第九章 方差分析及回归分析

§1 单因素的方差分析

§2 双因素的方差分析

§1 单因素试验的方差分析

试验结果（试验指标）：考察事物的结果

因素：影响试验指标的条件

单因素试验：为了考虑某个因素A对所考察的随机变量X的影响，在试验中让其他因素保持不变，仅让因素A改变。

因素A所处的状态称为水平。

一些例子：

- 温度变化对疾病传染的影响
- 电池充满时间对电池容量的影响

1. 单因素试验的数学模型

设在单因素试验中，所考察的因素A有s个水平$A_1, \dots, A_s$，在水平A_j下进行$n_j(\geq 2)$次独立试验，得到结果：

水平	A_1	A_2	…	A_s
样本观测	X_{11} X_{21} $\vdots$ $X_{n_1 1}$	X_{12} X_{22} $X_{n_2 2}$	…	X_{1s} X_{2s} $\vdots$ $X_{n_s s}$
样本总和	$T_{.1}$	$T_{.2}$	…	$T_{.s}$
样本均值	$\bar{X}_{.1}$	$\bar{X}_{.2}$	…	$\bar{X}_{.s}$
总体均值	μ_1	μ_2	…	μ_s

$$T_{.j} = \sum_{i=1}^{n_j} X_{ij}$$

$$\bar{X}_{.j} = \frac{1}{n_j}\sum_{i=1}^{n_j} X_{ij}$$

假设：各水平$A_j(j=1,\ldots,s)$下的样本$X_{ij}\sim\left(\mu_j,\sigma^2\right)$，$i=1,2,\ldots,n_j,j=1,2,\ldots,s$，且相互独立。

故$X_{ij}-\mu_j$可看成随机误差，记$X_{ij}-\mu_j=\epsilon_{ij}$，则

$$\begin{cases} X_{ij}=\mu_j+\epsilon_{ij} \\ \epsilon_{ij}\sim N(0,\sigma^2),\text{各}\epsilon_{ij}\text{独立} \\ i=1,2,\ldots,n_j,j=1,2,\ldots,s \end{cases}$$

其中μ_j与σ^2均为未知参数，上式即为单因素试验方差分析的数学模型

检验 s 个总体$N(\mu_1,\sigma^2),\dots,N(\mu_s,\sigma^2)$的均值是否相等，即检验：

$$H_0{:}\,\mu_1=\cdots=\mu_s,\ H_1{:}\,\mu_1,\dots,\mu_s\text{不全相等}$$

总平均：$\mu=\frac{1}{n}\sum_{j=1}^{s}n_j\mu_j$ （其中$n=\sum_{j=1}^{s}n_j$ ）

水平A_j的效应：A_j的总体平均值与总平均的差异

$$\delta_j=\mu_j-\mu\ (j=1,2,\dots,s)$$

$$\boxed{\mu_j = \mu + \delta_j}$$

数学模型可写为：

$$\begin{cases} X_{ij} = \mu + \delta_j + \epsilon_{ij} \\ \epsilon_{ij} \sim N(0, \sigma^2), \text{各}\epsilon_{ij}\text{独立} \\ \sum_{j=1}^{s} n_j \delta_j = 0 \\ i = 1, 2, \dots, n_j, j = 1,2, \dots, s \end{cases}$$

假设检验：

$$\begin{cases} H_0: \delta_1 = \delta_2 = \cdots = \delta_s = 0 \\ H_1: \delta_1, \delta_2, \dots, \delta_s \text{不全为}0 \end{cases}$$

2. 平方和的分解

总偏差平方和：全部试验数据间的差异

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

- 水平A_j下的样本平均值：

$$\bar{X}_{.j} = \frac{1}{n_j}\sum_{i=1}^{n_j} X_{ij}$$

- 样本总平均值：

$$\bar{X} = \frac{1}{n}\sum_{j=1}^{s}\sum_{i=1}^{n_j} X_{ij}$$

总平均：

$$\mu = \frac{1}{n}\sum_{j=1}^{s} n_j \mu_j$$

$$S_T = \sum_{j=1}^{s}\sum_{i=1}^{n_j}\left(X_{ij} - \bar{X}_{.j} + \bar{X}_{.j} - \bar{X}\right)^2$$

$$= \sum_{j=1}^{s}\sum_{i=1}^{n_j}\left(X_{ij} - \bar{X}_{.j}\right)^2 + \sum_{j=1}^{s}\sum_{i=1}^{n_j}\left(\bar{X}_{.j} - \bar{X}\right)^2$$

$$+ 2\sum_{j=1}^{s}\sum_{i=1}^{n_j}\left(X_{ij} - \bar{X}_{.j}\right)\left(\bar{X}_{.j} - \bar{X}\right)$$

$$\sum_{j=1}^{s}\sum_{i=1}^{n_j}\left(X_{ij} - \bar{X}_{.j}\right)\left(\bar{X}_{.j} - \bar{X}\right) = \sum_{j=1}^{s}\left(\bar{X}_{.j} - \bar{X}\right)\sum_{i=1}^{n_j}\left(X_{ij} - \bar{X}_{.j}\right)$$

$$= \sum_{j=1}^{s}\left(\bar{X}_{.j} - \bar{X}\right)\left(\sum_{i=1}^{n_j} X_{ij} - n_j\bar{X}_{.j}\right) = 0$$

$$S_T=\sum_{j=1}^{s}\sum_{i=1}^{n_j}\left(\bar{X}_{.j}-\bar{X}\right)^2+\sum_{j=1}^{s}\sum_{i=1}^{n_j}\left(X_{ij}-\bar{X}_{.j}\right)^2=S_A+S_E$$

其中，误差平方和

$$S_E=\sum_{j=1}^{s}\sum_{i=1}^{n_j}\left(X_{ij}-\bar{X}_{.j}\right)^2$$

效应平方和

$$S_A=\sum_{j=1}^{s}\sum_{i=1}^{n_j}\left(\bar{X}_{.j}-\bar{X}\right)^2$$

$$=\sum_{j=1}^{s}n_j\left(\bar{X}_{.j}-\bar{X}\right)^2=\sum_{j=1}^{s}n_j\bar{X}_{.j}^2-n\bar{X}^2$$

3. S_E, S_A的统计性质：

(1) $S_E/\sigma^2 \sim \chi^2(n-s)$

由$\frac{\sum_{i=1}^{n_j}\left(X_{ij}-\bar{X}_{.j}\right)^2}{\sigma^2} \sim \chi^2(n_j-1)$, 及分布的可加性得

$$S_E/\sigma^2 \sim \chi^2\left(\sum_{j=1}^{s}(n_j-1)\right)$$

(2) $E[S_E]=(n-s)\sigma^2$

(3) $E[S_A]=\sum_{j=1}^{s} n_j\delta_j^2+(s-1)\sigma^2$

（利用$\bar{X}_{.j} \sim N\left(\mu_j, \frac{\sigma^2}{n_j}\right)$, $\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$及$S_A$的定义形式）

(4) S_A与S_E相互独立

(5) 当H_0为真时，

$$\frac{S_A}{\sigma^2} \sim \chi^2(s-1)$$

(6) 当H_0为真时，

$$F = \frac{S_A/(s-1)}{S_E/(n-s)} \sim F(s-1, n-s)$$

4. 假设检验的拒绝域

$$\begin{cases} H_0: \delta_1 = \delta_2 = \cdots = \delta_s = 0 \\ H_1: \delta_1, \delta_2, \ldots, \delta_s \text{不全为} 0 \end{cases}$$

H_0为真时，$\frac{S_A}{s-1}$是σ^2的无偏估计

（否则，$\mathrm{E}\left[\frac{S_A}{s-1}\right] > \sigma^2$ ）

无论H_0是否为真，$\frac{S_E}{n-s}$是σ^2的无偏估计

检验问题的拒绝域为：

$$F = \frac{S_A/(s-1)}{S_E/(n-s)} \geq F_\alpha(s-1, n-s)$$

单因素试验方差分析表：

方差来源	平方和	自由度	均方	F比
因素A	S_A	$s-1$	$\bar{S}_A=\dfrac{S_A}{s-1}$	$\bar{S}_A/\bar{S}_E$
误差	S_E	$n-s$	$\bar{S}_E=\dfrac{S_E}{n-s}$	
总和	S_T	$n-1$		

通过检验样本的F比是否落在拒绝域中来决定是否拒绝原假设。

计算S_T、S_A、S_E的简便公式：

记：$T_{.j}=\sum_{i=1}^{n_j}X_{ij}\,,j=1,2,\ldots,s$

$$T_{..}=\sum_{j=1}^{s}\sum_{i=1}^{n_j}X_{ij}$$

则：$S_T=\sum_{j=1}^{s}\sum_{i=1}^{n_j}X_{ij}^2-\frac{T_{..}^2}{n}$

$$S_A=\sum_{j=1}^{s}\frac{T_{.j}^2}{n_j}-\frac{T_{..}^2}{n}$$

$$S_E=S_T-S_A$$

例1 为了比较5种治疗某疾病的药，将30个病人分成5组，每组6人，每组病人使用一种药，记录病人从使用药物开始至痊愈所需的时间，得下表：

药物	治愈所需天数
1	5，8，7，7，10，8
2	4，6，6，3，5，6
3	6，4，4，5，4，3
4	7，4，6，6，3，5
5	9，3，5，7，7，6

问：所有药物的效果是否有差别？

解：

$$H_0: \mu_1 = \mu_3 = \mu_4 = \mu_5, \qquad H_1: \mu_1, \ldots, \mu_5 \text{不全相等}$$

$$s = 5, n_1 = \cdots = n_5 = 6, n = 30, \sum_{j=1}^{s}\sum_{i=1}^{n_j} X_{ij}^2 = 1047$$

$$T_{.1} = 45, T_{.2} = 30, T_{.3} = 26, T_{.4} = 31, T_{.5} = 37, T_{..} = 169$$

方差来源	平方和	自由度	均方	F比
因素A	$S_A = 36.467$	$s-1$=4	$\bar{S}_A = 9.117$	3.90
误差	$S_E = 58.500$	$n-s$=25	$\bar{S}_E = 2.34$	
总和	$S_T = 94.967$	$n-1$=29		

$> F_{0.05}(4,25) = 2.76$
故拒绝H_0

5. 未知参数的估计

$\widehat{\sigma^2} = \frac{S_E}{n-s}$ 是σ^2的无偏估计

$\hat{\mu} = \bar{X}$, $\widehat{\mu_j} = \bar{X}_{.j}$分别是 μ, μ_j 的无偏估计

$$(E[\bar{X}] = \mu,\ E\left[\bar{X}_{.j}\right] = \frac{1}{n_j}\sum_{i=1}^{n_j} E\left[X_{ij}\right] = \mu_j)$$

$\widehat{\delta_j} = \bar{X}_{.j} - \bar{X}$是$\delta_j$的无偏估计，且有$\sum_{j=1}^{s} n_j\widehat{\delta_j} = 0$

当拒绝H_0时，需要作出总体$N(\mu_j, \sigma^2)$和$N(\mu_k, \sigma^2)$的均值差$\mu_j - \mu_k = \delta_j - \delta_k$的区间估计。

由于，$E[\bar{X}_{.j}-\bar{X}_{.k}]=\mu_j-\mu_k$，$D[\bar{X}_{.j}-\bar{X}_{.k}]=\sigma^2\left(\frac{1}{n_j}+\frac{1}{n_k}\right)$

$$\widehat{\sigma^2}=\frac{S_E}{n-s},\quad \frac{S_E}{\sigma^2}\sim\chi^2(n-s)$$

所以，

$$\frac{(\bar{X}_{.j}-\bar{X}_{.k})-(\mu_j-\mu_k)}{\sqrt{\bar{S}_E(\frac{1}{n_j}+\frac{1}{n_k})}}\sim t(n-s)$$

可得$\mu_j-\mu_k=\delta_j-\delta_k$的置信区间为：

$$\left(\bar{X}_{.j}-\bar{X}_{.k}\pm t_{\alpha/2}(n-s)\sqrt{\bar{S}_E(\frac{1}{n_j}+\frac{1}{n_k})}\right)$$

例2 求例1中未知参数$\sigma^2, \mu, \delta_j (j = 1,2,3,4,5)$的点估计，以及$\mu_1 - \mu_2, \mu_3 - \mu_5$的置信度为0.95的置信区间。

解：

σ^2的估计：$\widehat{\sigma^2} = \frac{S_E}{n-s} = 2.34$

μ的估计：$\hat{\mu} = \bar{X}$=5.633

μ_j的估计分布为：7.5, 5, 4.333, 5.167, 6.167

δ_j的估计分布为：$1.867, -0.633, -1.3, -0.467, 0.534$

$\sqrt{\bar{S}_E(1/n_j + 1/n_k)} = 0.8832$

查表得：$t_{0.025}(25) = 2.0595$

$$\left(\bar{X}_{.j}-\bar{X}_{.k}\pm t_{\alpha/2}(n-s)\sqrt{\bar{S}_E\left(\frac{1}{n_j}+\frac{1}{n_k}\right)}\right)$$

$\mu_1-\mu_2,\mu_3-\mu_5$的置信度为0.95的置信区间分别为：

（0.681, 4.319）和（-3.653, -0.015）

这说明μ_1与μ_2, μ_3与μ_5的差异都很显著。

	平方和	自由度	均方	F比
因素A	$S_A=36.467$	$s-1$=4	$\bar{S}_A=9.117$	3.90
误差	$S_E=58.500$	$n-s$=25	$\bar{S}_E=2.34$	
总和	$S_T=94.967$	$n-1$=29		

$>F_{0.05}(4,25)=2.76$
故拒绝H_0方差来源

§2 双因素试验的方差分析

一、双因素等重复试验的方差分析

有两个因素作用于试验指标

因素A有r个水平$A_1, \dots, A_r$

因素B有s个水平$B_1, \dots, B_s$

对每对组合(A_i, B_j)都做$t(\geq 2)$次试验，

称为双因素等重复试验。假定所有试验相互独立。

例：机票价格与服务质量对航空运输的影响

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

$X_{ijk} \sim N(\mu_{ij}, \sigma^2)$,
$i = 1, \ldots, r, j = 1, \ldots, s$

数学模型：

$$\begin{cases} X_{ijk} = \mu_{ij} + \epsilon_{ijk} \\ \epsilon_{ijk} \sim N(0,\sigma^2), \text{各}\epsilon_{ijk}\text{独立} \\ i = 1,\dots,r, j = 1,\dots,s, k = 1,\dots,t \end{cases}$$

其中μ_{ij}与σ^2均为未知参数。

总平均：$\mu = \frac{1}{rs}\sum_{i=1}^{r}\sum_{j=1}^{s}\mu_{ij}$

$$\mu_{i.} = \frac{1}{s}\sum_{j=1}^{s}\mu_{ij}\,,\ i = 1, \dots, r$$

$$\mu_{.j} = \frac{1}{r}\sum_{i=1}^{r}\mu_{ij},\ j = 1, \dots, s$$

水平A_i的效应：$\alpha_i = \mu_{i.} - \mu$

水平B_j的效应：$\beta_j = \mu_{.j} - \mu$

$$\mu_{ij} = \mu + \alpha_i + \beta_j + \gamma_{ij}$$

其中，$\gamma_{ij} = \mu_{ij} - \mu_{i.} - \mu_{.j} + \mu$ 称为水平A_i和B_j的交互效应。

数学模型可写为：

$$
\begin{cases}
X_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \epsilon_{ijk} \\
\epsilon_{ijk} \sim N(0, \sigma^2), \qquad \text{各}\epsilon_{ijk}\text{独立} \\
i = 1, \ldots, r, j = 1, \ldots, s, k = 1, \ldots, t \\
\sum_{i=1}^{r} \alpha_i = \sum_{j=1}^{s} \beta_j = \sum_{i=1}^{r} \gamma_{ij} = \sum_{j=1}^{s} \gamma_{ij} = 0
\end{cases}
$$

其中$\mu, \alpha_i, \beta_j, \gamma_{ij}, \sigma^2$均为未知参数。

假设检验：

$$\begin{cases} H_{01}: \alpha_1 = \alpha_2 = \cdots = \alpha_r = 0 \\ H_{11}: \quad \alpha_1, \alpha_2, \ldots, \alpha_r \text{不全为0} \end{cases}$$

$$\begin{cases} H_{02}: \beta_1 = \beta_2 = \cdots = \beta_s = 0 \\ H_{12}: \quad \beta_1, \beta_2, \ldots, \beta_s \text{不全为0} \end{cases}$$

$$\begin{cases} H_{03}: \gamma_{11} = \gamma_{11} = \cdots = \gamma_{rs} = 0 \\ H_{13}: \quad \gamma_{11}, \gamma_{12}, \ldots, \gamma_{rs} \text{不全为0} \end{cases}$$

平方和分解：

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

总偏差平方和$S_T = \sum_{i=1}^{r}\sum_{j=1}^{s}\sum_{k=1}^{t}\left(X_{ijk} - \bar{X}\right)^2$

误差平方和$S_E = \sum_{i=1}^{r}\sum_{j=1}^{s}\sum_{k=1}^{t}\left(X_{ijk} - \bar{X}_{ij.}\right)^2$

效应平方和

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

交互效应平方和$S_{A\times B} = t\sum_{i=1}^{r}\sum_{j=1}^{s}\left(\bar{X}_{ij.} - \bar{X}_{i..} - \bar{X}_{.j.} + \bar{X}\right)^2$

可以证明，$S_T, S_E, S_A, S_B, S_{A\times B}$的自由度分别为$rst-1$, $rs(t-1)$, $r-1$, $s-1$, $(r-1)(s-1)$

当H_{01}: $\alpha_1=\alpha_2=\cdots=\alpha_r=0$ 为真时，可以证明

$$F_A=\frac{S_A/(r-1)}{S_E/(rs(t-1))}\sim F(r-1,rs(t-1))$$

在显著性水平α下，假设H_{01}的拒绝域为：

$$F_A\geq F_\alpha(r-1,rs(t-1))$$

类似地，H_{02}: $\beta_1=\beta_2=\cdots=\beta_s=0$ 的拒绝域为：

$$F_B=\frac{S_B/(s-1)}{S_E/(rs(t-1))}\geq F_\alpha(s-1,rs(t-1))$$

当H_{03}:$\gamma_{11}=\gamma_{11}=\cdots=\gamma_{rs}=0$为真时，可以证明

$$F_{A\times B}=\frac{S_{A\times B}/((r-1)(s-1))}{S_E/(rs(t-1))}\sim F((r-1)(s-1),rs(t-1))$$

在显著性水平α下，假设H_{03}的拒绝域为：

$$F_{A\times B}\geq F_{\alpha}((r-1)(s-1),rs(t-1))$$

双因素等重复试验方差分析表：

方差来源	平方和	自由度	均方	F比
因素A	S_A	$r-1$	$\bar{S}_A=\dfrac{S_A}{r-1}$	$F_A=\dfrac{\bar{S}_A}{\bar{S}_E}$
因素B	S_B	$s-1$	$\bar{S}_B=\dfrac{S_B}{s-1}$	$F_B=\dfrac{\bar{S}_B}{\bar{S}_E}$
交互作用	$S_{A\times B}$	$(r-1)(s-1)$	$\bar{S}_{A\times B}=\dfrac{S_{A\times B}}{(r-1)(s-1)}$	$F_{A\times B}=\dfrac{\bar{S}_{A\times B}}{\bar{S}_E}$
误差	S_E	$rs(t-1)$	$\bar{S}_E=\dfrac{S_E}{rs(t-1)}$	
总和	S_T	$rst-1$		

例 3 某化工企业为了提高产量，选了3种不同浓度、4种不同温度做试验。在同一浓度与温度组合下各做2次试验，其数据如下表所示，在显著性水平$\alpha = 0.05$下不同浓度和不同温度以及它们之间的交叉作用对产量有无显著性影响？

$A\backslash B$	B_1	B_2	B_3	B_4
A_1	14, 10	11, 11	13, 9	10, 12
A_2	9, 7	10, 8	7, 11	6, 10
A_3	5, 11	13, 14	12, 13	14, 10

解：$r = 3, s = 4, t = 2$

方差来源	平方和	自由度	均方	F比
A	44.33	2	$\bar{S}_A = 22.17$	$F_A = 4.09$
B	11.50	3	$\bar{S}_B = 3.83$	$F_B = 0.78$
$A \times B$	37.00	6	$\bar{S}_{A \times B} = 6.17$	$F_{A \times B} = 0.63$
误差	65.00	12	$\bar{S}_E = 5.42$	
总和	147.83	23		

而$F_{0.05}(2,12) = 3.89$，因此只有因素A的F比$F_A = 4.09 > 3.89$，即只有因素A是显著的，浓度不同对产量有显著影响，而温度以及浓度和温度交互作用对产量无显著性影响，所以为了提高产量必须控制好浓度。

例4 火箭使用四种燃料、三种推进器实验，每种燃料和推进器的组合各发射两次，是在显著性水平0.05的条件下，检验不同燃料、不同推进器下的射程是否有显著差异？交互作用是否显著？

		推进器（B）		
		B1	B2	B3
燃料(A)	A1	58.2	56.2	65.3
		52.6	41.2	60.8
	A2	49.1	54.1	51.6
		42.8	50.5	48.4
	A3	60.1	70.9	39.2
		58.3	73.2	40.7
	A4	75.8	58.2	48.7
		71.5	51.0	41.4

解：需要检验假设：

$$H_{01}: \alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = 0$$
$$H_{02}: \beta_1 = \beta_2 = \beta_3 = 0$$
$$H_{03}: \gamma_{11} = \gamma_{11} = \cdots = \gamma_{43} = 0$$

	B1		B2		B3		
A1	58.2	T_{ij} (110.8)	56.2	(97.4)	65.3	(126.1)	334.3
	52.6		41.2		60.8		
A2	49.1	(91.9)	54.1	(104.6)	51.6	(100)	296.5
	42.8		50.5		48.4		
A3	60.1	(118.4)	70.9	(144.1)	39.2	(79.9)	342.4
	58.3		73.2		40.7		
A4	75.8	(147.3)	58.2	109.2	48.7	(90.1)	346.6
	71.5		51.0		41.4		
	468.4		455.3		396.1		1319.8

总偏差平方和

$$S_T=\sum_{i=1}^{4}\sum_{j=1}^{3}\sum_{k=1}^{2}\left(X_{ijk}-\bar{X}\right)^2=2638.29833$$

效应平方和

$$S_A=6\sum_{i=1}^{4}(\bar{X}_{i..}-\bar{X})^2=261.67500$$

$$S_B=8\sum_{j=1}^{3}\left(\bar{X}_{.j.}-\bar{X}\right)^2=370.98083$$

交互效应平方和

$$S_{A\times B}=2\sum_{i=1}^{4}\sum_{j=1}^{3}\left(\bar{X}_{ij.}-\bar{X}_{i..}-\bar{X}_{.j.}+\bar{X}\right)^2=1768.69250$$

误差平方和

$$S_E=S_T-S_A-S_B-S_{A\times B}=236.95000$$

方差来源	平方和	自由度	均方	F比
因素A(燃料)	261.67500	3	87.2250	$F_A = 4.42$
因素B(推进器)	370.98083	2	185.4904	$F_B = 9.39$
交互作用	1768.69250	6	294.7821	$F_{A\times B} = 14.9$
误差	236.95000	12	19.7458	
总和	2638.29833	23		

结论：

$F_{0.05}(3,12) = 3.49 < F_A, F_{0.05}(2,12) = 3.89 < F_B$, 所以在显著性水平0.05下拒绝假设$H_{01}$和$H_{02}$，即推进器和燃料的影响都是显著的

$F_{0.05}(6,12) = 3.00 < F_{A\times B}$，因此拒绝假设$H_{03}$，交互作用也是显著的

二、双因素无重复试验的方差分析

双因素试验中，对每对水平组合只做一次试验（因素之间的交互作用不存在或可以忽略）

数学模型：

$$
\begin{cases}
X_{ij} = \mu + \alpha_i + \beta_j + \epsilon_{ij} \\
\epsilon_{ij} \sim N(0, \sigma^2), \qquad \text{各}\epsilon_{ij}\text{独立} \\
i = 1, \dots, r, j = 1, \dots, s \\
\sum_{i=1}^{r} \alpha_i = \sum_{j=1}^{s} \beta_j = 0
\end{cases}
$$

其中$\mu, \alpha_i, \beta_j, \sigma^2$均为未知参数。

假设检验：

$$\begin{cases} H_{01}: \alpha_1 = \alpha_2 = \cdots = \alpha_r = 0 \\ H_{11}: \quad \alpha_1, \alpha_2, \ldots, \alpha_r \text{不全为0} \end{cases}$$

$$\begin{cases} H_{02}: \beta_1 = \beta_2 = \cdots = \beta_s = 0 \\ H_{12}: \quad \beta_1, \beta_2, \ldots, \beta_s \text{不全为0} \end{cases}$$

双因素无重复试验方差分析表：

方差来源	平方和	自由度	均方	F比
因素A	S_A	$r-1$	$\bar{S}_A=\dfrac{S_A}{r-1}$	$F_A=\dfrac{\bar{S}_A}{\bar{S}_E}$
因素B	S_B	$s-1$	$\bar{S}_B=\dfrac{S_B}{s-1}$	$F_B=\dfrac{\bar{S}_B}{\bar{S}_E}$
误差	S_E	$(r-1)(s-1)$	$\bar{S}_E=\dfrac{S_E}{(r-1)(s-1)}$	
总和	S_T	$rs-1$		

当$H_{01}: \alpha_1 = \alpha_2 = \cdots = \alpha_r = 0$ 为真时，可以证明

$$F_A = \frac{S_A/(r-1)}{S_E/((r-1)\,(s-1))} \sim F(r-1, (r-1)\,(s-1))$$

在显著性水平α下，假设H_{01}的拒绝域为：

$$F_A \geq F_\alpha(r-1, (r-1)(s-1))$$

类似地，

$$F_B = \frac{S_B/(s-1)}{S_E/((r-1)\,(s-1))} \sim F(s-1, (r-1)\,(s-1))$$

$H_{02}: \beta_1 = \beta_2 = \cdots = \beta_s = 0$ 的拒绝域为：

$$F_B = \frac{S_B/(s-1)}{S_E/((r-1)\,(s-1))} \geq F_\alpha(s-1, (r-1)(s-1))$$

方差分析的作用：

1. 通过对试验数据的统计分析，推断造成试验数据间差异的原因是试验水平差异还是随机误差的影响。
2. 推断哪些因素的影响是显著的。
3. 分析出最佳的试验水平（固定模型），或估计总体变量的参数（随机模型）。

方差分析与假设检验的区别：

方差分析能同时检验多个总体的某个参数（如均值）是否相等，而假设检验每次只能检验两个总体的某个参数是否相等。

*正交试验设计

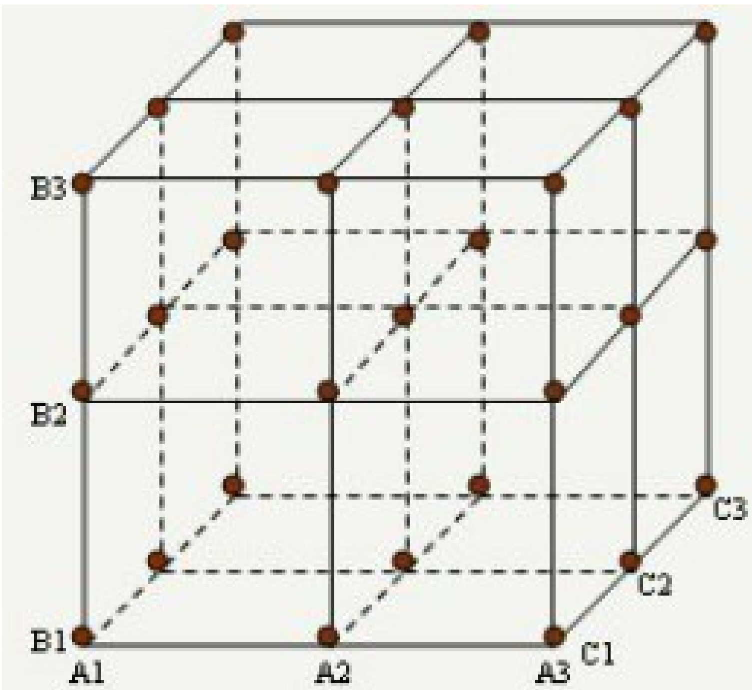
正交试验设计：研究和处理多因素试验的一种方法，利用正交表安排试验，通过少量试验获得满意的试验结果。

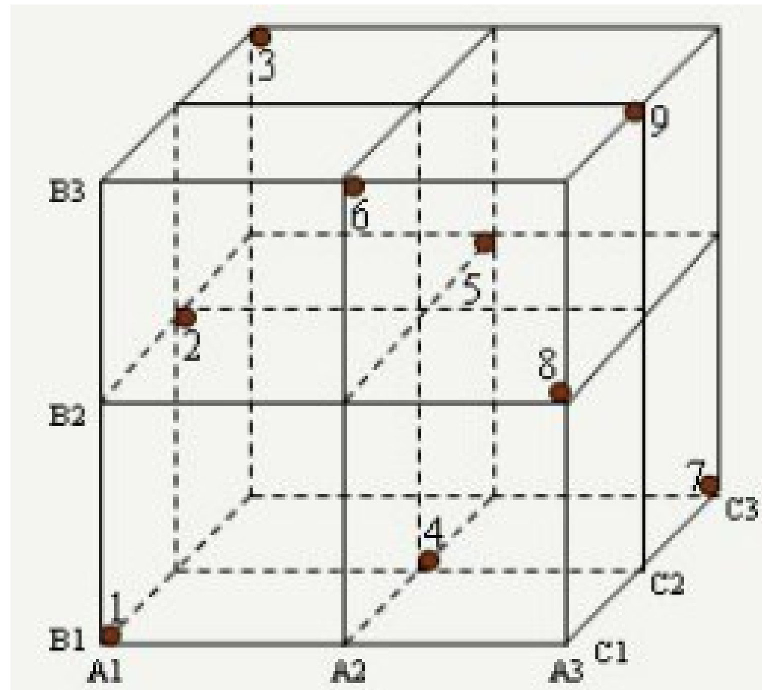
基本方法：

(1) 安排试验方案

(2) 分析试验结果

全面试验法 vs. 正交试验法





- 3个因素×3个水平

正交表：$L_9(3^3)$

编号 \ 因素	A	B	C
1	1	1	1
2	1	2	2
3	1	3	3
4	2	1	2
5	2	2	3
6	2	3	1
7	3	1	3
8	3	2	1
9	3	3	2

表的纵列数（可安排因素的最多个数）

正交表的代号

$$L_9(3^3)$$

表的横行数（需做试验的次数）

表示每一因素的水平个数

正交表的特点：

(1) 每一列中，不同数字出现的次数相同；

(2) 任意两列中，将同一行的两个数字看成有序数对时，每种数对出现的次数相等；

(3) 一般正交表$L_p(n^m)$中，$p=m(n-1)+1$。

正交表的优点：

用正交表安排试验时，各因素的各种水平搭配是均衡的。

正交试验的设计步骤：

(1) 确定试验要考虑的因素，对所做的试验了解少，因素应多取一些，反之可以少取一些；

(2) 确定每个因素的变化水平，重要的因素水平可以多一些；

(3) 根据试验条件、成本等，估算试验次数；

(4) 选取正交表。当试验精度要求高、因素多、交互作用多时，应选择大的正交表；

(5) 进行试验，测定试验指标；

(6) 对试验结果进行计算分析，得出合理结论。

作 业

- 概率论与数理统计

  pp. 265-266, #2, #4, #6