢定：

$$\begin{cases} H_0：\mu_1 = \mu_2 = \mu_3 = \mu \\ H_1：\mu_1, \mu_2, \mu_3 \text{ 不全相等} \end{cases}$$

(2)計算各平方和

| 地區 | 產量 | | | | $\bar{x}_j$ | $S_j^2$ |
|---|---|---|---|---|---|---|
| 北部 | 1 | 3 | 2 | | 2 | 1 |
| 中部 | 3 | 2 | 4 | 2 | 2.75 | 0.917 |
| 南部 | 7 | 6 | 4 | | 5.67 | 2.333 |
| | | | | | $\bar{\bar{x}} = 3.4$ | $\sigma_T^2 = 3.24$ |

$$SST = n_T \cdot \boldsymbol{\sigma}_T^2 = (n_T - 1)S_T^2 = 10 \times 3.24 = 32.4$$

$$SSE = \sum_{j=1}^{k} (n_j - 1) \cdot S_j^2 = 2 \times 1 + 3 \times 0.917 + 2 \times 2.33 = 9.417$$

$$SSA = SST - SSE = 32.4 - 9.417 = 22.983$$

# 案例1 解說(b)

(3)計算ANVOA Table

| 變異來源 | 平方和 | 自由度 | 平均平方和 | F檢定 |
|---|---|---|---|---|
| 因子(組間) | SSA=22.983 | 2 | MSA=11.492 | $F^* = \frac{MSA}{MSE} = \frac{11.492}{1.345} = 8.544$ |
| 隨機(組內) | SSE=9.417 | 7 | MSE=1.345 | |
| 總和 | SST=32.4 | 9 | | |

在顯著水準0.05下，$F^* = 8.544 > F_{0.05,2,7} = 4.74$，故拒絕$H_0$，即北、中、南三區稻米產量有顯著差異。

# 案例2

某企業想要研究三種新進員工職前教育訓練課程(1)電視教學；(2)講師講習；(3)實地觀摩，對學習效果是否有影響。於是將新進員工隨機分成三組，分別施以不同職前教育訓練課程，結訓後的實習成績如下表所示，請問在顯著水準0.05下，三種課程方式是否會影響實習成績？

| 訓練課程 | 實習成績 | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 電視教學 | 70 | 83 | 88 | 92 | 85 | 80 | 90 | |
| 講師講習 | 76 | 85 | 80 | 90 | 85 | 88 | 90 | 94 |
| 實地觀摩 | 82 | 80 | 75 | 89 | 70 | 72 | | |

「現代統計學」p205，吳柏林 著，五南圖書

# 案例2 解說

(1)建立假設檢定。假設各訓練方式實習成績平均為：$\mu_1$ ~ 電視教學；$\mu_2$ ~ 講師講習；$\mu_3$ ~ 實地觀摩；

$$\begin{cases} H_0: \mu_1 = \mu_2 = \mu_3 = \mu \\ H_1: \mu_1, \mu_2, \mu_3 \text{ 不全相等} \end{cases}$$

(2)計算ANVOA Table

| 訓練課程 | 實習成績 | | | | | | | | $\bar{x}_j$ |
|---|---|---|---|---|---|---|---|---|---|
| 電視教學 | 70 | 83 | 88 | 92 | 85 | 80 | 90 | | 84 |
| 講師講習 | 76 | 85 | 80 | 90 | 85 | 88 | 90 | 94 | 86 |
| 實地觀摩 | 82 | 80 | 75 | 89 | 70 | 72 | | | 78 |

| 變異來源 | 平方和 | 自由度 | 平均平方和 | F檢定 |
|---|---|---|---|---|
| 因子(組間) | SSA=229 | 2 | MSA=114.5 | $F^* = \frac{MSA}{MSE} = \frac{114.5}{45.44} = 2.52$ |
| 隨機(組內) | SSE=818 | 18 | MSE=45.44 | |
| 總和 | SST=1047 | 20 | | |

(3)顯著水準0.05下，$F^* = 2.52 < F_{0.05,2,18} = 3.55$，故接受$H_0$，即此三種教育訓練課程對實習成績並無影響。

# 案例3

市調公司接受一委託案，調查市面上相同屬性的四種不同品牌飲料，其銷售情形是否有顯著差異。該公司乃選擇20個消費傾向類似的地區，且每一品牌飲料隨機指定其中五個不重複的地區做調查。下列資料是每一品牌飲料在各該地區平均每一千人口的銷售箱數：

| 品牌 | 銷售箱數 | | | | |
|---|---|---|---|---|---|
| 甲 | 31 | 28 | 30 | 27 | 29 |
| 乙 | 26 | 28 | 25 | 29 | 27 |
| 丙 | 31 | 29 | 32 | 32 | 31 |
| 丁 | 27 | 25 | 28 | 24 | 26 |

(A)市調人員擬以變異數分析做研究，試問應有哪些前提假設？

(B)若在顯著水準0.05下，品牌銷售量是否有顯著差異？

「應用統計學 二版」p347，李德治、童惠玲，博碩文化

# 案例3 解說(a)

(A)變異數分析需滿足下列三個條件：

1. 常態性假設～每個小母體分配均為常態分配。
2. 同質性假設～各常態分配的變異數皆相等。
3. 獨立性假設～抽樣方法為簡單隨機抽樣，被抽樣的樣本彼此獨立。

(B) (1)建立假設檢定：

$$\begin{cases} \mathrm{H}_0: \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu \\ \mathrm{H}_1: \mu_1, \mu_2, \mu_3, \mu_4 \text{不全相等} \end{cases}$$

(2)計算ANVOA Table ：

| 品牌 | 銷售箱數 | | | | | 樣本數 | 樣本和 | 樣本平均數 |
|---|---|---|---|---|---|---|---|---|
| 甲 | 31 | 28 | 30 | 27 | 29 | 5 | 145 | 29 |
| 乙 | 26 | 28 | 25 | 29 | 27 | 5 | 135 | 27 |
| 丙 | 31 | 29 | 32 | 32 | 31 | 5 | 155 | 31 |
| 丁 | 27 | 25 | 28 | 24 | 26 | 5 | 130 | 26 |

總平均 $\bar{\bar{x}} = {}^{(29+27+31+26)}/_{4} = 28.25$

# 案例3 解說(b)

(2)計算ANVOA Table

| 變異來源 | 平方和 | 自由度 | 平均平方和 | F檢定 |
|---|---|---|---|---|
| 品牌(組間) | SSA=73.75 | 4-1=3 | MSA=24.58 | $F^* = \frac{MSA}{MSE} = \frac{24.58}{2.25} = 10.926$ |
| 隨機(組內) | SSE=36 | 20-4=16 | MSE=2.25 | |
| 總和 | SST=109.75 | 20-1=19 | | |

其中 $$SSA = \sum_{j=1}^{k}\sum_{i=1}^{n_j}(\bar{x}_j - \bar{\bar{x}})^2 = \sum_{j=1}^{k} n_j\bar{x}_j^2 - n_T\bar{\bar{x}}^2$$

$$= [5(29)^2 + 5(27)^2 + 5(31)^2 + 5(26)^2] - 20(28.25)^2 = 73.75$$

SSE = SST – SSA = 109.75 - 73.75 = 36

顯著水準0.05下，$F^*$ =10.926 > $F_{0.05,3,16}$=3.24，故拒絕$H_0$，即此四種品牌銷售均數有顯著差異。

# The End