

【统计简单学】

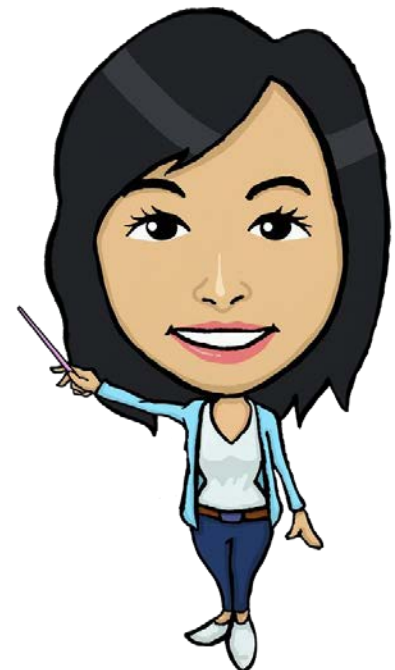
第七单元 变异数分析

授课教师：唐丽英 教授

新竹交通大学 工业
工程与管理学系

第七单元 内容大纲

- 第一部份：变异数分析介绍
 - 变异数分析简介
 - 变异数分析之统计假设
 - 变异数分析之执行步骤
 - 变异数分析之计算流程
- 第二部份：变异数分析范例



第一部份：变异数分析介绍

变异数分析简介

- 何谓变异数分析？

- 变异数分析（**Analysis of Variance, ANOVA**）是两组独立样本t检定（假设两个群体变异数相等）之延伸。
- 此法可**同时检定三个或三个以上**独立群体的平均数间是否有显著差异，即：检定多个群体平均数($\mu_1=\mu_2=\mu_3=\dots=\mu_a$)是否相等。

- 例1：

比较三所学校学生每月平均花费是否有显著差异					
交通大学		清华大学		台湾大学	
交大学生1	\$5,500	清大学生1	\$7,200	台大学生1	\$8,500
交大学生2	\$6,500	清大学生2	\$5,100	台大学生2	\$6,800
...
交大学生99	\$3,800	清大学生99	\$6,900	台大学生99	\$5,200
交大学生100	\$7,200	清大学生100	\$4,300	台大学生100	\$7,300

变异数分析之统计假设

1) a组数据是随机抽自a个独立之常态母体

2) $\sigma_1^2 = \sigma_2^2 = \dots = \sigma_a^2$

变异数分析之执行步骤

1) 設立虛無假說 (H_0) 及對立假說 (H_a)

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

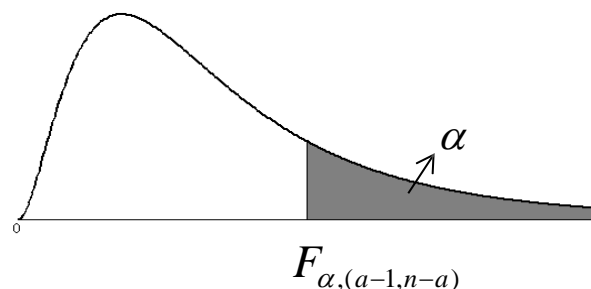
H_a : 至少有一 μ 值不同。

2) 指定顯著水準(level of significance, α .)

3) 決定適當之檢定樣本量(test statistic): F test

$$F = \frac{MS_B}{MS_E}$$

4) 決定棄卻域: 查 F-表, 自由度 d.f. = $(a-1, N-a)$, 其中 $N = n_1 + n_2 + \cdots + n_a$



5) 下結論-推翻虛無假說(reject H_0)或不推翻虛無假說, 並將此結論按題意引申。

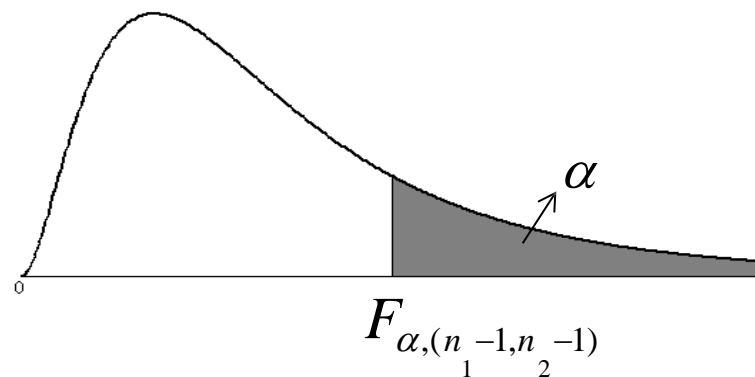
F分布 (F-Distribution)

- 由两个变异数相等之独立常态群体中，以重覆取样方式，分别随机抽取样本大小为 n_1 及 n_2 之二组样本，则

$\frac{S_1^2}{S_2^2}$ 会服从 **F分布**，

其自由度为 **$v_1 = n_1 - 1$** ， **$v_2 = n_2 - 1$** 。

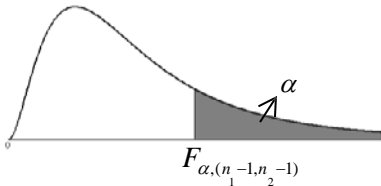
$$\text{即： } \frac{S_1^2}{S_2^2} \sim F_{v_1, v_2} = F_{n_1 - 1, n_2 - 1}$$



F分布 (F-Distribution)

- 例2：假设 $n_1 = 7, n_2 = 13$ ，利用下列表查 $F_{0.01} = ?$

$\alpha = 0.01$



		Numerator Degrees of Freedom							
v ₂	v ₁	1	2	3	4	5	6	7	8
	1	4052.2	4999.5	5403.4	5624.6	5763.7	5859.0	5928.4	5981.1
Denominator Degrees of Freedom	2	98.50	99.00	99.17	99.25	99.30	99.33	99.36	99.37
	3	34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.49
	4	21.20	18.00	16.69	15.98	15.52	15.21	14.98	14.80
	5	16.26	13.27	12.06	11.39	10.97	10.67	10.46	10.29
	6	13.75	10.92	9.78	9.15	8.75	8.47	8.26	8.10
	7	12.25	9.55	8.45	7.85	7.46	7.19	6.99	6.84
	8	11.26	8.65	7.59	7.01	6.63	6.37	6.18	6.03
	9	10.56	8.02	6.99	6.42	6.06	5.80	5.61	5.47
	10	10.04	7.56	6.55	5.99	5.64	5.39	5.20	5.06
	11	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74
	12	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50
	13	9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30
	14	8.86	6.51	5.56	5.04	4.69	4.46	4.28	4.14
	15	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00

【解】 $F_{0.01, (6, 12)} = 4.82$

变异数分析之计算流程

1) 符号：

	1	2	3	· · ·	i	· · ·	a
样本大小	n_1	n_2	n_3	· · ·	n_i	· · ·	n_a
样本和	$y_{1\cdot}$	$y_{2\cdot}$	$y_{3\cdot}$	· · ·	$y_{i\cdot}$	· · ·	$y_{a\cdot}$
样本平均数	y_1	y_2	y_3	· · ·	y_i	· · ·	y_a

变异数分析之计算流程

2) 公式

$$\text{总变异} = \text{组间变异} + \text{组内变异}$$

$$\text{总变异} = SS_T = \sum_i \sum_j (y_{ij} - \bar{y}_{..})^2 = \sum_i \sum_j y_{ij}^2 - CM$$

$$\text{其中 } CM = y_{..} / N, (N = n_1 + n_2 + \dots + n_a)$$

$$\text{组间变异} = SS_B = \sum_t n_i (\bar{y}_{i.} - \bar{y}_{..})^2 = \sum_t y_i^2 / n_i - CM$$

$$\begin{aligned} \text{组内变异} = SS_E &= \sum_i \sum_j (y_{ij} - \bar{y}_{i.})^2 = \sum_i \sum_j y_{ij}^2 - \sum_i y_i^2 / n_i \\ &= SS_T - SS_B \end{aligned}$$

变异数分析之计算流程

3) 变异数分析表(ANOVA table)

变异来源	平方和	自由度	均方	F值
组间之变异	SS_B	$a-1$	$MS_B = SS_B / (a-1)$	$F = MS_B / MS_E$
组内之变异	SS_E	$N-a$	$MS_E = SS_E / (N-a)$	
总变异	SS_T	$N-1$		

$$(N = n_1 + n_2 + \dots + n_a)$$

第二部份：变异数分析范例

变异数分析范例（一）

- **例3**：某实验室研究人员欲探讨四种不同之催化剂对某混合液体的浓缩度(Concentrations) 是否有显著之差异。其实验结果如下表所示，试分析之。

单位：浓缩度

催化剂1	催化剂2	催化剂3	催化剂4
16	18	20	11
11	17	24	10
17	13	25	15
12	15	23	
	12		

变异数分析范例（一）

【解】

	催化剂1	催化剂2	催化剂3	催化剂4	合计
样本大小 n_i	4	5	4	3	N=16
样本和 y_i	56	75	92	36	$y_{..} = 259$
样本平均数 y_i	14	15	23	12	

1) H_0 ：四种催化剂间无差异

H_a ：四种催化剂间有差异

2) $\alpha=0.05$

变异数分析范例（一）

3) 检定统计量：F=?

ANOVA 表

变异来源	平方和	自由度	均方	F值
组间之变异	264.44	3	88.15	13.22 检定统计量
组内之变异	80.00	12	6.67	
总变异	344.44	15		

$$(1) CM(\text{校正因子}) = \frac{y_{..}^2}{N} = \frac{259^2}{16} = 4192.5625$$

$$(2) \sum_i \sum_y y_{ij}^2 = 16^2 + 11^2 + 17^2 + \cdots + 11^2 + 10^2 + 15^2 = 4537$$

$$(3) \sum_i \frac{y_i^2}{n_i} = \frac{56^2}{4} + \frac{75^2}{5} + \frac{92^2}{4} + \frac{36^2}{3} = 4457$$

$$SST = \sum_i \sum_y y_{ij}^2 - CM = (2) - (1) = 344.4375$$

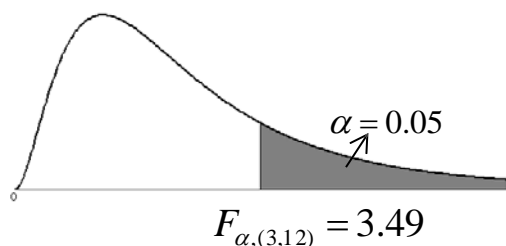
$$SSB = \sum_i \frac{y_i^2}{n_i} - CM = (3) - (1) = 264.4375$$

$$SSE = SST - SSB = 80.00$$

变异数分析范例（一）

$$\alpha = 0.05$$

4) 弃却域：



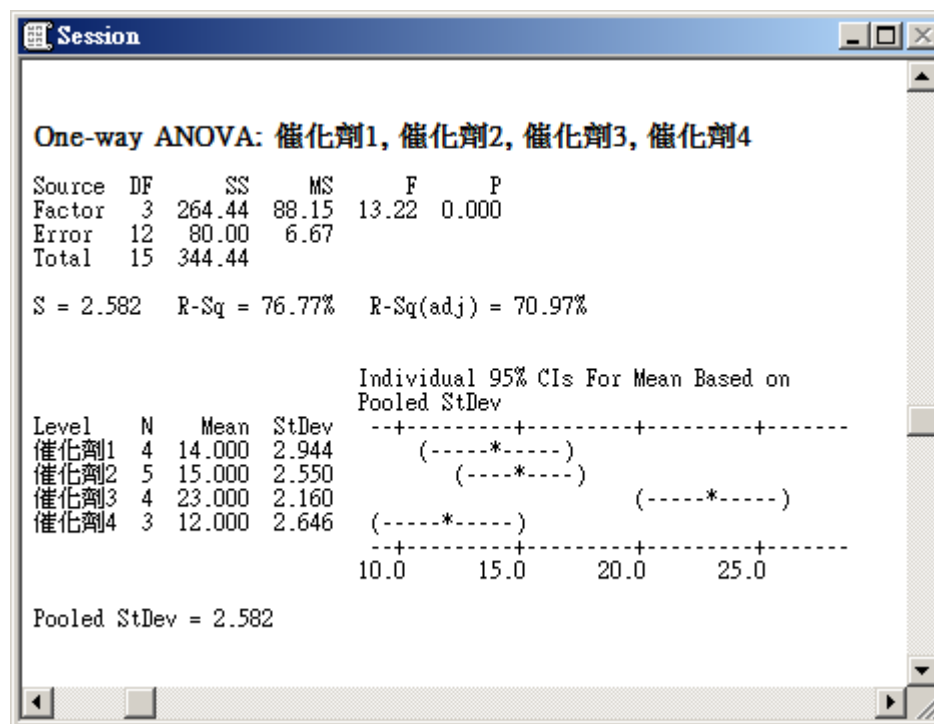
因 $F > F_{\alpha, (3, 12)}$ ，拒绝 H_0

$v_1 \backslash v_2$						
		1	2	3	4	5
1		161.4476	199.5	215.7073	224.5832	230.1619
2		18.5128	19	19.1643	19.2468	19.2964
3		10.128	9.5521	9.2766	9.1172	9.0135
4		7.7086	6.9443	6.5914	6.3882	6.2561
5		6.6079	5.7861	5.4095	5.1922	5.0503
6		5.9874	5.1433	4.7571	4.5337	4.3874
7		5.5914	4.7374	4.3468	4.1203	3.9715
8		5.3177	4.459	4.0662	3.8379	3.6875
9		5.1174	4.2565	3.8625	3.6331	3.4817
10		4.9646	4.1028	3.7083	3.478	3.3258
11		4.8443	3.9823	3.5874	3.3579	3.2058
12		4.7472	3.8853	3.4903	3.2603	3.1082
13		4.6672	3.8056	3.4105	3.1803	3.0282

5) 结论：四种催化剂有显著之差异。

变异数分析范例（一）

【Minitab 计算机报表】



藉由计算机报表中各组样本平均数之95%信赖区间，可以进一步找出哪两个或哪几个群体平均数间有显著差异。

结论：四种催化劑中，**催化劑3之浓缩度显著高于其他三种**。
催化劑1、催化劑2及催化劑3之间则无显著差异。

变异数分析范例（二）

- **例4**：某行动电话业者推出某种新手机，且欲比较此手机在台北、台中、高雄、宜兰四个地区的销售量是否有明显差异，他由四个地区的销售量中随机抽取三周的销售量，得资料如下表所示。

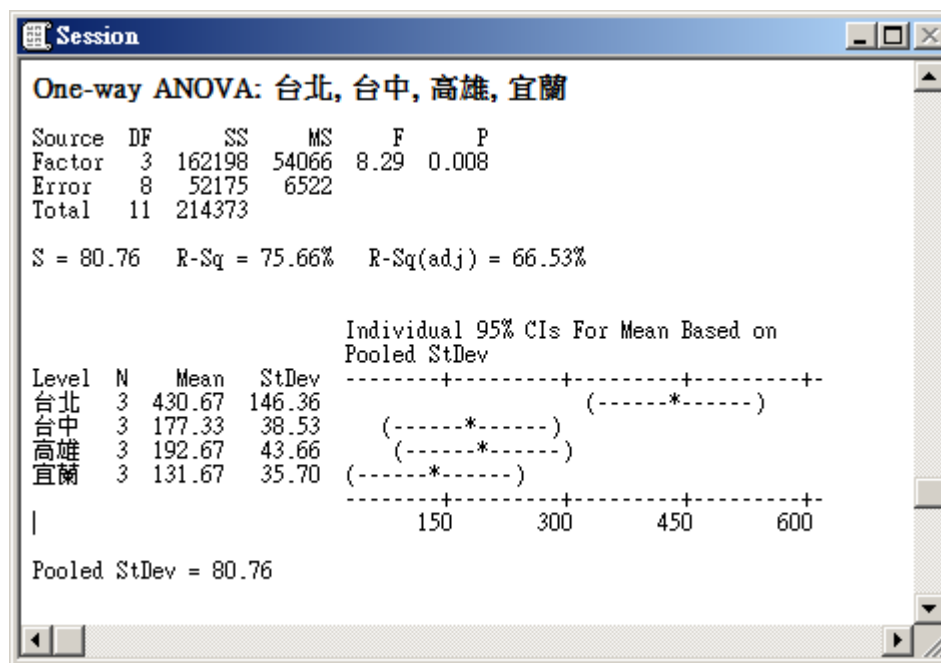
地区	销售量		
台北	276	567	449
台中	143	170	219
高雄	234	197	147
宜兰	94	136	165

试问新手机在这四个地区的销售量是否有显著差异？

变异数分析范例（二）

- 例4：利用 Minitab 电脑报表求解：

【Minitab 计算机报表】



1. 由上表可知四个地区新产品的平均销售量间有显著差异（其p值=0.008<0.05）。
2. 进一步检视四个区域之信赖区间，可知**台北地区的平均销售量显著地高于其他三个地区之平均销售量**，而台中地区、高雄地区与宜兰地区之平均销售量间并无显著差异。

变异数分析范例（三）

- 例5：某品保工程师欲比较A牌、B牌与C牌灯泡之平均寿命间是否有显著差异。现分别用这三种厂牌之灯泡各进行20个加速寿命测试实验，得到灯泡寿命数据如下表所示。

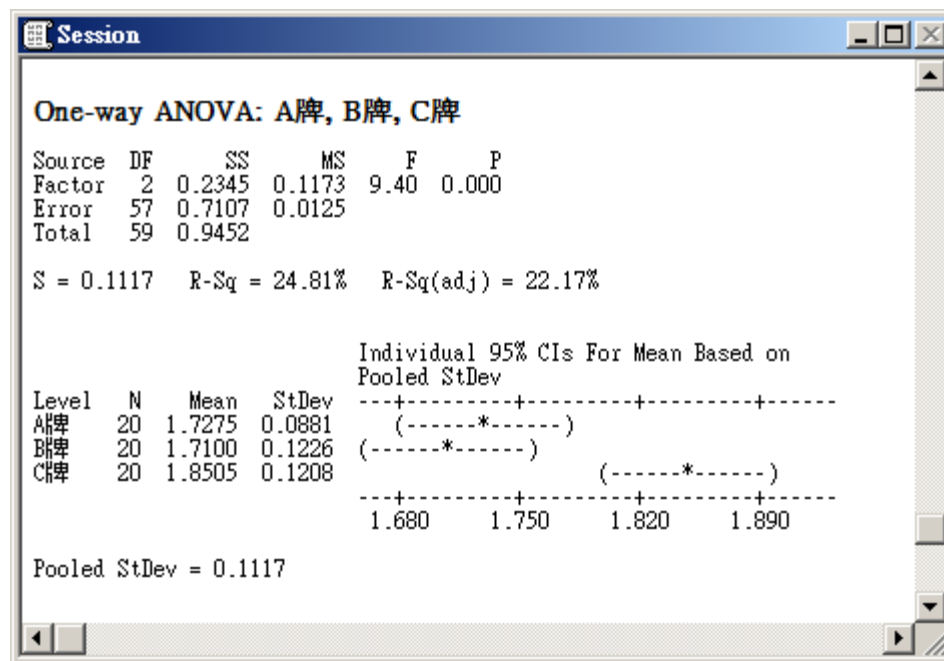
厂牌	产品寿命（千小时）																			
A牌	1.72	1.64	1.73	1.84	1.78	1.82	1.81	1.80	1.73	1.56	1.64	1.73	1.84	1.81	1.80	1.73	1.64	1.73	1.56	1.64
B牌	1.91	1.70	1.78	1.90	1.74	1.53	1.57	1.61	1.50	1.62	1.70	1.78	1.86	1.57	1.61	1.78	1.70	1.78	1.86	1.70
C牌	1.64	2.02	1.88	1.83	1.76	1.91	1.70	1.78	2.00	1.74	2.02	1.88	1.83	1.70	1.78	1.88	2.02	1.88	1.74	2.02

试检定这三种不同厂牌灯泡的平均寿命间是否有显著差异。

变异数分析范例（三）

- 例5：利用 Minitab 电脑报表求解：

【Minitab 计算机报表】



- 由上表可知三种厂牌灯泡的平均寿命间有显著差异（ $p\text{值}=0.000 < 0.05$ ）。
- 进一步检三个厂牌之之信赖区间，可知 **C厂牌的灯泡平均寿命显著的高于A、B两种厂牌**，而A、B两种厂牌的灯泡平均寿命间并无显著差异。

本单元结束

第七单元 简单回顾

简单回顾

- 变异数分析
- 变异数分析之统计假设
- 变异数分析之执行步骤：
 - F分布
- 变异数分析中各变异量之计算：
 - 总变异
 - 组间变异
 - 组内变异
 - 变异数分析表