

# 应用统计



---

## 第11讲 方差分析与实验设计

比较不同正态分布样本的均值差异



# 农药的有效性

---

试验号\杀虫率%\农药	A1	A2	A3	A4	A5	A6
1	87	90	56	55	92	75
2	85	88	62	48	99	72
3	80	87			95	81
4		94			91	



## 灯泡质量比较

灯丝\使用寿命\灯泡	1	2	3	4	5	6	7	8
甲	1600	1610	1650	1680	1700	1700	1780	
乙	1500	1640	1400	1700	1750			
丙	1640	1550	1600	1620	1640	1600	1740	1800
丁	1510	1520	1530	1570	1640	1680		



## 两个正态总体的假设检验

---

$x_1, \dots, x_m$  来自正态总体  $N(\mu_1, \sigma^2)$ ,  $y_1, \dots, y_n$  来自正态总体  $N(\mu_2, \sigma^2)$ , 考虑

检验问题  $H_0: \mu_1 = \mu_2$  VS  $H_1: \mu_1 \neq \mu_2$

$$\bar{x} - \bar{y} \sim N\left(\mu_1 - \mu_2, \left(\frac{1}{m} + \frac{1}{n}\right)\sigma^2\right), \quad \frac{(m-1)s_x^2 + (n-1)s_y^2}{\sigma^2} \sim \chi^2(m+n-2)$$

$$t = \sqrt{\frac{mn(m+n-2)}{m+n}} \frac{\bar{x} - \bar{y} - (\mu_1 - \mu_2)}{\sqrt{(m-1)s_x^2 + (n-1)s_y^2}} \sim t(m+n-2)。$$

方差不等的时候较难处理



# 方差分析

---

## Analysis of variance

From Wikipedia, the free encyclopedia

In statistics, **analysis of variance (ANOVA)** is a collection of statistical models, and their associated procedures, in which the observed variance is partitioned into components due to different explanatory variables. In its simplest form ANOVA gives a statistical test of whether the means of several groups are all equal, and therefore generalizes Student's two-sample  $t$ -test to more than two groups.

## History

Ronald Fisher first used variance in his 1918 paper *The Correlation Between Relatives on the Supposition of Mendelian Inheritance*<sup>[2]</sup>. His first application of the analysis of variance was published in 1921<sup>[3]</sup>. Analysis of variance became widely known after being included in Fisher's 1925 book *Statistical Methods for Research Workers*.



# 单因子方差分析 (one-way ANOVA)

## 因素（因子）与水平

试验号\杀虫率%\农药	A1	A2	A3	A4	A5	A6
1	87	90	56	55	92	75
2	85	88	62	48	99	72
3	80	87			95	81
4		94			91	



## 单因素试验结果

试验次数 水平	1	2	...	$n_i$
$A_1$	$X_{11}$	$X_{12}$	...	$X_{1n_1}$
$A_2$	$X_{21}$	$X_{22}$	...	$X_{2n_2}$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
$A_r$	$X_{r1}$	$X_{r2}$	...	$X_{rn_r}$



## 方差分析模型

---

因素  $A$  取  $r$  个水平:  $A_1, A_2, \dots, A_r$

$r$  个正态水平:  $X_1, X_2, \dots, X_r$ ,  $X_i \sim N(\mu_i, \sigma^2) (i = 1, 2, \dots, r)$

假设方差相等

水平  $A_i$  下进行  $n_i$  ( $n_i \geq 2$ ) 次独立试验:  $X_{i1}, \dots, X_{in_i}$

$$X_{ij} \sim N(\mu_i, \sigma^2) (j = 1, 2, \dots, n_i, i = 1, 2, \dots, r)$$

$$\varepsilon_{ij} = X_{ij} - \mu_i \sim N(0, \sigma^2) (j = 1, 2, \dots, n_i, i = 1, 2, \dots, r)$$

$$X_{ij} = \mu_i + \varepsilon_{ij}$$





# 方差分析理论模型

---

假设检验问题

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

VS

$H_1: \mu_1, \cdots, \mu_r$  中至少有两个不相等



# 方差分析模型

$$n = \sum_{i=1}^r n_i$$

理论总均值

$$\mu = \frac{1}{n} \sum_{i=1}^r n_i \mu_i$$

$\alpha_i = \mu_i - \mu$  ( $i = 1, 2, \dots, r$ ): 水平  $A_i$  对试验指标的效应

$$\left\{ \begin{array}{l} X_{ij} = \mu + \alpha_i + \varepsilon_{ij} \\ \sum_{i=1}^r n_i \alpha_i = 0 \\ \varepsilon_{ij} (j = 1, 2, \dots, n_i, i = 1, 2, \dots, r) \text{ 相互独立, } \varepsilon_{ij} \sim N(0, \sigma^2) \end{array} \right.$$

$\alpha_i$  为水平间的差异, 而  $\varepsilon_{ij}$  为水平内部的差异的随机变量

$$H_0: \alpha_1 = \alpha_2 = \dots = \alpha_r = 0$$



## 检验统计量

---

$$n = \sum_{i=1}^r n_i$$

样本总均值

$$\bar{X} = \frac{1}{n} \sum_{i=1}^r \sum_{j=1}^{n_i} X_{ij} = \frac{1}{n} \sum_{i=1}^r n_i \bar{X}_i$$

总偏差平方和

$$S_T = \sum_{i=1}^r \sum_{j=1}^{n_i} (X_{ij} - \bar{X})^2$$

可以把偏差理解为某种信息



# 方差分析

---

$$\begin{aligned} s_T &= \sum_{i=1}^r \sum_{j=1}^{n_i} (X_{ij} - \bar{X})^2 \\ &= \sum_{i=1}^r \sum_{j=1}^{n_i} \left[ (X_{ij} - \bar{X}_i) + (\bar{X}_i - \bar{X}) \right]^2 \\ &= \sum_{i=1}^r \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_i)^2 + \sum_{i=1}^r \sum_{j=1}^{n_i} (\bar{X}_i - \bar{X})^2 + 2 \sum_{i=1}^r \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_i)(\bar{X}_i - \bar{X}) \end{aligned}$$

$$\sum_{i=1}^r \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_i)(\bar{X}_i - \bar{X}) = \sum_{i=1}^r (\bar{X}_i - \bar{X}) \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_i) = 0$$

$$s_T = \sum_{i=1}^r \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_i)^2 + \sum_{i=1}^r \sum_{j=1}^{n_i} (\bar{X}_i - \bar{X})^2$$



# 方差分析

---

$$S_T = S_e + S_A$$

$$S_e = \sum_{i=1}^r \sum_{j=1}^{n_i} \left( X_{ij} - \bar{X}_i \right)^2$$

水平内样本方差的求和  
也即误差平方和

$$S_A = \sum_{i=1}^r \sum_{j=1}^{n_i} \left( \bar{X}_i - \bar{X} \right)^2 = \sum_{i=1}^r n_i \left( \bar{X}_i - \bar{X} \right)^2$$

水平间样本方差的求和



## 误差平方和 (组内平方和)

$$\begin{cases} X_{ij} = \mu + \alpha_i + \varepsilon_{ij} \\ \sum_{i=1}^r n_i \alpha_i = 0 \\ \varepsilon_{ij} \text{ 相互独立, } \varepsilon_{ij} \sim N(0, \sigma^2) \end{cases}$$

$$s_e = \sum_{i=1}^r \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_i)^2$$

$$X_{ij} = \mu + \alpha_i + \varepsilon_{ij}$$

$$\bar{X}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} (\mu + \alpha_i + \varepsilon_{ij}) = \mu + \alpha_i + \bar{\varepsilon}_i, \quad \bar{\varepsilon}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} \varepsilon_{ij}$$

$$s_e = \sum_{i=1}^r \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_i)^2 = \sum_{i=1}^r \sum_{j=1}^{n_i} (\varepsilon_{ij} - \bar{\varepsilon}_i)^2$$

$$E(s_e) = \sum_{i=1}^r E \left[ \sum_{j=1}^{n_i} (\varepsilon_{ij} - \bar{\varepsilon}_i)^2 \right] \stackrel{\varepsilon_{ij} \sim N(0, \sigma^2)}{=} \sum_{i=1}^r (n_i - 1) \sigma^2 = (n - r) \sigma^2$$



## 因素平方和 (组间平方和)

$$\begin{cases} X_{ij} = \mu + \alpha_i + \varepsilon_{ij} \\ \sum_{i=1}^r n_i \alpha_i = 0 \\ \varepsilon_{ij} \text{ 相互独立, } \varepsilon_{ij} \sim N(0, \sigma^2) \end{cases}$$

$$s_A = \sum_{i=1}^r \sum_{j=1}^{n_i} (\bar{X}_i - \bar{X})^2 = \sum_{i=1}^r n_i (\bar{X}_i - \bar{X})^2$$

$$\bar{X} = \frac{1}{n} \sum_{i=1}^r \sum_{j=1}^{n_i} (\mu + \alpha_i + \varepsilon_{ij}) = \mu + \bar{\varepsilon}, \quad \bar{\varepsilon} = \frac{1}{n} \sum_{i=1}^r \sum_{j=1}^{n_i} \varepsilon_{ij}$$

$$s_A = \sum_{i=1}^r n_i (\bar{X}_i - \bar{X})^2 \stackrel{\bar{X}_i = \mu + \alpha_i + \bar{\varepsilon}_i}{=} \sum_{i=1}^r n_i (\alpha_i + \bar{\varepsilon}_i - \bar{\varepsilon})^2$$

$$\begin{aligned} E(s_A) &= E \sum_{i=1}^r n_i \left( \alpha_i^2 + 2\alpha_i (\bar{\varepsilon}_i - \bar{\varepsilon}) + (\bar{\varepsilon}_i - \bar{\varepsilon})^2 \right) \\ &= \sum_{i=1}^r n_i \alpha_i^2 + E \sum_{i=1}^r n_i (\bar{\varepsilon}_i - \bar{\varepsilon})^2 = \sum_{i=1}^r n_i \alpha_i^2 + (r-1) \sigma^2 \end{aligned}$$



## 假设检验

---

$$H_0: \mu_1 = \mu_2 = \cdots = \mu_r \quad (H_0: \alpha_1 = \alpha_2 = \cdots = \alpha_r = 0)$$

原假设成立时

$$E\left(\frac{s_A}{r-1}\right) = \sigma^2 \quad \frac{s_A}{\sigma^2} \sim \chi^2(r-1)$$

$$\frac{s_e}{\sigma^2} \sim \chi^2(n-r), \quad \frac{\cancel{s_A} / (r-1)}{\cancel{s_e} / (n-r)} \sim F(r-1, n-r)$$

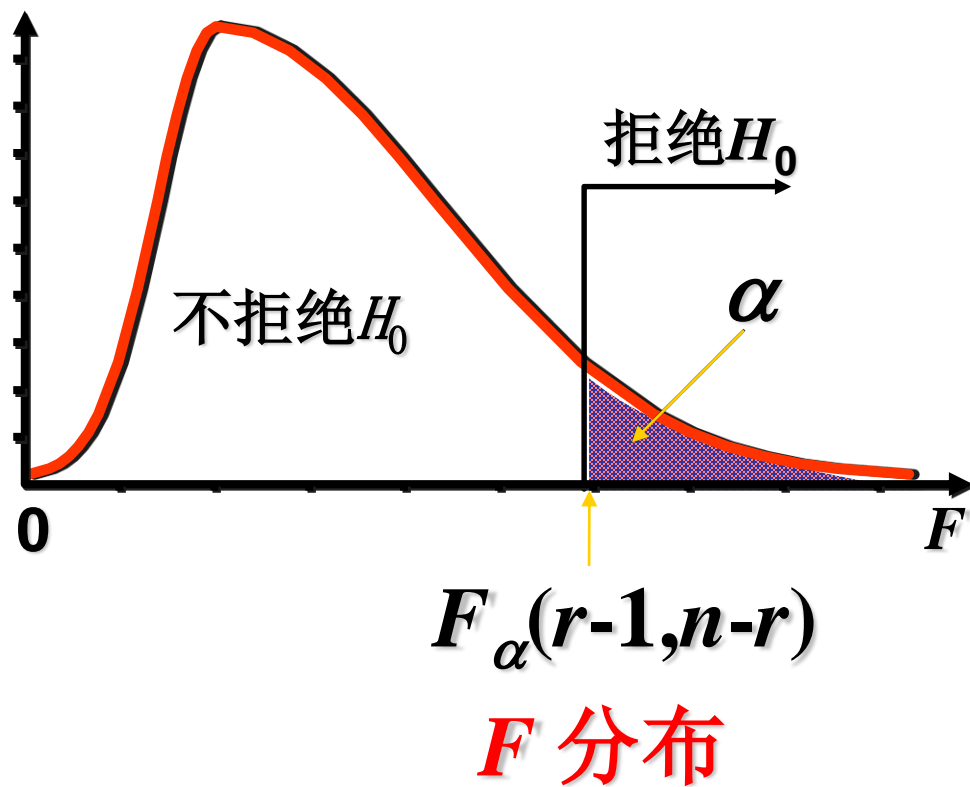




# 方差分析表

方差 来源	平方和	自由度	均值	<b>F值</b>  F(r-1,n-r)
因素	$S_A$	$r-1$	$\frac{S_A}{r-1}$	$\frac{\frac{S_A}{r-1}}{\frac{S_e}{n-r}}$
误差	$S_e$	$n-r$	$\frac{S_e}{n-r}$	
总和	$S_T$	$n-1$		

# F分布与拒绝域



# 杀虫剂对比

试验号\杀虫率%\农药	A1	A2	A3	A4	A5	A6
1	87	90	56	55	92	75
2	85	88	62	48	99	72
3	80	87			95	81
4		94			91	

## ANOVA

杀虫率

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	3794.500	5	758.900	51.162	.000
Within Groups	178.000	12	14.833		
Total	3972.500	17			

## 灯丝使用寿命对比

灯丝\使用寿命\灯泡	1	2	3	4	5	6	7	8
甲	1600	1610	1650	1680	1700	1700	1780	1800
乙	1500	1640	1400	1700	1750			
丙	1640	1550	1600	1620	1640	1600	1740	
丁	1510	1520	1530	1570	1640	1680		

### ANOVA

使用寿命

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	39776.456	3	13258.819	1.638	.209
Within Groups	178088.929	22	8094.951		
Total	217865.385	25			



# 方差齐性检验

---

单因素方差分析模型

因素  $A$  取  $r$  个水平:  $A_1, A_2, \dots, A_r$

$r$  个正态水平:  $X_1, X_2, \dots, X_r$ ,  $X_i \sim N(\mu_i, \sigma^2) (i = 1, 2, \dots, r)$

方差分析模型的F检验对正态性的偏离具有一定的稳健性，  
也就是，正态性假设不满足时，对均值相等的F检验影响较小

但是，F检验对方差齐性的偏离较为敏感，

所以， $r$ 个方差的齐性检验很有必要



## Hartley检验

---

- 仅适用于个水平样本量相等的场合

因素  $A$  取  $r$  个水平:  $A_1, A_2, \dots, A_r$

$r$  个正态水平:  $X_1, X_2, \dots, X_r$ ,  $X_i \sim N(\mu_i, \sigma^2) (i = 1, 2, \dots, r)$

$r$  个水平 下进行的独立试验次数均相等:  $n_1 = n_2 = \dots = n_r$

Hartley, H.O. The Use of Range in Analysis of Variance,  
*Biometrika*, 1950 (37), 271–280.

# Hartley检验

$$F_{\max} = \frac{\max\{S_1^2, S_2^2, \dots, S_k^2\}}{\min\{S_1^2, S_2^2, \dots, S_k^2\}}$$

## Critical Values of $F_{\max}$ for Hartley's Homogeneity of Variance Test

The upper value in each box is for  $\alpha = 0.05$ . The lower value is for  $\alpha = 0.01$ . The test assumes that there are equal sample sizes in each group (n). For unequal sample sizes, use the smaller of the df for the two variances being compared.

DF (n-1)	Number of treatments (k)										
	2	3	4	5	6	7	8	9	10	11	12
2	39.0 199	87.5 448	142 729	202 1036	266 1362	333 1705	403 2063	475 2432	550 2813	626 3204	714 3605
3	15.4 47.5	27.8 85.0	39.2 120	50.7 151	62.0 184	72.9 216	83.5 249	93.9 281	104 310	114 337	124 361
4	9.6 23.2	15.5 37.0	20.6 49.0	25.2 59	29.5 69	33.6 79	37.5 89	41.1 97	44.6 106	48.0 113	51.4 120
5	7.2 14.9	10.8 22.0	13.7 28.0	16.3 33	18.7 38	20.8 42	22.9 46	24.7 50	26.5 54	28.2 57	29.9 60
6	5.82 11.1	8.38 15.5	10.4 19.1	12.1 22	13.7 25	15.0 27	16.3 30	17.5 32	18.6 34	19.7 36	20.7 37



## Bartlett检验

---

$$S_i^2 = \frac{1}{n_i - 1} \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_i)^2 = \frac{Q_i}{f_i}, \quad i = 1, 2, \dots, r$$

$$\text{误差均方MS}_e = \frac{1}{f_e} S_e = \frac{1}{f_e} \sum_{i=1}^r \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_i)^2 = \frac{1}{f_e} \sum_{i=1}^r Q_i = \sum_{i=1}^r \frac{f_i}{f_e} S_i^2$$

$$\text{其中 } f_e = f_1 + f_2 + \dots + f_r = \sum_{i=1}^r (n_i - 1) = n - r$$

$$r \text{ 个样本方差的几何平均GMS}_e = \left[ (S_1^2)^{f_1} (S_2^2)^{f_2} \dots (S_r^2)^{f_r} \right]^{1/f_e}$$

$$\text{GMS}_e \leq \text{MS}_e,$$

等号仅在所有的 $S_i^2$ 均相等时成立，而且各个 $S_i^2$ 的差异越大，这两个平均值的差异就越大





## Bartlett检验

---

$$\text{GMS}_e \leq \text{MS}_e,$$

等号仅在所有的 $S_i^2$ 均相等时成立，而且各个 $S_i^2$ 的差异越大，这两个平均值的差异就越大

$$\text{拒绝域的形式: } W = \left\{ \ln \frac{\text{MS}_e}{\text{GMS}_e} > d \right\}$$

$$\text{Bartlett证明: } \frac{f_e}{C} (\ln \text{GMS}_e - \ln \text{MS}_e) \sim \chi^2(r-1)$$

$$\text{其中 } C = 1 + \frac{1}{3(r-1)} \left( \sum_{i=1}^r \frac{1}{f_i} - \frac{1}{f_e} \right)$$

Bartlett, M. S. (1937). Properties of sufficiency and statistical tests.

*Proceedings of the Royal Statistical Society, Series A* (160), 268–282

## 双因素方差分析

双因素多水平等重复试验结果表

$B$ $A$	$B_1$	$B_2$	$\dots$	$B_s$
$A_1$	$X_{111}, X_{112}, \dots, X_{11t}$	$X_{121}, X_{122}, \dots, X_{12t}$	$\dots$	$X_{1s1}, X_{1s2}, \dots, X_{1st}$
$A_2$	$X_{211}, X_{212}, \dots, X_{21t}$	$X_{221}, X_{222}, \dots, X_{22t}$	$\dots$	$X_{2s1}, X_{2s2}, \dots, X_{2st}$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
$A_r$	$X_{r11}, X_{r12}, \dots, X_{r1t}$	$X_{r21}, X_{r22}, \dots, X_{r2t}$	$\dots$	$X_{rs1}, X_{rs2}, \dots, X_{rst}$

因素  $A$  取  $r$  个水平, 因素  $B$  取  $s$  个水平,  
共有  $sr$  种不同的水平组合  $(A_i, B_j)$ ,  $sr$  个总体  $X_{ij}$   
每种水平组合独立做  $t$  次试验



## 双因素方差分析实例

种子\亩产量\农药	B1	B2	B3	B4
A1	173, 172	174, 176	177, 179	172, 173
A2	175, 173	178, 177	174, 175	170, 171
A3	177, 175	174, 174	174, 173	169, 169



## 方差分析模型

---

$X_{ij} \sim N(\mu_{ij}, \sigma^2)$  ( $i = 1, 2, \dots, r, j = 1, 2, \dots, s$ ) , 相互独立

$X_{ijk} \sim N(\mu_{ij}, \sigma^2), k = 1, 2, \dots, t$

$\varepsilon_{ijk} = X_{ijk} - \mu_{ij} \sim N(0, \sigma^2)$  ( $i = 1, 2, \dots, r, j = 1, 2, \dots, s, k = 1, 2, \dots, t$ )

$$\begin{cases} X_{ijk} = \mu_{ij} + \varepsilon_{ijk} \\ \varepsilon_{ijk} \sim N(0, \sigma^2) \end{cases}$$



# 方差分析模型

理论总均值  $\mu = \frac{1}{rs} \sum_{i=1}^r \sum_{j=1}^s \mu_{ij}$

$$\mu_{i*} = \frac{1}{s} \sum_{j=1}^s \mu_{ij}, \quad \alpha_i = \mu_{i*} - \mu, \quad (i=1, 2, \dots, r) \quad A_i \text{对试验结果的效应}$$

$$\mu_{*j} = \frac{1}{r} \sum_{i=1}^r \mu_{ij}, \quad \beta_j = \mu_{*j} - \mu, \quad (j=1, 2, \dots, s) \quad B_j \text{对试验结果的效应}$$

$$\sum_{i=1}^r \alpha_i = 0, \quad \sum_{j=1}^s \beta_j = 0, \quad \gamma_{ij} = (\mu_{ij} - \mu) - \alpha_i - \beta_j$$

其中  $\mu_{ij} - \mu$  反映水平组合  $(A_i, B_j)$  对试验指标的总效应

$\gamma_{ij}$  称为  $A_i$  与  $B_j$  对试验指标的交互效应

$\alpha_i$ 、 $\beta_j$  有时也称为相应水平的主效应

## 交互作用实例（大豆亩产）

磷肥 氮肥	$P_1 = 0$	$P_2 = 3$	$P_1 \rightarrow P_2$
$N_1 = 0$	300	340	$340 - 300 = 40$
$N_2 = 5$	330	440	$440 - 330 = 110$
$N_1 \rightarrow N_2$	30	100	$440 - 300 = 140$



## 双因素方差分析的数学模型

$$\sum_{i=1}^r \gamma_{ij} = \sum_{j=1}^s \gamma_{ij} = 0, \quad \mu_{ij} = \mu + \alpha_i + \beta_j + \gamma_{ij}$$

$$\left\{ \begin{array}{l} X_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \varepsilon_{ijk} \quad (i = 1, 2, \dots, r, j = 1, 2, \dots, s, k = 1, 2, \dots, t) \\ \sum_{i=1}^r \alpha_i = 0, \quad \sum_{j=1}^s \beta_j = 0, \quad \sum_{i=1}^r \gamma_{ij} = \sum_{j=1}^s \gamma_{ij} = 0 \\ \varepsilon_{ijk} \sim N(0, \sigma^2), \text{ 且所有的 } \varepsilon_{ijk} \text{ 相互独立} \end{array} \right.$$



## 假设检验

---

$$H_{01} : \alpha_1 = \alpha_2 = \cdots = \alpha_r = 0$$

$$H_{02} : \beta_1 = \beta_2 = \cdots = \beta_s = 0$$

$$H_{03} : \gamma_{ij} = 0, \quad (i = 1, 2, \cdots, r, \quad j = 1, 2, \cdots, s)$$





# 方差分析

---

$$\bar{X} = \frac{1}{rst} \sum_{i=1}^r \sum_{j=1}^s \sum_{k=1}^t X_{ijk}$$

$$\bar{X}_{ij*} = \frac{1}{t} \sum_{k=1}^t X_{ijk}, \quad \bar{X}_{i**} = \frac{1}{st} \sum_{j=1}^s \sum_{k=1}^t X_{ijk}, \quad \bar{X}_{*j*} = \frac{1}{rt} \sum_{i=1}^r \sum_{k=1}^t X_{ijk}$$

$$S_T = \sum_{i=1}^r \sum_{j=1}^s \sum_{k=1}^t (X_{ijk} - \bar{X})^2, \quad S_e = \sum_{i=1}^r \sum_{j=1}^s \sum_{k=1}^t (X_{ijk} - \bar{X}_{ij*})^2$$

$$S_A = st \sum_{i=1}^r (X_{i**} - \bar{X})^2, \quad S_B = rt \sum_{j=1}^s (\bar{X}_{*j*} - \bar{X})^2$$

$$S_{A \times B} = t \sum_{i=1}^r \sum_{j=1}^s (\bar{X}_{ij*} - \bar{X}_{i**} - \bar{X}_{*j*} + \bar{X})^2$$



# 方差分析

---

$$S_T = S_e + S_A + S_B + S_{A \times B}$$

$$\bar{\varepsilon} = \frac{1}{rst} \sum_{i=1}^r \sum_{j=1}^s \sum_{k=1}^t \varepsilon_{ijk}, \quad \bar{\varepsilon}_{ij*} = \frac{1}{t} \sum_{k=1}^t \varepsilon_{ijk}$$

$$\bar{\varepsilon}_{i**} = \frac{1}{st} \sum_{j=1}^s \sum_{k=1}^t \varepsilon_{ijk}, \quad \bar{\varepsilon}_{*j*} = \frac{1}{rt} \sum_{i=1}^r \sum_{k=1}^t \varepsilon_{ijk}$$

$$\bar{X}_{ij*} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \bar{\varepsilon}_{ij*}, \quad \bar{X} = \mu + \bar{\varepsilon}$$

$$\bar{X}_{i**} = \mu + \alpha_i + \bar{\varepsilon}_{i**}, \quad \bar{X}_{*j*} = \mu + \beta_j + \bar{\varepsilon}_{*j*}$$



# 方差分析

---

$$S_T = S_e + S_A + S_B + S_{A \times B}$$

$$S_e = \sum_{i=1}^r \sum_{j=1}^s \sum_{k=1}^t (\epsilon_{ijk} - \bar{\epsilon}_{ij*})^2$$

$$S_A = st \sum_{i=1}^r (\alpha_i + \bar{\epsilon}_{i**} - \bar{\epsilon})^2$$

$$S_B = rt \sum_{j=1}^s (\beta_j + \bar{\epsilon}_{*j*} - \bar{\epsilon})^2$$

$$S_{A \times B} = t \sum_{i=1}^r \sum_{j=1}^s (\gamma_{ij} + \bar{\epsilon}_{ij*} - \bar{\epsilon}_{i**} - \bar{\epsilon}_{*j*} + \bar{\epsilon})$$



## 检验统计量

---

$$E(s_e) = rs(t-1)\sigma^2$$

$$E(s_A) = (r-1)\sigma^2 + st \sum_{i=1}^r \alpha_i^2$$

$$E(s_B) = (s-1)\sigma^2 + rt \sum_{j=1}^s \beta_j^2$$

$$E(s_{A \times B}) = (r-1)(s-1)\sigma^2 + t \sum_{i=1}^r \sum_{j=1}^s \gamma_{ij}^2$$



## 检验统计量

---

$$E(s_e) = rs(t-1)\sigma^2$$

$$\text{当 } H_{01} \text{ 成立时, } s_A / \sigma^2 \sim \chi^2(r-1)$$

$$\text{当 } H_{02} \text{ 成立时, } s_B / \sigma^2 \sim \chi^2(s-1)$$

$$\text{当 } H_{03} \text{ 成立时, } s_{A \times B} / \sigma^2 \sim \chi^2((r-1)(s-1))$$



# 方差分析表

方差来源	平方和	自由度	均方	F 值
因素 $A$	$s_A$	$r - 1$	$\bar{s}_A = s_A / (r - 1)$	$F_A = \bar{s}_A / \bar{s}_e$
因素 $B$	$s_B$	$s - 1$	$\bar{s}_B = s_B / (s - 1)$	$F_B = \bar{s}_B / \bar{s}_e$
交互作用 $A \times B$	$s_{A \times B}$	$(r - 1)(s - 1)$	$\bar{s}_{A \times B} = s_{A \times B} / (r - 1)(s - 1)$	$F_{A \times B} = \bar{s}_{A \times B} / \bar{s}_e$
误差	$s_e$	$rs(t - 1)$	$\bar{s}_e = s_e / rs(t - 1)$	
总和	$s_T$	$rst - 1$		



## 双因素方差分析实例

种子\亩产量\农药	B1	B2	B3	B4
A1	173, 172	174, 176	177, 179	172, 173
A2	175, 173	178, 177	174, 175	170, 171
A3	177, 175	174, 174	174, 173	169, 169



## 方差分析表

Dependent Variable: 亩产量

Source	平方和	自由度	Mean Square	F	Sig.
种子	8.083	2	4.042	4.409	.037
农肥	90.833	3	30.278	33.030	.000
种子 * 农肥	51.917	6	8.653	9.439	.001
Error	11.000	12	.917		
Total	161.833	23			





# 为什么需要试验设计

钢铁厂要提高铁水温度

水平\因素	A 焦比	B 风压/133pa	C 底胶高度/m
1	1:16	170	1.2
2	1:18	230	1.5
3	1:14	200	1.3



## 试验设计

---

在多因素、多水平的试验中，如果对每个因素的各个水平都互相搭配进行全面的试验，需要做的试验次数会很多。

例如，2个8水平的因素： $8^2 = 64$ 次试验，

6个3水平的因素，需要 $3^6 = 6561$ 次试验

在保证试验效果的前提下，尽可能减少试验次数



## 正交表 $L_9(3^4)$

(4因素3水平)

试验号\列号	1	2	3	4
1	1	1	1	1
2	1	2	2	2
3	1	3	3	3
4	2	1	2	3
5	2	2	3	1
6	2	3	1	2
7	3	1	3	2
8	3	2	1	3
9	3	3	2	1

类似有  $L_4(2^3)$ ,  $L_{12}(2^{11})$ ,  $L_{27}(3^{13})$  等

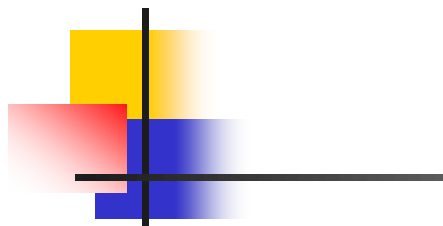


## 混合水平的正交表

正交表  $L_8(4^1 \times 2^4)$

试验号	列 号				
	1	2	3	4	5
1	1	1	1	1	1
2	1	2	2	2	2
3	2	1	1	2	2
4	2	2	2	1	1
5	3	1	2	1	2
6	3	2	1	2	1
7	4	1	2	2	1
8	4	2	1	1	2

# 钢水温度



因素 试验号	1 A	2 B	3 C	铁水温度/℃	铁水温度值 减去 1350
1	1	1	1	1365	15
2	1	2	2	1395	45
3	1	3	3	1385	35
4	2	1	2	1390	40
5	2	2	3	1395	45
6	2	3	1	1380	30
7	3	1	3	1390	40
8	3	2	1	1390	40
9	3	3	2	1410	60
$K_1$	95	95	85	极差绝对值越大说明 该因素影响越显著	
$K_2$	115	130	145		
$K_3$	140	155	120		
$\kappa_1 \left( = \frac{K_1}{3} \right)$	31.7	31.7	35.0		
$\kappa_2 \left( = \frac{K_2}{3} \right)$	38.3	43.3	48.3		
$\kappa_3 \left( = \frac{K_3}{3} \right)$	46.7	41.7	40.0		
极 差	15.0	11.6	13.3		
优 方 案	$A_1$	$B_2$	$C_2$		

# 交互作用表


$L_8(2^7)$ 两列之间的交互作用表为表 4.3.3.

表 4.3.3

列号 列号	1	2	3	4	5	6	7
(1)		3	2	5	4	7	6
(2)			1	6	7	4	5
(3)				7	6	5	4
(4)					1	2	3
(5)						3	2
(6)							1
(7)							

所以  $L_8(2^7)$ 的表头设计可为

列号	1	2	3	4	5	6	7
因子	A	B	$A \times B$	C	$A \times C$	$B \times C$	$A \times B \times C$



例 4.3.3 某橡胶配方考虑因子水平表如表4.3.6.

表 4.3.6

因子 水平	促进剂总量(A)	炭墨品种(B)	硫磺分量(C)
1	1.5	天津耐高磨	2.5
2	2.5	天津耐高磨与长春 硬炭黑并用	2.0

试验指标:弯曲次数(越多越好).

表 4.3.7 橡胶配方试验计算表

列号 试验号	A 1	B 2	A×B 3	C 4	A×C 5	B×C 6	7	弯曲(万次) $y_i$
1	1	1	1	1	1	1		1.5
2	1	1	1	2	2	2		2.0
3	1	2	2	1	1	2		2.0
4	1	2	2	2	2	1		1.5
5	2	1	2	1	2	1		2.0
6	2	1	2	2	1	2		3.0
7	2	2	1	1	2	2		2.5
8	2	2	1	2	1	1		2.0
$k_{1j}$	7	8.5	8	8	8.5	7		$\sum y_i = 16.5$
$k_{2j}$	9.5	8	8.5	8.5	8	9.5		
$k_{1j} - k_{2j}$	-2.5	0.5	-0.5	-0.5	0.5	-2.5		
	$B_1$					$B_2$		
$C_1$	$\frac{1.5+2.0}{2} = 1.75$					$\frac{2.0+2.5}{2} = 2.25$		
$C_2$	$\frac{2.0+3.0}{2} = 2.5$					$\frac{1.5+2.0}{2} = 1.75$		



$$Q = Q_A + Q_B + Q_C + Q_{A \times B} + Q_{A \times C} + Q_{B \times C} + Q_e.$$

Q按下式计算：

$$Q = \sum_{i=1}^n y_i^2 - \frac{1}{n} \left( \sum_{i=1}^n y_i \right)^2,$$

$Q_A, Q_B, Q_C$  按下式计算（因  $n_i = \frac{n}{r_j}, \bar{y} = \frac{1}{n} \sum_{i=1}^{r_j} k_{ij}$ ）：

$$Q_j = \sum_{i=1}^{r_j} n_i \bar{y}_i^2 - n \bar{y}^2 = \frac{r_j}{n} \sum_{i=1}^{r_j} k_{ij}^2 - \frac{1}{n} \left( \sum_{i=1}^{r_j} k_{ij} \right)^2.$$

当  $r_j=2$  时，

$$Q_j = \frac{2}{n} (k_{1j}^2 + k_{2j}^2) - \frac{1}{n} (k_{1j} + k_{2j})^2 = \frac{1}{n} (k_{1j} - k_{2j})^2,$$

$Q_{A \times B}, Q_{A \times C}, Q_{B \times C}$  亦按上式计算，这是因  $A \times B, A \times C, B \times C$  也像  $A, B, C$  一样只占一列。如果它们占两列，则它们分别为相应两列的平方和之和。

表 4.3.8 方差分析表

方差来源	平方和	自由度	方差	F 值	显著性
A	$Q_A=0.78125$	1	$Q_A/1=0.78125$	25	*
B	$Q_B=0.03125$	1	$Q_B/1=0.03125$	1	
C	$Q_C=0.03125$	1	$Q_C/1=0.03125$	1	
$A \times B$	$Q_{A \times B}=0.03125$	1	$Q_{A \times B}/1=0.03125$	1	
$A \times C$	$Q_{A \times C}=0.03125$	1	$Q_{A \times C}/1=0.03125$	1	
$B \times C$	$Q_{B \times C}=0.78125$	1	$Q_{B \times C}/1=0.78125$	25	*
误差	$Q_e=0.03125$	1	$Q_e/1=0.03125$		
总和	$Q=1.71875$	7			



## 正交试验设计

---

20世纪20、30年代，费希尔在试验设计方面做出了一系列先驱性的贡献。

20世纪上半叶，正交设计方法已经在数学界中提出。

到40年代后期，日本统计学家田口玄一博士首次将正交设计方法应用到日本的电话机试验上。到1970年，日本使用正交设计方法项目超过100万次，其中1/3效果显著。

20世纪70年代以来，我国应用正交设计取得一大批优秀成果。



# 作业

---

## 习题四（200-201页）

1, 2, 3