双原因方差分析措施

● 双原因试验的方差分析

在实际应用中，一种试验成果（试验指标）往往受多种原因的影响。不但这些原因会影响试验成果，而且这些原因的不同水平的搭配也会影响试验成果。

例如：某些合金，当单独加入元素A或元素B时，性能变化不大，但当同步加入元素A和B时，合金性能的变化就尤其明显。

统计学上把多原因不同水平搭配对试验指标的影响称为交互作用。交互作用在多原因的方差分析中，把它当成一种新原因来处理。

我们只学习两个原因的方差分析，更多原因的问题，用正交试验法比较以便。

无交互作用的双原因试验的方差分析

数学模型

假设某个试验中，有两个可控原因在变化，原因A有a个水平，记作A_1，A_2，…，A_a；原因B有b个水平，记作B_1，B_2，….B_b；则A与B的不同水平组合A_iB_j（i=1，2，…，a；j=1，2，…，b）共有ab个，每个水平组合称为一种处理，每个处理只作一次试验，得ab个观察值X_{ij}，得双原因无反复试验表

双原因无反复（无交互作用）试验资料表

原因B / 原因A	B_1	B_2	$\cdots$	B_b	$T_{i\cdot}=\sum_{j=1}^{b}X_{ij}$	$\overline{X}_{i\cdot}=T_{i\cdot}/b$
A_1	X_{11}	X_{12}	$\cdots$	X_{1b}	$T_{1\cdot}$	$\overline{X}_{1\cdot}$
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
A_a	X_{a1}	X_{a2}	$\cdots$	X_{ab}	$T_{a\cdot}$	$\overline{X}_{a\cdot}$
$T_{\cdot j}=\sum_{i=1}^{a}X_{ij}$	$T_{\cdot 1}$	$T_{\cdot 2}$	$\cdots$	$T_{\cdot b}$	$T=\sum_{i=1}^{a}\sum_{j=1}^{b}X_{ij}$	
$\overline{X}_{\cdot j}=T_{\cdot j}/a$	$\overline{X}_{\cdot 1}$	$\overline{X}_{\cdot 2}$	$\cdots$	$\overline{X}_{\cdot b}$		$\overline{X}=\frac{1}{ab}T$

➢ 无交互作用的双原因试验的方差分析

基本假设（1）X_{ij} 相互独立；

（2）$X_{ij} \sim N\left(\mu_{ij}, \sigma^2\right)$，（方差齐性）。

线性统计模型 $X_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij}$

其中 $\mu = \dfrac{1}{ab}\sum_{i=1}^{a}\sum_{j=1}^{b}\mu_{ij}$ 全部期望值的总平均

$\alpha_i = \dfrac{1}{a}\sum_{j=1}^{b}\mu_{ij} - \mu = \mu_{ig} - \mu$ 水平A_i对试验成果的效应

$\beta_j = \dfrac{1}{b}\sum_{i=1}^{a}\mu_{ij} - \mu = \mu_{gj} - \mu$ 水平B_j对试验成果的效应

$\varepsilon_{ij} = X_{ij} - \mu_{ij}$ 试验误差

$$\alpha_i = \frac{1}{a}\sum_{j=1}^{b}\mu_{ij} - \mu = \mu_{i\mathrm{g}} - \mu$$ 水平$\mathbf{A_i}$对试验成果的效应

$$\beta_j = \frac{1}{b}\sum_{i=1}^{a}\mu_{ij} - \mu = \mu_{\mathrm{g}j} - \mu$$ 水平$\mathbf{B_j}$对试验成果的效应

$$\varepsilon_{ij} = X_{ij} - \mu_{ij}$$ 试验误差

特征：$\sum_{i=1}^{a}\alpha_i = 0;\ \ \sum_{j=1}^{b}\beta_j = 0;\ \ \varepsilon_{ij} \sim N\left(0, \sigma^2\right)$

要分析原因A，B的差别对试验成果是否有明显影响，即为检验如下假设是否成立：

$$H_{01} : \alpha_1 = \alpha_2 = \alpha_3 = 0$$

$$H_{02} : \beta_1 = \beta_2 = \mathrm{L} = \beta_b = 0$$

➢ 总离差平方和的分解定理

仿单原因方差分析的措施，考察总离差平方和

$$SS_T=\sum_{i=1}^{a}\sum_{j=1}^{b}\left(X_{ij}-\overline{X}\right)^2$$

可分解为：$SS_T=SS_A+SS_B+SS_E$

$$SS_A=b\sum_{i=1}^{a}\left(\overline{X}_{i.}-\overline{X}\right)^2$$

称为原因A的离差平方和，反应原因 A 对试验指标的影响。

$$SS_B=a\sum_{j=1}^{b}\left(\overline{X}_{.j}-\overline{X}\right)^2$$

称为原因B的离差平方和，反应原因 B 对试验指标的影响。

$$SS_E=\sum_{i=1}^{a}\sum_{j=1}^{b}\left(X_{ij}-\overline{X}_{i.}-\overline{X}_{.j}+\overline{X}\right)^2$$

称为误差平方和，反应试验误差对试验指标的影响。

若假设 H_{01}, H_{02} 成立，则：$X_{ij} \sim N(\mu, \sigma^2)$

可推得：$\frac{SS_T}{\sigma^2} \sim \chi^2(ab-1)$　　$\frac{SS_A}{\sigma^2} \sim \chi^2(a-1)$

$\frac{SS_B}{\sigma^2} \sim \chi^2(b-1)$　　$\frac{SS_E}{\sigma^2} \sim \chi^2((a-1)(b-1))$

将 $\frac{SS_T}{\sigma^2}, \frac{SS_A}{\sigma^2}, \frac{SS_B}{\sigma^2}, \frac{SS_E}{\sigma^2}$ 的自由度分别记作

df_T, df_A, df_B, df_E ，则

$$F_A = \frac{SS_A / df_A}{SS_E / df_E} = \frac{MS_A}{MS_E} \sim F((a-1), (a-1)(b-1))$$

$$F_B = \frac{SS_B / df_B}{SS_E / df_E} = \frac{MS_B}{MS_E} \sim F((b-1), (a-1)(b-1))$$

$$F_A = \frac{SS_A / df_A}{SS_E / df_E} = \frac{MS_A}{MS_E} \sim F\left(\left(a-1\right),\left(a-1\right)\left(b-1\right)\right)$$

$$F_B = \frac{SS_B / df_B}{SS_E / df_E} = \frac{MS_B}{MS_E} \sim F\left(\left(b-1\right),\left(a-1\right)\left(b-1\right)\right)$$

对给定的检验水平 α ，

当 $F_A > F_\alpha\left(\left(a-1\right),\left(a-1\right)\left(b-1\right)\right)$ 时，

拒绝H_{01}，即A 原因的影响有统计意义。

当 $F_B > F_\alpha\left(\left(b-1\right),\left(a-1\right)\left(b-1\right)\right)$ 时，

拒绝H_{02}，即B 原因的影响有统计意义。

F 右侧检验

双原因（无交互作用）试验的方差分析表

方差起源	平方和	自由度	均方和	F 值	F 值临介值
原因A	SS_A	df_A	$MS_A=\dfrac{SS_A}{df_A}$	$F_A=\dfrac{MS_A}{MS_E}$	$F_\alpha((a-1),(a-1)(b-1))$
原因B	SS_B	df_B	$MS_B=\dfrac{SS_B}{df_B}$	$F_B=\dfrac{MS_B}{MS_E}$	$F_\alpha((b-1),(a-1)(b-1))$
误差	SS_E	df_E	$MS_E=\dfrac{SS_E}{df_E}$		
总和	SS_T	df_T			

注意 $df_E=df_T-df_A-f_B$， $SS_E=SS_T-SS_A-SS_B$

各原因离差平方和的自由度为水平数减一，总平方和的自由度为试验总次数减一。

双原因（无交互作用）试验的方差分析表

简便计算式：

$$SS_A = D_A - p,\ SS_B = D_B - p$$

$$SS_E = R - D_A - D_B + p,\ SS_T = R - p$$

其中：$D_A = \left(\sum_{i=1}^{a} T_{i.}^2 \right) \Big/ b,\qquad p = T^2 / ab,$

$$D_B = \left(\sum_{j=1}^{b} T_{.j}^2 \right) \Big/ a,\qquad R = \sum_{i=1}^{a} \sum_{j=1}^{b} X_{ij}^2$$

例1 设甲、乙、丙、丁四个工人操作机器Ⅰ、Ⅱ、Ⅲ各一天，其产品产量如下表，问工人和机器对产品产量是否有明显影响？

机器 B / 工人 A	Ⅰ	Ⅱ	Ⅲ	$T_{i\cdot}=\sum_{j=1}^{b}X_{ij}$	$\overline{X}_{i\cdot}=T_{i\cdot}/b$
甲	50	63	52	165	55.0
乙	47	54	42	143	47.7
丙	47	57	41	145	48.3
丁	53	58	48	159	53.0
$T_{\cdot j}=\sum_{i=1}^{a}X_{ij}$	197	232	183	$T=612$	
$\overline{X}_{\cdot j}=T_{\cdot j}/a$	49.3	58.0	45.8		$\overline{X}=51$

解 基本计算如原表

$$R=\sum_{i=1}^{a}\sum_{j=1}^{b}X_{ij}^{2}=31678 \qquad D_A=\frac{1}{b}\sum_{i=1}^{a}T_{i.}^{2}=23495$$

$$D_B=\frac{1}{a}\sum_{j=1}^{b}T_{.j}^{2}=42040.67 \qquad p=\frac{T^2}{ab}=31212$$

$$SS_T=R-p=466 \qquad df_T=n-1=11$$

$$SS_A=D_A-p=114.67 \qquad df_A=a-1=3$$

$$SS_B=D_B-p=318.5 \qquad df_B=b-1=2$$

$$SS_E=SS_T-SS_A-SS_B=32.83 \qquad df_E=df_A\cdot df_b=6$$

$$MS_A = SS_A / df_A = 38.223$$

$$MS_B = SS_B / df_B = 159.25$$

$$MS_E = SS_E / df_E = 5.47$$

$$F_A = MS_A / MS_E = 6.98$$

$$F_B = MS_B / MS_E = 29.10$$

$$F_{0.01}(3,6) = 9.78$$

$$F_{0.05}(3,6) = 4.76$$

$$F_{0.01}(2,6) = 10.92$$

$$F_{0.05}(3,6) < F_A < F_{0.01}(3,6) \qquad F_B > F_{0.01}(2,6)$$

结论：工人对产品的产量有明显影响，

机器对产品的产量有极明显影响。

例2：某厂对生产的高速钢铣刀进行淬火工艺试验，考察回火温度A和淬火温度B两个原因对强度的影响。今对两个原因各3个水平进行试验，得平均硬度见表：

试验成果 Bj / Ai	B1（1210‘C）	B2（1235’C）	B3（1250‘C）
A1（280’C）	64	66	68
A2（300‘C）	66	68	67
A3（320’C）	65	67	68

假设：不同组合水平下服从正态分布、相互独立、方差相等。

所需要处理的问题是：全部Xij的均值是否相等。

方差分析表：

方差起源	离差平方和	自由度	F值	F0.05(2,4)	F0.01(2,4)	明显性
原因A	1.56	2	FA=1.01	6.94	18.0	
原因B	11.56	2	FB=7.46	6.94	18.0	*
试验误差	3.1	4				
总误差	16.22	8				

$$SST=\sum_{i=1}^{3}\sum_{j=1}^{3}X_{ij}^{2}-\frac{T^{2}}{3\times 3}=16.22 \qquad SSA=\frac{1}{3}\sum_{i=1}^{3}T_{i.}^{2}-\frac{T^{2}}{3\times 3}=1.56$$

$$SSB=\frac{1}{3}\sum_{j=1}^{3}T_{.j}^{2}-\frac{T^{2}}{3\times 3}=11.56 \qquad SSE=SST-SSA-SSB=3.1$$

$F_A<F_{0.05}(2,4)$ **A影响不明显。** $F_{0.05}(2,4)<F_B<F_{0.01}(2,4)$ **B影响明显，因为**

高速钢洗刀的硬度越大越好，所以因素B可取B3水平，即淬火温度1250‘C为好，因素A水平的拟定，应考虑经济方便，取A1水平为好。

【例3】有四个品牌的彩电在五个地域销售，为分析彩电的品牌(原因A)和销售地域(原因B)对销售量是否有影响，对每个品牌在各地域的销售量取得下列数据，见下表。试分析品牌和销售地域对彩电的销售量是否有明显影响？

- 不同品牌的彩电在各地域的销售量数据

品牌(原因A)	销售地域(原因B)				
	B_1	B_2	B_3	B_4	B_5
A_1	365	350	343	340	323
A_2	345	368	363	330	333
A_3	358	323	353	343	308
A_4	288	280	298	260	298

四、双原因方差分析例题

1、对原因*A*提出的假设为

- H_0: $\mu_1=\mu_2=\mu_3=\mu_4$
（品牌对销售量没有影响）
- H_1: μ_i (i =1,2, … , 4) 不全相等
（品牌对销售量有影响）

2、对原因*B*提出的假设为

- H_0: $\mu_1=\mu_2=\mu_3=\mu_4=\mu_5$
（地域对销售量没有影响）
- H_1: μ_j (j =1,2,…,5) 不全相等
（地域对销售量有影响）