

---

# ANOVA

## Analysis of variance

## 變異數分析

**Jong Yih Kuo**

**[jykuo@ntut.edu.tw](mailto:jykuo@ntut.edu.tw)**

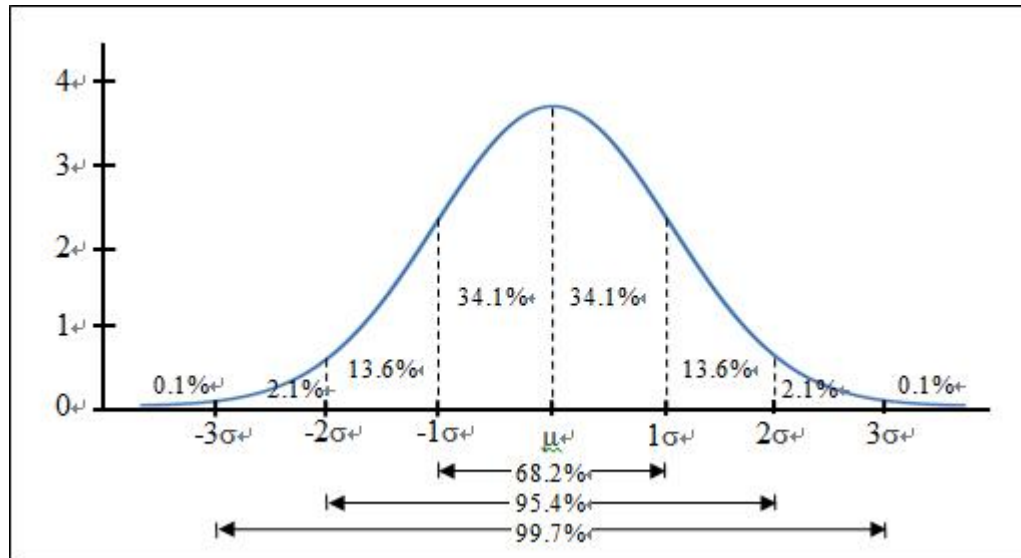
**Department of Computer Science and  
Information Engineering  
National Taipei University of Technology**

# 基本概念

- ❑ ANOVA 透過對變異數的分析，同時檢定 $k$ 個母體之平均數 $\mu_i$ ， $i = 1, 2, \dots, k$ ，是否有顯著的差異，或檢定因子 (Factor)/自變數對依變數是否有顯著的影響。
- ❑ T-test只能用在兩個組別的檢測，若三個或三個以上的組別需要檢驗時就必須使用ANOVA
- ❑ ANOVA無法比較哪些組別有顯著的差異

# 三大假設

- ❑ 每個母體均為常態分布
- ❑ 來自每個母體的隨機樣本均互為獨立
- ❑ 每個母體的變異數須同質(變異數無顯著差異)



# 檢定步驟

## □ 設定虛無假設與對立假設

- $H_0: \mu_0 = \mu_1 = \dots = \mu_n$
- $H_1$ : 至少有兩個平均值不相等

$n_j$  = 第j組的樣本數  
 $\bar{X}_j$  = 第j組的平均值

## □ 計算F值

$$F = \frac{MSB}{MSW} = \frac{\frac{SS_B}{k-1}}{\frac{SS_W}{n-k}}$$

$$SS_B = \sum_{j=1}^k n_j (\bar{X}_j - \bar{\bar{X}})^2$$

$$SS_W = \sum_{j=1}^k \sum_{i=1}^{n_j} (X_{ji} - \bar{X}_j)^2$$

k = 組別數量, n = 總樣本數

SS = 平方和, B代表組間、W代表組內

$df_b$  = 組間自由度 = k-1,  $df_w$  = 組內自由度 = n-k

MSB/MSW =  $SS_B/SS_E$  除以各自的自由度

SS: Sum of Squares

MSB: Mean Sum of Square Between Groups

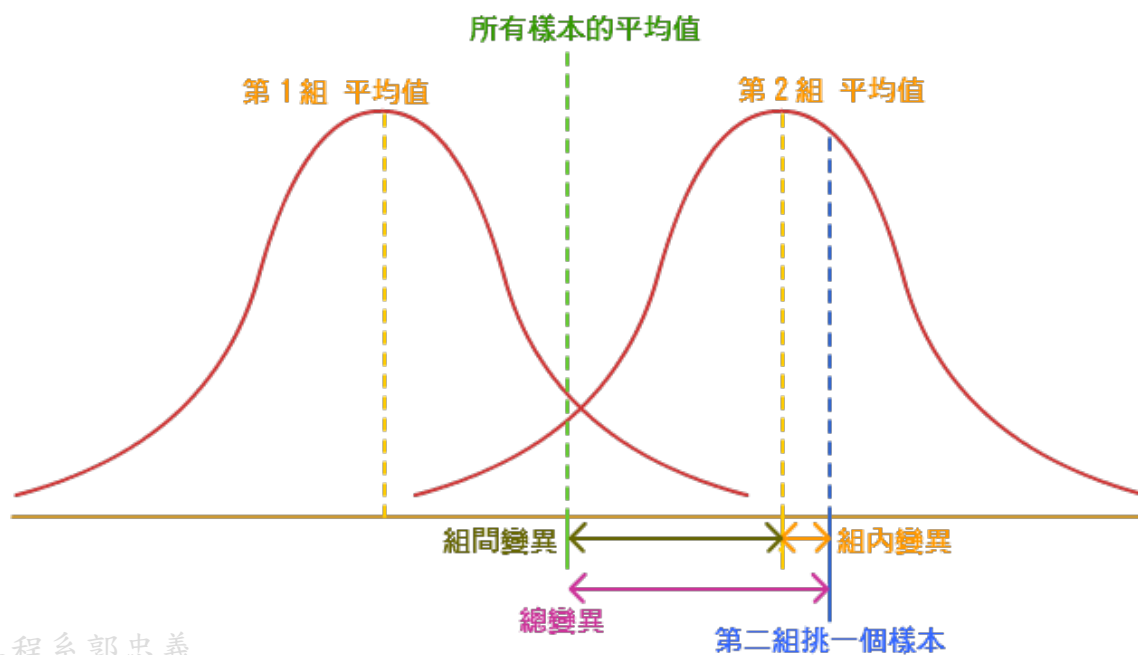
MSW: Mean Sum of Square Within Groups

(各組平均值－總平均值)<sup>2</sup>的總和

(組內各樣本值－該組平均值)<sup>2</sup>的總和

# 檢定步驟

- 當  $F$  值越大，表示組平均數的分散情形越嚴重。
- 設定顯著水準  $\alpha$
- 判斷或推翻虛無假設
  - 若  $F > F_{Crit}$ ，即可拒絕虛無假設。
  - $F < F_{Crit}$ ，不拒絕虛無假設。



# 範例一

- 假設某種健康的植物，抽出15顆並分成三組，每一組都在一種溫度下安全生長並記錄壽命，想了解溫度對於植物的壽命有無顯著影響？

2	10	10
3	8	13
7	7	14
2	5	13
6	10	15

- 這三組樣本的平均數分別為：
  - $<10^{\circ}\text{C}$  : 4     $10^{\circ}\text{C}\sim 20^{\circ}\text{C}$  : 8     $>20^{\circ}\text{C}$  : 13
  - 三樣本平均值(15個值的平均值) : 8.3

# 範例一

## □ 組間變異(SSB)：

2
3
7
2
6

10
8
7
5
10

10
13
14
13
15

$$4 - 8.3 = (-4.3)^2 \quad 8 - 8.3 = (-.3)^2 \quad 13 - 8.3 = (4.7)^2$$

$$18.8 + .1 + 21.8 = 40.7$$

$$40.7 \times 5 = 203.3$$

## □ 組內變異(SSE)：

sample

2	- 4 = -2 <sup>2</sup>
3	- 4 = -1 <sup>2</sup>
7	- 4 = 3 <sup>2</sup>
2	- 4 = -2 <sup>2</sup>
6	- 4 = 2 <sup>2</sup>

$$\begin{array}{r} 4 \\ 1 \\ 9 \\ 4 \\ 4 \\ \hline 22 \end{array}$$

sample

10	- 8 = 2 <sup>2</sup>
8	- 8 = 0 <sup>2</sup>
7	- 8 = -1 <sup>2</sup>
5	- 8 = -3 <sup>2</sup>
10	- 8 = 2 <sup>2</sup>

$$\begin{array}{r} 4 \\ 0 \\ 1 \\ 9 \\ 4 \\ \hline 18 \end{array}$$

sample

10	- 13 = -3 <sup>2</sup>
13	- 13 = 0 <sup>2</sup>
14	- 13 = 1 <sup>2</sup>
13	- 13 = 0 <sup>2</sup>
15	- 13 = 2 <sup>2</sup>

$$\begin{array}{r} 9 \\ 0 \\ 1 \\ 0 \\ 4 \\ \hline 14 \end{array}$$

$$22 + 18 + 14 = 54$$

# 範例一

□ 總變異數(SST) = SSB+SSE

2	-	8.3	= -6.3	40.1
3	-	8.3	= -5.3	28.4
7	-	8.3	= -1.3	1.8
2	-	8.3	= -6.3	40.1
6	-	8.3	= -2.3	5.4
10	-	8.3	= 1.7	2.7
8	-	8.3	= -0.3	0.1
7	-	8.3	= -1.3	1.8
5	-	8.3	= -3.3	11.1
10	-	8.3	= 1.7	2.8
10	-	8.3	= 1.7	2.8
13	-	8.3	= 4.7	21.8
14	-	8.3	= 5.7	32.1
13	-	8.3	= 4.7	21.8
15	-	8.3	= 6.7	44.4
sum				257.3



# 範例一

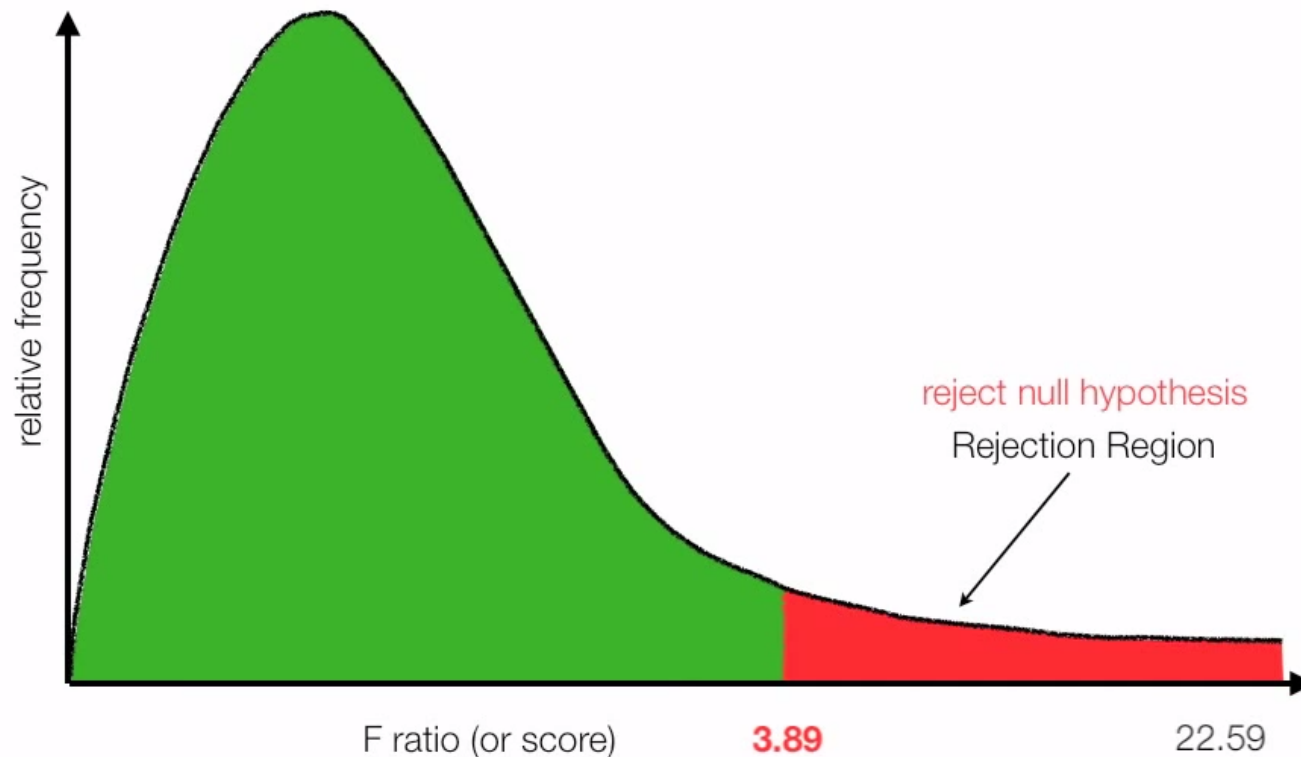
- 組間自由度( $df_b$ )：
  - 組別數量 $-1=3-1=2$
- 組內自由度( $df_w$ )：
  - 總樣本數 $-$ 組別數量 $=15-3=12$
- 總自由度( $df_s$ ) $=df_b+df_w=n-1$ 
  - $n$ 為總樣本數

$$F = \frac{203.3 \div 2}{54 \div 12} = \frac{101.667}{4.5} = 22.59$$

□  $F(2,12)$  且顯著水準=0.05，查下表：(F Distribution)

df w \ df b	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	161.5	199.5	215.7	224.6	230.2	234.0	236.8	238.9	240.5	241.9	243.9	246.0	248.0	249.1	250.1
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40	19.41	19.43	19.45	19.45	19.46
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.74	8.70	8.66	8.64	8.62
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.91	5.86	5.80	5.77	5.75
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.68	4.62	4.56	4.53	4.50
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.00	3.94	3.87	3.84	3.81
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.57	3.51	3.44	3.41	3.38
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.28	3.22	3.15	3.12	3.08
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.07	3.01	2.94	2.90	2.86
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.91	2.85	2.77	2.74	2.70
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.79	2.72	2.65	2.61	2.57
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.69	2.62	2.54	2.51	2.47
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.60	2.53	2.46	2.42	2.38
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.53	2.46	2.39	2.35	2.31
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.48	2.40	2.33	2.29	2.25

- $F > 3.89 \rightarrow$  分析結果為顯著，可拒絕虛無假設
- $F < 3.89 \rightarrow$  分析結果為不顯著，不可拒絕虛無假設



# 範例二

- 了解不同年齡層 A組20~29歲，B組30~39歲，C組40~49歲，對筆記型電腦喜好程度是否有顯著差異(顯著水準為0.05)，隨機抽取年齡層各5個人，以1-10的分數請他們評分如下：

A	B	C
8	8	4
4	5	4
5	9	6
5	7	6
4	9	5

## 範例二

$$\bar{A} = \frac{8 + 4 + 5 + 5 + 4}{5} = 5.2$$

$$\bar{B} = \frac{8 + 5 + 9 + 7 + 9}{5} = 7.6$$

$$\bar{C} = \frac{4 + 4 + 6 + 6 + 5}{5} = 5$$

$$\text{總平均} = \frac{5.2 + 7.6 + 5}{3} = 5.9$$

$$SSB = 5(5.2 - 5.9)^2 + 5(7.6 - 5.9)^2 + 5(5 - 5.9)^2 = 20.95$$

$$\begin{aligned} SSE = & (8 - 5.2)^2 + (4 - 5.2)^2 + (5 - 5.2)^2 + (5 - 5.2)^2 + (4 - 5.2)^2 \\ & + (8 - 7.6)^2 + (5 - 7.6)^2 + (9 - 7.6)^2 + (7 - 7.6)^2 + (9 - 7.6)^2 \\ & + (4 - 5)^2 + (4 - 5)^2 + (6 - 5)^2 + (6 - 5)^2 + (5 - 5)^2 = 26 \end{aligned}$$

$$df_b = 3 - 1 = 2$$

$$df_w = 15 - 3 = 12$$

$$F = \frac{\frac{20.95}{2}}{\frac{26}{12}} = 4.83$$

$$F_{0.05}(2, 12) = 3.89$$

$$\because F > F_{0.05}(2, 12)$$

$\therefore$  分析結果為顯著，拒絕 $H_0$