2.1 方差分析法

王正明　易泰河

系统工程学院　军事建模与仿真系

2019 年 11 月 15 日

引言

例

为了提高某化工产品的转化率，考虑 3 个三水平因子：

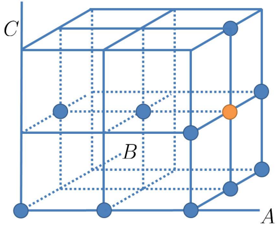
- 反应温度 A，$A_1 = 80℃$，$A_2 = 85℃$，$A_3 = 90℃$；
- 反应时间 B，$B_1 = 90\mathtt{min}$，$B_2 = 120\mathtt{min}$，$B_3 = 150\mathtt{min}$；
- 用碱量 C，$C_1 = 5\%$，$C_2 = 6\%$，$C_3 = 7\%$.

如何安排这一试验？

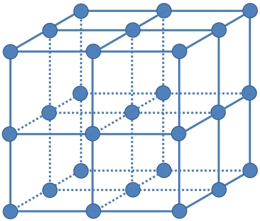
- 因子设计或析因设计：同时考虑多个试验因子，每个因子设定为若干个水平.
- 试验目的：处理比较、因子筛选和系统优化.

引言

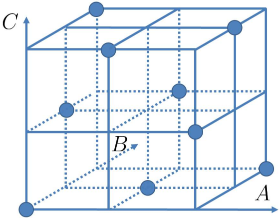
(1) 单因子轮换



(2) 全面实施



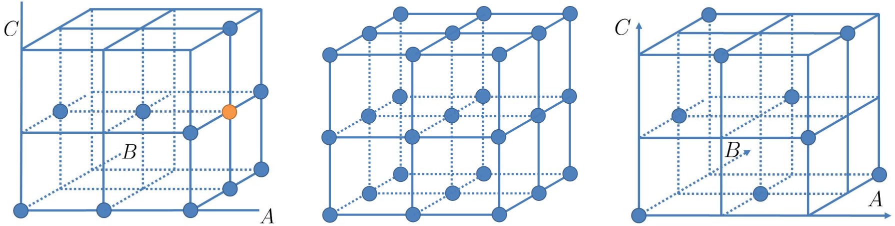
(2) 部分实施



- 部分实施：挑选一部分有代表性的处理组成试验方案；
- 如何挑选有代表性的点？如何分析数据？

引言

(1) 单因子轮换　　(2) 全面实施　　(2) 部分实施



- 部分实施：挑选一部分有代表性的处理组成试验方案；
- 如何挑选有代表性的点？如何分析数据？

引言

例 (手枪工艺)

制造某新型手枪共有 A_1, A_2, A_3, A_4 四种不同工艺, 为研究它们之间的差异, 命 a, b, c, d, e 五个战士打靶, 每人提供 400 发子弹. 如何安排试验? 命中频率数据如下, 四种工艺是否有差异?

	A_1	A_2	A_3	A_4
a	0.60	0.59	0.71	0.72
b	0.80	0.81	0.88	0.86
c	0.68	0.64	0.80	0.79
d	0.68	0.70	0.81	0.82
e	0.59	0.60	0.73	0.72
和	3.36	3.34	3.93	3.91
平均	0.672	0.668	0.786	0.782

引言

例 (手枪工艺)

制造某新型手枪共有 A_1, A_2, A_3, A_4 四种不同工艺, 为研究它们之间的差异, 命 a, b, c, d, e 五个战士打靶, 每人提供 400 发子弹. 如何安排试验? 命中频率数据如下, 四种工艺是否有差异?

	A_1	A_2	A_3	A_4
a	0.60	0.59	0.71	0.72
b	0.80	0.81	0.88	0.86
c	0.68	0.64	0.80	0.79
d	0.68	0.70	0.81	0.82
e	0.59	0.60	0.73	0.72
和	3.36	3.34	3.93	3.91
平均	0.672	0.668	0.786	0.782

第二章 因子设计

2.1 方差分析法

2.1.1 单因子试验的方差分析

2.1.2 双因子试验的方差分析

2.1.3 三因子试验的方差分析

2.1.4 多重比较与对照

2.2 2^k 因子设计及其部分实施

2.3 3^k 因子设计及其部分实施

2.4 正交设计的一般讨论

本节教学目的

1. 理解方差分析的基本思想;
2. 学会单因子、二因子试验的方差分析;
3. 能够利用 R 进行方差分析.

2.1 方差分析法

2.1.1 单因子试验的方差分析

(1) 固定效应模型

(2) 方差分析

(3) 参数估计

2.1.2 双因子试验与交互效应

(1) 固定效应模型及其参数估计

(2) 交互效应

(3) 方差分析

水平	观察值			
A_1	y_{11}	y_{12}	$\cdots$	y_{1n_1}
A_2	y_{21}	y_{22}	$\cdots$	y_{2n_2}
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
A_a	y_{a1}	y_{a2}	$\cdots$	y_{an_a}

- 问题: A 的变动是否会带来响应的波动?
- 假设: 不同处理下响应值来自于方差相同的正态总体, 且每次试验结果都互相独立.
- 思路: 若不同处理下正态总体均值相等, 则响应的波动完全由噪声因子引起, 否则响应的波动还包含试验因子的效应.
- 总波动统计量 $=$ 不同因子效应统计量 $+$ 误差效应统计量.

2.1 方差分析法

2.1.1 单因子试验的方差分析

(1) 固定效应模型

(2) 方差分析

(3) 参数估计

2.1.2 双因子试验与交互效应

(1) 固定效应模型及其参数估计

(2) 交互效应

(3) 方差分析

- 将 y_{ij} 分解为处理 A_i 的影响部分 μ_i 和误差 ε_{ij}

$$\begin{cases} y_{ij} = \mu_i + \varepsilon_{ij}, & \varepsilon_{ij} \overset{\text{i.i.d.}}{\sim} N(0, \sigma^2), \\ i = 1, \cdots, a, & j = 1, \cdots, n_i. \end{cases}$$

- 一般平均 $\mu := \frac{1}{n}\sum_{i=1}^{a} n_i\mu_i$; A_i 的效应 $\tau_i := \mu_i - \mu$, 得到单因子试验的固定效应模型:

$$\begin{cases} y_{ij} = \mu + \tau_i + \varepsilon_{ij}, & \varepsilon_{ij} \overset{\text{i.i.d.}}{\sim} N(0, \sigma^2), \\ \sum_{i=1}^{a} n_i\tau_i = 0, \ i = 1, \cdots, a, \ j = 1, \cdots, n_i. \end{cases}$$

- 将 y_{ij} 分解为处理 A_i 的影响部分 μ_i 和误差 ε_{ij}

$$\begin{cases} y_{ij} = \mu_i + \varepsilon_{ij}, & \varepsilon_{ij} \overset{\text{i.i.d.}}{\sim} N(0, \sigma^2), \\ i = 1, \cdots, a, & j = 1, \cdots, n_i. \end{cases}$$

- 一般平均$\mu := \frac{1}{n}\sum_{i=1}^{a} n_i \mu_i$; A_i 的效应$\tau_i := \mu_i - \mu$, 得到单因子试验的固定效应模型:

$$\begin{cases} y_{ij} = \mu + \tau_i + \varepsilon_{ij}, \quad \varepsilon_{ij} \overset{\text{i.i.d.}}{\sim} N(0, \sigma^2), \\ \sum_{i=1}^{a} n_i \tau_i = 0, \ i = 1, \cdots, a, \ j = 1, \cdots, n_i. \end{cases}$$

2.1 方差分析法

2.1.1 单因子试验的方差分析

(1) 固定效应模型

(2) 方差分析

(3) 参数估计

2.1.2 双因子试验与交互效应

(1) 固定效应模型及其参数估计

(2) 交互效应

(3) 方差分析

方差分析:

$$H_0 : (\tau_1, \tau_2, \cdots, \tau_a) = \mathbf{0} \text{ vs } H_1 : (\tau_1, \tau_2, \cdots, \tau_a) \neq \mathbf{0}.$$

1. 什么是 p 值?
2. 什么是统计显著性 (statistical significance)?
3. 两类错误分别是什么?

- 记 $y_{\cdot\cdot} := \sum_{i=1}^{a}\sum_{j=1}^{n_i} y_{ij}$, $\bar{y}_{\cdot\cdot} := \frac{y_{\cdot\cdot}}{n}$, $y_{i\cdot} := \sum_{j=1}^{n_i} y_{ij}$, $\bar{y}_{i\cdot} := \frac{y_{i\cdot}}{n_i}$.
- 响应的波动用总偏差平方和表示:

$$SS_T := \sum_{i=1}^{a}\sum_{j=1}^{n_i}(y_{ij}-\bar{y}_{\cdot\cdot})^2 = \sum_{i=1}^{a}\sum_{j=1}^{n_i} y_{ij}^2 - \frac{y_{\cdot\cdot}^2}{n}$$
$$= \sum_{i=1}^{a} n_i(\bar{y}_{i\cdot}-\bar{y}_{\cdot\cdot})^2 + \sum_{i=1}^{a}\sum_{j=1}^{n_i}(y_{ij}-\bar{y}_{i\cdot})^2$$

- 因子平方和 $SS_A := \sum_{i=1}^{a} n_i(\bar{y}_{i\cdot}-\bar{y}_{\cdot\cdot})^2 = \sum_{i=1}^{a}\frac{y_{i\cdot}^2}{n_i} - \frac{y_{\cdot\cdot}^2}{n}$;
- 误差平方和 $SS_E := \sum_{i=1}^{a}\sum_{j=1}^{n_i}(y_{ij}-\bar{y}_{i\cdot})^2 = SS_T - SS_A$.

- 记 $y_{\cdot\cdot} := \sum_{i=1}^{a}\sum_{j=1}^{n_i} y_{ij}$, $\bar{y}_{\cdot\cdot} := \frac{y_{\cdot\cdot}}{n}$, $y_{i\cdot} := \sum_{j=1}^{n_i} y_{ij}$, $\bar{y}_{i\cdot} := \frac{y_{i\cdot}}{n_i}$.
- 响应的波动用总偏差平方和表示:

$$SS_T := \sum_{i=1}^{a}\sum_{j=1}^{n_i}(y_{ij}-\bar{y}_{\cdot\cdot})^2 = \sum_{i=1}^{a}\sum_{j=1}^{n_i} y_{ij}^2 - \frac{y_{\cdot\cdot}^2}{n}$$

$$= \sum_{i=1}^{a} n_i(\bar{y}_{i\cdot}-\bar{y}_{\cdot\cdot})^2 + \sum_{i=1}^{a}\sum_{j=1}^{n_i}(y_{ij}-\bar{y}_{i\cdot})^2$$

- 因子平方和 $SS_A := \sum_{i=1}^{a} n_i(\bar{y}_{i\cdot}-\bar{y}_{\cdot\cdot})^2 = \sum_{i=1}^{a}\frac{y_{i\cdot}^2}{n_i} - \frac{y_{\cdot\cdot}^2}{n}$;
- 误差平方和 $SS_E := \sum_{i=1}^{a}\sum_{j=1}^{n_i}(y_{ij}-\bar{y}_{i\cdot})^2 = SS_T - SS_A$.

- 记 $y_{\cdot\cdot} := \sum_{i=1}^{a}\sum_{j=1}^{n_i} y_{ij}$, $\bar{y}_{\cdot\cdot} := \frac{y_{\cdot\cdot}}{n}$, $y_{i\cdot} := \sum_{j=1}^{n_i} y_{ij}$, $\bar{y}_{i\cdot} := \frac{y_{i\cdot}}{n_i}$.
- 响应的波动用总偏差平方和表示:

$$SS_T := \sum_{i=1}^{a}\sum_{j=1}^{n_i}(y_{ij}-\bar{y}_{\cdot\cdot})^2 = \sum_{i=1}^{a}\sum_{j=1}^{n_i} y_{ij}^2 - \frac{y_{\cdot\cdot}^2}{n}$$
$$= \sum_{i=1}^{a} n_i(\bar{y}_{i\cdot}-\bar{y}_{\cdot\cdot})^2 + \sum_{i=1}^{a}\sum_{j=1}^{n_i}(y_{ij}-\bar{y}_{i\cdot})^2$$

- 因子平方和 $SS_A := \sum_{i=1}^{a} n_i(\bar{y}_{i\cdot}-\bar{y}_{\cdot\cdot})^2 = \sum_{i=1}^{a}\frac{y_{i\cdot}^2}{n_i} - \frac{y_{\cdot\cdot}^2}{n}$;
- 误差平方和 $SS_E := \sum_{i=1}^{a}\sum_{j=1}^{n_i}(y_{ij}-\bar{y}_{i\cdot})^2 = SS_T - SS_A$.

- $\mathbb{E}(SS_A) = \sum_{i=1}^{a} n_i\tau_i^2 + (a-1)\sigma^2$, 当 H_0 成立时, 因子 A 的均方和$MS_A := SS_A/(a-1)$ 是 σ^2 的无偏估计.
- $\mathbb{E}(SS_E) = (n-a)\sigma^2$, 误差均方和$MS_E := SS_E/(n-a)$ 是 σ^2 的无偏估计.
- 当检验统计量

$$F := \frac{MS_A}{MS_E} = \frac{SS_A/(a-1)}{SS_E/(n-a)}.$$

大到一定程度时, 拒绝 H_0.

定理

如果原假设 $H_0 : \tau_1 = \cdots = \tau_a = 0$ 成立, 则

$$F := \frac{MS_A}{MS_E} = \frac{SS_A/(a-1)}{SS_E/(n-a)} \sim F(a-1, n-a).$$

来源	平方和	自由度	均方和	F 值
因子 A	$SS_A = \sum_{i=1}^a \frac{y_{i.}^2}{n_i} - \frac{y_{..}^2}{n}$	$f_A = a-1$	$MS_A = \frac{SS_A}{f_A}$	$F = \frac{MS_A}{MS_E}$
误差	$SS_E = SS_T - SS_A$	$f_E = n-a$	$MS_E = \frac{SS_E}{f_E}$	
总	$SS_T = \sum_{i=1}^a \sum_{j=1}^{n_i} y_{ij}^2 - \frac{y_{..}^2}{n}$	$f_T = n-1$		

- 自由度是诸平方和所对应的 χ^2 分布的自由度, 满足 $f_T = f_A + f_E$.

例 (cont. I)

制造某新型手枪共有 A_1, A_2, A_3, A_4 四种不同工艺, 为研究它们之间的差异, 命 a, b, c, d, e 五个战士打靶, 命中频率数据如下:

	A_1	A_2	A_3	A_4
a	0.60	0.59	0.71	0.72
b	0.80	0.81	0.88	0.86
c	0.68	0.64	0.80	0.79
d	0.68	0.70	0.81	0.82
e	0.59	0.60	0.73	0.72
和	3.36	3.34	3.93	3.91
平均	0.672	0.668	0.786	0.782

试对本例进行方差分析.

(1) 手动计算:

$$SS_T = \sum_{i=1}^{a}\sum_{j=1}^{n_i} y_{ij}^2 - \frac{y_{\cdot\cdot}^2}{n} = 0.16106,$$

$$SS_A = \sum_{i=1}^{a} \frac{y_{i\cdot}^2}{n_i} - \frac{y_{\cdot\cdot}^2}{n} = 0.06618,$$

$$F = \frac{SS_A/(4-1)}{(SS_T - SS_A)/(20-4)} = 3.72 > F_{0.05}(3,16) = 2.67.$$

(2) R 添加包 stats 提供了方差分析函数 aov():

```
aov(formula, data = NULL, projections = FALSE,
    contrasts = NULL, ...)
```

(1) 手动计算:

$$SS_T = \sum_{i=1}^{a}\sum_{j=1}^{n_i} y_{ij}^2 - \frac{y_{..}^2}{n} = 0.16106,$$

$$SS_A = \sum_{i=1}^{a} \frac{y_{i.}^2}{n_i} - \frac{y_{..}^2}{n} = 0.06618,$$

$$F = \frac{SS_A/(4-1)}{(SS_T - SS_A)/(20-4)} = 3.72 > F_{0.05}(3,16) = 2.67.$$

(2) R 添加包 stats 提供了方差分析函数 aov():

```
aov(formula, data = NULL, projections = FALSE,
   contrasts = NULL, ...)
```

2.1 方差分析法

2.1.1 单因子试验的方差分析

(1) 固定效应模型

(2) 方差分析

(3) 参数估计

2.1.2 双因子试验与交互效应

(1) 固定效应模型及其参数估计

(2) 交互效应

(3) 方差分析

- 未知参数 $(\mu, \tau_1, \cdots, \tau_a, \sigma^2)$, 共$a+1$个.
- 当诸 ε_{ij} 互相独立, 且满足 $\mathbb{E}(\varepsilon_{ij}) = 0$ 时,

$$\widehat{\sigma^2} = \frac{SS_E}{n-a}$$

是 σ^2 的无偏估计.

- 未知参数 $(\mu, \tau_1, \cdots, \tau_a, \sigma^2)$, 共$a+1$个.
- 当诸 ε_{ij} 互相独立, 且满足 $\mathbb{E}(\varepsilon_{ij}) = 0$ 时,

$$\widehat{\sigma^2} = \frac{SS_E}{n-a}$$

是 σ^2 的无偏估计.

(1) 最小二乘估计:

- 使

$$Q(\mu, \tau_1, \cdots, \tau_a) := \sum_{i=1}^{a} \sum_{j=1}^{n_i} (y_{ij} - \mu - \tau_i)^2$$

  达到最小的 $\hat{\mu}$ 与 $\hat{\tau}_i$.

- 对参数求导, 并令导数为 0, 得到

$$\begin{cases} \widehat{\mu} = \bar{y}_{\cdot\cdot} \\ \widehat{\tau}_1 = \bar{y}_{1\cdot} - \bar{y}_{\cdot\cdot} \\ \qquad \vdots \\ \widehat{\tau}_a = \bar{y}_{a\cdot} - \bar{y}_{\cdot\cdot} \end{cases}$$

(2) 极大似然估计:

- 当假设 $\varepsilon_{ij} \overset{\text{i.i.d.}}{\sim} N(0,\sigma^2)$ 成立时，诸参数的似然函数为

$$L(\mu_1,\cdots,\mu_a,\sigma^2)=\left(\sqrt{2\pi}\sigma\right)^{-n}\exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{a}\sum_{j=1}^{n_i}(y_{ij}-\mu_i)^2\right\}.$$

- 负对数似然函数:

$$\ell(\mu_1,\cdots,\mu_a,\sigma^2)=\frac{n}{2}\ln(2\pi)+\frac{n}{2}\ln(\sigma^2)+\frac{1}{2\sigma^2}\sum_{i=1}^{a}\sum_{j=1}^{n_i}(y_{ij}-\mu_i)^2.$$

- 对参数求导并令导数为 0，得到

$$\begin{cases}\widehat{\mu}_i=\bar{y}_{i\cdot}, & i=1,\cdots,a,\\ \widehat{\sigma}^2=\dfrac{1}{n}\displaystyle\sum_{i=1}^{a}\sum_{j=1}^{n_i}(y_{ij}-\bar{y}_{i\cdot})^2.\end{cases}$$

(3) 最小二乘估计与极大似然估计的区别:

- 最小二乘估计只需假定诸 ε_{ij} 独立同分布, 均值为 0, 方差有限. 而极大似然估计需要知道误差的具体分布.
- 设 $\hat{\boldsymbol{\theta}}$ 是参数 $\boldsymbol{\theta}$ 的极大似然估计, $f(\boldsymbol{\theta})$ 是 $\boldsymbol{\theta}$ 的连续函数, 则 $f(\hat{\boldsymbol{\theta}})$ 是 $f(\boldsymbol{\theta})$ 的极大似然估计. 根据这一性质, 可得到总体均值 μ 和效应 τ_i 的极大似然估计为:

$$\hat{\mu} = \bar{y}_{\cdot\cdot}, \qquad \hat{\tau}_i = \bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot}, i = 1, \cdots, a.$$

- 极大似然估计同时给出所有参数的估计, 最小二乘估计分别处理 τ_i 和 σ^2.

例 (cont. II)

制造某新型手枪共有 A_1, A_2, A_3, A_4 四种不同工艺, 为研究它们之间的差异, 命 a, b, c, d, e 五个战士打靶, 命中频率数据如下:

	A_1	A_2	A_3	A_4
a	0.60	0.59	0.71	0.72
b	0.80	0.81	0.88	0.86
c	0.68	0.64	0.80	0.79
d	0.68	0.70	0.81	0.82
e	0.59	0.60	0.73	0.72
$y_{i\cdot}$	3.36	3.34	3.93	3.91
$\bar{y}_{i\cdot}$	0.672	0.668	0.786	0.782

例 (cont. II)

固定效应模型:

$$\begin{cases} y_{ij} = \mu + \tau_i + \varepsilon_{ij},\ \varepsilon_{ij} \overset{\text{i.i.d.}}{\sim} N(0, \sigma^2), \\ i = 1, \cdots, 4, \quad j = 1, \cdots, 5, \\ \tau_1 + \tau_2 + \tau_3 + \tau_4 = 0. \end{cases}$$

诸参数的估计为

$$\begin{aligned} \widehat{\mu} &= \bar{y}_{\cdot\cdot} = 0.727, \\ \widehat{\tau}_1 &= \bar{y}_{1\cdot} - \bar{y}_{\cdot\cdot} = -0.055, \\ \widehat{\tau}_2 &= \bar{y}_{2\cdot} - \bar{y}_{\cdot\cdot} = -0.057, \\ \widehat{\tau}_3 &= \bar{y}_{3\cdot} - \bar{y}_{\cdot\cdot} = 0.059, \\ \widehat{\tau}_4 &= \bar{y}_{4\cdot} - \bar{y}_{\cdot\cdot} = 0.055. \\ SS_E &= 0.09488 \end{aligned}$$

2.1 方差分析法

2.1.1 单因子试验的方差分析

(1) 固定效应模型

(2) 方差分析

(3) 参数估计

2.1.2 双因子试验与交互效应

(1) 固定效应模型及其参数估计

(2) 交互效应

(3) 方差分析

二因子试验等重复情形:

$$\begin{cases} y_{ijl} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijl}, \quad \varepsilon_{ijl} \overset{\text{i.i.d.}}{\sim} N(0, \sigma^2), \\ i = 1, \cdots, a, \quad j = 1, \cdots, b, \quad l = 1, \cdots, m. \end{cases}$$

- α_i 是 A_i 的主效应, β_j 是 B_j 的主效应, $(\alpha\beta)_{ij}$ 表示处理 (A_i, B_j) 对 y 的交互效应或交互作用.
- 将 y_{ijl} 分解为:

$$\begin{aligned} y_{ijl} = \bar{y}_{\cdot\cdot\cdot} &+ (\bar{y}_{i\cdot\cdot} - \bar{y}_{\cdot\cdot\cdot}) + (\bar{y}_{\cdot j\cdot} - \bar{y}_{\cdot\cdot\cdot}) \\ &+ (\bar{y}_{ij\cdot} - \bar{y}_{i\cdot\cdot} - \bar{y}_{\cdot j\cdot} + \bar{y}_{\cdot\cdot\cdot}) + (y_{ikl} - \bar{y}_{ij\cdot}), \end{aligned}$$

- 猜测应有估计: $\hat{\mu} = \bar{y}_{\cdot\cdot\cdot}$, $\hat{\alpha}_i = \bar{y}_{i\cdot\cdot} - \bar{y}_{\cdot\cdot\cdot}$, $\hat{\beta}_j = \bar{y}_{\cdot j\cdot} - \bar{y}_{\cdot\cdot\cdot}$, $\widehat{(\alpha\beta)}_{ij} = \bar{y}_{ij\cdot} - \bar{y}_{i\cdot\cdot} - \bar{y}_{\cdot j\cdot} + \bar{y}_{\cdot\cdot\cdot}$, $\hat{\varepsilon}_{ijl} = y_{ikl} - \bar{y}_{ij\cdot}$.

二因子试验等重复情形:

$$\begin{cases} y_{ijl} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijl}, \quad \varepsilon_{ijl} \overset{\text{i.i.d.}}{\sim} N(0, \sigma^2), \\ i = 1, \cdots, a, \quad j = 1, \cdots, b, \quad l = 1, \cdots, m. \end{cases}$$

- α_i 是 A_i 的主效应, β_j 是 B_j 的主效应, $(\alpha\beta)_{ij}$ 表示处理 (A_i, B_j) 对 y 的交互效应或交互作用.
- 将 y_{ijl} 分解为:

$$\begin{aligned} y_{ijl} = \bar{y}_{\cdots} &+ (\bar{y}_{i\cdot\cdot} - \bar{y}_{\cdots}) + (\bar{y}_{\cdot j\cdot} - \bar{y}_{\cdots}) \\ &+ (\bar{y}_{ij\cdot} - \bar{y}_{i\cdot\cdot} - \bar{y}_{\cdot j\cdot} + \bar{y}_{\cdots}) + (y_{ikl} - \bar{y}_{ij\cdot}), \end{aligned}$$

- 猜测应有估计: $\hat{\mu} = \bar{y}_{\cdots}$, $\hat{\alpha}_i = \bar{y}_{i\cdot\cdot} - \bar{y}_{\cdots}$, $\hat{\beta}_j = \bar{y}_{\cdot j\cdot} - \bar{y}_{\cdots}$, $\widehat{(\alpha\beta)}_{ij} = \bar{y}_{ij\cdot} - \bar{y}_{i\cdot\cdot} - \bar{y}_{\cdot j\cdot} + \bar{y}_{\cdots}$, $\hat{\varepsilon}_{ijl} = y_{ikl} - \bar{y}_{ij\cdot}$.

- 双因子试验的固定效应模型

$$
\begin{cases}
y_{ijl} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijl}, \quad \varepsilon_{ijl} \overset{\text{i.i.d.}}{\sim} N(0, \sigma^2), \\
i = 1, \cdots, a, j = 1, \cdots, b, l = 1, \cdots, m. \\
\sum_{i=1}^{a} \alpha_i = \sum_{j=1}^{b} \beta_j = 0, \\
\sum_{i=1}^{a} (\alpha\beta)_{ij} = 0, \quad j = 1, \cdots, b, \\
\sum_{j=1}^{b} (\alpha\beta)_{ij} = 0, \quad i = 1, \cdots, a.
\end{cases}
$$

2.1 方差分析法

2.1.1 单因子试验的方差分析

(1) 固定效应模型

(2) 方差分析

(3) 参数估计

2.1.2 双因子试验与交互效应

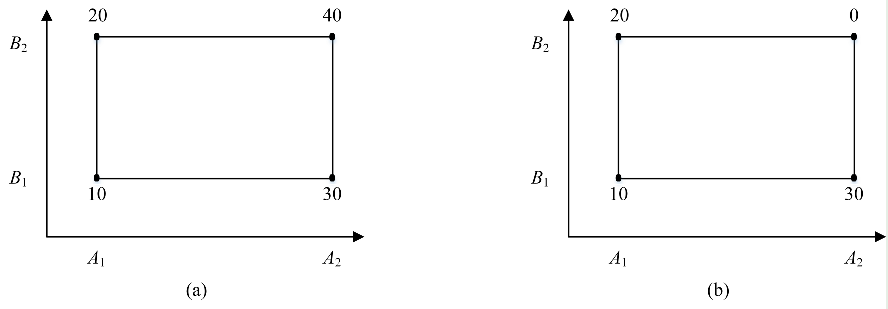
(1) 固定效应模型及其参数估计

(2) 交互效应

(3) 方差分析

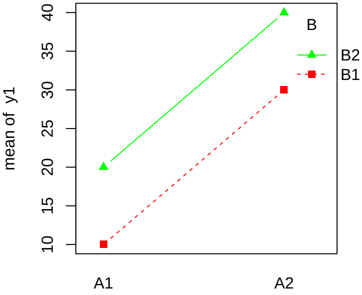
例

设有两个二水平因子 A 和 B, 它们的两个水平分别为 A_1, A_2 和 B_1, B_2. 每个组合试验一次, 试验结果如下

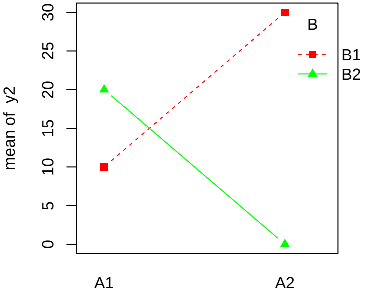


A 与 B 之间有无交互效应?

- 计算法：交互效应的估计值为 0 说明无交互效应；
- 画图法：利用 R 函数 `interaction.plot()` 得到交互效应图



(a)

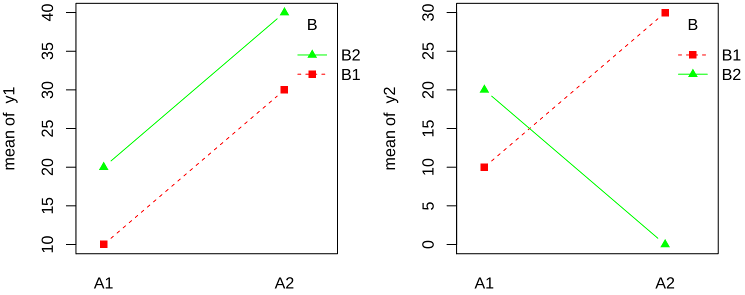


(b)

(a) 中两因子无交互效应　　(b) 中两因子有交互效应

- 两种方法都缺乏定量依据！

- 计算法：交互效应的估计值为 0 说明无交互效应；
- 画图法：利用 R 函数 `interaction.plot()` 得到交互效应图



(a) (b)

(a) 中两因子无交互效应　　(b) 中两因子有交互效应

- 两种方法都缺乏定量依据！

2.1 方差分析法

2.1.1 单因子试验的方差分析

(1) 固定效应模型

(2) 方差分析

(3) 参数估计

2.1.2 双因子试验与交互效应

(1) 固定效应模型及其参数估计

(2) 交互效应

(3) 方差分析

两因子方差分析: 检验主效应及交互效应是否显著.

- 响应的离差平方和 SS_T 分解:

$$SS_T = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{l=1}^{m}(y_{ijl} - \bar{y}_{\cdot\cdot\cdot})^2 = SS_A + SS_B + SS_{A\times B} + SS_E,$$

- $SS_A = mb\sum_{i=1}^{a}(\bar{y}_{i\cdot\cdot} - \bar{y}_{\cdot\cdot\cdot})^2$, $SS_B = ma\sum_{j=1}^{b}(\bar{y}_{\cdot j\cdot} - \bar{y}_{\cdot\cdot\cdot})^2$,
- $SS_{A\times B} = m\sum_{i=1}^{a}\sum_{j=1}^{b}(\bar{y}_{ij\cdot} - \bar{y}_{i\cdot\cdot} - \bar{y}_{\cdot j\cdot} + \bar{y}_{\cdot\cdot\cdot})^2$,
- $SS_E = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{l=1}^{m}(\bar{y}_{ijl} - \bar{y}_{ij\cdot})^2$.

两因子方差分析: 检验主效应及交互效应是否显著.

- 响应的离差平方和 SS_T 分解:

$$SS_T = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{l=1}^{m}(y_{ijl} - \bar{y}_{\cdots})^2 = SS_A + SS_B + SS_{A\times B} + SS_E,$$

- $SS_A = mb\sum_{i=1}^{a}(\bar{y}_{i\cdot\cdot} - \bar{y}_{\cdots})^2$, $SS_B = ma\sum_{j=1}^{b}(\bar{y}_{\cdot j\cdot} - \bar{y}_{\cdots})^2$,
- $SS_{A\times B} = m\sum_{i=1}^{a}\sum_{j=1}^{b}(\bar{y}_{ij\cdot} - \bar{y}_{i\cdot\cdot} - \bar{y}_{\cdot j\cdot} + \bar{y}_{\cdots})^2$,
- $SS_E = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{l=1}^{m}(\bar{y}_{ijl} - \bar{y}_{ij\cdot})^2$.

- SS_E 服从自由度为 $ab(m-1)$ 的 χ^2 分布.
- 当假设 $H_{A_0}: \alpha_1 = \cdots = \alpha_a = 0$ 成立时, $SS_A \sim \chi^2(a-1)$;
- 当假设 $H_{B_0}: \beta_1 = \cdots = \beta_b = 0$ 成立时, $SS_B \sim \chi^2(b-1)$;
- 当假设 $H_{(A\times B)_0}: (\alpha\beta)_{ij} = 0, i = 1, \cdots, a, j = 1, \cdots, b$ 成立时, $SS_{A\times B} \sim \chi^2\left((a-1)(b-1)\right)$;
- 诸平方和的均方为

$$MS_A = SS_A/(a-1), \quad MS_B = SS_B/(b-1),$$

$$MS_{A\times B} = SS_{A\times B}/[(a-1)(b-1)], \quad MS_E = SS_E/[ab(m-1)].$$

方差来源	自由度	平方和	均方	F 值
A	$a-1$	SS_A	MS_A	MS_A/MS_E
B	$b-1$	SS_B	MS_B	MS_B/MS_E
$A\times B$	$(a-1)(b-1)$	$SS_{A\times B}$	$MS_{A\times B}$	$MS_{A\times B}/MS_E$
误差	$ab(m-1)$	SS_E	MS_E	
总和	$n-1$	SS_T		

例

考察电池的最大输出电压受极板材料 A 与环境温度 B 的影响，两个因子均取 3 个水平，每个处理重复 4 次，试验数据如下:

		因子 B			$y_{i\cdot\cdot}$
		15℃	25℃	35℃	
因子 A	1	130 155 174 180 (639)	34 40 80 75 (229)	20 70 82 58 (230)	1098
	2	150 188 159 126 (623)	136 122 106 115 (479)	25 70 58 45 (198)	1300
	3	138 110 168 160 (576)	174 120 150 139 (583)	96 104 82 60 (342)	1501
$y_{\cdot j\cdot}$		1838	1291	770	$y_{\cdots} = 3899$

例 (cont.)

来源	平方和	自由度	均方和	F 值
A	6767.06	2	3383.53	6.73
B	4753509	2	23767.70	47.25
$A\times B$	13180.44	4	3295.11	6.55
误差	13580.75	27	520.99	
总和	81063.64	35		

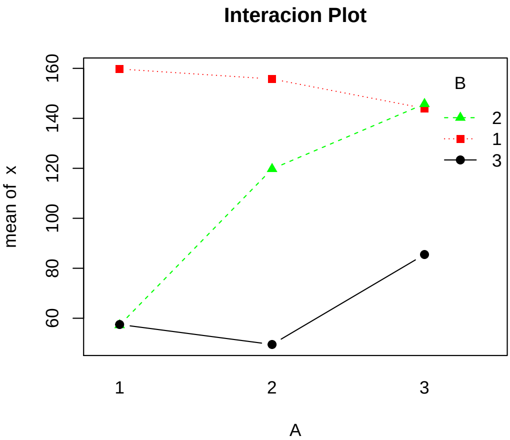
由于 $F_{0.95}(2,27)=3.35$, $F_{0.95}(4,27)=2.73$, 所以因子 A、因子 B 和交互效应 $A\times B$ 当 $\alpha=0.05$ 时都是显著的.

利用 R 进行两因子方差分析:

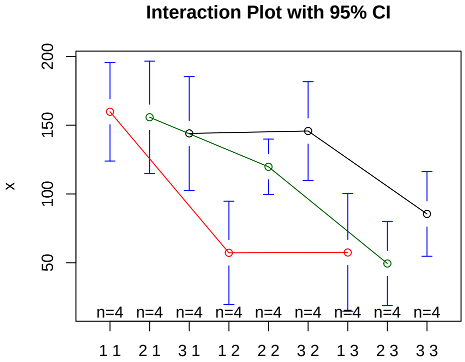
```
x <- c(130, 155, 174, 180, 150, 188, 159, 126, 138, 110, 168, 160, 34, 40, 80, 75, 136, 122, 106, 115, 174, 120, 150, 139, 20, 70, 82, 58, 25, 70, 58, 45, 96, 104, 82, 60);
A <- factor(c(rep(1:3,each = 4), rep(1:3,each=4), rep(1:3,each=4)));
B <- factor(rep(1:3, each = 12));
aov_Data <- data.frame(A, B, x);
aov_Result <- aov(x ~ A * B, data = aov_Data);
# aov_Result <- aov(x ~ A+B+A:B, data = aov_Data);
summary(aov_Result)
#>             Df Sum Sq Mean Sq F value Pr(>F)
#> A            2   6767    3384   6.727 0.004261 **
#> B            2  47535   23768  47.253 1.52e-09 ***
#> A:B          4  13180    3295   6.551 0.000807 ***
#> Residuals   27  13581     503
#> ---
#> Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

R 提供多种方式展示两因子之间的交互效应.

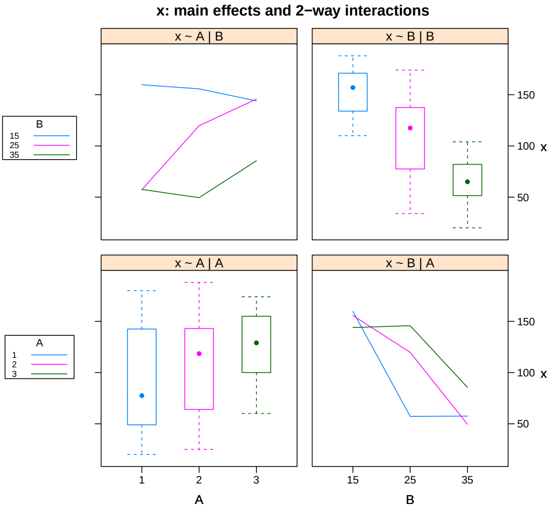
```
interaction.plot(A, B, x, type = "b", col = c("red"
    , "green", "black"), pch = c(15, 17, 19), main =
    "Interacion Plot")
```



```r
library(gplots);
plotmeans(x~interaction(A, B, sep = " "), connect = 
    list(c(1, 4, 7), c(2, 5, 8), c(3, 6, 9)), col = 
    c("red", "darkgreen", "black"), main = "
    Interaction Plot with 95% CI", xlab = "A and B 
    Combination")
```



```r
library(HH);
interaction2wt(x ~ A*B);
```



总结

- 多因子试验与因子设计
- 方差分析的基本思想
- 偏差平方和的分解
- 检验统计量的构造

习题

1. 利用最小二乘法获得双因子线性可加模型中参数的估计.
2. 证明双因子试验的离差平方和分解式.
3. 利用最小二乘法获得三因子线性可加模型中参数的估计.
4. 手枪工艺例中, 如果将五个战士当作一个五水平的因子, 则它是一个双因子试验, 能否对其进行方差分析检验是否存在交互效应, 并利用 R 绘制交互效应图.